

Modern Aspects of Solid State Chemistry

Modern Aspects of Solid State Chemistry

Edited by

C. N. R. Rao

*Professor of Chemistry
Indian Institute of Technology
Kanpur, India*

 PLENUM PRESS • NEW YORK–LONDON • 1970

**This volume comprises the lectures given at a Winter School
in Solid State Chemistry, supported by the National Council for
Science Education, New Delhi (which operates in cooperation
with the U.S. National Science Foundation and the
Indian University grants commission), held at Kanpur, India,
in November and December of 1969.**

Library of Congress Catalog Card Number 76-133422

SBN 306-30506-2

ISBN-13: 978-1-4684-1877-4 e-ISBN-13: 978-1-4684-1875-0
DOI: 10.1007/978-1-4684-1875-0

© 1970 Plenum Press, New York

softcover reprint of the hardcover 1st edition 1970

A Division of Plenum Publishing Corporation
227 West 17th Street, New York, N.Y. 10011

United Kingdom edition published by Plenum Press, London
A Division of Plenum Publishing Corporation, Ltd.
Donington House, 30 Norfolk Street, London W.C. 2, England

All rights reserved

No part of this publication may be reproduced in any form
without written permission from the publisher

PREFACE

The three natural streams of present-day chemistry are Structure, Dynamics and Synthesis and all these three elements are essential for the study of materials, particularly in the solid state. The solid state provides challenging opportunities for illustrating and applying principles of chemistry to systems of academic interest and technological importance. There are several practising solid state chemists in universities and research laboratories, but the subject has not yet become part of the formal training program in chemistry. Being one of the new frontiers of chemistry, Solid State Chemistry has a tremendous future and undoubtedly demands the active involvement of many more chemists.

A Winter School in Solid State Chemistry was organized at the Indian Institute of Technology, Kanpur, to promote this area and to develop curricular material. Solid State Chemistry being highly interdisciplinary in nature, the lecturers and participants at the Winter School had widely different backgrounds and interests. It was my great desire that the lecture material from the Winter School should become available to a larger body of students, teachers and research workers interested in the solid state and hence this volume.

Some of the areas which have not been covered in this volume are crystallography and crystal chemistry, materials preparation including synthetic routes, organic solids, color centers, spectroscopy and defect chemistry associated with impurity ions doped in host crystals, and the role of solid state chemistry in developing devices. There were lectures on some of these topics at the Winter School, but unfortunately we could not obtain these contributions in time for inclusion in this volume. If the material on some of these topics is taken from other sources, I have a feeling that the present volume can provide much of the basic material required for a course in Solid State Chemistry.

I have tried to edit the chapters and link them up wherever possible. In spite of this effort, the readers will undoubtedly notice differences in language, symbols and notations as well as the breadth and depth of coverage among the chapters. I feel that these shortcomings may not be very serious since each chapter is a unit by itself. I sincerely hope that this volume will find response from all those interested in the study of the solid state.

Kanpur, India
May 7, 1970

C.N.R. RAO

CONTRIBUTORS TO THE VOLUME*

ANDERSON J. S., University of Oxford, Oxford
BHIDE V. G., National Physical Laboratory, Delhi
CHAKRAVORTY D., Indian Institute of Technology, Kanpur
HONIG J. M., Purdue University, Lafayette, Indiana
KHAN D. C., Indian Institute of Technology, Kanpur
MAHANTY J., Indian Institute of Technology, Kanpur
MENYUK N., Lincoln Laboratory, M. I. T., Lexington, Mass.
NARASIMHAN P. T., Indian Institute of Technology, Kanpur
PARASNIS A. S., Indian Institute of Technology, Kanpur
RAMAKRISHNA V., Indian Institute of Technology, Delhi
RAMASAstry C., Indian Institute of Technology, Madras
RAO C. N. R., Indian Institute of Technology, Kanpur
RAY R. K., Indian Institute of Technology, Kanpur
SASTRI M. V. C., Indian Institute of Technology, Madras
SHARAN R., Indian Institute of Technology, Kanpur
SINGRU R. M., Indian Institute of Technology, Kanpur
SUBBARAO E. C., Indian Institute of Technology, Kanpur
VIJAYARAGHAVAN R., Tata Institute of Fundamental Research, Bombay
YOGANARASIMHAN S. R., Indian Institute of Technology, Delhi

* Only Winter School Lecturers have been listed.

ACKNOWLEDGEMENTS

The Winter School in Solid State Chemistry, held at the Indian Institute of Technology, Kanpur, between November 25 and December 15, 1969, was supported and sponsored by the National Council for Science Education (NCSE), New Delhi. I am most grateful to Professor D.S. Kothari, Chairman, University Grants Commission and NCSE, Dr. R.D. Deshpande, Chief Executive of the NCSE and members of the NCSE Chemistry Panel, particularly Professor T.R. Seshadri and Prof. N.V. Subba Rao, for their generous support. But for the grant from the NCSE it would not have been possible to support the travel and stay of a majority of the participants and lecturers.

The travel of the visiting professors from abroad, Professors J.S. Anderson, J.M. Honig and N. Menyuk was supported by the Kanpur Indo-American Program and I am highly thankful to Mr. Gilbert Oakley, Jr., the Leader of the Program. The Winter School would not have been the same without these visiting professors.

The Winter School received considerable support from Dr. P.K. Kelkar, Director, and Dr. M.S. Muthana, Deputy Director, of this Institute, to whom my thanks are due. The cooperation and assistance of the Department of Chemistry of this Institute, in particular from the Department Head, Professor M.V. George, are gratefully acknowledged.

I would like to express my sincere thanks to all the lecturers at the Winter School for their enthusiastic collaboration; most of them have put in considerable effort in writing the chapters in this volume. Some of the lecturers who could not contribute to this volume, but contributed to the Winter School significantly are Professor S. Ramaseshan and Dr. A.K. Singh of the National Aeronautical Laboratory, Bangalore, Dr. J. Shankar and Dr. M.D. Karkhanavala of the Bhabha Atomic Research Centre, Bombay, Dr. A.K. Sridhar of the Solid State Physics Laboratory, New Delhi, and Dr. T.M. Srinivasan and Dr. M.M. Chaudhri of this Institute.

The Winter School was lively and had high intellectual content. There was an average of 100 participants in a lecture. I am thankful to the participants from the various universities and research laboratories of this country for their participation.

The members of my research group, particularly Mr. M. Natarajan, Mr. S. Ramdas and Dr. G.V. Subba Rao, have assisted me greatly in organizing the Winter School and in bringing out this volume; to these gentlemen, I offer my thanks. I am thankful to Mr. M. Natu who has provided considerable help in the office of the Winter School and in the preparation of the manuscript and to Messrs. N. Ahmad, R.D. Tripathi and R.D. Singh who have devoted several hours in typing the manuscripts.

I am most thankful to Mr. Robert Ubell and the Plenum Publishing Corporation, New York, for their enthusiastic cooperation in bringing out this volume.

Finally, I would like to thank my wife, Indu, and my children, Suchitra and Sanjay Srinivas, who have sacrificed considerably in allowing me to fully devote my time and efforts to my profession.

CONTENTS

	Page No.
1. LATTICE DYNAMICS by <u>J. Mahanty</u>	1
Introduction	
Theory of Small Oscillations	
Effect of Lattice Symmetry	
Density of Vibrational States	
Vibrational Thermodynamic Properties of Crystals	
X-ray and Neutron Scattering	
2. COHESIVE ENERGIES OF IONIC SOLIDS by <u>C.N.R. Rao</u>	17
Born-Haber Cycle	
Born Treatment of Ionic Solids	
Some Related Aspects	
3. DEFECT CHEMISTRY AND NON-STOICHIOMETRIC COMPOUNDS by <u>J.S. Anderson</u>	29
Point Defect Equilibria in Crystals	
The Systematics of Non-Stoichiometric Compounds	
Grossly Non-Stoichiometric Compounds	
Ferrous Oxide: A Case History	
Strong Defect Ordering; Intermediate Superstructure Phases	
Elimination of Defects: Crystallographic Shear	
Imperfections of Ordered Structures (Lattice Imaging)	
Conclusions and Outstanding Problems	
4. THEORY OF POINT DEFECTS IN IONIC CRYSTALS by <u>C.N.R. Rao</u> and <u>S. Ramdas</u>	107
Introduction to Point Defects in Ionic Crystals	
Experimental Studies of Point Defects	
Theoretical Calculations of the Formation Energies of Schottky Defects	
Defect Interaction Energies	
Migration Energies	
Interstitial Defects	
Concluding Remarks	

5. DIELECTRIC PROPERTIES by <u>E.C. Subbarao</u>	129
Definitions and Units	
Mechanisms of Polarization	
Alternating Current Phenomena	
Ferroelectrics	
6. MAGNETISM by <u>N. Menyuk</u>	159
Non-Interacting Atoms	
Magnetic Interactions	
Long-Range Order	
Spiral Configurations	
Magnetic Atoms in a Crystal	
Superexchange Interactions	
Collective d-Electron Model	
7. NEUTRON DIFFRACTION AND SOLID STATE PROPERTIES by <u>D.C. Khan</u>	219
Introduction	
Thermal Neutrons	
Elastic Scattering	
Inelastic Scattering	
Conclusions	
8. MAGNETIC RESONANCE by <u>P.T. Narasimhan</u>	231
Introduction	
Nuclear Moments and Nuclear Magnetic Resonance	
Spin-Lattice Relaxation	
Bloch Equations and Transverse Relaxation	
Experimental Arrangements	
NMR in Non-metallic Solids	
Quadrupolar Splittings in NMR	
NMR in Metals	
Chemical Shift and Spin Coupling in Solids	
Pure Quadrupole Resonance in Solids	
Electron Paramagnetic Resonance	
Concluding Remarks	
9. NUCLEAR MAGNETIC RESONANCE IN SUPERCONDUCTORS AND DILUTE ALLOYS by <u>R. Vijayaraghavan</u>	261
NMR in Superconductors	
NMR in Dilute Alloys	

CONTENTS	xi
10. MAGNETOSTRICTION IN MATERIALS by <u>R.K. Ray</u>	273
Introduction	
Behaviour of Common Materials	
Conclusions	
11. MÖSSBAUER EFFECT AND DYNAMICS OF ATOMIC MOTION IN SOLIDS by <u>V.G. Bhide</u>	287
Resonance Fluorescence	
Mössbauer's Experiment	
Resonance Absorption of Gamma Rays and Dynamics of Atomic Motion	
Experimental Results	
Concluding Remarks	
12. POSITRON ANNIHILATION IN SOLIDS by <u>R.M. Singru</u>	329
Introduction	
Methods and Results	
13. ELEMENTS OF ORDER-DISORDER THEORY AND DIVERSE APPLICATIONS by <u>J.M. Honig</u>	343
Statement of the Problem	
The Zero Order Approximation (Langmuir Isotherm)	
The Fermi-Dirac Distribution Function	
Decomposition of the Lattice into Subfigures	
The Fowler-Guggenheim Adsorption Isotherm	
Thermodynamic Properties of Binary Mixtures in the Bragg-Williams Approximation	
14. DIFFUSION IN SOLIDS by <u>V. Ramakrishna</u>	359
The Diffusion Coefficient	
Intrinsic Diffusion Coefficients	
Self-Diffusion Coefficients	
Experimental Procedures	
15. IONIC CONDUCTIVITY by <u>C. Ramasastry, Y.V.G.S. Murty</u> <u>and B.S.V.S.R. Acharyulu</u>	377
Conduction Mechanisms in Ionic Crystals	
Anisotropy of Conduction in Ionic Crystals	

16. INORGANIC GLASSES by <u>D. Chakravorty</u>	391
Constitution of Glass	
Glass Formation	
Inorganic Glass Forming Systems	
Transport Properties of Glasses	
17. ELECTRICAL PROPERTIES OF SOLID CATALYSTS by <u>M.V.C. Sastri, V. Srinivasan and</u> <u>B. Viswanathan</u>	425
Historical Introduction	
Classification of Solid Catalysts based	
on Electronic Properties	
Inverse Mixed Catalysts	
Conclusions	
18. DISLOCATIONS AND SOLID STATE REACTIONS by <u>A.S. Parannnis</u>	447
Dislocations, Their Structure and Properties	
The Role of Dislocations in Chemical Reactions	
19. SOLID STATE REACTIONS by <u>S.R. Yoganarasimhan</u>	465
Role of Defects in the Reactivity of Solids	
Solid-Gas Reactions: Oxidation of Metals	
Solid-Solid Reactions	
Thermal Decomposition of Solids	
20. BAND AND TRANSPORT THEORIES IN SOLIDS by <u>J.M. Honig</u>	497
Introduction to Band Structure Concepts	
(Particle-in-the-Box)	
Kronig-Penney Model	
Bloch Functions for the Circular Chain	
Band Structure for a Circular Chain in the	
Tight Binding Approximation	
Pauli Exclusion Principle and Fermi-Dirac	
Statistics	
Heat Capacity of a Metal	
Classification of Materials	
Elementary Discussion of Various Transport	
Phenomena	
Fermi Levels and Conductivity in Intrinsic	
and Extrinsic Semiconductors	
Transport Effects in Materials Characterized	
by 'Hopping' Electrons	
Brief Survey of Experimental Results	

CONTENTS	xiii
21. SPECTROSCOPY OF METALS AND SEMICONDUCTORS by <u>R. Sharan</u>	567
Introduction	
Macroscopic Theory and Experimental Methods	
Microscopic Theory of Optical Properties	
Optical Properties of Metals	
Optical Properties of Semiconductors	
Conclusions	
22. PHASE TRANSFORMATIONS IN SOLIDS, by <u>C.N.R. Rao</u>	589
Introduction	
Thermodynamic Considerations	
Structural Changes in Transformations	
Kinetics of Transformations	
Order-Disorder Transitions	
Martensite Transformations	
Ferroelectric Transformations	
Magnetic Transitions	
Semiconductor-Metal Transitions	
High Pressure Transformations	
Born Treatment of Phase Transformations in Alkali Halides	
Comments on Experimental Techniques	

LATTICE DYNAMICS

J. Mahanty

Department of Physics

Indian Institute of Technology, Kanpur, India

I. INTRODUCTION

Lattice dynamics which deals with the small oscillations of atoms in crystalline solids about their equilibrium positions is based on the same concepts as are relevant to the understanding of vibrations of atoms in molecules. The energy levels of a molecule can be approximated by superposing a series of vibrational levels on each electronic level. A suitable theoretical basis is provided by the Born-Oppenheimer approximation, using which one can show that each low-lying electronic energy level corresponds to a certain equilibrium arrangement of the constituent atoms, and small displacements of atoms from these positions generate harmonic restoring forces leading to coupled oscillations of the atoms. These oscillations can be classified in terms of normal modes, and a class of the energy levels of the molecule in the neighbourhood of the above electronic level is obtained by adding to it the energies corresponding to the vibrational quanta associated with the normal modes. The states near the ground state of the molecule are certainly approximated well by the above approach, though for the higher electronic excited states the approach becomes less exact.

One class of the low-lying levels of a crystalline solid can be expressed similarly, in terms of the vibrational quanta associated with the small oscillations of the atoms about their equilibrium positions. The equilibrium positions of the atoms have a significance similar to that of the atoms in a molecule, as discussed above. The main point of difference is that whereas the low-lying electronic energy levels are sharp and corresponding

to each electronic level there are vibrational levels, in the case of solids the electronic levels are not sharp and it is not possible to associate vibrational levels with electronic levels in any simple way. However, it is reasonable to assume that when the atoms in the solid are in their equilibrium positions the entire system is in the state of lowest energy, and a particular group of energy levels lying immediately above this energy is given by the vibrational quanta associated with lattice vibrations. This group of levels is of immense significance in so far as the thermodynamic properties of the solid are concerned. Many other properties associated with the interaction of the solid with external radiation of various kinds (x-rays, neutrons, etc.) are also influenced by lattice vibrations. We shall discuss, in these lectures, some of the essential features of the subject of lattice dynamics.

II. THEORY OF SMALL OSCILLATIONS

Let us consider a system of N interacting atoms. In the ground state they will be in their respective equilibrium positions at R_1, R_2, \dots, R_N . The potential energy of the assembly is a function of the atomic coordinates, of the form $V(R_1, R_2, \dots, R_N)$. This function of $3N$ variables has a minimum (by hypothesis) when $r_1 = R_1, r_2 = R_2, \dots, r_N = R_N$. In terms of the coordinates $u_\alpha(1) = (r_1 - R_1)_\alpha, \alpha = 1, 2, 3$, which are the Cartesian components of the displacement from equilibrium of the 1-th atom, the potential energy can be expanded in a Taylor series,

$$\begin{aligned} V(r_1, r_2, \dots, r_N) &\approx V(R_1, R_2, \dots, R_N) + \frac{1}{2} \sum_{\alpha, \beta; l, l'} \left. \frac{\partial^2 V}{\partial u_\alpha(l) \partial u_\beta(l')} \right|_{u=0} X \\ &\quad u_\alpha(l) u_\beta(l') \\ &\approx V_0 + \frac{1}{2} \sum_{\alpha, \beta; l, l'} A_{\alpha\beta}(l, l') u_\alpha(l) u_\beta(l'); \end{aligned} \quad (1)$$

where

$$A_{\alpha\beta}(l, l') = \left. \frac{\partial^2 V}{\partial u_\alpha(l) \partial u_\beta(l')} \right|_{u=0} \quad (2)$$

The equations of motion, therefore, are given by (with m_l being the mass of the 1-th atom)

$$m_1 \ddot{u}_\alpha(l) = - \frac{\partial V}{\partial u_\alpha(l)} = - \sum_{\beta, l} A_{\alpha\beta}(l, l') u_\beta(l'). \quad (3)$$

$A_{\alpha\beta}(l, l')$ is the effective restoring force in the α -direction on the l -th atom when the l' -th atom is displaced by a unit distance in β -direction. We thus have a set of $3N$ linear equations to solve. These equations are best handled by matrix methods. We define the following matrices

$$\underline{U} = \begin{bmatrix} \vdots \\ \dot{u}_\alpha(l) \\ \vdots \end{bmatrix}, \text{ i.e., } \underline{u}(\alpha, l) \equiv u_\alpha(l), \hat{A} \equiv \begin{bmatrix} A_{\alpha\beta}(l, l') \end{bmatrix}, \hat{M} = \begin{bmatrix} m_1 & & \\ & \ddots & \\ & & m_1 \end{bmatrix} \dots \quad (4)$$

Then (3) can be written as

$$\hat{M} \ddot{\underline{U}} = - \hat{A} \underline{U}. \quad (5)$$

(5) can be written in a more symmetric form by defining

$$\underline{U} = \hat{M}^{-\frac{1}{2}} \underline{Y} \quad (6)$$

when (5) becomes

$$\ddot{\underline{Y}} = - \hat{M}^{-\frac{1}{2}} \hat{A} \hat{M}^{-\frac{1}{2}} \underline{Y} = - \hat{B} \underline{Y} \quad (7)$$

where

$$\hat{B} = \hat{M}^{-\frac{1}{2}} \hat{A} \hat{M}^{-\frac{1}{2}}. \quad (8)$$

\hat{A} and \hat{B} matrices have several interesting properties. \hat{A} is a symmetric matrix and so is B . Furthermore, if all the atomic displacements in the β -direction are equal, the effect would be a translation of the system leading to no restoring force. Hence

$$\sum_{l'} A_{\alpha\beta}(l, l') = 0. \quad (9)$$

Other rules obeyed by the matrix elements of \hat{A} can also be established.

Eqn. (7) can be solved by a substitution

$$\underline{Y} = e^{\pm i\omega t} \underline{X} \quad (10)$$

Then, (7) becomes

$$\hat{B} \underline{X} = \omega^2 \underline{X} \quad \text{or} \quad [\hat{B} - \hat{I} \omega^2] \underline{X} = 0. \quad (11)$$

Thus the allowed frequencies (squared) are the eigenvalues of \hat{B} and for each frequency the appropriate \underline{X} is the corresponding eigenvector. These are called the normal mode frequencies of the system. To obtain the frequencies one has to solve the secular equation,

$$|\hat{B} - \hat{I} \omega^2| = 0 \quad (12)$$

which is a polynomial of $3N$ -th degree in ω^2 , and so we have $3N$ normal mode frequencies.

A general state of the system can be expressed as a superposition of normal modes,

$$\underline{U}(t) = \sum_{\mu=1}^{3N} M^{-\frac{1}{2}} \underline{X}_{\mu} \left\{ a_{\mu} e^{i\omega_{\mu} t} + b_{\mu} e^{-i\omega_{\mu} t} \right\}. \quad (13)$$

Even though each normal mode is a sinusoidal function of t , a linear combination of them need not have a simple time-dependence.

III. EFFECT OF LATTICE SYMMETRY

The main physical characteristics of a crystal lattice are associated with the symmetry in the arrangement of the atoms. This symmetry reflects itself in the structure of the A and B matrices. We shall examine this point taking the simple example of a linear chain of masses and springs, and show that the A and B matrices that occur here have such structural features that their eigenvectors and eigenvalues can be found easily.

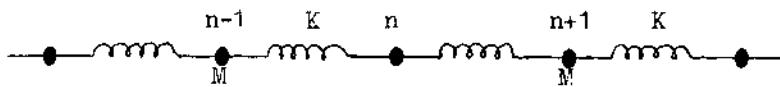


Fig. 1

Let us consider a chain of N particles connected by springs (as in Fig. 1), each particle being of mass M , and the spring constant of each spring is K . We shall arrange such that the last particle is connected to the first particle by a spring of the same type, so that if u_n is the displacement from equilibrium of the n -th particle,

$$u_n = u_{n+N} \text{ for all } n. \quad (14)$$

Eqn. (14) represents what is known as the periodic boundary

condition on particle displacements. It turns out that for large N the exact nature of the boundary condition is not very important as far as the distribution of the values of ω is concerned. But use of the periodic boundary condition renders the mathematics of lattice dynamics somewhat more manageable without loss of generality.

One can write the equations of motion of the particles as,

$$\begin{aligned} M \ddot{u}_n &= K(u_{n+1} - u_n) + K(u_{n-1} - u_n) \\ n = 1, 2, \dots, N \quad (\text{with } u_{N+1} \equiv u_1 \text{ and } u_0 \equiv u_N \\ &\quad \text{as a result of the periodic boundary condition}). \end{aligned} \quad (15)$$

The N equations represented by (15) can be put in the form of (11) where

$$B = \begin{bmatrix} 1 & 2 & 3 & 4 & \dots & N-1 & N \\ 2K/M & -K/M & 0 & 0 & \dots & 0 & -K/M \\ -K/M & 2K/M & -K/M & 0 & \dots & 0 & 0 \\ 0 & -K/M & 2K/M & -K/M & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ N & -K/M & 0 & 0 & 0 & \dots & -K/M & 2K/M \end{bmatrix} \quad (16)$$

B is a typical example of a cyclic matrix, which has the feature that each row is obtained from the previous row by translating everything by one step to the right. One of the properties of a $N \times N$ cyclic matrix is that the n -th component of its k -th eigenvector is given by (when properly normalized),

$$\phi_k(n) = \frac{1}{\sqrt{N}} e^{ikn} \quad . \quad (17)$$

We can thus choose x_n , the n -th component of \underline{x} in (11) in the form

$$x_n = e^{ikn} v \quad . \quad (18)$$

The periodic boundary condition of u_n implies the same condition on x_n , so that

$$x_{n+N} \equiv x_n \quad (19)$$

or

$$e^{ikN} = 1$$

or

$$k = \frac{2\pi p}{N}, \quad p = 1, 2, 3, \dots, N. \quad (20)$$

Obviously, if an integral value of p exceeding N is taken, it will reproduce the form of x_n for $p' < N$, where $(p-p')$ is a multiple of N . Hence the useful set of k values is given by (20).

A more general choice of x_n would be of the form

$$x_n = \sum_k a_k \phi_k(n) \quad (21)$$

where the summation is over the allowed values of k as given by (20). a_k is the coefficient representing the amplitude of the normal mode characterised by the index k that occurs in the particular state of the chain in which x_n is the displacement of the n -th particle. Knowing x_n for all n and using the orthogonality property of ϕ_k for different k , we get the relation

$$a_k = \frac{1}{\sqrt{N}} \sum_n x_n e^{ikn}. \quad (22)$$

The eigenvalue $\omega^2(k)$ of \hat{B} corresponding to the eigenvector ϕ_k is obtained trivially by noting that

$$\begin{aligned} (\hat{B} \phi_k)_n &= \sum_n B_{nn} \phi_k(n) \\ &= \frac{1}{\sqrt{N}} \left[\frac{2K}{M} e^{ikn} - \frac{K}{M} e^{ik(n+1)} - \frac{K}{M} e^{ik(n-1)} \right] \\ &= \frac{2K}{M} (1 - \cos k) \phi_k(n) . \end{aligned} \quad (23)$$

$$\text{Thus, } \hat{B} \phi_k = \frac{2K}{M} (1 - \cos k) \phi_k$$

$$\text{and hence, } \omega^2(k) = \frac{2K}{M} (1 - \cos k) . \quad (24)$$

$$\text{Also, } \omega(k) = 2\sqrt{\frac{K}{M}} \sin(k/2) . \quad (25)$$

Using (13) we can write the most general state of the chain in the form

$$U(t) = \sum_k \left[a_k \phi_k e^{i\omega(k)t} + b_k \phi_k e^{-i\omega(k)t} \right] \quad (26)$$

where the constants a_k and b_k can be determined by suitable initial conditions on $\underline{U}(0)$ and $\dot{\underline{U}}(0)$.

This example serves to illustrate the great simplification in the mathematics of (11) using the symmetry properties of \hat{B} matrix, which in their turn are related to the symmetry of the spatial arrangement of the particles. This is entirely analogous to the way one classifies the molecular normal modes in terms of the irreducible representations of the point group of the molecule - but we shall not dwell on the group theoretic aspects here.

The effect of lattice symmetry is felt in two respects. Firstly, the normal modes assume the character of propagating waves, each characterized by a wave number k . Secondly, the normal mode frequencies are functions of k . The results obtained for the linear chain can be generalized to the case of realistic three-dimensional lattices as follows.

Let us take a crystal with N^3 unit cells, in each of which there are p atoms, the mass of the s -th atom being M_s . The analogue of (3) would then be,

$$M_s \ddot{u}_\alpha (\underline{l}, s) = - \sum_{\beta, s', \underline{l}'} A_{\alpha\beta} (\underline{l}, s; \underline{l}', s') u_\beta (\underline{l}', s') \quad (27)$$

where $u_\alpha (\underline{l}, s)$ is the α -component of the displacement of the s -th atom in the \underline{l} -th cell (\underline{l} has to be a vector now with integer components). The periodic boundary condition in this case would imply,

$$u_\alpha (\underline{l}, s) \equiv u_\alpha (l_1, l_2, l_3; s) \equiv u_\alpha (l_1 + N, l_2 + N, l_3 + N; s) \quad (28)$$

for all α and s , $\{l_i\}$ being the components of \underline{l} . The matrix \hat{A} , whose elements are $A_{\alpha\beta} (\underline{l}, s; \underline{l}', s')$, is cyclic with respect to the indices \underline{l} and \underline{l}' , i.e., its elements depend on $|\underline{l} - \underline{l}'|$. We shall write them as $A_{\alpha\beta} (|\underline{l} - \underline{l}'|; s; s')$. The arguments preceding (18) are applicable in this case also, so that we can write.

$$u_\alpha (\underline{l}, s) = \frac{y_\alpha (\underline{l}, s)}{\sqrt{M_s}} ; \quad y_\alpha (\underline{l}, s) = x_\alpha (\underline{l}, s) e^{+i\omega t}$$

and

$$x_\alpha (\underline{l}, s) = e^{i\underline{k} \cdot \underline{l}} v_\alpha (s) \quad (29)$$

where \underline{k} is a vector having components k_1, k_2, k_3 , each satisfying (20). In (29) the transformation of (6) has been used explicitly. Eqn. (27) then becomes,

$$\omega^2 v_\alpha(s) = \sum_{\beta, s'} D_{\alpha\beta}^{ss'}(\underline{k}) v_\beta(s') \quad (30)$$

where

$$D_{\alpha\beta}^{ss'}(\underline{k}) = \sum_l e^{i\underline{k} \cdot \underline{l}} \frac{A_{\alpha\beta}(l; s, s')}{\sqrt{\frac{M}{s} \frac{M}{s'}}}. \quad (31)$$

The secular equation corresponding to (30) is,

$$| D_{\alpha\beta}^{ss'}(\underline{k}) - \omega^2 \delta_{\alpha\beta} \delta_{ss'} | = 0. \quad (32)$$

The dimension of the matrix $\hat{D}(\underline{k})$ having elements $D_{\alpha\beta}^{ss'}(\underline{k})$ is $3p$, so that for each value of \underline{k} , (32) would give $3p$ values of $\omega^2(\underline{k})$. We represent these by an index r , so that $\omega_r(\underline{k})$ is the r -th root of (32), $r = 1, 2, \dots, 3p$. Since \underline{k} itself can have N^3 values (each of its components having N values as given in (20)), the total number of frequencies is $3N^3p$, though they are now broken up into $3p$ branches.

It can be shown that the frequencies in three of the $3p$ branches start with the value $\omega = 0$ and the rest never become zero. These three branches are called the acoustic branches of the crystal. The reason for this nomenclature is that in these branches the relationship between ω and \underline{k} (i.e., $\omega(\underline{k})$) is linear for small \underline{k} , which is a feature of acoustic waves in elastic media.

The structure of the normal modes now is naturally more complicated, but we can write the $(l; \alpha, s)$ -component of the (\underline{k}, r) -th normal mode as

$$\phi_{\underline{k}, r}(l; \alpha, s) = e^{i\underline{k} \cdot \underline{l}} Q_r^{\underline{k}}(\alpha, s) \quad (33)$$

where $Q^{\underline{k}}$ having components $Q_r(\alpha, s)$ is an eigenvector of \hat{D} as defined in (31). One can then write the analogue of (13) using $\phi_{\underline{k}, r}$ for any general vibrational state of the lattice.

IV. DENSITY OF VIBRATIONAL STATES

$\omega_r(\underline{k})$ is a function of the discrete variable \underline{k} , whose components are given by (20). Because of the occurrence of N in the denominator in (20), for large N the values of \underline{k} become densely spaced, and one can regard $\omega_r(\underline{k})$ as almost a continuous function of \underline{k} defined in the range of values,

$$0 \leq k_i \leq 2\pi, i = 1, 2, 3. \quad (34)$$

A simple geometrical representation of the allowed values of \underline{k}

can be in the form of the vectors that one can get by joining the origin to the lattice points formed by filling a cube of side length 2π with small cubes of side $2\pi/N$. The density of allowed \underline{k} -values inside the cube of length 2π is evidently,

$$\sigma(\underline{k}) = \frac{N^3}{(2\pi)^3} . \quad (35)$$

In fact, one could construct an entire "reciprocal lattice" of lattice points formed by cubes of side length 2π , and any point in this space would be an allowed value of \underline{k} (in the limit of $N \rightarrow \infty$). However, one can always subtract appropriate number of integral multiples of 2π from each component of \underline{k} to get a vector which lies entirely inside the first cell of volume $(2\pi)^3$ - and since \underline{k} occurs in the various equations in the form of $\exp(i\underline{k} \cdot \underline{l})$, this vector within the first cell will be completely equivalent to the original \underline{k} . Thus we need to specify the N^3 useful values of \underline{k} within the first cell only.

In the picture here we have treated both \underline{l} and \underline{k} as dimensionless vectors. It is more usual to treat \underline{l} as a lattice vector having the dimension of length and consequently \underline{k} assumes the dimension of reciprocal of length. Furthermore, since \underline{k} and $\underline{k} + \underline{g}$ (where \underline{g} is a "reciprocal lattice vector") are equivalent, one can, by subtracting suitable reciprocal lattice vectors from \underline{k} , define all values of \underline{k} in a symmetric region surrounding the origin in reciprocal lattice - called the first Brillouin zone, or simply, the Brillouin zone. These details can be seen in any text book^{2,3}, but we shall stick to the present convention for simplicity.

Since $\omega_r(\underline{k})$ is an almost continuous function of \underline{k} , it would be appropriate to talk in terms of a "Frequency distribution function" $g_r(\omega)$ defined by,

$$g_r(\omega) d\omega \equiv \text{Number of frequencies in the } r\text{-th branch lying in the interval } \omega \text{ and } \omega + d\omega . \quad (36)$$

Any additive function of the normal mode frequencies, such as the thermodynamic properties arising out of lattice vibrations, can thus be expressed as,

$$A \equiv \sum_{\omega} f(\omega) \equiv \sum_r \int_0^{\infty} g_r(\omega) f(\omega) d\omega . \quad (37)$$

The problem of evaluating $g_r(\omega)$ knowing $\omega_r(\underline{k})$ as a function of \underline{k} can be tackled as follows. We shall discuss the problem in terms of evaluation of $G_r(x)$, where $x = \omega^2$, since $\omega_r^2(\underline{k})$ is directly obtained from the secular equation. $G_r(x)$ is related to $g_r(\omega)$ by the relation

$$g_r(\omega) = 2\omega G_r(\omega^2) = 2\omega G_r(x). \quad (38)$$

Let

$$G_r(x) = \int_{-\infty}^{+\infty} e^{ixt} F_r(t) dt \quad (39)$$

where the characteristic function $F_r(t)$ is given by,

$$\begin{aligned} F_r(t) &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-ixt} G_r(x) dx \\ &= \frac{1}{2\pi} \sum_{\underline{k}} e^{-it\omega_r^2(\underline{k})} \end{aligned}$$

Since the allowed values of \underline{k} are dense, the summation over \underline{k} can be replaced by integration using the density $\rho(\underline{k})$ of (35), i.e.,

$$\begin{aligned} F_r(t) &= \frac{1}{2\pi} \int d^3k \rho(\underline{k}) e^{-it\omega_r^2(\underline{k})} \\ &= \frac{N^3}{(2\pi)^4} \iiint_0^{2\pi} dk_1 dk_2 dk_3 e^{-it\omega_r^2(\underline{k})} \end{aligned} \quad (40)$$

Once the characteristic function $F_r(t)$ is determined using (40), (39) and (38) would give us the frequency distribution function, which is in fact a measure of the density of the vibrational states of the crystal arising out of the vibration of the r -th branch.

Actual computation of frequency distribution function is almost never done by the method given above because of computational difficulties, and many simpler computational methods have been developed¹. But certain model-independent features of $G_r(\omega^2)$ and hence of $g_r(\omega)$ can be established by use of (38), (39) and (40). The most important features in this respect are the critical points (or Van Hove singularities, after Van Hove who first pointed them out) which occur in $G_r(\omega^2)$ at the maxima, minima and saddlepoints of $\omega_r^2(\underline{k})$ as a function of \underline{k} . The behaviour of $G_r(\omega^2)$ in the neighbourhood of the critical points can be established by examining the asymptotic form of $F_r(t)$ for large t , which is obtained from (40) by the method of stationary phase. To illustrate the method, we will take the example of a three dimensional lattice and consider the behaviour of $G_r(\omega^2)$ near ω_{\min}^2 . In that neighbourhood, assuming that the minimum occurs at $\underline{k} = \underline{k}_0$

$$\omega_r^2(\underline{k}) = \omega_{\min}^2 + \frac{1}{2} \sum_{i,j} \frac{\partial^2 \omega_r^2(\underline{k})}{\partial k_i \partial k_j} |_{\underline{k}_0} \times (\underline{k} - \underline{k}_0)_i (\underline{k} - \underline{k}_0)_j \quad (41)$$

Substituting this in (40) we get the contribution to the asymptotic

form of $F_r(t)$ from the minimum of $\omega_r^2(\underline{k})$ in the form

$$F_r(t) \propto t^{-3/2} e^{-it\omega_{\min}^2},$$

and substituting this in (39) we obtain

$$\begin{aligned} G_r(\omega^2) &\propto \int \frac{e^{i(x-\omega_{\min}^2)t}}{t^{3/2}} dt \\ &\propto (\omega^2 - \omega_{\min}^2)^{\frac{1}{2}}. \end{aligned} \quad (42)$$

For the acoustic branches, since $\omega_{\min}^2 = 0$, $G_r(\omega^2) \propto \omega$ and using (38) we get

$$g(\omega) \propto \omega^2. \quad (42a)$$

More detailed discussion on these aspects, i.e., the behaviour of $g(\omega)$ at other critical points can be seen in literature¹.

V. VIBRATIONAL THERMODYNAMIC PROPERTIES OF CRYSTALS

Having evaluated the vibrational frequencies, we can express the low-lying vibrational energy states as arising out of an approximate Hamiltonian,

$$H \approx \sum_{r,\underline{k}} \hbar \omega_r(\underline{k}) \{ a_r^+(\underline{k}) a_r(\underline{k}) + \frac{1}{2} \}.$$

The operators $a_r^+(\underline{k})$ and $a_r(\underline{k})$, the "creation" and "annihilation" operators of the vibrational quanta, are related to the normal coordinates through the following equations:

$$a_r(\underline{k}) = \frac{1}{\sqrt{2\hbar\omega_r(\underline{k})}} [T_r(\underline{k}) + i\omega_r(\underline{k}) \dot{T}_r(\underline{k})] \quad (43a)$$

$$a_r^+(\underline{k}) = \frac{1}{\sqrt{2\hbar\omega_r(\underline{k})}} [T_r(\underline{k}) - i\omega_r(\underline{k}) \dot{T}_r(\underline{k})] \quad (43b)$$

and $T_r(\underline{k})$ is defined by an analogue of (13),

$$U = \sum_{\underline{k},r} T_r(\underline{k},t) \hat{M}^{-\frac{1}{2}} \phi_{\underline{k},r}. \quad (44)$$

In (44) the time dependence of each normal mode $T_r(\underline{k})$ is of the

form $e^{\pm i\omega_r(\underline{k})t}$, and is implicitly contained in $T_r(\underline{k}, t)$.

The eigenvalues of the "number operator" $[a_r^+(\underline{k}) a_r(\underline{k})]$ are positive integers, so that the general expression for the vibrational energy of the system is

$$E \left(\{n_r(\underline{k})\} \right) = \sum_{r,\underline{k}} \hbar \omega_r(\underline{k}) \left[n_r(\underline{k}) + \frac{1}{2} \right] \quad (45)$$

where the positive integer $n_r(\underline{k})$ denotes the state of excitation of the normal mode vibration corresponding to the indices r and \underline{k} .

The partition function of the crystal would be

$$\begin{aligned} Z &= \prod_{r,\underline{k}} \sum_{n_r(\underline{k})=0}^{\infty} e^{-\hbar \omega_r(\underline{k}) \left[n_r(\underline{k}) + \frac{1}{2} \right] / k_B T} \\ &= \prod_{r,\underline{k}} \frac{1}{2 \sinh (\hbar \omega_r(\underline{k}) / 2k_B T)}. \end{aligned}$$

The internal energy is given by,

$$\bar{E} = - \frac{\partial}{\partial (1/k_B T)} \ln Z = \sum_{r,\underline{k}} \frac{\hbar \omega_r(\underline{k})}{2} \coth \left(\frac{\hbar \omega_r(\underline{k})}{2k_B T} \right) \quad (46)$$

Other thermodynamic properties can also be expressed as additive functions of the normal mode frequencies, in a form analogous to (46). For all such functions (37) is applicable; for instance, (46) can be written in the form

$$\begin{aligned} \bar{E} &= \sum_r \int_0^{\infty} g_r(\omega) \frac{\hbar \omega}{2} \coth \left(\frac{\hbar \omega}{2k_B T} \right) d\omega \\ &= \sum_r \int_0^{\infty} g_r(\omega) \left[\frac{\hbar \omega}{2} + \frac{\hbar \omega}{e^{\hbar \omega / k_B T} - 1} \right] d\omega \quad (47) \end{aligned}$$

At low temperatures the contribution to the above integral will be primarily from the region $\omega \sim 0$. In this neighbourhood only the acoustic branches matter, and for them, as indicated in (42a), $g_r(\omega) \propto \omega^2$.

Hence, the temperature-dependent part of the internal energy is,

$$\bar{E}(T)_{T \text{ Small}} \propto \int_0^{\omega_{\max}} \omega^3 \left(e^{\hbar \omega / k_B T} - 1 \right)^{-1} d\omega$$

$$\propto \left[\frac{k_B T}{\hbar} \right]^4 \int_0^\infty x^3 (e^x - 1)^{-1} dx$$

where $x = (\hbar \omega)/(k_B T)$. (The upper limit of the integral can be made ∞ at low temperatures without significant error.) Hence,

$$\bar{E}(T)_{T \text{ Small}} \propto T^4 \quad (48)$$

and the specific heat C_v would then be

$$C_v(T)_{T \text{ Small}} = \frac{\partial}{\partial T} \bar{E}(T)_{T \text{ Small}} \propto T^3. \quad (49)$$

This is the celebrated Debye T^3 - law of specific heats of solids at low temperatures.

The above analysis is just to emphasize that we have, through lattice dynamics, a reasonable framework for understanding the vibrational thermodynamic properties of crystals.

VI. X-RAY AND NEUTRON SCATTERING

We shall discuss here in somewhat general terms the effect of lattice vibrations on scattering of x-rays or neutrons from crystals. This scattering is due to a potential $V(\underline{r})$ that the incoming neutron sees in the crystal. One can write

$$V(\underline{r}) = \sum_{\underline{l}} v_a(\underline{r} - \underline{R}_{\underline{l}}) \quad (50)$$

where $v_a(\underline{r} - \underline{R}_{\underline{l}})$ is the potential that the incoming neutron experiences due to the crystal atom located at the position $\underline{R}_{\underline{l}}$. For a perfect crystal at rest

$$\underline{R}_{\underline{l}} = \underline{l} \cdot \quad (51)$$

But for a vibrating crystal,

$$\underline{R}_{\underline{l}} = \underline{l} + \underline{u}_{\underline{l}} \cdot \quad (52)$$

In the perfect crystal at rest,

$$V(\underline{r}) = V(\underline{r} + \underline{l}), \quad (53)$$

i.e., the potential will have the periodicity of the lattice. Such a potential can be expanded in a Fourier series,

$$V(\underline{r}) = \sum_{\underline{g}} e^{i\underline{g} \cdot \underline{r}} a_{\underline{g}} \quad (54)$$

where \underline{g} is a reciprocal lattice vector, as mentioned in Section IV.

The scattering process can be regarded as a transition from the initial state to a final state, each of which is a product of the plane wave state of the incoming or outgoing particle and a lattice vibrational state obtained by specifying the quantum number $n_r(\underline{k})$ for all r and \underline{k} . Thus,

$$|\Psi_i\rangle = |\underline{p}_i; \{n_r^i(\underline{k})\}\rangle \propto e^{i\frac{\underline{p}_i}{\hbar} \cdot \underline{r}} |\{n_r^i(\underline{k})\}\rangle \quad (55)$$

$$|\Psi_f\rangle = |\underline{p}_f; \{n_r^f(\underline{k})\}\rangle \propto e^{i\frac{\underline{p}_f}{\hbar} \cdot \underline{r}} |\{n_r^f(\underline{k})\}\rangle \quad (56)$$

where \underline{p}_i and \underline{p}_f are the incoming and outgoing momentum of the neutron. The scattering amplitude would be given by,

$$S(\underline{p}_i, \underline{p}_f) \propto \langle \Psi_f | V | \Psi_i \rangle \quad (57)$$

For the perfect crystal at rest, one can use (54) for V to obtain, firstly, the fact that the lattice part of the initial and final states must be the same and secondly,

$$\begin{aligned} S(\underline{p}_i, \underline{p}_f) &\propto \int e^{i(\frac{\underline{p}_i - \underline{p}_f}{\hbar}) \cdot \underline{r} + i\underline{g} \cdot \underline{r}} d^3r \\ &\propto \delta\left(\frac{\underline{p}_i - \underline{p}_f}{\hbar} + \underline{g}\right). \end{aligned} \quad (58)$$

That is, the scattering amplitude will be non-zero only if

$$\underline{p}_i - \underline{p}_f = \underline{g} \hbar. \quad (59)$$

This is the so-called elastic scattering process; (59) is the same as Bragg's law of reflection.

Where there are lattice vibrations, using (52) in (50) we get,

$$V(\underline{r}) \approx \sum_{\underline{l}} v_a(\underline{r} - \underline{l}) + \sum_{\underline{l}} u_{\underline{l}} \cdot \nabla v_a(\underline{r} - \underline{l}) + O(u_{\underline{l}}^2) \quad (60)$$

The contribution of the first term to the scattering amplitude is essentially the same as (58). The second term is linear in $u_{\pm 1}$, and using (44) one can express it linearly in T_r and hence linearly in terms of $a_r(\underline{k})$ and $a_r^+(\underline{k})$. As is well known, these operators have nonvanishing matrix elements between vibrational states which differ in quantum number by ± 1 . The second term thus leads to nonvanishing matrix element between the final and initial states only if,

$$\underline{p}_i - \underline{p}_f = \underline{g} \cdot \underline{n} \pm \underline{k} \quad (61)$$

where \underline{k} is the phonon wave vector. Scattering characterised by the selection rule of (61) is obviously inelastic scattering with emission or absorption of one phonon in the process.

This short discussion is just to illustrate how knowledge of the vibrational states of the crystal helps in understanding its physical properties.

REFERENCES

1. Maradudin, A.A., Montroll, E.W., and Weiss, G.H., Theory of Lattice Dynamics in the Harmonic Approximation, Solid State Physics (Suppl. 3), Academic Press, New York, 1963.
2. Ziman, J., Principles of Theory of Solids, Cambridge University Press, 1964.
3. Kittel, C., Introduction to Solid State Physics (3rd Edition), John Wiley, New York, 1966.

COHESIVE ENERGIES OF IONIC SOLIDS

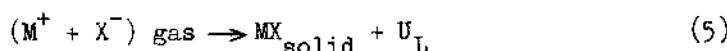
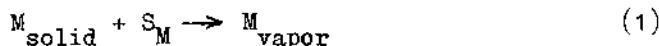
C. N. R. Rao

Department of Chemistry

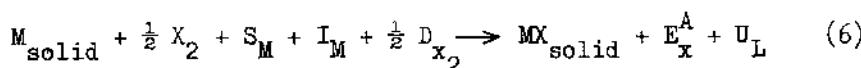
Indian Institute of Technology, Kanpur, India

I. BORN-HABER CYCLE

The cohesive energy of an ionic solid is the energy required to separate a gram molecule of ions at equilibrium distances in a solid to infinite distances apart. This energy can be evaluated from thermochemical data employing the Born-Haber Cycle. Let us consider the series of reactions that would give the heat of formation of an ionic solid MX



Adding the above,



If the heat of sublimation S_M , the ionization potential I_M , the

dissociation energy D_{X_2} , electron affinity E_x^A and the heat of formation ΔH_f are known experimentally, the cohesive energy, U_L , per an ion pair of MX can be calculated

$$-U_L = S_M + I_M + 1/2D_{X_2} - E_x^A - \Delta H_f \quad (8)$$

Note that all the energy quantities in 1 to 7 are per atom or molecule. The experimental values of the energy quantities for a few ionic solids are listed in Table 1.

Table 1

Born-Haber Cycle Data (in Kcal/mole) for Some Alkali Halides

	$-\Delta H_f$ (a)	I_M (b)	S_M (a)	$1/2D_{X_2}$ (a)	$-E_x^A$ (b)	$-U_L$ (a)(c)
LiF	145.7	124.4	38.4	18.9	82.1	242.3
LiCl	96.0	124.4	38.4	28.9	85.8	198.9
NaCl	98.2	118.4	25.9	28.9	85.8	182.6
NaBr	86.0	118.4	25.9	26.8	80.5	173.6
NaI	68.8	118.4	25.9	25.5	72.4	163.2
KCl	104.2	100.0	21.5	28.9	85.8	165.8
RbCl	103.4	96.3	19.5	28.9	85.8	159.3
CsCl	106.9	89.7	28.9	85.8	155.4	

(a) The data are at 298°K

(b) The data are at 0°K

(c) A correction of -3.0 KCal for the reaction,
 $M(g) + X(g) \xrightarrow{298^\circ\text{K}} M(g) + X(g) \xrightarrow{0^\circ\text{K}}$, has been
 added (in all the cases) to obtain U_L at 298°K.

II. BORN TREATMENT OF IONIC SOLIDS

The first attempts to calculate the cohesive energy of ionic solids were by Born and Madelung who made the basic assumption that an ionic crystal may be considered as built up of positive and negative ions whose charge distribution is spherically symmetric. Thus, the force between two ions would only depend on the distance. If one considers a row of positive and negative ions, + - + - + - +, where the distance between adjacent ions is r , it becomes readily apparent that the coulombic energy becomes progressively lower as the contributions from the ions farther away are added; the net contribution from attractive terms is always larger than that from the repulsive term as given by equation 9.

$$E_{\text{coulomb}} = -\frac{2e^2}{r} (1 - 1/2 + 1/3 - 1/4 + 1/5 - 1/6 + \dots) \quad (9)$$

where e is the charge per ion; the first two terms here give the attractive energy in a molecule of NaCl.

In reality, crystals contain three dimensional arrays of ions and the coulombic energy will be given by an expression considerably more complex than equation 9. Since coulomb forces decrease slowly with distance, it would not suffice to consider only a few ions around a central ion. The coefficient of e^2/r is therefore given by a series which is determined by the crystal structure. In the case of NaCl, the series is

$$E_{\text{coulomb}} = -\frac{e^2}{r} \left[\frac{6}{\sqrt{1}} - \frac{12}{\sqrt{2}} + \frac{8}{\sqrt{3}} - \frac{6}{\sqrt{4}} + \frac{24}{\sqrt{5}} - \dots \right] \quad (10)$$

Note that a Na^+ ion is surrounded by 6 Cl^- ions at a distance r , 12 Na^+ ions at a distance $\sqrt{2} r$, 8 Cl^- ions at $\sqrt{3} r$ etc. Since the coefficient of e^2/r is a constant, the general expression for E_{coulomb} may be written as

$$E_{\text{coulomb}} = -\alpha_M e^2/r \quad (11)$$

where α_M is called the Madelung constant. Series of the type 10 have been calculated for various structures and the values of α_M for a few typical structures are: NaCl, 1.74756; CsCl, 1.76267; ZnS (sphalerite), 1.63805; ZnS (wurtzite), 1.64132; CaF₂, 5.03878; Cu₂O, 4.44248. In equation 11, e represents the electronic charge times the valence of the ions under consideration.

It was realized quite early by Born that repulsive forces must be present in an ionic solid to prevent it from collapsing. The repulsive energy between an ion and other ions was first assumed to be, $E_{\text{rep}} = B/r^n$, where n and B are constants characteristic of the ionic solid. The total energy of one ion pair therefore becomes,

$$E_1 = \alpha_M e^2/r + B/r^n \quad (12)$$

For a crystal containing N positive and N negative ions,

$$E = N(-\alpha_M e^2/r + B/r^n) \quad (13)$$

When the crystal is in equilibrium at absolute zero, $r = r_0$ and the energy will be minimum; at this minimum,

$$(dE/dr)_{r=r_0} = 0 \quad (14)$$

From equations 13 and 14

$$B = (\alpha_M e^2/n)r_0^{n-1} \quad (15)$$

The cohesive energy U_L is therefore,

$$U_L = -N \frac{e^2}{r_o} \left[1 - \frac{1}{n} \right] \quad (16)$$

In this equation, the only unknown is the repulsive exponent n .

The repulsive exponent, n , was derived by Born from compressibility data. The compressibility at absolute zero, β_0 , is defined as

$$1/\beta_0 = V_o (d^2E/dV^2)_{V=V_o} \quad (17)$$

where V_o is the crystal volume corresponding to r_o . Since $V = Ncr^3$ where c is a constant for a particular type of structure,

$$(dE/dV) = (1/3Ncr^2) (dE/dr) \quad (18)$$

$$\text{and } (d^2E/dV^2) = (1/9Nc^2r^2) d/dr [1/r^2(dE/dr)] \quad (19)$$

Combining equations 13, 14, 15 and 19 we obtain,

$$n = 1 + (9cr_o^4/\beta_o e^2 \alpha_M) \quad (20)$$

If β_0 is known for an ionic solid, the value of the repulsive parameter n can thus be calculated. Some of the values of n obtained from Slater's compressibility data are LiF, 5.9; LiCl, 8.0; LiBr, 8.7; NaCl, 9.1 and NaBr, 9.5. The value of n varies for different ionic solids. An examination of equation 16 shows that the contribution from the repulsive term is much smaller than the coulombic term. The value of n is not very critical as far as the calculation of the lattice energy is concerned.

Equations (18) and (19) are valid only for static crystals. In more refined treatments of the equation of state, the vibrational parts of the thermodynamic functions of the solid (at a given temperature and pressure) are approximately taken into account. The Hildebrand equation of state assumes the vibrational energy of the solid to be a purely temperature-dependent function.

Thus, we can write,

$$dU_L/dV = -P + T \eta / \beta \quad (21)$$

Where η is the volume thermal expansion. Equation (8) can thus be written as

$$V(d^2U_L/dV^2) = 1/\beta + T/\beta^2 [(\partial\beta/\partial T)_P + \eta/\beta (\partial\beta/\partial P)_T] \quad (22)$$

On the other hand the Mie-Gruneisen equation of state¹⁶ assumes (on the basis of the Debye model) that the vibrational free energy of the solid divided by its temperature to be a function of only the ratio between the temperature and a purely volume-dependent characteristic temperature:

$$\frac{dU_L}{dV} = -P + W^{\text{vib}} \eta / C_v \beta \quad (23)$$

$$\text{and } V \left(\frac{d^2 U_L}{dV^2} \right) = 1/\beta + \left(W^{\text{vib}} / C_v \beta^2 \right) \left[\left(\frac{\partial \beta}{\partial T} \right)_P + \eta / \beta \left(\frac{\partial \beta}{\partial P} \right)_T \right] + V \eta^2 / C_v^2 \beta^2 \left[T C_v - W^{\text{vib}} - T W^{\text{vib}} / C_v \left(\frac{\partial C_v}{\partial T} \right)_V \right] \quad (24)$$

Here C_v and W^{vib} are the heat capacity at constant volume and the vibrational energy respectively. At temperatures somewhat lower than the Debye characteristic temperatures, equations (23) and (24) should be used for a solid whose Gruneisen parameter $V\eta/C_v\beta$ is volume dependent.

The repulsive forces between two ions depend on the electron distribution in the ions; the number of electrons in the outer shells in an ion would therefore determine the magnitude of the repulsive term. Pauling has proposed values of n on the basis of the occupation of electronic shells. The values for ions with different closed-shell (inert gas) configurations are as follows

He	Ne	Ar	Kr	Xe
5	7	9	10	12

For an ionic solid MX the value of n would be the average of the values for M and X. Thus, n for NaCl is 8 while it is 7 for NaF.

The lattice energies calculated from equation 16 agree well with the experimental values. Some of the calculated and experimental values (in eV) are compared below

	calc	obs		calc	obs
LiCl	8.4	8.6	LiBr	7.9	8.2
NaCl	8.0	7.9	NaBr	7.5	7.5
KCl	7.1	7.1	KBr	6.8	6.8
RbCl	6.9	7.0	RbBr	6.6	6.6
CsCl	6.5	6.7	CsBr	6.2	6.4

With the increasing impact of quantum mechanics in the study of the chemical binding and interatomic forces, it was soon recognized that the early Born treatment of ionic solids needed refinements. Accordingly, Born and Mayer suggested that the repulsive

potential is better represented in the exponential form,

$$E_{\text{rep}} = Be^{-r/\rho}, \quad (25)$$

where B and ρ are constants. An additional term was included to represent the attractive energy arising from London dispersion (van der Waals) forces which are operative in all states of matter by virtue of dipole-dipole and dipole-multipole interactions. The interaction energy between two ions or atoms of closed shell electronic configuration is approximately given by,

$$E_{\text{dd}} = -(3 p_1 p_2 / 2r^6) (I_1 I_2 / I_1 + I_2) \quad (26)$$

where p_1 and p_2 are the polarizabilities and I_1 and I_2 are the ionization potentials of the two ions. Equation 26 accounts only for part of the van der Waals energy, that due to dipole-dipole interactions, and may be written in the form $-C/r^6$ for an ionic solid. The total van der Waals energy is given by a rapidly converging infinite series; it would suffice, however, to include the next higher term due to the dipole-quadrupole interaction. This term is of the form D/r^8 .

The total van der Waals terms is thus given by,

$$E_{\text{vdw}} = -(C/r^6) - (D/r^8) \quad (27)$$

Although E_{vdw} forms a small part of the cohesive energy of an ionic solid, it has been found to be important in explaining the relative stabilities of different structures of ionic solids and in determining the magnitudes of defect energies. The van der Waals coefficients C and D are written as,

$$C = S'_6 C_{+-} + S''_6 (C_{++} + C_{--})/2 \quad (28)$$

$$D = S'_8 D_{+-} + S''_8 (D_{++} + D_{--})/2 \quad (29)$$

where C_{++} , D_{++} etc. are constants for the interactions of ions with the charges indicated by the subscripts and the constants S'_6 , S''_6 etc. are lattice sums which have been evaluated by Jones and Ingham. The values of these sums and some typical values of the van der Waals coefficients for NaCl and CsCl are listed below:

$C \times 10^{60}$ ergs	$D \times 10^{76}$ ergs				Lattice Sums							
	C_{++}	C_{--}	C_{+-}	C	D_{++}	D_{--}	D_{+-}	D	S'_6	S''_6	S'_8	S''_8
NaCl	1.68	116	11.2	180	0.8	233	13.9	180	6.5952	1.8067	6.1457	0.8002
CsCl	1.52	129	129	1530	278	260	250	2600	8.7088	3.5445	8.2077	2.1476

Another refinement introduced by Born and Mayer was to include the zero point energy, ϕ_0 , in the lattice energy expression. The energy retained by the lattice at absolute zero is calculated by regarding it as a system of coupled oscillators.

$$\phi_0 = (9/4)N h \nu_D \quad (30)$$

where ν_D is the Debye maximum frequency. The total Born-Mayer expression for the lattice energy of an ionic solid may now be written as

$$-U_L = (\alpha e^2/r) + (C/r^6 + D/r^8) - Be^{-r/\rho} - \phi_0 \quad (31)$$

The constant B in the repulsive term in equation 27 was first evaluated by Huggins in terms of adjusted crystal "basic" radii; Ladd and Lee have modified the method so that it becomes applicable to cases where the basic radii are now known. The expression for E_{rep} is given as

$$E_{rep} = M_1 b_{+-} \exp(-r/\rho) + 1/2M_2(b_{++} + b_{--}) \exp(-ar/\rho) \quad (32)$$

where M_1 and M_2 are the number of first and second neighbours and ar is the distance of second neighbours. In equation 28 we have assumed that the repulsive energy is the sum of the two-body repulsive energies and the parameter ρ is independent of the ionic species in a crystal. The repulsive energy can be evaluated if the three parameters b_{+-} , $(b_{++} + b_{--})$ and a are known; the number of parameters can be decreased by making use of the additivity of the basic radii of the ions in a family of solids. The approximate additivity of the interionic distance is given as,

$$r = r_+ + r_- + \Delta \quad (33)$$

where Δ varies slightly from crystal to crystal.

By comparing 32 with the expression for r obtained from 33, we find that the parameter b_{+-} may be written as

$$b_{+-} = \beta_{+-} b \exp(r_+ + r_-)/\rho \quad (34)$$

where b is a parameter constant and β_{+-} is the interaction weightage term. Following Huggins and Mayer, equation 32 may be rewritten as,

$$E_{rep} = M_1 \beta_{+-} b_+ b_- \exp(-r/\rho) + \frac{1}{2} M_2 (\beta_{++} b_+^2 + \beta_{--} b_-^2) \exp(-ar/\rho) \quad (35)$$

where $b_+ = b^{1/2} \exp(r_+/\rho)$ and $b_- = b^{1/2} \exp(r_-/\rho)$. The β_{ij} coefficients, according to Pauling, have the form,

$$\beta_{ij} = 1 + (Z_i/n_i) + (Z_j/n_j) \quad (36)$$

where Z and n are respectively the charge of the ion and the number of electrons in the outermost shell of the ion. A common value for ρ for the members of a family of salts makes b an arbitrary constant and one gets the Huggins-Mayer form for the repulsive energy:

$$B(r) = M \beta_{+-} b_+ b_- \exp(-r/\rho) + M' (\beta_{++} b_+^2 + \beta_{--} b_-^2) \exp(-ar/\rho) \quad (37)$$

Assuming a value of 0.345 for ρ , the characteristic b_+ parameters for the five alkali ions and the characteristic b_- parameters for the four halogen ions were determined; these parameters gave a good fit of the interionic distances in the twenty alkali halides. However, Huggins-Mayer form does not provide a good agreement for the compressibilities of the salts. To get over this difficulty, Cubiciotti allowed ρ to vary with the type of halide ion (i.e. taking a common value of ρ for all the alkali halides formed by a given halide ion). Unfortunately, he used adiabatic compressibilities and a common value for b for all the salts which was no longer valid once ρ was allowed to vary.

Tosi has recently reported a set of parameters using a two parameter equation (25) taking into account only the first neighbour repulsions and a set of basic radii which were modified over those of Goldschmidt; his results indicated an increase in the magnitude of the cation radii and a decrease in the values of anion radii, this observation has been justified by recent theoretical consideration.

Douglas has introduced a repulsive energy expression (eqn. 38) where he uses different hardness parameters depending on the nature of the ion. He assumes a common value of b for all the salts and determines the values of the basic radii and

$$B(r) = \beta_{+-} b M \exp[(r_+ + r_- - r)/\rho_{+-}] + 1/2M'b [\beta_{++} \exp(2r_+ - ar)/\rho_{++} + \beta_{--} \exp(2r_- - ar)/\rho_{--}] \quad (38)$$

ρ_{ij} . Douglas found that $\rho_{+-} = \rho_{++} = \rho_{--}$ for the lithium salts and he later assumed this observation to be true for all the other alkali halides. If we substitute this condition in equation (38) and write the corresponding first and second derivative of the lattice energy with respect to the interionic distance, we get,

$$\sigma' = \left\{ br/\rho_{+-} \right\} \left\{ M \beta_{+-} \exp \left[(r_+ + r_-)/\rho_{+-} \right] + \frac{1}{2M} \left[\beta_{++} \exp \left(2r_+/\rho_{++} \right) + \beta_{--} \exp \left(2r_-/\rho_{--} \right) \right] \right\} \exp \left(-r/\rho_{+-} \right) \quad (39)$$

$$\tau = \left\{ br^2/\rho_{+-}^2 \right\} \left\{ M \beta_{+-} \exp \left((r_+ + r_-)/\rho_{+-} \right) + \frac{1}{2M} \left[\beta_{++} \exp \left(2r_+/\rho_{++} \right) + \beta_{--} \exp \left(2r_-/\rho_{--} \right) \right] \right\} \exp \left(-r/\rho_{+-} \right) \quad (40)$$

We define a quantity δ , as,

$$\alpha_M e^2/r + C/r^6 + D/r^8 - U_L \quad (41)$$

and we get from the equations (21) and (22)

$$\tau/\sigma' = \sigma/\delta \quad (42)$$

Where τ , σ' and δ are all experimental quantities. This relation, however, is not found to hold good in all the alkali halides. Thus the assumption $\beta_{++} = \beta_{--} = a \beta_{+-}$ may not be valid. Tosi's suggestion to use this relation to derive from eqn. (25) an expression for the repulsive energy taking into account the second neighbour repulsions, is hence incorrect.

Other repulsive models have been proposed in the literature to account for the phase transitions of alkali halides. Some aspects of these repulsive models will be briefly discussed later by the author in the chapter on phase transitions of solids. Mention must be made here about the suggestion of Rao and coworkers that $\beta_2/\beta_1 = a$ (where a is the structure constant) in the four parameter repulsive term. This again leads to the condition (given by eqn. (42)) on the experimental quantities σ' and τ . A final remark on the Born treatment of alkali halides would be in order. Even though the Born model can approximately reproduce the lattice energy differences of the NaCl and CsCl phases of alkali halides, it does not permit quantitative predictions about polymorphic transitions with the repulsive models proposed hitherto.

III. SOME RELATED ASPECTS

Quantum mechanics was first applied to the calculation of cohesive energies of ionic crystals by Hylleraas for LiH and by Landshoff for NaCl. Recently, Löwdin has calculated the cohesive energies of a number of alkali halides. Landshoff calculated the energy of the NaCl structure relative to the energy of free ions;

while the accuracy of the method is questionable the calculated value agrees with the experimental value. Löwdin calculates the cohesive energies directly from the free ion wave functions employing a one-electron scheme in which the total wave functions are approximated by single Slater determinants. Since alkali halides have ions with filled shells both V.B. and M.O. procedures give the same results. In his recent work, Löwdin has indicated the importance of the kinetic energy term and the use of the virial theorem in the calculation of cohesive energies. The calculated values of Lowdin agree well with the experimental values.

The experimental and calculated values of the lattice energies of various ionic solids may be found in a number of reviews which have appeared in the literature. Values of Madelung constants and the repulsive and other Born parameters required for the calculation of lattice energies may be found in the references listed at the end of this chapter. It should be noted that the form of the potential energy expression and the values of the parameters are of great importance in the calculation of the formation and interaction energies of defects as well as surface energies of ionic solids.

We shall now briefly examine the problem of surface energies of ionic crystals. In a finite ionic crystal, the ions at (and near) the surface are present in an environment different from that in the interior. The surface energy of a finite crystal may be defined as the difference between the cohesive energy of the crystal and the product of the number of molecules in the crystal and the cohesive energy of an infinite crystal. The surface energy can be further classified into the edge, corner and face energies, all of which depend on the crystallographic plane(s) associated with the edge, the corner or the face. The ions at the surface differ from those in the bulk since the forces on the ions at the surface are not balanced by the normal lattice distances. This leads to a rearrangement (relaxation) of the ions at the surface. The face energy σ is defined as $\sigma = \sigma^0 + \Delta\sigma$, where, σ^0 is the energy of the surface when the ions are held rigidly and $\Delta\sigma$ is a small correction term to account for the relaxation of the surface ions. Further, $\sigma^0 = U_s \omega^{-1}$ where U_s is the energy of a pair of ions on the face of the crystal and ω is the area; ω depends on the face of the crystal. Thus, for the (100) and (110) faces of NaCl, ω is $2r_0^2$ and $2\sqrt{2} r_0^2$ respectively. The term U_s is calculated by employing the Born-Mayer type of expression with the appropriate lattice sum (which are different from those used for the calculation of lattice energies). The coordination numbers in the repulsive term are similarly different for the surface ions.

The author is thankful to Mr. S. Ramdas for his assistance in the preparation of the manuscript.

REFERENCES

1. M. Born and K. Huang, "Dynamic Theory of Crystal Lattices"; Oxford University Press, 1954.
2. M.F.C. Ladd and W.H. Lee, in "Progress in Solid State Chemistry", Ed. by H. Reiss, Vol. 1,2 and 3, Pergamon Press, Oxford.
3. T.C. Waddington in "Advances in Inorganic Chemistry and Radiochemistry", Ed. by H.J. Emeleus and A.G. Sharpe, Academic Press, New York, 1, 157 (1959).
4. M.P. Tosi, in "Solid State Physics", Ed. by F. Seitz and D. Turnbull, Academic Press, New York, 16, 1 (1964).
5. K.J. Rao, G.V. Subba Rao and C.N.R. Rao, J. Phys. C (Proc. Phys. Soc.) Ser. 2, 1, 1134 (1968) and the references cited there in.
6. T.B. Douglas, J. Chem. Phys., 38, 2461 (1963).
7. M. Born and J.E. Mayer, Z. Physik, 75, 1 (1932).
8. M.L. Huggins and J.E. Mayer, J. Chem. Phys., 1, 643 (1933).
9. M.L. Huggins, J. Chem. Phys., 5, 143 (1937); 15, 212 (1947).

DEFECT CHEMISTRY AND NON-STOICHIOMETRIC COMPOUNDS

J. S. Anderson

Inorganic Chemistry Laboratory

University of Oxford, Oxford, U. K.

I. POINT DEFECT EQUILIBRIA IN CRYSTALS

Crystallographic theory treats the structure of crystals in terms of the space lattice - a set of nodes or points in real space, related by unit translation vectors and symmetry operations that define the geometry and dimensions of the unit cell. Each point of the space lattice is associated with a group of atoms - the basis group - that defines the composition and detailed structure of the crystal. The effect of the translational and symmetry operations on this basis group is an identity operation: the space lattice concept implies that the environment of any atom in the basis group is identical throughout the crystal. There is no place in lattice theory for deviations from the ideal order or for variations in compositions and atomic positions from one unit cell to another.

In real crystals, on the other hand, there is inescapable evidence that there are imperfections in order and in the occupation of lattice sites. The atomic transport properties of crystal - diffusion, ionic conductivity, diffusion controlled chemical reaction - require that atomic sites may be vacant, or that atoms may occupy equilibrium positions not proper to the basis group. Such vacancies, interstitial atoms or atoms occupying 'wrong' sites (atoms of element A replacing atoms of B on crystallographic sites) constitute point defects. Crystals of many compounds which, by all operational criteria, are single phase materials may vary in composition over a certain phase range, within which their thermodynamic properties, unit cell dimensions and other properties are functions of their composition. The evidence that this is so comes mainly

from three sorts of observable properties: variation in cell dimensions while retaining the average structure and symmetry of an idealised stoichiometric crystal; variation of electronic properties with composition, and - most important - bivariant thermodynamic behaviour across the phase range: the chemical potentials of the components are continuous monotonic functions of composition. Such non-stoichiometric crystals necessarily involve defects of some kind: variability in the occupation of crystallographic sites. In the first approach to non stoichiometric compounds we consider how far they can be interpreted in terms of point defects.

Defect crystals can be treated theoretically as one aspect of the statistical thermodynamics of solutions. Because of the regular structure of crystals, the constitution of solid solutions must be discussed in terms both of the atomic species involved and the sites that they occupy. Three types of solid solution system can be recognised. (1) True solid solutions, in which one kind of atom is substituted by another with the same charge or valence properties. Examples are the mixed crystals of the alkali halides ($\text{Na}_{1-x}\text{K}_x\text{Cl}$; $\text{Na}(\text{Cl}_{1-x}\text{Br}_x)$, $(\text{Na}_{1-x}\text{K}_x)(\text{Cl}_{1-y}\text{Br}_y)$). Associated with differences in effective ionic size there will be changes in mean lattice spacing, local distortions and possibly a limit to the extent of possible substitution. Nevertheless, the electrostatic environment of every ion is unaffected by the substitution, and the balance between atoms and sites is unchanged. (2) Anomalous solid solutions may be formed when one kind of atom is replaced by another of compatible effective radius, but bearing a different ionic charge. Thus solid solutions containing a few per cent of the divalent ion are formed in the systems $\text{KCl}-\text{SrCl}_2$, $\text{AgCl}-\text{CdCl}_2$. In each case, M^{2+} substitutes for M^+ . The overriding consideration is then the maintenance of balance between anionic and cationic charges in the crystal: we must have the substitution $\text{M}^{2+} \rightarrow 2\text{M}^+$, and it is at once apparent that the balance between atoms and sites is lost. In the examples cited, each substitution creates a vacant cation site: $\text{M}^{2+} + \text{V} \rightarrow 2\text{M}^+$. In the mixed crystals of CaF_2 and YF_3 , the replacement $\text{Y}^{3+} \rightarrow \text{Ca}^{2+}$ leaves the cation sublattice filled, and the additional anion is accommodated in an interstitial position. Anomalous solid solution automatically creates defects. (3) Non-stoichiometric compounds, if the same element - usually the metallic component - can function in two different valence states, the possibility at once arises of a situation similar to (2), but with the added feature that the relative proportions of the two valence states, and therefore the composition of a given crystal and the consequent concentration of defects, can be changed continuously by a shift in a redox equilibrium. Thus, for such oxide systems as Fe_{1-x}O or UO_{2+x} the defect concentration and composition are dependent on the chemical potential of oxygen.

For such non-stoichiometric compounds as UO_{2+x} or Fe_{1-x}O , the

resulting bivariant phase range is very wide, and may not include the idealised composition. (cf. the iron-rich limit of $\text{Fe}_{0.950}$ at 1000-1100°C, or the phase $\text{VC}_{0.73}\text{-VC}_{0.87}$). In the majority of crystalline compounds, any range of composition is not chemically measurable, at least at low temperatures; they are described as line phases in equilibrium diagrams, but subtle variability of composition may be measured by other means - e.g. by changes in electronic properties. It is the task of theory to express the inter-relation of composition, thermodynamic potentials and electronic properties quantitatively, in terms of statistical thermodynamics.

To discuss the equilibria, an agreed symbolism is needed. This must contain three bits of information: the structure element - an atom or a vacancy, the site occupied - a proper site on one or other sublattice, or an interstitial site -, and the relative charge borne by the species. The last is defined as follows: Removal of a cation from an ionic structure would leave a zero charge at a site which properly bears a charge $+Z_{\text{Me}}$. The same effect on the electrostatic energy of the crystal would be attained by superimposing a charge of $-Z_{\text{Me}}$ on a normally occupied cation site. The cation vacancy thus bears a charge $-Z_{\text{Me}}$ relative to the normal regular periodicity of + and - ionic charges in the crystal. Similarly, an anion vacancy bears a relative charge of $+Z_{\text{e}}$. Removal of a neutral atom from a cation or anion site would leave the normal positive or negative charge associated with that site, albeit delocalised over the nearest neighbour atoms - a V centre or an F centre - and would create a (relative) neutral vacancy. The point defect is then completely specified by the symbol:

relative charge
(Structure element)
site

The symbols \times , $'$, $'$ are used to denote neutral, positive and negative charged defects, to differentiate between the relative and the absolute charges located at the site concerned.

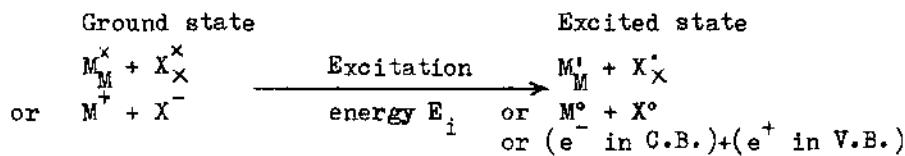
Considering the case of a compound MX, with univalent elements M and X, the possible types of point defect are:

Proper structure elements Possible point defects

M_M^x	X_X^x	Vacancies: V_M^x V_M' V_X^x V_X'
		Interstitials: M_I^x M_I' X_I^x X_I'
		Substitutions: M_X^x X_M^x

The charge state of defects can be considered from another, more fundamental standpoint. Any real crystal is a periodic

assemblage of a very large number of atoms and must be treated statistically; the electronic states must be expressed in terms of the band model. In the case of an ionic crystal, the band structure approaches the extreme limit of the tight binding approximation: the electron states in the filled valence band are highly localised and little different from the atomic orbitals of the isolated ions. In the conduction band there may be a quasi-continuum of states, or the situation may again be roughly described in terms of localised states. In the latter case, the valence band - conduction band excitation process is equivalent to a charge transfer excitation:



Interruption of the regular lattice periodicity creates discrete new localised electron states within the band gap. These can be represented by hydrogen-like wave functions, but since they are created in a medium of high dielectric constant, the radial extension of the s-like wave function is large and the lowest orbital embraces the near-neighbour sites of the defect. An anion vacancy gives rise to a set of localised levels below, and possibly very close to, the lower edge of the conduction band. The lowest energy state is that in which the lowest level is occupied - i.e. it is the relative neutral vacancy V_X^X . Excitation of an electron to the conduction band requires the relatively small energy E_b and furnishes a conduction electron: $V_X^X \rightarrow V^+ + e^-$ (in C.B.). Such levels are described as donor levels. Discrete levels just below the lower edge of the conduction band are similarly created by interstitial cations, so that the lowest state of an interstitial cationic defect is the (relative) neutral state M_I^X representing an interstitial atom M^0 ; ionisation of such a defect to an interstitial cation usually involves an exceedingly small energy increment ($M^0 \rightarrow M^+$) and may be represented $M_I^X \rightarrow M_I^+ + e^-$ (in C.B.). Cationic vacancies give rise to a set of levels above the top of the valence band. In this case, the excitation and ionisation processes are most conveniently expressed in terms of positive holes, e^+ . The lowest state of the cationic vacancy is that in which the discrete level is occupied by a positive hole; ionisation of the cation vacancy represents excitation of the positive hole into the valence band, with an excitation energy E_a , (a statement equivalent to the excitation of an electron from the valence band into the vacant level) - $V_{M^+}^X \rightarrow V_M^+ + e^+$ (in V.B.). Interstitial anions similarly give rise to levels above the valence band. Cation vacancies or anionic interstitials are said to create acceptor levels. At the absolute zero, all electrons must be in the lowest available state;

at higher temperatures the distribution of electrons between valence band, conduction band and discrete levels is determined by Fermi Dirac statistics. If, as shown in Fig. 1, the level V_X^*

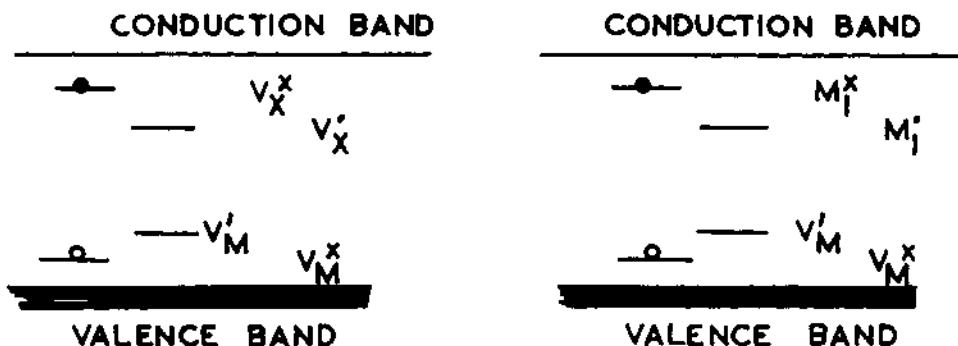


Fig. 1 Schematic band structure of semiconducting or insulating compound with discrete impurity levels.

lies near the top of the band gap and V_M^* near the bottom, the state of lowest energy will be that in which electrons drop from the (occupied) V_X^* levels to the (unoccupied) V_M^* levels: $V_X^* + V_M^* \rightarrow V_X + V_M$. In such a case - the usual situation in highly ionic solids at least - the ground state of the system corresponds to the presence of ionised defects.

We consider, for simplicity, a uni-univalent crystal of idealised structure-type MX , and with the real composition $MX_{1+\delta}$. The treatment is general, and can readily be extended to the general case of a crystal M_aX_b , or to crystals with atoms of higher valency. In the latter case, account must be taken of two or more ionization states for the defects. Neglecting for the moment the ionisation state of defects, and writing as N_{ij} the number of defects of type i affecting atoms of type j , the number of M atoms and X atoms, N_M , N_X , in a crystal with $S_M = S_X = S$ sites of each kind can be written as

$$N_M = S - N_{VM} + N_{IM} - N_{XM} + N_{MX}$$

$$N_X = S - N_{VX} + N_{IX} - N_{MX} + N_{XM}$$

and the deviation from ideal stoichiometry $MX_{1.0000}$ will be dependent on $N_X - N_M = (N_{VM} - N_{VX}) + (N_{IX} - N_{IM}) + 2(N_{XM} - N_{MX})$.

In principle, the internal equilibrium concentrations of defects within the crystal, and the equilibrium of the crystal with its environment, which determines the stoichiometric ratio $N_M:N$,

must be obtained by setting up and maximising the partition function for the system. As long as the concentration of defects is small, the partition function for the imperfect crystal Z_{imp} can be separated into two parts: a thermal part Z_t which, to a first approximation, may be taken as equal to the partition function for the perfect stoichiometric crystal Z_t^* , and a part Z_d representing the partition function associated with the defects:

$$Z_{\text{imp}} = Z_t^* \times Z_d$$

$$\ln Z_{\text{imp}} = \ln Z_t^* + \ln Z_d$$

Here Z is $Z(N_M, N_X, \text{all } N_{ij}, T, \mu_M, \mu_X)$, but N_M, N_X and N_{ij}, μ_M and μ_X are inter-related. Then the free energy change in going from the perfect crystal ($N_M = N_X, N_{ij} = 0$) to the defect crystal can be written as

$$-\Delta G_d = kT \ln Z_d$$

More strictly, it cannot be assumed that Z_t is strictly unchanged from Z_t^* , and an additional factor allowing for the change in the thermal part of the partition function should be introduced:

$-\Delta G_d = kT \ln Z_d + \ln(Z_t/Z_t^*)$. This correction will be neglected for the present. The external equilibrium of the crystal with its environment is then given by the relations

$$\frac{\partial G_d}{\partial N_M} = \mu_M/kT, \quad \frac{\partial G_d}{\partial N_X} = \mu_X/kT,$$

the chemical potentials of the components being inter-dependent through the Gibbs-Duhem relation.

Implicit in these expressions is the assumption that each point defect of a particular kind introduced into the crystal raises the energy of all quantum states of the system by the same amount. As long as the defects are dilute, so that they do not interact with each other, we may associate the creation of each defect (ij) with the expenditure of a characteristic energy increment E_{ij} . Interaction with the environment, whereby the stoichiometry is changed, involves the transfer of M (or X) atoms between the gas and normal lattice sites, which must be created or annihilated in equivalent numbers. The corresponding energy of reactions therefore involves the change in Madelung energy, the creation or annihilation of a defect and the changes in chemical binding energy - ionisation energies or bond energies - resulting from the changed balance of cation and anion numbers. In a complete treatment, the partition function for atoms and point defects must be supplemented by an electronic part, embodying the Fermi Dirac statistics of the distribution of electrons and positive holes between valence band, conduction band and defect levels.

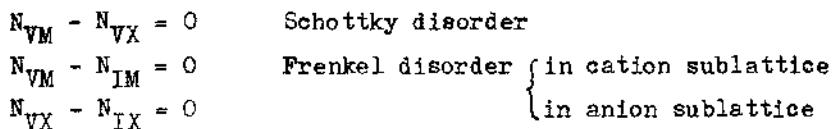
Evaluation of the defect partition function, for the general

case in which all modes of point defects are considered, is very cumbersome. It can readily be seen, however, that a simplification of the problem is permissible in most cases. The partition function for defects is made up of the product of a series of terms of the form

$$\sum_{N_{ij}} g(N_{ij}) \exp -[N_{ij} E_{ij}/kT]$$

where $g(N_{ij})$, the statistical weight of the term for N_{ij} of the specified defects, represents the number of equivalent configurations - e.g. the numbers of ways that N_{ij} defects can be distributed over the relevant number of lattice sites. This will be given by a straightforward combinatorial expression, and yields directly the contribution of the defects (ij) to the configurational entropy. It can be foreseen that the equilibrium concentration of each species of defect, obtained by maximising the partition function, will contain a factor $\exp -[E_{ij}/kT]$. If, as is likely to be the case, the endothermicities of different defect types differ substantially, these probability factors will be of greatly different magnitude for different defect types. It is then permissible to regard the defect equilibrium in the stoichiometric crystal as involving only one type of complementary defects; in the nonstoichiometric crystal, stoichiometric unbalance arises from inequality between the concentrations of the defects that constitute the dominant and only significant mode of intrinsic disorder.

For the stoichiometric crystal, writing $N_M = N_X$, the important limiting cases are:

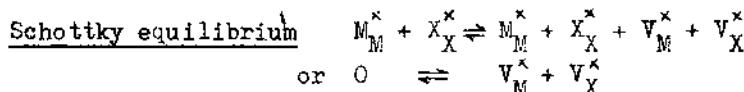


For non-stoichiometric crystals:

Metal excess: $N_{VM} < N_{VX}$ or $N_{VM} < N_{IM}$ or $N_{VX} > N_{IX}$

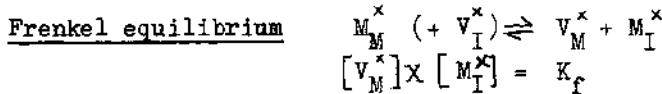
Non-metal excess: $N_{VM} > N_{VX}$ or $N_{VM} > N_{IM}$ or $N_{VX} < N_{IX}$

The formation of complementary defect pairs in a crystal (regarded as a closed system) can then be treated as a quasi-chemical reaction, to which an equilibrium constant can be assigned:



In an infinite crystal, creation of an additional pair of sites has no effect on the concentrations (more strictly, the activities) of normally occupied lattice sites, so that the second,

abbreviated form of the defect creating reaction is permissible. Writing the concentration of defects from this point onwards as the number per unit volume, we have $[V_M^x] \times [V_X^x] = K_S$



The relation to a more formal statistical thermodynamic treatment is clear. Writing

$$K = \exp(-\Delta G_d/kT) = \exp(\Delta S_d/k) \exp(-\Delta H_d/kT)$$

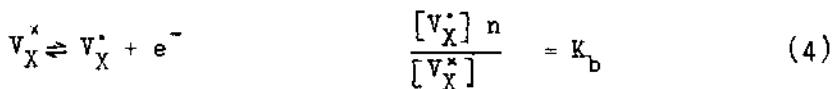
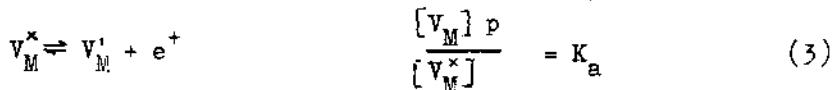
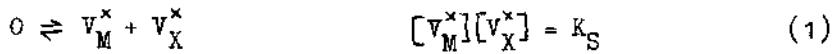
the first term is the configurational entropy, represented by the microcanonical part of the partition function considered above; the second term contains the energetic cost of creating a defect pair (e.g. $E_{yM} + E_{yX}$ in the case of a Schottky defect pair) and the volume energy associated with an increase in the number of sites in the crystal.

It may be noted that since ΔH_d has been assumed to be very different for Schottky disorder, cationic Frenkel disorder and anionic Frenkel disorder in any one crystal, the corresponding equilibrium constants will have very different temperature dependence. The assumption that a single defect type is important, although valid at lower temperatures, may no longer be strictly valid at high relative temperatures - e.g. approaching the melting point of an ionic or polar crystal. Because of the highly unfavourable electrostatic environment arising from the interchange of metal and nonmetal atoms in any crystal where there is a considerable difference in the electronegativity of the components, it is usually accepted that substitutional disorder is highly improbable in polar crystals.

Kröger-Vink Treatment of Equilibria. Given that the native disorder in a crystal can be regarded as of one type, or at least dominated by one type, the complete defect equilibria, including the ionisation of defects, can readily be discussed from the quasi-chemical standpoint. We consider a uni-univalent crystal MX with native Schottky equilibrium; exactly similar treatments can readily be worked out for crystals with Frenkel defects in the cation or the anion sublattice. For extension of the method to include multiply charged defects or more general formulae, reference should be made to F.A. Kröger, "The Chemistry of Imperfect Crystals".

The essentials comprise (i) the intrinsic inner equilibria of point defects and of electron distribution; (ii) the ionisation of defects; (iii) the transfer reaction, whereby atoms of M or of X are exchanged between sites in the crystal and a gaseous phase, at given chemical potential, in equilibrium with the crystal; and (iv) a charge balance condition.

We may then write down (i) the equations and equilibrium constants for the intrinsic disorder, in terms of neutral defects - equations (1) and (2); (ii) equations and equilibrium constants for defect ionisation (equations (3) and (4)), which may be combined with (1) and (2) to express the intrinsic disorder in terms of ionised defects (equation 1a).



(1) can then be rewritten in the alternative form:



From (1), (2), (3) and (4)

$$K_S = \frac{K_i}{K_a K_b} K'_S$$

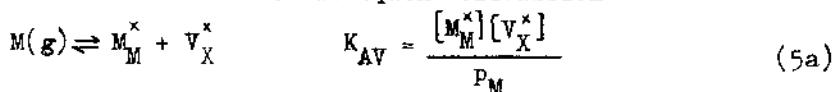
The ionised Schottky constant K'_S is the property that determines the intrinsic Schottky disorder in that large proportion of solids in which the defects are ionised in the ground state.

In an exactly stoichiometric crystal ($\delta = 0$ in $MX_{1+\delta}$)

$$\frac{[V_M^+]_0}{[V_X^+]_0} = \xi, \text{ the degree of native disorder}$$

$$\text{Then } K'_S = \xi^2$$

The transfer reaction can be written in terms of either component; the resulting expressions are thermodynamically related and equivalent, and the convenient form depends on whether the chemical potential of the metal or the nonmetal is the experimentally measured variable in any case. Thus, in studying the oxides, sulphides, etc. it is usual to control and measure the nonmetal equilibrium pressure, and to express the chemical potentials in terms of the diatomic molecular species O_2 , S_2 etc. The form (5b) is therefore used in the subsequent discussion.



$$\frac{1}{2} K_2(g) = V_M^x + X_X^x \quad K_{B_2} V = \frac{[V_M^x]}{\frac{1}{2} P_{X_2}^2} \quad (5b)$$

$$[V_M^x] = K_{B_2} V \cdot P_{X_2}^{\frac{1}{2}} = R \quad (5c)$$

The over-riding requirement is that of charge balance: equivalence between the total concentrations of all positively and negatively charged entities:

$$p + [V_X^+] = n + [V_M^+] \quad (6)$$

With this set of concurrent equilibria it is now possible to recognise boundary conditions, under which the charge balance is completely determined by a single pair of charged species. The equilibrium concentrations of defects and carriers can then be evaluated within each range. With two positively and two negatively charged species in (6), three ranges of behaviour are possible:

I Stoichiometric excess of metal $[V_X^+] \gg [V_M^+]$, $n \gg p$
 $[V_X^+] = n \quad (7)$

III Stoichiometric excess of nonmetal $[V_X^+] \ll [V_M^+]$, $n \ll p$
 $[V_M^+] = p \quad (8)$

II Near stoichiometry. There are now two possibilities

IIa If $K_s' > K_i$ $[V_X^+] = [V_M^+] > n, p$
Boundary condition $[V_X^+] \approx [V_M^+] \quad (9)$

IIb If $K_i > K_s'$ $n, p > [V_X^+], [V_M^+]$
Boundary condition $n \approx p \quad (10)$

Considering range I, we have from (7), $[V_X^+] = n$

$$[V_M^+] = R \quad \text{in all ranges}$$

$$[V_X^+] = K_s R^{-1} = \frac{K_i}{K_a K_b} K_s' R^{-1} \quad (11)$$

$$[V_X^+] n = [V_X^+]^2 = n^2 = K_b [V_X^+] \quad (12)$$

$$[V_X^+] = n = \left(\frac{K_i}{K_a} K_s' \right)^{\frac{1}{2}} R^{-\frac{1}{2}} \quad (12)$$

$$[V_M^+] = \frac{K_s'}{[V_X^+]} = \left(\frac{K_a}{K_i} K_s' \right)^{\frac{1}{2}} R^{\frac{1}{2}} \quad (13)$$

$$p = \frac{K_i}{n} = \left(\frac{K_a K_i}{K'_s} \right)^{\frac{1}{2}} R^{-\frac{1}{2}} \quad (14)$$

All concentrations are thereby specified, and similar manipulation yields the concentrations of carriers and defects, in each of the other ranges, in terms of the equilibrium constants. The results are tabulated below.

Range	I	IIa	IIb	III
Condition	$[V_X^+] = n$	$[V_X^+] = [V_M^+]$	$n = p$	$[V_M^+] = p$
$[V_M^+]$	R	R	R	R
$[V_X^+]$	$\left(\frac{K_i K'_s}{K_a K_b} \right) R^{-1}$	$\left(\frac{K_i K'_s}{K_a K_b} \right) R^{-1}$	$\left(\frac{K_i K'_s}{K_a K_b} \right) R^{-1}$	$\left(\frac{K_i K'_s}{K_a K_b} \right) R^{-1}$
$[V_M^+]$	$\left(\frac{K_a K'_s}{K_i} \right)^{\frac{1}{2}} R^{\frac{1}{2}}$	$K_s^{\frac{1}{2}}$	$\left(\frac{K_a}{K_i^{\frac{1}{2}}} \right) R$	$K_a^{\frac{1}{2}} R^{\frac{1}{2}}$
$[V_X^+]$	$\left(\frac{K_i K'_s}{K_a} \right)^{\frac{1}{2}} R^{-\frac{1}{2}}$	$K_s^{\frac{1}{2}}$	$\left(\frac{K'_s K_i^{\frac{1}{2}}}{K_a} \right) R^{-1}$	$\frac{K'_s}{K_a^{\frac{1}{2}}} R^{-\frac{1}{2}}$
n	$\left(\frac{K_i K'_s}{K_a} \right)^{\frac{1}{2}} R^{-\frac{1}{2}}$	$\left(\frac{K_i K'_s}{K_a} \right)^{\frac{1}{2}} R^{-1}$	$K_i^{\frac{1}{2}}$	$\frac{K_i}{K_a^{\frac{1}{2}}} R^{-\frac{1}{2}}$
p	$\left(\frac{K_a K_i}{K'_s} \right)^{\frac{1}{2}} R^{\frac{1}{2}}$	$\frac{K_{a_1}}{K_s^{\frac{1}{2}}}$	R	$K_a^{\frac{1}{2}} R^{\frac{1}{2}}$

The (approximate) concentrations, N_i , per unit volume, can be plotted logarithmically against $\log R$ ($= \frac{1}{2} \log P_X + \text{constant}$), as in Figs. 2 and 3 for the two cases of a wide band gap solid ($K' > K_i$) and an intrinsic semiconductor ($K_i < K'_s$). The former is of particular interest for the non-stoichiometric behaviour of oxides, halides and similar compounds.

Considering first the wide band gap case, it is clear that there is a range of equilibrium chemical potentials over which the composition of the crystal remains, to a first approximation, stoichiometric. To a more exact approximation this is not true: the stoichiometric solid, with $[V_M^+] = [V_X^+] = K_a^{\frac{1}{2}}$, exists only at a particular chemical potential of X_2 , given by $R_o = (K_i K'_s)^{\frac{1}{2}} / K_a$.

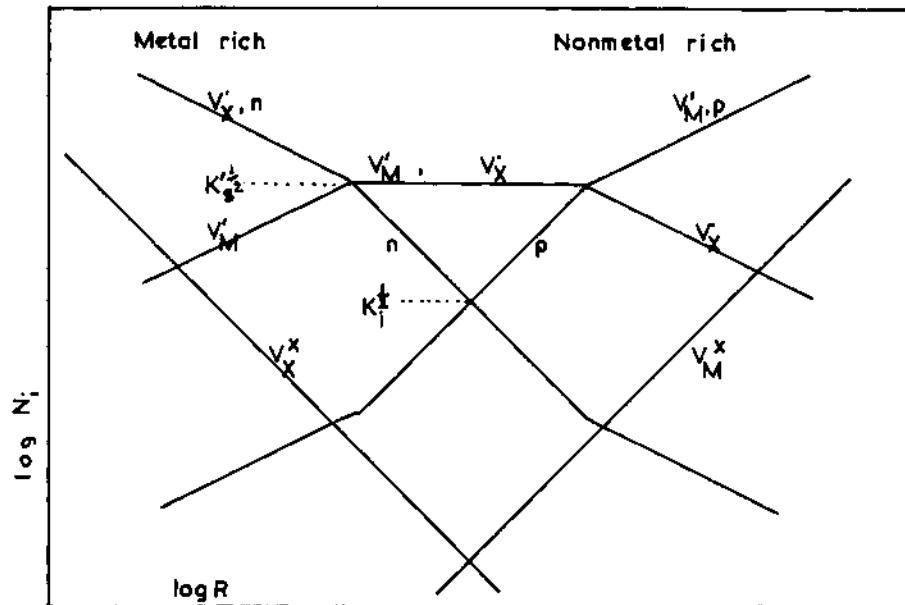


Fig. 2

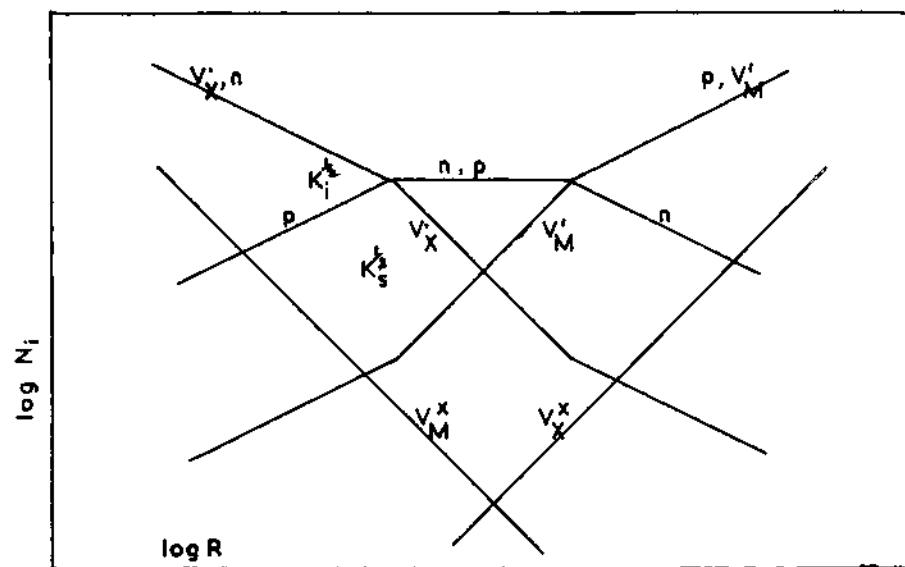


Fig. 3

At all other equilibrium pressures there is some deviation from ideal composition, revealed by changes in the carrier concentrations n , p , with a change-over from n-type to p-type properties at the stoichiometric point. From the geometry of the diagram it is evident that the width of this central range is given by

$$\Delta \log p_{X_2} = 2 \log (K'_s / K_i) \quad (15)$$

The extent to which the stoichiometry of the crystal actually varies over the central range is readily established. The true junctions between ranges I and II, II and III could be defined as:

$$\begin{aligned} \text{I/II} \quad n + [V_M^{\circ}] &= [V_X^{\circ}] \text{ with } n = [V_M^{\circ}] \\ \text{II/III} \quad p + [V_X^{\circ}] &= [V_M^{\circ}] \text{ with } p = [V_M^{\circ}] \quad (\text{Fig. 4}) \end{aligned}$$

Then, at the I/II junction

$$\begin{aligned} [V_M^{\circ}] &= \frac{1}{2} [V_X^{\circ}] = \frac{1}{\sqrt{2}} K_s^{\frac{1}{2}} = \frac{1}{\sqrt{2}} \xi \\ [V_X^{\circ}] &= \sqrt{2} \xi \\ \delta = [V_M^{\circ}] - [V_X^{\circ}] &= -\frac{1}{\sqrt{2}} \xi \quad (16) \end{aligned}$$

Similarly, at the II/III extremity of the central range

$$\delta = +\frac{1}{\sqrt{2}} \xi$$

Hence, over the central range, the stoichiometric defect varies only by $\pm 0.7\xi$ over a wide range of chemical potentials. If the intrinsic disorder is small, the crystal is, in an operational, chemical sense, invariable in composition. This is the meaning of a line phase.

It is of interest to consider some typical values. For typical oxides, halides etc. K'_s , or K'_f are determined by the energy of creation of a defect pair, ΔE_s or ΔE_f . These energies are typically in the range 2 to 5 ev., as determined by calculation or from the activation energies for diffusion or ionic conduction in pure and doped crystal. The values of ξ at any temperature (e.g., 0.8 Tm) are very small. Hence the line phase range of behaviour at 0.8 Tm covers a range of composition that is barely measurable by any chemical means. The intrinsic carrier concentration is determined by the band gap, ΔE_i , typically 5 to 10 ev. The intrinsic carrier concentration is therefore exceedingly small indeed - such compounds are typical insulators - and the central range, measured by ξ^2/n_0^2 , is likely to be very wide.

Data for KBr illustrate the situation. For KBr at 600°C, $K'_s = 10^{-14}$. Hence $\xi \sim 10^{-7}$, or $\sim 10^{15}$ vacancies/cm³. K_i at 600° is about 3.3×10^{-35} ($n_0 = p_0 \sim 5.7 \times 10^{-18}$, or $\sim 10^5$ carriers/cm³).

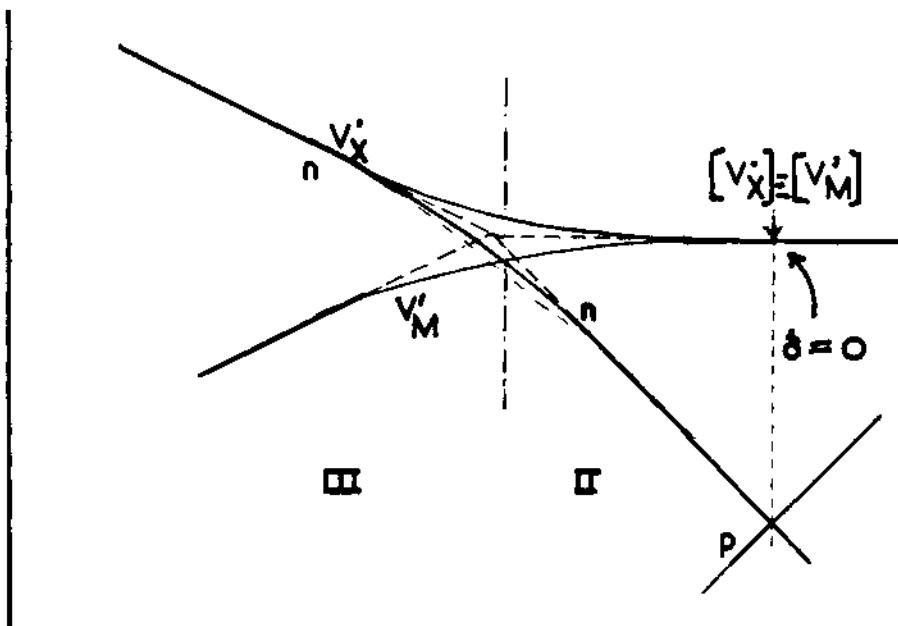


Fig. 4. Detail of the equilibrium concentrations at the junction of ranges I and II. $[V'_X] = [V'_M] + n = 2[V'_M]$

On the simplest model - leaving out some complexities in the defect structure that arise from association of defects into defect complexes - these data lead to a bromine pressure of $\sim 10^{-39}$ atm. at the lower (metal rich) end of the central range, and a pressure of $\sim 10^3$ atm. at the upper end of the range. Over this range of 10^{42} in the bromine activity, the total stoichiometric variability is of the order $\pm 7 \times 10^{-8}$.

We next consider the effect of changing temperature (Fig. 5). Raising the temperature naturally shifts the dissociation equilibrium - e.g. the equilibrium pressure over the stoichiometric solid - to higher pressures. K_s' (and therefore ξ) and K_i are highly temperature dependent, and since $\Delta E_i > \Delta E_s'$, the intrinsic carrier concentration increases more steeply than the intrinsic disorder. As a consequence, the central line-phase range of behaviour becomes narrower at higher temperatures and, because of the increase in ξ , the absolute deviation from ideal stoichiometry within that range becomes larger. If the accessible temperature range is long enough - e.g. for high melting solids such as the metallic oxides - we may

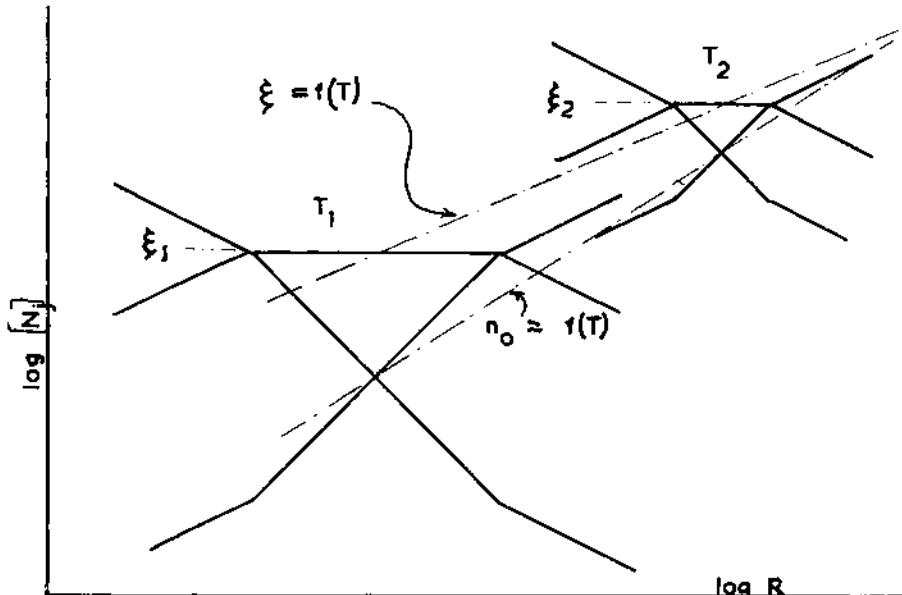


Fig. 5. Displacement of equilibria when temperature is raised from T_1 to T_2 , schematic. Increase in $\xi(T)$ and contraction of central nearly stoichiometric range.

expect to find a progressive change from line phase behaviour to detectable nonstoichiometry.

The other typical situation is that $K_i > K_s$. In this case, there is a central range of behaviour over which the electronic properties are insensitive to stoichiometric composition while the stoichiometry varies continuously with the activity of the components across the whole equilibrium diagram. Over the central range, the stoichiometric defect δ will vary between the limits $\pm 1/\sqrt{2} n_0$ where $n_0 (= K_s^2)$ is the intrinsic carrier concentration. For a narrow band gap solid, n_0 may be high, and it would be expected that such crystals would display nonstoichiometric behaviour.

In fact, the model must not be pressed too far. There is, indeed, a real distinction between ionic (wide band gap) solids and quasimetallic (narrow band gap) or intrinsic semiconductor solids. However, the conduction behaviour of the latter class depends on the detailed band structure, which the model used here has reduced to its simplest schematic form. The central range of behaviour has a breadth determined by K_i/K_s' , and the intrinsic

disorder may be small, so that there may still be a wide range of chemical potentials over which the stoichiometric defect δ remains small. Experimental evidence shows that electronic conductive properties are by no means always coupled with appreciable stoichiometric variability. Intrinsic semiconductors and quasimetallic compounds can be line phases.

We may finally note that the effects of impurity atoms or dopants can be included within this Kröger-Vink treatment. The effect of incorporating a foreign species F is to add a set of extra terms to the charge balance, site balance and electron equilibrium equations. The magnitude of the effect necessarily depends on the concentration $[F]$, especially as compared with the native defect and carrier concentrations $K_s^{1/2}$, $K_i^{1/2}$. A complete representation of the equilibria requires a 3-dimensional isothermal diagram, with $\log [N_i]$, $\log R$ and $\log [F]$ as coordinates. If $[F]$ becomes large enough, the possible formation of ordered ternary compounds must be considered.

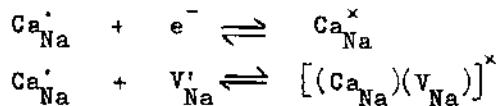
The total range of possibilities, which have been analysed by Kröger (loc. cit.) is therefore quite complex, but three basic types of effect can be recognised.

1. Controlled Atomic Imperfection. This arises when F is an ion of different valence from the ions of the host crystal, both F and the host cations being effectively constant in valency (i.e. $K_i \ll K_s^{1/2}$). Thus incorporation of Ca^{2+} in NaCl represents insertion of Ca^{2+} ions on Na^+ sites, as Ca_{Na}^+ .

The charge balance equation now becomes

$$n + [V_{\text{Na}}'] = p + [V_{\text{Cl}}'] + [\text{Ca}_{\text{Na}}^+] \quad (17)$$

and the Ca_{Na}^+ may trap electrons or ionised cation vacancies:



so that the total calcium concentration is

$$[\Sigma \text{Ca}] = [\text{Ca}_{\text{Na}}^+] + [\text{Ca}_{\text{Na}}^x] + [(\text{Ca}_{\text{Na}}^+)(V_{\text{Na}}')]^x \quad (18)$$

When the alternative boundary conditions are applied, the extreme situations $n = [V_{\text{Cl}}']$ (for $n > [\Sigma \text{Ca}]$) and $p = [V_{\text{Na}}']$ still apply, but if $[\Sigma \text{Ca}] > K_s^{1/2}$ (and $K_s^{1/2}$ for NaCl is small), the condition $[V_{\text{Na}}'] = [V_{\text{Cl}}']$ disappears: the situation is dominated over a very wide range by the condition (19):

$$[\text{Ca}_{\text{Na}}^+] = [V_{\text{Na}}']. \quad (19)$$

The effect of the foreign ions is to introduce compensating defects in equivalent amount.

2. Controlled Electronic Imperfection or Controlled Valence. If the added species is of effectively constant valence, but the host cations of variable valence, the compensation process is one that shifts the electronic equilibria and, in effect, changes the valence state of the host. Thus, for solid solutions of Li_2O in NiO , the basic process is the introduction of Li^+ on Ni^{2+} sites:

$$\text{Li}'_{\text{Ni}}$$

The charge balance equation then becomes

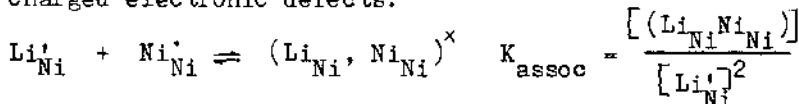
$$[\text{Li}'_{\text{Ni}}] + n + [\text{V}'_{\text{Ni}}] + 2[\text{V}''_{\text{Ni}}] = p + [\text{V}'_0] + 2[\text{V}''_0] \quad (20)$$

and, if $[\text{Li}] > K_s^{\frac{1}{2}}$, the dominant condition over a very wide range of oxygen pressures is the compensation equation

$$[\text{Li}'_{\text{Ni}}] = p \quad (21)$$

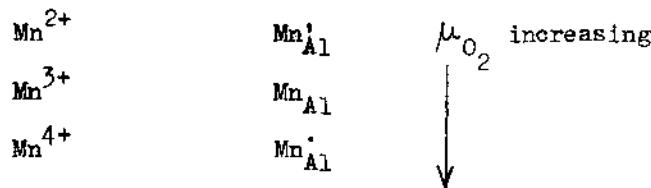
In a chemical sense, the positive hole can be identified as a Ni^{3+} cation : Ni'_{Ni} ; the compensation process is the replacement of 2Ni^{2+} by $\text{Li}^+ + \text{Ni}^{3+}$. The form of compensation equation written above emphasises that incorporation of Li^+ displaces the intrinsic carrier equilibrium and confers p-type conductivity, with the carrier concentration uniquely determined by the concentration of dopant.

There is again the possibility of association between oppositely charged electronic defects:



If the equilibrium constant K_{assoc} is significantly large, Li^+ and Ni^{3+} will pair up strongly on neighbouring cation sites. As the concentration of such associated sites is increased by increasing the lithium concentration, they can interact and mutually order, with the formation of ordered nuclei of the LiNiO_2 structure. The solid state reaction whereby the ternary phase is formed is the consequence of interactions between the defects (in this instance electronic defects) in a defective, random solid solution.

3. Induced Valence. If the host cations are invariable in valency, but the foreign ion can readily assume several valence states (i.e. if its successive ionisation potentials are suitable), the compensation process may differ, as chemical potentials are changed, through changes in the charge state of the foreign ions. Thus, in the introduction of manganese into alumina, three ionisation states of manganese must be considered:



The native disorder in Al_2O_3 is extremely small indeed : we simplify consideration by assuming that the important species are $\text{V}_{\text{Al}}^{\text{m}}$, $\text{V}_{\text{O}}^{\cdot}$ and $\text{V}_{\text{O}}^{\cdot\cdot}$. The native carrier concentration is exceedingly small: $K_i \ll K_s$

The charge balance condition can then be written as

$$[\text{Mn}_{\text{Al}}^{\cdot}] + n + 3[\text{V}_{\text{Al}}^{\text{m}}] = [\text{Mn}_{\text{Al}}^{\cdot}] + p + [\text{V}_{\text{O}}^{\cdot}] + 2[\text{V}_{\text{O}}^{\cdot\cdot}] \quad (22)$$

and the manganese balance as

$$[\Sigma \text{Mn}] = [\text{Mn}_{\text{Al}}^{\cdot}] + [\text{Mn}_{\text{Al}}^{\cdot}] + [\text{Mn}_{\text{Al}}^{\cdot}] \quad (23)$$

It then appears that the significant boundary conditions are :

$$\text{Extreme reducing conditions } n = [\text{V}_{\text{O}}^{\cdot}] \quad [\text{Mn}_{\text{Al}}^{\cdot}] \approx [\Sigma \text{Mn}] \quad (24)$$

$$\text{Reducing conditions } [\text{Mn}_{\text{Al}}^{\cdot}] = [\text{V}_{\text{O}}^{\cdot}] \quad [\text{Mn}_{\text{Al}}^{\cdot}] + [\text{Mn}_{\text{Al}}^{\cdot}] = [\Sigma \text{Mn}] \quad (25)$$

$$\text{A very wide middle range } [\text{Mn}_{\text{Al}}^{\cdot}] = [\text{Mn}_{\text{Al}}^{\cdot}] \quad [\text{Mn}_{\text{Al}}^{\cdot\cdot}] \approx [\Sigma \text{Mn}] \quad (26)$$

$$\text{Oxidising conditions } [\text{Mn}_{\text{Al}}^{\cdot}] = 3[\text{V}_{\text{Al}}^{\text{m}}] \quad [\text{Mn}_{\text{Al}}^{\cdot}] + [\text{Mn}_{\text{Al}}^{\cdot\cdot}] = [\Sigma \text{Mn}] \\ \dots \quad (27)$$

Because of the exceedingly small value of K_i , there is a very wide range of oxygen equilibrium pressures over which the foreign atom is constrained to adopt the charge state that is appropriate to the host crystal.

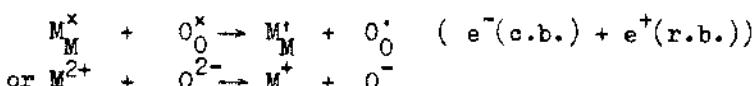
II. THE SYSTEMATICS OF NON-STOICHIOMETRIC COMPOUNDS

In Section I, there has been sketched out a completely general theory, based on statistical thermodynamics and applicable, in principle, to all ionic and covalent or metallic compounds. It shows that the perfectly stoichiometric crystal does not have a unique status, but represents the case of exactly balanced concentrations of intrinsic defects, and it predicts variability of composition as a general property. In practice, most compounds show no chemically detectable variability of composition, but behave as true line phases in equilibrium systems, while a relatively smaller number of compounds, mostly compounds of the transition elements, the lanthanide and actinide elements, do form nonstoichiometric phases with a broad and chemically significant

range of composition.

The reason (or, more strictly, one important reason) for line phase behaviour has been pointed out. The degree of intrinsic disorder ξ in the stoichiometric crystal depends on the energy of production of a complementary defect pair, ΔE_s or ΔE_f . If ΔE is large (e.g. for MgO), the degree of disorder is exceedingly small at any temperature; even if ΔE is relatively small - e.g., 2 e.v. per defect pair - ξ can attain a value of only 10^{-4} to $X \times 10^{-3}$ at the melting point of typical ionic solids. ξ enters as a scaling factor into the charge of stoichiometric composition produced by a given change in chemical potential.

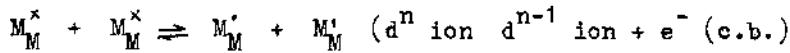
A second scaling factor comes from the equilibrium constant for the component transfer reaction (5). Transfer of a component atom between the crystal and the vapour phase involves not only the creation of a point defect, but also a valence change. The energy of reaction thus includes an ionisation energy and a resulting change in electrostatic lattice energy. In a formal sense, these energy quantities are contained in, and indeed determine, the band gap ΔE_1 . Regarded from a chemical standpoint, a crystal built up from $[s^2 p^6]$ or $[d^{10}]$ cations and inert gas type anions (e.g. oxides, sulphides, halides, etc.) has its filled valence band built up from these states. Excitation of an electron to the conduction band is essentially a charge transfer excitation. The anion valence band is likely to provide the uppermost-lying level for excitation; there are no energetically available acceptor states of the anions so that - in terms of a localised state description - the electron in the conduction band must be associated with a cation state. Thus, in an oxide MO,



In chemical terms: any inert gas anion can be oxidised, but not reduced, any inert gas cation or $[d^{10}]$ cation can be reduced, but not oxidised further. The energies of such charge transfer processes correspond to optical transitions in the short wave or far ultra-violet. Because a change of stoichiometry implies creation of positive holes or introduction of electrons, the same considerations apply: compounds of the kind under consideration can incorporate a metal-excess more readily than a non-metal excess, but the energetics of the valence change process are highly unfavourable. This is another way of stating that the central range, of nearly invariable composition, is wide, and the stoichiometric variability displayed only by changes in electron or positive hole concentration.

In transition metal compounds, however, the partially occupied

d state may provide a multiplet of levels lying between the non-metal valence band and the conduction band. Optical charge transfer, excitation to the conduction band and changes in stoichiometry may now involve these levels:



The transition energy is less than the band gap for a $[d^{10}]$ or $[s^2p^6]$ cation compound of similar structure; hence ΔE_i and the energy of reaction in the transfer reaction are both smaller. Variability of stoichiometry in the solid is related to the same factors as promote variability of valence states throughout chemistry.

From this qualitative recognition of a difference between compounds that are almost stoichiometrically defined and those in which mixed valence states are energetically possible, an important question emerges: how far is it permissible to use the defect theory, as developed, to describe the structure and properties of significantly non-stoichiometric compounds?

As developed, the point defect theory is based on the statistical thermodynamics of dilute, regular solutions, and can be strictly applied only to low concentrations of defects and correspondingly small departures from stoichiometry. It contains two assumptions that are certainly untrue at high concentrations: that the energy per defect is independent of the number of defect per unit volume, and that the configurational entropy is not only independent of the particular way a given number of defects is distributed over the sites of the crystal, but is correctly calculated by a simple combinatorial expression.

For crystals with a low concentration of intrinsic or stoichiometric defects (e.g. for $\delta > 10^{-4} - 10^{-5}$ or $10^{17} - 10^{18}$ defects ion^{-3}), the point defect idea gives a good structural and quantitative description of solids. It is applicable, for example, to colour centres in alkali halides, doped and non-stoichiometric semiconductors etc. Even at the levels encountered in alkali halides, there is evidence from optical properties, e.s.r. and other techniques and both like and oppositely charged defects may associate in pairs or larger aggregates. Where specific association reactions can be formulated, these may be included in the set of equilibria. With some reservations, the range of applicability of the theory may be extended, and probably reaches its valid limit for such compounds as NiO_{1+x} , $\text{Cu}_2\text{O}_{1+x}$ (which has a composition range up to about $\text{Cu}_2\text{O}_{1.003}$ at 1000°C) or SnTe (narrow band gap semiconductor, $K_i > K_s$, with $\delta_{(\max)} \pm 0.01$).

By contrast, in systems that are of particular chemical

interest, δ may be very large and a simple point defect treatment would assume that a high proportion of sites are defects:

TABLE II.1

Compound	Composition range	Apparent defect concentration
"FeO" FeO _{1.00} non-existent at ordinary pressure	Fe _{0.95} O - Fe _{0.85} O at 1300°K	5%-15% cation vacancies
UO _{2+x}	UO _{2.00} -UO _{2.24} (1400°K)	0-1 interstitial O atom per unit cell
	UO _{1.86} -?(2800°)	7% O vacancies
PrO _{2-x} phase above 900°K	PrO _{1.70} -PrO _{2.0}	0-15% O vacancies
TiO High temperature form	TiO _{0.84} to TiO _{1.30} = Ti _{0.92} O _{0.78} -Ti _{0.76} O _{1.00}	Up to 22-24% cation and anion vacancies

In some cases, but by no means all, the intrinsic disorder in the stoichiometric compound is extremely high: thus in both TiO_{1.0} and VO_{1.0}, 15% of anion sites and cation sites are unoccupied; they may be formulated as Ti_{0.85}O_{0.85}, VO_{0.85}O_{0.85}. All these compounds, on the basis of their diffraction patterns, have been assigned the simple rocksalt and fluorite structures, and it has commonly been assumed that they could be described as having randomly distributed point defects. If this were so, the energy of creation of point defects in these structures would, by implication, be very small; there is no evidence to support such an implausible inference. The close average spacing of vacant sites must certainly give rise to strong mutual interaction effects, and the question arises whether, in these concentrated defect systems, the point defect description is applicable at all.

Closely related to this question is the problem of order and disorder in apparently defective crystals. A perfect crystal would necessarily have the ideal stoichiometric composition, since every site in the crystal would be 'properly' occupied, as required by the crystal structure. If defects are present - e.g. if sites on one or other sublattice are vacant, or if atoms occupy interstitial sites that should 'properly' be vacant - complete order is lost. The degree of randomness can then be specified by some ordering parameter, which may be defined in several ways². Taking any atom in the crystal as a reference point, we may define a conditional probability or correlation coefficient P(n) as the probability that

the n^{th} neighbour site is properly occupied by the species of atom required by the idealised crystal structure. If the concentration of defects in one sublattice is x_p (i.e. if a fraction x_p of sites on that sublattice is vacant, or if a fraction x_p of interstitial sites is occupied instead of being vacant), the absolute probability of finding a site on that sublattice correctly occupied is $(1 - x_p)$. If the defects are randomly distributed, the correlation coefficient $P(n) = (1 - x_p)$ at least for widely separated pairs of sites, corresponding to the maximum attainable degree of randomness.

It is possible, however, for the defects originally introduced to array themselves in regularly spaced order, as a result of the interaction forces. They thereby establish a new sublattice, which is properly occupied by empty sites (or by interstitial atoms) when compared with the original crystal structure. The unit translation vectors of this new sublattice define a unit cell which is rationally related to, and is some multiple of, the unit cell of the parent structure: it is a superstructure.

This new superstructure is the criterion by which order must be judged. If superstructure ordering is perfect, the structure is free from defects: unoccupied sites or interstitial atoms have become structure elements of the crystal and an essential part of the ordering pattern. Imperfection of order can then arise in two ways. Additional atoms can be introduced, or additional vacant sites created in the superstructure. These are true point defects. There may also be the possibility of an order-disorder transformation, exactly analogous to the order-disorder transformations in intermetallic compounds. The disordered state represents a randomisation of defects over lattice sites, and the transformation must take place at a temperature directly related to the gain of energy that results from attainment of order. These general considerations are related to real systems in Section V.

It is necessary, however, to consider the problem of order more closely. What has been considered so far is essentially long range order: correlation between the occupation of crystal sites that are widely separated. The long range correlation coefficient may rapidly approach the values characteristic of complete randomness even although the short range order is high - i.e. although correlation coefficients taken over near neighbours approach the values for the ordered state. In the limit, short range order could be perfect over short distances within small regions or microdomains which are, however, quite uncorrelated with their neighbours. In such a case, the only atoms with defective environment would be those on or close to the boundaries between microdomains.

Applying these general ideas to nonstoichiometric compounds, it is clear that the existence of the nonstoichiometric phase, with

an averaged unit cell, implies that long range order is lost. There is, however, always the latent possibility of ordering: either by complete ordering of the crystal as it is, if the defects occupy a rational fraction of lattice sites, or by a phase reaction whereby the crystal breaks down into two, coexisting, ordered structures. Such an ordering process would replace the nonstoichiometric phase by a succession of ordered intermediate phases of definite structure and therefore of definite composition. These would be mixed valence phases with rational, but possibly complex formulae. Alternatively, the intermediate ordered structure might form tolerating a significant degree of disorder. In that case the intermediate phases would themselves display a significant, but small, stoichiometric variability.

Both the thermodynamic and the kinetic aspects of this reasoning claim attention. According to the Third Law $S \rightarrow 0$ as $T \rightarrow 0^\circ\text{K}$, for a crystalline solid. For a nonstoichiometric crystal, the entropy has a thermal part, which goes to zero, and a configurational part which is independent of the thermal motion of the atoms. Thus as $T \rightarrow 0$, $S \rightarrow S_{\text{config}}$; there is a residual entropy, as there is for a frozen glass, and the nonstoichiometric state is strictly metastable at the absolute zero. Attainment of true equilibrium involves either complete ordering or disproportionation into a pair of ordered phases with zero configurational entropy. It must be inferred that nonstoichiometric compounds should give place to stoichiometric 'line phases' in all systems at low temperatures and that they are inherently high temperature phases. Attainment or order, or disproportionation into a pair of ordered phases, however, involves diffusion of vacancies or interstitial atoms within the crystal. Solid state diffusion processes have high activation energies in almost all systems except the metal hydrides, and they are effectively frozen at low relative temperatures. The possibility of achieving the low temperature equilibrium state thus depends kinetically on there being a range of temperature below some order-disorder temperature but above the temperature at which diffusion effectively ceases. The latter is purely on operational definition, measured by the mean diffusion length over which atomic place exchange processes are operative during a particular measurement. It may clearly have a different significance in a laboratory experiment or over geological time. Herein lies part of the significance of mineralogical chemistry, as providing evidence of the crystal structures that approximate to ideal equilibrium.

III. GROSSLY NON-STOICHIOMETRIC COMPOUNDS

It is now practicable to consider the occurrence of non-stoichiometry amongst the compounds of the transition, lanthanide and actinide elements in the light of the principles set out above. If point defects are to be replaced by some newer concept, the first

question arising is the real structure of defects in crystals. Only when this is known is it possible, in principle, to formulate the internal and external equilibria of the crystal in thermodynamic terms. At the present time, the real nature of defects or defect centres is known only for a few groups of substances. Good thermodynamic measurements have been carried out on very few systems, almost entirely restricted to a few nonstoichiometric oxides. With some reservations imposed by the lack of firm knowledge, it is possible to group the nonstoichiometric compounds into:

- (i) Systems which must still be formally described in terms of point defects.
- (ii) Systems based on defect complexes, randomly arranged.
- (iii) Systems in which ordering processes dominate the observed behaviour.

Systems With Apparent Random Point Defects. Into this category come the grossly non-stoichiometric carbides, nitrides etc. of the transition metals, the metal hydrides and a few other groups of binary compounds. It is probably not accidental that these compounds are themselves metallic in character. The creation of a point defect - for example, a carbon vacancy in the rock salt structure of TiC_x ($0.5 < x < 1.0$) - does not involve a change in ionisation states or coulomb forces between altervalent ions, but rather changes in the Fermi energy resulting from a change in the electron : atom ratio. There is no good theory to define how the density of states function will be modified by changes in the occupancy fraction of non-metal sites, or by the creation of vacancies on metal atom sites, so that only a very general theoretical treatment of the stability and thermodynamics of such systems is possible. There will be interaction forces between atoms or vacancies, but these are likely to be short-range forces, approximately expressed by interactions between nearest neighbour sites only, in the form of a site preference energy \underline{W} :

$$\underline{W} = E_{CC} + E_{VV} - 2E_{VC} \quad (28)$$

where E_{CC} , E_{VV} and E_{VC} are the contributions to the total energy of the system by an adjacent pair of occupied sites, an adjacent pair of vacancies, or a vacancy next to an occupied site respectively. The behaviour of the system then depends on sign and magnitude of \underline{W} . If $\underline{W} > 0$, the vacant sites effectively repel one another and are randomised as far as is consistent with maximising their average spacing. For suitable rational fractions of vacant sites, and at lower temperatures, the minimum repulsive energy can be attained by regular ordering. This situation appears to hold for the non-stoichiometric carbides, which are truly random structures at high temperatures but may form discrete superstructure phases at lower temperatures. Thus vanadium carbide VC_x ($0.73 < x < 0.87$) gives rise to several superstructure phases, based on multiples of the rock

salt sub-cell (e.g. V_8C_7 or $V_{32}C_{28}$, a $2 \times 2 \times 2$ superstructure), with the vacant carbon sites ordered in a helical array.

If $W < 0$, sites of like occupancy effectively attract each other and tend to cluster³. At high temperatures, the distribution is random, but perturbed by a measure of site pairing. Below some critical temperature directly related to W , however, the system unmixed into a pair of coexisting phases: these phases, arising from the clustering of filled sites and of empty sites respectively are themselves phases of variable composition. As the temperature is raised, their compositions approach one another until they merge at the critical temperature, demarcating a closed 2-phase univariant region of the equilibrium diagram. This type of behaviour, with a critical temperature above which there is a very broad non-stoichiometric range, is encountered in a number of non-stoichiometric systems, and is rather representative of the behaviour of the metal hydrides⁴. Thus the lanthanides form two limiting hydrides: LnH_2 , with the fluorite structure and all hydrogen atoms on tetrahedral sites, and LnH_3 , in which the vacant octahedral sites are also occupied. Between these is a non-stoichiometric range of composition, with the octahedral sites partially occupied, and in this the structure could be described equally well as of the LnH_2 fluorite type with interstitial hydrogen, or as of the LnH_3 type with octahedral vacancies. The site preference energy or interaction energy W is small in these and in hydride structures generally - of the order of 1.5-2 Kcals.mole⁻¹, and the critical miscibility temperature is low. Below T_C , the system can unmix into a defect fluorite phase LnH_{2+x} and a complementary defect LnH_{3-y} phase, the chemical potential of hydrogen being the same for two defect configurations, corresponding to two compositions in equilibrium below T_C . For systems such as this, the introduction of a nearest neighbour interaction energy into the statistical thermodynamic treatment reproduces the observed properties quite well.

In regarding non-stoichiometric compounds of these kinds as point defect systems, it must be borne in mind, however, that they are dynamic, not static systems: under the conditions of true equilibrium, self diffusion processes are occurring freely in the crystal, and the average life time of an atom on a particular site is short. Thus the hydrides, in particular, give narrow line proton n.m.r. signals, and the self diffusion rate has been measured, for a number of metal hydrides, from the resonance line width. With a short life time on any site, and short range interaction forces, all long range order is lost; what is not so certain is whether, on the short time scale imposed by the self diffusion processes, there is some vestige of short range order - some measure of correlation between the occupancy of nearest neighbour or near neighbour sites. In the carbides, at least, the transition from a non-stoichiometric phase at high temperatures to ordered phases at low temperatures

suggests that there may indeed be some short range ordering, which disappears progressively as the site life time decreases.

Another group of substances that can be included in the concentrated point defect type is that of ternary cage or tunnel structures, represented by the tungsten bronzes M_xWO_3 or the titanium bronzes Na_xTiO_2 . In these, the host oxide provides an invariant framework, with holes or channels large enough to accommodate atoms of the alkali metals or other highly electropositive elements. The best known compounds of this type are the cubic tungsten bronzes M_xWO_3 ($M = Na, Li, Ca, Cd, Pb$, lanthanide metals etc.) which, in structural terms, are defect perovskites ABO_3 with incomplete occupation of the A sites. Alternatively, they can be described as having the ReO_3 structure with a proportion of the 12-coordinate sites filled. In conformity with what is known of the band structure of the ReO_3 type, they are metals: the atoms on the A sites contribute their valence electrons to the conduction band of the crystal, and the maximum proportion of A sites that can be occupied depends on the number of electrons per A atom introduced into the conduction band. Thus for Na_xWO_3 , $0.4 < x < 0.95$; for the lanthanide bronzes (3 electrons per lanthanide atom), the maximum occupation of A sites is around 15 per cent. The relevant question is whether the A sites are occupied at random, or in an ordered manner. There is some evidence that, in the high-sodium bronzes (e.g. in $Na_{0.75}WO_3$) there is a high degree of ordering, but at low A site occupancy it remains an open question whether the distribution is random, or whether there is clustering into ordered domains and vacant domains.

Defect Complexes. The implicit assumption in a point defect approach is that the creation of a vacancy or introduction of an interstitial atom produces only minimal relaxation effects - small shifts of atomic positions resulting from the redistribution of coulomb forces - in the surrounding structure. This assumption underlies the calculation of defect energies, but its validity must now be questioned, at least for defect concentrations large enough to introduce significant interaction effects. The real structure of a defect centre has been established only for a very few non-stoichiometric phases, and can be inferred with some confidence for a few more: in each case the point defect concept has to be replaced by that of a complex, involving association of point defects and valence defects, changes in coordination and displacement of atoms from the sites of the idealized parent structure, and the size of the defect complex embraces one or several unit cells of the parent structure. Although the evidence relates at present only to oxide structure of two types, the principle may well be general.

Modifications and Extensions of Point Defect Theory. In its usual form, point defect theory takes no account of interactions; the Kröger-Vink treatment can be extended without difficulty to

include such processes as the pairing or more extensive association of vacancies on next neighbour sites, or other interactions that can be expressed as quasi chemical equilibria. To each of these an equilibrium constant can be assigned, and the resulting concentrations can be included in the equilibrium diagram. At this level, however, the treatment is essentially empirical, and not based on rigorous statistical thermodynamics. In principle, a rigorous treatment could be framed by including all pairwise and multicentre interactions, but such an approach is at present quite impracticable.

Consider a crystal with $N (= N_A + N_B + \dots)$ atoms, containing N_D defects. N_D represents all the different kinds of 'anomalously' occupied lattice sites - vacant sites, interstitial atoms etc. - and is not necessarily the number that might be inferred from a simple consideration of the stoichiometry (compare the titanium oxide system with high concentrations of vacant sites on both the metal and the non metal sublattices, or UO_2+X and its defect complex, with the inferences that might be drawn purely from a comparison of the X-ray density and the measured density). In the usual treatment, the partition function (and hence the free energy) of the crystal is derived from a random arrangement of defects:

$$G(N, N_D, T) = G(N) + G(N, N_D) \quad (29)$$

and is taken to be independent of the particular way in which defects are distributed over the lattice sites. If, however, there are interactions between defects, both the enthalpy and the configurational entropy may depend on the actual distribution of defects. Denoting a particular configuration of defects as $\{N_D\}$, the free energy should more correctly be expressed as

$$G(N, N_D, \{N_D\}) = G(N) + G(N, N_D) + G(\{N_D\}) \quad (30)$$

where the first term is the free energy of the perfect crystal, the second terms is the free energy of defect formation as calculated for an interaction - free system, and the third (the configurational free energy) contains the effects of interaction terms, manifested in the formation of defect clusters, short range ordering and, in some cases, superstructure ordering.

The problem, therefore, is to express the configurational free energy in a meaningful and tractable way. This is a difficult problem. Three basic approaches have been used. The first is that referred to above in connection with the metal hydrides, in which a nett attractive interaction energy is assumed to operate between vacancies or interstitials on nearest neighbour sites. It amounts to assuming that the energy of defect creation varies with the defect concentration $[N_D]$, but that the configurational entropy is independent of $[N_D]$. It leads to an interpretation of limited

phase ranges and the closure of the miscibility gap. The second approach is to assume a repulsive energy between defects on nearest neighbour sites: in the extreme form, each defect excludes a certain number of sites from occupation⁵. This is equivalent to the statement that the energy is a defect on any permitted site is the same but that the configurational entropy is reduced, and is no longer given by a simple permutation of defects over all the sites of the crystal.

The third line of approach is to attempt to allow explicitly for the variation of interaction energy with the distance between point defects or defect clusters - in ionic crystals the interactions between defects and alter-valent ions are long range coulomb forces - and for the fact that, at equilibrium, the defects will take up such positions in the crystal as will minimize the free energy. Expressed in rigorous terms, this is a problem closely related to the Ising problem, which cannot be solved for a three dimensional case. Nevertheless, an attempt to find a computational solution has recently been made by Atlas⁶. For the procedure used, reference should be made to the original papers, but it may be noted that the key idea is that, in a crystal lattice, the periodicity of the structure implies that the interaction energy between any pair of charged defects take on some one of a discrete set of values. The crucial step in the argument is the procedure - open to criticism - whereby the resultant interaction energy of any one defect with all the others in its environment is transformed into a single-valued quantity. Since this single valued nett interaction energy ascribed to a particular defect also determines a volume within which no other defect can be situated, the resulting expression for the free energy specifies not only how the chemical potential varies with composition, but also the configurational entropy. In general, the distribution optimises itself on a radial distribution function, but for certain defect concentrations the free energy may be minimized by an approach to order: such effects show up in a substantial decrease in the configurational entropy at certain concentrations. Atlas developed the theory for defective fluorite structures, and applied it to the measured thermodynamic data for the oxygen-deficient CeO_{2-X} and oxygen-excess UO_{2+X} phases. Theoretical models which leave a number of freely adjustable parameters have little meaning; in this case, all the relevant quantities could be fixed by fitting the theoretical expressions to two experimental points, chosen at will anywhere in the T-p-X diagram. On this basis, the theoretical model fitted the observed behaviour of both systems over the whole range of temperature and composition quite well. In the CeO_{2-X} system, below 900°K, it reproduced the observed unmixing into ordered phases; in both systems the behaviour of the configurational entropy, as a function of composition, suggested that some substantial measure of order was retained, around the compositions of the intermediate phases,

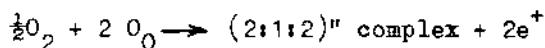
even in the broad non-stoichiometric ranges CeO_X ($1.7 < x < 2.0$) and UO_{2+x} ($0 < x < 0.25$). How far these conclusions are reliable depends, however, on the validity of the critical steps in the development of the statistical thermodynamic model.

Non-Stoichiometric Phases Based on Defect Complexes. As has already been indicated, our knowledge of the real microstructure of non-stoichiometric compounds is very meagre. There is some direct and indirect evidence for the defect structure in the ferrous oxide (wüstite) phase, which may perhaps be typical of ionic oxides with the rock salt structure, and in the non-stoichiometric phases derived from the fluorite structure. In each case, the point defect is to be replaced by a highly structured group of associated defects - a defect complex - indicative of very strong interactions.

The "wüstite phase will be considered rather fully in Section IV, as a 'case history'. In this place we shall consider the defect structure of the fluorite-related phases. The fluorite structure can be described in terms of a face centred cubic cell with cations at the 000 and related positions, and anions at the tetrahedral sites between them ($\frac{1}{4}\frac{1}{4}\frac{1}{4}$, $\frac{1}{4}\frac{3}{4}\frac{1}{4}$ etc.). A set of octahedral sites at $0\frac{1}{2}0$ etc. is therefore unoccupied. The defect properties (intrinsic disorder, diffusion, ionic conductivity) of CaF_2 itself and the fluorite-structure oxides can be interpreted, at least at fairly low relative temperatures T/T_m (T_m - melting point, °K) as arising from disorder in the anion sublattice only. Anion-excess fluorite phases MX_{2+x} are represented by the UO_{2+x} phase, solid solutions of YF_3 in CaF_2 etc.; anion deficient fluorites by the CeO_{2-x} , PrO_{2-x} the high temperature UO_{2-x} , the solid solutions of rare earth oxides with ThO_2 , solid solutions of rare earth oxides and oxyfluorides $\text{MO}_{1.5+}$ MOF, the CaO or Y_2O_3 stabilized zirconia solid solutions etc. In every case, the density and cell dimensions are compatible with a description in terms of point defects in the anion sublattice only - interstitial anions or anion vacancies respectively. The cation sublattice remains fully populated, and only minimal shifts in the cation positions appear to be involved.

In the hyperstoichiometric fluorite phases, therefore, the point defect description postulates interstitial anions, and the natural assumption was that up to one additional anion could enter the unit cell at the $\frac{1}{2}\frac{1}{2}\frac{1}{2}$ position. This assumption, however, encounters the difficulty that it implies a 30 per cent reduction in the normal anion-anion distance. From a neutron diffraction study of $\text{UO}_{2.12}$ and of the fully ordered phase U_4O_9 (in which there is one additional oxygen atom in each unit cell, but with a complex superstructure), Willis deduced that the actual structure is more complex⁷. The additional anion is not located at the $\frac{1}{2}\frac{1}{2}\frac{1}{2}$ position;

instead, it is displaced by about 1 Å in the [110] direction. At the same time, two adjacent oxygen ions are displaced from their fluorite sites, by similar distances along the [111] directions (Fig. 6). This is a minimum description of what is no longer a point defect but a defect complex or extended defect involving the reconstruction of a whole unit cell. As described, it involves three interstitial atoms, in two kinds of site, and two anion vacancies, and may be symbolised as a 2:1:2 complex. The complex may be more extended than has been described - e.g., a 3:2:3 complex - but the general principle is not open to doubt. In UO_{2-x} , such extended defects are randomly distributed; in U_4O_9 they are ordered. The complex, embracing a whole unit cell, no longer has cubic symmetry, and it is in part from the correlations between defect complexes of different orientation that the complicated and unsolved superstructure of U_4O_9 (a 4x4x4 cubic cell, with 576 oxide ions) results. Even in the complex as deduced by Willis, the interatomic distances are short. We may therefore enquire whether the positive holes:



are trapped as O^- ions, thereby reducing the effective ionic radii,

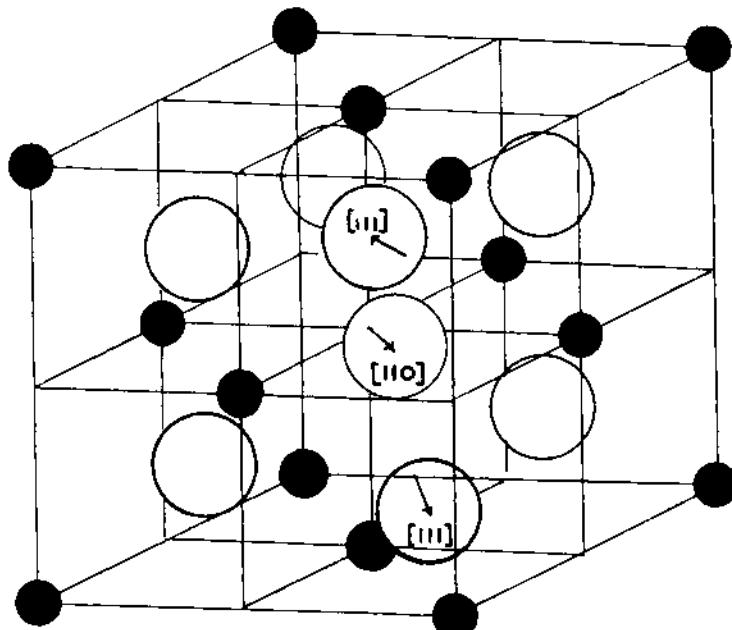


Fig. 6. The 2:1:2 interstitial 'Willis cluster' in UO_{2+x}

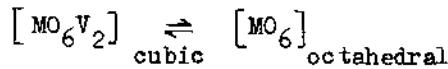
rather than as U^{5+} ions. Against this is evidence that the formation of such defect complexes does not depend on variable valency of the component ions. From a study of diffuse neutron scattering - probably the most sensitive technique for determining the structure of isolated defects - Fender and Cheetham, at Oxford, have found that in the $\text{CaF}_2\text{-YF}_3$ solid solutions the interstitial fluorine anions create the same kind of defect complexes as were identified by Willis in UO_{2+x} . The defect complex may, therefore, be the general description for hyperstoichiometric fluorite structure, at least in highly ionic compounds where coulomb interactions are dominant. It has, however, been established from n.m.r. studies that in dilute $\text{CaF}_2\text{-YF}_3$ solid solutions, with ca 10^{-4} of additional anions, the extra fluorine atoms are indeed on octahedral sites. The transition from point defects to extended defects thus appears to be a concentration effect, arising from generalized long range, and not purely local or steric, interactions.

For the hypostoichiometric fluorites, MX_{2-x} , the evidence is indirect, but strongly indicative of association of empty anion sites in pairs about the same cation. The fluorite structure can be regarded as built up from cubic $[\text{MO}_8]$ polyhedra, linked through their corners. The extreme of anion deficiency is represented by the fully ordered Type C rare earth oxides, with a $2 \times 2 \times 2$ superstructure in which the empty anion sites form non-intersecting strings along the four [111] directions of the crystal. Every cation is then 6-coordinate, at the centre of distorted octahedra of oxygen, which may be represented as $[\text{MO}_6\text{V}_2]$ and are of two kinds. Between this fully ordered structure and the perfect fluorite structure are the mixed valence oxides LnO_{2n-2} and the ternary oxides closely related to them, and the non-stoichiometric oxides LnO_{2-x} of Ce, Pr, Tb and U; closely related in structure are the pyrochlores $\text{A}_2\text{B}_2\text{O}_7$ and their solid solutions. In both kinds of derivative structure, the essential feature is the six coordinate cation, derived from the $[\text{MO}_8]$ polyhedron of the fluorite structure by association of vacant anion sites in pairs. Which structural course is taken as the anion: cation ratio changes - e.g. as the ratio of M^{3+} to M^{4+} in MO_{2-x} , or the ratio of O^{2-} to F^- in $\text{MO}_{1.5}\text{-MOF}$ solid solutions is changed - depends upon the characteristic stereochemistry (the site preference energy) of the cations.

If the site preference energy is small or zero - i.e. if all cations are identical as in $\text{MO}_{1.5-x}\text{F}_{2x}$ ($0 < x < 0.5$), or similar and interconvertible as in CeO_{2-x} - relaxation of anions from the sites of the fluorite structure is small; the 6-coordinate cation remains in an essentially cubic environment, but its coordination polyhedron has two empty anion sites at cube diagonal corners. This automatically implies that six of the cubic coordination polyhedra linked to it are of the type $[\text{MO}_7\text{V}]$, with 7-coordinate cations^{8,9}. The

resulting complex (Fig. 7), which might be symbolized $\{[\text{MO}_6\text{V}_2][\text{M}'\text{O}_7\text{V}]\}_6$ appears to be the basic unit of the anion deficient fluorites. It is itself a space-filling unit; packed together with their trigonal axes aligned, these units build up the $\text{MM}'_6\text{O}_{12}$ structure of Pr_7O_{12} , Y_6UO_{12} etc. Linked by corner sharing with $[\text{MO}_8]$ polyhedra, they may build up the multiplicity of ordered phases now recognised in the composition range $\text{MX}_{1.71} - \text{MX}_{2.0}$ ($X = \text{O}$ or $\text{O} + \text{F}$) - implying order both in spatial distribution and in orientation - or, if disordered, the non-stoichiometric phases and solid solutions.

If two species of cations are present with sufficient differences in ionic radius, relaxation may be more far reaching, but still depends upon the association of vacancies in pairs. The pyrochlore structure $\text{A}_2\text{B}_2\text{O}_7$ ($= \text{MO}_{1.75}$) is very closely related to that of fluorite, but the A cations are now in a regular octahedral environment, the B cations in a cubic environment. A pair of vacancies must again associate about one cation, followed by a relaxation process which could be symbolized



and which involves a significant shift in some of the atomic

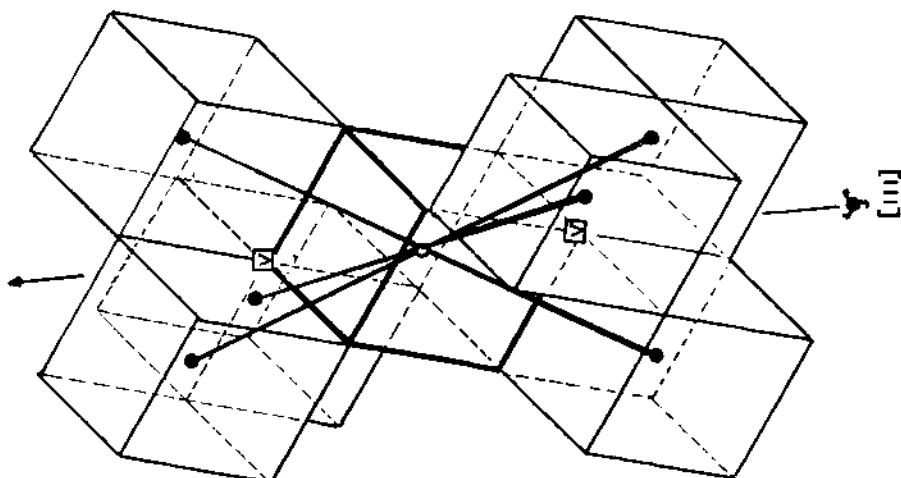


Fig. 7. The defect cluster in anion-deficient fluorite structures, as postulated by Bevan. Pairing of vacancies forms a unit of 7 cubic coordination polyhedra: MO_6V_2 linked to 6 MO_7V .

positions. The smaller cation adopts the octahedral coordination, but if the difference in ionic radius between A and B cations is not large, the energetic preference for one or other environment is rather small. The displacement of atoms from the fluorite sites is also quite small.

These factors can be identified as governing the properties of a number of fluorite-related ternary oxide systems. Coherence between an element of pyrochlore structure and the fluorite structure is possible because only small atomic displacements are involved. Such is probably the nature of the defect complex in the oxygen-deficient stabilized zirconias. There is some evidence that in lime stabilized zirconia, $\text{Ca}_x\text{Zr}_{1-x}\text{O}_{2-x}$, small domains of pyrochlore are formed, with an equilibrium size that depends upon the temperature.

The role of site preference energy can be seen in the rare earth - zirconia systems. The energetics imposes a preference, and not an absolute exclusion of cations from sites of different coordination. The stabilised fluorite phase has a certain composition range, probably incorporating rare earth cations in defect complexes of the kind indicated above. The pyrochlore phase also has a solid solution range, in which a proportion of either A or B cations must be in sterically unfavourable sites. At high temperatures, the behaviour depends upon the magnitude of the site preference energy. Where this is high (as in compounds $\text{L}_2\text{Zr}_2\text{O}_7$ of the early lanthanides, with large ionic radii), the pyrochlore phase is stable up to the melting point ($T > 2200^\circ\text{C}$), implying that the distinction between cubic [LO_8] and regular octahedral [ZrO_6] polyhedra is maintained: site preference energy is large as compared with thermal energies. As the radii of the lanthanide cations decrease along the series, so the site preference energy diminishes. The cations can become scrambled between the 8-coordinate and the 6-coordinate sites, and the distinction between the regular octahedral and the quasi-cubic [MO_6V_2] configuration for the latter becomes less significant. Since these depend on the site preference energy, these effects are highly temperature dependent. What is observed is that above some critical temperature, which decreases progressively along the lanthanide series, the pyrochlore phase transforms to a defect fluorite structure: the steric requirements of the cations are no longer able to displace the oxygen atoms from the averaged positions of the fluorite structure, and the defect structure reverts to that already discussed for anion-deficient fluorites (Fig. 8).

IV. FERROUS OXIDE: A CASE HISTORY

Phase Equilibria. Because of its technological importance,

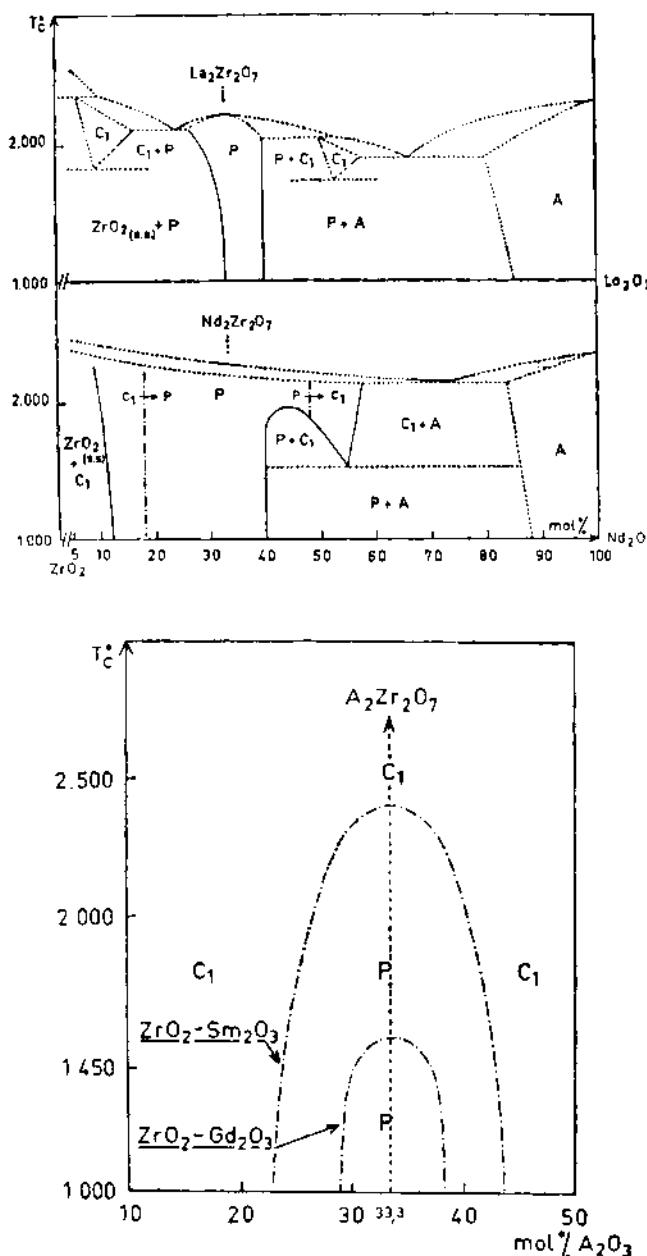


Fig. 8. (a) Stability of pyrochlore phases up to the melting point for light rare earths of large ionic radius. (b) Transformation pyrochlore defective fluorite at progressively lower temperatures as the ionic radius of the rare earth decreases.

its inherent interest and its deceptive appearance of simplicity, ferrous oxide or wüstite is one of the most intensively studied defect structures. The phase diagram is well established. Below 570°C, the wüstite phase is thermodynamically unstable; metallic iron and magnetite, Fe_3O_4 , coexist as a stable phase pair. Above 570°C, the wüstite phase has a progressively broadening composition range. The iron-rich limiting composition changes little, from $\text{FeO}_{1.059}$ at the peritectic point to a maximum iron content $\text{FeO}_{1.049}$ at about 910°C. The oxygen-rich limit shifts progressively from $\text{FeO}_{1.059}$ at the peritectic point to $\text{FeO}_{1.200}$ in equilibrium with solid Fe_3O_4 and the liquid oxide at 1424°C. Thus the stoichiometric oxide lies outside the stable composition range: the solid is inherently highly defective.

Oxide of any composition within the stable existence range therefore undergoes a phase reaction when it is cooled, unmixing into iron and Fe_3O_4 . Such a solid state reaction is dependent on diffusion and, since the activation energy for solid state diffusion processes is high, the rate of the unmixing reaction is strongly dependent on the temperature. Diffusion is, in fact, relatively slow, and by rapid cooling the wüstite structure can be retained at the ordinary temperature as a quenched, metastable species. Many of the structural and other properties of wüstite have been investigated with the quenched, metastable material. It must be borne in mind, however, that some diffusion and adjustment towards the low temperature equilibrium state takes place during cooling. Even with the highest attainable quenching rates, the mean diffusion path during quenching can be shown to be, in most cases, of the order of tens of Angstroms. Structural conclusions drawn from experiments on quenched samples of any high temperature non-stoichiometric phases must therefore be treated with some reservations; no measurements on quenched samples are a complete substitute for investigations carried out under equilibrium conditions. In the case of wüstite, diffusion is important at quite low temperatures; changes resulting from diffusion, considered below, become apparent on annealing at temperatures as low as 150°C, and proceed relatively rapidly above 350°C.

Thermodynamics. The relative stability of phases, and the stability range of a non-stoichiometric phase, is best discussed by reference to free energy - composition diagrams. Consider any phase of composition A_mB_n . The composition can be expressed as molefractions $x_A = (m/m+n)$, $x_B = (n/m+n)$ and the standard free energy in terms of the partial molar free energies of the components:

$$\Delta G^\circ = x_A \bar{G}_A + x_B \bar{G}_B \quad (31)$$

in place of ΔG° , the free energy per mole, we may introduce a

quantity $g = (1/mn)\Delta G^\circ$, the free energy per atom. Then for any phase of variable composition (and in terms of defect theory that implies, in principle, every crystalline phase) we may construct a plot of g against x_A (or x_B) (Fig. 9). It is shown in thermodynamic texts that a tangent to the free energy curve at any composition extrapolated to the ordinates $x_p = 1$ ($x_B = 0$), $x_B = 1$ ($x_A = 0$) has as its intercepts the corresponding partial molar free energies \bar{G}_A , \bar{G}_B respectively.

A line phase has an extremely narrow paraboloidal free energy curve; its free energy varies steeply with changes of composition, and the tangent point shifts very little on the composition axis even for large changes in the equilibrium chemical potential of its components. For a non-stoichiometric phase, the free energy curve is broad and usually asymmetric; with changes in chemical potential, the tangent point sweeps across a significant range of compositions.

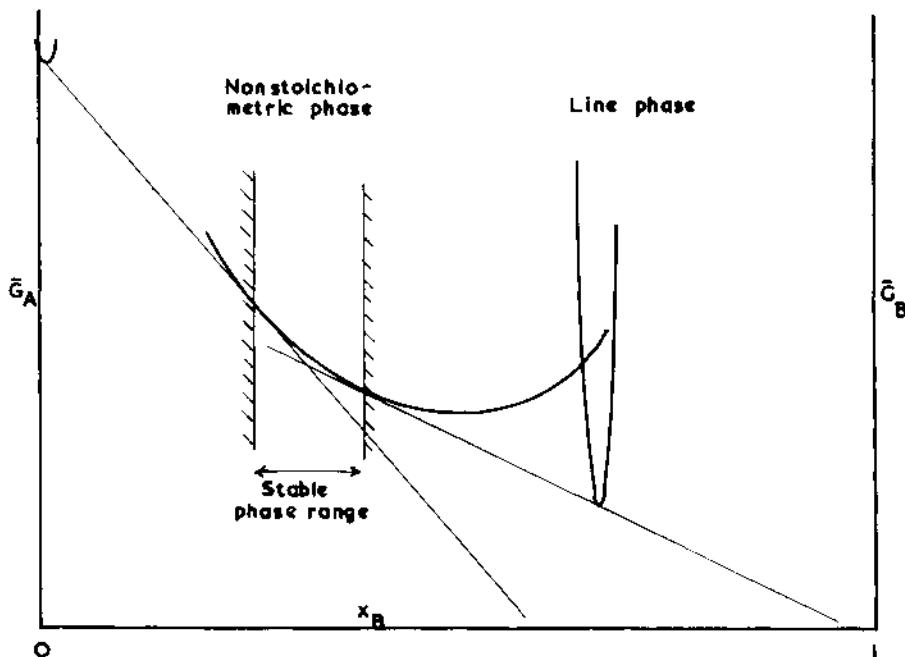
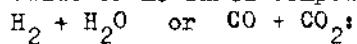


Fig. 9. Free energy isotherm for system with a nonstoichiometric phase in equilibrium with a line phase or with component A.

The condition that two phases - e.g. the successive compounds in a binary system - should coexist in equilibrium in that the chemical potentials of the components are the same in both phases. From the free energy diagram it is clear that this implies that they have a common tangent. Quite apart from any question of absolute stability, it is evident that the stable phase range of any non-stoichiometric phase is determined by the compositions at which common tangents can be drawn to the free energy curves of the next higher and next lower phases in the phase succession. In the case of wüstite this means the compositions at which the chemical potentials of iron and oxygen attain the values that they have in Fe_3O_4 and in metallic iron (saturated with oxygen) respectively. In so far as the entropy per atom differs in the different phases - and here the configurational entropy of the nonstoichiometric phase is the important factor - the relative positions of the free energy curves alters with temperature. For wüstite above 570°C, the situation is as in Fig. 9; below 570°, the free energy curve of wüstite lies wholly above the common tangent to the $\text{Fe}(0)$ and Fe_3O_4 free energy curves; the phase is metastable, but obtainable. Similarly, it cannot be assumed that the stoichiometric $\text{FeO}_{1.00}$ is incapable of existence, but rather that this composition may correspond to a point in the unstable region of the diagram (see Section IV).

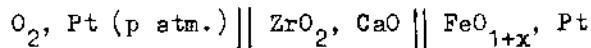
Whilst it is the free energy that strictly determines stabilities, the quantities that are accessible to direct measurement are the partial molar free energies G_A , G_B , or chemical potentials μ_A , μ_B . For wüstite, and for numerous other oxide systems, these have been determined by two main methods: (a) By equilibrating oxide of measured composition with a suitable gas mixture - e.g.



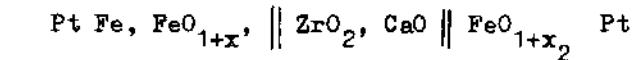
$$1. \quad 2\text{H}_2\text{O} \rightleftharpoons 2\text{H}_2 + \text{O}_2 \quad P_{\text{O}_2} = K_1 \left(\frac{P_{\text{H}_2\text{O}}}{P_{\text{H}_2}} \right)^2$$

$$2. \quad 2\text{CO}_2 \rightleftharpoons 2\text{CO} + \text{O}_2 \quad P_{\text{O}_2} = K_2 \left(\frac{P_{\text{CO}_2}}{P_{\text{CO}}} \right)^2$$

(b) By using solid state galvanic cells. Certain oxides, particularly the defect fluorite structure such as ZrO_2 stabilized with CaO , or ThO_2 doped with Y_2O_3 , are pure unipolar ionic conductors below 1000-1300°C, at not too low oxygen potentials, and conduct the current by transport of O^{2-} ions. They may therefore be used as the intermediate electrolyte in high temperature galvanic cells such as



or



Reference electrode I Working electrode II

The cell reaction involves the transport of one mole of O_2 per 4 Faradays of electricity, between a reference electrode at a constant, known oxygen potential and an electrode at which the oxygen potential is to be measured. Thus, for the second cell shown, the measured e.m.f. is

$$E = \frac{1}{4} \left[\bar{G}_{O_2}(FeO_{1+x_1}) - \bar{G}_{O_2}(FeO_{1+x_2}) \right] = \frac{RT}{4} \ln \frac{P_{O_2}(FeO_{1+x_2})}{P_{O_2}(FeO_{1+x_1})} \dots \quad (32)$$

The galvanic cell method, in particular, has given results of great precision. The variation of \bar{G}_{O_2} with composition of the oxide has been determined either from cells with successively varied compositions, or by progressively producing small changes in the composition of the oxide in the working electrode by coulometric titration - i.e. by passage of a measured quantity of electricity through the cell by applying an external e.m.f.

At 1100°C , the equilibrium oxygen pressure changes from $10^{-13.2}$ to $10^{-10.7}$ atmospheres across the wüstite phase range. At high temperatures, $\log P_{O_2}$ (ie μ_{O_2}) is a nearly linear function of the composition. Below 1000° , the dependence of chemical potential upon composition displays features that have to be correlated with the defect structure; these are considered in IV.4.

A Point Defect Interpretation. X-ray diffraction data for quenched wüstite indicate a face centred cubic structure, with cell dimensions a_0 about 4.5\AA ; the intensities of the diffraction lines are consistent with the rock salt, B1 structure. The unit cell of the B1 structure contains 4 molecules - i.e. the wüstite cell should contain 4 Fe and 4 O atoms. Since the Fe:O ratio is actually irrational, the measured unit cell dimensions must refer to an averaged cell, for which there are three possible ways in which the composition might vary

Table IV.1

Fe sub lattice	O sublattice	No. atoms per unit cell		Point Defect interpretation
		Z_{Fe}	Z_O	
I Perfect	Imperfect	4	> 4	Interstitial O
II Imperfect	Perfect	< 4	4	Fe vacancies
III Imperfect	Imperfect	< 4	> 4	Unbalanced Schottky type Defects

As the O:Fe ratio increases, the cell dimension decreases progressively. Jette and Foote¹⁰ established that the general defect structure of wustite is subtracted by measuring the cell dimensions and density for four compositions of wustite quenched from 1000°C. If the average unit cell contains m Fe and n O atoms, the theoretical density for a cell of side a Å would be

$$d_{Xray} = \frac{55.84m + 16n}{0.6023} \cdot \frac{1}{a^3} \quad (33)$$

Expressing the composition in terms of the extreme defect models I and II, the resulting X-ray densities could be compared with the observed values. The comparison is shown in Table IV.2 and Fig. 10a.

Table IV.2

Fe:O ratio	a_0 , Å	d_{obs}	d_I	d_{II}
0.945	4.301	5.728	6.074	5.740
0.93	4.292	5.658	6.136	5.706
0.92	4.285	5.624	6.181	5.687
0.91	4.282	5.613	6.210	5.652

For defect model III, the calculated densities would fall between the extreme cases I and II.

It is evident that the density of wustite accords closely with model II: it is an iron-deficient phase, to be represented as

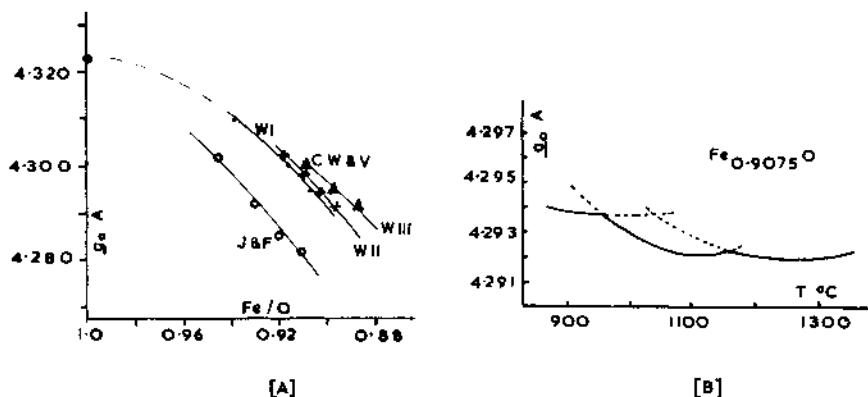


Fig. 10. (a) Cell dimensions of Fe_{1-x}O . Data of Jette and Foote and, for wustite quenched from different temperatures, of Carel, Weigert and Vallet. (b) Cell dimensions of $\text{Fe}_{0.9075}\text{O}$ quenched from different temperatures in the W-I, W-II and W-III ranges.

Fe_{1-x}O rather than as FeO_{1+y} and the necessary implication is that some sites in the cation sublattice are vacant. The most immediate inference is that the deficiency in iron (x in Fe_{1-x}O) represents the fraction of cation vacancies.

From Jette and Foote's results, a is a linear function of composition expressed in weight - per cent oxygen, but not linear in terms of the formula Fe_{1-x}O (or of FeO_{1+y}), as would be expected from the Vegard's law behaviour of most solid solutions and non-stoichiometric phases. The compositions examined by them were close to the iron-rich phase limit. When a wider range of compositions is examined, the variation of cell size with composition is found to be more complex. Levin and Wagner (1966) found that their results could be fitted to a polynomial from which the cell size of the hypothetical $\text{Fe}_{1.000}\text{O}$ can be extrapolated as $a_0 = 4.336 \text{ \AA}$. According to Carel, Weigel and Vallet¹¹, however, the cell dimensions depend systematically on the temperature from which a sample is quenched, as well as on its composition. Thus, for one and the same sample ($\text{Fe}_{0.9070}\text{O}$), shock quenched from temperatures in the range 900-1340°C, the values fall consistently on three distinct curved segments of the plot (Fig. 10b). For samples of different composition quenched from the same high temperature (1340°C), a_0 was a nearly linear function of composition expressed as FeO_{1+y} - i.e., if the composition is expressed in terms of the iron deficiency, Fe_{1-x}O , the relation must be represented by a polynomial expression. The curve fell, however, into two segments with a sharp break and change of slope at the composition

$\text{Fe}_{0.94}\text{O}^0$. These results, which can be correlated with inferences drawn from precise thermodynamic studies would imply that wüstite cannot be regarded simply as a solid solution of point defect vacancies in a B1 structure. They can be understood only if there is some ultramicrostructure involving internal equilibria which are themselves displaced by changes in the composition and the temperature of equilibration.

Defect Equilibria. If wüstite is based upon point defect cation vacancies, the transfer reaction for change of composition can be written



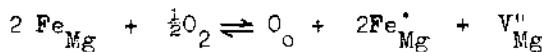
The positive holes would then be localised at cations occupying normal octahedral sites (Fe^{2+} converted to Fe^{3+}) as $\text{Fe}_\text{Fe}^{\cdot}$.

$$\text{Since } p = 2[\text{V}_\text{Fe}^{\text{u}}]$$

$$[\text{V}_\text{Fe}^{\text{u}}] = K p_{\text{O}_2}^{1/6} \quad (34)$$

This pressure dependence is observed for iron-rich wüstites, $\text{Fe}_{0.95}\text{O}-\text{Fe}_{0.90}\text{O}$ at temperatures above 1000°C ; the p -type conductivity of the samples follows the same law. At lower temperatures, however, and for wüstite richer in oxygen, the simple $p_{\text{O}_2}^{1/6}$ relationship is no longer valid.

There are also anomalies in the oxidation-reduction behaviour of solid solutions of FeO in other B1 type oxides - e.g. $\text{FeO}-\text{MgO}$ solid solutions. In such a solid solution, the activity a_V of cation vacancies is approximately the fraction of all cation sites vacant - i.e. is given by $[\text{V}^{\text{u}}]$; the activity of iron cations is no longer nearly unity, but is given by the fraction of all cation sites occupied by iron atoms. For dilute solutions of FeO in MgO this must be represented as $[\text{Fe}_\text{Mg}^{\text{u}}]_{\text{Total}}$. The transfer reaction must be explicitly written as



and the equilibrium is expressed by

$$\frac{[\text{V}_\text{Mg}^{\text{u}}][\text{Fe}_\text{Mg}^{\cdot}]^2}{[\text{Fe}_\text{Mg}^{\text{u}}]^2} = K p_{\text{O}_2}^{\frac{1}{2}} \quad (35)$$

or since $[Fe_{Mg}^{\cdot}] = 2[V_{Mg}^{\cdot}]$ and $[Fe_{Mg}^{\cdot}] = [Fe_{Mg}^{\cdot}]_{\text{Tot}} - [Fe_{Mg}^{\cdot}]$

$$\frac{[Fe_{Mg}^{\cdot}]^3}{\{[Fe_{Mg}^{\cdot}]_{\text{Tot}} - [Fe_{Mg}^{\cdot}]\}^2} = K' p_{O_2}^{\frac{1}{2}} \quad (36)$$

If we write, for the fraction of iron cations oxidised to the +3 state at a given oxygen pressure $[Fe^{\cdot}] / [Fe]_{\text{Tot}} = r$

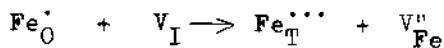
$$\frac{r^3}{(1-r)^2} [Fe_{Mg}^{\cdot}]_{\text{Tot}} = \text{const.} \quad (37)$$

It follows that as $[Fe]_{\text{Tot}}$ decreases, the degree of oxidation at a given oxygen pressure should increase. The reverse is the case experimentally¹².

The question then arises whether the equilibria would be better expressed involving associated defect complexes. Within the framework of the point defect model, such complexes could originate in the trapping of positive holes by vacancies, and would include such species as $\{Fe_{Fe}^{\cdot}, V_{Fe}^{\cdot}\} \{ (FeFe)_2 V_{Fe}^{\cdot} \}^*, \{ (FeFe)_n V_{Fe}^{\cdot} \}^{(n-1)}$ etc. but it has proved impossible to get a self-consistent interpretation of the behaviour of wüstite and its solid solutions on this basis.

The newer structural information considered in IV.5 suggests, however, that account should be taken of local reconstructive relaxation about the defects - e.g. the shift of Fe_{Fe}^{\cdot} atoms from the octahedral sites of the B1 structure to interstitial sites on which the coordination number is changed from 6 to 4.

This may be represented



where $Fe_O^{\cdot} = Fe_{Fe}^{\cdot}$, a Fe^{3+} atom on a normal octahedral site, and $Fe_T^{\cdot\cdot\cdot}$ is a Fe^{3+} atom on a 4-coordinate, tetrahedral interstitial site. We now have

$$[V_{Fe}^{\cdot}] = 2[Fe^{\cdot}]_{\text{total}} + [Fe_T^{\cdot\cdot\cdot}] \quad (38)$$

and the resulting complex $(Fe_T^{\cdot\cdot\cdot}, V_{Fe}^{\cdot})^*$ can enter into further complexing reactions by association with vacancies, clustering etc. The first mentioned reaction would give the negatively charged complex $\{(Fe_T^{\cdot\cdot\cdot})(V_{Fe}^{\cdot})\}^-$ and, by electron trapping, its neutral state, effectively $2\{(Fe_O^{\cdot})(Fe_T^{\cdot\cdot\cdot})(V_{Fe}^{\cdot})\}^*$. The structural evidence is that this species is likely to be an important one in the equilibria,

and Kofstad¹³ has shown that, if it may be further assumed that the occupation of one tetrahedral site blocks the occupation of the neighbouring sites, the equilibrium data at lower temperatures and higher oxygen contents can be reasonably well accounted for. Kofstad's site blocking assumption is not fully consistent with the structural evidence, and the inclusion of defect complexes in the equilibria cannot reconcile the observed and the predicted behaviour of FeO-MgO solid solutions. There is, in fact, evidence from their magnetic properties that the Fe³⁺ ions in these solutions are clustered together even in dilute solid solutions, contrary to the assumption made in any dilute regular solution treatment, that they are uniformly distributed.

It has been assumed thus far that the chemical potential of oxygen is a continuous, although not necessarily a linear, function of the composition across the phase field. Kofstad's analysis of the equilibria was made on that basis. In 1965, however, Vallet and Raccah concluded that the assumption is incorrect. From their measurements by a gas equilibration method they inferred that although the system is truly bivariant, the $\bar{G}(T,X)$ values reveal discontinuous changes with changes in composition and temperatures¹⁴. This conclusion has recently been confirmed by Fender and Riley, by very precise galvanic cell measurements¹⁵. The discontinuities were unambiguously shown by coulometric titrations which changed the oxide composition isothermally by very small increments. In isothermal measurements, the G-X curves show abrupt changes in slope; for oxide of constant composition, the \bar{G} -T curves are similarly divided into rectilinear segments by changes in slope, corresponding to small changes in $\Delta\bar{H}$ and $\Delta\bar{S}$.

The effect of these discontinuities in the partial molar thermodynamic properties is to subdivide the wüstite phase field into three areas, wüstite I, II and III (Fig. 11). The conclusion is inescapable that these reflect changes in the predominant defect structure. The boundaries between the regions correspond to 2nd order phase transformations, and these would appear to be more like order-disorder transformations than uniformly progressive shifts in the internal equilibrium.

The Real Defect Structure of Wüstite. Since the work of Jette and Foote it has been generally assumed that wüstite has the B1 structure, like MgO, with cation vacancies. This structure accounts for the diffraction pattern to a first approximation, but more recent work by neutron diffraction, X-ray diffraction and neutron scattering has shown that this description is oversimplified.

Roth¹⁶ determined the magnetic structure of wüstite by neutron diffraction, using material of composition $Fe_{0.944}O$ and $Fe_{0.926}O$

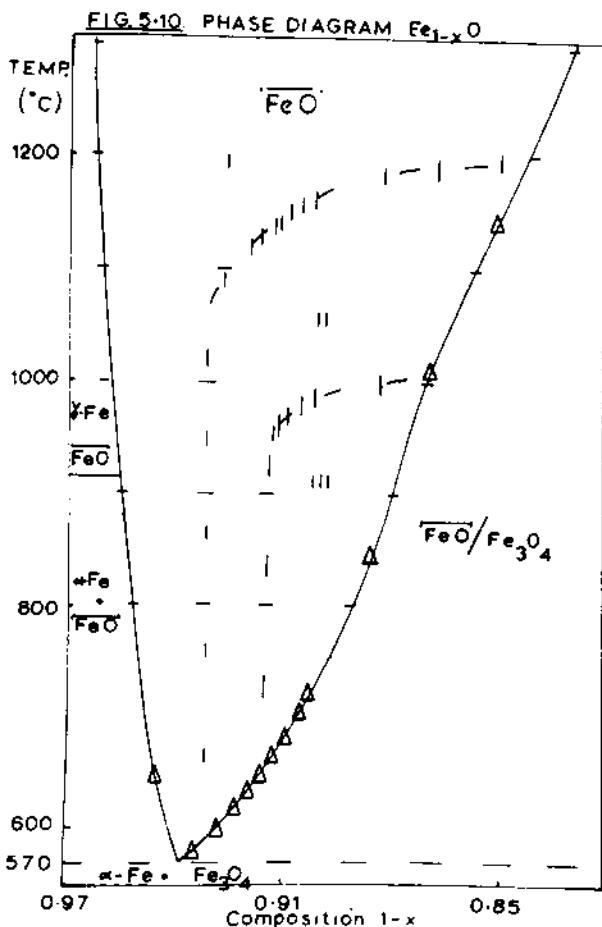


Fig. 11. Phase diagram of the Fe-O system, "wustite region.

prepared by quenching from 1000°C. The magnetic unit cell is cubic, with a side $2a_0$, and the spins are lines up normal to the (111) plane. At room temperature, the diffraction pattern arises both from nuclear scattering and from magnetic scattering by the paramagnetic Fe^{2+} and Fe^{3+} ions. Calculations based on a random distribution of Fe^{3+} ions and a vacancy concentration equal to the stoichiometric defect (i.e. on the point defect model) could not account for the measured absolute and relative intensities of the Bragg peaks; the data could be fitted only by shifting a fraction of the cations to the tetrahedral interstitial, sites, thereby creating new vacancies as represented by the equations given in Section IV. From the site occupancy fractions that best fitted the Bragg peak intensities, it could be inferred that $[V_{\text{Fe}}] \sim 2 [Fe_I]$ - i.e. each vacancy created by the primary transfer reaction traps a positive hole and is converted to a Roth Complex.

This Roth complex, $\{[\text{Fe}_T^{\text{2+}}(\text{V}_{\text{Fe}})^{\text{3+}}]\}$, with adjacent octahedral sites occupied by Fe^{2+} and Fe^{3+} ions, can be regarded as a small element of the Fe_3O_4 structure - one octant of the Fe_3O_4 unit cell. Since Roth found no coherent scattering from the cations on tetrahedral sites, it may be concluded that the complexes are randomly arranged. On the other hand, their concentration in oxygen-rich wüstite would be high, so that they must cluster together. The first and most obvious assumption was that these magnetite-like elements of structure would cluster to form microdomains of magnetite, coherently grown - and mobile - within the FeO structure.

That conclusion has not been borne out by subsequent work. Manenc first drew attention to the fact that the description of wüstite as a defect B1 structure ignored certain extra features in the X-ray diffraction pattern. Quenched specimens especially in the oxygen-rich region of the phase diagram display additional lines; on annealing at low temperatures the disproportionation reaction is preceded by the development of strong superstructure lines, indicative of some ordering process. The superlattice is cubic, and is an irrational multiple of the basic wüstite cell ($a_{\text{s.l.}} \sim 2.7 a_0$) so that the ordering must be imperfect and grouped about at least two values of the repeat distance within the structure. A more complete structural analysis was carried out by Koch and Cohen¹⁷. For $\text{Fe}_{0.904}\text{O}$, quenched from within the wüstite III range, they identified two sets of satellite lines, one of which changed during annealing and could be ascribed to pre-precipitation changes within the wüstite crystal. Single crystal X-ray work showed that the intensities of all lines could be accounted for only by the presence of large defect clusters - an association of 13 octahedral vacancies with 4 tetrahedral iron cations (Fe^{3+}) (Fig. 12). This amounts to complete reconstruction on one unit cell of the B1 structure into a unit, not of Fe_3O_4 structure, but of zinc blende-like structure. This cluster has a large net anionic charge, and must trap a balancing ionic atmosphere of Fe_{Fe} on the octahedral sites of the immediately adjacent planes of the crystal structure. Thus the real size of the defect cluster is a unit $2 \times 2 \times 2$ times the wüstite B1 cubic cell. If all the cation vacancies are involved in the formation of such clusters, $29\% (= 4/14)$ of the Fe^{3+} cations should be found in tetrahedral sites. Because the clusters are charged, coulomb forces impose some measure of ordering; Koch and Cohen found ordering on a superlattice of side $3.5 a_0$. High temperature X-ray work showed that the superlattice reflexions disappeared, but diffuse scattering remained, indicative of some persistence of short range order.

Independent support for the Koch-Cohen cluster model comes from the work of Childs and Fender at Oxford, using diffuse neutron

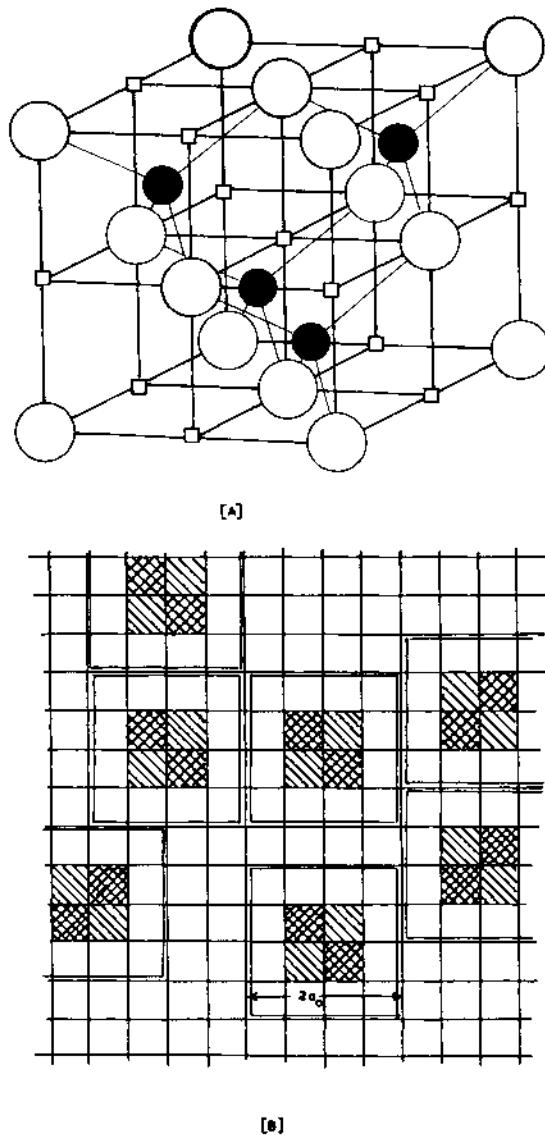


Fig. 12. The Koch cluster of 13 vacant octahedral cation sites and 4 iron atoms in tetrahedral sites. (b) Schematic arrangement of Koch clusters in the NaCl-type structure of FeO. Each cluster preempts a $2 \times 2 \times 2$ unit of the B1 structure. Sites not involved in the clusters or their balancing sheets of counter-ions remain as elements of unperturbed B1 structure. Clusters may be randomized or partly ordered on a superstructure.

scattering. From the Bragg law $\lambda = 2d \sin\theta$ it is clear that radiation of wave length $\lambda > 2d_{100}$ can give rise to no Bragg reflexions. Long wave length neutrons therefore produce no diffraction pattern, but are scattered incoherently from isolated defects. The angular distribution of scattered neutrons can then be compared with that calculated for the effect of alternative defect structures. Childs examined material with composition $\text{Fe}_{0.91}\text{O}$ to $\text{Fe}_{0.94}\text{O}$, and compared the scattering with that computed (a) for isolated vacancies; (b) for complexes of the type $\{\text{Fe}_0^{\cdot} \text{V}^{\cdot}\}^x$ as postulated by Flood; (c) for Roth clusters $\{\text{Fe}_T^{\cdot\cdot\cdot} (\text{V}_T^{\cdot}\text{Fe})_2^{\cdot}\}^{\frac{1}{2}}$; (d) for a quasi-element of Fe_3O_4 , $\{\text{Fe}_T^{\cdot\cdot\cdot} (\text{V}_T^{\cdot}\text{Fe})_4^{\cdot}\}^{\frac{1}{2}}$; and (e) the Koch-Cohen cluster, together with "multiclusters" formed by association of these units at various lattice spacings. The results showed clearly that, of the models examined, the defect centre must be identified with the Koch-Cohen cluster; also that the observed diffuse neutron scattering must be explained by the clustering of clusters, imperfectly correlated in spacing but corresponding, in the oxygen-richest samples examined, to a mean spacing of above 2.6 \AA_0 .

Results obtained by Koch and Cohen and by Childs both refer to wüstite quenched from inside the wüstite III range. For this material, the structure and size of the unit cluster is independent of the composition; change in the O:Fe ratio only alters the fraction of the total number of B1 lattice sites that is included within the clusters, and hence the average spacing of the clusters. There is some residual short range order of clusters and, at low temperatures (including the metastable state below 570°) a considerable degree of long range order is reached. In principle, the ordering of clusters with varying numbers of unperturbed FeO-like sheets of atoms between them should give rise to a family of metastable iron oxides, distinct from magnetite.

If the transitions within the wüstite range, $\text{WIII} \rightleftharpoons \text{WII} \rightleftharpoons \text{WI}$, involve changes in microstructure, and since they proceed from left to right as the temperature is raised or as the stoichiometric deficit of iron is decreased, the most immediate hypothesis is that they represent progressive dissociation of the defect complexes into smaller units. Wüstite III retains no more than imperfect short range order, so that the first stage of breakdown would presumably be the dissociation of Koch-Cohen clusters into smaller complexes. It is not yet clear whether the $\text{WIII} \rightleftharpoons \text{WII}$ transition is the conversion of Koch-Cohen clusters to Roth complexes, and $\text{WII} \rightleftharpoons \text{WI}$ the conversion of Roth complexes into what are, in effect, random point defects (as in defect carbides etc. - see Section III). Such an interpretation for the wüstite I region appears unlikely, but direct evidence is lacking.

Stoichiometric FeO and the Disproportionation of Wüstite. When

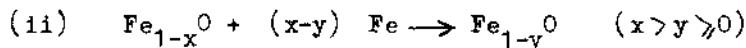
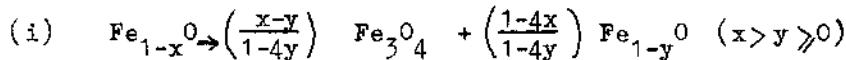
quenched, metastable wüstite is annealed at 150-500°C, diffusion is not too slow for internal adjustments of equilibrium leading ultimately to the disproportionation. Nucleation of Fe_3O_4 , and especially of metallic iron, is kinetically hindered, however, so that pre-precipitation processes can be observed.

Carel, Weigel and Vallet found that when oxygen-rich wüstite was slowly cooled, the diffraction pattern showed that two cubic B1 phases were present. One was an oxygen-richer wüstite, the other had $a_0 = 4.328\text{\AA}$, close to the extrapolated cell dimensions of stoichiometric $\text{Fe}_{1.00}\text{O}$. Koch and Cohen, and Manenc, found that strong superstructure ordering developed below the stable phase range. For the observed superstructure, this would imply the enrichment of some regions in Koch-Cohen clusters at the expense of others - i.e. a spontaneous partial unmixing into iron-rich and oxygen-rich domains, without involving nucleation. This is suggestive of a spinodal unmixing process, occurring by self-amplifying fluctuations of composition with a periodic distribution in space. Manenc found quite independent evidence for this unmixing domain structure by transmission electron microscopy¹⁸. Oxygen-rich wüstite, annealed at 150°C, formed a lamellar structure, with iron-rich and oxygen-rich domains separated by {100} walls. It would therefore appear that solid state diffusion processes, involving migration of tetrahedral iron atoms and cation vacancies, prepare for the disproportionation reaction by tending towards the perfect Koch-Cohen superstructure and, by depleting certain domains of defects, forming an oxide that approximates to stoichiometric FeO .

As stated earlier, the free energy corresponding to the composition $\text{Fe}_{1.00}\text{O}$ lies on the metastable extension of the free energy curve of wüstite. This free energy curve, however, relates to the phase equilibria under the customary conditions - i.e. under a total hydrostatic pressure of 1 atmosphere or less. The effect of pressure as an environmental variable for defect systems must be considered. For most purposes in solid state work, it is relatively unimportant whether the free energy used is the Helmholtz free energy $E-TS$ or the Gibbs free energy $E + PV + TS$ - i.e. whether the constant volume or a constant pressure system is considered. P here is the total pressure; if P is small, $P\Delta V$ is negligible compared with the other terms in the free energy change of any process. This is no longer true for pressures of the order of 10 kilobars; the compressibilities of solids are no longer negligible and changes in volumes arising from solid state reactions may make significant contributions to the free energy change, through the $P\Delta V$ term. Equilibria must be considered in terms of the Gibbs free energy. Schottky type point defect equilibria are non-conservative in lattice sites: they require an expansion of the crystal and this statement is valid even if the point defects created are themselves transformed into defect complexes. Defect equilibria that are

non-conservative must therefore be displaced at high pressures. Apart from some experimental results for the TiO and VO phases, little work has been carried out on this subject.

In the case of ferrous oxide, any displacement of the phase limits depends upon the free energy change at constant temperature in the reactions



(i) represents a possible shift in upper phase boundary composition by a pressure induced disproportionation, (ii) the shift in lower phase boundary composition by incorporation of iron - i.e. by annihilation of vacancies. Neglecting the compressibilities of the solid phases, which are inherently small and appear only as a difference term, ΔV_r can be calculated from the cell dimensions and densities, for the formation of stoichiometric FeO ($y=0$) by both disproportionation and reaction with iron. ΔV_r is < 0 for both reactions and is of the order -0.3 \AA^3 per molecule of Fe_{1-x}^0 ($x \sim 0.05$), so that equilibria must be displaced to the right at sufficiently high pressures.

Reaction (i) has not been studied. For reaction (ii), Katsura, Iwasaki and Kimura (1967) found that $\text{Fe}_{0.923}^0$ was converted to stoichiometric $\text{Fe}_{1.00}^0$, with $a_0 = 4.323 \text{ \AA}$, at 770°C under 36 Kbar pressure¹⁹. At 1000° , the same result is obtained at about 15 Kbar. The conditions of greatest interest, where the phase range is progressively shifted and narrowed, remain to be explored.

V. STRONG DEFECT ORDERING: INTERMEDIATE SUPERSTRUCTURE PHASES

Repeated reference has been made, in the previous sections, to the ordering of point defects or extended defects as a consequence of their interactions. In principle, the configurational entropy of a non-stoichiometric crystal could be removed (as required, by the Third Law, for true low temperature equilibrium) in two ways: by disproportionation into two phases with compositions and structures corresponding to two different, but definite, valence states of the metallic component, or by arranging the defects in a regular periodicity, so as to define a new set of unit translation vectors which would enclose some multiple of the unit cell of the parent structure. In the latter case, the product would be a mixed valence compound. If the new unit cell is large, the composition may be defined by the ratio of large numbers of metal atoms and non metal atoms respectively, and so may appear irrational. If

there are alternative superstructure arrangements, corresponding to different atomic ratios, there can be a succession of ordered intermediate phases with distinct unit cell formulae but closely similar atomic ratios. The change in chemical potentials and other thermodynamic properties from one phase to the next would necessarily be small. Their diffraction patterns would differ little, being dominated by reflexions from the sub-cell that is common to all - the unit cell of the parent structure.

A system of this kind would closely simulate the properties of a truly nonstoichiometric, thermodynamically bivariant phase. A sufficiently refined crystallographic study of well characterised material is probably the ultimate way of discriminating between the alternatives. This problem has been recognised during the past 10-15 years and, as a result, it has been established both that ordered intermediate phases are formed in many systems and that these may undergo reversible transformations into nonstoichiometric phases. In general, it is found that ordering in the intermediate phases is based on not too large a multiple of the parent structure sub-cell, but in principle every conceivable atomic ratio could give rise to an ordered structure by adopting a sufficiently large repeating unit. Establishment of superstructure order would then depend, however, on the balance between long range interactions and defect energies. Few kinds of long range interactions forces in crystals can be recognized. Coulomb interactions are the most important, and it is understandable that defect ordering is most important for compounds in which the binding forces are largely ionic - i.e. in which the mixed valence character can properly be described in terms of ions in different, localized charge states.

Both extended defects (defect complexes etc.) and point defects may undergo ordering. The most important series of intermediate compounds are those in which vacant sites on either the anion sublattice or, more commonly the cation sublattice, form a superstructure, and these form the basis for the following discussion. The essential features of such structures are (a) that every site of the superstructure corresponds to a site of the parent structure; (b) the sites of the cation (or anion) sublattice are now of two kinds: those that are properly occupied and those that are properly empty in the fully ordered superstructure.

It follows that the empty sites are no longer defect vacancies; they are essential elements of the structure. Superstructure ordering assimilates the defects into the regular structure. To avoid confusion, it is advisable not to use the term 'vacancy' or the symbol V for these empty sites which determine the translation vectors and symmetry of the new cell. They will be designated Z sites. Defect creation, disorder, and departure from ideal stoichiometry must now be considered in terms of equilibrium not of

a binary, but a ternary system.

The sequence of intermediate phases in the rare earth oxides $\text{CeO}_{1.5}$ - CeO_2 , $\text{PrO}_{1.5}$ - PrO_2 , $\text{TbO}_{1.5}$ - TbO_2 are of this kind, with Z sites and occupied sites in the anion sublattice while the cation sublattice is, to a first approximation, undisturbed. A second example, to be more fully considered, is the sequence of phases and structures formed in the chalcogenides of the transition metals, between the compositions $\text{MX}-\text{MX}_2$ or $\text{MX}-\text{MX}_{1.5}$. There is an extensive series of compounds of the elements from vanadium to nickel that represents a transition between the B8 (NiAs) and C6 (CdI_2) structures. The chalcogen atoms, in both end members and intermediate compounds, form a hexagonal close packed structure that is undisturbed, except for small dimensional changes, by changes in the number of cation sites filled. The cations occupy octahedral interstices, in hexagonal layers such that every (0001) set of octahedral sites is occupied in the B8 structure, giving the layer sequence -M-X-M-X-M-X--, while alternate layers of octahedral sites are occupied in the C6 structure, giving the layer sequence -X-M-X X-M-X X-M-X --. Since the non-metal lattice is invariant throughout the series, it can be omitted from the discussion; defect structures and ordering can be considered in terms of the cation lattice alone. In the C6 structure, filled cation sheets and sheets of empty, potential cation sites alternate along the c axis; in the series of compounds referred to, the C6 to B8 transition involves the progressive filling of the potential cation sites in alternating cation sheets, preserving the sequence -F-I-F-I--, where F is a filled and I an incompletely filled cation sheet. Formation of ordered phases is then essentially a 2-dimensional problem: the ordered distribution of M sites and Z sites within the hexagonal arrays of the incompletely filled sheets. In the extensive series of compounds now known to conform to the simple stacking principle, the ordered phases of known structure correspond to ordering with simple, rational ratios of M and Z sites²⁰. Table V.1 illustrates this, showing the distribution of M and Z sites not in the true unit cell prescribed by the plane group symmetry of the ordering pattern (Fig. 13), but in a larger of hexagonal sites.

In some of the other transition metal chalcogenides - e.g. the sulphides of Ti, Zr, Nb, etc. - the structural relations are rather less simple, with mixed hexagonal and cubic stacking along the c axis and changes of stacking sequence at certain compositions. The common feature is that partial occupancy is restricted to certain sheets of cation sites.

Defects in such ordered structures as those of Fig. 13 must now be considered. Intrinsic disorder, due to exchange of occupants between type M and type Z sites can be written.

Table V.1

Compound	Filled	Incomplete	Unit composition
MX	M_{12}	M_{12}	$M_{24}X_{24}$
M_7X_8	M_{12}	M_9Z_3	$M_{21}X_{24}$
M_5X_6	M_{12}	M_8Z_4	$M_{20}X_{24}$
M_3X_4	M_{12}	M_6Z_6	$M_{18}X_{24}$
M_2X_3	M_{12}	M_4Z_8	$M_{16}X_{24}$
M_5X_8	M_{12}	M_3Z_9	$M_{15}X_{24}$
MX_2	M_{12}	Z_{12}	$M_{12}X_{24}$



The empty M site is a true point defect vacancy; the misplaced M atom on a Z site could loosely be regarded as a sort of interstitial. This is better regarded as a form of substitutional disorder, analogous to the exchange of atomic positions in an ordered intermetallic compound such as CuZn; as such, it leads to an order-disorder transformation, with a lambda-point specific heat anomaly, at an elevated temperature that depends on the site-exchange energy.

Extrinsic point defects can be introduced by loss or gain of non-metal, in the form of M_Z or V_M defects, for M:X ratios greater or less than that of the perfectly ordered superstructure phase. The concentration of defects concerned in the defect equilibria, in diffusion processes etc. is very much smaller than the concentration of empty cation sites. Above the order-disorder transition, there is no distinction between M and Z sites; the effect of changing the stoichiometry is to change the average occupancy fraction for all sites. It can therefore be predicted that above the order-disorder temperature these compounds may display a wide nonstoichiometric composition range, below that temperature, the several ordered phases are likely to have narrow composition ranges. This conclusion is in accord with the known properties of the transition metal chalcogenides. The real structure of the nonstoichiometric phases above the order-disorder temperature is not fully explained. Long range order is completely lost, but the fraction of vacant sites is too high for a satisfactory description in terms of point defect vacancies. If there is any residue of short range order, the real distribution of occupied and vacant cation sites might be better described as corresponding to small areas or microdomains, differing in their local ordering

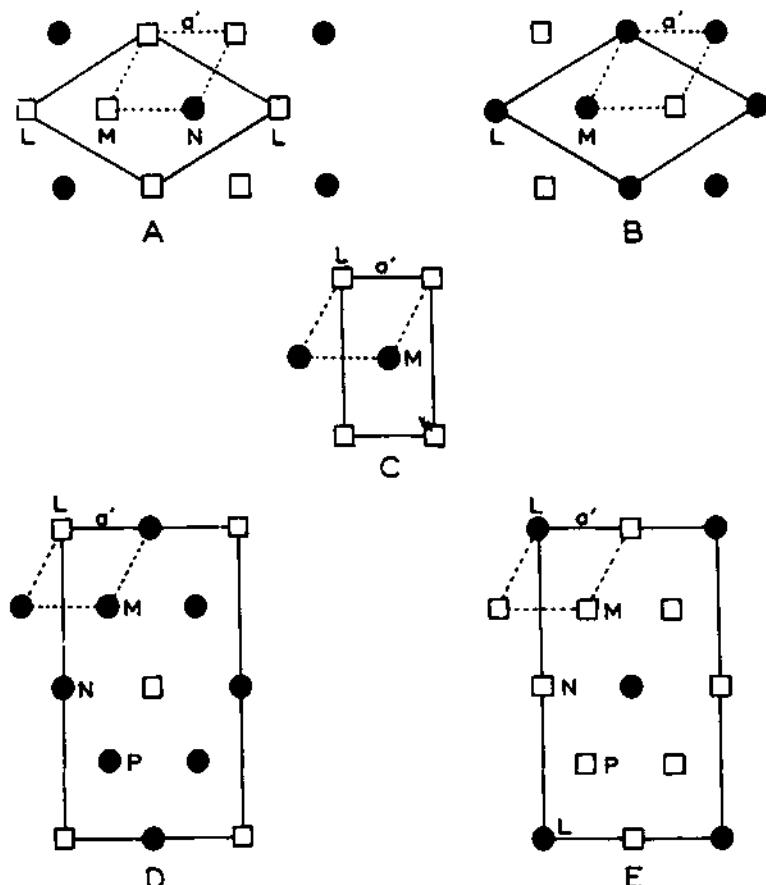


Fig. 13. Modes of ordering of filled and unfilled cation sites (Z sites) recognized in the B8/C6 series of transition metal chalcogen compounds. (a) M_2X_3 (b) M_5X_6 (c) M_3X_4 (d) M_7X_8 (e) M_5X_8 .

pattern and therefore in their local composition. Experimental evidence about this is lacking.

Attention has been restricted, thus far, to disorder within the incomplete cation layers, while preserving the regular stacking of complete and incomplete layers. Stacking disorder is also possible. If complete and incomplete cation sheets were randomly distributed along the c axis, the result would be to randomise vacancies completely, at high temperatures. Lattice energy calculations by Bertaut for the particular case of Fe_7S_8 , and by Burch, at Oxford, for the more general case of nonstoichiometric B8/C6 structures suggest that there is a strong energetic preference for

localising site occupancy into complete and incomplete cation sheets, so that complete randomization appears improbable. What is possible is the occasional irregular fault in stacking - a local switch to another stacking sequence, juxtaposition of two filled or two partially filled sheets, or the occasional interpolation of an incompletely filled sheet that is, indeed ordered, but ordered in a different pattern (and so having a different composition) from the incomplete sheets proper to the nominal composition of the compound.

Non-stoichiometry in these highly ordered structures can thus arise in several distinct ways. Some of these phases - e.g. the titanium sulphides with their complex stacking sequences - do show quite wide stoichiometric variability (Fig. 14), but the structure of the nonstoichiometric phases has not been completely worked out in any case²¹.

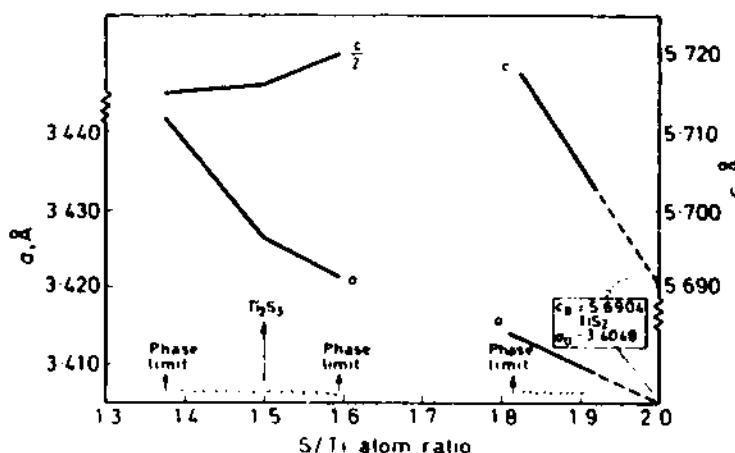


Fig. 14. Cell dimension-composition relations in the titanium sulphides TiS_x showing variability of composition in nearly ordered structures based on alternation of filled and incomplete cation sheets.

VI. ELIMINATION OF DEFECTS: CRYSTALLOGRAPHIC SHEAR

Section V has dealt with mixed valence intermediate phases in which defects are assimilated into the structure by ordering them on a regular array. Relative to the parent structure, this ordering process is conservative of lattice sites. A second group of ordered intermediate phases, now well recognised in the Chemistry of the transition metal oxides but probably not restricted to oxide chemistry, is non conservative in lattice sites: the ordering

process eliminates defects more or less completely by such a rearrangement of coordination polyhedra that the number of sites in the non-metal sublattice is decreased.

This type of ordering gives rise to extensive homologous series of intermediate phases, with progressively changing M:O ratio and closely related structures in the oxides of Ti, Nb, Mo, W and probably Re²². Figure 15 illustrates the principle, in the structure of the oxide Mo_8O_{23} - a (010) projection. The basic structure is of the ReO_3 type, with $[\text{MO}_6]$ octahedral groups linked at all their corners to similar groups. Along certain crystallographic directions, and at regular spacings, the linkage of $[\text{MO}_6]$ octahedra is changed from apex-sharing to edge sharing. This involves some distortion of the octahedra, and brings the cations close together (possibly involving direct cation-cation binding); it retains essentially the same coordination environment for the cations but with, locally, a lowered O:M ratio. Except along these defined crystallographic directions, the ReO_3 -type structure is retained; across the sheets of differently linked octahedra, two slabs of ReO_3 structure appear to be mutually displaced, so as to bring cations of one slab into positions that would be interstitial sites for the adjacent slab. The in-plane oxygen positions are (except for small displacements due to distortions) continuous throughout the structure. This virtual displacement of slabs of the structure is termed crystallographic shear; the layers of differently linked octahedra define the shear plane and the mutual displacement of the ReO_3 -like slabs is defined by a shear vector. To produce a change in stoichiometry - i.e. to eliminate oxygen sites - this shear vector must have a component normal to the shear plane; a displacement parallel to the shear plane produces an antiphase boundary, without change of total composition.

The stoichiometry of the shear phase then depends on three factors:

- a) The nature of the crystallographic shear; the manner in which polyhedra are fused together at the shear plane (Figs. 15,16). The shear phases can be regarded, in a sense, as coherent intergrowths of the parent oxide with a lower oxide - thus, in the shear phases derived from the rutile structure, the shear plane can be considered as a single slice of the corundum structure $\text{MO}_{1.5}$, intergrown with the MO_2 parent structure.
- b) The width of the slabs of parent structure between successive shear planes - i.e. the length of the unbroken rows of 'properly' linked octohedra. In principle, this could vary indefinitely, up to any number; it is limited by (c).
- c) The regular spacing of shear planes. Only if the shear planes

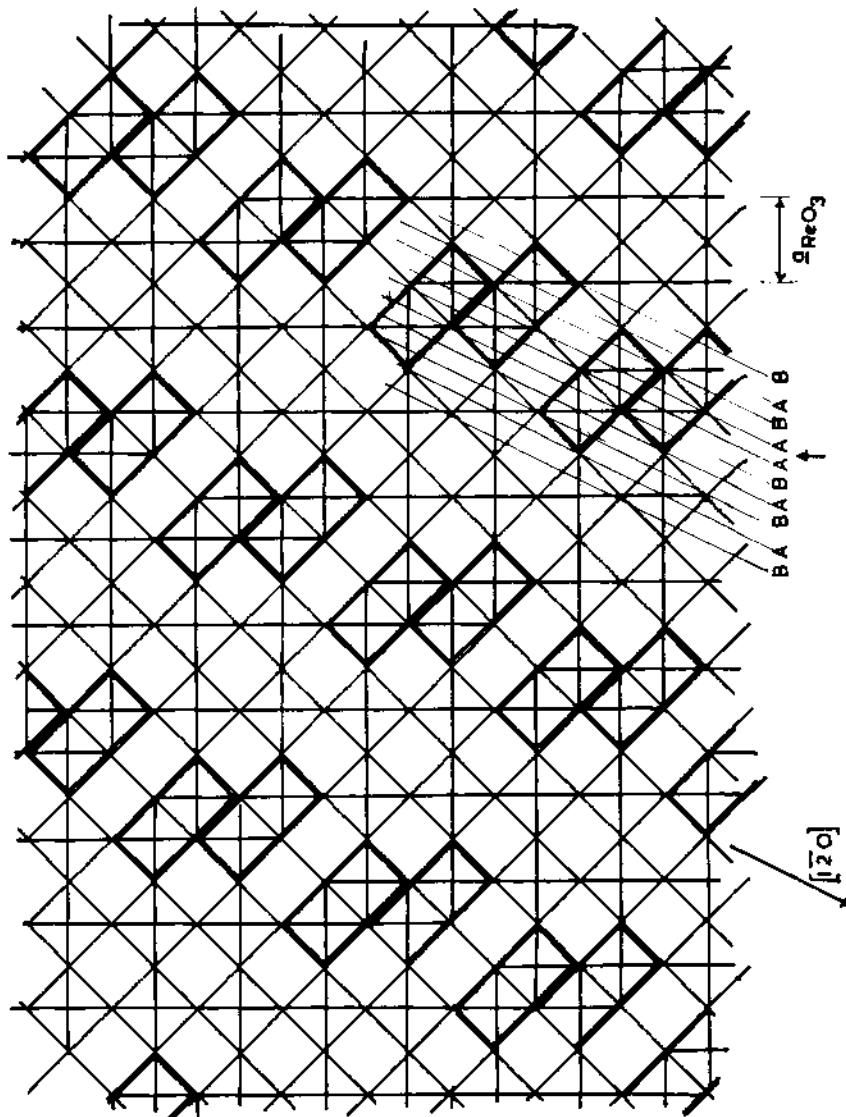


FIG. 15. The crystallographic shear structured of Mo₈O₂₃. Slabs of ReO₃ structure with rows 8 octahedra long. Crystallographic shear on (120)ReO₃. Alternation of type A (MoO₂) and type B (O) layers interrupted at the shear plane.

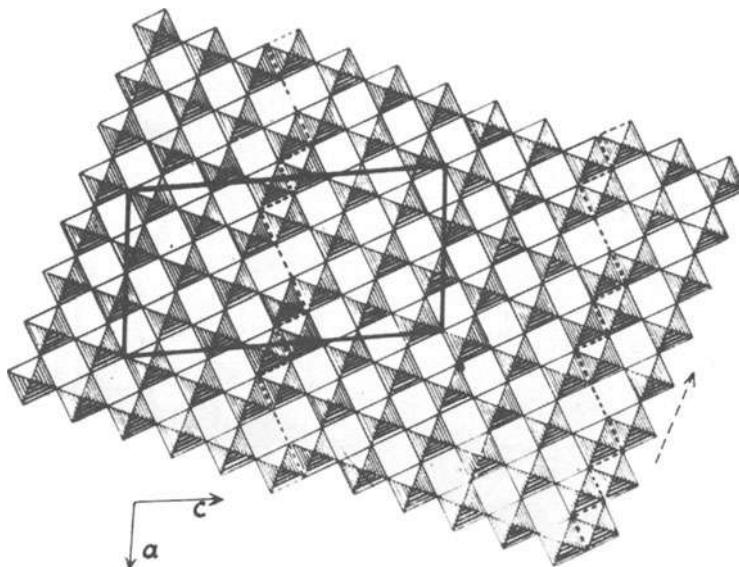


Fig. 16. The structure of $W_{20}O_{58}$ with shear on $(130)_{ReO_3}$.

recur at absolutely regular intervals is the composition of the crystal defined. The evidence is that the known shear phases behave as compounds of invariable composition, so that this condition is fulfilled. This regularity of shear plane spacing necessarily involves long range interaction forces, and is not yet fully explained.

A parent structure MO_x can then generate a family of shear phases M_nO_{xn-m} where m is determined by factor (a) and n by factor (b).

Table VI.1 shows the characteristic parameters for the best known groups of shear structure oxides, and brings out the very subtle changes in over-all composition that can correspond to distinct crystallographic structures.

Even more subtle adjustments of geometrical structure to changes in the metal: nonmetal ratio are provided by the block structure phases^{23,24,25}. In these, two sets of shear planes define rectangular columns of ReO_3 -type structure, infinite along one direction (the b axis), and measuring $m \times n$ octahedra along the a and c directions of the ReO_3 structure. Through the shear planes these are spliced to a second set of columns, the octahedra of which are displaced by half an octahedron diagonal along the b direction. The two sets of columns from a space-filling rectangular array; the junctions between them provide a set of tetrahedral sites. Each

Table VI.1

Parent Structure	Rearrangement of Mo ₆ octahedra at shear plane	Shear plane	Shear vector	Homologous series	Examples
ReO ₃ type	apex edge sharing	($\bar{1}20$) ($\bar{1}30$)	$\frac{1}{2}\langle110\rangle$ $\frac{1}{2}\langle110\rangle$	M _n O _{3n-1} M _n O _{3n-2}	Mo ₈ O ₂₃ , (Mo,W) ₁₄ O ₄₁ W ₂₀ O ₅₈ , W ₄₀ O ₁₁₈
TiO ₂ (rutile) type	edge face sharing	($\bar{1}21$) ($\bar{1}32$)	$\frac{1}{2}\langle10\bar{1}\rangle$ $\frac{1}{2}\langle10\bar{1}\rangle$	M _n O _{2n-1} M _n O _{2n-1}	Ti _n O _{2n-1} , V _n O _{2n-1} (4 n 9) Cr ₂ Ti _n O _{2n+3} Ti _n O _{2n-1} (10 n 40)
ReO ₃ Double shear or block structure	apex edge	(100) (001)	$\frac{1}{2}\langle11\bar{1}\rangle$ $\frac{1}{2}\langle11\bar{1}\rangle$		Nb oxides Nb ₂₈ O ₇₀ (H-Nb ₂ O ₅) to Nb ₁₂ O ₂₉ (NbO ₂ .417) Ternary (Nb, Ti), (Nb,W), (Nb,Mo) oxides Nb oxide fluorides

set of columns may be isolated, or may be linked by edge sharing to a similar column with its cations at the same level in known block structures they are so fixed either in pairs or in infinite strings. The geometry of each block can then be represented by a symbol $(m \times n)_p$ ($p = 1, 2$ or ∞), and the composition of a block is thereby defined as $M_{nmp} O_{2nmp-p(m+n)+4}$. The second set of blocks may have the same or different values for the geometrical parameters m, n, p ; the only constraint is that the spliced assembly fills space. When account is taken of the possibilities of regular intergrowth (or regular alternation along rows and files of blocks), exceedingly small changes in M:O ratio can be accommodated by rearrangement of the block structure whilst retaining perfect order.

To preserve electrostatic charge balance, changes in block structure imply changes in the average charge per cation. Block structures MO_n ($n < 2.5000$) are formed by the reduction of Nb_2O_5 (Nb^{5+} and Nb^{4+} cations), or by ternary oxides in the system $Nb_2O_5-TiO_2$ (i.e. with Nb^{5+} and Ti^{4+} ions). Oxides MO_n ($n > 2.5000$) are formed in the ternary systems $Nb_2O_5-WO_3$ and $Nb_2O_5-MoO_2$ (Nb^{5+} and W^{6+} or Mo^{6+}). There is some evidence that vacancies can be produced at the tetrahedral sites, but in general their behaviour is that of line phases. They are formed readily at high temperatures ($1300 - 1600^\circ C$) from mixtures of the oxides, or by slight reduction of Nb_2O_5 , and although a complete rearrangement of shear planes is needed for the conversion of any phase to another block structure of different composition, the reactions take place with surprising ease. Shear plane mobility is high, but the mechanism of migration is not altogether satisfactorily understood. As far as is known, the cations of differing valency are randomly distributed, without segregation of ions of high charge or low charge in the shear planes.

An alternative way of looking at shear structures is that they arise from the abstraction of a complete lattice sheet containing only oxygen atoms²⁶. In the ReO_3 structure, ($\bar{1}20$) lattice sheets are of two kinds: those that contain both cations and oxygen atoms, of composition MO_2 - type A, and those containing oxygen atoms - type B. The sequence of stacking is ...A B A B A B... In the shear structures (Fig. 15) the sequence of stacking is ...A B A B A A B A B A B... A type B sheet has been abstracted, and the structure closed up along the shear vector $\frac{1}{2}\langle 110 \rangle$. The ($\bar{1}30$) lattice sheets are also of two kinds: A' sheets of composition MO and B' sheets, oxygen only, of composition O. In W_2O_5 etc. the ReO_3 sequence ...A'B'A'B'A'B'... is replaced by ...A'B'A'B'A'B'A'B'... where a B' sheet has been removed

and two A' sheets have collapsed together to give the puckered structure of the shear plane (Fig. 16). These considerations lead to a plausible mechanism for the formation of shear structure in the reduction of the parent oxides - not proved, but consistent with electron microscope observations of the growth of shear planes. The proposed mechanism envisages the creation of oxygen vacancies by the reduction process, followed by aggregation of a few vacancies into a vacancy disc lying in the shear plane. Collapse of the vacancy disc along the shear vector then produces an area of shear plane, bounded by a dislocation loop. The distortions at the dislocation may energetically favour both diffusion of oxygen and abstraction of oxygen at that locus; dislocation climb would drain vacancies from the surrounding structure, to extend the area of collapsed shear structure. In either case, the extension of the dislocation loop can carry the shear plane through the crystal. Given the conversion of point defects to shear planes, a certain degree of reduction of an oxide MO_n prescribes a corresponding density of shear planes per unit volume, lying on all the crystallographically equivalent orientations - e. g. on {120}. Their parallel alignment and regular spacing then involve slower diffusive movements under the operation of rather long range forces.

With this concept in mind, it is reasonable to enquire whether defects can be eliminated by crystallographic shear from appropriate structures, within the composition limits of a nonstoichiometric phase. This could involve either isolated shear planes extending through the crystal, or shear-collapsed discs bounded by dislocation rings. It has been found, in electron microscope studies by Tilley, at Oxford, and by Hyde and Bursill in Western Australia, that random shear planes are present in very slightly substoichiometric TiO_2 . The observations relate, however, to samples quenched from 1000°C and it is not certain that the observed structure is identical with the high temperature equilibrium state. Dislocations are regarded as 'non-thermodynamic' defects, and it is not clear whether the concept of a disc of shear structure bounded by a dislocation can be brought within the treatment of equilibrium thermodynamics.

All the shear structures investigated hitherto have been oxides of a restricted number of transition metals. It is extremely probable, however, both that defects may be eliminated by crystallographic shear from oxides other than

those of ReO_3 , TiO_2 and MoO_3 types and that shear structures will be recognized in other series of metallic compounds.

VII. IMPERFECTIONS OF ORDER IN ORDERED STRUCTURES

It will be clear from the foregoing discussion that the ultramicrostructure of crystals has to be deduced from the properties of material in bulk, and that statements about short range order are almost entirely based on inferences rather than observation. Even if a crystal structure is carried out on a small crystal, 10^{-2}cm in dimensions, the experimental sample contains $10^{15} - 10^{16}$ unit cells and the finer details of structure may be averaged out. In this section a new development in structural studies will be discussed; the possibility of investigating ultramicrostructure at the level of the individual unit cell. This possibility is opened up by the application of electron microscopy to the study of crystals and ordered structures.

Direct Lattice Imaging. The electron microscope provides for the observation of both the magnified image and the diffraction pattern of a thin crystalline specimen (of thickness up to perhaps 1000 \AA). Because of the short wave length of the electrons used ($\lambda = 0.038 \text{ \AA}$ for 100 KV electrons), the Ewald sphere is of large radius and diffraction angles are small. When account is taken of the prolongation of reciprocal lattice points normal to the thin crystal, many crystal planes may be brought simultaneously into the Bragg reflecting orientation. For an electron beam directed along a main zone axis, a stationary crystal presents a whole reciprocal lattice sheet as its diffraction pattern. In X-ray diffraction a reciprocal lattice sheet must be mapped by rotating or oscillating the crystal.

In some respects the electron diffraction pattern presents more accessible information than the other diffraction techniques. The symmetry, dimensions and population of the reciprocal lattice are displayed directly. Because the scattering of electrons by atoms is very efficient, and not so heavily weighted in favour of atoms of high atomic number as is the scattering of X-rays, reflections due chiefly to scattering by the lighter atoms are more readily observable.

'Shape factors' are important. For an infinite crystal - or, effectively, for any macroscopic crystal - any reciprocal lattice point hkl is a true point, but if a coherently scattering region is very small in any dimension, its transform in reciprocal space is extended along the normal to that direction by an amount $l^* \propto L^{-1}$, where L is the dimension of the region in real space. The reciprocal lattice transform of a thin sheet is a spike in reciprocal space, normal to the sheet; that of a needly shaped crystal is a disc. In virtue of this feature, faults in crystalline order contribute to the diffraction pattern - e.g. planar faults such as edge dislocations or stacking faults, antiphase boundaries, thin coherently scattering regions that differ in phase from neighbouring regions etc. The diffraction spots may be drawn out in streaks, or additional reflexions observed. Finally, the diffuse scattering from uncorrelated defects etc. is more intense for electrons than for X-ray diffraction.

The relation between the reciprocal lattice and the real space lattice is that each is the Fourier transform of the other. Recombination of the undeflected electron beam (the (000) reciprocal lattice point) with selected reciprocal lattice points produces an image by transformation of the selected Fourier components only. If the (000) beam is recombined with, e.g. the (001) and higher orders in $(00\bar{1})$, the image is a fringe pattern directly representing the operating set of $(00\bar{1})$ lattice planes. The contrast in such a fringe system depends upon the distribution of the structure factor across the unit cell. For such compounds as the shear phases, in which one lattice translation is given directly by the spacing of shear planes, and in which the scattering amplitude is highly concentrated at the shear planes, the image contrast corresponds with the shear plane spacing. Furthermore, the phase grating treatment of diffraction by thin crystals (Cowley and Moodie 1960) shows that at the correct focal position (the zero order phase grating image) the image contains all the information present in the selected Fourier components of the object, including the aperiodic features.

This means that direct lattice imaging (a) presents for direct measurement the interplanar spacings at the level of the individual unit cell, and (b) displays any deviations from perfect regularity of spacing and scattering power. This statement must be qualified

by recognition of the fact that even a thin crystal is perhaps 10-100 unit cells thick: the fringe contrast depends upon the total scattering in a column of the structure. Particularly in shear structures, however, it is valid to assume uniformity in depth; shear planes are inherently planar, since jogs would be associated with a high fault energy. In some other instances - superstructure ordering in tunnel structures etc. - the inferences may be less direct. Nevertheless, direct lattice imaging, which is not new and has been used, for example in the direct observation of edge dislocations, and as a measure of resolving power of electron microscopes, opens up new possibilities of studying problems of long and short range ordering.

Results of Direct Lattice Imaging. Relatively few workers have as yet exploited the possibilities of direct lattice imaging. The most important contributions have been made by Allpress, and examples taken largely from his work will serve to show the power of the method and the new insights on crystalline order that it reveals^{27,28,29,30}. Much of this work has been concerned with the block structure oxides; in these, the very direct relationship between the geometry of the structure and the numbers of metal- and oxygen sites in the individual blocks makes it possible to convert fringe spacings unambiguously into compositions.

To simplify the interpretation of the lattice images, it is convenient to introduce a schematic representation of the block structures (Fig. 17) representing each block by a rectangle drawn through the centres of the outside metal atoms. Whereas in the single shear structures the shear planes are sheets of higher than average atomic density running right across the crystal, the blocks are linked together through their shear planes in a staggered configuration. Lines through the atoms in tetrahedral sites then define the directions of highest atomic density, and these divide the whole structure into files and rows of blocks. The spacing between the fringes is then determined (a) primarily by the number of octahedra along each edge of the block - the parameters m and p in the formula, and (b) secondarily, by the linkage of blocks at the same level. Any structure that is based on blocks of two sizes, in some regular stacking sequence will then show two characteristic features in its electron diffraction pattern and its lattice image: a long spacing corresponding to the true repeat distance of the structure, and an intermediate set of fringes corresponding to the sequence of individual blocks in the stacking.

Fig. 18 shows the electron diffraction patterns of crystals isolated from Nb_2O_5 - WO_3 preparations rich in Nb_2O_5 ($\text{Nb}_2\text{O}_5 : \text{WO}_3 > 10:1$). The only phases identified by X-ray crystallography in this region were $\text{WNb}_{12}\text{O}_{33}$, with the structure $(4 \times 3)_1 + (4 \times 3)_1$

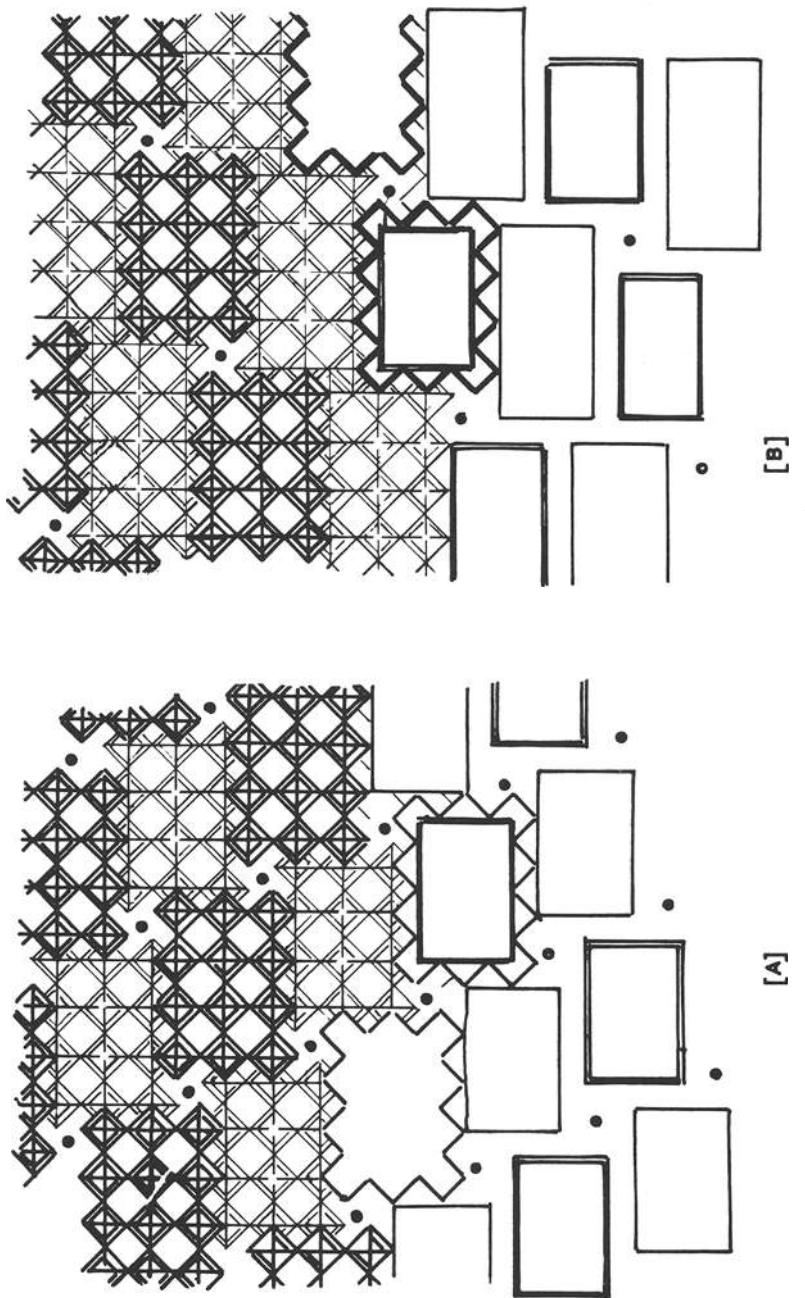


Fig. 17. Block or double shear structures. (a) $(4 \times 3)_1 + (4 \times 3)_1$ blocks in $\text{WNb}_{12}\text{O}_{33}$.
(b) $(4 \times 3)_1 + (5 \times 3)_\infty$ blocks in $\text{Nb}_{28}\text{O}_{70}$.

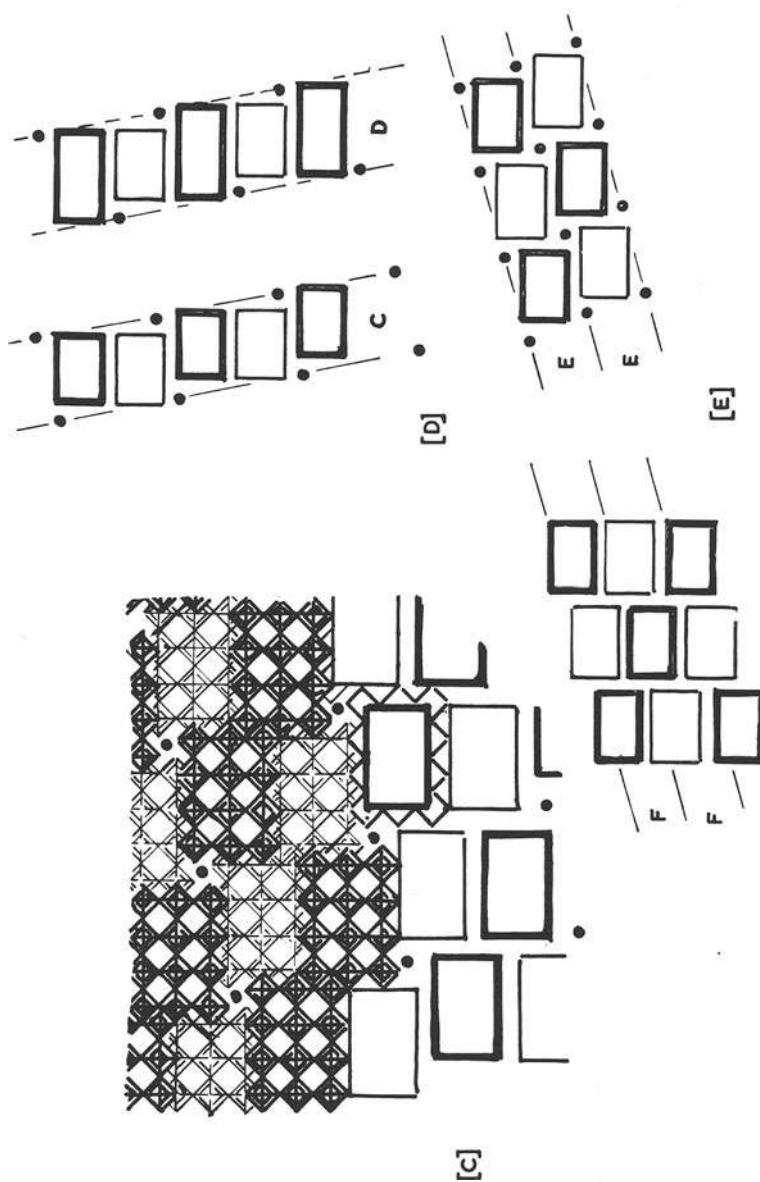


Fig. 17. (c) $(4 \times 3)_2 + (4 \times 3)_2$ blocks in $TiNb_{24}O_{62}$. (d) Characteristic widths of files of blocks as found in the $TiNb_{24}O_{62}$ (c) and the $Nb_{28}O_{70}$ (d) structures. (e) Characteristic widths of rows of blocks involving (F) or not involving (E) tetrahedral sites.

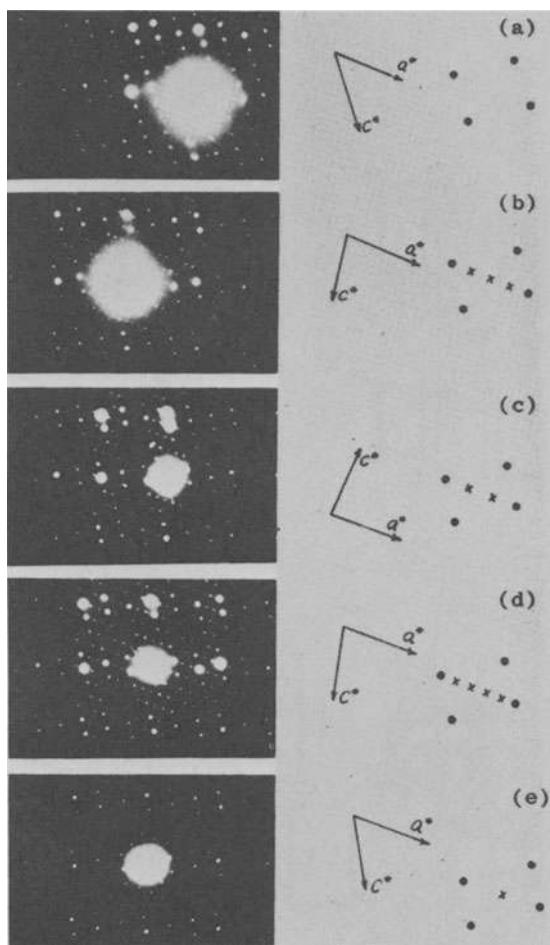


Fig. 18. Electron diffraction patterns in the $\text{W}_{103}\text{-Nb}_2\text{O}_5$ series.
 (a) $\text{WNb}_{12}\text{O}_{33}$ (b) New compound $\text{WNb}_{26}\text{O}_{68}$. (c) New compound $\text{WNb}_{40}\text{O}_{103}$
 (d) New compound $\text{WNb}_{68}\text{O}_{173}$, intergrown with $\text{WNb}_{40}\text{O}_{103}$ (e) $\text{Nb}_{28}\text{O}_{70}$.

(Fig. 18 (a)) and $\text{H-Nb}_2\text{O}_5$, $\text{Nb}_{28}\text{O}_{70} (4x3)_1 + (5x3)_{\infty}$ (Fig. 18 (e)). Their diffraction patterns are very similar, but in (e) the a^* spacing is halved. This is at once explicable in terms of the two kinds of blocks involved. Denoting the rows of $(4x3)_1$ blocks, with interposed tetrahedral atoms, as type A, and rows of $(5x3)_{\infty}$ blocks with no tetrahedral atoms as type B, diffraction pattern (a) corresponds to the stacking: $\text{WNb}_{12}\text{O}_{33} - \underline{\text{A}}-\overline{\text{A}}-\underline{\text{A}}-\overline{\text{A}}-\underline{\text{A}}-\overline{\text{A}}-$, and (e) to $\text{Nb}_{28}\text{O}_{70} \underline{\text{B}}-\overline{\text{A}}-\underline{\text{B}}-\overline{\text{A}}-\underline{\text{B}}-\overline{\text{A}}-$, where large and small letters distinguish the blocks with metal atoms at levels 0 and $\pm \frac{b}{3}$ respectively. For $\text{WNb}_{12}\text{O}_{33}$, reflections $(\underline{h}00)$ are extinguished by symmetry considerations for h odd; this is not the case for $\text{Nb}_{28}\text{O}_{70}$ and hence the a^*

spacing is halved in the diffraction pattern of the latter. In patterns (b), (c), (d), the reciprocal lattice spacing found in $\text{WNb}_{12}\text{O}_{33}$ appears to be divided by 4, 3 and 5 respectively. In conjunction with the other diffraction evidence, these patterns must be interpreted as due to new phases with more complex stacking sequences:

- (b) Sequence $\overbrace{\text{A-B-A-A-B-A-A}}^4$ 4 block repeat $\text{WNb}_{26}^{0.68}$
- (c) $\overbrace{\text{A-B-A-A-B-A-A-B-A}}^6 = 3$ block repeat $\text{WNb}_{40}^{0.103}$
- (d) $\overbrace{\text{A-B-A-A-B-A-B-A-A-B-A-B-A}}^{10} = 5$ block repeat $\text{WNb}_{68}^{0.173}$

The formulae of the new phases follow directly from the identification of the block stacking. The phases can be regarded from a slightly different viewpoint as completely ordered intergrowth of layers of $\text{WNb}_{12}\text{O}_{33}$ structure with layers of $\text{Nb}_{28}^{0.70}$ structure.

This concept of coherent intergrowth is important because one and the same physical crystal may show regions in which one structure changes to another without any discontinuity. Indeed, the new phase $\text{WNb}_{68}^{0.173}$ was not found as a single crystal, but only as domains in crystal which would, therefore, have an indeterminate and variable total composition. We must then enquire into the scale on which coherent intergrowth can occur; for this, direct lattice imaging is revealing.

Fig. 19 shows ($\bar{h}00$) lattice fringes obtained from $\text{Nb}_2\text{O}_5\text{-TiO}_2$ preparations, rich in Nb_2O_5 . Fig. 19(a) is the lattice image of $\text{Nb}_{28}^{0.70}$, with regularly spaced fringes 17 Å apart (calc. 16.8 Å) corresponding to the distance between tetrahedral atoms fixed by the $(4 \times 3)_1$ blocks of the structure. A crystal from a $\text{TiO}_2+40\% \text{Nb}_2\text{O}_5$ sample (Fig. 19(b)) shows predominantly the 17 Å spacing of $\text{Nb}_{28}^{0.70}$, but with occasionally interposed fringes of 15 Å spacing which correspond to a file of $(4 \times 3)_2$ blocks (calc. spacing 14.8 Å). A slice of $\text{TiNb}_{24}^{0.62}$ (structure $2 \times (4 \times 3)_2$, one or two unit cells thick) is occasionally intergrown. In Fig. 19(c), from a composition $\text{TiO}_2 + 26\% \text{Nb}_2\text{O}_5$, the lattice image is made up mostly of double fringes with a periodicity of 32 Å (= 17 + 15 Å), but includes a slice 3 unit cells thick defined by a triple fringe pattern of dimensions 47 Å (= 17 + 2 × 15). If we denote the files of blocks $(4 \times 3)_2$ as type C, and those of the $(4 \times 3)_1$ blocks as type D, we have the stacking sequences:

in Nb_2O_5 -D-D-D-D-

in $\text{TiNb}_{24}^{0.62}$ -C-C-C-C-

The new fringe patterns, and the formulae necessarily assigned to

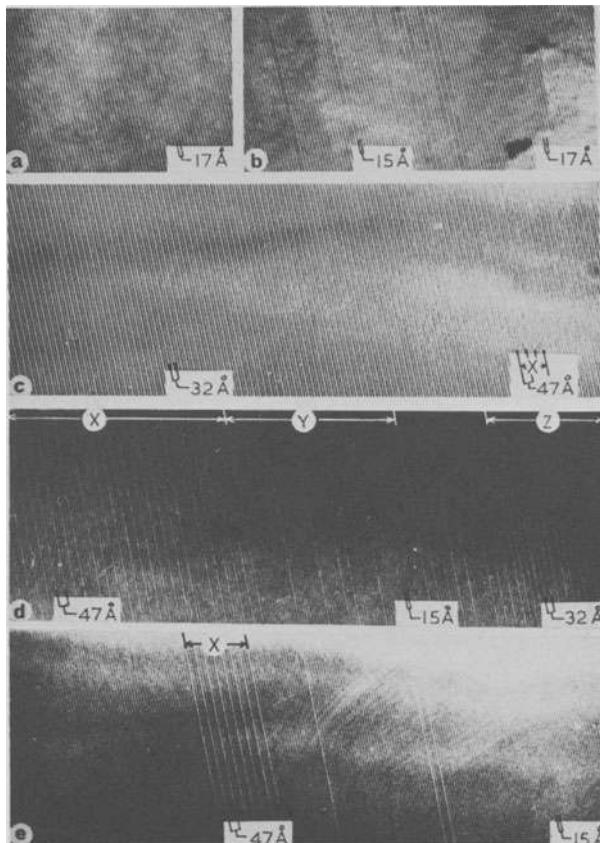


Fig. 19. ($\bar{h}00$) lattice fringes from crystals in the TiO_2 - Nb_2O_5 series. (a) $Nb_{28}O_{70}$. (b) Coherent intergrowth of $Nb_{28}O_{70}$ and $TiNb_{24}O_{62}$ in $TiO_2 + 40$ Nb_2O_5 . (c) $TiNb_{52}O_{132}$ with small domain (X) of $TiNb_{38}O_{97}$. (d) Coherent domains of (mostly) $TiNb_{38}O_{97}$ (X), $TiNb_{24}O_{62}$ (Y) and $TiNb_{52}O_{132}$ (Z). (e) A domain of $TiNb_{38}O_{97}$ in $TiNb_{24}O_{62}$.

the structures are then

32 Å fringes sequence	$\overbrace{C-D-C-D-C-D-}$	$TiNb_{52}O_{132}$
47 Å fringes	$\overbrace{C-C-D-C-C-D-}$	$TiNb_{38}O_{97}$

In Fig. 19(d), a sample of total composition close to $TiNb_{52}O_{132}$, there are completely coherent regions of $TiNb_{38}O_{97}$ (x), $TiNb_{24}O_{62}$ (y) and $TiNb_{52}O_{132}$ (z), but these show occasional faults one cell wide. Similarly in Fig. 19(e) from a crystal of supposedly

pure $TiNb_{24}O_{62}$ there are both isolated faults and a strip of $TiNb_{38}O_{97}$, seven unit cells wide.

What these lattice images show is that there were fluctuations of composition - in particular variations of the Ti:Nb ratio - between different parts of the same crystal. Because different structures, for different Ti:Nb or W:Nb ratios, are geometrically compatible, so that they can intergrow coherently and with no strain, local adjustments of composition by short range diffusion could line up deviant compositions in strips or areas, giving rise to a particular block structure, whereas long range uniformity was not attained. In that sense the crystals were not in a true equilibrium state, at least at low temperatures. Faulted strips can be formed in this way, either running across a considerable width of crystal or terminating within the crystal, by intersection with another row of aberrant block structure, or at a dislocation.

Fig. 20, showing the $(00\bar{1})$ lattice image of a crystal of $W_5Nb_{16}O_{55}$ is of interest in this connection. The regular fringe spacing of 18.5\AA , corresponding to the long dimension of the $(5\times 4)_1 + (5\times 4)_1$ blocks, show isolated spacings of 15\AA and 22\AA . The former corresponds to the $(4\times 4)_1$ blocks of $W_3Nb_{14}O_{44}$ (i.e. a tungsten-rich strip), the latter to $(6\times 4)_1$ blocks of a new phase $W_7Nb_{18}O_{66}$ (i.e. niobium-rich). A strip of crystal 131 unit cells

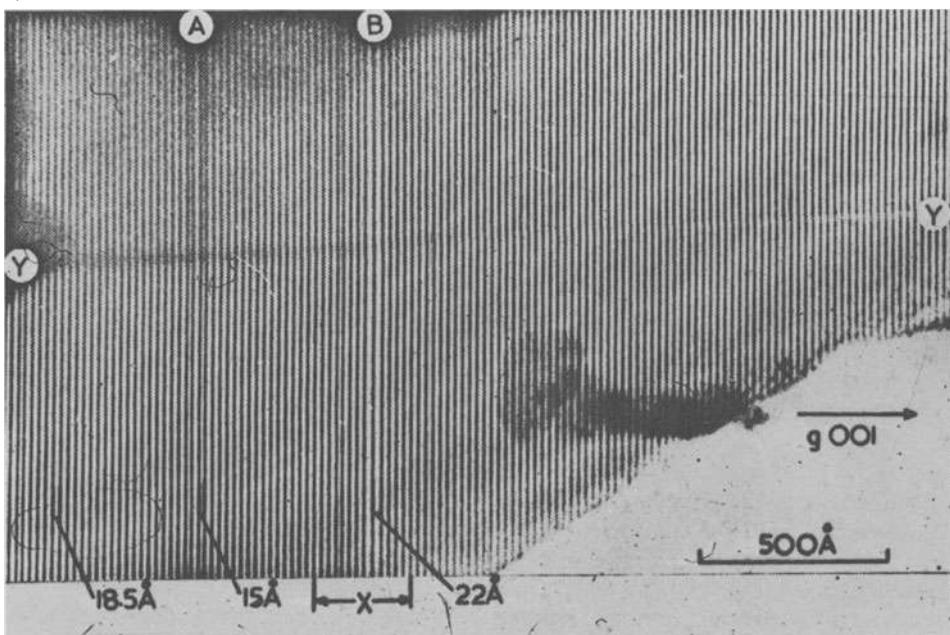


Fig. 20. Insertion of occasional strips of a different block structure in the regular sequence of (001) fringes (18.5 \AA spacing) in $W_5Nb_{16}O_{55}$.

wide across a faulted region (about 0.25 microns) included 125 'proper' spacings of 18.5 \AA , 4 spacings of 15 \AA and 2 spacings of 22 \AA . The total composition, based on the formulae assigned to the fault strips, is $\text{Mo}_{2.6187}$, as compared with the composition $\text{Mo}_{2.6190}$ for the perfect $\text{W}_5\text{Nb}_{16}\text{O}_{55}$ crystal. The aberrant rows of blocks arise from a long fluctuation in the composition - probably a metastable condition, and evidence of incomplete diffusive mixing.

Less extensive work has been carried out on the single-shear structures than on the block structures, but the two essential findings apply to these also: coherent intergrowth of different members of the series both in domains and as fault strips, and the recognition of new and more closely spaced members of homologous series. The shear phases based on rutile, discovered by Magneli and Anderson, had the {121} shear plane and extended from Ti_{4}O_7 to $\text{Ti}_{9}\text{O}_{17}$. In apparently perfect, homogeneous crystals grown by chemical vapour transport, intergrowths of several members of the series are found, both as domains large enough to give superlattice reflexions and, as shown by streaking of the diffraction spots, in narrow bands similar to the intergrowths described above. Of particular interest is the structure found for very slightly reduced TiO_2 . Material reduced to a composition just lower than the supposed limiting composition of the rutile phase at $\text{TiO}_{1.993}$ does not consist of rutile + the uppermost Magneli phase, but shows the diffraction patterns corresponding to a new family of shear structures, with the alternative, predicted {132} shear plane. Detailed work by Hyde and Bursill has proved that the shear vector is the predicted $\frac{1}{2} \langle 011 \rangle$. This part of the phase diagram cannot be resolved into a succession of clearly defined two phase regions: in any one crystal, at a given degree of reduction, several different super-lattices may be found, corresponding to oxides $\text{Ti}_{n}\text{O}_{2n-1}$ with $n = 10$ to about 40 i.e. coherently growing oxides $\text{Ti}_{10}\text{O}_{19}$ to about $\text{Ti}_{40}\text{O}_{79}$ with closely spaced compositions. Direct lattice imaging of material of composition about $\text{TiO}_{1.990}$ shows that the shear planes defining the composition of successive strips of structure are almost randomly intergrown; the distribution of their spacings lies on a histogram which peaks at some average composition. Within crystals that are nominally pure rutile phase, isolated {132} shear planes may be found. It has been disputed whether oxygen deficient rutile has oxygen vacancies or interstitial titanium. As studied at room temperature, rutile appears to eliminate point defects by crystallographic shear. Caution must be exercised in extrapolating this conclusion to high temperatures where true defect equilibrium can be established: at low temperatures TiO_2 undoubtedly has an infinitesimally narrow true composition range, and all these phenomena may be complicated by ex-solution processes and metastable states as the microstructure adjusts itself to the change from 1000°C to room temperature.

In addition to this work on double shear and single shear oxides, some exploratory work has been carried out that bears on the establishment of order in potentially random structures. The observations relate to the apparently nonstoichiometric ternary oxides $(NbO)_xMo_3$ related in structure to the tetragonal tungsten bronzes. This structure (Fig. 21) is based on fully corner linked $[Mo_6]$ octahedra (where M = W, Nb or Ta in such proportions as to give

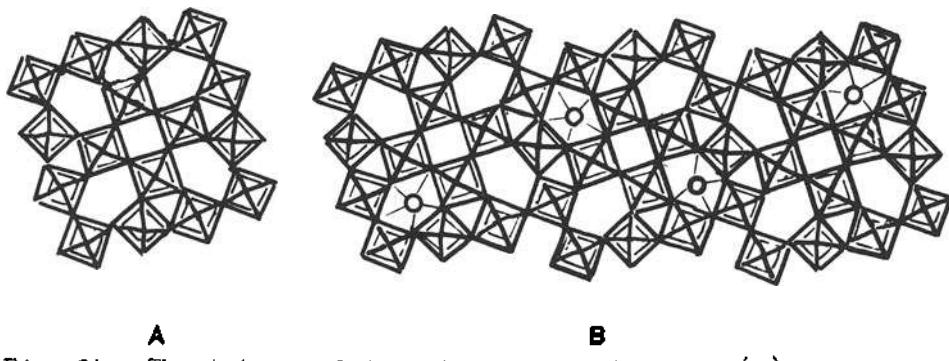


Fig. 21. The tetragonal tungsten bronze structure (a) and the derived ordered structure of $Nb_{16}W_{18}O_{94}$, one of the series $(NbO)_xMo_3$.

the right stoichiometry to the host framework), so arranged as to enclose pentagonal tunnels. The size of these tunnels is such that they can accommodate large ions (e.g. K^+ in the tetragonal tungsten bronzes) or NbO groups. In the latter case it is probable that any one tunnel is either completely empty or completely filled, so that its niobium atoms are brought into 7-fold, pentagonal bipyramidal coordination with oxygen. In the compounds $(NbO)_xMo_3$, only a fraction of the pentagonal tunnels is populated. If the tunnels are filled at random, the average symmetry is that of the tetragonal tungsten bronzes, with a unit cell about $12.5 \times 12.5 \text{ \AA}$, one octahedron in depth. If the tunnels are filled in an ordered manner, a superstructure is formed, as in the compound of defined stoichiometry $(NbO)_{0.133}W_{0.4}O_3$ or $Nb_{16}W_{18}O_{94}$ shown in Fig. 21(b). Fig. 22 shows the lattice image of a thin fragment of $Nb_2O_5 \cdot 3WO_3$, or $(NbO)_{0.111}W_{0.667}Nb_{0.333}O_3$, formed by combining the (000) ray with the ($\bar{h}00$) and ($00\bar{l}$) reflexions of the basic tetragonal bronze structure. The lattice image clearly displays (a) domains where only the $12.5 \times 12.5 \text{ \AA}$ grid of the host lattice is visible, corresponding to domains of disordered structure; (b) individual unit cells within those domains which differ in contrast - a difference arising, perhaps, from the distinction in scattering

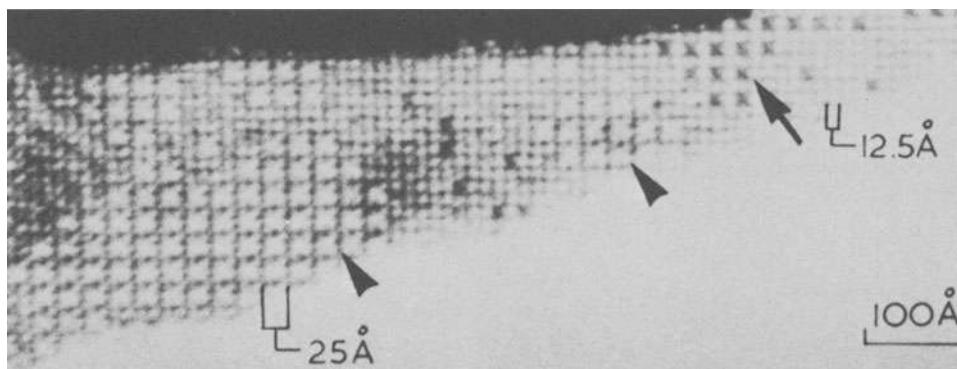


Fig. 22. Lattice image of $\text{Nb}_2\text{O}_5 \cdot 3\text{WO}_3$: (h00) and (001) fringes. 12.5 \times 12.5 \AA subcell, with domains ordered into a 2 \times 2 superlattice.

power between cells with empty and cells with filled tunnels; and (c) domains with a 25 \times 25 \AA grid, which must arise from ordering on a 2 \times 2 superstructure.

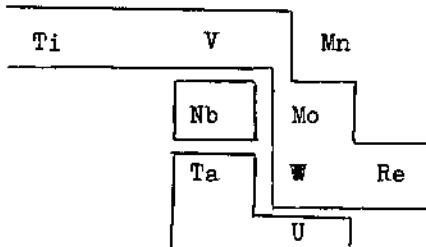
The results presented above come from the exploratory stage of a technique which offers great promise. It may well lead to a much deeper insight into the real constitution of defect structures.

VIII. CONCLUSIONS AND OUTSTANDING PROBLEMS

Broad generalisations not infrequently turn into rather crude approximations as scientific knowledge advances. The point defect concept appeared at first to offer a rational interpretation for the structure and properties of inorganic solids. As can now be seen, this is no longer a tenable viewpoint: defects are real, but at high concentrations one is no longer able to discuss structure and properties in terms of point defects alone. A general theory of defect equilibria, leading to a unified statistical thermodynamic treatment, presents great difficulties.

One problem is that the preferred ultramicrostructure of defect solids shows a high degree of chemical specificity: elements from the same group of the Periodic Table may resolve the ordering and structure of defects in different ways. Their atoms do not behave only as hard-sphere cations, even in structures derived from maximum valence state compounds. It may, probably, be inferred that the radial extension and overlap properties of their valence shell orbitals are important. Thus, in binary oxides of elements in Groups IV - VII of the Periodic Table, shear structures are formed by Ti and V in the rutile type (and, probably, in the intermediate

oxides between V_2O_5 and VO_2), by Mo, W and probably Re. MnO_2 and



the other rutile oxides are not known to form shear structures. ZrO_2 and HfO_2 are of different structures, and conform to the defect fluorite structure pattern. Nb^V and Nb^{IV} with the same electronic configuration and V^V and V^{IV} alone forms block structures in its (IV,V) mixed valence and ternary oxides. Ta oxides do not

display crystallographic shear, but give rise to yet another, almost infinitely variable, set of structures in which the relative numbers of 6- and 7-fold coordinate cations are changed with the M:O ratio, while retaining complete order. Uranium (in U_2O_8 and some ternary phases) almost certainly follows the same pattern.

Amongst the B1 structure oxides, there is a similar chemical specificity. There is a break in the general character between TiO and MnO :



CrO has not been characterised. TiO and VO are metallic, the oxides of elements to the right of chromium are essentially ionic. In FeO , there are well defined clusters and the same is true for MnO . It may prove true for CoO also. TiO and VO are vacancy structures, as has been seen, and there are similar remarkably high vacancy concentrations across the phase ranges of both compounds. These two compounds are, however, almost immiscible in the solid state: the range of solid solution is only about 3% at either end. The most immediate explanation is that the short range ordering of vacancies is totally different in the two cases. NbO , with a phase range $NbO_{0.98}$ to $NbO_{1.03}$, and a structure represented by $Nb_{\frac{3}{4}} O_{\frac{3}{4}}$, is immiscible with both VO and TiO .

A second question is provoked by such evidence as is given in Section VII, that domains of one structure, possibly only one or two unit cells in the important dimensions, may form and intergrow in another structurally and dimensionally compatible matrix. This question is the philosophical, as distinct from the operational, definition of a definite compound. Table VIII.1 lists the block structures so far recognised in the range of O:M ratios 2.45-2.60. Some of these phases have been recognised only as intergrown domains. The free energy-composition diagram of such a system is shown schematically in Fig. 23. It consists of an exceedingly closely spaced set of line-phase free energy curves. In principle, each pair can coexist in equilibrium at a particular chemical potential of the components. The tangents to pairs of adjacent free energy

Table VIII.1

Compound	O:M ratio
Ti ₅ Nb ₄₄ O ₁₂₀	2.4490
Ti ₃ Nb ₃₄ O ₉₁	2.4595
Ti Nb ₂₄ O ₆₂	2.4800
Ti Nb ₃₈ O ₉₇	2.4872
Ti Nb ₅₂ O ₁₃₂	2.4906
Nb ₂₈ O ₇₀	2.5000
W Nb ₆₈ O ₁₇₃	2.5073
W Nb ₄₀ O ₁₀₃	2.5122
W Nb ₂₆ O ₆₈	2.5185
W Nb ₁₂ O ₃₃	2.5385
W ₂ Nb ₁₄ O ₄₄	2.5882

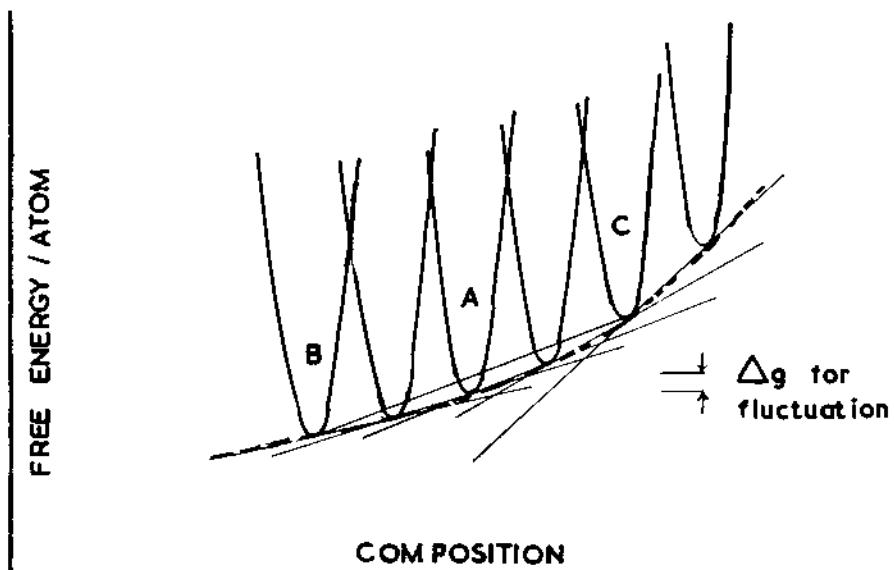


Fig. 23. Free energy-composition isotherm of a system with a succession of closely spaced intermediate phases: narrow composition range only represented. Constant chemical potential tangent lines merge into continuous envelope. The metastable coexistence of B with C, for total composition the same as phase A, raises the free energy by the small increment Δg .

curves, defining the chemical potential, must inevitably differ in slope by very small amounts. The problem is whether such a situation is distinguishable from a continuous variation of chemical potential across a phase range, for which the operationally determined free energy curve would be the envelope shown in the Figure. Furthermore, a perturbation of the system, such as some disproportionation of one phase A into a pair of near phases B and C would increase the free energy of the system by a very small amount. On the one hand, this explains why the driving force behind attainment of perfectly uniform composition may be so small that intergrown domains persist even after long annealing. On the other hand it raises the question whether, in a perfectly ordered crystal of phase A, local inhomogeneities with their consequent structural changes (e.g. realinement of shear planes) could arise by spontaneous fluctuations at high temperatures.

In some respects, the discussion of these question is restricted because sufficient and reliable structural and thermodynamic information is available only for rather few substances and types of systems. As that knowledge is extended, a new unifying viewpoint may emerge.

REFERENCES

General

Kröger F.A., The Chemistry of Imperfect Crystals. (1964) N. Holland, Amsterdam.

Mandelcorn E. (Editor), Nonstoichiometric Compounds (1964) Academic Press New York, Chapters 3 and 5.

Rees A.L.G., Chemistry of the Defect Solid State (1954) Methuen, London.

Rabenau A. (Editor), Problems of Nonstoichiometry (1970) N. Holland, Amsterdam.

Specific

1. Kröger F.A. J., Phys. Chem. Solids, 29, 1889 (1968).
2. Elcock E.W., Order-Disorder Phenomena (1956) Methuen, London.
3. Anderson J.S., Proc. Roy. Soc., A 185, 69 (1945).
4. Messen C.E. and Wan-Hoi Hung G., J. Phys. Chem., 72, 3958 (1968).

5. Hagemark K. and Broli M., *J. Inorg. Nucl. Chem.*, 28, 2837 (1966).
6. Atlas L.M., *J. Phys. Chem. Solids*, 29, 91, 1349 (1968).
7. Willis B.T.M., *Nature*, 197, 755 (1963); *J. Physique*, 25, 431 (1964).
8. Hyde B.G., Bevan D.J.M. and Eyring L., *Phil. Trans. Roy. Soc.*, A 259, 583 (1966).
9. Bevan D.J.M., Cameron R.S., Mann A.W., Brauer G. and Roether U., *Inorg. Nucl. Chem. Letters* 4, 241 (1968).
10. Jette E.R. and Foote F., *J. Chem. Phys.*, 1, 29 (1933).
11. Carel C., Weigel D. and Vallet P., *Compt. rend.*, 260, 4325 (1965).
12. Brynestad J. and Flood H. *Z. Elektrochem.*, 62, 953 (1958).
13. Kofstad P. and Hed A.Z., *J. Electrochem. Soc.*, 115, 102 (1968).
14. Vallet P. and Raccah P., *Compt. rend.*, 258, 3679 (1964), *Mem. Sci. Rev. Met.* 62, 1 (1965).
15. Fender B.E.F. and Riley F.D., *J. Phys. Chem. Solids*, 30, 793 (1969).
16. Roth W.L., *Acta Cryst.*, 13, 140 (1960).
17. Koch F. and Cohen J.B., *Acta Cryst.*, B 25, 275.
18. Herai T., Thomas B., Marenc J. and Benard J., *Compt. rend.*, 258, 4528 (1964).
19. Katsura T., Iwasaki B. and Kimura S., *J. Chem. Phys.*, 47, 4559 (1967).
20. Chevreton M., *Bull. Soc. Franc. Mineral. Crist.*, 90, 592 (1967).
21. Jeannin Y. and Benard J., *Advances in Chemistry*, 39, 191 (1963).
22. Wadsley A.D. in *Non-Stoichiometric Compounds* (Ed. Mandelcorn, 1964) Chap.
23. Wadsley A.D., *Helv. Chem. Acta. Wener Centenary Issue*, 207 (1967).
24. Andersson S., Mumme W.G. and Wadsley A.D., *Acta Cryst.*, 21, 802 (1966).

25. Roth R.S. and Wadsley A.D., *Acta Cryst.*, 19, 26, 724 (1965).
26. Anderson J.S. and Hyde B.G., *J. Phys. Chem. Solids*, 28, 1393 (1967).
27. Allpress J.G., Sanders J.V. and Wadsley A.D., *Acta Cryst.*, B 25, 1156 (1969).
28. Allpress J.G. and Wadsley A.D., *J. Solid State Chem.*, 1, 28 (1969).
29. Allpress J.G., *J. Solid State Chem.*, 1, 66 (1969).
30. Allpress J.G., *Mater. Res. Bull.*, 4, 107 (1969).

THEORY OF POINT DEFECTS IN IONIC CRYSTALS

C. N. R. Rao and S. Ramdas

Department of Chemistry

Indian Institute of Technology, Kanpur, India

I. INTRODUCTION TO POINT DEFECTS IN IONIC CRYSTALS

Any deviation from a perfect periodic lattice or structure in a crystal is an imperfection. There are three main types of imperfections, point, line and plane imperfections¹, of which we shall deal only with the point defects in this chapter. The point imperfections encountered in ionic solids are: (i) chemical impurities (ii) vacant lattice sites and (iii) interstitial ions. Real crystals are always imperfect in some respect. In thermal equilibrium in an otherwise perfect crystal a certain number of lattice vacancies or interstitials are always present, because the entropy is increased by the presence of disorder in the structure and at a finite temperature, the equilibrium condition of a crystal happens to be a state of minimum free energy.

$$F = E - TS. \quad (1)$$

Many properties of ionic solids are controlled as much by imperfections as by the nature of the host crystal which may act only as a matrix for the imperfections. Some of the physical properties affected by imperfections are conductivity, color, luminescence, diffusion of atoms through crystals and mechanical properties. In fact, experimental studies on ionic conductivity are responsible for much of the information on intrinsic lattice disorder (see chapter by C. Ramasastry and coworkers). The intrinsic defects in an ionic crystal are essentially of the following types: (i) Schottky defects: A disorder of this kind implies the missing of two oppositely charged ions from their normal lattice sites, (ii) Frenkel defects: Here, a cation leaves its normal lattice site and

occupies an interstitial position. Since our discussion will be mainly concerned with alkali halides which show only Schottky defects we shall examine the energetics of the vacancy production in detail. The same arguments can be extended to Frenkel disorder encountered in silver or copper halides.

If E_k is the energy required to remove an ion from a site inside the crystal to a site on the surface, nE_k would be the increase in internal energy associated with the production of n isolated vacant sites. According to the Boltzmann formula, the increase in entropy on creating n vacancies is,

$$S = k \log \frac{N!}{(N-n)! n!} \quad (2)$$

where N is the total number of sites. The change in the Helmholtz free energy is given by,

$$F = E - TS = nE_k - kT \log \frac{N!}{(N-n)! n!} \quad (3)$$

At equilibrium (under constant volume conditions), F must be a minimum with respect to changes in n . In ionic crystals it is more favourable to form a separated pair of positive and negative ion vacancies and the formation of pairs keeps the surface of the crystal electrostatically neutral. Thus, we have,

$$n = N \exp (-E_s/2kT) \quad (4)$$

where E_s is the Schottky defect formation energy.

At thermal equilibrium, the concentrations of cation and anion vacancies are governed by the product relation,

$$x_c x_a = \exp (-\Delta F_s/kT) \equiv x_o^2 \quad (5)$$

Thus, if the concentration of either type of vacancy is varied by means of doping with an aliovalent impurity, the concentration of the other is changed in such a way as to still satisfy eqn. (5). This relation assumes a random distribution which is not completely correct because of the long range coulomb interaction between charged defects. Lidiard² has taken this factor into account in modifying eqn. (5) by employing the Debye-Hückel theory of electrolytes. The modified form of the equation is,

$$x_c x_a = x_o^2 \exp [q^2 \kappa / \epsilon kT (1 + \kappa R)] \quad (6)$$

where x_o is the solubility product given by eqn. (5) and κ is

the Debye-Hückel screening constant given by

$$\kappa^2 = 8 \pi q^2 x_0 / V \epsilon kT. \quad (7)$$

Here, q is the effective charge of the vacancies, V is the volume per molecule of the pure salt, R is the 'distance of closest approach' of two vacancies and ϵ is the dielectric constant.

The situation is slightly different when we dope a crystal with an aliovalent impurity. Let us consider the case of doping an alkali halide with a divalent cation where it goes into substitutional solid solution. If the concentration of the impurity is c , then $x_c = c + x_a$. From eqn. (6) we have,

$$x_c = \frac{c}{2} \left[1 + (1 + 4x_0^2/c^2)^{\frac{1}{2}} \right] \quad (8)$$

Similarly, for x_a we can write

$$x_a = \frac{c}{2} \left[(1 + 4x_0^2/c^2)^{\frac{1}{2}} - 1 \right] \quad (9)$$

When $x_0 \ll c$ the concentration of defects is controlled by the amount of dopant added and consequently the concentration of anions is decreased by the addition of impurity and the effect on the ionic conductivity would depend on the carriers with higher mobility. Because of the oppositely charged cation vacancy and the divalent cation (dopant) they form a bound pair and the concentration of such pairs x_k is given by,

$$x_k/x_c(c-x_k) = K_1 \exp \left[-q^2 \kappa / \epsilon kT(1 + \kappa R) \right] \quad (10)$$

after taking into account the Debye-Hückel approximation for long-range correlations resulting from the coulombic interactions. Larger complexes of impurities and vacancies may also occur and are important at low temperatures and at high impurity concentrations.

II. EXPERIMENTAL STUDIES OF POINT DEFECTS

Experimental investigations of electrical conductivity and diffusion on ionic crystals, particularly alkali halides², have been carried out by many authors. These studies provide fairly reliable information on defect energies, defect concentrations and other defect properties. Because of the simplicity and wide applicability of the techniques, we shall discuss the ionic conductivity and the diffusion measurements in some detail (see ref.2

for a detailed discussion).

Perhaps the most convenient method of studying lattice vacancies is to study them in mixed crystals of alkali halides containing divalent additions; e.g., with the addition of controlled amounts of CaCl_2 in KCl , a K^+ vacancy is formed for each Ca^{2+} ion doped in the crystal. The ionic conductivity of such a crystal at different temperatures (and as a function of the dopant concentration at a given temperature) gives important defect parameters. Two important conclusions from the ionic conductivity data can be derived. (i) At not too high temperatures, the ionic conductivity at a given temperature is directly proportional to the amount of divalent addition and the cation vacancies play a key part in the diffusion process. This part of the conductivity curve ($\log \sigma$ vs $1/T$), called the extrinsic region, directly gives the migration barrier height for the cation, because,

$$\sigma = (N_0 e^2 p_v \nu a^2/kT) \exp(-E/kT) \quad (11)$$

where ν is the atomic vibrational frequency, a is the lattice constant, p_v the concentration of vacancies and N_0 the total number of ions per unit volume of the appropriate species in the crystal. In the extrinsic range, the number of vacancies, p_v , is independent of temperature. In the intrinsic range, the proportion of vacancies in an ionic crystal is temperature-dependent and is given by,

$$n_v = \exp(-E_s/2kT) \quad (12)$$

Thus, the slope of plot of $\log \sigma$ vs $1/kT$ in the intrinsic range is given approximately by $E_{\text{mig}}^+ + 1/2 E_s$. On combining measurements in the extrinsic and intrinsic range we can determine the energy of formation of a Schottky pair as well as the jump activation energy. A typical ionic conductivity plot is shown in Fig. 1 and the results derived from such plots in alkali halides are given in Table II.1.

In an accurate evaluation of the conductivity plot, one should take into account the Debye-Hückel correction; when this is included, eqn. (11) becomes,

$$\begin{aligned} \sigma &= N_0 e (x_c \mu_c + x_a \mu_a) \\ &= (4 N_0 \nu_c a^2 e^2 / kT) \exp(-\Delta F_c / kT) (c/2) \left[(1 + 4 x_0^2/c^2)^{\frac{1}{2}} + 1 \right] \\ &\quad + (4 N_0 \nu_a a^2 e^2 / kT) \exp(-\Delta F_a / kT) (o/2) \left[(1 + 4 x_0^2/c^2)^{-\frac{1}{2}} + 1 \right] \end{aligned} \quad (13)$$

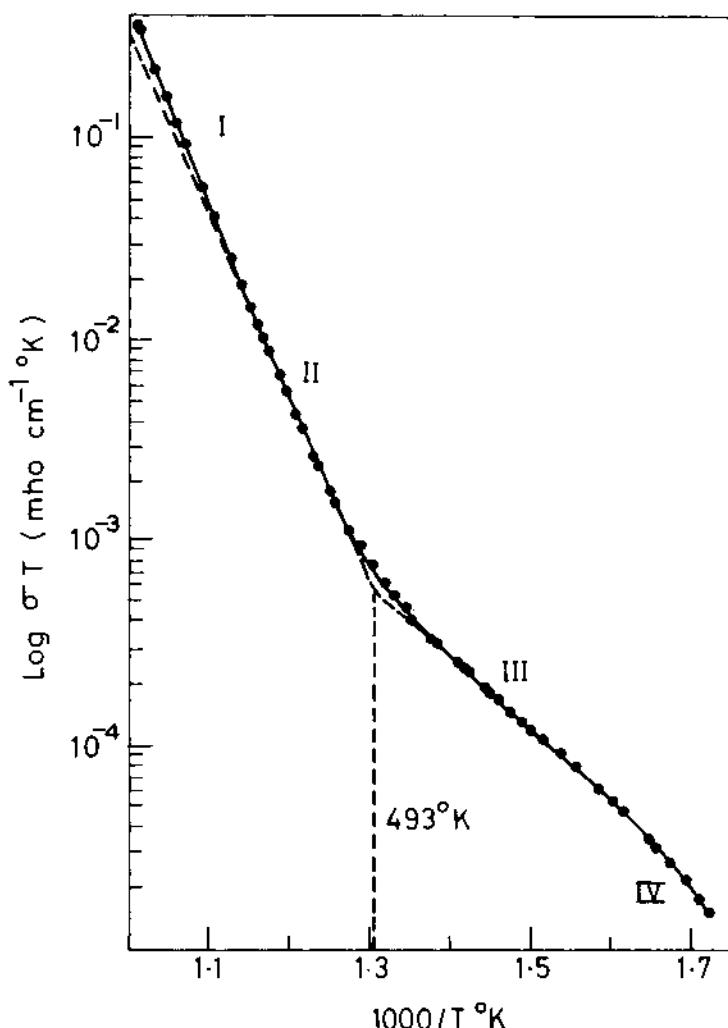


Fig. 1. The ionic conductivity of 'pure' NaCl as a function of temperature ($\log \sigma T$ vs $10^3/T$). (Taken from Ref. 5). Intrinsic conduction occurs in stages I and II and stage III corresponds to conduction by cation vacancies present as a result of the divalent cation impurities. Vacancies become associated with the impurities to form neutral bound pairs in stage IV.

Here, we have assumed the presence of both cation and anion migration. However, in the intrinsic region, if we assume that the cation vacancies are far more mobile than the anion vacancies then eqn. (13) becomes,

TABLE II.1

Experimental Enthalpies of Formation of Schottky Defects*

Compound	Formation Enthalpy E_s	Vacancy Migration Enthalpy E_{mig}^+	
		Cation	Anion
LiF	2.34 - 2.68	0.65	1.1
LiCl	2.2	0.41	-
LiBr	1.8	0.39	-
LiI	1.34	0.38	-
NaCl	2.18 - 2.38	0.66 - 0.76	0.9 - 1.1
NaBr	1.72	0.80	1.18
KCl	2.26 - 2.31	0.71	0.95 - 1.04
KBr	2.30 - 2.53	0.62 - 0.67	0.87 - 0.95
KI	1.60	1.21	-
CsCl	1.86	0.6	0.34
CsBr	2.0	0.58	0.27
CsI	1.9	0.58	0.3

*For original references see Ref. 5.

$$\sigma T = (4N_o \nu_c a^2 e^2/k) \exp(\Delta S_c/k + S_s/k) \exp(-E_{\text{mig}}^+/kT - E_s/2kT) \quad (14)$$

At low temperatures ($c \gg x_o$), we have,

$$\sigma T = (4N_o \nu_c a^2 e^2/k) \exp(\Delta S_c/k) \exp(-E_{\text{mig}}^+/kT) \quad (15)$$

The actual situation is, however, more complex than the treatment given above. At higher temperatures, there is a slight upward curvature which can be attributed to anion conductivity and this feature is also characteristic of Debye-Hückel effects and should be taken into account in the detailed analysis of the ionic conductivity plots. Further, it is advisable to obtain least square fits of the data for all the temperature regions.

Cation self-diffusion by vacancies in NaCl lattices is given by³

$$D = kT f \mu_c x_c / e$$

$$= 4 \nu_c a^2 f x_c \exp(\Delta E_{\text{mig}}^+ / k) \exp(-\Delta S_c / k) \quad (16)$$

where f is the correlation factor. A similar equation can be written down for the anion. Comparison of ionic conductivity data with diffusion measurements gives the correlation factor f and the mechanism of diffusion. Direct determination of the diffusion constant has been carried out in ionic crystals³, using radioactive tracer techniques in which the diffusion of a known initial distribution of radioactive ions is followed as a function of time or distance. It is, however, found in practice that the diffusion constant thus measured and the one inferred from ionic conductivity by means of the Einstein relation do not agree within experimental error, suggesting that a diffusion mechanism which does not involve the transport of charge may also be involved; e.g., the diffusion of pairs of positive and negative ion vacancies would provide such a unit and also the diffusion of an associated divalent ion-vacancy complex. Some of the difference can be accounted for by the increased probabilities for a backward jump⁴. For this process the tracer probability will be low by a fractional amount of the order of $1/Z$ where Z is the number of nearest neighbours. Measurements on doped crystals where the single anion vacancy concentration is small, allow the measurement of the vacancy pair component of the diffusion. In the extrinsic region cation diffusion proceeds both by cation vacancies and neutral impurity-vacancy pairs. Some of the typical defect parameters deduced from the diffusion measurements are also listed in Table II.1.

Other important experimental methods used in the study of ionic crystals are the relaxation methods, NMR, ESR (in the study of transition metal ions in defect complexes) and the accurate simultaneous measurements of macroscopic expansion $\Delta L/L$ and of X-ray lattice parameter expansion $\Delta a/a$ made as the crystal is heated. The last study mentioned gives the total fraction of vacancy defects less the fraction of interstitials generated in the heating up, however, the high temperatures involved often make this a difficult technique to apply to ionic crystals. The detailed descriptions of the above mentioned studies are available in many research papers and several of them have been discussed briefly by Barr and Lidiard⁵.

III. THEORETICAL CALCULATIONS OF THE FORMATION ENERGIES OF SCHOTTKY DEFECTS

The success of Born's classical ionic model in explaining many of the fundamental properties like compressibilities, ionic

radii and elastic constants of ionic crystals, especially those of alkali halides, drew the attention of many workers in extending the model to explain the defect properties of these crystals. The extension of the model for the theoretical estimation of many of the defect properties has been reviewed by Barr and Lidiard⁵ and in an NBS Monograph⁶. Most of the studies reported in the literature^{2,5,6} are on the calculations of Schottky defect formation energies in alkali halides. When a positive or negative ion is removed from its normal lattice site, the nearest ions surrounding the defects are displaced outwards from the lattice positions due to the virtual opposite charge at the defect site (Fig. 2). The positive and negative ions that are far removed from the defect site also move inwards or outwards according to the effective charge on the defect because of the long range nature of the coulomb field.

Jost⁸ was the first to point out the gain due to polarization of the lattice arising out of the ionic displacements around defects. Jost calculated the Schottky formation energy as the difference between the lattice energy E_L and the polarization energy E_p . The lattice was evaluated by treating the vacant site as a spherical cavity of radius R_+ (or R_-) of the missing cation (or anion) in a continuous medium of dielectric constant ϵ (equal to the static dielectric constant of the substance). In the medium surrounding the hole, there is an electric field, $E = -e/\epsilon r^2$, because of the effective charge on the defect. The electric potential at the center of the cavity is $(-e/R_{\pm})(1-1/\epsilon)$ and the

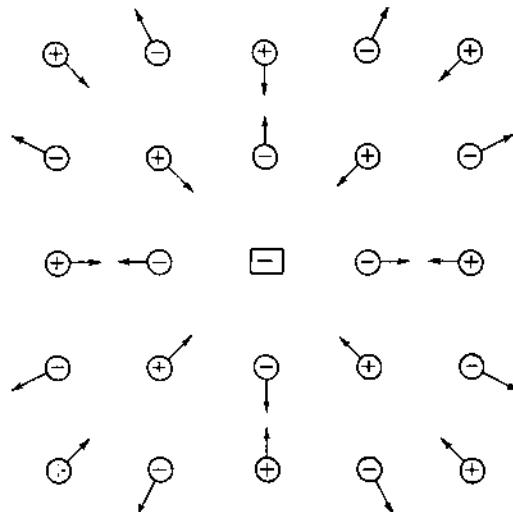


Fig. 2. Displacement of ions around a vacancy.

gain in energy from polarization of the lattice around both cation and anion vacancy is, therefore, $(e^2/2)(1-1/\epsilon)(1/R_+ + 1/R_-)$. Thus, we have,

$$E_s = E_L - (e^2/2)(1-1/\epsilon)(1/R_+ + 1/R_-) \quad (17)$$

This treatment, though crude, clearly demonstrates why E_s is always much smaller than E_L .

Mott-Littleton Method. Owing to the nature of the long range electrical interactions and the presence of strong electric fields at the vacant site ($\sim 10^8$ v/cm) and also because of the lack of information on the radius of the vacant sites, Mott and Littleton⁹ simplified the treatment by dividing the crystal into region I and region II. Region I includes the defect and its immediate neighbours as desired (inclusion of each shell away from the defect in region I constitutes a different degree of approximation), while the second includes the rest of the crystal. The advantage in this division is that for the region II ions, the displacements and electronic moments of the ions are small enough that a harmonic approximation may be used. In region I, on the other hand, the harmonic approximation is not adequate and the assumed potential functions represent the ionic interactions explicitly. We shall now describe the Mott and Littleton method of calculating the E_s in some detail. The Schottky formation energy is given by, $E_s = E_+ + E_- - E_L$, where E_+ and E_- are the energy of formation of the positive and negative ion vacancy. Our main attempt would be to calculate E_+ and E_- taking into account the ionic displacements and electronic moments. The moments and displacements of the ions in region II are obtained from the macroscopic polarization P given by the equation,

$$\vec{P} = (1/4\pi)(1-1/\epsilon)q \vec{r}/r^3 \quad (18)$$

Dividing P in proportion to the electronic and displacement polarizabilities, we see that the total amount (electronic + displacement) associated with an ion at position \vec{r} from the defect is,

$$V/4\pi \left[(1-1/\epsilon)(\alpha_{\pm} + \alpha)/(2\alpha_{\pm} + 2\alpha) \right] q \vec{r}/r^3 = M'_{\pm} q r_o^3 \vec{r}/r^3 \quad (19)$$

The ionic displacements are given by,

$$V/4\pi \left[(1-1/\epsilon)\alpha/(2\alpha_{\pm} + 2\alpha) \right] (\pm q/ze) \vec{r}/r^3 = M'(\pm q/ze)r_o^3 \vec{r}/r^3 \quad (20)$$

α_+ , α_- are the electronic polarizabilities and α is the displacement polarizability which can be calculated from the Born model. Thus, for ions of charge $\pm Ze$,

$$\alpha = Z^2 e^2 / p \quad (21)$$

where the force constant p is given by,

$$p = f \left[\phi''(r) + 2 \phi'(r)/r \right] \quad (22)$$

Here, f is 4 for NaCl structure, $16/3$ for CaCl structure, and $8/3$ for ZnS structure; $\phi(r)$ is the total non-coulombic potential given by,

$$\phi(r) = b \beta_{ij} \exp \left[(r_i + r_j - r)/\rho \right] - (c_{+-}/r^6) - (d_{+-}/r^8) \quad (23)$$

where b and ρ are Born repulsion coefficients, r_i , r_j are the crystal radii of the ions and β_{ij} is the Pauling's coefficient. The last two terms are the van der Waals terms (see the chapter by C.N.R. Rao on ionic crystals). Here, the potential only due to the first neighbours are included since $\phi(r)$ is insignificant beyond second neighbours.

We can now write down the equations of equilibrium assuming the form (eqn. 23) for the non-electrostatic interactions of the ions of region I; for the first order approximation we choose region I containing the vacant site and its first six neighbours. If the displacements of the ions in region I are taken as ξr_o and the induced electronic moment as $m \epsilon r_o$, then the outward electrostatic force, F_e , on $\langle 100 \rangle$ ion for a substance $M^+ X^-$ (with NaCl structure) can be computed as,

$$F_e = e^2/r_o^2 \left\{ (5/4 + \sqrt{2})/(1 + \xi)^2 - 4(1 + \xi)/[1 + (1 + \xi)^2]^{3/2} \right. \\ \left. - 1/(2 + \xi)^2 + 2.3713 m_{\pm}^+/(1 + \xi)^3 - (0.388 M_{\mp}^+ + 1.965 M_{\pm}^+) \right\} \quad (24)$$

for cation and anion vacancy respectively. The first three terms result from the ionic displacements of the five neighbours, the fourth term from the electronic moments on these ions and the last term from the displacements + electronic moments of region II ions. The non-electrostatic force F_r is given by,

$$F_r = (-b_1/\rho) \exp \left[-(1 - \xi - M'/4) r_o/\rho \right] + 4b_1(\xi + M'/2\sqrt{2}) r_o/(\rho d_1)$$

$$x \exp(-d_1/\rho) + 6c_{+-} / [r_0(1-\xi - M'/4)]^7 - 24 r_0(\xi + M'/2\sqrt{2})/d_1^8 \quad (25)$$

where

$$b_1 = \beta_{+-} b \exp[(r_+ + r_-)/\rho] \text{ and } d_1 = r_0 [(1-M'/2\sqrt{2})^2 + (\xi + M'/2\sqrt{2})^2]^{\frac{1}{2}}.$$

The total force is given by $F_e + F_r$ and this now contains two unknowns, namely ξ and m . The electronic moment for the region I ions is given by,

$$m_{\pm} er_0 = \alpha_{\pm} (F_e/e) \quad (26)$$

By substituting this equation for m we can obtain the relaxation ξ by using the equilibrium condition, $F_e + F_r = 0$.

From the values of ξ and m thus obtained, the electrostatic energy E_e , and the non-coulombic energy, E_{nc} , at the positive ion vacancy can be written as

$$E_{e_{\pm}} = e^2/r_0 \left\{ \alpha_M - 6/(1+\xi_{\pm}) + 6m_{\pm}/(1+\xi_{\pm})^2 + 6.3346 M'_{\pm} + 4.1977 M'_{\pm} \right\} \quad (27)$$

and

$$E_{nc_{\pm}} = 6b_1 \exp[-(1+\xi_{\pm}) r_0/\rho] - 6c_{+-} / [r_0(1+\xi_{\pm})]^6 \quad (28)$$

Here + and - denote the positive and negative ion vacancy. The terms preceding M'_{\pm} are the lattice summations of the type

$$S_{\pm} = \sum_{\pm}^* 1/r_{\pm}^4, \quad (29)$$

the asterisk (*) denoting that the first neighbour ions are excluded from the sum. Thus, we have,

$$E_{\pm}^V = E_{e_{\pm}} - E_{nc_{\pm}} \quad (30)$$

An alternative method of obtaining relaxations and electronic moments is by the energy minimization procedure proposed by Kurozawa¹⁰ and subsequently modified by Boswarwa and Lidiard¹¹. In this procedure, we write an energy function W of the defect solid

in the form,

$$W = W_1 + W_2(\xi, m) + W_3(\xi, m, \eta, \mu) + W_4(\eta, \mu) \quad (31)$$

Here, W_1 is the energy of the defect in a rigid non-polarizable lattice and the other terms represent the change in energy brought about by relaxation and polarization of the lattice; η and μ are the displacements and moments in region II. The term $W_2(\xi, m)$ is the energy of region I alone, $W_3(\xi, m, \eta, \mu)$ is the energy of interaction between region I and II and $W_4(\eta, \mu)$ is the energy of region II along. At equilibrium, $(\partial W / \partial \xi) = 0$; $(\partial W / \partial m) = 0$; $(\partial W / \partial \eta) = 0$ and $(\partial W / \partial \mu) = 0$. We thus obtain the displacements and moments. The main advantage in this procedure is that we are able to add directly the change in the energy upon relaxation to the energy of the defect configuration with respect to the perfect lattice. Both the force balance and the energy minimization methods should give identical results. This can, however, be realized only when the description of displacements and moments is exact; in any approximate treatment the two approaches may differ. Some of the typical results of the theoretically estimated E_g values are given in Table III.1; these values may be compared with the experimental values in Table II.1.

Refinements in Theoretical Calculations. In ionic crystals, the polarization of region II contributes appreciably to the energy of formation of a vacancy and we should, therefore, pay particular attention to this description. In fact, many of the modifications of the Mott-Littleton approach are related to the treatment of the polarization of region II ions¹¹. In principle, the best description will be to solve the equations of equilibrium for region II by Kanzaki's method¹² of lattice statics, but this has not been attempted in detail.

Upto this point, we have considered the vacancy only as a source of electrical polarization; in addition to its effective charge, the vacancy will generally have a different radius r_v and the surrounding ions of the lattice will undergo elastic displacements. The direction of these displacements is the same for negative and positive ions. Hence, the distance between (000) and (100) ions is no longer $r_o = r_+ + r_-$, but $r_o + x = r_v + r_-$. The relative displacement of the neighbours is, therefore, $\xi_{(100)} = x/r_o$ and from the theory of elastic continuum, the displacement of the far ions at the position \vec{r} ($n m l$) is given by

$$\xi_{(nml)} = \xi_{(100)} / (n^2 + m^2 + l^2) \quad (32)$$

Each displaced ionic charge gives a dipole moment of the magnitude,

TABLE III.1

Theoretically estimated formation energies of Schottky defects*

Substance	Theoretical E_s
LiF	1.24
LiCl	1.08
LiBr	0.86
LiI	0.64
NaF	2.29 - 2.97
NaCl	1.30 - 2.20
NaBr	1.31 - 2.01
NaI	1.35 - 1.60
KF	2.04 - 2.51
KCl	1.75 - 2.26
KBr	1.68 - 2.14
KI	1.70 - 1.92
RbF	1.85 - 2.26
RbCl	1.98 - 2.24
RbBr	1.87 - 2.11
RbI	1.75 - 1.97
CsF	1.65 - 2.11
CsCl	1.03 - 1.78
CsBr	1.03 - 1.65
CsI	1.11 - 1.61

*See Ref. 5 for references to original papers.

$$\mu_{(nml)} = r_o e \xi_{(nml)} = r_o e \xi / (n^2 + m^2 + l^2) \quad (33)$$

where $\xi = \xi_{(100)} \cdot \frac{(e^2/r_o^2)}{(\epsilon_0^2/r_o^2)} \xi_{(1.965 - 0.388)}$, in the equation for F_e (eqn. 24) and $(e^2 \xi / r_o)$ (6.3346 - 4.1977) in the polarization part of E_e in eqn. (28). Here, the relaxation ξ is not broken into electrical and elastic parts and it denotes directly the relaxation of the nearest neighbour ions. Though Brauer's modifi-

cation gave better E_s values compared to other theoretical estimates, it gives a poor description of the elastic strain field. Hardy and Lidiard¹⁴ have regarded the solid as an isotropic elastic continuum and accordingly the ion at distance \vec{r} is displaced radially outwards by the distance,

$$x = \xi r_0 = k r_0^{3/2} \quad (34)$$

Here, k measures the elastic strength of the vacancy; k is fixed by ensuring that the displacements assumed for region II join smoothly on to those in region I thus, can be written as,

$$\xi = k + M' \quad (35)$$

whereas according to Brauer¹³ $\xi = k$ which overemphasizes the elastic component in the displacement of more distant ions. This modification improves the E_s values by about 0.2 eV. This approximation is favoured because the sum of the elastic strengths for positive and negative ion vacancies $k_+ + k_-$ obtained by this treatment agrees well with experimentally estimated value for the free volume of formation of a vacancy.

The relative motion of the positive and negative ion against one another leads to a distortion of the otherwise spherical charge distribution around each ion and this creates additional electronic dipoles, namely, "deformation dipoles", which actually oppose those induced by an electric field. Szigeti pointed out that the effect of these dipoles on the dielectric constant is as though the actual charge e is replaced by an effective charge e^* which is less than e for most of the alkali halides. Empirical values of e^* are known from the long wave optical lattice vibration⁵. Thus, the total polarizability is $\alpha_+ + \alpha_- + 2 e^{*2}/p$; according to the deformation dipole model of Hardy and Lidiard¹⁴ the ionic displacement polarizability becomes e^*e/p and the contribution of the deformation dipole to the total polarizability is given by $-2 e^* (e - e^*)/p$, which is centered mostly on the anions. Inclusion of this alters M'_+ and M'_- and the additional deformation on the region I ions is then estimated by lattice dynamical studies in line with the treatment of Hardy and Lidiard¹⁴.

Another important factor to be considered in these calculations is the contribution due to overlap repulsion in the potential energy expression of a solid. The use of Goldschmidt radii give poor estimate of E_s and recent workers use Tosi's radii which seem to fit experimental data better⁶. The importance of the van der Walls term has been realized only in recent times^{15,16} and this also improves the E_s by about 0.1 to 0.2 eV. Several forms of the repulsion potential including the Born-Mayer-Verwey form have been employed in the literature and their relative merits

have been discussed^{11,15,16}.

One of the criticisms against Mott-Littleton calculations is the use of continuum approximation right from the second shell of ions^{11,15,16}. Though the inclusion of the elastic terms in Lidiard's treatment¹⁴ removes this ambiguity to some extent, a large region I would be mathematically more exact. Such calculations have been done recently¹⁷, but they show small differences between E_s^V and E_s^U whereas the usual procedures predict a difference of about 0.3 eV in accordance with experimental observations.

The immediate neighbours of charged defects are subjected to strong electric fields and the induced electronic moments of the nearest ions would, therefore, be less than those calculated from low field polarizabilities. In fact, it has been shown that by arbitrarily reducing the polarizability of the normal value of O²⁻ ion in MgO by 1/3 gives a better E_s value. A good test of this phenomenon has been made in the case of Li⁺ in KCl where the Li⁺ is known to occupy an off-center position⁶. The use of the central dipole approximation implied in the Mott and Littleton method leads to negative divergence in the energy of such systems. This difficulty was overcome by the use of variable polarizability in proportion to the overlap of the ions. Probably such modifications are necessary in the case of lithium halides which consistently give low estimates of E_s .

IV. DEFECT INTERACTION ENERGIES

In the creation of Schottky defects, we have considered the cation and anion vacancies to be far apart to exclude any mutual interaction. However, at higher vacancy concentrations, the cation and anion vacancies may come nearer and form a vacancy pair. When they are at nearest neighbour distances, we call the configuration a ground state vacancy pair. The order of the excited states corresponds to the distance between the two vacancies. There is sufficient experimental evidence for the existence of vacancy pairs from diffusion studies (the pairs do not contribute to conductivity because of their neutral charge). If we assume absence of lattice polarization, then the interaction energy between the two defects is given by, $E_c = -e^2/\epsilon r_d$, where r_d is the distance between the defects. Because of the unsymmetrical nature of such defects, however, there is considerable lattice polarization and consequently the interaction energy is lower than E_c and reach the value of E_c only at a far distance. Since there is no reliable experimental method for determining the interaction energies either between two vacancies¹⁸ or between a vacancy and a divalent impurity¹⁹, considerable attention has been

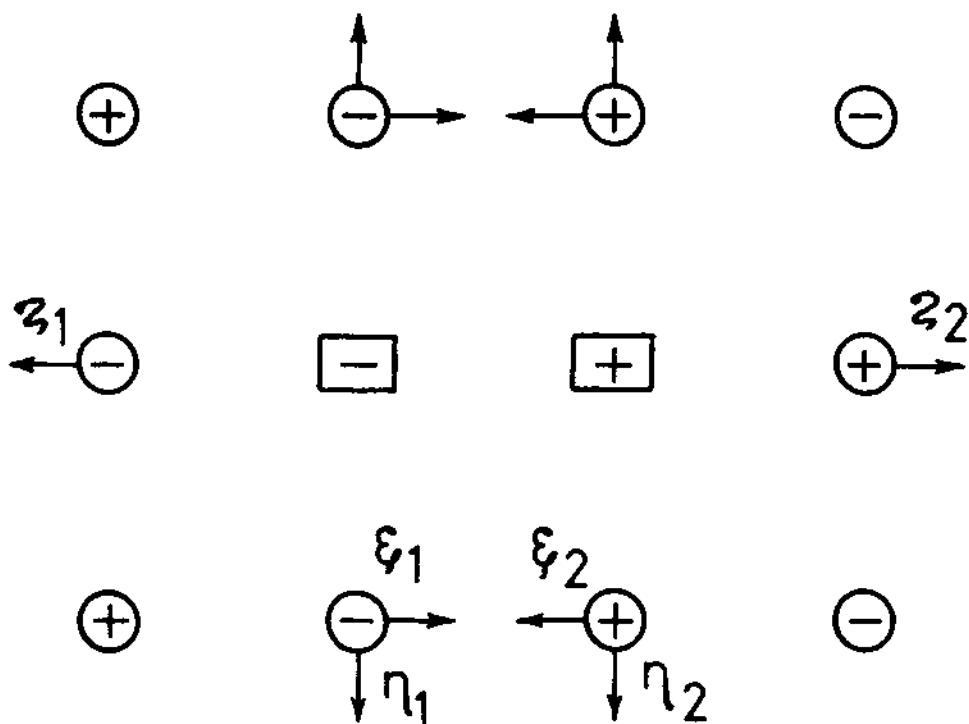


Fig. 3. Relaxations around a bound ion pair.

focussed on the theoretical estimation of these quantities especially in the ground state.

Energy for the ground state interaction of a vacancy pair, E_g^V , is calculated as the difference in energies for the creation of a positive ion vacancy in a perfect lattice and the energy to create the same in the neighbourhood of a negative ion vacancy. The latter is obtained as the average of the potential energy at the site of the ion before (E_1^V) and after (E_2^V) its removal (Fig. 3). Most of the calculations on E_g^V have been carried out by giving a uniform (restricted) relaxation^{15,18} η by setting $\xi_1 = \xi_2 = 0$ and $\eta_1 = \eta_2 = z_1 = z_2 = \eta$. Though the difference in the values of E_g^V by the free relaxation¹⁶ (as shown in Fig. 3) and by restricted relaxation approximations is quite negligible, the free relaxation configuration is theoretically more justified. Interaction energies in the ground state for impurity-vacancy complexes¹⁹ and vacancy pairs^{15,16,18} are shown in Tables IV.1 and IV.2.

TABLE IV.1

Theoretical estimates of ground state
binding energies of vacancy pairs

Salt	E_g^{vp} (in eV)
LiCl	0.37
NaCl	0.54 - 0.60
KCl	0.65 - 0.76
KBr	0.58
KI	0.50
RbCl	0.83
RbBr	0.73
RbI	0.55

TABLE IV.2

Binding energies (in eV) of cation vacancies
to divalent cations in NaCl and KCl

Solute	Salt	NaCl		KCl	
		Theory	Expt.	Theory	Expt.
Cd		0.38	0.3-0.6	0.32	0.5
Ca		0.38	0.3-0.6	0.32	0.5
Sr		0.43	0.5	0.38	0.4

V. MIGRATION ENERGIES

On the application of an electric field, cations and anions move to their respective adjacent vacant sites and thus contribute to the ionic conductivity. In alkali halides, anion migration

becomes important only at high temperatures since the activation energy for migration is considerably more than that for cation migration. The energy for vacancy migration, E_{mig} , of an alkali halide is generally evaluated as the difference in the energy of the crystal in the saddle point configuration (Fig. 4) and the energy of the solid containing a single vacancy, i.e., the difference between the energy to extract two next nearest neighbour ions and to introduce one of them into the saddle point configuration between the two neighbouring vacancies and the energy to extract a single ion from a perfect crystal. Most of the theoretical studies^{15,20,21} have been carried out on the migration through the <100> direction and they do not take into consideration the elastic terms and deformation dipoles in the long polarization of the far ions. The magnitude of the migration barrier height has been proved to be highly dependent on the kind of repulsive potentials used^{5,6}. Most of the theoretical studies, however, do not give the expected difference between the cation and anion migration barrier heights. This is likely to be due to the inadequate treatment of the polarization effects⁵; it is also likely that the use of the exponential form for the repulsive potential at shorter distances may not be appropriate⁶. Some authors have calculated barrier heights employing the inverse form of the Born-Mayer-Verwey repulsion potential of the type, $B r^{-12}$, but there appears to be no

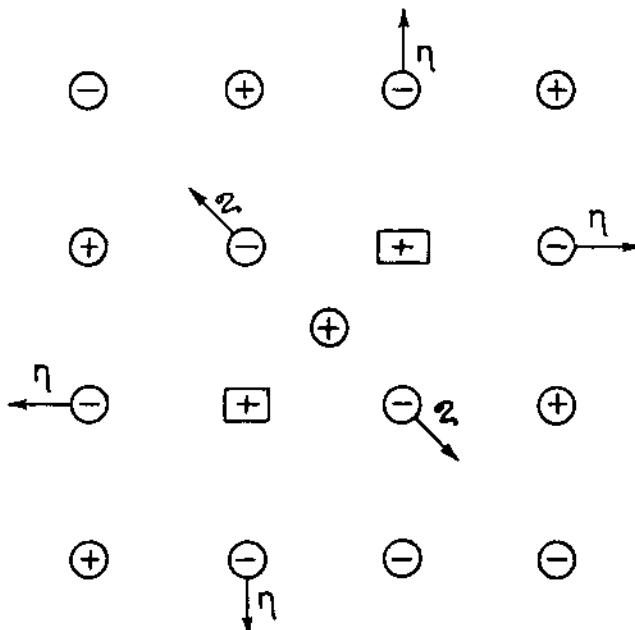


Fig. 4. Saddle point configuration for migration of a cation to a single vacancy.

justification for using such potentials in the solid phase⁵. The calculated barrier heights for migration of cations and anions into single vacancies for some of the alkali halides are listed in Table V.1.

TABLE V.1

Activation energies (in eV) for the
migration into single vacancies^(a)

Salt	Cation vacancy		Anion vacancy	
	Theory	Expt.	Theory	Expt.
NaCl	0.30*	0.66-0.76	0.72*	0.90-1.11
	0.70-0.92		0.75-1.00	
KCl	0.67*	0.71	0.76	0.95-1.04
	0.77-0.96		0.80-1.21	
RbCl	0.87-1.02		0.99-1.12	

* With BMV potential for overlap repulsion

(a) See Refs. 5 and 6 for greater details.

At higher temperatures, due to the increased concentration of vacancy pairs, migration of a cation or anion into a vacancy pair has been noted and calculations show comparable or higher activation energies compared to the single vacancy migration^{15,22}; diffusion studies also seem to justify this trend.

VI. INTERSTITIAL DEFECTS

Calculations on interstitial defects are relatively few. Calculations of formation energies in CaF₂ by Franklin²³ show predominant anion Frenkel defects. Other important systems are the silver halides which show cation Frenkel defects. A few calculations have been reported on the formation and migration of interstitial defects in such systems; the formation energies are nearly half the value one generally encounters with Schottky defects. The calculation of migration barrier heights for interstitial migration through different paths²⁴ shows evidence for migration through interstitialcy replacement in a <111> direc-

tion both by collinear and non-collinear paths. These calculations are not sufficiently rigorous to warrant definite conclusions. Some of the available values of the energies of formation of interstitial defects are listed in Table VI.1.

TABLE VI.1

Frenkel defect formation energies

Compound	E_f (in eV)		Experimental
	Cation	Anion	
CaF_2	2.7 ± 0.4	7.5 ± 0.8	2.8
SrF_2	2.5 ± 0.4	8.6 ± 0.8	2.3
BaF_2	2.3 ± 0.4	8.8 ± 0.8	1.9
AgCl	-	$1.83 - 3.23$	1.2-1.7

VII. CONCLUDING REMARKS

In spite of various attempts to theoretically calculate the energies of formation, interaction and migration of defects, we still have a number of questions unanswered. For example, we do not know the best form of the repulsive potential to be employed in various situations; the use of very hard repulsion term (provided by BMV potential) in certain instances seems to be quite arbitrary. The calculated Schottky formation energies in the case of lithium halides are particularly low. It is likely that the point dipole model fails (in such systems) to describe contributions from polarization term properly. Thus, the case of Li^+ doped KCl where the Li^+ is in off-center position provides an instance where there is definite need to employ a different treatment of the polarization contribution^{25,26}.

Point defect calculations are truly applicable to completely ionic solids (such alkali halides) but would not be valid in oxides and other systems where there is considerable covalency; calculations become particularly difficult when large displacements occur as in MgO (same is true of lithium halides). Theoretical calculations on non-stoichiometric defects would be much more difficult. Even in stoichiometric oxides like TiO , VO and NbO where vacancy concentrations are large, it would be difficult to extend theore-

tical calculations on the basis of point defect theory of the type discussed earlier in this chapter. Firstly, the lack of adequate parameters like ionic radii, compressibilities, etc. and the presence of the considerable overlap between ions would render the estimation of overlap repulsion terms difficult. Besides, in such systems, there will be considerable defect ordering (formation of clusters) which undoubtedly result from large interaction energies between defects. We feel that low values for the formation energies of Schottky pairs²⁷ are not the cause for the high concentrations of defects in these oxides. We feel that it is worthwhile to examine the origin of such defect ordering. Another interesting aspect would be to calculate the distribution of various excited states of vacancy pairs in ionic solids; such calculations would provide a nearly complete partition function for a Schottky solid. Considerable work is needed for the proper estimation of defect volumes, particularly in the case of anisotropic defects like bound vacancy pair. In addition to some of the problems mentioned here, point defect calculations on ionic solids with 8-8, 4-4 coordination will be fruitful; in some of these ionic solids, we do not even know whether Schottky or Frenkel defects are predominant.

REFERENCES

1. C. Kittel, "Introduction to Solid State Physics", 3rd edition, John Wiley and Sons, New York, 1966.
2. A.B. Lidiard, *Handbuch der Physik*, 20, 246 (1957).
3. J.R. Manning, "Diffusion Kinetics for Atoms in Crystals", Van Nostrand, Princeton, 1968.
4. R.E. Howard and A.B. Lidiard, "Reports on Progress in Physics", Vol. 27, Institute of Physics and Physical Society, London, 1964.
5. L.W. Barr and A.B. Lidiard, "Defects in Ionic Crystals", *Physical Chemistry - An Advanced Treatise*, Academic Press, Vol. 10, 1968.
6. "Calculations of Properties of Vacancies and Interstitials", *Proceedings of Skyland Conference*, National Bureau of Standards, Misc. Publn. No. 287, 1966.
7. M. Born and K. Huang, "Dynamical Theory of Crystal Lattices", Oxford University Press, 1954.
8. W. Jost, *J. Chem. Phys.*, 1, 466 (1933).

9. N.F. Mott and M.J. Littleton, Trans. Faraday Soc., 34, 485 (1938).
10. T. Kurozawa, J. Phys. Soc., Japan, 13, 153 (1962).
11. I.M. Boswarwa and A.B. Lidiard, Phil. Mag., 15, 825 (1967).
12. H. Kanzaki, J. Phys. Chem. Solids, 2, 24 (1957); *ibid*, 2, 37 (1957).
13. P. Brauer, Z. Naturf., A7, 372 (1952).
14. J.R. Hardy and A.B. Lidiard, Phil. Mag., 15, 825 (1967).
15. K.J. Rao and C.N.R. Rao, Phys. Stat. Solidi, 28, 157 (1968).
16. S. Ramdas and C.N.R. Rao (to be published).
17. A. Scholz, Phys. Stat. Solidi, 25, 285 (1968).
18. M.P. Tosi and F.G. Fumi, Nuovo cim., 7, 95 (1958).
19. M.P. Tosi and G. Airoldi, Nuovo Cim., 8, 584 (1958).
20. R. Guccione, M.P. Tosi and M. Asdente, J. Phys. Chem. Solids, 10, 162 (1959).
21. M.P. Tosi and M. Doyama, Phys. Rev., 151, 642 (1966).
22. K. Tharmalingham and A.B. Lidiard, Phil. Mag., 6, 1157 (1961).
23. A.D. Franklin, J. Phys. Chem. Solids, 29, 823 (1968).
24. K. Tharmalingham, J. Phys. Chem. Solids, 25, 255 (1964).
25. R.J. Quigley and T.P. Das, Solid State Comm., 5, 487 (1967).
26. R.D. Hatcher and G.J. Dienes, Phys. Rev., 134, A214 (1964).
27. P.G. Dickens, R. Heckingbottom and J.W. Linnett, Trans. Faraday Soc., 64, 1489 (1968).

DIELECTRIC PROPERTIES

E. C. Subbarao

Department of Metallurgical Engineering

Indian Institute of Technology, Kanpur, India

I. DEFINITIONS AND UNITS

The response of a material at the electronic, atomic, molecular and microscopic level to an applied electric field constitutes its dielectric properties. Polarization and dielectric loss are the phenomena of interest and they are studied as a function of frequency (from d.c. to near optical frequencies) and of temperature. We will first introduce the necessary dielectric parameters¹⁻⁴.

Maxwell's equations give

$$\underline{D} = \epsilon \underline{E} \quad (1)$$

where \underline{D} is the dielectric displacement or electric flux density (coul/m^2), \underline{E} is the electric field strength, volts/m and the proportionality constant and ϵ is the dielectric constant of the medium, farad/m.

If the medium to which the electric field is applied is free space (vacuum), the proportionality constant of eqn. (1) is called the dielectric constant of vacuum, ϵ_v , of value 8.854×10^{-12} farad/m. The dielectric constant of a material may now be expressed by ϵ_r relative to that of vacuum by

$$\epsilon_r = \frac{\epsilon}{\epsilon_v} \quad (2)$$

so that eqn. (1) can be rewritten as

$$\underline{D} = \epsilon_v \epsilon_r \underline{E} \quad (3)$$

Table I.1 Dimensions and Units³

Quantity	Symbol	Defining equation	Dimensions in mks system		Multiplication factor for converting ratio-lized mks units to Esu units	Ratio esu/emu (c=3x10 ¹⁰)
			Primary unit	Derived unit		
Capacitance	C	$C = \epsilon/V$	$\text{sec}^2 \text{coul}^2 / \text{kgm}^2$	farad	9×10^{11}	1×10^{-9}
Complex dielectric constant	ϵ^*	$\epsilon^* = \epsilon' - i\epsilon''$	$\text{sec}^2 \text{coul}^2 / \text{kgm}^2$	farad/m	$36\pi \times 10^9$	$4\pi \times 10^{-11}$
Complex permeability	μ^*	$\mu^* = \mu' - i\mu''$	$\text{kgm}^2 / \text{coul}^2$	henry/m	$\frac{1}{36\pi} \times 10^{-13}$	$\frac{1}{4\pi} \times 10^7$
Conductivity	σ	$\sigma = J/E$	$\text{sec}^2 \text{coul}^2 / \text{kgm}^2$	mho/m	9×10^9	1×10^{-11}
Current	I	$I = dQ/dt$	coul/sec	amp	3×10^9	1×10^{-1}
Current density	J	$J = dI/dA$	coul/sec m^2	amp/m^2	3×10^{11}	1×10^1
Dielectric constant	ϵ	$\epsilon = \frac{D}{E}$	$\text{sec}^2 \text{coul}^2 / \text{kgm}$	farad/m	$36\pi \times 10^9$	$4\pi \times 10^{-11}$
Dissipation factor	D	$D = \tan \delta$				
Electric charge	Q	Primary unit			3×10^9	1×10^{-1}
Electric dipole moment	P	$P = Qd$	coul m		3×10^{-1}	1×10^1
Electric field strength	E	$E = F/q$, or $E = pE$	$\text{kgm/sec}^2 \text{coul volt/m}$	volt/m	$\frac{1}{3} \times 10^{-4}$	1×10^6
Displacement (Electric flux density)	D	$D = \int_V P dV$	coul/m^2	farad/m^2	$12\pi \times 10^5$	$4\pi \times 10^{-5}$

Table I.1 ...contd...

Electric loss factor	$\epsilon'' = \frac{J_{\text{loss}}}{\omega e}$	$\text{sec}^2 \text{coul}^2 / \text{kgm}^3$	farad/m	$36\pi \times 10^9$	$4\pi \times 10^{-11}$	c^2
Electric polarization	$P = D - \epsilon_v E$	coul/m^2	farad/m	$12\pi \times 10^5$	$4\pi \times 10^{-5}$	c
Electric potential	$\phi = - \int_{\infty}^r E \cdot dl$	$\text{kgm}^2/\text{sec}^2 \text{coul}$	volt/m	$\frac{1}{300}$	1×10^8	$1/c$
Electric susceptibility	$\chi_e = \epsilon_r^{-1}$	$\text{kgm}^2/\text{sec}^2$	volt			
Electrostatic energy	$U = \frac{1}{2} \int E \cdot dl$	$\text{kgm}^2/\text{sec}^2$	joule	1×10^7	1×10^7	1
Force	$F = d(mV)/dt$	kgm/sec^2	newton	1×10^5	1×10^5	1
Loss tangent	$\tan \delta = \frac{\epsilon''}{\epsilon'}$					
Power	$P = \frac{dw}{dt}$	$\text{kgm}^2/\text{sec}^3$	joule/sec = watt	1×10^7	1×10^7	1
Q of dielectric (quality factor)	$Q = \frac{1}{\tan \delta}$					
Relative dielectric constant ϵ_r	$\epsilon_r = \frac{\epsilon'}{\epsilon_v}$					
Resistance (d.c.)	$R = V/I$	$\text{kg} \cdot \text{m}^2/\text{sec} \cdot \text{coul}^2 \text{ ohm}$		$\frac{1}{9} \times 10^{-11}$	1×10^9	$1/c^2$
Resistivity	$\rho = 1/\sigma'$	$\text{kg} \cdot \text{m}^3/\text{sec} \cdot \text{coul}^2 \text{ ohm m}$		$\frac{1}{9} \times 10^{-9}$	1×10^{11}	$1/c^2$
Torque	$T = F \times d$	$\text{kg} \cdot \text{m}^2/\text{sec}^2$	newton m	1×10^7	1×10^7	1

Note that the relative dielectric constant, ϵ_r , is a dimensionless quantity.

The application of an electric field to a material causes a displacement of electric charges, giving rise to the creation or reorientation of dipoles in the material. This property is designated as polarization, P , defined as dipole moment per unit volume, coul m/m³ or coul/m². When an electric field is applied to a material, the dielectric displacement is the sum of the effect on vacuum and that on the material, so that eqn. (1) becomes

$$\underline{D} = \epsilon_v \underline{E} + \underline{P} \quad (4)$$

$$= \epsilon_v \epsilon_r \underline{E} \quad (3)$$

$$\text{Thus, } \underline{P} = \epsilon_v (\epsilon_r - 1) \underline{E} \quad (5)$$

$$= \epsilon_v \chi_e \underline{E} \quad (6)$$

where χ_e is electric susceptibility, a dimensionless quantity. It is seen that \underline{D} and \underline{P} have the same units.

The polarization P of a material may be expressed in terms of its elementary dipole moments, p , (coul m) by

$$\underline{P} = N \underline{p} \quad (7)$$

where N is the number of dipoles per unit volume. The electrical pliability of a material in response to an electric field E is responsible for unit dipole moments, so that

$$\underline{p} = \alpha \underline{E} \quad (8)$$

where α is the polarizability, coul m²/volt or farad m². Thus,

$$\alpha = \frac{\underline{p}}{\underline{E}} = \frac{\epsilon_v \chi_e \underline{E}}{N \underline{E}} = \frac{\epsilon_v \chi_e}{N} \quad (9)$$

The defining equations for the dielectric and magnetic parameters, the primary and derived units, and the conversion of the mks to the esu and emu units are summarized in Table I.1 adopted from Von Hippel³.

The application of an electric potential V to a parallel plate condenser (Fig. 1) results in the accumulation of charges on the plates, Q , such that

$$Q = CV \quad (10)$$

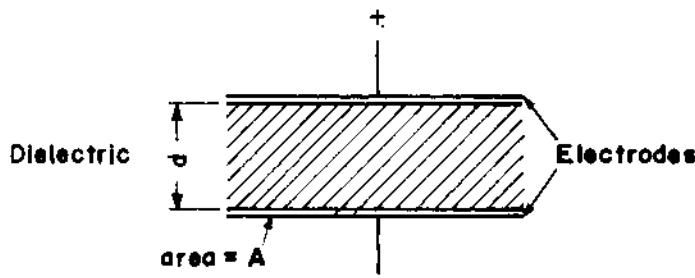


Fig. 1. Parallel plate capacitor

where C is the proportionality constant, called capacitance, in farads when Q is expressed in coul and V is volts. If the space between the parallel plates is vacuum, eqn. (10) becomes

$$Q' = C_v V \quad (11)$$

where C_v is the capacitance of free space, which is given by the geometrical factors viz., A , the area of the condenser plates, and d , the separation of the plates. Neglecting fringing effects,

$$C_v = \frac{A}{d} \epsilon_v \quad (12)$$

so that, in the case of a condenser with a dielectric filling the space between the parallel plates,

$$C = C_v \frac{\epsilon}{\epsilon_v} = C_v \epsilon_r \quad (13)$$

$$\text{and } Q = C_v \epsilon_r V \quad (14)$$

Knowledge of the geometric factors yields the value of C_v and this, together with a measurement of charge of a condenser for known applied voltage, gives the relative dielectric constant of the dielectric material.

II. MECHANISMS OF POLARIZATION

The centers of gravity of positive charges and of negative charges coincide in neutral atoms and symmetric molecules. Application of an electric field causes relative displacement of these charges, leading to the creation of dipoles and polarization. Unsymmetric arrangement of atoms in a molecule results in a dipole even in the absence of an external field and in those cases, the applied electric field tends to orient the permanent dipole moments parallel to the field direction. In the following is discussed

the four basic polarization mechanism, viz. electronic, ionic orientational and interfacial (or space charge).

Electronic Polarization, P_e . The application of an electric field E displaces the center of the positive charges $+Ze$ and the center of the negative charges $-Ze$ by a distance d (Fig. 2). At

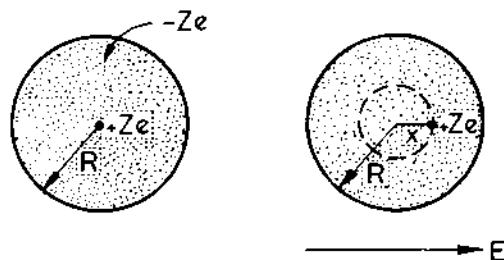


Fig. 2. (a) Positive ($+Ze$) and negative ($-Ze$) charges in a neutral atom of radius R . (b) The relative displacement of the positive and negative charges by a distance d ($d \ll R$) when an electric field E is applied.

equilibrium, the force of the electric field ($Z e E$) and the coulombic force between the charges are equal. That is,

$$Z e E = (Ze)^2 \frac{d}{\epsilon_v 4\pi R^3} \quad (15)$$

where R is the radius of the electron cloud

$$Z e d = \epsilon_v 4\pi R^3 E = p_e \quad (16)$$

$$p_e = N p_e = N 4\pi R^3 E \quad (17)$$

$$\text{and } \alpha_e = \epsilon_v 4\pi R^3 \quad (18)$$

Thus the electronic polarization is dependent only on the atomic radius, and increases with atomic volume. Since atomic volumes are nearly independent of temperature, p_e is also independent of temperature.

Ionic Polarization, P_i . In a diatomic molecule, such as NaCl, the ions Na^+ and Cl^- are separated by a distance l . When a field E is applied, ions are displaced from their equilibrium positions, so that the dipole becomes elongated to $l + d$. The polarization resulting from the additional displacement of the ions is called the ionic polarization, P_i and is again independent of temperature.

Orientational Polarization, P_o . An electric field E applied

in the x direction tends to align the permanent dipole designated by $+ \mathbf{q}$ and $- \mathbf{q}$ making an angle θ with the x axis. The potential energy, U , of the dipole of moment p_p is given by

$$U = - |p_p| |E| \cos\theta \quad (19)$$

The aligning effect of the field on the dipoles is counteracted by the randomizing influence of temperature. The number of dipoles pointing at a solid angle $d\Omega$ is given by the Boltzmann expression

$$N = A e^{-U/kT} d\Omega \quad (20)$$

$$= A e^{-|p_p| |E| \cos\theta/kT} 2\pi \sin\theta d\theta \quad (21)$$

$$\text{since } d\Omega = 2\pi \sin\theta d\theta \quad (22)$$

Each of these dipoles constitutes an orientational polarization of $|p_p| \cos\theta$ in the x direction, so that the average moment, \bar{p}_p , of each molecule is

$$\bar{p}_p = \frac{\int_0^\pi A p_p \cos\theta e^{|p_p| |E| \cos\theta/kT} 2\pi \sin\theta d\theta}{\int_0^\pi A e^{-|p_p| |E| \cos\theta/kT} 2\pi \sin\theta d\theta} \quad (23)$$

The numerator is the orientational polarization and the denominator is the total number of dipoles. Substituting $|p_p| |E|/kT$ by x and $\cos\theta$ by y , eqn. (23) reduces to

$$\bar{p}_p = \bar{p}_p [\coth x - \frac{1}{x}] \quad (24)$$

$$= \bar{p}_p L(x) \quad (25)$$

where $L(x)$ is called the Langevin function, after the man who first developed this expression for magnetic dipole moments in 1905. $L(x)$ is plotted as a function of x in Fig. 3. It is seen that at large values of x (i.e. at low temperatures and high fields), $L(x)$ approaches a value of 1, or saturation. For normal fields and temperatures, $x \ll 1$ and then $L(x)$ can be approximated by the tangent of the curve of Fig. 3 near the origin i.e. by $x/3$. Then,

$$\bar{p}_p = \frac{\bar{p}_p^2 E}{3 kT} \quad (26)$$

$$P_o = N \bar{p}_p^2 E/3 kT = N \alpha E \quad (27)$$

$$\text{and } \alpha_o = \bar{p}_p^2/3 kT \quad (28)$$

It is thus seen that of the three polarization terms discussed so far, only P_o is temperature dependent.

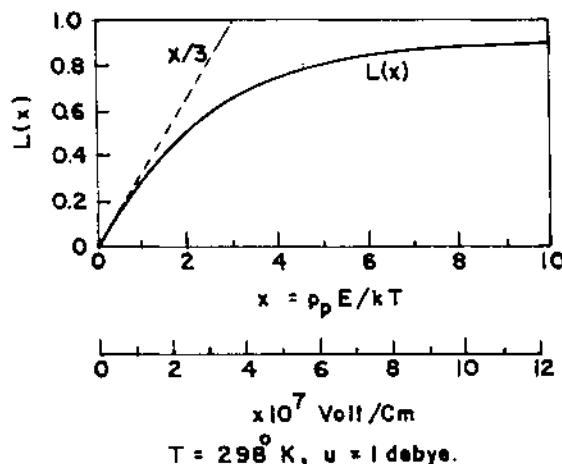


Fig. 3. Langevin function of dipole orientation.

Interfacial or Space Charge Polarization³, P_s . The charge displacements responsible for the three kinds of polarization discussed so far are small (less than atomic diameters or interatomic distances). On the other hand, charge carriers in some dielectrics, particularly multiphase materials, can migrate considerable distance under the action of an electric field. These charges may pile up at the interfaces or at the electrodes and give rise to space charges and an apparent microscopic field (Fig. 4). This manifests itself as an enhanced dielectric constant and the corresponding polarization is called interfacial or space charge polarization, P_s .

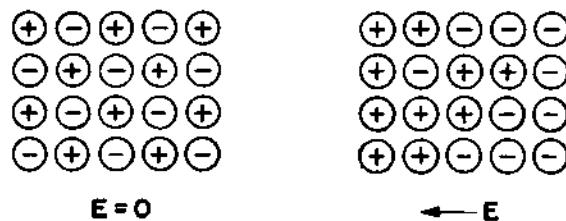


Fig. 4. Space charge polarization

Total Polarization. The total polarization of a multiphase material containing permanent dipoles is given by

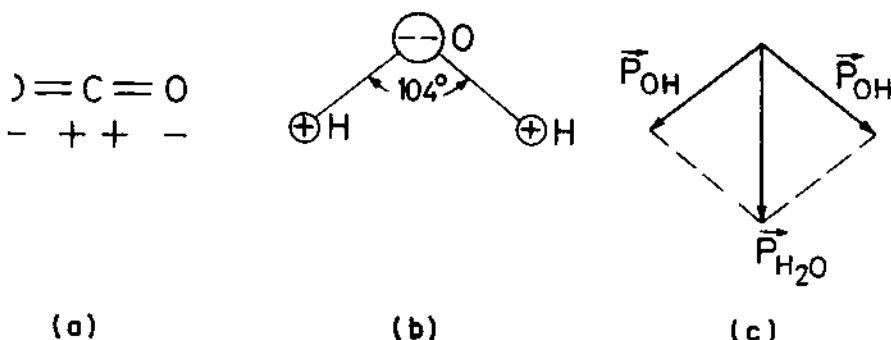


Fig. 5. (a) Symmetric CO_2 and (b) unsymmetric H_2O molecule (c) the effective dipole moment of H_2O molecule.

$$P = P_e + P_i + P_o + P_s \quad (29)$$

and that of a single phase dielectric containing permanent dipoles

$$P = P_e + P_i + P_o \quad (30)$$

$$= P_e + P_i + \left(\frac{N p_p^2 E}{3 kT} \right) \quad (31)$$

Thus the polar dielectrics (possessing permanent dipole moments) can be distinguished from the nonpolar ones from the temperature variation of polarization. Carbon dioxide with its collinear bond is an example of a nonpolar material whereas water molecule with a bond angle of 104° has a permanent dipole moment^{2,6} (Fig. 5). The intercept of the P vs $1/T$ plot gives P_e and P_i and the slope gives $N p_p^2 E / 3k$, from which the dipole moment can be estimated. Fig. 6 shows the data for symmetric CH_4 and CCl_4 and asymmetric CH_3Cl and CH_2Cl_2 molecules⁷.

The orientation of dipoles becomes hindered in most liquids on freezing⁸ (Fig. 7a). In some cases, however, the dipoles may be oriented upto some temperature below the melting point, as in HCl ⁸ (Fig. 7b).

Local Field³. A dipole in a condensed phase, such as a liquid and a solid, experiences not only the external field E but also the electric field due to the polarization medium around it. Therefore the local field, also called effective field or internal field, E' consists of

$$E' = E + E_1 + E_2 \quad (32)$$

where E_1 = the field due to the dipole chains surrounding the

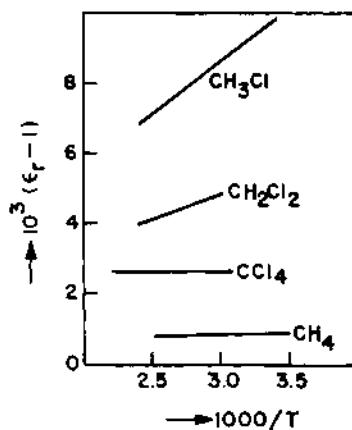


Fig. 6. Variation of relative dielectric constant as a function of temperature for symmetric (CH_4 and CCl_4) and unsymmetric (CH_3Cl and CH_2Cl_2) molecules.

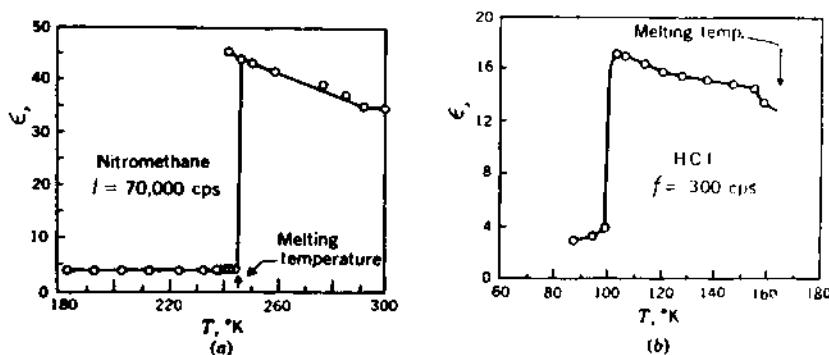


Fig. 7 (a) Dielectric constant of nitromethane (CH_3NO_2) is independent of temperature in the solid state and large and temperature-dependent in the liquid state due to the orientation of permanent dipoles. Electronic, ionic and orientational polarization contribute in the liquid state whereas only the first two contribute in the solid state.
 (b) The increase in dielectric constant of HCl on freezing is due to increase in density. Molecular rotation continues in the solid state until a lower temperature viz. 100°K , when there is a sudden decrease in dielectric constant.

spherical cavity constructed around the reference atom or dipole (Fig. 8) and $E_2 =$ the field due to the polarization of the atoms or molecules within the cavity (Fig. 8). It can be shown³ that

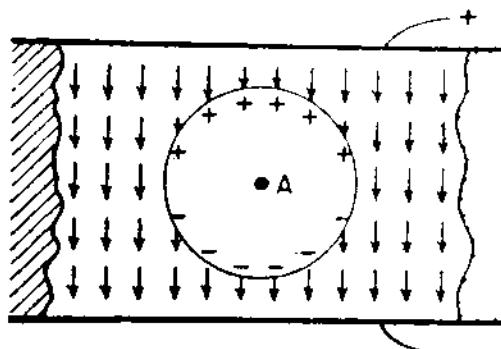


Fig. 8. Model for calculation of internal field.

$$E_1 = \frac{P}{3\epsilon_v} = \frac{\epsilon_v (\epsilon_r - 1)E}{3\epsilon_v} = \frac{(\epsilon_r - 1)E}{3} \quad (33)$$

and it may be assumed that, in many materials,

$$E_2 = 0 \quad (34)$$

so that the local field becomes

$$\begin{aligned} E' &= E + \frac{P}{3\epsilon_v} = E + \frac{\epsilon_r E}{3} - \frac{E}{3} \\ &= \frac{(\epsilon_r + 2)}{3} E \end{aligned} \quad (35)$$

$$P = N\alpha E' = \frac{N\alpha (\epsilon_r + 2)E}{3} = \epsilon_v (\epsilon_r - 1)E \quad (36)$$

so that we have a relation between the polarizability for unit volume, $N\alpha$ and the relative dielectric constant of the material, ϵ_r ,

$$\frac{N\alpha}{\epsilon_v} = \frac{\epsilon_r - 1}{\epsilon_r + 2} \quad (37)$$

For gases at low pressures, $\epsilon_r - 1 \ll 1$, so that $\epsilon_r + 2$ may be approximated by 3. Then,

$$\frac{N\alpha}{\epsilon_v} = \epsilon_r - 1 = \chi_e \quad (38)$$

where χ_e is the electric susceptibility of the gas.

We may rewrite eqn. (37) in terms of polarizability for mole (molar polarization), $\overline{\alpha}$, instead of polarizability for unit volume, so that

$$\overline{M} = \frac{N_0 \alpha}{3\epsilon_v} = \frac{\epsilon_r - 1}{\epsilon_r + 2} = \frac{\epsilon_r - 1}{\epsilon_r + 2} \cdot \frac{M}{\rho} [\text{m}^3] \quad (39)$$

where N_0 is the Avogadro's number, 6.023×10^{26} per Kg mole and ρ is the density in Kg m^{-3} , using the relation

$$N_0 = \frac{NM}{\rho} = 6.023 \times 10^{26} \quad (40)$$

Equation (39) is the Clausius-Mosotti equation and a parallel one exists in the optical case using refractive index, n , instead of relative dielectric constant. The corresponding equation called the Lorenz-Lorentz equation is given by,

$$\overline{M} = \frac{N_0 \alpha}{3\epsilon_v} = \frac{n^2 - 1}{n^2 + 2} \cdot \frac{M}{\rho} \quad (41)$$

where \overline{M} is now called molar refraction.

III. ALTERNATING CURRENT PHENOMENA IN DIELECTRICS

When a time varying electric field of the form

$$E = E_0 e^{i\omega t} \quad (42)$$

is applied to a dielectric material, the response is not entirely instantaneous, so that the proportionality constant, ϵ , in eqn. (1) should be a complex quantity, ϵ^* , instead of a simple, real quantity. Then eqn. (1) becomes

$$D = \epsilon^* E \quad (43)$$

$$\text{where } \epsilon^* = \epsilon' - i\epsilon'' \quad (44)$$

ϵ' being the real part of the dielectric constant and ϵ'' being the imaginary part of the dielectric constant. The imaginary part arises due to the time lag between the response and the stimulus and the time lag may be depicted by a phase angle δ (Fig. 9) such that, for small values of δ ,

$$\tan \delta = \frac{\epsilon''}{\epsilon'} \quad (45)$$

Since E is dependent on frequency and time (eqn. 42), both the real and imaginary parts of the dielectric constant (ϵ' and ϵ'') vary with frequency and time. The time dependence is characterised by a time constant or relaxation time τ , equal to the time in which, on removal of the impressed field, the electrical response (D or P) drops to $1/e$ of the original value. Rewriting eqn. (44) to show the frequency and time dependence gives us¹⁰,

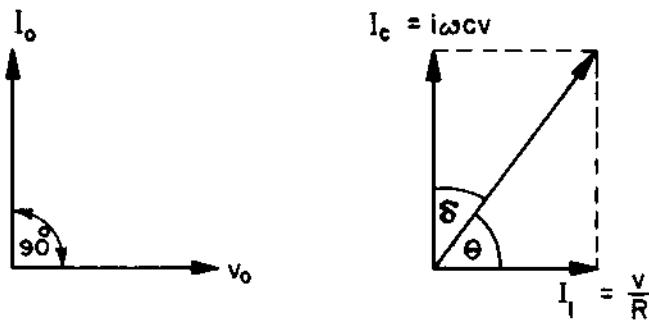


Fig. 9. Current-voltage relation of (a) an ideal capacitor and (b) a capacitor containing a dielectric with loss.

$$\epsilon^* = \left(\epsilon_\infty + \frac{\epsilon_s - \epsilon_\infty}{1 + \omega^2 \tau^2} \right) + i \left(\frac{\epsilon_s - \epsilon_\infty \omega \tau}{1 + \omega^2 \tau^2} \right) \quad (46)$$

where ϵ_s is the static (or low frequency) dielectric constant and ϵ_∞ is the optical (or high frequency) dielectric constant. Then,

$$\epsilon' = \epsilon_\infty + \frac{\epsilon_s - \epsilon_\infty}{1 + \omega^2 \tau^2} \quad (47)$$

$$\text{or } \frac{\epsilon' - \epsilon_\infty}{\epsilon_s - \epsilon_\infty} = \frac{1}{1 + \omega^2 \tau^2} \quad (48)$$

$$\epsilon'' = \frac{(\epsilon_s - \epsilon_\infty) \omega \tau}{1 + \omega^2 \tau^2} \quad (49)$$

$$\text{or } \frac{\epsilon''}{\epsilon_s - \epsilon_\infty} = \frac{\omega \tau}{1 + \omega^2 \tau^2} \quad (50)$$

Combining eqns (45), (47) and (49),

$$\begin{aligned} \tan \delta &= \frac{\epsilon''}{\epsilon'} = \frac{(\epsilon_s - \epsilon_\infty) \omega \tau}{1 + \omega^2 \tau^2} \cdot \frac{1 + \omega^2 \tau^2}{(1 + \omega^2 \tau^2) + (\epsilon_s - \epsilon_\infty)} \\ &= \frac{(\epsilon_s - \epsilon_\infty) \omega \tau}{\epsilon_\infty \omega^2 \tau^2 + \epsilon_s} \end{aligned} \quad (51)$$

Eqns (48) and (50), are called Debye equations. It is instructive

to plot ϵ' , ϵ'' and $\tan \delta$ as a function of $\omega\tau$ (Fig. 10). As $\omega \rightarrow 0$, $\epsilon' \rightarrow \epsilon_s$, and $\epsilon'' \rightarrow 0$ and $\tan \delta \rightarrow 1/\epsilon_s$. At high frequencies, i.e., as $\omega \rightarrow \infty$, $\epsilon' \rightarrow \epsilon_\infty$, $\epsilon'' \rightarrow 0$ and $\tan \delta \rightarrow 0$. At an intermediate frequency when $\omega = 1/\tau$, $\epsilon' = (\epsilon_s + \epsilon_\infty)/2$, $\epsilon'' = (\epsilon_s - \epsilon_\infty)/2$ and $\tan \delta = (\epsilon_s - \epsilon_\infty)/(\epsilon_s + \epsilon_\infty)$. Thus, an inflection in ϵ' and a peak value of $(\epsilon_s - \epsilon_\infty)/2$ in ϵ'' are observed at $\omega\tau = 1$, $\tan \delta$ versus $\omega\tau$ goes through a peak of $(\epsilon_s - \epsilon_\infty)/2\sqrt{\epsilon_s \epsilon_\infty}$ at $\omega\tau = \sqrt{\epsilon_s/\epsilon_\infty}$. Quite often, the polarization process in real

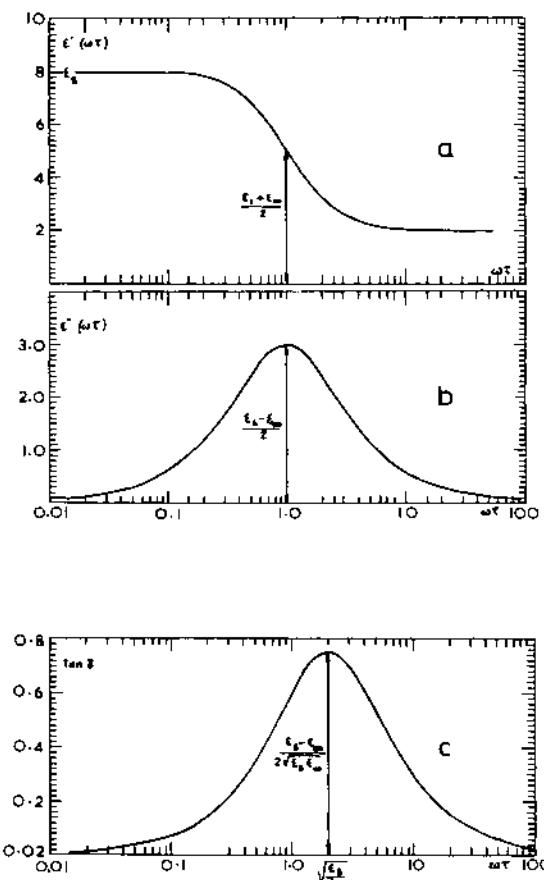


Fig. 10. (a) The real part (ϵ') and (b) the imaginary part (ϵ'') of a dielectric with $\epsilon_s = 8$ and $\epsilon_\infty = 2$, as a function of (c) Variation of loss tangent ($\tan \delta$) of the same dielectric with $\omega\tau$

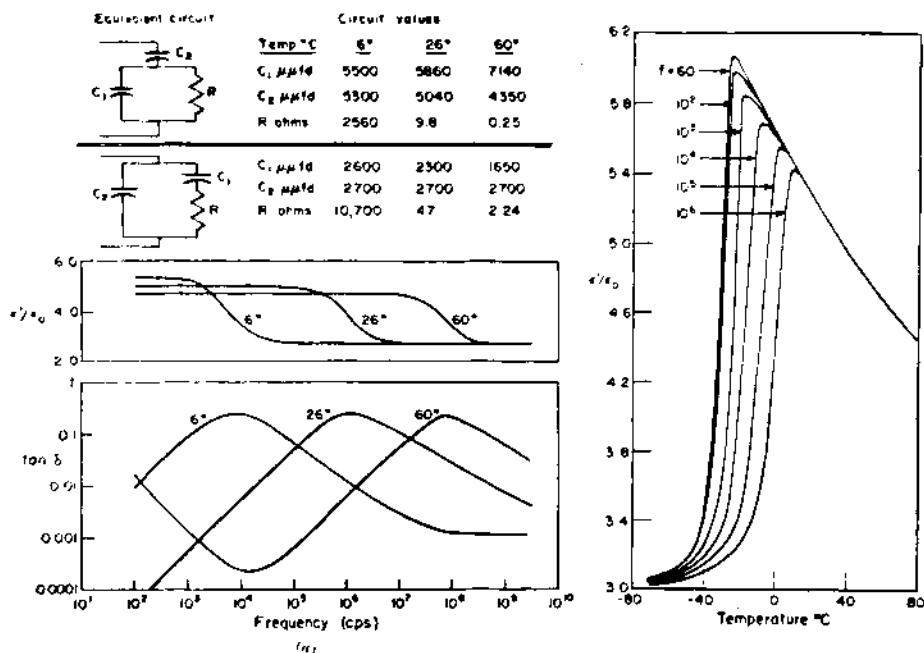


Fig. 11. Variation of dielectric constant (ϵ') and loss ($\tan \delta$) of 'Arcolor' transformer oil with frequency and temperature.

materials is not characterized by a single relaxation time, but a spectrum of relaxation times. In that case, the plots of the Debye equations deviate from the ideal; for example, the ϵ'' versus frequency curve shows a broadening, as illustrated in Fig. 11 for the case of 'Arcolor' transformer oil¹¹ and in Fig. 12 for vulcanized rubber¹².

Another interesting way of treating dielectric data is to plot ϵ'' as a function of ϵ' at different frequencies, as was suggested by Cole and Cole¹³. Such a Cole-Cole plot for an ideal Debye dipolar liquid is a semicircle with its center on the ϵ' axis and a radius of $(\epsilon - \epsilon_\infty)/2$; for vulcanized rubber with a distribution of relaxation times, it is a semi-circle with its center below the ϵ' axis, the radius making an angle of $\alpha\pi/2$ with the ϵ' axis where α is a measure of the distribution of relaxation times¹¹ (Fig. 12). Debye¹⁴ has related the relaxation time to viscosity of dipolar liquids and therefore γ can be

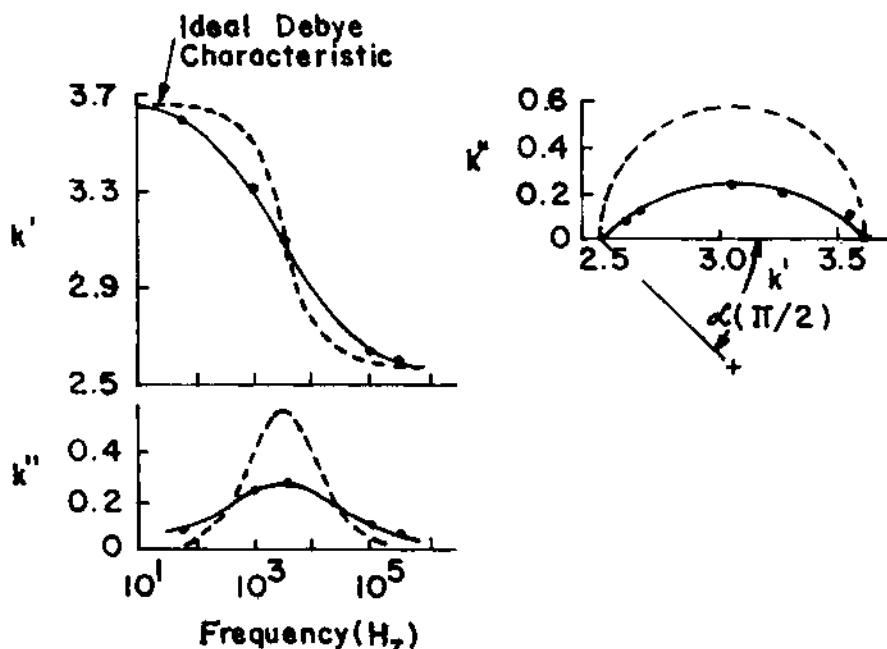


Fig. 12. Frequency response characteristics of vulcanized rubber indicating distribution of relaxation times.

expected to increase with decreasing temperature. In fact, making use of a model consisting of two positions of equal energy for the dipole, separated by a barrier of ΔE , it can be shown¹⁰ that the relaxation time τ follows an Arrhenius type rate equation of the form

$$\tau = \tau_0 e^{-\Delta E/kT} \quad (52)$$

where τ_0 is a constant and the other symbols have their usual significance. The broadening of the relaxation spectrum and the shift of the frequency corresponding to ϵ''_{\max} with temperature of Arcolor transformer oil is shown in Fig. 11a. Further lowering of the temperature freezes the dipole moments and as a result the dielectric constant drops¹¹ (Fig. 11b) as discussed earlier.

Dielectric relaxation is not restricted to liquids only. It can be present in solids also. Strontium titanate, SrTiO_3 , has cubic perovskite type structure. Solid solutions of the type $(1 - x)\text{SrTiO}_3 + x\text{Bi}_{2/3}\text{TiO}_3$ may be formed in which the Ti and oxygen sublattices are completely filled but for each Sr^{2+} ion replaced, only $2/3 \text{ Bi}^{3+}$ ions are introduced to maintain charge neutrality. Consequently for $x \text{ Sr}^{2+}$ ions replaced, $1/3 x$ cation vacancies are

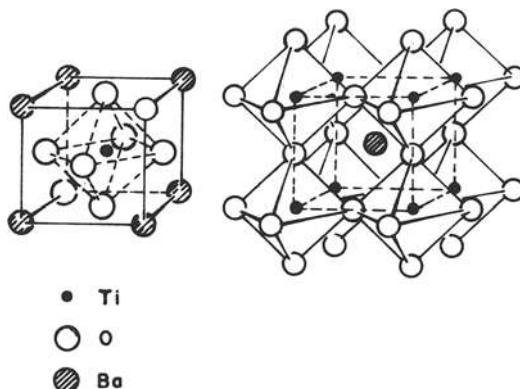


Fig. 13. Cubic perovskite structure (BaTiO_3)

created in Sr^{2+} sublattice. These vacancies carry a net effective negative charge of -2, whereas each Bi^{3+} ion in the Sr site carries a net effective positive charge of +1; there is a coulombic attraction between these two oppositely charged species, thus forming a dipole. An examination of the crystal structure of the perovskite lattice (Fig. 13) reveals six equivalent dipole directions. The large number of lattice defects that can be introduced in these materials and the possibility of orienting the resulting dipoles constitutes a large orientational polarization, increasing the dielectric constant very significantly¹⁵ (Fig. 14). Relaxation dielectrics based on perovskite type structures have been extensively studied recently¹⁶⁻²⁰.

Introduction of aliovalent impurities in alkali halides generate vacant lattices, which form dipoles with the impurity ions. The reorientation of these dipoles by applied fields leads to dielectric absorption²¹⁻²⁵. The relaxation of cation-vacancy pairs in ThO_2 doped with CaO is examined by dielectric and mechanical loss measurements and considered from group theoretical standpoint by Wachtman.

The rutile structure of TiO_2 may be modified by partial replacement of Ti^{4+} ions by Mg^{2+} ion, the charge neutrality maintained by introducing large Ba^{2+} ions into some of the tunnel sites in the structure. Ba^{2+} ions can be moved rather easily in the tunnels leading to large dielectric absorption in the tunnel direction and very low losses in the direction perpendicular to the tunnels²⁶.

In the case of Li - doped transition metal oxides, such as NiO , electrons are believed to hop from one lattice site to the next. The crystallography, electrical conductivity and mechanical relaxa-

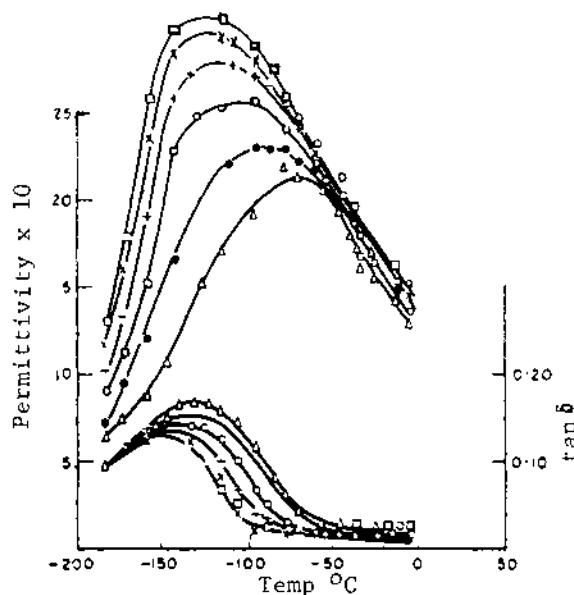


Fig. 14. Dielectric constant and loss tangent of a single crystal of (Sr - Bi) titanate, as a function of frequency and temperature (frequency range used 30 Hz to 500 kHz).

tion effects in these materials have been studied²⁷⁻²⁹.

Amorphous materials such as inorganic glasses exhibit dielectric losses due to several mechanisms. These can be distinguished by their characteristic temperature and frequency dependence³⁰⁻³². A comprehensive survey of dielectric loss mechanisms is presented by Meakins³³ and Volger³⁴.

We had earlier considered four polarization mechanisms. It is reasonable to expect that a different relaxation time is associated with each of these processes. The variation of ϵ' and ϵ'' with frequency is schematically shown in Fig. 15. With increasing frequency, the inertia effects successively eliminate the space charge, orientational and ionic contributions and finally only the electronic contribution remains. Thus, the dielectric constant at high frequencies, ϵ_∞ , is equal to n^2 , where n is the refractive index. A dielectric loss peak is observed at a frequency where one or the other of the polarization mechanisms begin to damp out. Thus, the contribution of the different polarization mechanisms may be separated by a study of dielectric properties as a function of frequency at constant temperature, and as a function of temperature at a fixed frequency, and also of refractive index.

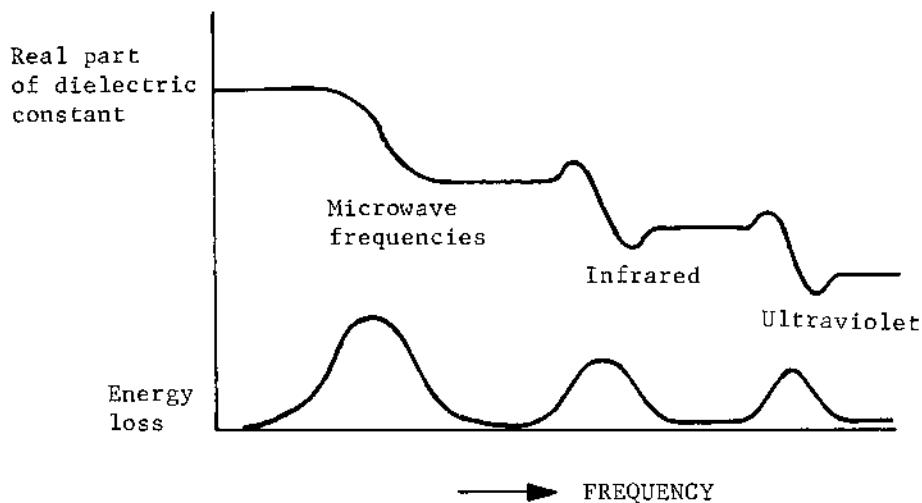


Fig. 15. Variation of dielectric constant and dielectric loss as a function of frequency.

IV. FERROELECTRICS³⁵⁻³⁸

When the center of gravity of the positive charges does not coincide with the center of gravity of the negative charges in a crystal, a dipole results. The dipole moment arises spontaneously in such a non-centrosymmetric crystal, accompanied by a phase transition. If the direction of the permanent dipoles in a material can be reversed by an applied electric field, then such a material is said to be a ferroelectric and the phenomenon is called ferroelectricity. These terms are adopted from the analogous ferromagnets and ferromagnetism, because of the similarity of many of the characteristics. The chief characteristics are discussed below with particular reference to BaTiO_3 , which has been studied most extensively.

The main classes of ferroelectrics are listed in Table IV.1.

Properties of Barium Titanate. Barium titanate, BaTiO_3 , crystallizes with perovskite type structure (Fig. 16). It has cubic symmetry above 120°C . On cooling, it successively transforms to tetragonal at 120°C , orthorhombic at 0°C and rhombohedral at -80°C . BaTiO_3 is ferroelectric at all temperatures upto 120°C , above which it is paralelectric. The directions of polarization (polar axis) is along a cube edge i.e. (001) in the tetragonal phase, along the face diagonal i.e. (110) in the orthorhombic phase and along a body diagonal i.e. (111) in the rhombohedral

Table IV.1

Name	Chemical formula	Curie Temp. °C	Spontaneous polarization 10^{-6} c/cm^2
Rochelle salt	$\text{NaKC}_4\text{H}_4\text{O}_6 \cdot 4\text{H}_2\text{O}$	upper +24 lower -18	0.25
KDP	KH_2PO_4	-150	4.7
Barium Titanate	BaTiO_3	+120	26.0
Triglycine sulfate	$(\text{NH}_2\text{CH}_2\text{COOH})_3 \cdot \text{H}_2\text{SO}_4$	+49	2.8
Sodium Nitrite	NaNO_2	+160	7.0

phase. BaTiO_3 crystals can be grown either by a floating zone method or by a flux method (with KF as flux)³⁹. Barium titanate ceramics can be produced by the usual ceramic (or powder metallurgy) techniques and solid state sintering of a mixture of BaCO_3 and TiO_2 . Temperature variation of the dielectric properties of single crystal⁴⁰ and ceramic barium titanate³ is shown in Fig. 17. The dielectric constant of the non-cubic crystal is different in different directions as shown, whereas in the polycrystalline

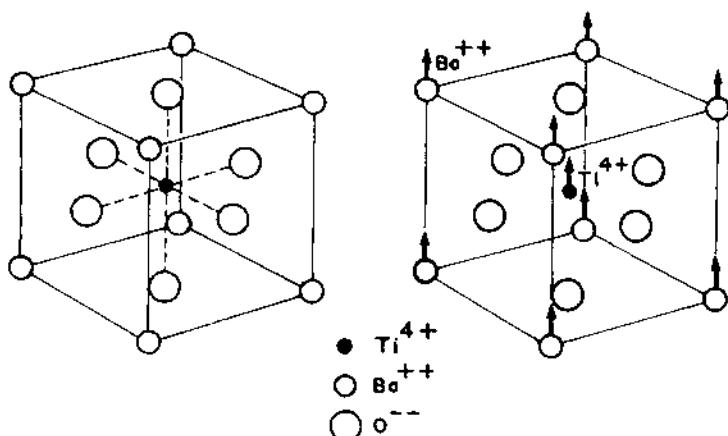


Fig. 16. Structure of cubic perovskite and ionic displacement in BaTiO_3 .

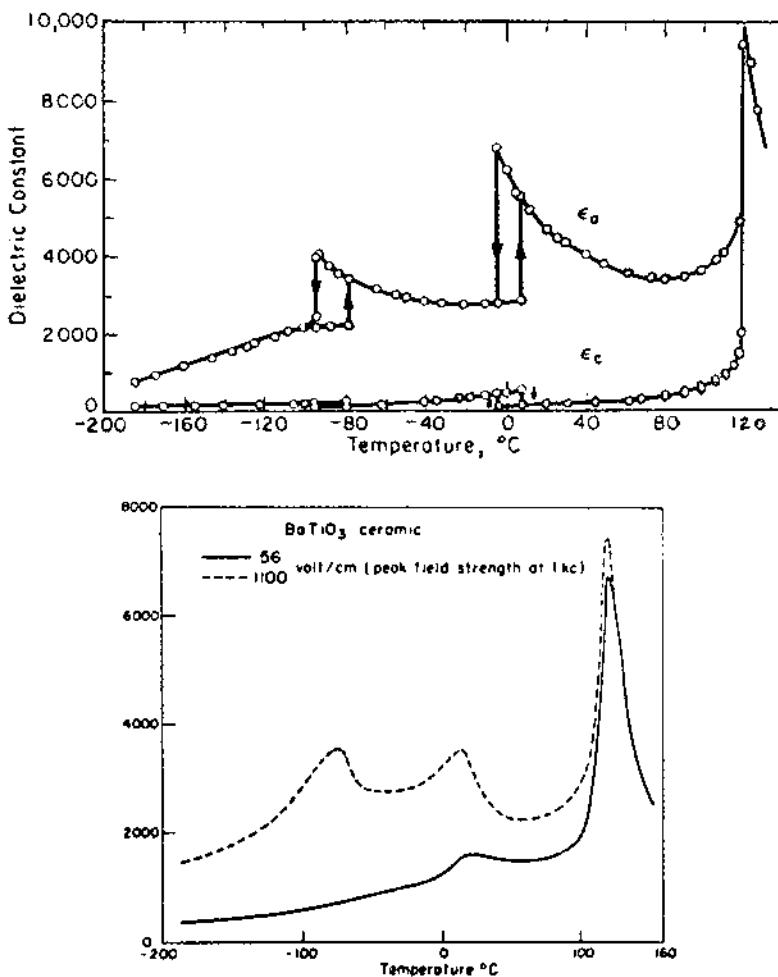


Fig. 17. Temperature dependence of dielectric constant of (a) single crystal and (b) ceramic BaTiO_3 .

material the average property is measured. Sudden changes in dielectric constant are observed in the vicinity of the phase changes, though the changes are not as sharp in the ceramic as in the single crystal. The high dielectric constant of ceramic barium titanate makes it an attractive material for miniature condensers. However, from a practical stand point, the marked temperature dependence of the dielectric properties are undesirable. If a capacitor material is sought with a small temperature coefficient in the neighbourhood of room temperature, the transition at 0°C may be shifted down and then that at 120°C moved to higher temperature.

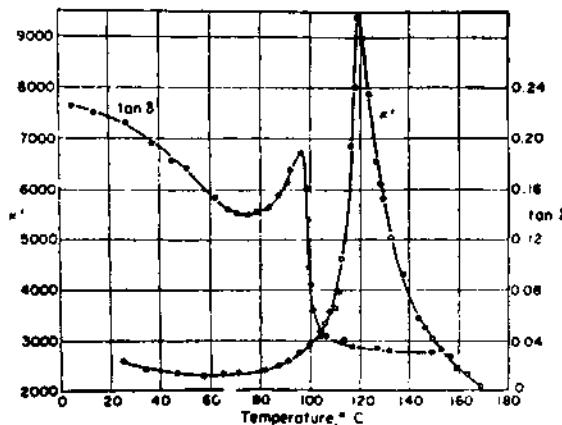


Fig. 18. Dielectric constant and loss tangent of ceramic barium titanate as a function of temperature.

by appropriate solid solution formation (e.g. by replacing Ba^{2+} by Pb^{2+})^{41,42}, or by controlling the microstructure to obtain a small grain size. The variation of $\tan \delta$ and ϵ' of ceramic barium titanate is also shown in Fig. 18.

Application of an alternating electric field gives rise to a hysteresis loop. The relationship between D or P and E is non-linear in the ferroelectric region, whereas it is linear in the paraelectric region. The (B-E) hysteresis loop of a ferroelectric is similar to that of the B-H loop of the ferromagnetic material. The important parameters of the hysteresis loop, shown in Fig. 19, are the spontaneous polarization (P_s , extension of the saturated

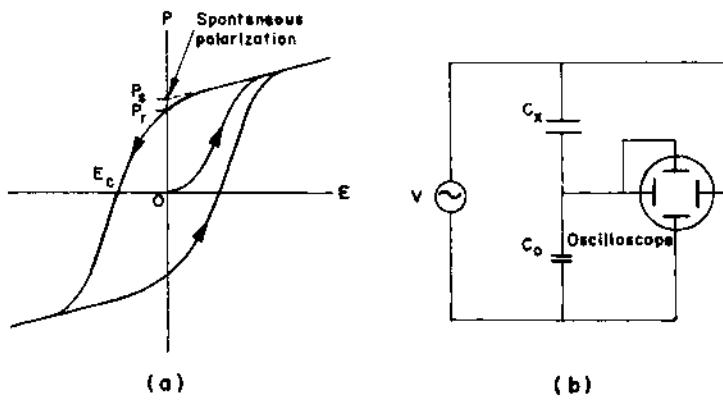


Fig. 19. (a) Hysteresis loop of a ferroelectric.
(b) A circuit for obtaining hysteresis loop.

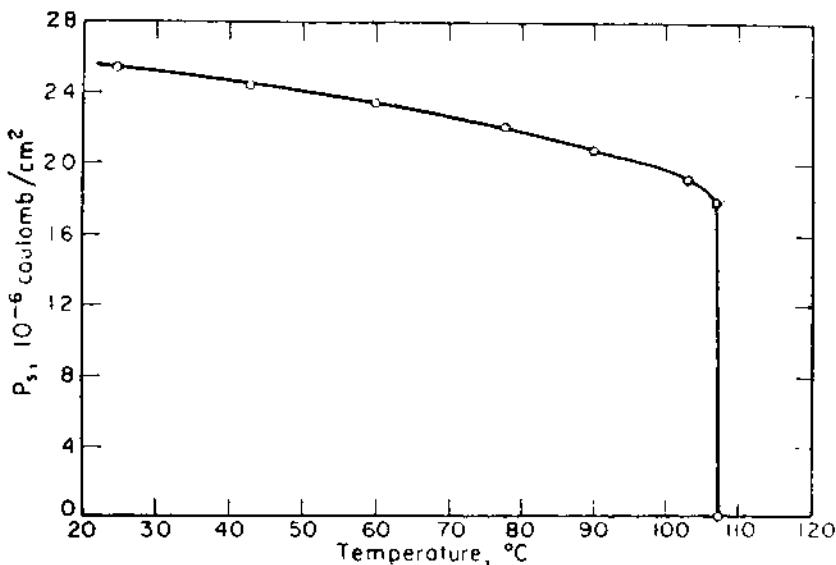


Fig. 20. Temperature variation of spontaneous polarization of BaTiO_3 .

portion of the loop to zero field), remanent polarization (P_r , the polarization remaining when the field is reduced to zero), and coercive field (E_c , the field necessary to reduce the polarization to zero). The shape of the loop is more nearly a rectangle in the case of a single crystal than the rounded one for a polycrystalline material. A square loop material may be used for computer memory cores and other switching circuits. The variation of spontaneous polarization with temperature⁴⁶ is shown in Fig. 20. It is seen that P_s disappears at the ferroelectric-paraelectric transition temperature.

The electrostatic energy of a crystal is lowered if the crystal has the directions of spontaneous polarization randomly distributed. This is accomplished by the formation of domains, or regions with uniform polarization. Ferroelectric domains are separated from each other by domain walls (Fig. 21a and b). Note the analogy with grains and grain boundaries. Domain walls can be delineated by etching with appropriate etchants⁴⁴, or by depositing charged, colloidal particles from a suspension⁴⁵. Ferroelectric crystals are transparent and hence domain walls and their movement under the influence of applied electric and mechanical stress can be observed under a polarising microscope (Fig. 21). The shape of the hysteresis loop can be understood in terms of domain processes. When a small field is applied, domains with the polar direction more favourably oriented with respect to the applied field are nucleated and these grow at the expense of less favourably oriented

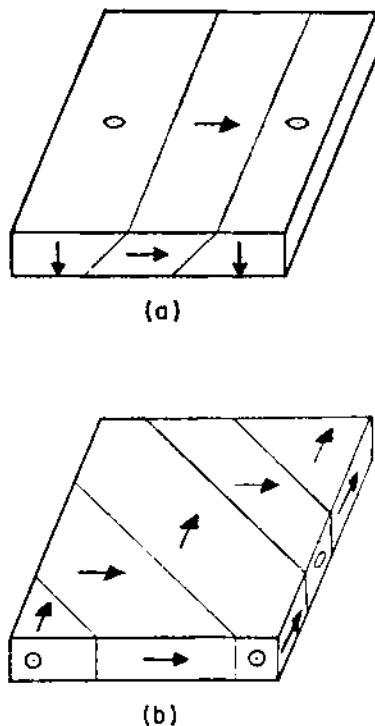


Fig. 21. Domain structure of BaTiO_3 (a) and (b) schematic; the others are photomicrographs.

ones, with increased magnitude of the field. By the time the steep portion of the loop is traversed all the domains would have one of their easy directions of polarization close to that of applied field. Further increase in the field rotates the direction of polarization of the crystal such that the polar axes coincide with the field direction. In this state the maximum polarization is reached. Decrease of the electric field causes only partial reorganization of domain structure, so that when $E \rightarrow 0$, a certain amount of polarization, called remanent polarization, is retained. Application of an electric field of an opposite sign nucleates domains whose polar axes are at 180° to the previous ones and when

$E = E_c$, the directions of polarization are randomized so that $P = 0$. Further increase of the field leads to the growth of the appropriate domains until saturation is attained in this direction. Similar arguments will complete the loop. The switching times associated with domain reversal in ferroelectrics are of the order of 10^{-6} sec. It is this fact coupled with the ease with which electric fields can be applied that makes ferroelectrics interesting in switching circuits and memory applications.

Mosotti Catastrophe. From eqns. (35), (36) and (37) we obtain,

$$P = \frac{N\alpha E}{1 - N\alpha/3\epsilon_v} \quad (53)$$

$$\text{Since, } \chi_e = \epsilon_r - 1 = \frac{P}{\epsilon_v E} = \frac{N\alpha/\epsilon_v}{1 - N\alpha/3\epsilon_v} \quad (54)$$

When $N\alpha/3\epsilon_v = 1$, polarization in eqn. (53), susceptibility and relative dielectric constant in eqn. (54) go to infinity. If the electronic and ionic polarizabilities are negligible compared to the orientational contribution, then from eqn. (28),

$$\frac{N\alpha}{3\epsilon_v} \cdot \frac{N\alpha_0}{3\epsilon_v} = \frac{N\alpha}{3\epsilon_v} \times \frac{p_p^2}{3kT} \quad (55)$$

Eqn. (55) holds at high temperature. As the temperature is decreased, a critical temperature, called the Curie temperature, T_c , is reached when the randomizing effect of the temperature is counterbalanced by the orienting effect of the internal field. Then,

$$\frac{N\alpha}{3\epsilon_v} = 1 = \frac{N\alpha}{3\epsilon_v} \times \frac{p_p^2}{3kT_c} \quad (56)$$

$$\text{or } T_c = N p_p^2 / 9\epsilon_v k \quad (57)$$

so that eqn. (54) becomes

$$\chi_e = \epsilon_r - 1 = \frac{3 T_c}{T - T_c} \quad (58)$$

Eqn. (58) is called the Curie-Weiss law and is encountered in ferromagnetism also. Above T_c , the thermal agitation keeps dipole directions randomized and the material is paraelectric. Below T_c , the internal field increases the polarization and the polarization, in turn, increases the local field and the material overcomes the effect of temperature and becomes spontaneously polarized i.e. it

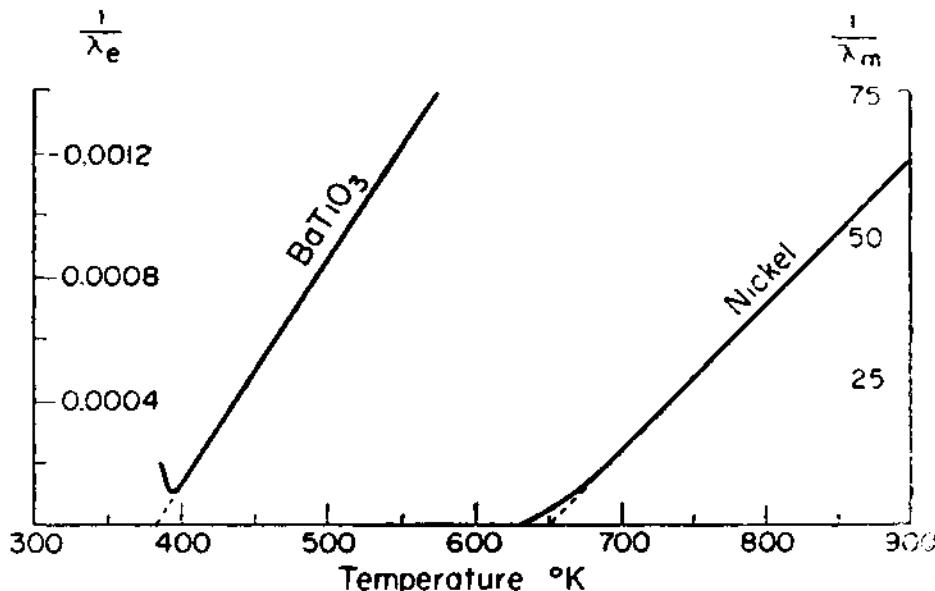


Fig. 22. Variation of electric susceptibility of BaTiO_3 with temperature.

becomes a ferroelectric.

The Curie-Weiss law, eqn. (58), is sometimes written in a slightly different form,

$$\epsilon = \epsilon_{\infty} + \frac{C}{T - T_c} \quad (59)$$

where C is called the Curie constant $= \frac{N p_p^2}{3 \epsilon_v k}$

and the other symbols have their usual significance. A plot of eqn. (59) for BaTiO_3 in Fig. 22 gives $C = 1.5 \times 10^5 \text{ } ^\circ\text{K}$ and $T_c = 120^\circ\text{C}$.

The most important explanation for the origin of ferroelectricity in BaTiO_3 is the lattice dynamical theory, detailed by Cochran⁴⁶. A transverse optical mode has been identified whose frequency changes with temperature as $\omega_T^2 \propto (T - T_c)$, so that when $T \rightarrow T_c$, $\omega_T \rightarrow 0$. Since $\epsilon \propto 1/\omega^2$, as $T \rightarrow T_c$, $\epsilon \rightarrow \infty$ giving rise to paraelectric - ferroelectric transition. The existence of a soft mode has been found in SrTiO_3 ,⁴⁷ but the ferroelectricity of this compound has not been clearly established. However, such a soft mode has not been detected in established ferroelectrics such as BaTiO_3 , using neutron inelastic scattering, Raman

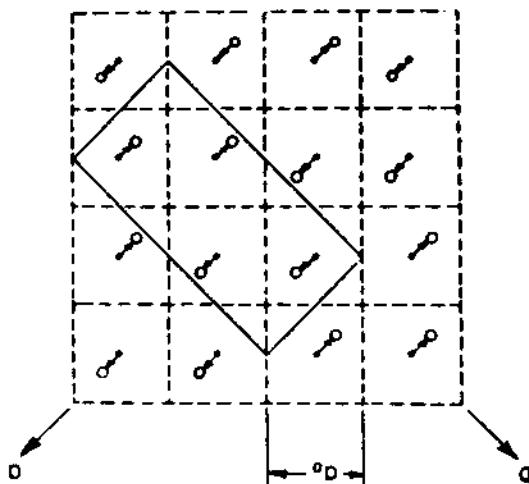


Fig. 23. Antiferroelectric structure of PbZrO_3 .

spectroscopy, etc.

When the internal field gives rise to a parallel alignment of the dipoles in adjacent unit cells, ferroelectricity results. In some cases, the internal field gives rise to an antiparallel alignment of dipoles in adjacent unit cells, and then antiferroelectricity is the result (Fig. 23).

REFERENCES

1. A.J. Dekker, Electrical Engineering Materials, Prentice-Hall, Inc. Englewood Cliffs, N.J., 1959, Chap. 2 and 3.
2. C. Wert and R.M. Thomson, Physics of Solids, McGraw-Hill Book Co., New York, 1964, Chap. 16 and 17.
3. A.R. Von Hippel, Dielectrics and Waves, John Wiley & Co., New York, 1954, Chap. I, 8, II.1, 2, 3, 16, 22, 23, 25, 26, 27, 28.
4. L.V. Azaroff and J.J. Brophy, Electronic Processes in Materials, McGraw-Hill Book Co., New York, 1963, Chap. 12.
5. J.H. Van Vleck, The Theory of Electric and Magnetic Susceptibilities, Oxford University Press, 1932.
6. N. Nussbaum, Electronic and Magnetic Behaviour of Materials, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1967, Chapter 3.

7. R. Sanger, Phys. Z., 27, 556 (1926).
8. C. Smyth and W. Wallis, J. Chem. Phys., 3, 557 (1935).
9. C. Smyth and C. Hitchcock, J. Am. Chem. Soc., 55, 1830 (1933).
10. V.V. Daniel, Dielectric Relaxation, Academic Press, New York, 1967.
11. A.R. Von Hippel, Molecular Science and Molecular Engineering, M.I.T. Press, Cambridge, 1959, Chap. 15.
12. W. Kauzmann, Revs. Mod. Phys., 14, 12 (1942).
13. K.S. Cole and R.H. Cole, J. Chem. Phys., 9, 341 (1941).
14. P. Debye, Polar Molecules, Chemical Catalog Co., New York, 1929; Dover Publications, New York, 1945.
15. L.E. Cross, Proc. Instu. Elect. Engrs. 109 B, Suppl., 22 (1962).
16. G.I. Skanair and N. Matveeva, Zh. Eksp. Teor. Fiz., 30, 1047 (1956).
17. G.A. Smolenskii, V.A. Isupov, A.I. Agranovskaya, and S.N. Popov, Soviet Phys. Solid State, 11, 2906 (1960).
18. A.N. Gubkin, A.M. Kashtanova, and G.I. Skanair, Fiz. Tverd. Tela, 3, 1110 (1961).
19. V.A. Isupov, Izv. Akad. Nauk. Tadzhik SSR, 20, 560 (1964).
20. S.V. Bogdanov, K.V. Kiseleva and V.A. Rassushin, Soviet Phys. Crystallogr., 10, 58 (1965).
21. J.S. Dryden and D.A.A.S. Narayana Rao, J. Chem. Phys., 25, 222 (1956).
22. A.B. Lidiard, Handbuch der Physik, Springer-Verlag, Berlin, (1957), Vol. 20.
23. J.S. Dryden and R.J. Meakins, Disc. Faraday Soc., 23, 39 (1957).
24. Y. Haven, Disc. Faraday Soc., 23, 76, 77 (1957).
25. R.A. Sack, Disc. Faraday Soc., 23, 78 (1957).
26. J.S. Dryden and A.D. Wadsley, Trans. Faraday Soc., 54, 1574 (1958).

27. R.R. Heikes and W.D. Johnston, *J. Chem. Phys.*, 26, 582 (1957).
28. W.D. Johnston, R.R. Heikes and D. Sastrich, *J. Phys. Chem. Solids*, 1, 1 (1958).
29. S. Von Houton, *J. Phys. Chem. Solids*, 23, 1045 (1962).
30. J.M. Stevens, *Handbuch der Physik*, Springer-Verlag, Berlin (1957), Vol. 20.
31. P.M. Sutton, *Prog. Dielectrics*, 2, 115 (1960).
32. D. Chakravorty, *Inorganic Glasses*, This Vol.
33. R.J. Meakins, *Prog. Dielectrics*, 3, 151 (1961).
34. J. Volger, *Prog. Semiconductors*, 4, 205 (1960).
35. F. Jona and G. Shirane, *Ferroelectric Crystals*, Pergamon Press, New York, 1962.
36. W. Kanzig, in *Solid State Physics*, eds. F. Seitz and D. Turnbull, 4, 1 (1967).
37. P.W. Forsbergh, in *Handbuch der Physik*, ed. G. Flugge, Springer-Verlag, Berlin, 17, 264 (1956).
38. E. Fatuzzo and W.J. Merz, *Ferroelectricity*, John Wiley, New York, 1967.
39. J.P. Remeika, *J. Am. Chem. Soc.*, 76, 940 (1954).
40. W.J. Merz, *Phys. Rev.*, 76, 1221 (1949).
41. G. Shirane and K. Suzuki, *J. Phys. Soc. Japan*, 6, 274 (1951).
42. S. Nomura and S. Sawada, *J. Phys. Soc. Japan*, 6, 36 (1951).
43. W.J. Merz, *Phys. Rev.*, 91, 513 (1953).
44. J.A. Hooton and W.J. Merz, *Phys. Rev.*, 98, 409 (1955).
45. G.L. Pearson and W.L. Feddman, *J. Phys. Chem. Solids*, 9, 28 (1959).
46. W. Cochran, *Adv. in Physics*, 9, 387 (1960); 10, 401 (1961).
47. R.A. Cowley, *Phys. Rev.*, 134, A 981 (1964).

MAGNETISM

N. Menyuk

Lincoln Laboratory*, Massachusetts Institute of Technology
Lexington, Massachusetts 02173, U.S.A.

All materials are magnetic in the sense that they respond in a measurable way to the application of a magnetic field. For isotropic diamagnetism and paramagnetism the applied field gives rise to a net magnetic moment which is directly proportional to the field, as expressed in the relationship $M = \chi H$, where M is the magnetic moment per unit volume, or the magnetization, H is the applied magnetic field and χ is the volume susceptibility. Although χ is treated as a scalar, materials exist in which H and M are not necessarily parallel, in which case χ is a tensor. That case will not be treated here.

I. NON-INTERACTING ATOMS

Diamagnetism. Materials in which χ is negative are called diamagnetic. The effect arises from the application of Lenz's Law to the motion of the atomic electrons. To first order in H the application of a magnetic field superposes an angular precession of frequency $\omega = -eH/2mc$ on the electron motion in the absence of H . This electron precession is equivalent to a current $i = -ze^2H/4\pi mc^2$, expressed in e.m.u., with z the number of electrons per atom. Since the magnetic moment μ of a current loop is given by iA , where loop area $A = \pi(\bar{x}^2 + \bar{y}^2)$ for the electrons perpendicular to the field,

* Supported in part by the Department of the Air Force.

$$\mu = - (ze^2H/4mc^2)(\bar{x}^2 + \bar{y}^2) . \quad (1)$$

Taking \bar{r}^2 as the mean square distance of the electrons from the nucleus for a spherically symmetric charge distribution, $\bar{r}^2 = \bar{x}^2 + \bar{y}^2 + \bar{z}^2 = \frac{3}{2}(\bar{x}^2 + \bar{y}^2)$. Therefore, the diamagnetic susceptibility for a system with N atoms per unit volume is given by

$$\chi = \frac{\mu N}{H} = - \frac{ze^2 N}{6mc^2} \bar{r}^2 . \quad (2)$$

This contribution to the susceptibility is small, and is present in all substances. Those materials in which this contribution to the susceptibility is dominant are called diamagnetic; they will not be discussed further. However, it should be noted that this effect is present in paramagnetic materials as well, and is sometimes a significant factor in the total observed susceptibility.

Classical Model of Paramagnetism. If an atom with a magnetic moment μ is placed in a magnetic field, there is a tendency for the moment to align itself parallel with the field to reduce the interaction energy

$$E = - \mu \cdot H = - \mu H \cos \theta . \quad (3)$$

Thermal effects will tend to prevent the alignment, and the net magnetization is given by

$$M = N\mu \overline{\cos \theta} \quad (4)$$

where $\overline{\cos \theta}$ is the average value at thermal equilibrium. Assuming a Maxwell-Boltzmann distribution law, the probability of finding an atom with its magnetic moment making an angle between θ and $\theta + d\theta$ (or solid angle $d\Omega$) is just

$$\overline{\cos \theta} = \int e^{-E/kT} \cos \theta d\Omega / \int e^{-E/kT} d\Omega .$$

Carrying the integration out over all directions and using Eqn. (3) yields

$$\overline{\cos \theta} = \int_0^\pi 2\pi e^{-\mu H \cos \theta / kT} \sin \theta \cos \theta d\theta / \int_0^\pi 2\pi e^{-\mu H \cos \theta / kT} \sin \theta d\theta$$

which has the solution (i.e., Ref. 1, p. 170)

$$\overline{\cos \theta} = \operatorname{ctnh} x - \frac{1}{x} \equiv L(x) \quad (5)$$

where Eqn. (5) defines the Langevin function $L(x)$ and $x = \frac{\mu H}{kT}$ is called the argument of the function. Therefore,

$$M = N \mu \left(\operatorname{ctnh} \frac{\mu H}{kT} - \frac{kT}{\mu H} \right) . \quad (6)$$

In general, $\operatorname{ctnh} x = \frac{1}{x} + \frac{x}{3} - \frac{x^3}{45} + \frac{7x^5}{2160} + \dots$ so, for $\mu H \ll kT$, all but the first two terms in the series expansion can be ignored, and

$$M \approx \frac{N \mu^2 H}{3 kT} \quad \text{or} \quad x = \frac{N \mu^2}{3 kT} . \quad (7)$$

In the above manner Langevin² derived the Curie Law $\chi = C/T$ on the basis of purely classical concepts, which is correct for the assumption of a unique magnetic moment for all similar atoms. However, if the atomic moment is taken as proportional to the angular momentum, this assumes an identical angular momentum for every similar atom. This is not a permissible assumption in classical Boltzmann statistics, and the full continuous range of electronic angular momenta from $-\infty$ to $+\infty$ should be considered. Van Leeuwen³ showed that when this is done, the resultant susceptibility is zero.

That, of course, is in contradiction with experimental results, and points out the inability of classical physics to describe magnetic phenomena. The basic difficulty lies in the assumption of a continuum of angular momenta of all possible values. This difficulty was overcome with the advent of quantum mechanics. We will therefore consider quantum mechanical effects in magnetism before continuing our discussion of paramagnetism.

Vector Model of the Atom. In dealing with the atom, quantum mechanics has shown that the 'state' of an electron orbiting the nucleus can be described by a set of discrete quantum numbers. In addition to the principal quantum number n , those of interest are related to the orbital angular momentum, spin angular momentum, total angular momentum and the components of the angular momenta along the applied field. According to the vector model of the atom the magnitude of orbital angular momentum of an electron is $\ell(h/2\pi)$ and the spin angular momentum is $s(h/2\pi)$, where h is Planck's constant ($=6.624 \times 10^{-27}$ erg-sec), ℓ is limited to the values $0, 1, 2, \dots, (n-1)$ and $s = \pm 1/2$. For an atom of m electrons we shall use the Russell-Saunders coupling scheme, taking the total orbital angular momentum operator L as due to the sum of the orbital angular momenta of the individual electrons; i.e., $\sum_{i=1}^m \ell_i$ and similarly for the total spin angular momentum $S = \sum_{i=1}^m s_i$, subject to the condition that $S = 0$ or a positive integer, or a positive half-integer, depending upon whether the system contains an even or odd number of electrons. The total angular momentum is then given by $J = \frac{h}{2\pi}$ where where $J = L + S$ and the quantum number J is limited to the values

0 or a positive integer, or a positive half integer depending on whether S is an integer or a half integer. Finally the magnetic quantum number m_J , m_L and m_S , with angular momenta given by $m_J(h/2\pi)$, $m_L(h/2\pi)$ and $m_S(h/2\pi)$, respectively represent the components of the total, orbital and spin angular momenta along the applied field. The magnetic quantum number m_J is limited to the values $J, J-1, J-2, \dots -J$, with m_L and m_S similarly restricted. Of the six quantum numbers, only four are independent and define the state of the atom, i.e., n, L, m_L, S or n, L, J, m_J .

The above discussion was based on the Bohr-Sommerfeld vector model of the atom. However, a quantum mechanical derivation shows that the magnitudes of the spin, orbital and total angular momentum as given above should be amended, and have the values $\sqrt{S(S+1)}(h/2\pi)$, $\sqrt{L(L+1)}(h/2\pi)$ and $\sqrt{J(J+1)}(h/2\pi)$ respectively.

Classically, the magnetic moment due to a revolving body of mass m , electrical charge $-e$ and angular momentum p is given by $\mu = -(e/2mc)p$. Applying this relationship to the orbital angular momentum yields $\mu_L = -(eh/4\pi mc)L$, bearing in mind that $|L| = \sqrt{L(L+1)}$. The quantity $(eh/4\pi mc)$ is known as the Bohr magneton, and is symbolized by μ_B , where $\mu_B = 9.27 \times 10^{-27}$ ergs/Oe. Thus $\mu_L = -L \mu_B$.

However, this relationship does not hold for μ_S . Dirac⁴ has shown on the basis of a relativistic quantum mechanical analysis that $\mu_S = -2S\mu_B$. Taking the total magnetic moment $\mu = \mu_L + \mu_S$ the resultant magnetic moment is no longer parallel to J , but precesses around J . Because of the precession we need only consider the component μ_J parallel to J . From Fig. 1, we see that $\mu_J = \mu_L \cos(L, J) + \mu_S \cos(S, J)$. Expressing μ_J in the form $\mu_J = -g\mu_B J$ and applying the law of cosines gives

$$g = 1 + \frac{J(J+1) + S(S+1) - L(L+1)}{2J(J+1)} \quad (8)$$

where g is called the Lande- g factor. For $L = 0 (J=S)$, $g = 2$ and for $S = 0 (J=L)$, $g = 1$. For both orbital and spin contributions to the total magnetic moment, g lies between these limits. The component of the magnetic moment parallel to an applied field is then

$$\mu_{m_J} = -g\mu_B m_J \quad . \quad (9)$$

Quantum-mechanical Theory of Paramagnetism. In Boltzmann statistics the probability of finding a quantity A in the i^{th} of n possible states (with energy E_i) at thermal equilibrium is given by

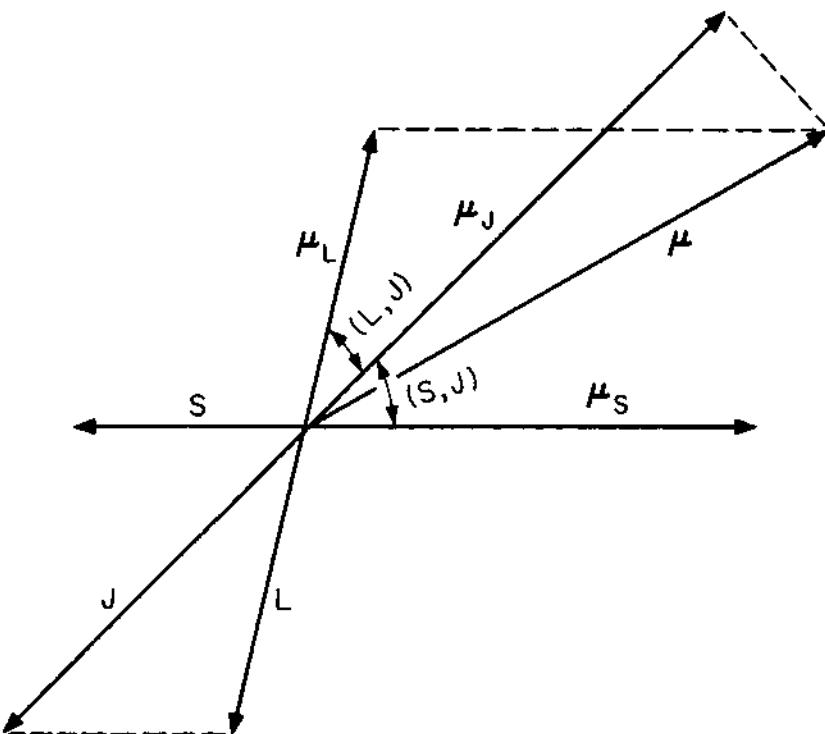


Fig. 1. Vector relationships amongst angular momenta and associated magnetic moments.

$$P(A_i) = A_i e^{-E_i/kT} / \sum_i e^{-E_i/kT} .$$

The thermal equilibrium value of A is therefore obtained by summing over all possible states to obtain

$$\langle A \rangle = \frac{\sum_i A_i e^{-E_i/kT}}{\sum_i e^{-E_i/kT}} . \quad (10)$$

The Hamiltonian of a single magnetic atom with angular momentum J and magnetic moment μ_J in the presence of an applied field is just

$$\mathcal{H} = -\mu_J \cdot H = -\mu_{m_J} \cdot H \quad (11)$$

with eigenvalues, as given by (9)

$$E_{m_J} = -g\mu_B m_J H \quad (12)$$

where m_J is limited to the discrete values $J, J-1, \dots -J$.

For a system of N identical non-interacting particles per unit volume, the magnetization is just

$$M = Ng\mu_B \langle m_J \rangle = Ng\mu_B \sum_{m_J=-J}^J m_J e^{g\mu_B m_J H/kT} / \sum_{m_J=-J}^J e^{g\mu_B m_J H/kT} . \quad (13)$$

The expression on the right hand side of Eqn. (13) has been evaluated and reduces to the form

$$M = Ng\mu_B J B_J(x) \quad (14)$$

where B_J is called the Brillouin function and is equal to

$$B_J(x) = \frac{2J+1}{2J} \coth \left(\frac{2J+1}{2J}x \right) - \frac{1}{2J} \coth \frac{x}{2J} \quad (15)$$

where $x = g\mu_B J H/kT$.

In the classical limit, when we let $J \rightarrow \infty$ and $\mu_B \rightarrow 0$ such that $g\mu_B \rightarrow \mu$ where μ is a fixed magnetic moment, we find that $B_\infty(x) = L(x)$ in accord with Eqn. (5).

For $x \ll 1$, which will be the case except at very high fields or low temperatures, we can expand the coth function as previously, and obtain

$$B_J(x) \approx \frac{J+1}{3J}x = \frac{J+1}{3J} (g\mu_B J H/kT). \quad (16)$$

Substituting into Eqn. (14) yields

$$M = \frac{Ng^2 \mu_B^2 J(J+1)}{3kT} H \quad (17)$$

or

$$x = \frac{M}{H} = \frac{C}{T} \text{ where } C = \frac{Ng^2 \mu_B^2 J(J+1)}{3k} \quad (18)$$

which is again Curie's law, but with the Curie constant now expressed as a function of the total angular momentum quantum number.

Defining the effective moment $\mu_{eff} = g[J(J+1)]^{1/2}$, from Eqn. (18) $\mu_{eff} = [(3kC)/(g^2 N)]^{1/2}$. When dealing with molar susceptibility, N is Avogadro's number and $(3k/N\mu_B^2)$ is fortuitously almost exactly equal to 8. Then the inverse of the slope obtained from a plot of

the inverse molar susceptibility versus temperature curve, χ_M , can be compared directly with the theoretical value $g^2J(J+1)/8$.

The derivation of Eqn. (18) assumed all the atoms or ions to have a unique value of J . In the 3d ionic materials which we will be considering, the energy difference between states of different J values are generally much greater than kT . Then only the lowest-energy J -state will be significantly populated, and the effect of admixing higher-energy J -states will give rise to a small temperature-independent "Van-Vleck" paramagnetism in addition to the much larger C/T term. This, as well as the cases in which the energy difference between J -states is of the order of, or much less than kT are discussed in detail by Van Vleck⁵. Collective electron considerations lead to a temperature independent "Pauli" paramagnetism which will be discussed further in a later section (Eqn. 66).

II. MAGNETIC INTERACTIONS

Ordered Structures. The existence of spontaneously magnetized ferromagnetic materials, such as iron and nickel, has been known for a long time. Knowledge of other forms of ordered magnetic structures is much more recent. Antiferromagnetism, wherein the atomic moments are arranged antiparallel, with a zero resultant net moment, was first introduced in 1932 by Néel.⁶ Néel introduced the concept of ferrimagnetism in 1948,⁷ which also involves an antiparallel arrangement of atomic moments, but with incomplete compensation so that a net spontaneous moment exists. In 1959 Yoshimori⁸, Kaplan⁹ and Villain¹⁰ independently noted that the magnetic energy of a system could sometimes be reduced by the introduction of a spiral or helical arrangement of the atomic moments.

Order of Magnitude of Molecular Field. To explain the spontaneous alignment of magnetic moments in ferromagnetic materials below a critical temperature T_c , called the Curie point, Weiss¹¹ postulated the existence of a strong internal 'molecular' field tending to align the moments in parallel. The order of magnitude of this molecular field H_W can be estimated by noting that this field is opposed by the thermal motion of the moments, being equal at the Curie point. Therefore $\mu H_W \propto kT_c$, so $H_W \approx kT_c/\mu$. For iron, $T_c \approx 10^3$, $\mu \approx 2\mu_B \approx 2 \times 10^{-20}$ and $k \approx 10^{-16}$, so $H_W \approx 5 \times 10^6$ Oe. This is a very powerful field; it cannot be accounted for by ordinary magnetic dipole interactions, since the magnetic field at an electron due to the magnetic moment of an electron on a neighboring atom is given by $H \approx \mu_B/r^3 \approx 10^3$ Oe.

Origin of Molecular Field - Exchange Energy. The source of the molecular field proved to be of purely quantum mechanical origin

and was not established until the work of Heisenberg¹² and of Dirac¹³. This effect can most simply be illustrated by consideration of the hydrogen molecule, consisting of two nuclei (a and b) and two electrons (1 and 2). For this case, the Schrödinger equation is given by

$$(H_0 + H') \Psi = E \Psi \quad (19)$$

$$H_0 = \frac{\hbar^2}{2m} (\nabla_1^2 + \nabla_2^2) - \frac{e^2}{r_1} - \frac{e^2}{r_2} \quad (20)$$

$$H' = -\frac{e^2}{r_{1b}} - \frac{e^2}{r_{2a}} + \frac{e^2}{r_{12}} \quad (21)$$

where r_{1b} and r_{2a} are the distances between the electrons 1 or 2 and nuclei b or a respectively, and r_{12} is the inter-electronic distance. The wave function can be expressed as the product of a spatial function ψ and a spin function χ

$$\Psi = \psi(r_1, r_2) \chi(\sigma_1, \sigma_2) \quad . \quad (22)$$

Before proceeding further it is necessary to introduce the further assumption of the Pauli exclusion principle, the quantum-mechanical formulation of which is that the total eigenfunction of a system with several electrons is antisymmetric in all its electrons. Therefore, if the spatial function is symmetric, the spin function must be antisymmetric, and vice versa. The spatial function may be expressed as either

$$\begin{aligned} \psi_s(r_1, r_2) &= [u_1(1)u_2(2) + u_2(1)u_1(2)] \\ \psi_a(r_1, r_2) &= [u_1(1)u_2(2) - u_2(1)u_1(2)] \end{aligned} \quad (23)$$

where the ψ_s and ψ_a correspond to symmetric and antisymmetric functions, respectively, and $u(1)$ and $u(2)$ are related to the hydrogen atom wave functions. The nature of this relationship depends on whether a Heitler-London (localized) or a molecular-orbital (itinerant) description is employed¹⁴.

The spin function of a single electron can have only two values, $+\frac{1}{2}$ and $-\frac{1}{2}$. These are symbolized by α and β respectively. For the two electron system we have been considering, there are four possible states. Three of them are symmetric, representing a triplet spin function ($S=1$) and correspond to a ferromagnetic spin arrangement. They are

$$\chi_s = \frac{1}{\sqrt{2}} \begin{bmatrix} \alpha(1) & \alpha(2) \\ \beta(1) & \beta(2) \end{bmatrix} \quad (24)$$

and a singlet ($S=0$) antisymmetric spin function corresponding to an antiferromagnetic spin arrangement.

$$\chi_a = \frac{1}{\sqrt{2}} \begin{bmatrix} \alpha(1) \beta(2) & -\alpha(2) \beta(1) \end{bmatrix} . \quad (25)$$

According to the Pauli exclusion principle, the total wave function is restricted to $\psi_s(r_1, r_2) \chi_s(\sigma_1, \sigma_2)$ or $\psi_a(r_1, r_2) \chi_a(\sigma_1, \sigma_2)$.

The energy of the system is dependent solely on the spatial function, and if we include the portion of the Hamiltonian H' as a perturbation, the energies of the singlet and triplet states are

$$E = 2E_0 + K \pm J \quad (26)$$

where the positive sign corresponds to the symmetric spatial function associated with the singlet spin function, the negative sign to the antisymmetric spatial function (triplet), E_0 is the energy of an unperturbed free atom and

$$K = \int |u_1(1)|^2 H' |u_2(2)|^2 d\tau_1 d\tau_2 \quad (27)$$

$$J = \int u_1^*(1) u_2^*(2) H' u_2(1) u_1(2) d\tau_1 d\tau_2 . \quad (28)$$

K is just a Coulomb energy term and J is called the exchange energy. From Eqn. (26) it is apparent that the relative stability of the symmetric and antisymmetric spatial states depends entirely on the sign of J . This exchange energy, which is seen to be of electrostatic origin, is the source of the molecular field.

Although the spin state does not directly effect the energy, it is a necessary condition of the Pauli exclusion principle that the symmetry of the spin function be fixed by the sign of J . It is therefore possible to treat the exchange energy as though it were due to interactions between the spins with the form

$$E_{12} = -2 J_{12} S_1 \cdot S_2 \quad (29)$$

where J_{12} is directly related to J while S_1 and S_2 are the spin angular momentum operators. Treating these operators as classical spins it is seen that a parallel arrangement of S_1 and S_2 is favored by a positive value of J_{12} and an antiparallel arrangement by a negative value of J_{12} .

For the hydrogen molecule we have been considering, the symmetric spatial function (the 'bonding' state) has the lowest energy. Therefore $J_{12} < 0$, and S_1 is antiparallel to S_2 , corresponding to the singlet state.

In dealing with the more general problem of a solid, Eqn. (29) is generalized to

$$H_{ex} = - \sum_{i,j} J_{ij} S_i \cdot S_j \quad (30)$$

where H_{ex} is known as the Heisenberg exchange Hamiltonian, J_{ij} is the exchange interaction between atoms i and j , S_i and S_j are the spins of these atoms and each interaction is counted twice by the unrestricted summation. A more complete discussion of exchange energy is available in many texts on magnetism and in review articles.^{14,15}

III. LONG-RANGE ORDER

Molecular Field Theory in Ferromagnets. In postulating the concept of the molecular field, Weiss assumed it to be proportional to the magnetization,

$$H_W = \lambda M \quad . \quad (31)$$

Taking the total magnetic field as the sum of the applied field H and the molecular field H_W , we can repeat the derivation of the magnetization for the case of paramagnetism (Eqn. 14) to obtain

$$M = Ng\mu_B J B_J(y) \quad (32)$$

but with the total field included in the argument of the Brillouin function; that is, $y = g\mu_B J(H + \lambda M)/kT$. It then follows that for $y \ll 1$

$$M = \frac{Ng^2\mu_B^2 J(J+1)}{3kT} (H + \lambda M) = \frac{C}{T} (H + \lambda M) \quad (33)$$

or

$$M = \frac{CH}{T - C\lambda} \quad . \quad (34)$$

Defining $C\lambda$ as the paramagnetic Curie point θ , and recalling that the susceptibility $\chi = M/H$

$$\chi = \frac{C}{T - \theta} \quad . \quad (35)$$

Equation (35) is the Curie-Weiss law, and rather accurately describes the paramagnetic behavior well above the Curie point. Strictly speaking, $T_c = 0$ within the molecular field approximation. However, this approximation is quite poor in the vicinity of the ferromagnetic Curie point and experimentally $\theta > T_c$.

Below T_c a ferromagnet exhibits a spontaneous magnetization even in the absence of an applied field. In this temperature region we can no longer use the approximation $y \ll 1$, but must return to Eqn. (32). Noting that at $T = 0$, $y \rightarrow \infty$ and $B_J(\infty) = 1$, we obtain $M_0 = NJg\mu_B$, where $M_0 = M$ at $T = 0$. Therefore

$$\sigma \equiv \frac{M}{M_0} = B_J(y) \quad (36)$$

where σ is called the reduced magnetization. The resultant value of σ as a function of the reduced temperature T/T_c is given by Smart¹⁶ for several values of J .

Weiss Molecular Field and Other Approaches to Study of Ferromagnetism. In addition to its conceptual simplicity, the Weiss molecular field approximation has the advantage of being relatively easy to calculate. It gives a good qualitative description of ferromagnetism and, as we shall discuss later, of anti-ferromagnetism and ferrimagnetism. It predicts the onset of a spontaneous magnetization, fairly reasonably approximates its temperature dependence and gives an accurate description of its high-temperature behavior ($T \gg T_c$). However, on closer inspection, a number of discrepancies appear between theory and experiment, particularly at low temperatures and in the vicinity of T_c .

The molecular field essentially considers a single ion acted upon by the average value of the magnetization. In this approximation all magnetic order disappears at the Curie point. In actual fact, considerable short range order persists above the Curie point and accounts for the discrepancy between T and θ . Improvements in the molecular field theory can be made by dealing with a cluster of atoms, calculating the interactions exactly within the cluster and then treating the interactions between the cluster and all the other atoms in the crystal by means of an effective field. The three cluster methods are the Oguchi, the constant coupling and the Bethe-Peierls-Weiss method. Although these methods, and particularly the latter two, yield improvements at high temperatures ($T > T_c$), they are less effective at low temperatures. In fact, the constant-coupling method fails to give complete saturation at $T = 0$, while the B-P-W method fails completely at low temperature. These methods as well as molecular-field theory are discussed thoroughly by Smart¹⁶.

Another method for studying the susceptibility above the Curie

point involves a high-temperature expansion of the form $\chi^{-1} \propto \sum_{n=1}^{\infty} a_n T^{-n}$. A highly accurate determination of χ can be obtained by considering a sufficiently large number of terms. However, since the evaluation of the coefficients become increasingly more difficult with increasing n , the general procedure is to evaluate a fairly small number of terms and then to use approximation techniques to determine the radius of convergence of the series. In the equation $\chi \propto (T-T_c)^{-\gamma}$ just above T_c , this approach has been able to account satisfactorily for the experimentally observed values of γ , which generally vary from 1.26 to 1.4. By comparison, molecular field theory predicts $\gamma = 1$. The general subject of high-temperature expansion and magnetic-transition phenomena are covered in a recent book by Stanley¹⁷.

None of the above methods are suitable for a study of low-temperature magnetic behavior. In this temperature region the spin-wave theory is most accurate. At $T = 0$, the individual spins can be thought of as completely lined up in parallel along the z direction. As the temperature is raised, thermal effects will give rise to deviations of the spins from the direction of their neighbors with a definite periodicity. These deviations in the $x-y$ plane are sinusoidal and are called spin-waves. By means of these spin waves it is possible to reduce M by $2u_B$ with much less energy cost than would be required by a spin flip. To lowest order, spin wave theory predicts that at low temperatures ($T \ll T_c$)

$$\frac{M(T)}{M_0} = 1 - AT^{3/2} \quad (37)$$

The $T^{3/2}$ dependence, as given in Eqn. (37), has been observed in many materials. Numerous discussions of spin waves exist in the literature, such as the one by Keffler and Loudon.¹⁸ A semi-classical derivation of Eqn. (37) is given by Kittel (Ref. 1, Appendix O).

Antiferromagnetism. In our discussion of exchange interactions it was noted that for $J_{ij} < 0$ in Eqn.(30), energy considerations favored an antiferromagnetic arrangement of the atomic moments. We will consider the simplest case in the molecular-field approximation, namely a two-sublattice structure in which half the atomic sites (a-sites) have their moments in one direction, and the other half (b-sites) have their moments oppositely directed. We further assume that a magnetic interaction exists only between a-site and b-site moments. The Weiss molecular field at sites a and b are given by

$$H_{wa} = -2\lambda M_b \quad \text{and} \quad H_{wb} = -2\lambda M_a \quad (38)$$

where the factor of 2 enters because there are only $N/2$ atoms at

each site, and the magnetization of each of the sublattices is

$$M_a = M_{ao} B_J(y_a) ; \quad M_b = M_{bo} B_J(y_b) \quad (39)$$

where $y_a = g\mu_B J(H+H_{wa})$ and $y_b = g\mu_B J(H+H_{wb})$. At high temperatures, as before, this leads to

$$M_a = (N/2) \frac{g\mu_B J(J+1)(H-2\lambda M_b)}{3kT} = \frac{C}{2T}(H-2\lambda M_b) \quad (40)$$

and similarly

$$M_b = \frac{C}{2T}(H-2\lambda M_a) . \quad (41)$$

Solving Eqns. (40) and (41) simultaneously yields

$$\frac{M_a M_b}{H} = X = \frac{C}{T+\theta} \quad (42)$$

where $\theta = C\lambda$. Therefore the susceptibility curve above the ordering temperature (called the Néel temperature in antiferromagnets and symbolized by T_N) has the same form as that given by the Curie-Weiss law (Eqn. 35), but with a negative intercept on the temperature axis.

At the transition temperature, there will be a spontaneous magnetization at each of the sublattices in the absence of an applied field. Setting $H = 0$ in Eqns. (40) and (41) we obtain

$$M_a = -M_b \quad \text{and} \quad T_N = -\theta . \quad (43)$$

Below T_N the individual sublattice moments will follow a Brillouin function, as indicated by (39), with $M_a = -M_b$. The susceptibility in the ordered region has been investigated by Van Vleck,¹⁹ who found that for a field parallel to the atomic moments the susceptibility $X_{||}$ is independent of temperature. However, for a field perpendicular to the atomic moment the susceptibility X_{\perp} below T_N decreases with decreasing temperature to zero at $T = 0$. Therefore, in a polycrystalline sample, averaging the susceptibility over all directions

$$X(0) = \frac{1}{3} X_{||}(0) + \frac{2}{3} X_{\perp}(0) = \frac{2}{3} X(T_N) . \quad (44)$$

The predicted inverse susceptibility vs. temperature curve in an antiferromagnet therefore appears as shown in Fig. 2a. A large number of materials have been found that exhibit the general behavior discussed above. A rather complete list with their basic properties as of 1961 is given in "Magnetism and the Chemical Bond" by Goodenough.²⁰

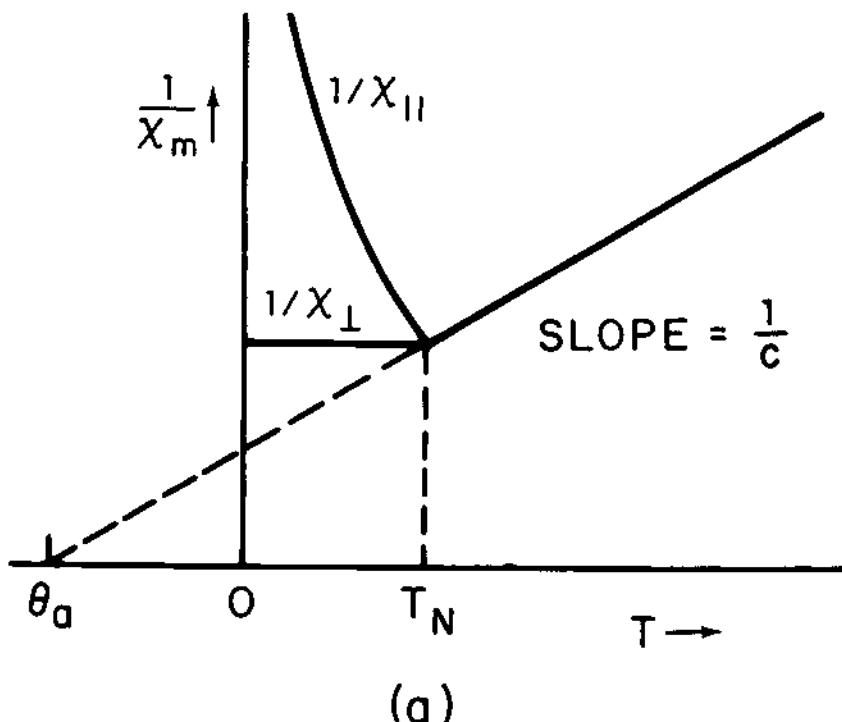


Fig. 2(a). Theoretical curves of inverse susceptibility vs. temperature for antiferromagnetism.

In many antiferromagnetic materials the ratio θ/T_N differs markedly from the value -1, as given by Eqn. (43). (for example $-\theta/T_N \approx 5.0$ in MnO.) The major fault lies in the assumption of two sublattices with interactions only between sites on different sublattices. Although this can apply to nearest-neighbor interactions in a bcc structure, it would be impossible to have such an arrangement in an fcc structure. To generalize to an antiferromagnet containing n sublattices, the total field acting on the i^{th} sublattice is given by

$$H_i = H + \sum_{j=1}^n \lambda_{ij} M_j . \quad (45)$$

Assuming $\lambda_{ii} = 0$, at high temperature this leads to n coupled linear equations of the form

$$M_i = \frac{C}{nT} (H + \sum_{j=1}^n \lambda_{ij} M_j) \quad i=1,2,3\dots n \quad (46)$$

It can then be shown (see Smart¹⁶ for details) that

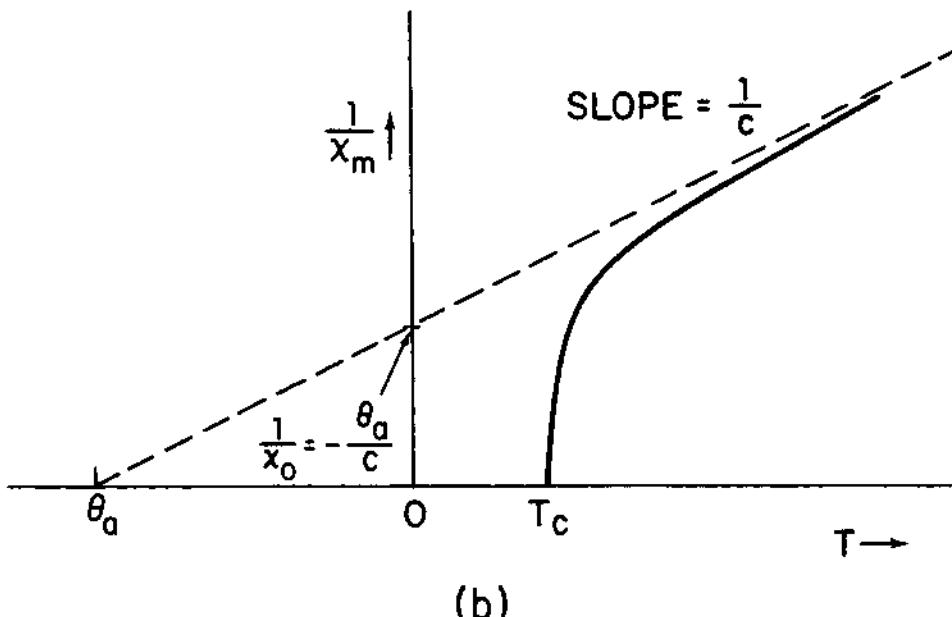


Fig. 2(b). Theoretical curves of inverse susceptibility vs. temperature for ferrimagnetism.

$$\theta = \frac{c}{n} \sum_{j=1}^n \lambda_{ij} \quad \text{and} \quad T_N = \frac{c}{n} \sum_{j=1}^n \epsilon_{ij} \lambda_{ij} \quad (47)$$

where $\epsilon_{ij} = \pm 1$. Further, $M_i = \epsilon_{ij} M_j$ and $\sum_{i=1}^n M_i = 0$.

Thus the paramagnetic Curie point remains the algebraic sum of the exchange interactions but can be different in magnitude from the ordering temperature.

Ferrimagnetism. In our considerations of ferro- and anti-ferromagnetism, we have dealt exclusively with magnetic ions or atoms on equivalent lattice sites. With this restriction removed, it is possible to have an antiparallel arrangement of these moments with a net spontaneous moment below the ordering temperature. This phenomenon was termed ferrimagnetism by Néel, who used the above model to explain the magnetic behavior of materials with the spinel structure, which is shown in Fig. 3. Although ferrimagnetism has since been observed in the garnet structure, in a series of rather complex hexagonal structures,²¹ and other structures (i.e., Cr_5S_6 , RbNiF_3), we will limit our considerations to spinels and follow

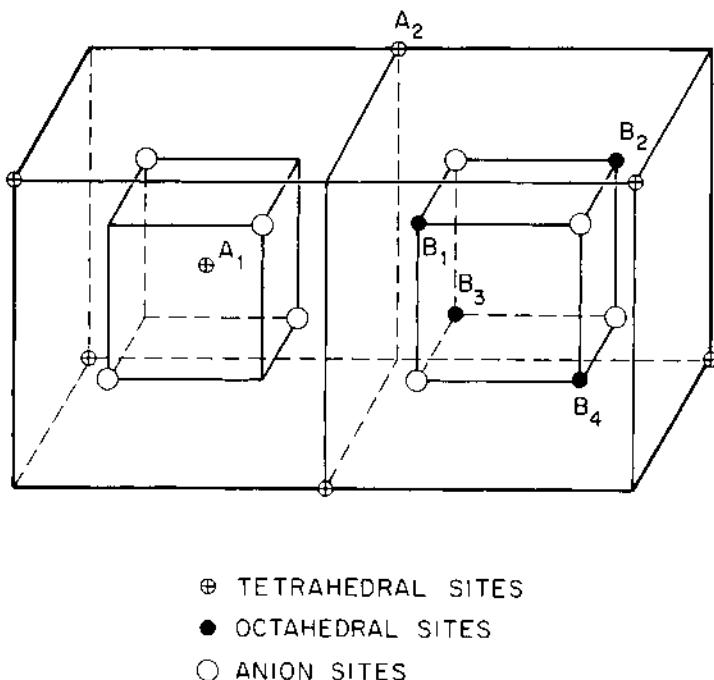


Fig. 3. Spinel lattice structure.

Néel's treatment.

As seen in Fig. 3, the spinel lattice consists of two types of lattice sites. They are the "A" sites, which are tetrahedrally coordinated, and the "B" sites, which have octahedral coordination. The formula unit for a spinel oxide is AB_2O_4 .

Following Néel, we consider only a single kind of magnetic ion with λ and μ the proportion of magnetic ions on A and B sites respectively. Magnetic interactions are limited to nearest-neighbors, with n representing the molecular field coefficient between near neighbor A-B interaction, αn the A-A interaction coefficient and βn the B-B interaction coefficient. The molecular fields at A and B sites are then

$$\begin{aligned}
 H_{wa} &= n (\alpha \lambda M_a - M_b) \\
 H_{wb} &= n (\beta \mu M_b - M_a)
 \end{aligned} \tag{48}$$

where the minus sign indicates we are considering only the case of anti-ferromagnetic A-B interactions. Néel has shown that in this

case, the high-temperature susceptibility can be expressed in the form

$$\frac{1}{X} = \frac{T-\theta_a}{C} - \frac{\sigma}{T-\theta} \quad (49)$$

where $\theta_a = -nC(2\lambda\mu - \lambda^2\alpha - \mu^2\beta)$
 $\sigma = n^2C\lambda\mu [\lambda(1+\alpha) - \mu(1+\beta)]^2$
 $\theta = nC\lambda\mu (2+\alpha+\beta)$

and C is the molar Curie constant.

Equation (49) represents an hyperbola with an asymptotic Curie point θ_a and a Curie point

$$T_c = \frac{nC}{2} [\lambda\alpha + \mu\beta + \sqrt{(\lambda\alpha - \mu\beta)^2 + 4\lambda\mu}] \quad (50)$$

Thus the predicted form of the inverse paramagnetic susceptibility of a ferrimagnetic material is given in Fig. 2b which accords well with observed behavior.

At temperatures below T_c the magnetization of each sublattice follows a Brillouin function.

$$M_a = M_a(0)B_J\left(\frac{\mu H_{wa}}{kT}\right) ; M_b = M_b(0)B_J\left(\frac{\mu H_{wb}}{kT}\right) \quad (51)$$

Since $H_{wa} \neq H_{wb}$, M_a and M_b will, in general, have different temperature dependences. Thus the relative contribution of the sublattice magnetizations to the total observed M, where

$$M = M_a + M_b \quad (52)$$

will vary as a function of temperature, with the net magnetization departing significantly from a Brillouin function. Néel investigated the nature of the net magnetization curves as functions of the parameters α , β , λ and μ . The predicted curves for the value $\lambda/\mu = 2/3$ are shown as functions of α and β in Fig. 4. It is seen that the predicted magnetization vs. temperature curves include ones in which the net magnetic moment increases with increasing temperature to some maximum value before falling to zero at the Curie point, and the possibility of a sign reversal below the Curie point. Although such curves were unknown at the time of Néel's paper, both types have since been observed (i.e., $MnCr_2S_4$ (type P) and Co_2VO_4 (type N)).

Equation (52), with M_a and M_b antiparallel, was also able to explain the low saturation moment values observed in ferrites at low temperatures. The ferrites have the formula $M^{2+}Fe_2^{3+}O_4$, where

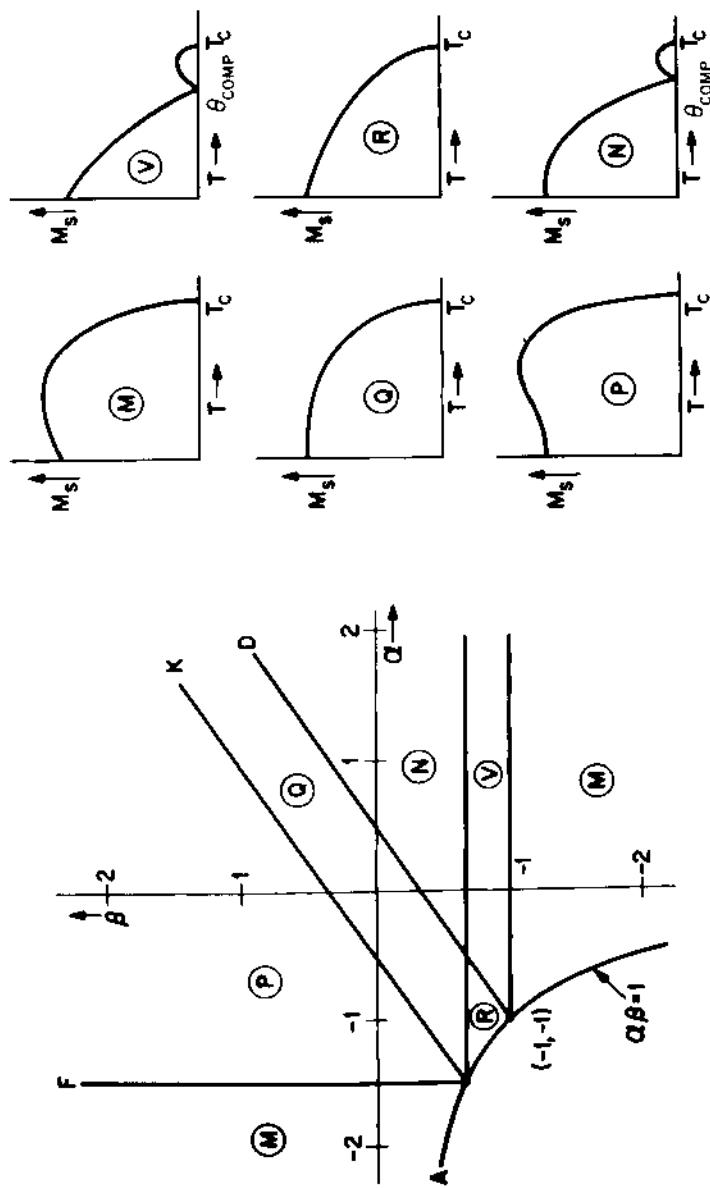


Fig. 4. Néel collinear ferrimagnetic model: (a) Separation of magnetization-curve regions of α - β plane for $\lambda/\mu = 2/3$. (b) General shape of magnetization curves within the corresponding regions.

M^{2+} is usually a transition-metal ion ($M = Mn, Fe, Co, Ni$). Since the total spin quantum number of Fe^{3+} is $S = 5/2$, the expected contribution of the Fe^{3+} to the total moment is $10\mu_B$ per formula unit, and there is an additional contribution of $2 S_M \mu_B$ per formula unit, where S_M is the total spin quantum number of cation M . However, instead of a saturation moment of $(10 + 2S_M)\mu_B$, a value much closer to $2S_M\mu_B$ was observed.²² This can be explained by assuming half the Fe^{3+} ions to be on the A-sites, half on the B-sites and the M^{2+} ions on the remaining B-sites. (Such an arrangement of cations is referred to as an inverse spinel.) With M_a antiparallel to M_b , this leads to an antiferromagnetic cancellation of the Fe^{3+} moments, leaving a net moment $M = 2S_M\mu_B$ /formula unit.

Thus the Néel theory was able to account for a large body of apparently anomalous behavior, namely the hyperbolic inverse susceptibility curve in the paramagnetic region, the anomalous magnetization curves and the low saturation moment.

Triangular Spin Arrangements. Despite these remarkable successes, the Néel theory contained some disturbing elements. Specifically, in the regions of Fig. 4 labelled R, V and T the moment at one of the sites is not saturated at $T = 0$, leading to the result $dM/dT \neq 0$ at $T = 0$, in violation of the third law of thermodynamics. In addition, it seemed unreasonable that the existence of large antiferromagnetic A-A and B-B interactions should lead to paramagnetism, as was the Néel result for region G. Yafet and Kittel²³ showed these problems could be resolved by removing the overly restrictive condition of a single A-site and single B-site sublattice. When this constraint is removed, the B-sites can be divided into B and B' sublattices, and the A-sites into A and A' sublattices. A large antiferromagnetic B-B interaction will then lead to a canting of B and B' sublattices relative to each other, and similarly for the A and A' sublattices. For the region of large antiferromagnetic A-A and B-B interactions (essentially region G), the material will become antiferromagnetic at low temperatures with A-sublattice moments antiparallel to those on the A' sublattice and B antiparallel to B'.

The Yafet-Kittel model, introduced in 1952, was the first departure from the a priori assumption of collinear spins in an ordered magnetic system. Triangular ordering of the B-sites has been observed in the tetragonally distorted spinel $CuCr_2O_4$,²⁴ and Plumier²⁵ has recently observed A-site canting in $MnCr_2S_4$.

IV. SPIRAL CONFIGURATIONS

Kaplan⁹ found that in cubic spinels with only nearest-

neighbor A-B and B-B interactions, the Yafet-Kittel configuration is locally unstable and hence cannot be the ground state. This led him to consider the simpler problem of a body-centered cubic lattice with first, second- and third-nearest neighbor interactions, and found the exchange energy could be minimized by a spiral spin arrangement. At the same time, Yoshimori⁸ and Villain¹⁰ independently discovered that the presence of competing exchange interactions could lead to magnetic spirals. The effect is perhaps most simply illustrated by considering a simple hexagonal lattice containing equivalent atoms at all sites that are ferromagnetically aligned within the basal planes, and experience a ferromagnetic nearest neighbor interaction J_1 ($J_1 > 0$) and an antiferromagnetic next-nearest neighbor interaction J_2 ($J_2 < 0$) along the c-axis.²⁶ Then, from Eqn. (30)

$$E_{ex} = -2(J_1 \cos \theta + J_2 \cos 2\theta) , \quad (53)$$

and minimizing the energy with respect to the angle θ , one obtains

$$\cos \theta = - \frac{J_1}{4J_2} \quad (54)$$

as the stable configuration for $|4J_2| > |J_1|$. The resultant spiral configuration, with θ defined by Eqn. (54), is shown in Fig. 5. In general, it should be noted that the wavelength λ , defined as the distance in which θ rotates through 360° , need not be an integral number of unit cells. Since their discovery on theoretical grounds, numerous spiral magnetic structures have been found by neutron diffraction experiments. The use of neutron diffraction techniques to determine magnetic structures are discussed by Dr. D.C. Khan in the following chapter.

The Classical Ground State Problem. The classical ground-state problem consists of treating the spins as vectors of fixed length in Eqn. (30) and then finding the set of such vectors that minimizes the Heisenberg exchange energy. For the case in which all the interactions are ferromagnetic, or for the case we considered previously of a two-sublattice material in which the only interaction is an antiferromagnetic one between the two sublattices, there are no competitive interactions, and the solution is intuitively obvious. However, when competitive interactions exist, a detailed investigation of the exchange energy equation is required. This is a many-body problem, for which exact solutions are normally unattainable. However, by using a technique first introduced by Luttinger and Tisza, Lyons and Kaplan²⁷ (L-K) were able to prove rigorously that in a Bravais lattice the classical ground state of the Heisenberg exchange Hamiltonian is always given by the planar spiral

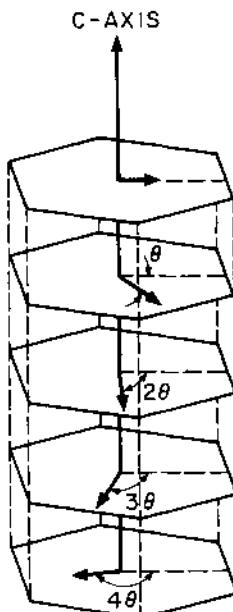


Fig. 5. Planar spiral configuration in a hexagonal lattice (after Enz, Ref. 26).

$$S_n = |S| (\hat{x} \sin k_o \cdot R_n + \hat{y} \cos k_o \cdot R_n) \quad (55)$$

where R_n defines the position of the atom in the n^{th} unit cell and k is a wavevector in the reciprocal space with wavelength given by $|k| = 2\pi/\lambda$ and k_o is the value of k which minimizes L in the equation

$$L = \sum_R J(R) e^{ik \cdot R} \quad (56)$$

The Luttinger-Tisza method has been used successfully in some non-Bravais lattices, such as the hexagonal and corundum structures, which contain equivalent atoms at all the sites. However, this approach does not work for non-Bravais lattices with arbitrarily different spin values. To overcome this difficulty, L-K introduced a generalized Luttinger-Tisza method (GLT) which made it possible to determine the classical ground state in some cases where the original approach failed.

Application to Spinels. A detailed investigation of the classical ground-state problem in cubic and tetragonally distorted normal spinels was undertaken by Kaplan, Dwight, Lyons and Menyuk²⁸⁻³¹ using the GLT method. (Normal spinels are defined as

those in which all the B-sites contain a single type of ion and all the A-sites contain a different ion. An example is $Mn^{2+}[Cr_2^{3+}]O_4^{2-}$, in which the Mn^{2+} ions are all at A-sites and the Cr^{3+} at B-sites.) For a cubic spinel with nearest neighbor antiferromagnetic A-B and B-B interactions, the ground-state can be expressed as a function of the single parameter

$$\underline{u} = \frac{4}{3} \frac{J_{BB}^S B}{J_{AB}^S A} . \quad (57)$$

It was found that for $0 < \underline{u} < \underline{u}_0 = 8/9$ the Néel configuration is the ground state. Precisely at the point at which the Néel state becomes unstable, a lower energy configuration was obtained consisting of conical spirals at each of the sublattices. Such spirals consist of two wavevectors; a $k = 0$ spin component along the cone axis and a spiraling $k \neq 0$ component perpendicular to the cone axis. In the spinel configuration there are three sets of such ferrimagnetic spirals, each with a different cone angle ϕ . They are ϕ_1 , defining the cone angle for the A-sites, and, with the site numbers as given in Fig. 3, ϕ_3 and ϕ_5 , which define the cone angles for sites B_1-B_2 and B_3-B_4 , respectively. The cone axes are collinear, with the spin component along the A-site axes anti-parallel to those of the B-site axes. The sites all have the same spiraling k vector along the [110] direction. All these values, as well as the phase angles, are single valued functions of \underline{u} , as shown in Fig. 6. The energy of this configuration is lower than that of Yafet and Kittel for all values of \underline{u} . Although unable to prove that the spiral configuration described above is the ground state, Lyons *et al.*²⁹ (LKDM) showed it to be locally stable for $\underline{u} < \underline{u}'' = 1.3$ but unstable for $\underline{u} > \underline{u}''$. It was also shown that for $\underline{u} < \underline{u}' = 1.35$ the LKDM configuration is the lowest energy magnetic spiral, a non-trivial result since it involves 14 parameters; i.e., $k_x, k_y, k_z, 6$ cone angles and 5 phase angles. These results indicate that the spiral configuration may be the ground state for $\underline{u} < \underline{u}''$, definitely is not the ground state for $\underline{u} > \underline{u}''$, and that the configuration which destabilizes the above mode must be more complex than a conical spiral, involving additional k -vector modes.

In a neutron-diffraction study of $MnCr_2O_4$, Hastings and Corliss³² found a complicated diffraction pattern corresponding closely to that predicted by LKDM. Taking the value $u = 1.6$ from the magnetization of the material, the non-zero k -vector was found to have the predicted value, and the relative intensities of the peaks corresponding to the spiraling component were also in accord with theory. The peak intensities corresponding to the axial components of the moment were also internally consistent with theory, but were about 20% larger than expected relative to the spiral peaks. Since the LKDM spiral is known to be unstable at $u = 1.6$,

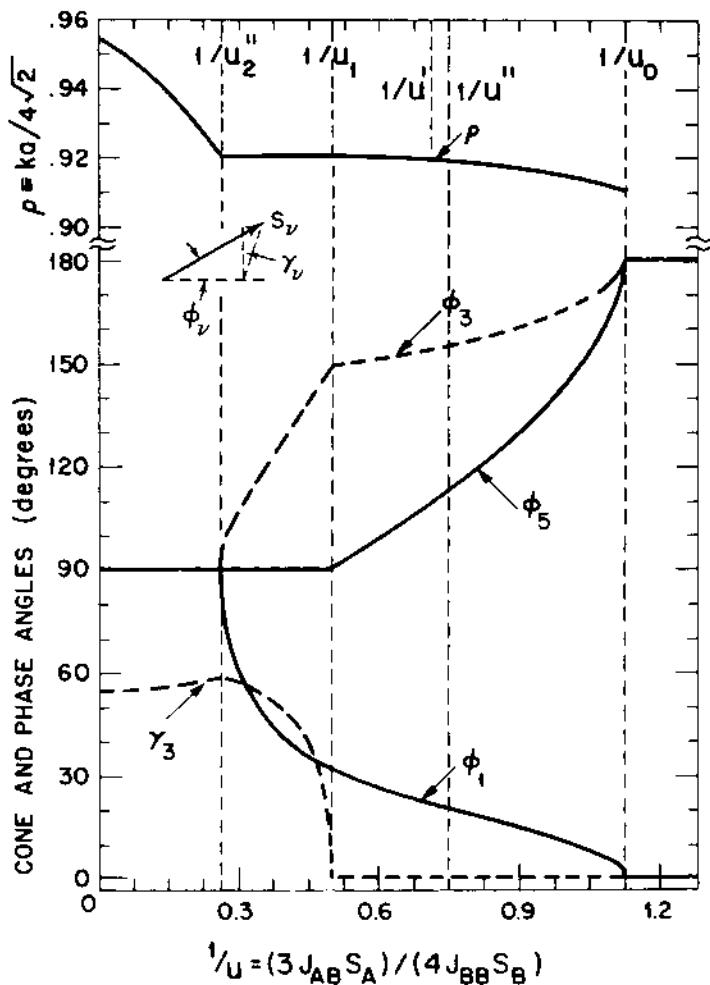


Fig. 6. Variation of cone angles (ϕ_v), phase angles (γ_v) and wave vector (k) with inverse interaction parameter ($1/u$) in spinels according to LKDM theory (Ref. 29).

the results on $MnCr_2O_4$ seem to indicate that this spiral is indeed the ground state for $u < 1.3$, and that there is no gross departure from it at higher u values. Additional confirmation of this was obtained from a neutron-diffraction study of $CoCr_2O_4$,³³ in which the theory was found to agree closely with the experimentally determined magnetic configuration shown in Fig. 7. However, the k -vector magnitude differed from theory by about 5%. It was hypothesized that this difference was due to the effect of more distant-neighbor interactions, which was ignored in the original

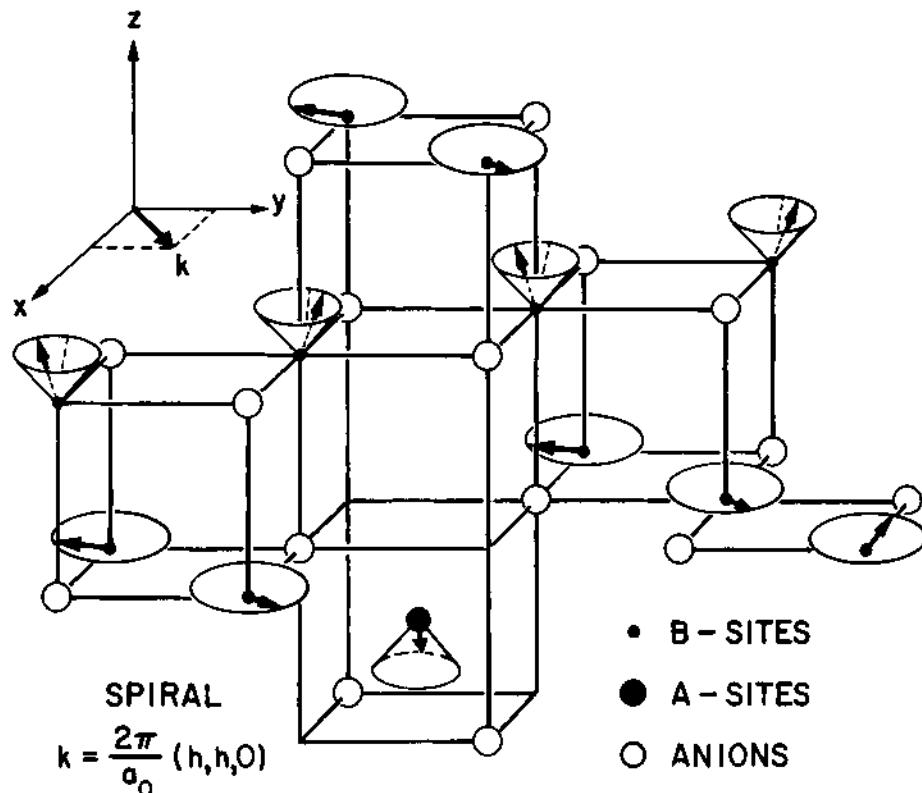


Fig. 7. Ferrimagnetic conical spiral configuration of CoCr_2O_4 .

theory. Apparent confirmation of this hypothesis was obtained by Dwight and Menyuk³⁴ who obtained agreement with both the peak intensities and the wavevector in CoCr_2O_4 when they considered the effect of next-nearest neighbor interactions. In addition, a molecular-field study of the predicted temperature dependence of magnetization in CoCr_2O_4 , assuming the spiral configuration at $T = 0$, agreed remarkably well with experiment, even to the presence of a kink at $T \approx 0.3T_c$, as shown in Fig. 8. The kink is apparently due to the rapid change in the phase angle γ_3 which occurs in this temperature region.

Studies of tetragonally distorted spinels showed the presence of regions in the interaction parameter space in which Yafet-Kittel type configurations are the ground state³⁰, in agreement with the

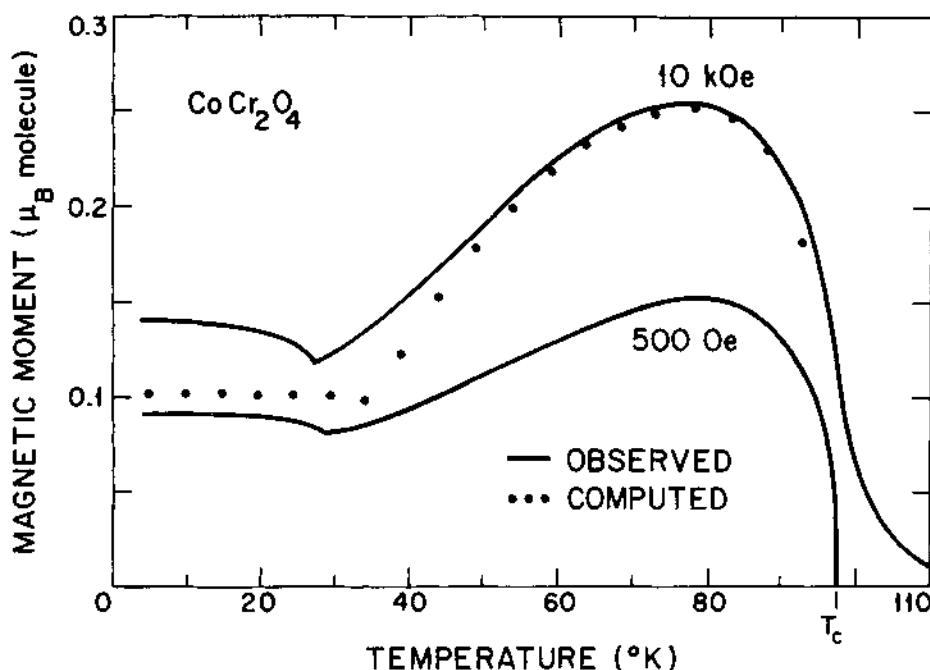


Fig. 8. Comparison between experimentally observed temperature variation of magnetization in CoCr_2O_4 and curve computed on the basis of molecular field theory (Ref. 34).

results observed on CuCr_2O_4 . There are also regions in which spirals along [h00] and [00l] directions are the ground state.³¹ To date, however, such configurations have not been observed experimentally. It should be mentioned that in MnV_2O_4 and CoV_2O_4 , which are normal cubic spinels, neutron diffraction experiments indicate spin configurations which are of the Yafet-Kittel type.^{35,36} This apparent discrepancy is due to the orbital degeneracy of the vanadium ion, which has a $(3d)^2$ configuration but an experimental vanadium-ion moment of $1.2 \mu_B$ rather than the $2.0 \mu_B$ predicted for its spin-only value. Interaction with the local trigonal component to the octahedral-site trigonal field introduces, in this case, large local anisotropies. Since the Heisenberg Hamiltonian derivation assumes a nondegenerate orbital state and also neglects crystalline anisotropy the spiral model is not applicable in this case.³⁶

V. MAGNETIC ATOMS IN A CRYSTAL

In the previous sections we have ignored the nonspherical effects which the presence of neighboring ions in a crystal have on the magnetic properties. To deal with this problem for the case of several electrons outside a closed shell, we start with the Schrödinger equation, $H\Psi = E\Psi$, where $H = H_0 + V_{el} + V_{cf} + V_{LS+Hex}$. Here, H_0 represents the Hamiltonian for a single outer electron moving in the average spherical potential due to the other outer and core electrons. This gives rise to hydrogenic wave functions and is correct for the case of a single outer electron. However this is no longer the case if there are two or more electrons in the outer shell. Then the term V_{el} must be added to account for the electrostatic interaction between outer electrons. This gives rise to an intra-atomic exchange effect as a consequence of the Pauli exclusion principle, which states that no two electrons on the same atom can have the identical set of quantum numbers. If two of the electrons (i,j) have antiparallel spins (i.e., have different spin quantum numbers), their spatial quantum numbers can be identical. If their spins are parallel, their spatial quantum numbers must be different, and on the average the electrons will be further removed from each other. Therefore the electrostatic repulsion energy $e^2/|r_i - r_j|$ will be lowered, so V_{el} tends to favor a parallel alignment of spins within the atom, with an energy difference (or "splitting") given by Δ_{ex} .

The term V_{cf} is introduced to deal with the perturbing effects of neighboring ions in the crystal. This problem is most simply approached by the ligand-field model,^{37,38} in which these effects are represented by an effective electrostatic field which reflects the symmetry of the lattice site and gives rise to energy level splittings Δ_{LF} . It is impossible to determine the magnitude of the splitting from first principle since the effective field is the sum of several effects (see footnote in Goodenough,²⁰ p. 52), and so must be determined experimentally.

For the case of non-zero orbital angular momentum L , there will be a coupling between L and S given by $V_{LS} = \lambda L \cdot S$, leading to a splitting $\Delta_{LS} \sim \lambda LS$. Hex is the Heisenberg Hamiltonian dealt with previously.

In dealing with the perturbing potentials V_{el} , V_{cf} and V_{LS} , it is necessary to know their magnitudes relative to each other, as reflected in the energy splitting attributable to these potentials. It is found that their relative magnitudes are strongly dependent upon the region of the periodic table in which the materials are located, as indicated in Table I.

Table I

	Δ_{ex}	Δ_{cf}	Δ_{L-S}
Fe-Group (3d)	1.25-5.0 eV	1.25-2.5 eV	0.01-0.1 eV
Pd-Group (4d)	0.5 -2.5 eV	2.5 -5.0 eV	0.05-0.25 eV
Pt-Group (5d)	0.5-2.5 eV	2.5-5.0 eV	0.1-0.6 eV
Rare-Earth (4f)	0.5-5.0 eV	~.02 eV	0.1-0.4 eV

In our discussion of ligand-fields, we will restrict ourselves to consideration of metallic cations of the first transition series; that is, the iron-group containing 3d outer-shell electrons. In this group, the relative energy splittings in fluorides and oxides, with the possible exception of trivalent cobalt or nickel or tetravalent ions, are $\Delta_{ex} > \Delta_{cf} > \Delta_{L-S}$. This is called the medium field case, for which Hund's rule is said to hold. Hund's rule states that of the terms given by equivalent electrons, those with the greatest value of $(2S+1)$ has lowest energy, and of those the lowest is the one with the greatest value of L. The first part of Hund's rule merely states that the outer-shell electrons of a single atom will tend to align ferromagnetically within the framework of the Pauli exclusion principle. This follows directly from the discussion of V_{el} given above. Except where otherwise stated we shall generally assume that Hund's rule holds in the succeeding discussion.

Hydrogenic Orbitals and Ligand Field Splitting. In solving the Schrödinger equation $H\psi = E\psi$ for the hydrogen atom, the spatial function $\psi(r, \theta, \phi)$ can be expressed as a product of functions $\psi(r, \theta, \phi) = R(r)\Theta(\theta)\Phi(\phi)$, where $R(r)$ is the radial function and $\Theta(\theta), \Phi(\phi)$ are the directional functions in polar coordinates. In a crystal with a single electron outside a closed shell one can use the solutions obtained from the hydrogen atom problem to describe the angular dependence of the orbitals. These are called hydrogenic orbitals. For s-electrons ($l=0$) the orbitals are spherical, i.e., there is no angular dependence. For p($l=1$) and d($l=2$) electrons there are three and five mutually orthogonal orbitals which can be represented in Cartesian coordinates by the functions

$$p_x \sim \frac{x}{r}; p_y \sim \frac{y}{r}; p_z \sim \frac{z}{r} \quad (58)$$

$$d_{xy} \sim \frac{xy}{r^2}; d_{yz} \sim \frac{yz}{r^2}; d_{zx} \sim \frac{zx}{r^2} \quad (59)$$

$$d_{x^2-y^2} \sim \frac{x^2-y^2}{r^2}; d_z^2 \sim \frac{2z^2-x^2-y^2}{r^2}.$$

The angular dependence of these functions are illustrated in Fig. 9.

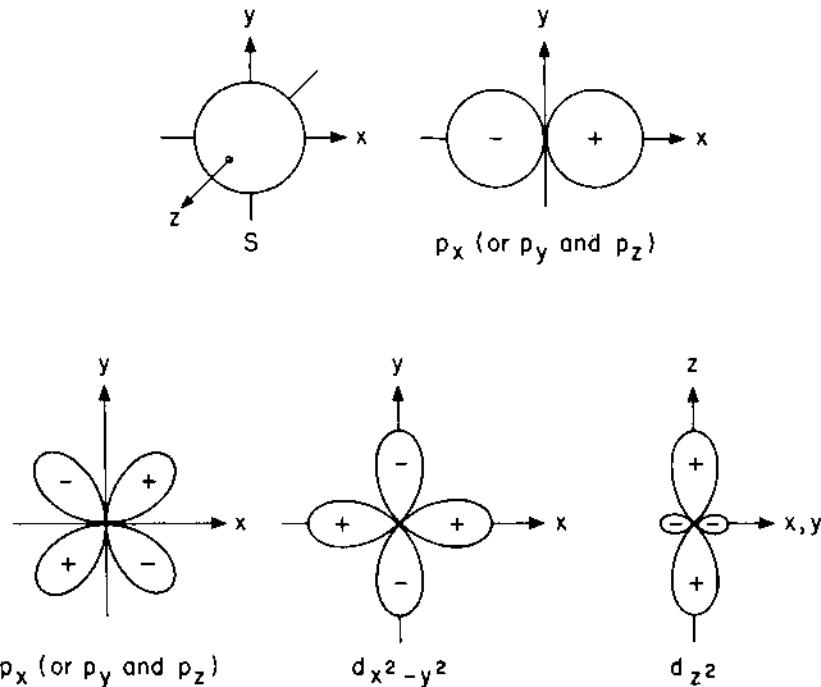


Fig. 9. Schematic illustration of angular dependence of hydrogenic s, p and d wave functions.

Let us now consider an ion containing a single 3d outer electron. For a free ion, all five orbitals as given above are degenerate. If we now place this ion in an octahedral interstice with negative ion neighbors, as indicated in Fig. 10a, electrons in the $d_{x^2-y^2}$ orbital will have their maximum density pointing directly at the ligands along the x and y axes. The d_{xy} orbital, on the other hand, will point away from the ligands, as in Fig. 10b. Therefore, the presence of the octahedral ligand field will tend to destabilize the $d_{x^2-y^2}$ orbital relative to the d_{xy} orbital. It is apparent that the d_{yz} and d_{zx} orbitals will remain degenerate with the d_{xy} orbital. Since $d_z^2 \sim \{ (z^2-x^2) + (z^2-y^2) \} / r^2$, it is also apparent that the d_z^2 orbital remains degenerate with that of $d_{x^2-y^2}$. Therefore, the effect on an octahedral environment is to split the five-fold degeneracy into a more stable three-fold degenerate level containing d_{xy} , d_{yz} and d_{zx} orbitals (henceforth referred to as t_{2g} orbitals) and a two-fold degenerate level consisting of the $d_{x^2-y^2}$

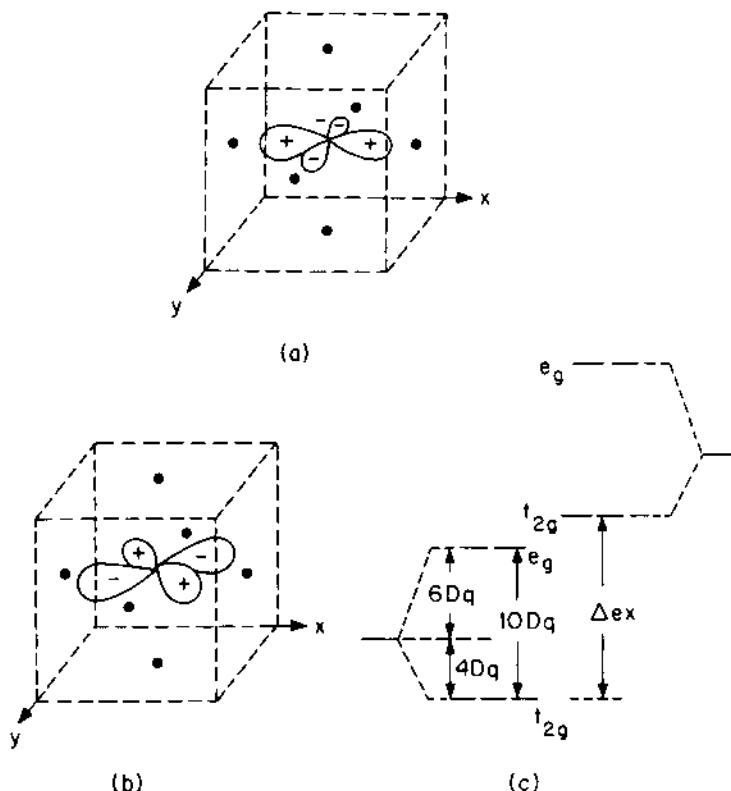


Fig. 10. d-orbitals in octahedral interstice with ligands indicated by black dots. (a) $d_{x^2-y^2}$ (e_g) orbital; (b) d_{xy} (t_{2g}) orbital; (c) resultant ligand-field splitting 10 Dq , with exchange splitting $\Delta_{\text{ex}} > 10 \text{ Dq}$ also shown.

and d_{z^2} orbitals (referred to as e_g orbitals). The splitting of the levels is shown in Fig. 10c where, following standard notation, Δ_{cf} is taken equal to 10 Dq . In the electrostatic point charge approximation, the energy of the d-state manifold is conserved, and the energy levels of the t_{2g} orbitals are lowered by 4 Dq and those of the e_g orbitals are raised by 6 Dq . In fact, the covalent mixing between cationic d and anionic s and p orbitals gives the principal contribution to 10 Dq . This does not affect the symmetry argument of the problem, but it now requires conservation of energy of all the admixed d, s and p states. This has the effect of raising the mean energy of the cationic d states relative to the anionic s and p states.

For the case of a tetrahedral interstice, the situation is

reversed. Now the e_g orbitals are stabilized relative to the t_{2g} orbitals by $10 D_q'$. However, since there are only 4 ligands, and since the t_{2g} orbitals do not point directly toward them, the total splitting in this case is smaller than that of an octahedral environment. From purely electrostatic considerations one would predict $(D_q)' = (4/9) D_q$.

With two 3d electrons in the outer shell, these electrons can combine in the free atom to form a number of terms, the lowest lying of which, by Hund's rule, is the 3F state, which is seven-fold degenerate. This term is further split by the ligand field into a low-lying 3-fold degenerate state, an additional 3-fold degenerate state and a non-degenerate state, as indicated in Fig. 11a. Although this is the correct way to approach the multiple 3d-electron problem, let us consider it from a one-electron point of view. Then we simply add the second electron into one of the

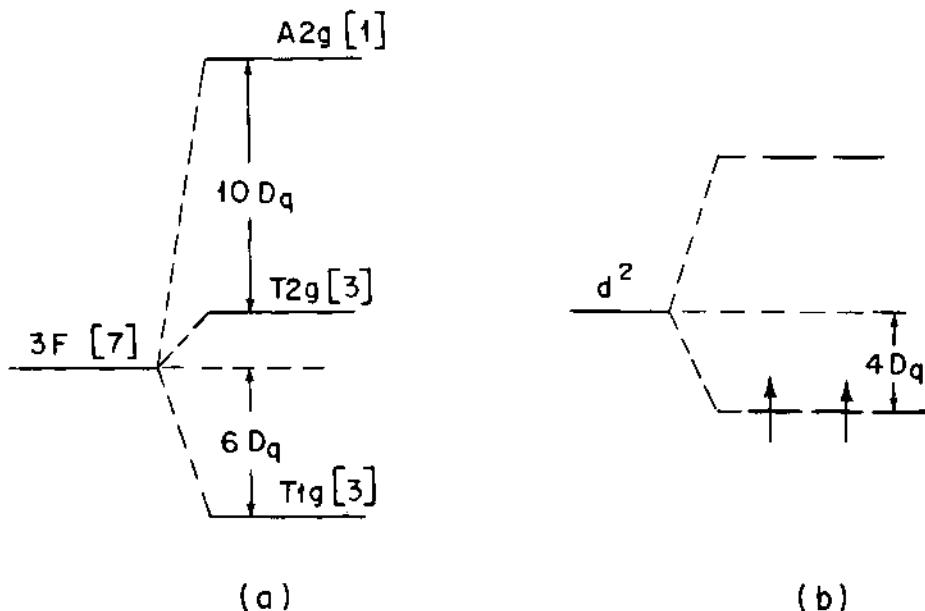


Fig. 11. Ligand-field splitting for two d-electrons in octahedral interstice: (a) two-electron model, with degeneracy of each level shown in brackets; (b) one-electron model.

empty t_{2g} orbitals with its spin parallel to that of the first electron. By Hund's rule, the antiparallel spin-states are destabilized by an amount Δ_{ex} . The one-electron model is shown in Fig. 11b, where the occupation of the t_{2g} orbitals by two parallel-spin electrons is indicated by arrows. Since there are three ways of occupying three orbitals, this state is also three-fold degenerate. The one-electron model is therefore able to describe the degeneracy of the $3d^2$ ground state. Although one should not use the one-electron model for detailed quantitative calculations, it correctly describes the major qualitative aspects of the problem and is much easier to visualize. We will therefore use it for the remainder of our discussion. The spin occupancy on the t_{2g} and e_g orbitals are then as given in Table II.

Table II

No. of 3d Electrons		Octahedral Site Arrangement		Tetrahedral Site Arrangement	
		t_{2g}	e_g	e_g	t_{2g}
1 (Ti^{3+}, V^{4+})	2_D	↑		↑	
2 (V^{3+}, Cr^{4+})	3_F	↑ ↑		↑ ↑	
3 (Cr^{3+}, Mn^{4+})	4_F	↑ ↑ ↑		↑ ↑	↑
4 (Mn^{3+}, Cr^{2+})	5_D	↑ ↑ ↑ ↑		↑ ↑	↑ ↑
5 (Fe^{3+}, Mn^{2+})	6_S	↑ ↑ ↑ ↑ ↑		↑ ↑	↑ ↑ ↑
6 (Co^{3+}, Fe^{2+})	5_D	↑↓↑↑↑	↑↑	↓↑	↑↑↑
7 (Co^{2+}, Ni^{3+})	4_F	↑↓↑↑↑	↑↑↑	↓↑↓	↑↑↑
8 (Ni^{2+})	3_F	↑↓↑↑↑	↑↑↑	↓↑↓	↓↑↑
9 (Cu^{2+})	2_D	↑↓↑↑↑	↓↑↑	↓↑↓	↓↑↑
10 (Cu^{1+}, Zn^{2+})	1_S	↑↓↑↑↑	↓↑↑	↓↑↓	↓↑↑

The above considerations have important applications with regard to site-preference, orbital angular momentum quenching and Jahn-Teller distortions, which will be discussed in turn.

Site Preference. From Fig. 10c it is apparent that for a cation containing p 3d electrons with m in e_g sites and n in t_{2g} sites ($m+n=p$) which we will describe by the notation $(t_{2g})^n(e_g)^m$, the ligand field (neglecting covalency effects) stabilizes an octahedral site by $(4n-6m)Dq$. Conversely, the ligand field

stabilization for a tetrahedral site is $(6m-4n)Dq'$. Therefore there is no ligand-field site-stabilization energy for d^0 , d^5 or d^{10} configurations in either site. The energy reduction in octahedral sites is then Dq multiplied by the following factors: 4 for d^1 and d^6 , 8* for d^2 and d^7 , 12 for d^3 and d^8 , 6 for d^4 and d^9 . For tetrahedral fields the energy reduction is Dq' times 6 for d^1 and d^6 , 12 for d^2 and d^7 , 8* for d^3 and d^8 , 4 for d^4 and d^9 .

Using the rather simple argument given above and obtaining estimates of Dq and Dq' from spectral data McClure³⁹ and Dunitz and Orgel⁴⁰ were able to achieve qualitative success in accounting for the observed cation distribution in spinels. For example, the large octahedral site stabilization energy for d^3 can account for the fact that chromites are normal spinels, while the lack of any stabilization in d^5 leads to inverse spinels in most ferrites.

There are, of course, other important contributions to site preference energy, such as Madelung energy, covalence effects and elastic energy. A discussion of the overall problem has been given by Goodenough²⁰.

Orbital Angular Momentum Quenching. In a central field, such as that for an isolated ion, the orbital angular momentum component L_z is a constant, giving rise to the orbital angular momentum quantum number m_l according to the equation

$$L_z \psi_m = -\frac{ih}{2\pi} \frac{\partial \psi_m}{\partial \phi} = m_l \psi_m , \quad (60)$$

where the subscript m corresponds to the value m_l .
 $\psi_0 = \psi_z^2 \sim (3\cos^2 \theta - 1)$, $\psi_{+1} = (\psi_{yz} + i\psi_{zx})/\sqrt{2} \sim \sin 2\theta \exp(\pm i\phi)$, and $\psi_{+2} = (\psi_{x^2-y^2} \pm i\psi_{xy})/\sqrt{2} \sim \sin^2 \theta \exp(\pm 2i\phi)$. However, the presence of a cubic field splits the ψ_{+2} orbitals into two real components, which give $m_l = 0$ when operated on by L_z , which is imaginary. Thus in a cubic field the orbital contribution to the magnetic moment of partially filled ψ_{+2} orbitals, which is proportional to the expectation value of L_z , is said to be quenched (a simple example of quenching of p electrons is illustrated in Appendix G of Ref. 1). As a result of this quenching, the observed moment of many ions is near their spin-only value; i.e., $J \approx S$ and $g \approx 2$. However, the ψ_{+1} orbitals are not split by cubic fields, and significant orbital moments can be obtained in octahedral d^1 , d^2 , d^6 and d^7 configurations and in tetrahedral d^3 , d^4 , d^8 and d^9 configurations. Further splittings may occur due to Jahn-Teller distortions from cubic symmetry, and whether the orbital contribution of partially filled ψ_{+1} orbitals is quenched or retained depends upon the sign of the distortion.

* Correct answer is 6 for multi-electron function, as indicated in Fig. 11a for octahedral case.

Jahn-Teller Distortions. The Jahn-Teller theorem states that if the ground state of an ion is orbitally degenerate, with no other perturbation present, the crystal will distort to a lower symmetry to remove the degeneracy.

To illustrate the physical basis of this effect, let us consider an Mn^{3+} ion on an octahedral site. It is a $3d^4$ ion with the configuration $(t_{2g})^3(e_g)^1$. The ground state is doubly degenerate, since the e_g electron can either be in the d_z^2 or the $d_{x^2-y^2}$ states. If the crystal were to distort tetragonally in such a way that $c/a < 1$, the e_g orbital along the z axis, d_z^2 , will be destabilized by an amount δ while $d_{x^2-y^2}^2$ will be stabilized by an equal amount, as shown in Fig. 12. The energy reduction δ supplies the

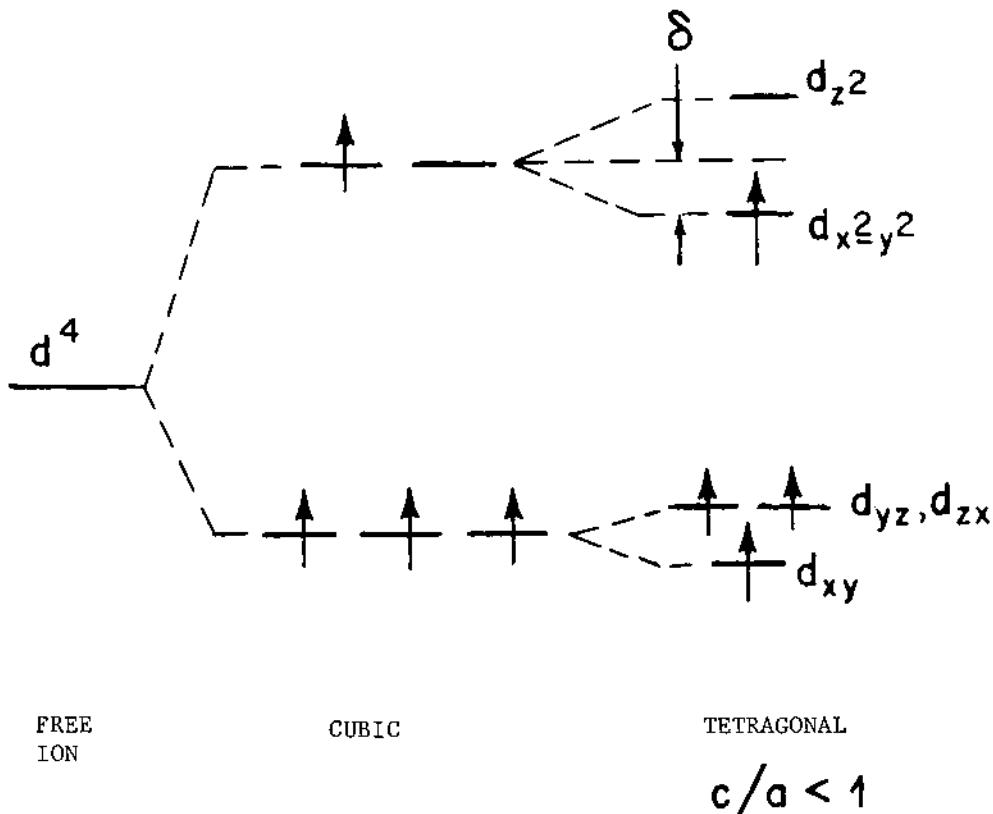


Fig. 12. Ligand-field splitting of d^4 state in octahedral interstice in cubic field and in tetragonal field with $c/a < 1$.

driving mechanism for the distortion, which proceeds until it is balanced by the elastic forces of the crystal. The corresponding splitting of the t_{2g} levels is also shown in Fig. 12. For a tetragonal distortion $c/a > 1$ the effect is reversed, with d_{z^2} stabilized. The above discussion depends only on the existence of an odd number of e_g electrons in the ground state, and so holds equally well for Cu^{2+} with the configuration $(t_{2g})^6(e_g)^3$ or for tetrahedral-site Fe^{2+} with configuration $(e_g)^3(t_{2g})^3$.

In first-order theory, $c/a < 1$, $c/a > 1$ or a linear combination of the two equally stabilize the octahedral site Mn^{3+} or Cu^{2+} ion, and it is necessary to consider higher-order elastic and electron-lattice coupling terms to predict the sign of the distortions. For an isolated octahedral complex, a $c/a > 1$ distortion should be more stable and has generally been found in spinels. For example, in Mn_3O_4 , $c/a \approx 1.16$; in CuFe_2O_4 , $c/a \approx 1.06$. However, tetrahedral-site Fe^{2+} ions in spinels⁴¹ and Mn^{3+} or Cu^{2+} ions in other crystals,⁴² such as the perovskite LaMnO_3 , may induce orthorhombic distortions.

Jahn-Teller distortions may also split a degeneracy in the t_{2g} orbitals; i.e., for d^1 , d^2 , d^6 and d^7 configurations in octahedral sites, and for d^3 , d^4 , d^8 and d^9 configurations in tetrahedral sites. However, since the t_{2g} orbitals are relatively non-bonding, smaller distortions can be expected with these configurations. Furthermore, these are just the configurations for which we noted that sizeable orbital effects are possible, so the sign of the Jahn-Teller distortion may be determined by spin-orbit coupling considerations.

Tetragonal distortion of a tetrahedral site $\text{Mn}^{3+}(e_g)^2(t_{2g})^2$ ion, for example, splits the triply degenerate t_{2g} levels into a doubly degenerate level containing the orbitals d_{xz} , d_{yz} and a non-degenerate d_{xy} orbital. Considerations similar to those carried out for the e_g splitting on an octahedral site show that the doubly degenerate level lies lowest for $c/a < 1$ as shown in Fig. 13. For $c/a > 1$, the situation is reversed, with d_{xy} stabilized relative to d_{xz} and d_{yz} .

It is clear that for $c/a < 1$, the orbital degeneracy together with any net azimuthal angular momentum is removed for a cation containing two or five t_{2g} electrons, while for one or four t_{2g} electrons a non-degenerate ground state requires $c/a > 1$. Thus a Jahn-Teller distortion that quenches the orbital angular momentum will give rise to an elongation of the unit cell for d^3 and d^8 ions on tetrahedral sites and a flattening for d^4 and d^9 ions. Distortions of this type have been observed in tetrahedrally coordinated $\text{Ni}^{2+}(d^8)$ and $\text{Cu}^{2+}(d^9)$ in the spinels NiCr_2O_4 ($c/a = 1.04$), NiRh_2O_4 ($c/a = 1.038$), CuCr_2O_4 ($c/a = 0.91$) and CuRh_2O_4 ($c/a = 0.91$). On

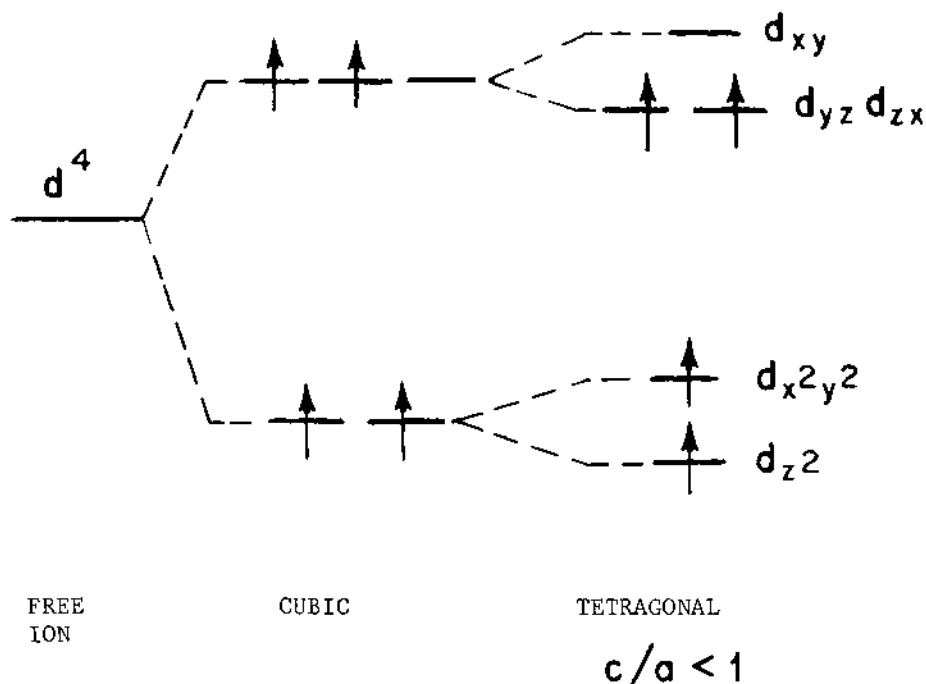


Fig. 13. Ligand-field splitting of d^4 state in tetrahedral interstices in cubic field and in tetragonal field with $c/a < 1$.

the other hand, distortions of the opposite sign preserve orbital degeneracy, which may be subsequently removed by spin-orbit splitting of the ψ_{+1} and ψ_{-1} levels.

Spin-Orbit Coupling. It was noted above that orbital effects are normally quenched by the ligand field, but that when the t_{2g} orbitals in cubic fields are partially filled, but not half-filled, this is not necessarily the case. For an unquenched orbital angular momentum, the spin-orbit energy $\lambda L \cdot S$ can be comparable with the Jahn-Teller stabilization energy. A Jahn-Teller distortion that conserves the spin-orbit coupling will only be effective below the Curie or Néel point, since the spins are randomly oriented above that temperature. However, once long range ordering occurs, the orbital states are coupled to the ordered spin arrangement via the spin-orbit interaction. Therefore the originally degenerate orbitals are no longer equally occupied; that is, the electron cloud no longer has cubic symmetry. This leads to a crystal distortion in order to lower the energy of the occupied orbitals. Since a

distortion of the type favored by a Jahn-Teller distortion removes the degeneracy completely, and hence quenches the orbital effect, the spin-orbit stabilized distortion will have the opposite sign. Hence there are two ways to distinguish between Jahn-Teller distortions with and without spin-orbit coupling: namely the sign of the distortion and the disappearance of spin-orbit distortions above the magnetic ordering temperature. Distortions due to spin-orbit effects have been observed in CoO and FeO.⁴²

Strong Ligand Fields ($\Delta_{CF} > \Delta_{ex}$). As noted in Table I, for elements in the second and third transition series (4d and 5d outer electrons), the ligand field splitting is greater than the exchange field splitting. This can also be true in oxides for trivalent cobalt and nickel, and tetravalent iron and cobalt. In that event Hund's rule no longer holds and the ions are said to be in a low-spin state. Low spin-states can occur for octahedral-site cations having d^n configuration with $4 < n < 7$ or for tetrahedral-site cations with $3 < n < 6$. Diamagnetic Co^{3+} ions with configuration $(t_{2g})^6(e_g)^0$ have been observed in a number of cobaltite spinels, such as Co_3O_4 . In the perovskite $LaCoO_3$, the low-spin state is only 0.08 eV below the high-spin state, which leads to surprising magnetic properties as a function of temperature.^{43,44} It is also possible to have a low-spin state to high-spin state transition at some temperature for which $\Delta_{cf} = \Delta_{ex}$. This apparently occurs in $MnAs_{0.94}P_{0.06}$, where there is evidence of an abrupt increase in Δ_{ex} on thermal expansion through a critical lattice parameter⁴⁵.

Magnetocrystalline Anisotropy. As we have seen, ligand field effects remove the spherical symmetry of the orbital angular momentum by lowering the energy for certain orbitals. Since the spins are coupled to the orbits via spin-orbit coupling and the orbits are coupled to the lattice, this leads to a dependence of the internal energy on the direction of the spontaneous magnetization relative to the lattice. For many $3d^n$ configurations the orbital effects are quenched, so that spin-orbit effects are zero to first order. However, higher order calculations recover some orbital moment by mixing with higher energy states. As a result, the cations with "quenched" orbitals have moderate anisotropy values. For unquenched orbital moments (i.e., Co^{2+} in octahedral sites) the anisotropy can be larger by two orders of magnitude.⁴⁶

The anisotropy energy reflects the symmetry of the crystal. In a cubic crystal, the anisotropy energy is expressed phenomenologically by

$$\mathcal{E}_b = K_1(\alpha_1^2 \alpha_2^2 + \alpha_2^2 \alpha_3^2 + \alpha_3^2 \alpha_1^2) + K_2 \alpha_1^2 \alpha_2^2 \alpha_3^2 \quad (61)$$

where α_1 , α_2 and α_3 are the direction cosines of the magnetization

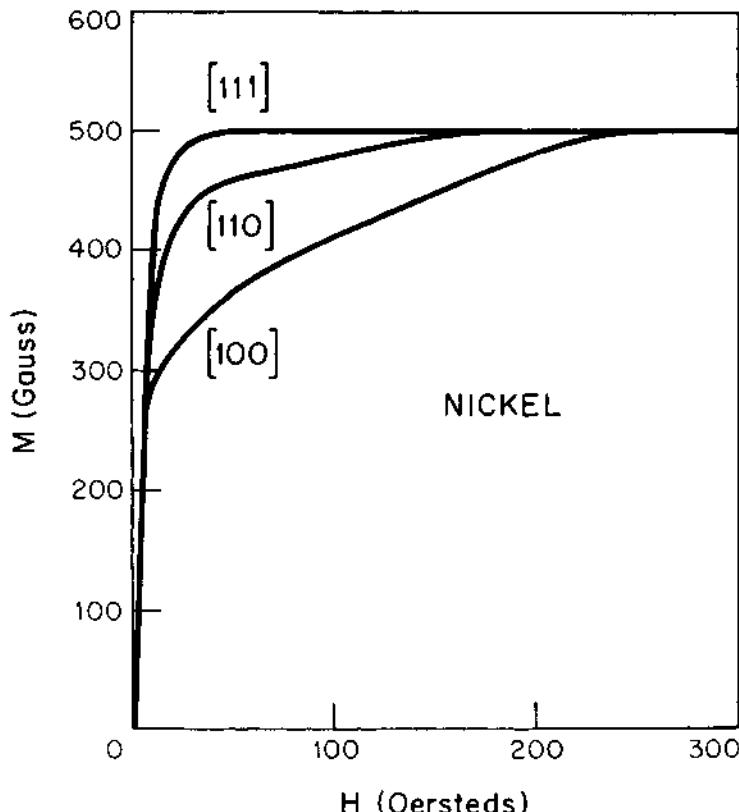


Fig. 14. Magnetization curves for single crystal of nickel along principal crystallographic axes at room temperature
[S. Kaya, Scientific Reports, Tôhoku Univ. 17, 639 (1928)].

direction relative to the cube edges of the lattice, and K_1 and K_2 are the first- and second-order anisotropy constants, respectively. Typically, values of $|K_1|$ are about 10^5 ergs/cm³. Strongly anisotropic magnetic materials (frequently referred to as hard magnetic materials) can have values of K_1 as high as 10^8 ergs/cc. The magnetic anisotropy of nickel, which at room temperature has the value $K_1 = -3.4 \times 10^4$ ergs/cm³ is illustrated in Fig. 14. In accord with Eqn. (61) the easy magnetization direction along the [111] corresponds to a negative value of K_1 .

VI. SUPEREXCHANGE INTERACTIONS

The term superechange is generally used with reference to magnetic insulators in which the magnetic interactions take place between localized electrons on neighboring cations. These inter-

actions may or may not be via an intervening anion. Our present understanding of this phenomenon has come primarily from the work of Anderson,^{15,47,48} Goodenough^{20,49,50} and Kanamori⁵¹.

Despite a large cationic separation, the magnetic interaction can be quite large if an intermediary anion is playing a major role in the effect. For example, the interaction in NiO that determines the magnetic ordering temperature T_N is a Ni-O-Ni interaction in which the Ni^{2+} ions are separated by $\sim 4 \text{ \AA}$, and here $T_N = 520^\circ\text{K}$. In order to gain insight into the role of the anion, we must consider the spatial relations between the cation and anion orbitals. As seen in Fig. 15a and b, the p_g orbital is orthogonal to the t_{2g} orbitals, but can have considerable overlap with an e_g orbital. Conversely, (Fig. 15c,d) the p_{π} orbital is orthogonal to the e_g orbitals but can overlap the t_{2g} orbitals. This overlap introduces covalent mixing between the anion and cation, thereby spreading the crystalline d orbitals out over the anions. This in turn makes possible a finite overlap of the crystalline d orbitals with the two Ni^{2+} ions on opposite sides of the anion.

Supereexchange occurs when the crystalline d orbitals of neighboring cations overlap each other, or more precisely have a finite matrix element $b_{ij} = (\psi_i(\mathbf{r}_i)|H|\psi_j(\mathbf{r}_j))$, which is called the transfer integral. Kramers⁵² first pointed out that this energy could couple electron transfer excited states into the ground state, and Anderson developed the formalism for this mechanism. Such an electron transfer involves a Coulomb energy expenditure U. For the case of two cations with half-filled e_g electrons overlapping an intermediate anion, as indicated in Fig. 16, the wave functions describing sites 1 and 2 become $\psi_1 = d_1 + \lambda p_g$ and $\psi_2 = d_2 + \lambda p_g$, where d_1 and d_2 are the unperturbed wave functions at sites 1 and 2 and λ represents the anion-cation covalent-mixing parameter. From the Pauli exclusion principle, the electron transfer can only occur if the electron spins at sites 1 and 2 are oriented anti-parallel. The energy for this antiferromagnetic arrangement has been shown by Anderson to be of the order $\Delta E \sim - (b_{12}^2/2U)$, where b_{12} is proportional to λ^2 .

In our hypothetical example we have not considered the effect of the t_{2g} orbitals. If each of the cations were d^3 ions at octahedral sites, the half-filled t_{2g} orbitals could have interacted via a p_{π} orbital at the anion. Their sign would have been the same as in our previous example, leading us to predict an antiferromagnetic d^3-d^3 interaction when the cation-anion-cation angle is 180° . For a d^5-d^5 180° interaction the effect would be the sum of the interactions considered above, so it too will be antiferromagnetic.

On the basis of similar, although frequently more sophisticated

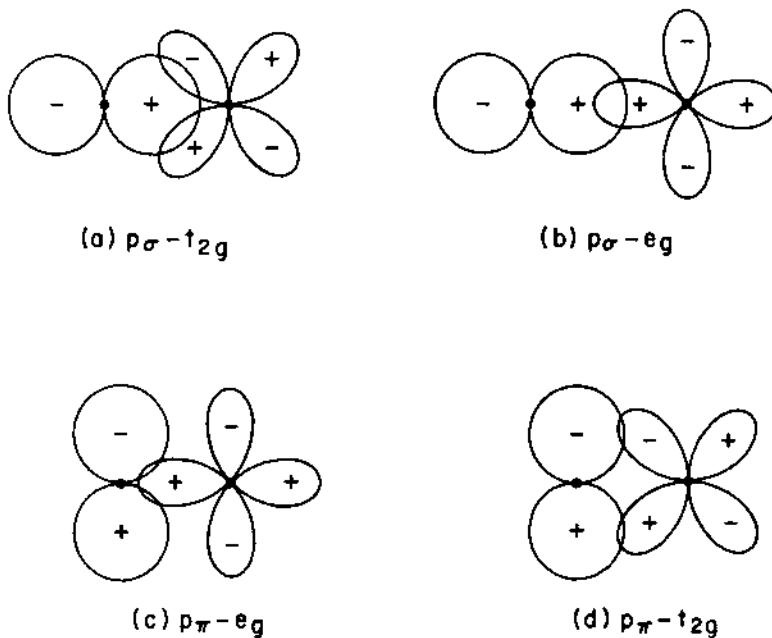
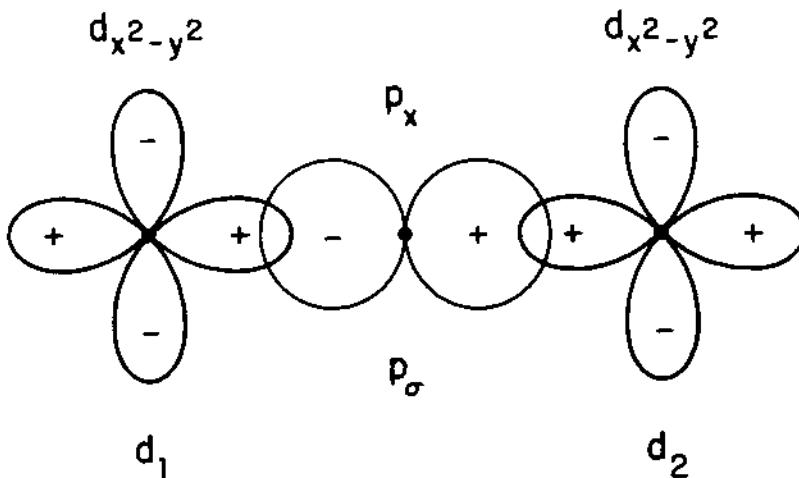


Fig. 15. Symmetry relationships between (a) p_{σ} and t_{2g} orbitals; (b) p_{σ} and e_g orbitals; (c) p_{π} and e_g orbitals; (d) p_{π} and t_{2g} orbitals.



$$\Psi_1 = d_1 + \lambda p_{\sigma}$$

$$\Psi_2 = d_2 + \lambda p_{\sigma}$$

Fig. 16. Overlap of $d_{x^2-y^2}$ cation orbitals with the p_{σ} orbital of intermediary anion.

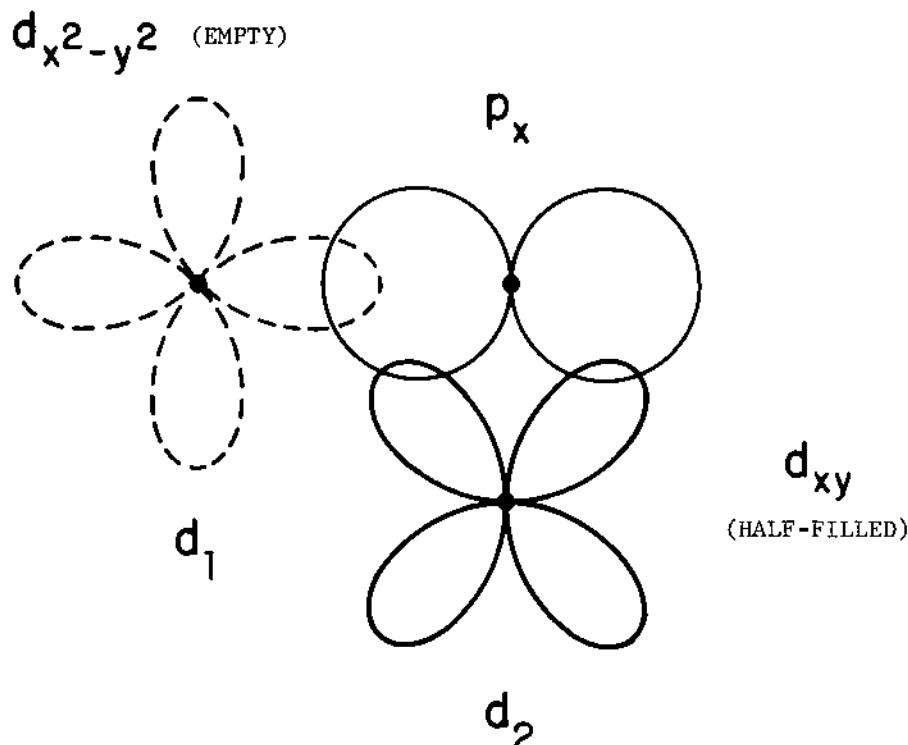


Fig. 17. 90° overlap of empty $d_{x^2-y^2}$ orbital (d_1) with half-filled d_{xy} orbital (d_2). Anion p_x orbital simultaneously forms σ -bond with d_1 -cation and π -bond with d_2 -cation.

considerations, Goodenough and Kanamori established the signs and qualitative magnitudes of the $d^n - \text{anion} - d^m$ superexchange interactions for various values of n and m (the Goodenough-Kanamori rules). For greater details, the reader is referred to the original articles as well as to Table XII and the accompanying discussion in Ref. 20.

They also considered the superexchange interaction expected for various 90° interactions. We will limit discussion here to the case of a d^3 -anion- d^3 interaction, with the cations at octahedral sites. In this case a single p orbital can simultaneously form a p_π bond with the half-filled t_{2g} orbitals of one cation and a p_σ bond with the empty e_g orbital of the other cation, as illustrated in Fig. 17. This cation-anion-cation interaction is ferromagnetic. In addition, there is the possibility of a direct overlap of the half-filled t_{2g}

orbitals of the two cations. This cation-cation interaction is antiferromagnetic⁵³ and is very sensitive to the cation-cation separation. Since there are both ferro- and antiferromagnetic interactions present, it is impossible to make an *a priori* prediction of the sign of the interaction. However, the direct cation-cation interaction is expected to be more sensitive to cation separation than the cation-anion-cation interaction, so one would predict that increasing the separation of the cations should lead to a relatively larger decrease of the antiferromagnetic interaction.

All the nearest-neighbor interactions are of this type in the normal spinels $M[\text{Cr}_2]\text{X}_4$, where M is a nonmagnetic cation and X represents the anion. And indeed, as shown in Fig. 18, for the

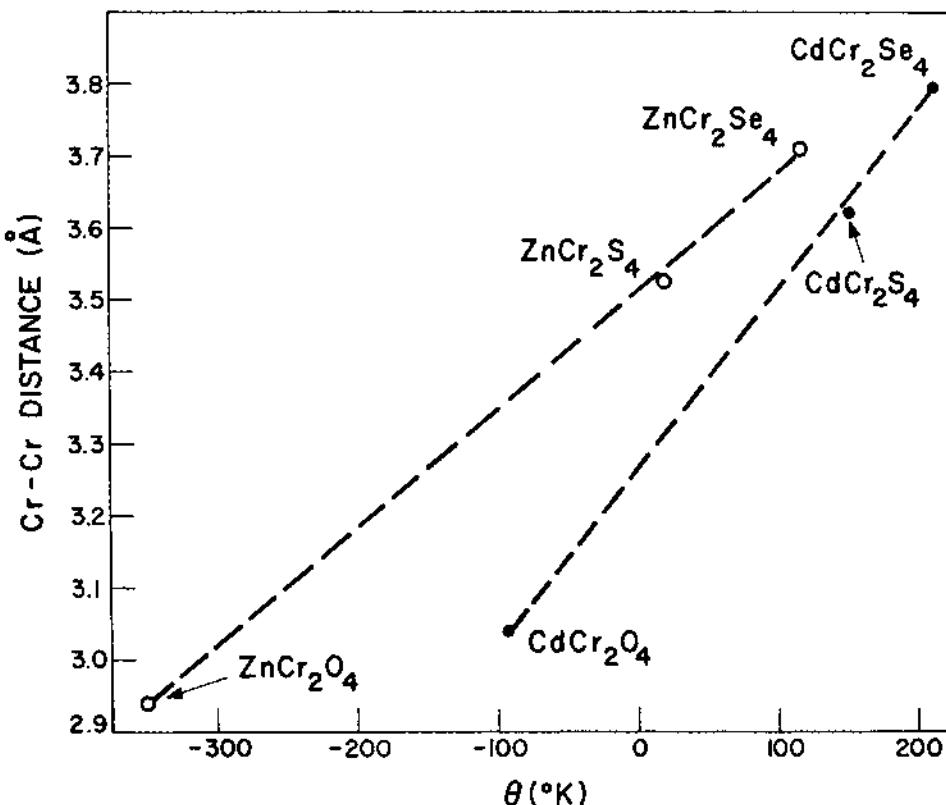


Fig. 18. Paramagnetic Curie temperatures of $M\text{Cr}_2\text{X}_4$ compounds as functions of Cr^{3+} - Cr^{3+} separation (Ref. 54).

systems wherein M = Zn or Cd and X = O, S or Se, it is found that the paramagnetic Curie temperatures, which are directly proportional to the sum of the magnetic interactions, become more positive with increasing cation-cation separation⁵⁴. Of the materials included in Fig. 18, only CdCr₂S₄ and CdCr₂Se₄ are ferromagnetic. The fact that ZnCr₂S₄ and ZnCr₂Se₄ have a positive value of θ_p indicates that the nearest-neighbor interactions are ferromagnetic in these materials as well, while their low temperature antiferromagnetic ordering can be explained by consideration of distant-neighbor magnetic interactions^{55,56}.

There are a few general statements that can be made regarding the relative magnitude of superexchange effects on the basis of symmetry and chemical considerations:

1) Since the overlap integral b for d_{z²}-p _{σ} orbitals are generally 25-50% larger than for d_{xy}-p _{π} orbitals, and the interaction strength is proportional to b²/U, the individual σ -type interactions should be about 1.5 to 2 times as great as π -type interactions.

2) Covalency effects between the cation and the anion should tend to increase the electron transfer effect ($b \sim \lambda^2$). This would lead one to expect that for ions with the same outer electron configuration, the interaction will be greatest for the highest valency cation. As pointed out by Anderson, this is borne out in comparisons of Mn²⁺ and Fe³⁺ materials, which show the Fe³⁺ exchange to be considerably larger than Mn²⁺ exchange. Covalency also increases on going to the right in the periodic table, which accounts for the increasing T_N on going from MnO to FeO, CoO, NiO. Fe³⁺ is also unique in being a common trivalent ion with σ interactions. Superexchange effects involving trivalent cations, such as Cr³⁺, normally involve the smaller π interactions. Thus oxides with high Curie temperatures contain Fe³⁺ and/or Ni²⁺ ions. On the same basis one would expect to obtain increasing covalency on going along the series O²⁻ → S²⁻ → Se²⁻. This is found to be the case for the compounds MnO, MnS and MnSe, which have Curie points of 122°K, 130°K, 173°K respectively. It is also consistent with the results shown in Fig. 18.

Double Exchange. In compounds containing magnetic cations of mixed valency it is possible to obtain energy stabilization by double exchange which occurs in the perovskite system (Ca_xLa_{1-x})(Mn_{1-x}Mn_x⁴⁺)O₃. Although the end members of the system are antiferromagnetic insulators, Jonker and Van Santen⁵⁷ found that a non-zero spontaneous magnetization exists between 0.1 < x < 0.5, which correlates with the region of high electrical conductivity. Zener⁵⁸ explained this by noting that the extra electron on the Mn³⁺ ion can travel to an Mn⁴⁺ ion. In this case, the Mn³⁺ ion

becomes an Mn⁴⁺ ion and the Mn⁴⁺ ion becomes the Mn³⁺ ion, so the final state is degenerate with the initial state. Unlike the superexchange effect, in which electron transfer involved the creation of an excited state at energy U above the ground state, double exchange can be treated by first-order perturbation theory. The presence of these mobile electrons are also the source of the high conductivity. Zener further assumed strong intra-atomic exchange, so the mobile electron is coupled parallel to the ionic spin, and that the mobile electrons do not change their spin orientation during the transfer process. Therefore the transfer can occur between two ions only if these ions are not antiparallel. Since the spin coupling reduces the ground state energy by permitting the mobile carriers to participate in the binding, double exchange favors ferromagnetic coupling.

This exchange effect has been investigated in considerable detail by Anderson and Hasegawa⁵⁹ and by deGennes⁶⁰, who noted that the transformation properties of the spin operators leads to an angular dependence of the double exchange energy of the form

$$E_d = -b \cos(\theta/2) . \quad (62)$$

This contrasts with superexchange, where second-order perturbation theory gives

$$E_s \sim \frac{b^2 \sin^2(\theta/2)}{U} \sim \frac{b^2}{U} \cos \theta + \text{const.} \quad (63)$$

which has the $\cos \theta$ dependence obtained in the usual Heisenberg exchange. Taking the transfer integral b as positive, Eqn. (62) always leads to a ferromagnetic interaction. However, the double exchange will normally be competing with an antiferromagnetic superexchange interaction of comparable magnitude. For a material with a mobile electron concentration c and Z bonds per cation participating in both double and superexchange effects, the energy is then⁶¹

$$E = E_s + E_d = NZJS^2 \cos \theta - NcZbc \cos(\theta/2) . \quad (64)$$

Minimization of Eqn. (63) yields

$$\cos(\theta/2) = \frac{bc}{4JS^2} \quad (65)$$

which defines the canting angle θ between successive ions. (For $bc/4JS^2 > 1$, a ferromagnetic arrangement is lowest.) This canting effect is quite different from the helical spin configurations previously considered, since those were caused by competing interactions amongst different neighbors. In general, double exchange will not be expected to cause spiral configurations. DeGennes pointed out that the work of Wollan and Koehler⁶¹ in the

system $(Ca_xLa_{1-x})(Mn_{1-x}^{3+}Mn_x^{4+})O_3$ gave results in accord with Eqn. (65) for $0 < x < 0.25$.

VII. COLLECTIVE d-ELECTRON MODEL

In the purely localized-electron model, the superexchange energy is $\sim b_{ij}^2/U \ll 1$, and it was noted that an electron transfer costs an amount of energy U , where U is the average Coulomb repulsion energy between electrons on the same atom. This is the energy that keeps the electrons localized, and the transfer involves a valence change as indicated in Fig. 19a.

However, if we modify the localized model by the introduction of bands, the electrons then gain kinetic energy in moving throughout the crystal. Taking the localized electron energy as the average band-state energy and placing the electron at the bottom of the band leads to an energy gain $\frac{1}{2}W_b$, where W_b is the band width. In the tight-binding approximation $W_b \sim b_{ij}$. For the case of wide bands the energy gain is greater than the Coulombic (or correlation) effect, and we have the metallic state. This is illustrated in Fig. 19d. In this case there is no spontaneous moment, and the material exhibits a small, temperature-independent paramagnetism called Pauli paramagnetism.

Pauli paramagnetism arises from the fact that only the small fraction of electrons which are within the order of kT of the top of the band can contribute to the paramagnetic susceptibility. Taking the fraction of electrons which can contribute equal to T/T_F , then from Eqn. (7)

$$\chi = \frac{\frac{N\mu^2}{3kT}}{\frac{T}{T_F}} = \frac{\frac{N\mu^2}{3kT_F}}{1} \quad (66)$$

where T_F is called the Fermi temperature (see Ref. 1, pp. 249-261) and is of the order $10^4 - 10^5$ K. The temperature dependence of susceptibility has been cancelled out, and the remaining susceptibility is about 2 orders of magnitude smaller than that of materials for which Curie's law holds.

The broad band limit represents the opposite extreme from the localized model previously considered. From here on we will address ourselves to considerations of magnetic behavior in the intermediate region between these limits. Our approach to the problem will closely follow that given in a number of articles by Goodenough.⁶²⁻⁶⁷ However, our discussion will be limited to the case of a single d electron per relevant d orbital ($n_d = 1$), which means there are an integral number of d-electrons for each transition-metal cation on equivalent lattice sites. The interaction in this case must be antiferromagnetic, and if we limit ourselves to the case of a single

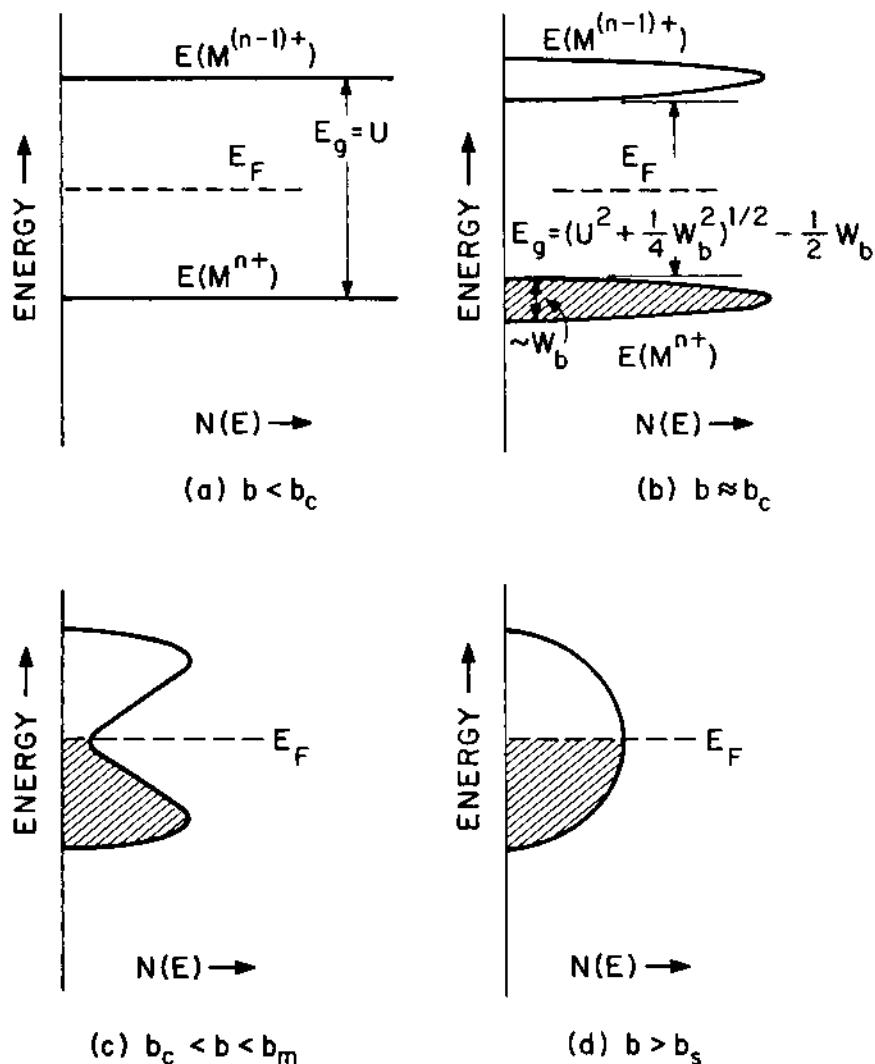


Fig. 19. Schematic representation of density of states $N(\epsilon)$ vs. energy ϵ for various values of b (Ref. 64).

exchange interaction J_{ij} arising from an overlap integral b , the Néel temperature

$$T_N \propto J_{ij} \propto \frac{b^2}{U} . \quad (67)$$

Now let us consider what happens as b increases and we enter the band regime. Increasing b corresponds to a broadening of the

band, leading to a decrease in the value of U , as indicated in Fig. 19b. Formally this reduction in U arises as a result of the screening effect, as

$$U = (|\phi_m(1)|^2 (e^2/r_{12}) \exp^{-\xi r_{12}} |\phi_m(2)|^2) \quad (68)$$

where ξ is the screening parameter which increases with increasing b . Therefore b^2/U is increasing, and since the superexchange perturbation calculation is based upon the assumption that $b^2/U \ll 1$, it is apparent that eventually a critical value $b = b_c$ will be reached beyond which this calculation (and therefore Eqn. 67) is no longer valid. However, for $0 < b < b_c$, Eqn. (67) holds and T_N increases with increasing b , as shown in Fig. 20.

For $b > b_c$, the two energy bands will continue to broaden, eventually overlapping at some new critical value $b = b_g$, as shown in Fig. 19c. Up to this point the electron correlation between

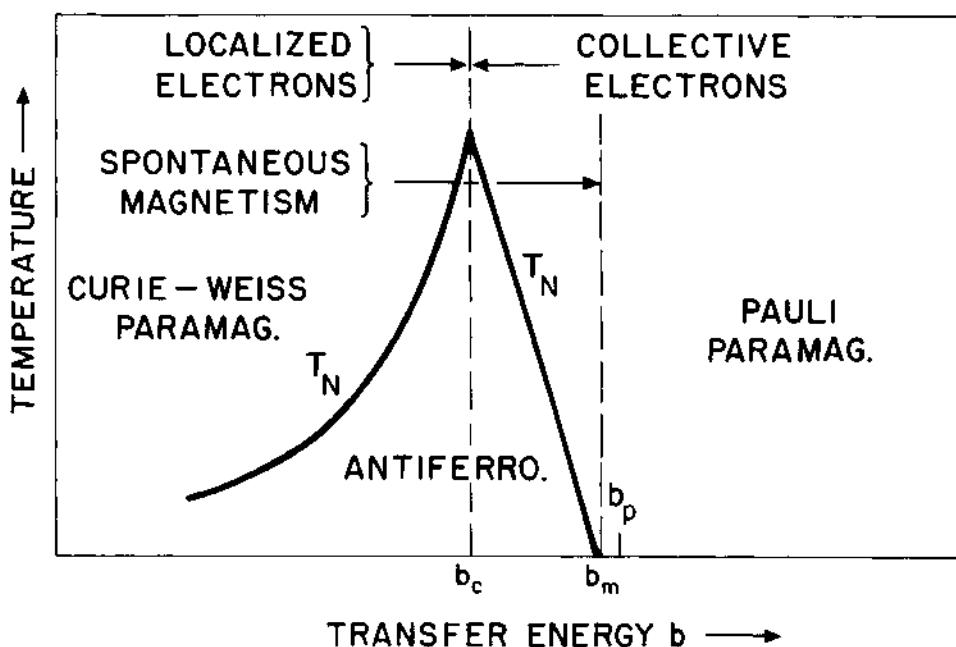


Fig. 20. Conceptual T-b diagram for $n_1 = 1$ (Ref. 65).

narrow band electrons should still be great enough to ensure a deep minimum in the density of states at the Fermi energy E_F . Since U/w_p decreases exponentially with increasing b , Mott⁶⁸ argued that the transition from a localized to a collective electron state should be extremely sharp.

On further increasing the transfer integral b , we reach the broad band limit discussed above. At and above this value of b , (b_p in Fig. 20) the material is Pauli paramagnetic with metallic conductivity ($b_p > b_g$). At some slightly lower value of $b = b_m \leq b_p$ long range magnetic order disappears. Therefore, the collective electron regime in which antiferromagnetism can occur is limited to the region $b_c < b < b_m$. Within this region the spontaneous atomic moment decreases with increasing b , and since T_N goes to zero at $b = b_m$, T_N must also decrease with increasing b . Since T_N increases with increasing b for $0 < b < b_c$, the maximum value of T_N occurs at $b = b_c$, as indicated in Fig. 20.

The reduction of T_N with increasing b for $b_c < b < b_m$ can be rationalized by noting that antiferromagnetic ordering between two sites leads to additional band splitting, reflecting the new translational symmetry. This additional splitting serves to stabilize the occupied states at the expense of the unoccupied states, and this stabilization can be considered the driving mechanism for the antiferromagnetism. However, this stabilization is decreased by thermal excitation of electrons into the conducting states, so that T_N will decrease with a decreasing energy gap, and hence with increasing b .

The paramagnetic susceptibility follows a Curie-Weiss law for $b < b_c$, and is temperature independent (Pauli paramagnetic) for $b > b_p$. In between, the susceptibility is in an ill-defined "mixed state" which is temperature dependent but with a magnitude and temperature dependence which decreases with increasing b . The important point is that the Curie-Weiss law is not necessarily valid in this region, so that any attempt to obtain a value of μ_{eff} from an inverse susceptibility versus temperature curve would give too large a value, and θ_p will not be related to the sum of the magnetic interactions.

It is of considerable interest to investigate the validity of the above ideas in a physical system, and in particular to establish the width of the transition region $b_c < b < b_m$. However, since there is no means of directly measuring the overlap integral b , it is necessary to do this indirectly by considering the parameters which effect b in a known manner and by investigating the variation of magnetic properties as a function of these parameters.

A large overlap integral may be due either to large, direct

cation-cation interactions or to strong cation-anion-cation interactions. In the case of direct interactions, $b = b(R)$ is strongly sensitive to atomic separation R . Therefore, an abrupt collective-electron to localized-electron transition at $b = b_c$ implies a critical separation R_c for such a transition⁶⁸. Goodenough has given an empirical expression for R_c in oxides containing 3d electrons⁶³.

$$R_c^{3d}(\text{M}^{m+}-\text{oxides}) = \{ 3.20 - .05m - .03(Z - Z_{Ti}) - .04S_i(S_i + 1) \} \text{ \AA} \quad (69)$$

where $(Z - Z_{Ti})$ is the increase in the cation atomic number over that of Ti and S_i is the net cation spin assuming localized electrons. The last three terms represent the radial contraction (and hence the variation in b) with valence, nuclear charge and intra-atomic exchange stabilization.

In the case of indirect cation-anion-cation interactions, from Eqns.(62) and (63), e_g-p_σ and $t_{2g}-p_\pi$ interactions lead respectively to overlap integral values

$$b_\sigma = N_\sigma^2 \lambda_\sigma^2 \quad \text{and} \quad b_\pi = N_\pi^2 \lambda_\pi^2 \quad (70)$$

where N_σ and N_π are normalization constants, while λ_σ and λ_π represent the anion-cation overlap.

We will investigate the relevance of the above concepts as applied to various 3d-cation monoxides with rock salt (B1) structure and to perovskites. These are instructive systems because both have isostructural members that contain, respectively, collective and localized electrons. In addition, one of these systems (the monoxides) offers an example wherein the direct cation-cation interaction can be of primary importance, while in perovskites the cation-anion-cation interaction plays the dominant role.

Rock Salt Structure. The physical properties of the monoxides of the first transition series are given in Table III. From the results of Table III it appears that MnO, FeO, CoO and NiO are describable by a localized-electron model, with a magnetic order consistent with an antiferromagnetic 180° cation-anion-cation superexchange interaction between cations with half-filled e_g orbitals. The expectation that b_σ increases with increasing Z is consistent with the observed increase in T_N with increasing Z , and indicates that $b_\sigma < b_c$ in TiO and VO, which have smaller Z values. Since it is clear from Table III that $b > b_m$ in TiO, the large value of b is presumably due to direct cation-cation interactions. This is further corroborated by the fact that $R < R_c$ in TiO.

Since $R < R_c$ in VO as well, one can also anticipate collective

Table III

<u>Structure</u>	<u>R(Å)</u>	<u>R_c(Å)</u>	<u>Transport</u>	<u>Magnetic</u>
TiO B1, 15% vac.	2.94	3.02	Metal	Pauli
VO B1, 15% vac	2.89	2.92	high cond.	Weak T- dependence
CrO Unstable	-	2.80	-	-
MnO B1	3.14	2.66	S.C.	Curie-Weiss, T _N =122°K
FeO B1	3.03	2.95	S.C.	C-W, 198°K
CoO B1	3.01	2.87	S.C.	C-W, 291°K
NiO B1	2.95	2.77	S.C.	C-W, 440°K
CuO Monoclinic	2.89	2.65	S.C.	C-W, 230°K

electron behavior in this material. However, the difference between R and R_c is much smaller, so it is of interest to determine the value of b relative to b_c and b_m, and to establish the presence of a minimum or an energy gap in the energy vs. N(ϵ) curve. To assist in this study, the materials VO_x have also been investigated, since the cubic rock-salt structure is stable over the region 0.8 < x < 1.30.

Conductivity measurements indicate VO to have a low resistivity ($\rho \sim 10^{-3} \Omega \text{-cm}$), but with a negative temperature coefficient⁶⁹, and an activation energy of the order of $5 \times 10^{-3} \text{ eV}$ ⁷⁰, that increases with x, particularly for x > 1.0. This indicates the presence of an energy gap E_g, so b < b_g < b_p in VO, the gap presumably arising from a splitting of the half-filled t_{2g} orbitals into narrow bands by electron correlations. It is assumed that the anion vacancies trap electrons in e_g or S states while the cation vacancies trap holes in t_{2g} states, so that the t_{2g} bands are half-filled for all x. The existence of two bands is further indicated by a change in the sign of the Seebeck coefficient in VO_x from negative for x < 1.02 to positive for x > 1.02. There is some disagreement between the magnetic data of Kawano *et al.*⁶⁹ and that of Banus *et al.* in the system VO_x⁷⁰. As seen in Fig. 21a, the samples of Kawano *et al.* indicate the presence of a Néel point for x > 1.14, while there is no such indication in the results of Banus *et al.* which are shown in Fig. 21b. Despite these differences there is a large area of agreement between the two sets of results. Both indicate increasing susceptibility with increasing x, and for x < 1 both show a slightly temperature sensitive susceptibility with no sign of long range order at the lowest temperatures.

The existence of an energy gap in VO indicates that b < b_g; the fact that VO is neither metallic nor truly Pauli paramagnetic requires that b < b_p; and finally, the lack of long range magnetic

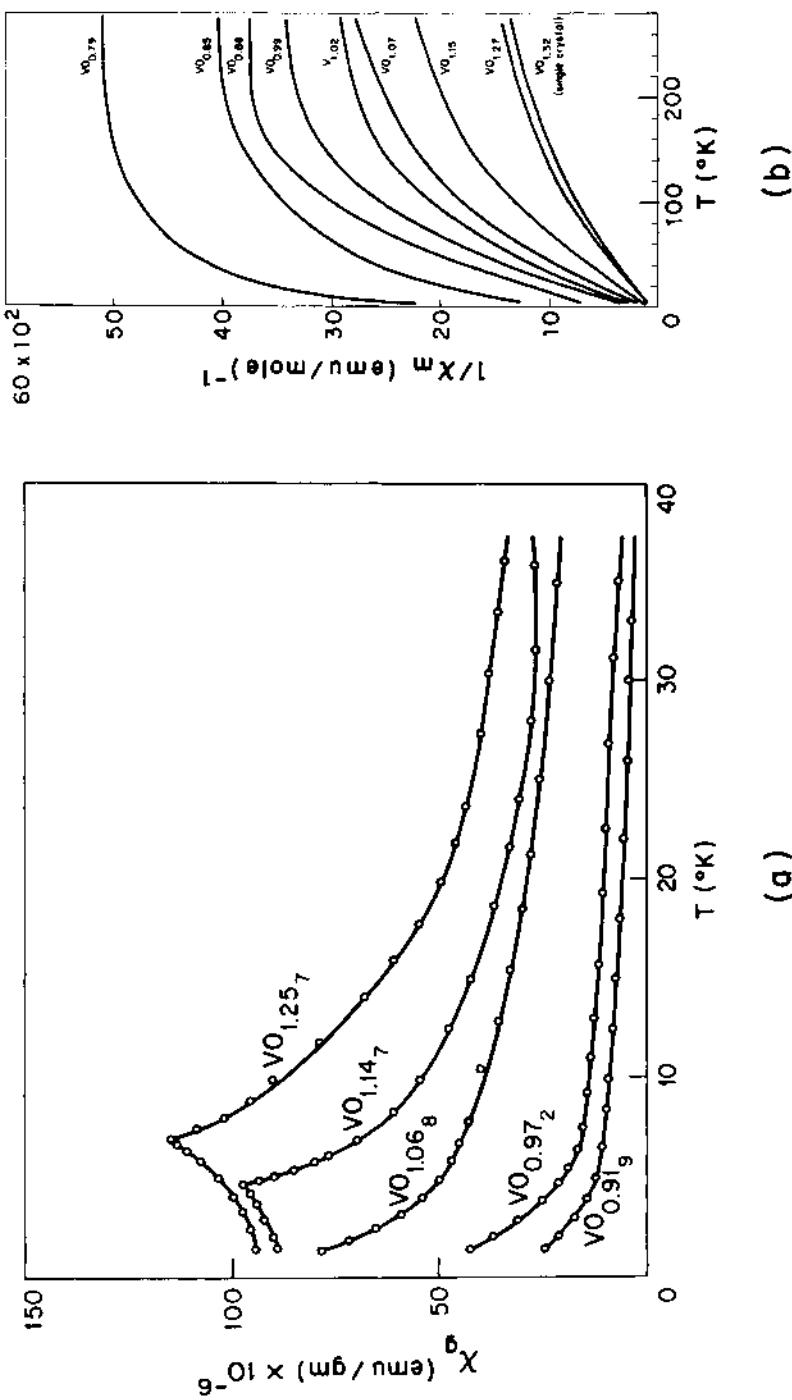


FIG. 21. Variation of susceptibility with temperature in VO_x as a function of x . (a) Ref. 69; (b) Ref. 70.

order requires that $b > b_m$. Therefore we can conclude that in VO, $b_m < b < b_g < b_p$. The striking increase in χ with x indicates increasing exchange enhancement typical of a value of b only slightly greater than b_m . It is therefore reasonable to assume that Δb is very small in VO, where $\Delta b = b - b_m$.

Perovskites. The perovskite structure is found in a large number of compounds with the chemical formula ABX_3 . The ideal cubic perovskite consists of a simple cubic B cation lattice structure with a large A cation at the body center and the anions at the center of each of the cube edges, as illustrated in Fig. 22. In many perovskites there are small orthorhombic or rhombohedral distortions from the ideal cubic structure, which frequently leads to a parasitic ferromagnetism superimposed on the basic antiferromagnetism.

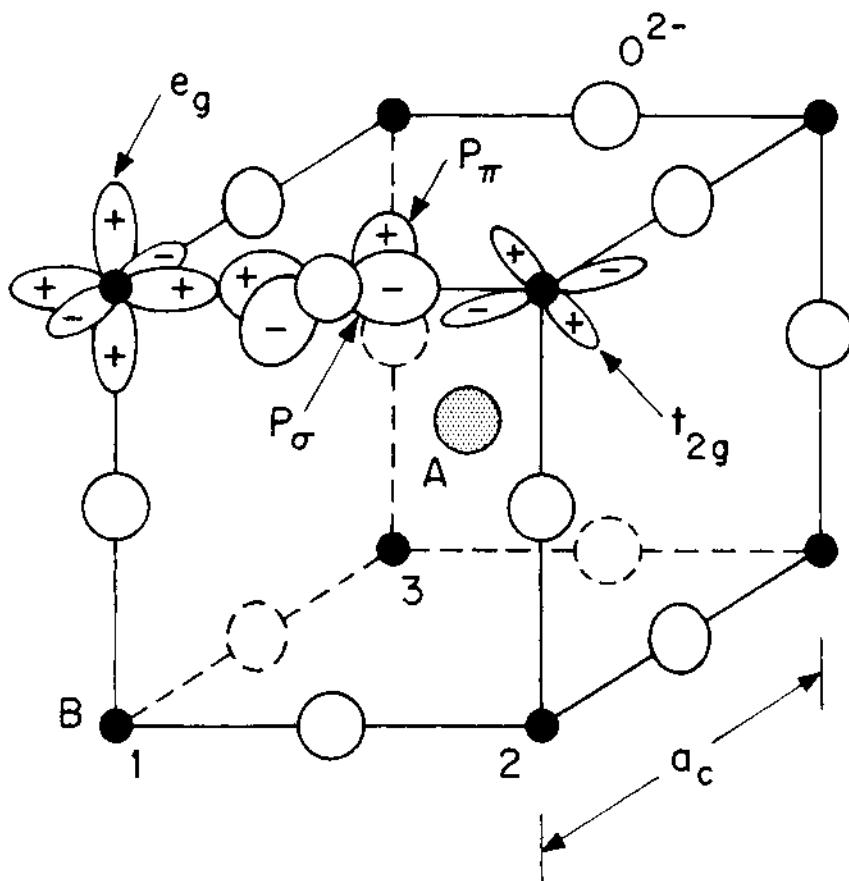


Fig. 22. Cubic perovskite structure, showing cation t_{2g} and e_g orbitals and anion p_π orbitals (Ref. 66).

Within this structure there are materials that are antiferromagnetic insulators, such as LaMnO_3 , and others that are metallic and Pauli paramagnetic, such as SrCrO_3 . Since the transition-metal B-cations are separated across the cube face by $R \approx 5.5 \text{ \AA}$, direct cation-cation interaction can be assumed small. Furthermore the large A ions are usually strongly electropositive, so that where metallic conductivity occurs, it is due to a b_{ij} arising from large cation-anion-cation interactions along the cube edge, of the types indicated in Fig. 22. That this is the case was demonstrated experimentally⁷¹ by the synthesis of a single crystal of ReO_3 , which has a structure like that of the perovskite with the A-site cation missing. It was found to be metallic, indicating that the A-site cation is not necessary for metallic conductivity. The ordered perovskite $\text{Sr}_2\text{Mg}^{2+}\text{Re}^{6+}\text{O}_6$, however, is semiconducting. Since the B-site ordering in this material disrupts the Re-O-Re interaction but not the direct Re-Re interactions, this agrees with the conclusion that the cation-anion-cation interaction along the cube edge is the major source of the large transfer integral.

This is a good system for study because of its simplicity relative to other ionic compounds. With the only important magnetic interaction expected to be the one along the cube edge, competing magnetic interactions do not arise as a complicating factor. In addition, a large number of perovskites can be prepared, and their magnetic properties compared. A number of oxides with perovskite structure that have been studied thus far are shown in Fig. 23. The net spin at the top of each column is the formal value for localized electrons, and the compounds in the last three columns on the right have cations in the low-spin state. The overlap integral is assumed to increase on going up any column. While not quantitative, it is based on the previously stated assumption that b increases with increasing covalence. Thus b is assumed to increase with increasing valence, and for the same valence state $b(5d) > b(4d) > b(3d)$. The A-site cation also effects the interaction, albeit indirectly, since the more ionic A-cation competes less strongly for the p_π orbital, resulting in a larger b_π^T value. Therefore we can expect, for example, $b_\pi^T(\text{Ba}) > b_\pi^T(\text{Ca})$. Finally, with occupied e_g and t_{2g} orbitals simultaneously present, $b_\sigma > b_\pi^T$.

All the materials within the dotted line $b_\sigma \approx b_c$ can be described by a localized model. All the compounds in the region $b_\sigma > b_c > b_\pi^T$ have empty e_g orbitals, so only π bonding need be considered. As $b_c > b_\pi^T$, these compounds also exhibit localized-electron behavior. Since there is not yet any clear cut experimental differentiation between the behavior of these two localized-electron regions, the location of the line $b_\sigma \approx b_c$, as given in Fig. 23, must be considered tentative. LaCoO_3 is shown twice since, as mentioned previously, this material undergoes a transition from the diamagnetic low-spin state at $T = 0^\circ\text{K}$ to high-spin state

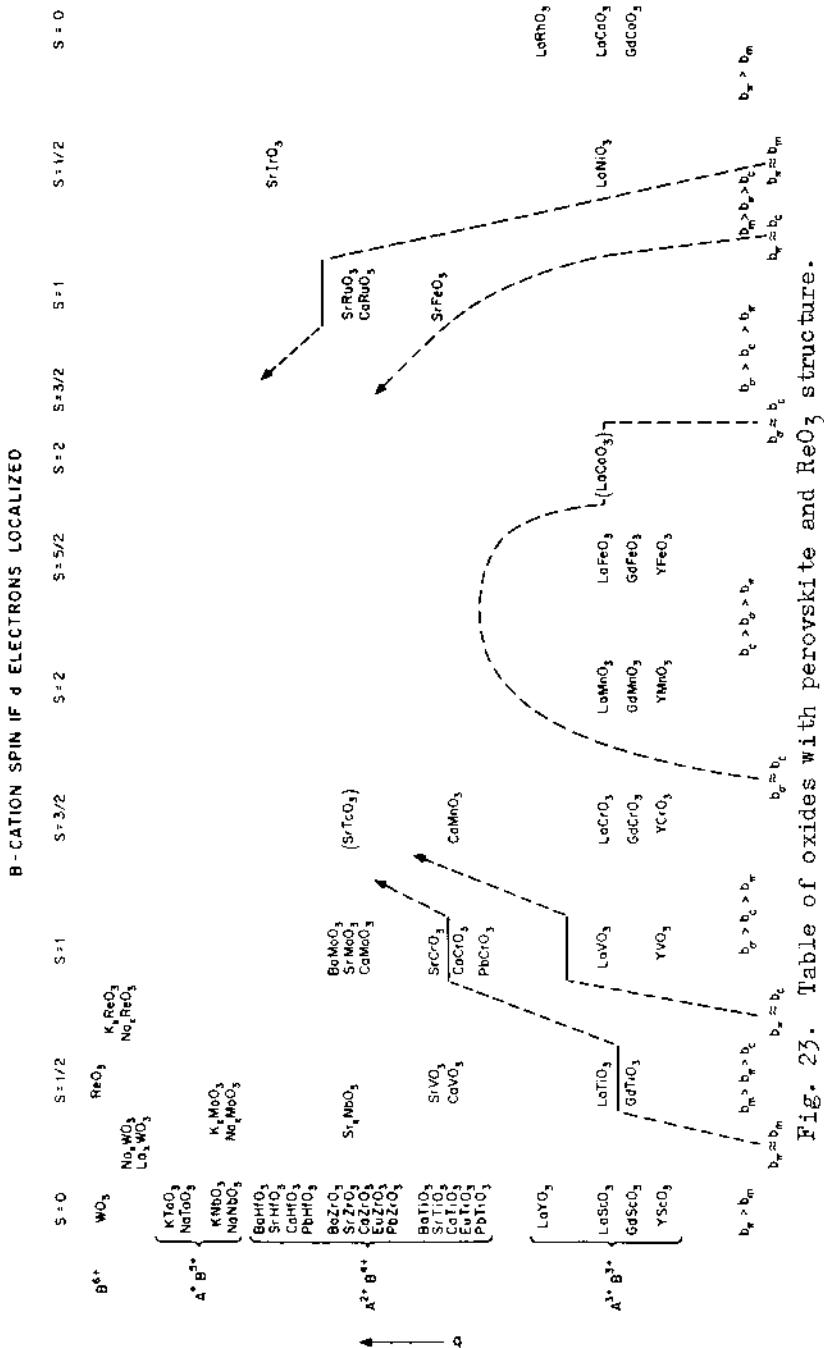


Fig. 23. Table of oxides with perovskite and ReO_3 structure.

at $T \sim 400\text{--}600^\circ\text{K}$.

The narrow region $b_m > b_{\pi} > b_c$ is more definitively established on the basis of observable changes in magnetic, electrical and crystallographic behavior. For example, in the third column, labelled $S = 1$, the top four compounds A MoO_3 ($\text{A} = \text{Ba}, \text{Sr}, \text{Ca}$) and SrCrO_3 are metallic and Pauli paramagnetic, ensuring that $b_{\pi} > b_p > b_m$. Continuing down the column, the remaining four compounds are antiferromagnetic with Néel points T_N of 90°K (CaCrO_3), 240°K (PbCrO_3), 137°K (LaVO_3) and 110°K (YVO_3). According to Fig. 20, these results imply collective-electron magnetism in CaCrO_3 ⁷², with $b_c < b_{\pi} < b_m$ and localized-electron magnetism in LaVO_3 and YVO_3 with $b_{\pi} < b_c$. PbCrO_3 remains cubic down to lowest temperatures, which is consistent with collective-electrons, while a spin-orbit distortion is predicted for localized moments at or below T_N ⁷³. Such a distortion is observed in LaVO_3 , confirming the localized model for that material.

It is of interest to note that a change of the A-site ion amongst $\text{A}^{2+}\text{CrO}_3$ compounds can introduce a shift from Pauli paramagnetism to antiferromagnetism with a relatively high value of T_N . Since the effect of the A-site ion on the magnetic interaction is indirect, involving only additional competition for the p_{π} orbitals, it is apparent that the region of collective electrons with long-range ordered magnetism ($b_c < b_{\pi} < b_m$) must be quite narrow.

In general, Fig. 23 indicates that the various areas are continuous; that high-spin values tend to stabilize the localized-electron state; that increased covalency of the transition-metal cation and of the anion tend to stabilize the collective-electron state; and finally, that increasingly ionic A-site cations also tend to stabilize the collective-electron state. Taken together, these results strongly support the principal assumptions that have been made.

An unfortunate aspect of the narrowness of the region $b_c < b_{\pi} < b_m$ is that relatively few compounds have a value of b_{π} which falls within this interval. The paucity of suitable materials makes a detailed study of magnetic properties within this region difficult. One approach to overcome this difficulty has been to use mixed chemical systems to control the degree of change effected. This approach has the disadvantage that chemical substitution can change more than one variable, making interpretation difficult. For example, an attempt to study the effect of lattice size on the Néel temperature of CaMnO_3 led MacChesney *et al.*⁷⁴ to substitute varying amounts of strontium for calcium. They found strontium substitution increased both the lattice cell length, a_0 , and T_N . An interpretation on the basis of size alone would lead to the conclusion $dT_N/da_0 > 0$. However, direct measurement utilizing pressure has established that

$dT_N/da_0 < 0$.⁷⁵ It is therefore apparent that the observed rise in T_N with increased strontium is due to the ionic character of strontium, and that this chemical factor is able to over-ride the size effect.

The use of pressure represents an alternate approach which is capable of studying the size dependence of T_N directly. Since the effect of increasing pressure in perovskites is to increase the value of b_{π} by decreasing the lattice dimensions, then according to Fig. 20, $dT_N/dp > 0$ for $b_{\pi} > b_c$ and $dT_N/dp < 0$ for $b_c < b_{\pi} < b_m$.⁷⁶ It has been found that $dT_N/dp < 0$ in CaCrO_3 , confirming the collective electron behavior of that material, while $dT_N/dp > 0$ in YCrO_3 as well as CaMnO_3 , in accord with their being localized-electron antiferromagnets.⁷⁵

The magnitude of the change in T_N at readily achievable hydrostatic pressures (~ 10 kbars) are usually quite small, dT_N/dp values of $\pm 0.5^\circ/\text{kbar}$ being typical. It is hoped that the simultaneous employment of both approaches will permit the use of chemical substitutions to make significant changes in the value of b_{π} within the region $b_c < b_{\pi} < b_m$, and then to separate the relative roles of chemical and size effects by means of pressure experiments.

I should like to express my thanks to J.B. Goodenough and K. Dwight for their help and advice in the course of preparing these lectures.

REFERENCES

1. C. Kittel, *Introduction to Solid State Physics*, John Wiley and Sons, New York, 1956.
2. P. Langevin, *J. de Physique* 4, 678 (1905).
3. J.H. van Leeuwen, *J. de Physique* (6) 2, 361 (1921).
4. P.A.M. Dirac, *Proc. Roy. Soc.* 117a, 610; 118a, 351 (1928).
5. J.H. Van Vleck, *The Theory of Electric and Magnetic Susceptibilities*, Oxford University Press, London, 1932.
6. L. Néel, *Ann. Phys.* [10] 18, 5 (1932).
7. L. Néel, *Ann. de Physique* 3, 167 (1948).
8. A. Yoshimori, *J. Phys. Soc. Japan* 14, 807 (1959).

9. T.A. Kaplan, Phys. Rev. 116, 888 (1959).
10. J. Villain, J. Phys. Chem. Solids 11, 303 (1959).
11. P. Weiss, J. de Physique 6, 667 (1907).
12. W. Heisenberg, Zeits. f. Physik, 38, 411 (1926).
13. P.A.M. Dirac, Proc. Roy. Soc. 112a, 661 (1926).
14. D.H. Martin, Magnetism in Solids, M.I.T. Press, Cambridge, Massachusetts, 1967.
15. P.W. Anderson, Theory of Magnetic Exchange Interactions, Exchange in Insulators and Semiconductors in Solid State Physics, Vol. 14, F. Seitz and D. Turnbull (eds.) Academic Press, New York, 1963.
16. J.S. Smart, Effective Field Theories of Magnetism, W.B. Saunders Company, Philadelphia, Pennsylvania, 1966.
17. H.E. Stanley, Critical Phenomena, W.A. Benjamin, New York, to be published.
18. F. Keffer and R. Loudon, J. Appl. Phys. 32, 2S (1961).
19. J.H. Van Vleck, J. Chem. Phys. 2, 85 (1941).
20. J.B. Goodenough, Magnetism and the Chemical Bond, Interscience Publishers, New York, 1963.
21. J. Smit and H.P.J. Wijn, Ferrites, John Wiley and Sons, New York, 1963.
22. E.W. Gorter, Philips Res. Repts. 9, 295, 321 (1954).
23. Y. Yafet and C. Kittel, Phys. Rev. 87, 290 (1952).
24. E. Prince, Acta Cryst. 10, 554 (1957).
25. R. Plumier, C.R. Acad. Sci. 268, 1549 (1969).
26. U. Enz, J. Appl. Phys. 32, 22S (1961).
27. D.H. Lyons and T.A. Kaplan, Phys. Rev. 120, 1580 (1960).
28. T.A. Kaplan, K. Dwight, D.H. Lyons and N. Menyuk, J. Appl. Phys. 32, 13S (1961).

30. N. Menyuk, K. Dwight, D.H. Lyons and T.A. Kaplan, Phys. Rev. 127, 1983 (1962).
31. K. Dwight, N. Menyuk, D.H. Lyons and T.A. Kaplan (to be published).
32. J.M. Hastings and L.M. Corliss, Phys. Rev. 126, 556 (1962).
33. N. Menyuk, K. Dwight and A. Wold, J. de Physique 25, 528 (1964).
34. K. Dwight and N. Menyuk, J. Appl. Phys. 40, 1156 (1969).
35. R. Plumier, C.R. Acad. Sci. 255, 2244 (1962).
36. K. Dwight and N. Menyuk, Proc. International Conference on Magnetism, Nottingham, 538 (1964).
37. L.E. Orgel, An Introduction to Transition Metal Chemistry, Ligand Field Theory, J. Wiley and Sons, New York, 1960.
38. C.J. Ballhausen, Introduction to Ligand Field Theory, McGraw-Hill Book Company, New York, 1962.
39. D.S. McClure, J. Phys. Chem. Solids 2, 311 (1957).
40. J.D. Dunitz and L.E. Orgel, J. Phys. Chem. Solids 2, 318 (1957).
41. J.B. Goodenough, J. Phys. Chem. Solids 25, 151 (1964).
42. J. Kanamori, J. Appl. Phys. 31, 14S (1960).
43. N. Menyuk, K. Dwight and P.M. Raccah, J. Phys. Chem. Solids 28, 549 (1967).
44. P.M. Raccah and J.B. Goodenough, Phys. Rev. 155, 932 (1967).
45. J.B. Goodenough, D.H. Ridgley and W.A. Newman, Proceedings International Conference on Magnetism, Nottingham, 542 (1964).
46. J. Kanamori, Anisotropy and Magnetostriiction of Ferromagnetic and Antiferromagnetic Materials, Chap. 4 in Magnetism, Vol. 1, G.T. Rado and H. Suhl (eds.), Academic Press, New York, 1963.
47. P.W. Anderson, Phys. Rev. 115, 2 (1959).
48. P.W. Anderson, Exchange in Insulators: Superexchange, Direct Exchange and Double Exchange, Chap. 2 in Magnetism, Vol. 1, G.T. Rado and H. Suhl (eds.), Academic Press, New York 1963.

49. J.B. Goodenough, *J. Phys. Chem. Solids* 6, 287 (1958).
50. J.B. Goodenough, *Phys. Rev.* 100, 564 (1955).
51. J. Kanamori, *J. Phys. Chem. Solids* 10, 87 (1959).
52. H.A. Kramers, *Physica* 1, 182 (1934).
53. J.B. Goodenough, *Phys. Rev.* 117, 1442 (1960).
54. N. Menyuk, K. Dwight, R.J. Arnott and A. Wold, *J. Appl. Phys.* 37, 1387 (1966).
55. F.K. Lotgering, Proceedings International Conference on Magnetism, Nottingham, 533, 1964.
56. K. Dwight and N. Menyuk, *Phys. Rev.* 163, 435 (1967).
57. G.H. Jonker and J.H. Van Santen, *Physica* 16, 337 (1950).
58. C. Zener, *Phys. Rev.* 82, 403 (1951).
59. P.W. Anderson and H. Hasegawa, *Phys. Rev.* 100, 675 (1955).
60. P.G. deGennes, *Phys. Rev.* 118, 141 (1960).
61. E.O. Wollan and W.C. Koehler, *Phys. Rev.* 100, 545 (1955).
62. J.B. Goodenough, Properties Imparted by d-Electrons, Chap. 4 in the Molecular Designing of Materials and Devices, A.R. von Hippel (ed.), M.I.T. Press, Cambridge, Massachusetts (1965).
63. J.B. Goodenough, *Czech. J. Phys.* 17, 304 (1967).
64. J.B. Goodenough, *J. Appl. Phys.* 39, 403 (1968).
65. J.B. Goodenough, *J. Phys. Chem. Solids* 30, 261 (1969).
66. J.B. Goodenough, Proceedings of 1969 Chania Conference in Crete (to be published).
67. J.B. Goodenough, Metallic Oxides in Progress in Solid State Chemistry, Howard Reiss (ed.), Pergamon Press, London (in press).
68. N.F. Mott, *Can. J. Phys.* 34, 1356 (1956).
69. S. Kawano, K. Kosuge and S. Kachi, *J. Phys. Soc. Japan* 21, 2744 (1966).

70. M. Banus, T.B. Reed and A. Strauss, private communication.
71. A. Ferretti, D.B. Rogers and J.B. Goodenough, *J. Phys. Chem. Solids* 26, 2007 (1965).
72. J.B. Goodenough, J.M. Longo and J.A. Kafalas, *Matl. Res. Bull.* 3, 471 (1968).
73. J.B. Goodenough, *Phys. Rev.* 171, 466 (1968).
74. J.B. MacChesney, H.J. Williams, J.F. Potter and R.C. Sherwood, *Phys. Rev.* 164, 779 (1967).
75. N. Menyuk, K. Dwight, J.A. Kafalas and J.B. Goodenough *J. Appl. Phys.* 40, 1324 (1969).
76. J.B. Goodenough, *Phys. Rev.* 164, 785 (1967).

SOLID STATE PROPERTIES AND NEUTRON DIFFRACTION

D. C. Khan

Department of Physics

Indian Institute of Technology, Kanpur, India

I. INTRODUCTION

Neutrons, are particles of mass about equal to that of protons, but without any electric charge. The precise mass of a neutron, as deduced from a combination of mass spectrometric and nuclear reaction data, is 1.008982 a.m.u. It has a spin 1/2 and a magnetic moment of -1.913148 nuclear magnetons¹. A neutron interacts with a nucleus with a very short range force. Also, because of the existence of the magnetic moment without the presence of charge it interacts with the atomic magnetic field due to the unpaired electrons of the outer shells without any accompanying coulomb effects. The study of the neutron scattering due to neutron-nucleus interaction gives information about the crystal structure and vibrational properties of the crystal whereas that due to magnetic interaction leads to the determination of the pattern of the spin arrangement in the lattice.

In the usual neutron diffraction experiment a monoenergetic collimated neutron beam is scattered through a solid. The scattering is elastic if no exchange of energy takes place between the neutrons and the atoms in the solid; if there is energy exchange, the scattering is inelastic. The intensity distribution pattern of the elastically scattered neutron as a function of the scattering angle reflect the periodicity of the scattering elements, nuclei or magnetic moments, in the solid lattice whereas the intensities of the individual peaks measure the strength of the interaction. These may be described as the static properties of the solid. The atomic vibrations in the solid, taken all together, can be analyzed

into a spectrum of plane waves. The energy associated with these waves are quantized, each energy quantum being called a phonon. During the passage of the neutrons through the solid, they may give energy quanta to the lattice producing more phonons or they may take energy quanta away from the lattice annihilating the corresponding phonon. This inelastic scattering is evidently connected intimately with the details of the phonon spectrum. Again, the magnetic spin arrangement in the solid at higher temperature deviates from its ideal form at absolute zero. This deviation or 'spin flip' can be described as the superposition of the waves of magnetic spin called magnons. Inelastic magnetic scattering gives information about the magnon spectrum.

II. THERMAL NEUTRONS

Thermal Neutrons and X-Rays. For the elastically scattered neutrons to have proper diffraction 'fringes' or a peak pattern, or the energy change of the inelastically scattered neutron to be easily measurable, the energy of the incident neutrons should be in the range 0.01 - 0.1 eV. Neutrons of this energy range are called thermal neutrons. The de Broglie wavelength of the thermal neutrons are of the order of the lattice parameter of crystals ($\sim 1\text{\AA}$). The x-rays have the same wavelength and hence x-ray diffraction and the neutron diffraction will have many similar features. The significant difference is in the nature of the interactions involved and the corresponding cross-sections. X-rays interact with the electrons through Coulomb forces and the scattering amplitude is proportional to Z , the atomic number, whereas the variation of the neutron-nucleus scattering amplitude with atomic number is small. Hence neutron diffraction techniques can precisely determine the position of light elements in the lattice in the presence of heavy elements in their environments. X-ray methods are imprecise here.

Thermal Neutron Sources and Detectors. The source of neutron for the scattering experiments have been the moderated nuclear fission reactors. The reactors used differ extensively as regards power, flux, moderators and hole geometries. Powers range from 0.6MW for the R1 reactor at Stockholm to 200 MW for the NRU reactor at Chalk River. Maximum thermal neutron flux varies from $1.5 \times 10^{12} \text{ cm}^{-2} \text{ sec}^{-1}$ (BEPO, Harwell) to $7.5 \times 10^{14} \text{ cm}^{-2} \text{ sec}^{-1}$ (HFBR Brookhaven). Fuel used have been enriched uranium in most cases. NRU uses natural uranium. Usually D_2O , graphite or water have been used as moderators. Recently pulsed reactors and accelerator pulsed neutron sources have been used².

The instrument extensively used for thermal neutron detection is the boron trifluoride filled proportional counter. These are

basically α -proportional counters, detecting the 2.3 mev α -particles emitted by the high cross-section (n, α) reaction of B^{10} . The efficiency of detection of α -particles approaches 100 per cent and an overall efficiency of 60 per cent is quite common. Other counters used are He-3 filled proportional counters and Boron or Lithium Scintillation detectors.

III. ELASTIC SCATTERING

Crystal and Spin Structure. Fig. 1. shows a typical experimental arrangement for elastic neutron scattering. The incident

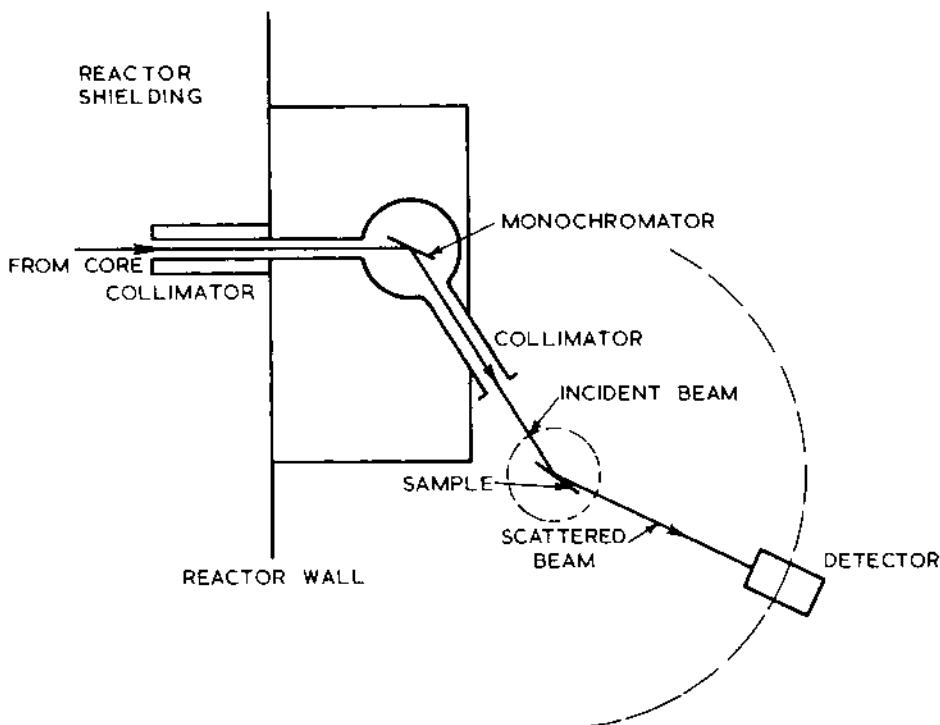


Fig. 1. A typical experimental arrangement for elastic neutron scattering (Diagrammatic representation).

beam from a reactor is collimated and allowed to fall on a crystal at a fixed angle θ . A single wavelength of this polychromatic beam satisfying the Bragg reflection condition gets scattered by 2θ . This monochromatic beam is collimated and allowed to fall on the experimental sample. If the sample is in powder form Debye-Scherrer type peaks will be detected in the scattered beam. If it is a single crystal, for suitable orientations of the crystal Bragg

peaks will be detected at Bragg angles. Fig. 2 shows the Bragg peaks for single crystal of CoO.

If the crystal is at a temperature above the transition temperature of magnetic ordering T_c (The Curie temperature for the ferromagnetic substances, and the Neel temperature for the anti-ferromagnetic and the ferrimagnetic substances), only the nuclear Bragg peaks will be present. Below T_c both nuclear and magnetic peaks will be present. The separation of the nuclear peak from the magnetic peak for unpolarized neutron beam naturally occurs if the magnetic unit cell is of different dimension than the atomic unit cell as in the case of the transition metal oxides (Fig. 2). If they are superposed they may be isolated by taking measurements above and below T_c .

For unpolarized neutron beams, the integrated power of each nuclear peak is proportional to F_n^2 and that for each magnetic peak is proportional to F_m^2 where

$$\begin{aligned} F_n^2 &= \left| \sum_p b_p \exp. 2\pi i (hx_p/a_0 + ky_p/b_0 + lz_p/c_0) \right|^2 e^{-2W} \\ F_m^2 &= \left| \sum_p p q_p \exp. 2\pi i (hx_p/a_0 + ky_p/b_0 + lz_p/c_0) \right|^2 e^{-2W} \end{aligned} \quad (1)$$

Here h, k, l are the Miller indices for the reflection plane, a_0, b_0, c_0 , the lattice constants, e^{-2W} , the Debye-Waller correction term and the summation is over the atoms in a unit cell. b_p is the average scattering length of the nuclei. p is the magnetic scattering length and $q^2 = \sin^2 \alpha$ where α is the angle between the scattering vector normal to the Bragg plane and the magnetization vector giving the direction of the spin axis.

A typical nuclear experiment seeks to find out the values of b_p 's if the positions of the atoms in the lattice (x_p, y_p, z_p) are known and vice versa. b_p 's are nuclear properties reflecting the strength and nature of the interaction between the neutron and the nucleus of the element. For a solid state experiment, the measurement of the positions of the atoms is more important. As mentioned earlier, for the presence of light and heavy elements together in the unit cell, this method is superior to x-ray method. For $(\sin \theta)/\lambda = 0.5 \text{ \AA}^{-1}$, the x-rays scattering factor of hydrogen is $0.02 \times 10^{-12} \text{ cm}$, that of Na is $1.14 \times 10^{-12} \text{ cm}$, for W $11.4 \times 10^{-12} \text{ cm}$, Au $12.5 \times 10^{-12} \text{ cm}$ and U $14.3 \times 10^{-12} \text{ cm}$. Now, the scattering length of hydrogen for neutron is $-0.38 \times 10^{-12} \text{ cm}$ compared to $0.35 \times 10^{-12} \text{ cm}$ for Na, $0.47 \times 10^{-12} \text{ cm}$ for W, $0.76 \times 10^{-12} \text{ cm}$ for Au and $0.85 \times 10^{-12} \text{ cm}$ for U³. Clearly, in the x-ray experiment, the contribution of

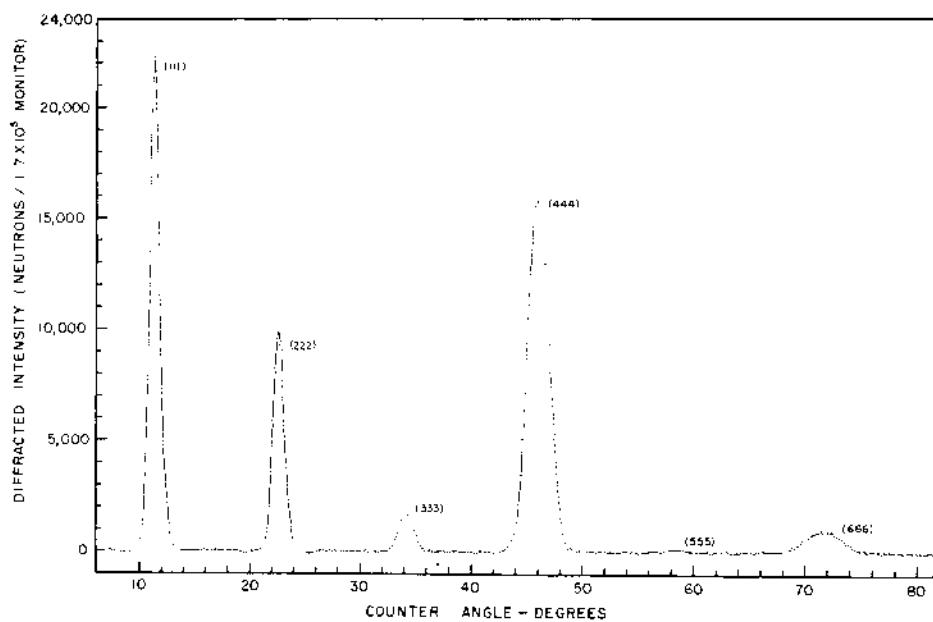


Fig. 2. Nuclear and magnetic neutron diffraction Bragg peaks for CoO. Peaks with odd indices are magnetic and those with even indices are nuclear.

hydrogen will be almost completely masked. The importance of this fact for the study of the structure of organic compounds in which hydrogen atoms are in abundance cannot be overemphasized.

The magnetic peaks reflect the long range ordering of the atomic magnetic moments in the crystal well below the transition temperature. The ordering is ferromagnetic if below T_c all the moments in a domain are aligned in the same direction. For anti-ferromagnetic substances the domain consists of sublattices; in each sublattice the moments are in the same direction, but for each sublattice there is another sublattice with moments in the opposite direction. However, different sublattice doublets in a domain may have different spin axis giving rise to complicated periodic pattern in an antiferromagnet. For ferrimagnetic substances the total magnetic moments of the two sublattices with opposite direction of moments do not balance, or there may exist two or more interpenetrating sublattices with magnetic moment axis in different directions.

The peak pattern as well as the intensities of the peaks in the scattered neutron beam will depend upon the magnetic ordering. Usual practice is to calculate the different peak intensities for various magnetic structure models and select the one that gives the best fit with the experimental data. In large number of crystals the moments of all the sublattices are either parallel or antiparallel to a single direction. Since q^2 in eqn. (1) is the square of the sine of the angle between this direction and the scattering vector, the problem then is to find out by trial a direction whose q^2 for different peaks will lead to theoretical intensities through eqn. (1) which will have best fit with the corresponding experimental intensities. A typical example is shown in Fig. 3 for NiO, where this technique leads to the conclusion that the spins are lying in the (111) plane and their direction is given by [110]. For more complicated spin patterns the expression for F_m in eqn. (1) has to be modified.

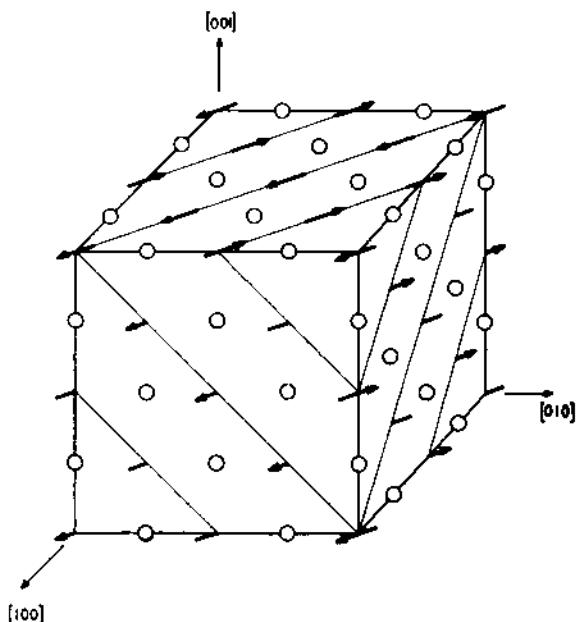


Fig. 3. Spin arrangement for NiO as determined by neutron diffraction experiment (W.L. Roth, J. Appl. Phys., 31, 2000 (1960)).

Magnetic Form Factor. Once the spin structure and hence the q for a crystal is known through a study of the low angle peaks, the eqn. (1) and the integrated intensities for the various Bragg peaks can be utilized to find p as a function of $(\sin \theta)/\lambda$. p is proportional to a quantity f , called the magnetic form factor. f

is a measure of the charge density distribution function of the unpaired electrons. Since the unpaired electrons are responsible for all the magnetic properties of the substance, the form factor and the unpaired electron distribution data are of fundamental importance in theoretical calculations of properties related to magnetism. A typical form factor curve is shown in Fig. 4. f values for large $(\sin \theta)/\lambda$ are quite important in the conversion

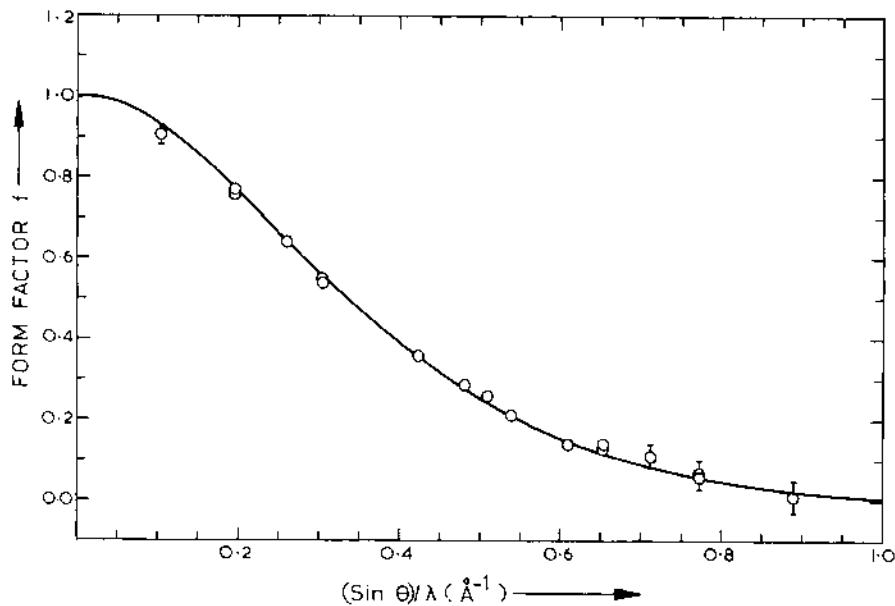


Fig. 4. The isotropic magnetic form factor of CoO.

expression from f to the charge density distribution of the unpaired electrons and hence need extremely careful measurement. The scattering of polarized neutron beams instead of unpolarized ones ensures a better precision and has been utilized to get the form factor of metals at the extreme range of $(\sin \theta)/\lambda \approx 0.08$, having a value as low as 0.022 with an accuracy of 0.002⁴.

IV. INELASTIC SCATTERING

Phonon Creation and Annihilation. The conventional theory of the lattice dynamics of a crystal assumes that the atoms in a crystal vibrate about equilibrium points with very small amplitudes such that the Taylor expansion of the potential function of the entire system about the equilibrium points may be cut off at the quadratic terms. This harmonic approximation leads to the solution of the classical vibration problem in terms of the normal modes, the displacement function corresponding to each of them having the form of a plane wave. The quantum mechanical solution attaches a definite quantum of energy to each of these plane waves. These quantized lattice waves are called phonons. A phonon is characterised by the propagation vector \underline{q} , the polarization vector $\epsilon_{q\lambda}$ and the frequency $\omega_{q\lambda}$. The dispersion relation

$$\omega_{q\lambda} = f(\underline{q}) \quad (2)$$

contains the force constants for inter-atomic interactions as parameters. Hence the experimentally determined dispersion curves form the basis for the analysis of the inter-atomic force constants.

If there are n atoms per unit cell of the experimental crystal, it can be shown that there are $3n$ branches of the dispersion curve. The three lowest branches are characterised by the fact that they have linear variation of ω with \underline{q} like acoustic waves near $\underline{q} = 0$ and these are called acoustic branches. The other $3(n - 1)$ branches are called optical branches, because the antiphase vibrations corresponding to these modes in ionic crystals give rise to volume polarizations which interact with the electromagnetic waves. Individual acoustic or optical branches differ from each other in their polarization.

For the elastic Bragg scattering, the momentum conservation equation is given by

$$\hbar \underline{q} = \hbar \underline{K} - \hbar \underline{K}_0, \quad (3)$$

where \underline{K}_0 is the propagation vector of the incident neutron, \underline{K} is the propagation vector of the scattered neutron, $\hbar \underline{q}$ is the momentum transferred to the lattice as a whole. If the incident beam is such that the Bragg scattering condition is not exactly satisfied, but the momentum conservation eqn. is

$$\hbar \underline{q} + \hbar \underline{Q} = \hbar \underline{K} - \hbar \underline{K}_0 \quad (4)$$

and the energy conservation equation is

$$\pm \hbar\omega = \frac{\hbar^2 K_0^2}{2m} - \frac{\hbar^2 K^2}{2m} \quad (5)$$

where ω and q are connected by the dispersion relation (2) for a phonon of the crystal, then the neutrons exchange energy and momentum with the lattice during their passage through the crystal causing generation (+ sign) or annihilation (- sign) of a phonon. If the experimental arrangement is such that K , K_0 and Q are known, then eqns. (4) and (5) can be utilized to find out q and ω . Hence if the inelastic scattering of neutrons are performed for various crystal orientations and beam directions, many phonons in the crystal can be identified in terms of (ω, q) and if the ω 's are plotted against the corresponding q 's, the dispersion curves (2) for the thermal vibrations of the crystal can be plotted.

A simple inelastic scattering experimental arrangement is shown in Fig. 5. It combines a polycrystalline Be filter producing a

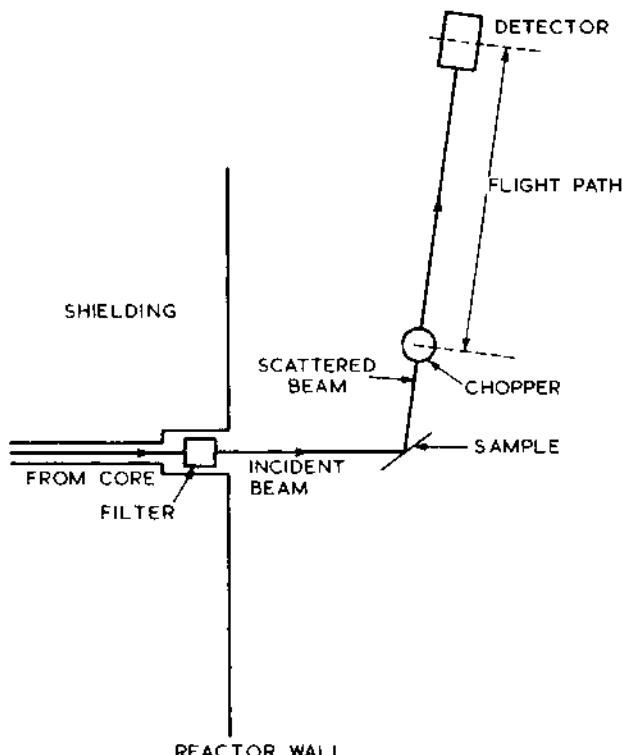


Fig. 5. An experimental arrangement for the inelastic neutron scattering (Diagrammatic representation).

beam of approximately monoenergetic neutrons (4-5 Å) and a chopper time-of-flight system to analyze the energy of the scattered neutrons. The chopper intercepts the scattered neutron beam and produces bursts of neutrons of short time duration on the other side. The simplest type of choppers is a rotating disc with a slot cut parallel to the axis of rotation near the periphery (Fermi type mechanical chopper). Time of flight technique consists of measuring the time the neutrons in the burst take to reach the detector. Since the time of flight is inversely proportional to the square root of energy, a time spectrum of the detected neutrons correspond to the energy spectrum of the neutron burst. In the experiment the orientation of the crystal is predetermined, the scattering direction is fixed. Hence K_0 and Q are known and the direction of K is known. The peak of the scattered spectrum gives the energy and so the magnitude of the scattered beam K and hence ω and q can be found out from eqns. (4) and (5). Fig. 6 shows the dispersion curve

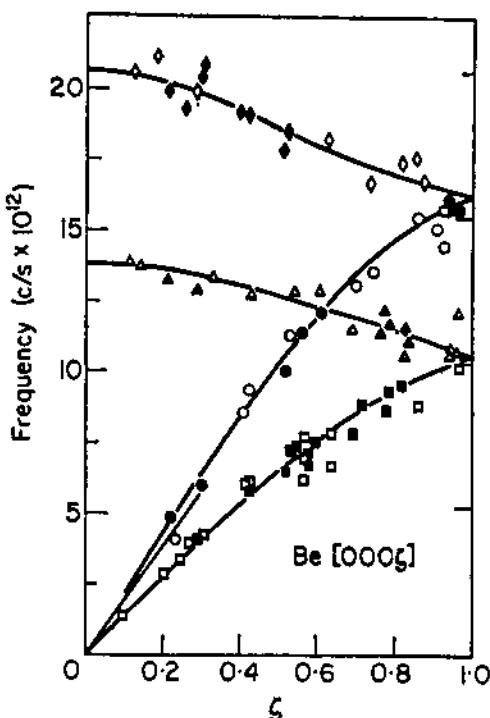


Fig. 6. The measured dispersion curve of beryllium (Schmunk et al. Phys. Rev., 128, 562 (1962)).

for Be along (0001) direction⁵. Both the acoustic mode (lower curves) and the optical mode (upper curves) of vibrations with their longitudinal and transverse branches are present as expected.

Magnon Creation and Annihilation. Experimental techniques similar to the study of the phonon dispersion law around magnetic Bragg scattering angles reveal the dispersion law for quantized spin waves or magnons. At absolute zero the spin system of a crystal is completely ordered. As the temperature increases, the spin disorder starts. Some of the spins have their Z-components changed. The propagation of this disorder is called the spin wave. This can be visualized semiclassically as the propagation of spin vectors with definite propagation vectors and frequencies through the medium. Each wave with a definite q and ω can be considered as a particle in that an energy E_q and a momentum p_q can be associated with it and as quantas they are referred to as magnons.

Since the phonons and the magnons may be studied by the same experimental arrangement, the phonon peaks have to be isolated from the magnons. When a magnetic field is applied normal to the scattering vector ($K - K_0$), the magnon peaks decrease and the phonon peaks augment due to a strong contribution of the magnetovibrational scattering. With complicated lattice, both acoustical and optical branches of the dispersion curve are found.

V. CONCLUSIONS

The long range order (antiferromagnetic, ferromagnetic etc.) in the magnetic domain breaks down as the transition temperature T_c or T_N is approached, but there remains transient short range order over smaller regions. The scattering around the transition temperature, called critical scattering, is predominantly elastic; however, it is in part inelastic because the short range order is dynamic. The critical elastic scattering gives information about the range of correlation whereas the inelastic scattering gives the relaxation time of the correlated spin arrangement.

Finally, it may be mentioned that the scattering phenomena discussed here are all coherent scattering, that is, there is a definite phase relationship between the de Broglie waves scattered at the different lattice sites. The incoherent part of the scattered beam are also studied leading to informations about important quantities as the magnetic form factor and the distribution of phonons.

REFERENCES

1. L.F. Curtiss, Introduction to Neutron Physics, Van Nostrand Co. Inc. Princeton, 1st ed. 1959, pp. 1-16.
2. S.J. Cocking and F.J. Webb in Thermal Neutron Scattering Ed. P.A. Egelstaff, Academic Press, London, 1965, pp. 142-190.
3. G.E. Bacon, Neutron Diffraction, 2nd Ed., Clarendon Press, Oxford, 1962, pp. 31-33.
4. R. Nathans and A. Paoletti, Phys. Rev. Lett., 2, 254 (1959).
5. R.E. Schmunk, R.M. Brugger, P.D. Randolph and K.A. Strong, Phys. Rev., 128, 562 (1962).

MAGNETIC RESONANCE

P. T. Narasimhan

Department of Chemistry

Indian Institute of Technology, Kanpur, India

I. INTRODUCTION

There are at least three main areas of magnetic resonance which are of interest to the solid state chemist. These are: nuclear magnetic resonance (NMR), nuclear quadrupole resonance (NQR) and electron paramagnetic resonance (EPR). We shall first deal with the fundamental principles involved in magnetic resonance¹⁻⁵ and then discuss a few applications of NMR, NQR and EPR to solid state chemistry. It has been found convenient to emphasize the principles of magnetic resonance from the NMR point of view although their validity is wider. Literature references given here are by no means exhaustive, since in this chapter we shall attempt to present a broad picture of the subject. Detailed references are given in areas where there is considerable current interest.

II. NUCLEAR MOMENTS AND NUCLEAR MAGNETIC RESONANCE

Nuclei with odd mass number A or nuclei with even mass number A but with odd charge number Z, possess spin angular momenta. The former class of nuclei have half-integral spins while the latter have integral spins. Nuclei with spin $I = 1/2$ and those with $I = 0$ have uniform spherically symmetric charge distribution while those with $I > 1/2$ have a non-spherically symmetric charge distribution and the deviation from spherical symmetry is characterized by a nuclear quadrupole moment, Q. The electron and such of those nuclei which have a non-zero spin have a magnetic moment $\vec{\mu}$ associated with their spin. In fact, the magnetic moment is simply

proportional to the spin angular momentum. For nuclei we have,

$$\vec{\mu}_N = \gamma_N \vec{J} = \gamma_N \vec{I}h = g_N \beta_N \vec{I} \quad (1)$$

γ_N is the so-called magnetogyric ratio (radians sec⁻¹ gauss⁻¹) and \vec{J} is the angular momentum and $h = h/2\pi$. g_N is the nuclear g-factor and β_N is the nuclear magneton defined as

$$\beta_N = \frac{e\hbar}{2M_p c} \quad (2)$$

with M_p = mass of proton and c the velocity of light. For electron, similarly, we can write

$$\vec{\mu}_e = g \beta_e \vec{J} \quad (3)$$

where g is the Lande g-factor (2.0023 for a free electron) and β_e is the Bohr magneton. In the following discussion we shall consider the case of nuclei in some detail since this will enable us to focus our attention clearly and to present the basic principles of NMR as well as a good part of EPR. It may be pointed out here that whereas NMR and EPR deal with energy levels of nuclei and electrons in a magnetic field NQR concerns itself primarily with the interaction ($e^2 qQ$) between the nuclear quadrupole moment Q and the surrounding electric field gradient (q). The common aspect for all these three is the magnetic dipole nature of the transitions between the energy levels concerned.

The interaction of a nucleus with magnetic moment $\vec{\mu}_N$ in a uniform external magnetic field \vec{H}_0 along the z-direction can be written as,

$$E = -\vec{\mu}_N \cdot \vec{H}_0 = -\gamma_N \vec{I} \cdot \vec{H}_0 \quad (4)$$

This interaction is often referred to as the Zeeman interaction for the nucleus. Since the z-component, I_z , of the nuclear spin takes on a set of quantized values m such that $m = I, I-1, \dots, -I$ we have a set of Zeeman energy levels with energy,

$$E = -\gamma_N \hbar m. \quad (5)$$

Thus, for a nucleus with $I = 1/2$ we have two energy levels with $m = \pm 1/2$ while for $I = 1$ we have three levels $m = \pm 1, 0$ and so on. The energy difference between adjacent levels with quantum numbers m and $m \pm 1$ is given by,

$$\Delta E = \gamma_N \hbar m = \hbar \omega = \hbar \nu \quad (6)$$

where ω is the circular frequency $2\pi\nu$. What has been stated

above is simply the Bohr relation which implies that transitions between adjacent levels will result in discrete lines. Here we have all levels equally spaced and hence all transitions between adjacent levels have the same frequency thus yielding a single line. Transitions between levels differing in m by one unit can indeed be caused by an oscillating magnetic field and it is customary to employ electromagnetic fields of the appropriate frequency ν for this purpose. Equation (6) shows that for a given nucleus the frequency ν is directly proportional to the applied magnetic field H_0 . Typically, in a magnetic field of 10,000 gauss the transition frequency for a proton ($I = 1/2$) is about 42 MHz. A few turns of copper wire wound on a sample tube and connected to a radio frequency oscillator serves the purpose. The axis of the coil is kept perpendicular to the axis of the uniform magnetic field H_0 . The linearly oscillating r.f. magnetic field ($2H_1 \cos \omega t$) produced in the sample coil can be thought of as a superposition of two circularly polarized fields (H_1) rotating in opposite directions. Now, from a classical point of view the nuclear moment not only spins on its own axis but also executes a precessional motion which describes a cone around the axis of the applied magnetic field \vec{H}_0 . The precession frequency, which is familiarly known as the Larmor frequency, is given by

$$\omega = \gamma H_0. \quad (7)$$

The classical and quantum mechanical approaches lead to the same result.

The effect of the applied r.f. field in the direction perpendicular to \vec{H}_0 is to produce two rotating magnetic fields. Only one of these rotating components has the same sense as the nuclear precession. When the frequency and sense of rotation of one of the r.f. field components is the same as that of the Larmor precession maximum torque is exerted by the rotating field. One can thus get a "handle" on the spinning nucleus and flip it to other allowed orientations with a corresponding change in energy. In quantum mechanical language, the effect of applying an oscillating field of the form $2H_1 \cos \omega t$ in the x direction is to produce a perturbation of the form

$$H_{\text{pert.}} = -2\sqrt{\hbar}H_1 I_x \cos \omega t; \quad (8)$$

the probability $P_{mm'}$ of a transition between levels m and m' due to this perturbation is

$$P_{mm'} = \sqrt{2} H_x^2 |\langle m | I_x | m' \rangle|^2 \delta(\omega_0 - \omega). \quad (9)$$

The δ function is used to denote the fact that transitions involve levels with no width. In reality, however, the lifetime of the

states is not infinite and hence linewidths have to be considered. This is usually done by including a "line-shape function" $g(\omega)$ and writing

$$P_{mm'} = \gamma^2 H_1^2 |\langle m | I_x | m' \rangle|^2 g(\omega). \quad (10)$$

The selection rule $\Delta m = \pm 1$ is thus obtained. Notice that the probability of induced transitions increases as the square of the perturbing field strength and $P_{m \rightarrow m'} = P_{m' \leftarrow m}$.

III. SPIN-LATTICE RELAXATION

The observation of magnetic resonance in bulk matter is facilitated by the establishment of Boltzmann distribution of the spins amongst the various Zeeman levels and here the relaxation processes play a very important role. To understand the phenomena let us consider a simple two-level spin system, as for example, the case of $I = 1/2$ (Fig. 1). Let P_{12} be the probability of transition from $-1/2$ to $+1/2$ level and similarly P_{21} . If N_1 and N_2 are the number of spins in the two levels we can write

$$\frac{dN_1}{dt} = N_2 P_{21} - N_1 P_{12} \quad (11)$$

For induced transitions $P_{12} = P_{21} = P$. Hence

$$\frac{dN_1}{dt} = P(N_2 - N_1) = -Pn \quad (12)$$

where n is the excess population ($N_1 - N_2$). We have then $N_1 = 1/2 (N + n)$ and $N_2 = 1/2 (N - n)$ where N is the total number of spins. We can also write

$$\frac{dn}{dt} = -2Pn \quad (13)$$

Solution of equation (13) yields

$$n = n(0) \exp - [2Pt] \quad (14)$$

with $n(0)$ the initial excess population at time $t = 0$. We thus see that if we start with an initial population difference this will disappear in the presence of the radiation field and net energy absorption will eventually go to zero. The spin system is rarely in isolation and is normally coupled to the surroundings or "lattice". At finite temperatures of the whole assembly the lattice has several degrees of freedom and the fluctuations of the surroundings produce a fluctuating field at the site of a spin. In fact, there may be a Fourier component in this fluctuating

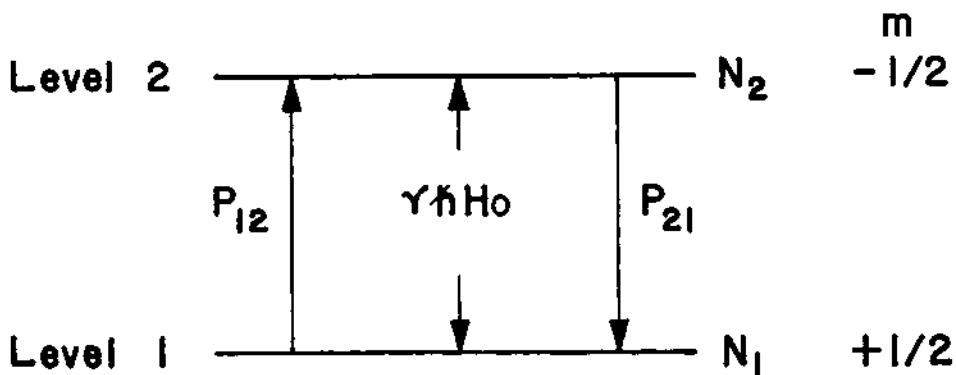


Fig. 1. The two-level spin system

field which is just right for inducing a transition. In other words, the spin system can transfer energy to the surroundings or take up energy from the surroundings by an amount equal to the Zeeman separation. Such lattice-induced transitions are also referred to as "radiationless" transitions. We shall denote the probability of a lattice-induced transition from level 1 to level 2 as R_{12} . Similarly, we have R_{21} . Writing the expression for dN_1/dt as before (see eqn. 11) we have

$$\frac{dN_1}{dt} = N_2 R_{21} - N_1 R_{12} \quad (15)$$

and

$$\frac{dn}{dt} = N(R_{21} - R_{12}) - n(R_{21} - R_{12}) \quad (16)$$

The lattice-induced transition probabilities depend on the ability of the spin system to transfer energy to the lattice. The analogy with heat transfer is clear. We can even think in terms of temperatures of the two systems with heat flow taking place until equilibrium is established. Considerations of this sort tell us that at equilibrium (denoted by superscript 0)

$$\frac{N_2^0}{N_1^0} = \exp - [\Delta E/kT] = \exp - [\sqrt{\hbar}H_0/kT] \quad (17)$$

The lattice-induced transition probabilities should be therefore in the ratio

$$\frac{R_{12}}{R_{21}} = \frac{n_2^o}{n_1^o} = \exp -[\sqrt{\kappa H_o}/kT] \quad (18)$$

Here we have $R_{12} \neq R_{21}$ at a given temperature and magnetic field. Solution of equation (16) can be obtained by writing it as

$$\frac{dn}{dt} = \frac{n^o - n}{T_1} \quad (19)$$

where $n^o = N \left(\frac{R_{21} - R_{12}}{R_{12} + R_{21}} \right)$ (20)

and

$$\frac{1}{T_1} = (R_{12} + R_{21}) \quad (21)$$

Solution of (19) has the form

$$n = n^o + c \exp - (t/T_1) \quad (22)$$

where c is the constant of integration. T_1 can be therefore interpreted as a characteristic time and is an inverse of transition probability.

For a bulk sample containing several spins with $I = 1/2$ the onset of magnetization can be described on a time scale starting from $t = 0$ in an unmagnetized state to the time t by the relation

$$n(t) = n^o \left\{ 1 - \exp - (t/T_1) \right\} \quad (23)$$

In the presence of both radiation-induced transitions we have

$$\frac{dn}{dt} = -2Pn + \frac{n^o - n}{T_1} \quad (24)$$

Under steady state conditions

$$n = \frac{n^o}{1 + 2PT_1} \quad (25)$$

The success of the magnetic resonance experiment lies in keeping the term $2PT_1$ as small as possible so that $n \approx n^o$. The applied r.f. field should not therefore disturb the populations very much from their equilibrium value. This point can be appreciated better if we discuss the rate of absorption of energy which is given by

$$P.n. \Delta E = n^o \cdot \Delta E \cdot \frac{P}{1 + 2PT_1} \quad (26)$$

Since from equation (10) we have $P \propto H_1^2$ we can increase the power

absorbed rapidly by increasing H_1 and when we reach a situation where $2PT_1 \gg 1$ the power absorbed by the spin system does not increase as fast as in the initial stages. The system is said to "saturate".

IV. BLOCH EQUATIONS AND TRANSVERSE RELAXATION

Continuing our discussion of the two-level spin system, we are now in a position to translate the population difference between the two levels in terms of the macroscopic magnetization M_z of the sample along the field. Indeed, we can write,

$$\frac{dM_z}{dt} = \frac{M^0 - M_z}{T_1} \quad (27)$$

where M^0 is the thermal equilibrium magnetization. However, it must be remembered that the macroscopic magnetic moment \vec{M} experiences a torque in the magnetic field \vec{H} which is due to the applied rotating field \vec{H}_1 and static field \vec{H}_0 . There is hence an additional term and we have

$$\frac{dM_z}{dt} = \frac{M^0 - M_z}{T_1} + \gamma (\vec{M} \times \vec{H})_z \quad (28)$$

where $(\vec{M} \times \vec{H})_z$ denotes $(M_x H_y - M_y H_x)$. T_1 can be therefore described as the characteristic time for the onset of magnetization in the z-direction when the sample is placed in the magnetic field. Equations similar to (28) can be written for the perpendicular components M_x and M_y involving a different time constant T_2 . It is appropriate therefore to refer to T_1 as the longitudinal relaxation time and T_2 as the transverse relaxation time. T_2 is also described as spin-spin relaxation time and the reason for this terminology will become clear later. We now proceed to write

$$\frac{dM_x}{dt} = \gamma (\vec{M} \times \vec{H})_x - \frac{M_x}{T_2} \quad (29)$$

$$\frac{dM_y}{dt} = \gamma (\vec{M} \times \vec{H})_y - \frac{M_y}{T_2} \quad (30)$$

The solution of these equations, known as Bloch equations, is simplified if we use the rotating coordinate system which has an angular frequency ω about the z-axis. The transverse components are then referred to as u and v with

$$u = M_x \cos \omega t - M_y \sin \omega t \quad (31)$$

$$v = -M_x \sin \omega t - M_y \cos \omega t \quad (32)$$

Steady state solutions (all time derivatives are zero) of the equations referred to rotating coordinate system yield expressions for u , v and M_z .

$$u = M^0 \sqrt{H_1 T_2} (\omega_0 - \omega) / D \quad (33)$$

$$v = -M^0 \sqrt{H_1 T_2} / D \quad (34)$$

$$M_z = M^0 [1 + T_2^2 (\omega_0 - \omega)^2] / D \quad (35)$$

$$\text{where, } D = 1 + T_2^2 (\omega_0 - \omega)^2 + \sqrt{H_1^2 T_1 T_2} \quad (36)$$

$$\text{and } \omega_0 = \sqrt{H_0} \quad (37)$$

These solutions in the rotating coordinate system can be transformed to the fixed coordinate system to yield the following expressions in the laboratory frame

$$M_x = \frac{1}{2D} M^0 \sqrt{T_2} [(\omega_0 - \omega) T_2 H_1 \cos \omega t + 2H_1 \sin \omega t] \quad (38)$$

$$M_y = -\frac{1}{2D} M^0 \sqrt{T_2} [2H_1 \sin \omega t T_2 (\omega_0 - \omega) - 2H_1 \cos \omega t] \quad (39)$$

$$M_z = \frac{M^0}{D} [1 + T_2^2 (\omega_0 - \omega)^2] \quad (40)$$

These are the steady state solutions of the Bloch equations. They are of general validity in both nuclear and electron paramagnetic resonance.

We have just now considered the effect of the field H_1 rotating in the x - y plane. In practice, this field is generated by a linearly oscillating field in the x axis of the form $H_x = 2H_1 \cos \omega t$ and $H_y = 0$. This is equivalent, however, to two circularly rotating fields whose sense of rotation is opposite to each other. As the frequency ω of the rotating field is varied, $\omega - \omega_0$ decreases and at $\omega = \omega_0$ we have large resonance effects as seen from equations (33)-(40). The other rotating component has usually negligible effect unless H_1 is large or H_0 is so small as to be comparable to H_1 . In such situations the other rotating component leads to a slight shift in the resonance frequency, the so-called Bloch-Siegert effect. In a classical picture, the rotating component which has the same sense as the precessing magnetic moment can exert the maximum torque in comparison to the component with the opposite sense.

The field $2H_1 \cos \omega t$ in the x -direction can be thought of as producing a magnetization which has in-phase (M') and out-of-phase

(M'') components. Writing these components in terms of susceptibilities, we have

$$M' = 2H_1 \cos \omega t. \chi' \quad (41)$$

$$M'' = 2H_1 \sin \omega t. \chi'' \quad (42)$$

χ' and χ'' are the so-called Bloch susceptibilities and the total complex susceptibility is

$$\chi = \chi' - i\chi'' \quad . \quad (43)$$

With $M^0 = \chi^0 H_0$ we can write

$$\chi' = \frac{\chi^0 \omega_0 T_2^2 (\omega_0 - \omega)}{2D} \quad (44)$$

$$\chi'' = \frac{\chi^0 \omega_0 T_2}{2D} \quad (45)$$

D is as defined in equation (36). If H_1 is small the expression for D simplifies to $[1 + T_2^2(\omega_0 - \omega)^2]$. The plot of χ' and χ'' vs $T_2(\omega_0 - \omega)$ is shown in Fig. 2.

The line shape for χ'' is that of a Lorentzian with the line shape function $g(\omega)$ given by

$$g(\omega) = \frac{2T_2}{1 + T_2^2 (\omega_0 - \omega)^2} \quad (46)$$

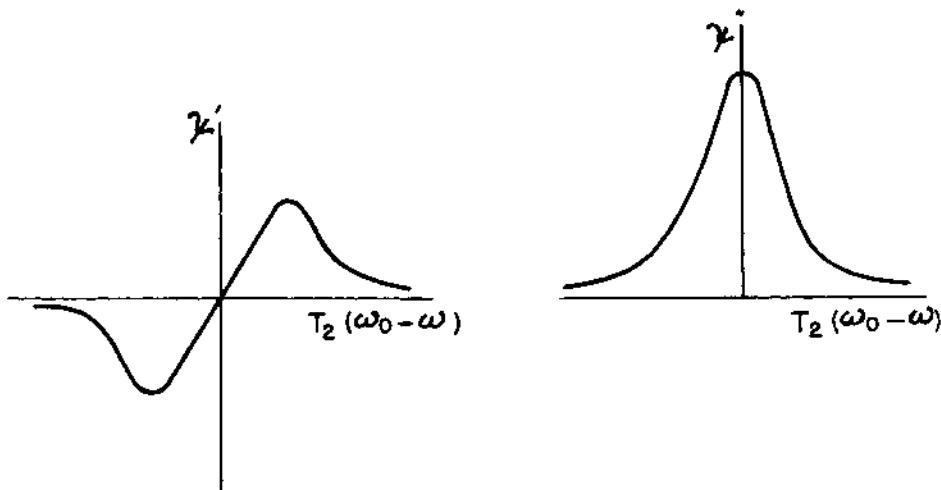


Fig. 2. Plot of χ' and χ'' against $T_2(\omega_0 - \omega)$.

The maximum value of $g(\omega)$ occurs at $\omega = \omega_0$ and

$$g(\omega)_{\max} = 2T_2 \quad (47)$$

The intimate connection between line width and T_2 , the transverse relaxation time, is clear. In this approach we have, however, implicitly assumed that the line width contribution due to T_2 is much greater than that due to T_1 .

It is of interest to note here that a spin system can be "prepared" in such a way that at an initial time $t = 0$ the spins are all precessing along \vec{H}_0 or in a perpendicular direction. The decay of magnetization in the longitudinal and transverse directions can be followed and in this manner T_1 and T_2 can be accurately studied. An elegant way of "preparing" the spins is to use a strong r.f. pulse such that

$$\sqrt{H_1 \Delta t} = \theta \quad (48)$$

Δt is the pulse width and θ the angle (in radians) through which the spins are flipped. Thus for $\theta = \pi/2$ we use a 90° pulse and so on.

V. EXPERIMENTAL ARRANGEMENTS

Two types of arrangements are commonly employed in observing magnetic resonance. One is the bridge method and the other the crossed-coil technique. In the bridge method the sample is placed in a uniform magnetic field and r.f. field is applied in a direction perpendicular to it. The sample coil (or cavity as the case may be depending on the frequency) forms part of a r.f. bridge and the set up is balanced initially at off-resonance. At resonance the sample absorbs energy and the balanced condition is upset and the detector output thus records resonance. Either the frequency or the magnetic field may be varied for resonance. In bridge type spectrometers the frequency is usually fixed and the magnetic field is varied. On account of the fact that only one coil is placed in the magnetic field the arrangement is also referred to as the "single coil" arrangement.

In the "crossed-coil" arrangement the axis of one coil is perpendicular to the uniform magnetic field axis while the axis of the second coil is perpendicular to both these axes. One coil (the transmitter coil) is used to produce the H_1 field while the other (the receiver coil) picks up the voltage induced in it by the precessing magnetic moments. This induced voltage is passed on to a radio receiver and the output presented in a suitable manner.

It is possible to observe the χ' and χ'' components in magnetic resonance experiments. Spectrometers can be tuned in such a way that the output is a measure of say the χ' component. This is the dispersion mode or u-mode while the other mode is the absorption mode or v-mode. For Lorentzian lines χ' and χ'' are related by the Kramers-Kronig relations and it is seen that the dispersion curve has the shape of the derivative of the absorption curve (see also Fig. 2).

The sensitivity of the spectrometer system is considerably enhanced by the use of modulation techniques. If the magnetic field at resonance is modulated at a frequency ω_m by an oscillator, the detector output is also modulated at ω_m . This modulated signal can be fed through a narrow band amplifier which is selectively tuned at ω_m and the output of the narrow band amplifier is passed on to a phase-sensitive detector. The phase-sensitive detector compares the phase of the detected resonance signal with the reference phase provided by the oscillator which modulates the field. If the phases match a d.c. voltage appears at the output of the phase-sensitive detector. This d.c. voltage is proportional to the resonance signal voltage. ω_m is normally in the audio frequency range and in electron paramagnetic resonance ω_m could be as high as 100 KHz. It is essential to keep the amplitude H_m of the modulation field small. Modulation broadening of the lines and consequent distortion of lines takes place at large H_m values. The sweeping of the H_o field through resonance should be also done not too fast since "slow-passage" conditions are necessary and further the phase-sensitive detector integrates the output with a finite time constant. Fast sweep of the field will not therefore yield true signal recordings. For phase-sensitive detector operation H_m should be kept smaller than the line width. Since line widths in solids are usually of the order of a few gauss conditions are not as critical as in the case of liquids. Magnetic field homogeneity restrictions are also not as severe as for liquids. If the spectrometer is tuned in the absorption mode the phase-sensitive detector output yields the derivative of absorption. In order to avoid saturation effects the r.f. field H_1 should be also kept low. The block diagram of a typical magnetic resonance spectrometer system is shown in Fig. 3. Pulsed spectrometers usually employ 90° and 180° pulses and the r.f. oscillator may be directly pulsed or an electronic "gate" used to admit the r.f. into the sample for a short duration. Broad band receivers are used to detect the pulsed signals. An "echo" will occur if proper pulse amplitudes and durations as well as sequences are employed. Thus the popular Carr-Purcell scheme employs 90°-180° pulses. Pulsed spectrometers are capable of yielding accurate T_1 and T_2 values. Fairly large r.f. pulse power is necessary for the study of solids.

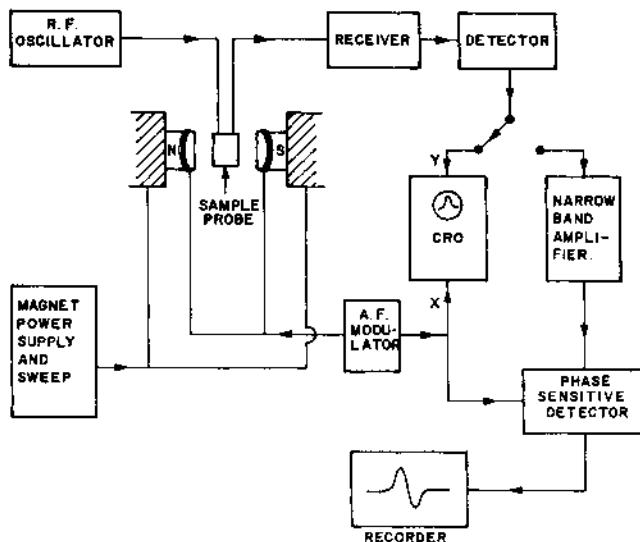


Fig. 3. Block Diagram of a Magnetic Resonance Spectrometer.

VI. NMR IN NON-METALLIC SOLIDS

Not all experimentally recorded magnetic resonance lines have the Lorentzian shape predicted by Bloch equations. A commonly observed line shape in solids is the Gaussian with a bell-shaped curve. Magnetic resonance lines can be characterized by their moments. The n^{th} moment is defined as

$$\langle \omega^n \rangle = \frac{\sum_{m,m'} \omega_{mm'}^n | \langle m | I_x | m' \rangle |^2}{\sum_{m,m'} | \langle m | I_x | m' \rangle |^2} \quad (49)$$

In terms of the absorption line shape function $f(\omega)$ we may write

$$\langle \omega^n \rangle = \frac{\int_0^\infty \omega^n f(\omega) d\omega}{\int_0^\infty f(\omega) d\omega} \quad (50)$$

The second and fourth moments with $n=2$ and $n=4$ respectively are often met with in the study of NMR in solids. The quantity $\langle \Delta \omega^n \rangle$ is also important here.

$$\langle \Delta \omega^n \rangle = \frac{\int_0^\infty (\omega - \langle \omega \rangle)^n f(\omega) d\omega}{\int_0^\infty f(\omega) d\omega} \quad (51)$$

We shall presently relate $\langle \Delta \omega^2 \rangle$ to the dipolar broadening in solids. Before doing so, let us consider the effect of H_1 on the observed signal.

From equations (33)-(35) it is clear that if H_1 is large the term $\gamma^2 H_1^2 T_1 T_2$ will be large and the transverse components and also χ' and χ'' decrease in magnitude. The resonance signal also decreases in intensity and this is the phenomenon of saturation (see also end of Sec. III). The term $\gamma^2 H_1^2 T_1 T_2$ is referred to as the saturation factor. Both u- and v-mode signals get affected in the same manner with large H_1 , as seen from the solutions of Bloch equations. Experimentally, however, it is found in solids that the dispersion mode signal saturates less easily than the absorption mode signal (see Chapter XII of Ref. 2, for example). In searching for new NMR signals in solids it is therefore advisable to operate the spectrometer in the dispersion mode and use as large a H_1 as possible.

The magnetic field experienced by a nuclear moment in a solid is not just the externally applied field value but is modified by the electron cloud around it (chemical shielding) and also the magnetic dipoles present near the nucleus of interest. We shall concern ourselves first with the dipolar field. A magnetic moment μ at a distance r away from a nucleus produces a local field $H_{loc} \approx \mu/r^3$. With $\mu = 10^{-23}$ erg/gauss = 10^{-3} B.M. and $r = 2\text{\AA}$ we see that $H_{loc} \approx 1$ gauss. This local field will depend upon the orientation of μ and the angular variation is given by the factor $(3 \cos^2 \theta - 1)$ where θ is the angle between the magnetic moment axis and the radius vector. Consequently, the local field may aid or oppose the externally applied field. Since in a solid there are several nuclei around a given nucleus we can expect a spread in the local field value and when we vary the external magnetic field value resonance condition may be obtained over a wide range of H_0 . The resonance line in a solid has therefore a larger width corresponding to a shorter T_2 . (vide also equation 46.) The relation between T_2 and H_{loc} can be understood by a different approach. If nuclei precess in phase at $t = 0$ they would get out of phase after some time due to local field inhomogeneity spoiling the resonance condition. After a time τ there will be significant dephasing since $\sqrt{H_{loc}} \tau \approx 1$. A rough estimate of T_2 is therefore

$$T_2 = \frac{1}{\sqrt{H_{loc}}} = \frac{r^3}{\sqrt{2} h} \quad (52)$$

$T_2 \approx 100\text{ sec.}$ for nuclei and in comparison we have $T_1 \approx$ secs (or hours !) in solids. If the externally applied magnetic field itself is inhomogeneous there will be a line broadening. This is a case of instrumental broadening.

The classical energy of interaction between two magnetic dipoles $\vec{\mu}_1$ and $\vec{\mu}_2$ can be expressed as

$$E = \frac{\vec{\mu}_1 \cdot \vec{\mu}_2}{r^3} - \frac{3(\vec{\mu}_1 \cdot \hat{r})(\vec{\mu}_2 \cdot \hat{r})}{r^5} \quad (53)$$

Here we have $\vec{\mu}_1 = \sqrt{\hbar I_1} \hat{r}$ and similarly $\vec{\mu}_2$. In quantum mechanical language the dipolar interaction Hamiltonian is

$$\underline{H}_d = \frac{\sqrt{I_1} \sqrt{I_2} \hbar^2}{r^3} [A + B + C + D + E + F] \quad (54)$$

where

$$A = I_{1z} I_{2z} (1 - 3 \cos^2 \theta) \quad (55)$$

$$B = -1/4 [I_1^+ I_2^- + I_1^- I_2^+] (1 - 3 \cos^2 \theta) \quad (56)$$

$$C = -3/2 [I_1^+ I_{2z} + I_{1z} I_2^+] \sin \theta \cos \theta e^{-i\phi} \quad (57)$$

$$D = -3/2 [I_1^- I_{2z} + I_{1z} I_2^-] \sin \theta \cos \theta e^{i\phi} \quad (58)$$

$$E = -3/4 I_1^+ I_2^+ \sin^2 \theta e^{-2i\phi} \quad (59)$$

and

$$F = -3/4 I_1^- I_2^- \sin^2 \theta e^{2i\phi} \quad (60)$$

Here we have used spherical polar coordinates and $I_1^+ = I_{1x} + i I_{1y}$ while $I_1^- = I_{1x} - i I_{1y}$.

Let us consider the case where the two interacting spins are identical with $I = 1/2$. There are three triplet functions:

$$\begin{aligned} T_{+1} &= \alpha(1) \alpha(2) \\ T_0 &= \alpha(1) \beta(2) + \alpha(2) \beta(1) \\ T_{-1} &= \beta(1) \beta(2) \end{aligned} \quad (61)$$

and one singlet function

$$S_0 = \alpha(1) \beta(2) - \alpha(2) \beta(1) \quad (62)$$

in this case. From symmetry considerations we see that there is no matrix element between singlet and triplet functions. Thus the only transitions that are present come from the triplet states. The three energy levels can be obtained by using the triplet wave functions and the Hamiltonian

$$\underline{H} = \underline{H}_{\text{Zeeman}} + \underline{H}_d \quad (63)$$

The result is shown in Fig. 4. The two transitions that are thus

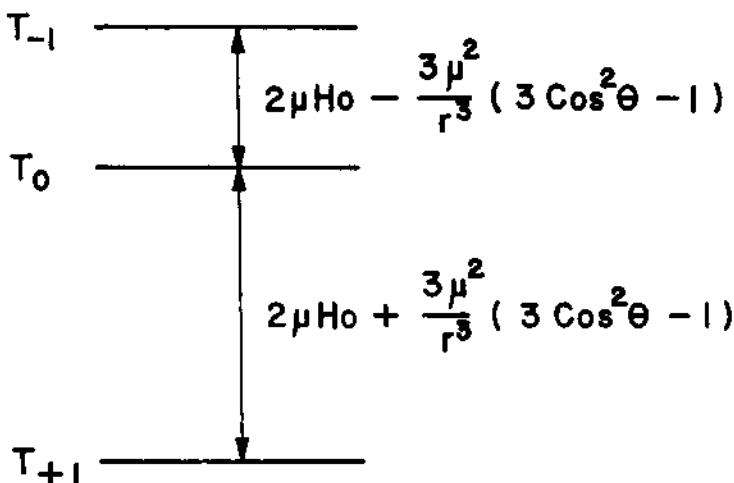


Fig. 4. Level separations in a system with two identical spins with dipolar interaction.

obtained correspond to

$$\hbar\nu = 2\mu_{H_0} \pm \frac{3\mu^2}{r^3} (3\cos^2\theta - 1) \quad (64)$$

In the absence of the dipolar coupling only one frequency ($\hbar\nu = 2\mu_{H_0}$) is obtained. Thus in a crystalline sample with pairs of identical interacting spins we get two resonance peaks at

$$H = H_0 \pm \frac{3}{2} \frac{\mu}{r^3} (3\cos^2\theta - 1) \quad (65)$$

By measuring the separation between the resonance doublets as a function of θ we can determine r . This idea was originated by Pake who first studied the proton-proton distance in $\text{CaSO}_4 \cdot 2\text{H}_2\text{O}$ single crystals. Theoretical expressions for the case of three and four interacting spins in triangular and tetrahedral configurations have been worked out by Pake, Gutowsky and Andrew^{1,2}. Several internuclear distances in crystals have been determined in this manner by NMR. Hydrogen bonding in solids⁶ has also been extensively studied by this approach. NMR has also provided valuable information on hydrogen-bonded ferroelectrics⁷ in the light of J.C. Slatters' theory. Deuterium resonance in deuterated crystals have also been studied in this manner. Blinc and coworkers⁸ have studied the phase transition problem in ferroelectrics with hydrogen bonds from the NMR point of view.

When there are several different pairs of interacting nuclei the superposition of various doublets as well as their differing

dependence on θ (eqn. 65) can create an experimental situation which is extremely complicated. Single crystal specimen are valuable from the point of view of angular variation studies and in powder specimen the orientation averaging curtails the information.

Van Vleck's moment method¹⁻³ is a powerful approach commonly employed in the analysis of NMR spectra of solids. In a solid consisting of N interacting spins the dipolar Hamiltonian is

$$\mathcal{H} = \frac{1}{2} \sum_{j=1}^N \sum_{k=1}^N \left[\frac{\vec{\mu}_j \cdot \vec{\mu}_k}{r_{jk}^3} - \frac{3(\vec{\mu}_j \cdot \vec{r}_{jk})(\vec{\mu}_k \cdot \vec{r}_{jk})}{r_{jk}^5} \right] \quad (66)$$

A theoretical calculation of the energy levels and transitions should in principle yield the spectrum for a solid. But this is not so in practice since such a calculation hinges on the solution of the many-body problem. The procedure suggested by Van Vleck is to "truncate" the dipolar Hamiltonian keeping only the A and B terms of equation (54). The spectrum is then calculated in terms of the moments (equations 49-51). The experimental moment values are then compared with the theoretical moments to yield important structural data. For a system of identical nuclei located at equivalent positions in a crystal the second moment line width expression is

$$\langle \Delta\omega^2 \rangle = \frac{3}{4} \gamma^4 \hbar^2 I(I+1) \sum_k \frac{(1-3 \cos^2 \theta_{jk})^2}{r_{jk}^6} \quad (67)$$

Moment expressions for other cases as well as powders are available in the literature¹⁻³.

One of the classic examples of the use of the method of moments is the study of proton NMR spectra of solid benzene and deuterated benzenes by Andrew and Eades⁹. These authors studied the temperature dependence of the second moment of the proton resonance and found that the second moment value of 9.7 gauss² at 80°K drops down to 1.6 gauss² at about 120°K. The observed second moment contains contributions from both intra- and intermolecular dipolar interactions. The interpretation of the variation in second moment is that with increase in temperature there is an increase in the amplitude of molecular motion along the hexad axis. Precise inter-proton distances could also be obtained for the molecule. The distance between adjacent protons in the molecule obtained from NMR is 2.495 ± 0.015 Å as compared to 2.473 ± 0.025 Å from x-ray results.

Extensive investigations of hindered rotation in solids of molecules or groups in molecules have also been carried out by

NMR. Hindered rotation of methyl and NH_4^+ groups in solids has been thus studied and clear evidence for the temperature dependence obtained. Work in this area follows largely the pattern originated by Gutowsky and Pake (See Ref. 1). Structural studies in solids have also been carried out using the moments method. As examples of systems studied in this manner mention may be made of diketene and the infusible white precipitate obtained by addition of HgCl_2 to aqueous ammonia. The latter compound has been shown by NMR to be $\text{NH}_2\cdot\text{HgCl}$. Crystals of monohydrates of sulfuric, nitric and perchloric acids contain H_3O^+ while oxalic acid is a normal hydrate. This conclusion has been drawn from NMR second moment studies (e.g., p. 35 of Ref. 4). Phase transitions in solids have been studied at high pressures by NMR and a comprehensive review on this subject is available¹⁰.

VII. QUADRUPOLAR SPLITTINGS IN NMR

In the NMR spectra of solids splittings besides those due to dipolar interactions may be seen in some instances. These are observed when the nuclear spin $I \geq 1$. Such nuclei possess a nuclear electric quadrupole moment Q and this quadrupole moment can interact with a non-vanishing electric field gradient produced by the electron cloud as well as other charges surrounding the nucleus in question. Denoting the principal components of the electric field gradient as q_{xx} , q_{yy} and q_{zz} the quadrupole interaction Hamiltonian can be written as¹¹,

$$\underline{H}_Q = \frac{e^2 Q}{4I(2I-1)} \left[q_{zz} (3\underline{I}_z^2 - \underline{I}^2) + (q_{xx} - q_{yy}) (\underline{I}_x^2 - \underline{I}_y^2) \right] \quad (68)$$

Writing $eq_{zz} = eq$ and $\gamma = (q_{xx} - q_{yy})/q_{zz}$ we have

$$\underline{H}_Q = \frac{e^2 Qq}{4I(2I-1)} \left[3\underline{I}_z^2 - \underline{I}^2 + \gamma (\underline{I}_x^2 - \underline{I}_y^2) \right] \quad (69)$$

γ is referred to as the asymmetry parameter. If the surrounding electric field gradient is axially symmetric i.e. $q_{xx} = q_{yy}$, $\gamma = 0$ and we then have

$$\underline{H}_Q = \frac{e^2 Qq}{4I(2I-1)} (3\underline{I}_z^2 - \underline{I}^2) \quad (70)$$

\underline{H}_Q can be treated as a perturbation on the Zeeman term if one is working at fairly large fields as in usual NMR experiments. For the case of $I = 3/2$ the quadrupole perturbation of the form given in equation (70) on the Zeeman levels leads to a triplet structure of the NMR spectrum with an intense central line flanked symmetrically by "satellite" lines. A measurement of the separation of these lines yields $e^2 Qq$ value from which q may be obtained. Crystal orientation studies can also help in obtaining the axes of the field gradient tensor. A value of $e^2 Qq/h = 334$ KHz has thus

been obtained for Na^{23} in NaNO_3 crystal¹². Deuterium quadrupole couplings have been largely measured in this manner¹³. Interpretation of such quadrupolar splittings in terms of charge distribution and crystal symmetry has been done in several cases and NMR study of solids has thus provided an additional avenue.

VIII. NMR IN METALS

The original observation of NMR in metallic copper by Knight^{1-3,14} has been followed by intense activity in this field. The resonance frequency of Cu^{63} in metallic copper was found by Knight to be slightly higher than that in diamagnetic copper salts. This difference in resonance frequency is referred to as Knight Shift and in most metals this situation prevails. The observed shift is proportional to the applied field but the fractional shift $K = \Delta H/H_0$ is independent of field and is usually almost independent of temperature also. It is also found that for elements with large atomic number the $\Delta H/H_0$ value increases. The range of K is 10^{-4} to 10^{-2} . The theoretical interpretation of the Knight shift employs the model of the conduction electron interacting with a magnetic nucleus via the hyperfine term:

$$H_{\text{h.f.}} = \frac{8\pi}{3} \gamma_e \gamma_N \hbar^2 \vec{I} \cdot \vec{S} \delta(\vec{r}) \quad (71)$$

where $\delta(\vec{r})$ is the Dirac δ -function denoting the probability of finding the conduction electron spin at the nucleus N. Referring to electron and nuclear co-ordinates as \vec{r}_t and \vec{R}_u respectively we may write equation (71) as

$$H_{\text{h.f.}} = \frac{8\pi}{3} \gamma_e \gamma_N \hbar^2 \sum_{t,u} \vec{I}_u \cdot \vec{S} (\vec{r}_t - \vec{R}_u) \quad (72)$$

For the conduction electron we can use the Bloch wave functions of the form

$$\chi_{\vec{k},s} = (\phi_{\vec{k}}(\vec{r}) e^{i\vec{k} \cdot \vec{r}}) (\sigma_s) \quad (73)$$

where on the right hand side σ_s is the spin function for the electron and the other term represents the space part. \vec{k} is the wave vector. We can evaluate the hyperfine interaction energy using proper antisymmetrized product wave functions. Thus, referring to the origin of the coordinate system as that of the nucleus with $\vec{R}_u = 0$ and with the magnetic field \vec{H}_0 along z-axis we have for the nucleus u,

$$E_{\text{h.f.}}^u = \sum_{\vec{k},s} \epsilon^u(\vec{k},s) = \frac{8\pi}{3} \gamma_e \gamma_N \hbar^2 I_{zu} \sum_{\vec{k},s} |\phi_{\vec{k},s}(0)|^2 m_s p(\vec{k},s) \quad (74)$$

$\phi_{\vec{k},s}(0)^2$ is the probability of finding a conduction electron with wave vector \vec{k} and spin s at the site of the nucleus and $p(\vec{k},s)$ is an occupation factor which is unity if the \vec{k},s are occupied or else is zero. m_s can take the values $+1/2$ or $-1/2$. The value of $p(\vec{k},s)$ depends on temperature of the electron system and the proper function to use here is the Fermi function $f(\vec{k},s)$. Since for the electron two m_s values are available the energy of interaction is

$$\begin{aligned} \xi^u(\vec{k},s) &= \frac{8\pi}{3} \gamma_e \gamma_N \hbar^2 I_{zu} [(1/2) f(\vec{k},1/2) + (-1/2) f(\vec{k},-1/2)] |\phi_{\vec{k}}(0)|^2 \\ &= \frac{8\pi}{3} \gamma_N \hbar I_{zu} |\phi_{\vec{k}}(0)|^2 [\gamma_e .1/2 \cdot f(\vec{k},1/2) - \gamma_e .1/2 \cdot f(\vec{k},-1/2)] \end{aligned} \quad (75)$$

The last term in equation (74) represents an effective magnetization and rewriting it in term of spin susceptibility of electrons we have

$$E_{h.f}^u = - \frac{8\pi}{3} \gamma_N \hbar I_{zu} \left[\sum_{\vec{k}} |\phi_{\vec{k}}(0)|^2 \chi_{\vec{k}}^s \right] H_0 \quad (76)$$

Equation (76) has been obtained using the fact that $\chi_{\vec{k}}^s H_0$ is the magnetization due to electron spins having m_s values $+1/2$ and $-1/2$ and thus performing the sum over s . The averaging over \vec{k} is more involved (see refs. 2 & 3) and can be performed to yield

$$E_{h.f}^u = - \frac{8\pi}{3} \gamma_N \hbar I_{zu} \langle |\phi_{\vec{k}}(0)|^2 \rangle_F \chi_e^s H_0 \quad (77)$$

$\langle \rangle_F$ denotes averaging at the top of the Fermi distribution and χ_e^s is the electron spin susceptibility. The Knight shift is then given by

$$K = \frac{\Delta H}{H_0} = \frac{8\pi}{3} \langle |\phi_{\vec{k}}(0)|^2 \rangle_F \chi_e^s \quad (78)$$

Atomic hyperfine data usually provide $|\phi_{\vec{k}}(0)|^2$ values in the free atom. The symbols P_F and P_A are usually employed in the literature for $|\phi_{\vec{k}}(0)|^2$ and $|\phi_{\vec{A}}(0)|^2$ respectively and P_F/P_A is denoted by ξ . Since s-electrons have large $|\phi_{\vec{k}}(0)|^2$ values Knight shift data provide us with a method of determining % s character of the conduction electrons. The discussion given above for Knight shift has considered only the isotropic type of hyperfine interaction, i.e. I.S. However, there is an anisotropic type of hyperfine interaction as well and if the site symmetry around the nucleus is low the nuclear-electron interaction should be really described in terms of the appropriate tensor. The Knight shift is then anisotropic and can manifest itself by asymmetric line shapes recorded in the NMR of metal nuclei. Several alloys and metals

have been thus investigated and important information on conduction electrons obtained. A reference may be made here to the large number of studies on transition metals and rare earth metals and their alloys with a view to determine their electronic structure¹⁵. Mössbauer studies have also supplemented the NMR information. It has been found in some of these cases that K can be negative and this has been explained in terms of the hyperfine interaction of the core s-electrons with conduction d-electrons. Additional factors such as spin and orbital magnetic moments of d-electrons have also been shown to contribute to K. In passing, we may note that the theoretical work of Korringa and Pines relate Knight shift to the spin-lattice relaxation in metals (see for ex. refs. 2,3).

A few words about line widths in metals and alloys may not be out of place here. In pure metals it is very often found that the NMR line is narrow while in alloys it is broad and almost lost in noise. Of the factors that contribute to this phenomenon the most important are exchange narrowing, quadrupole coupling and indirect spin coupling. The Van Vleck second moment is reduced for the case of 100% pure metal (all identical nuclei) while with isotope dilution even line width contributions increase significantly. The reduction in line width for the case of identical nuclei is referred to as exchange narrowing. If quadrupole effects are present the main resonance line is split and this splitting may broaden the line. Similarly, unresolved splittings due to electron-coupled (indirect) nuclear spin interactions (see also Sec. IX) may also broaden the NMR line in metals. Thus, Bloembergen and Rowland¹⁶ have studied the metal NMR of Tl²⁰³ and Tl²⁰⁵ mixed in various proportions and shown how as the "impurity" concentration increases the line widths change leading to sharp lines for the pure metals. A survey of the topic has been given by Rowland¹⁶. Translational diffusion of spins in metals with an increase in temperature has been studied by NMR by following the T₁ and T₂ values. Solid solutions of hydrogen in metals have been examined from this point of view. Various models for the relaxation mechanism have also been proposed^{2,15}.

IX. CHEMICAL SHIFT AND SPIN COUPLING IN SOLIDS

We have mentioned earlier that the externally applied magnetic field may be modified by the electron cloud surrounding the magnetic nucleus of interest. The Knight shift discussed in the previous section is only a special case of this type. For a given nuclear species in different chemical environments, the resonance frequency can be different due to the differences in the electronic environment and consequent differences in the magnetic shielding. Indeed, this is the basis of high resolution NMR spectroscopic studies

involving liquid and gas samples. The chemical shielding is expressed in the form

$$H_N = H_o (1 + \sigma) \quad (79)$$

where σ is the shielding constant which may be positive or negative. In general, σ is a tensor with principal components σ_{xx} , σ_{yy} and σ_{zz} but in liquids one gets an average $\sigma = 1/3 (\sigma_{xx} + \sigma_{yy} + \sigma_{zz})$. Now, for protons $\sigma \approx 10^{-5}$ while for F¹⁹ it is larger being in the range $10^{-2} - 10^{-3}$. Due to the narrow lines in fluids nuclei in different chemical environments give rise to characteristic lines usually separated from one another (chemical shift). An additional feature that is seen in the NMR spectra of fluids is the multiplet splittings due to spin-spin interaction of the form $J_{NN}, \vec{I}_N \cdot \vec{I}_{N'}$, between pairs of nuclei N, N'. Such well-resolved spectra are obtained with liquids and gases only since the dipolar interaction between nuclei, which gives rise to broad lines in solids, averages out to zero in fluids. In solids, however, such a study of chemical shifts and spin couplings is not possible. In certain favourable cases shielding anisotropy ($\Delta\sigma$) values have been obtained from conventional NMR studies in solids¹⁷. It has been pointed out by Andrew¹⁸ that if a single crystal specimen containing the nucleus of interest is in a magnetic field at the "magic angle" and rotated macroscopically at ultra centrifuge speeds the dipolar term can be averaged out leaving the chemical shift and spin coupling effects. Thus, in practice, it becomes feasible to study the "high resolution" NMR spectrum of a solid. The experiments of Andrew and co-workers in this area have been encouraging.

A rather elegant way of averaging out the dipolar and quadrupolar interactions in solids has been proposed by Waugh¹⁹. Here the pulse technique is utilized and by a proper sequence of pulses the nuclear moment is allowed to precess at the "magic angle". In this manner, the use of complicated ultra centrifuge turbines is avoided.

Considerable interest centers around the measurement of spin couplings in solids. Some theoretical investigations of electronic structure of solids have included a discussion of the role of various mechanisms for indirect spin-spin interaction here. One of the important results in this area was obtained several years ago by Ruderman and Kittel²⁰ who studied the spin coupling in metals by conduction electrons. This theory was further extended by Kasuya²¹ and Yosida²². The general theory is often referred to as the Ruderman-Kittel-Kasuya-Yosida (R-K-K-Y) theory. Both isotropic and anisotropic interactions of the nuclei and conduction electrons are considered in the theory.

X. PURE QUADRUPOLE RESONANCE IN SOLIDS

It was pointed out in Section VII that a nucleus with electric quadrupole moment Q can interact with the surrounding electric field gradient and the Hamiltonian for this interaction was denoted as eQ . We shall first consider the case of $\gamma = 0$.

Axially symmetry case ($\gamma = 0$). Here, we have,

$$\langle m | H_Q | m' \rangle = \frac{e^2 Q g}{4I(2I-1)} [3m^2 - I(I+1)] \delta_{mm}, \quad (80)$$

with $\delta_{mm} = 0$ if $m \neq m'$ and $\delta_{mm} = 1$ if $m = m'$. The quadrupole energy levels are thus given by

$$E_m = \frac{e^2 Q g}{4I(2I-1)} [3m^2 - I(I+1)] = A [3m^2 - I(I+1)] \quad (81)$$

For half-integral spins there are $I + 1/2$ doubly degenerate energy levels and for integral spins there are $I + 1$ energy levels of which I are doubly degenerate and the one with $m = 0$ is non-degenerate. Magnetic transitions between levels with m and $m \pm 1$ quantum numbers can be caused by a perturbation in the form of a r.f. field with H_1 in the x, y plane just as in the case of NMR. The transition frequency is given by

$$\omega_m = \frac{3A}{\pi} (2 | m | + 1) \quad (82)$$

The number of transition frequencies is $I - 1/2$ for half-integral spins and I for integral spins. The $\pm m$ levels are degenerate in the absence of magnetic field but a weak magnetic field can remove the degeneracy and additional lines now appear depending on the orientation of the electric field gradient axis with the external magnetic field. Zeeman studies are thus of help in determining the axis of the electric field gradient which in turn is of help in crystal structure studies and related problems.

Non-axially symmetric case ($\gamma \neq 0$). Here, we must consider two cases (i) integral spins and (ii) half-integral spins in a non-axially symmetric electric field gradient. The detailed results and energy levels are available in the literature²³. In the case of $I = 3/2$, unlike other cases, there is only one transition frequency. Thus a determination of the field gradient g and the asymmetry parameter γ is not possible for the $I = 3/2$ case unless a Zeeman study is made. In single crystal specimen a weak magnetic field H_0 with the r.f. field H_1 perpendicular to it will yield a maximum of four lines. The separation within the inner pair of lines (referred to as the α pair) and in the outer pair of lines (the β pair) varies with the orientation of the crystal in the magnetic field. Referred in terms of the angle

θ between \vec{H}_o and the electric field gradient symmetry axis and the azimuthal angle ϕ the condition for the inner pair of lines to coincide is

$$\sin^2 \theta = \frac{2}{3 - \gamma \cos^2 \phi} \quad (83)$$

Asymmetry parameter measurements can be thus made from Zeeman study of single crystals for $I = 3/2$ case. In the case of other nuclei, such as $I = 1$ case, there are sufficient number of pure NQR transitions (without Zeeman effect) to enable one to determine γ . Thus we have for $I = 1$

$$\begin{aligned} v_1 &= \frac{3}{4h} e^2 qQ (1 + 1/3 \gamma) \\ v_2 &= \frac{3}{4h} e^2 qQ (1 - 1/3 \gamma) \\ v_3 &= \frac{1}{2h} e^2 qQ \gamma \end{aligned} \quad (84)$$

An ingeneous method of determining γ from powder specimen in the case of $I = 3/2$ has been developed by Morino and Toyama²⁴. Here the r.f. field is applied parallel to H_o in a polycrystalline sample. Theoretical calculations predict the appearance of four transitions as shown in Fig. 5, where $v_H = \sqrt{H_o}/2\pi$.

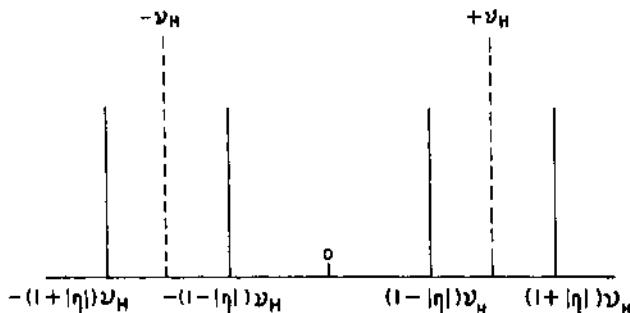


Fig. 5. NQR transitions with $\vec{H}_1 \parallel \vec{H}_o$; $I = 3/2$ case.

Due to the random orientation of the field gradient axes amongst various crystallites in the specimen the observed spectrum is a result of the powder averaging. Morino and Toyama have shown that the theoretically expected derivative spectrum for $\gamma \neq 0$ and with $\vec{H}_1 \parallel \vec{H}_o$ has the form shown in Fig. 6. By measuring the separation of the peaks as a function of applied field and extrapolating to infinite field strength γ can be obtained. The Morino-Toyama procedure thus avoids the use of single crystal specimen.

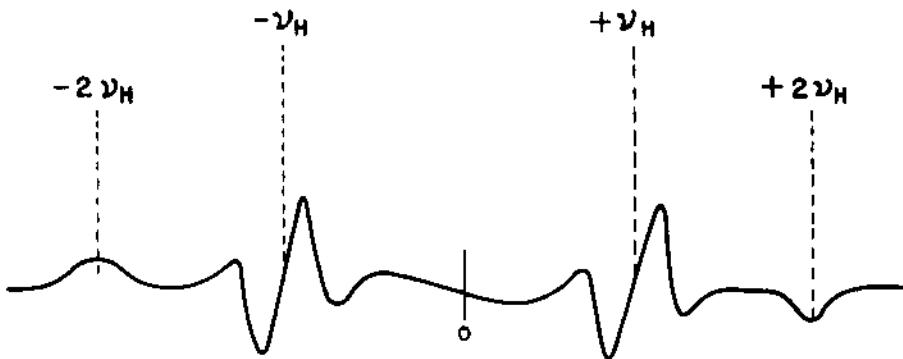


Fig. 6

There are several nuclei with $I > 1/2$ and hence possessing nuclear quadrupole moment eQ . The sign of eQ is determined by the nuclear properties and for prolate spheroid types (e.g., Br⁷⁹) it is positive while for oblate spheroid types (e.g., Cl³⁵) it is negative. The electric field gradient eQ due to electron cloud distribution around the quadrupolar nucleus described by a wave function Ψ and other nuclei with atomic number Z_i is,

$$eQ = -e \langle \Psi | \frac{3 \cos^2 \theta - 1}{r^3} | \Psi \rangle + e \sum_i Z_i \frac{(3 \cos^2 \theta - 1)}{r_i^3} \quad (85)$$

Owing to the nature of the operator only electrons with $l \neq 0$ need be considered. The field gradient reflects the charge distribution around the quadrupolar nucleus in the solid. For a spherically symmetric charge distribution as in Cl⁻, $e^2 Q q$ is zero since $q = 0$. It is clear therefore that if the site symmetry is high and the charge distribution is spherically symmetric no NQR signal will be observed. Discussions on ionic character of bonds have thus been developed using the fact that a covalently bonded nucleus in a molecule might have a q value which is non-zero while the ionic form has $q = 0$. The well-known Townes-Dailey approach²⁵ has been extensively employed in the literature for relating observed NQR frequencies to bond character. We thus have for a X-Cl bond

$$e^2 q Q = -(1-s)(1-i) e^2 q_0 Q \quad (86)$$

where s is the percentage s-character of the bond while i is its ionic character. q_0 is the atomic field gradient and the above equation assumes X to be less electronegative than Cl. On the other hand if X is more electronegative,

$$e^2 q Q = -(1-s) [1 + i(1 + 2\epsilon)] e^2 q_0 Q \quad (87)$$

where $\epsilon \sim 0.25$. It must be remembered that equations (86) and (87) and similar equations given in the literature are of an approximate nature only. Attempts at more exact calculations of q values in crystals have met with severe set backs due to the fact that the field gradients are considerably modified by deformation of closed shells by the nuclear quadrupole moment itself. This phenomenon was originally pointed out by Sternheimer and it is therefore necessary to write the observed e^2QQ as

$$e^2qq = e^2Qq_{\text{val}} (1 - R) + e^2Qq_{\text{latt}} (1 - \sqrt{\omega}) \quad (88)$$

q_{val} is the contribution from the valence electron with $l \neq 0$ (p electrons contribute most) while q_{latt} is the contribution from other charges external to the valence electron. R and $\sqrt{\omega}$ are Sternheimer shielding factors which are theoretically evaluated. $|R|$ lies in the range 0 to 1 while $\sqrt{\omega}$ values may be as high as 100! Due to the uncertainty in the actual R and $\sqrt{\omega}$ values to be used in crystals sophisticated discussions relating wave functions to q are to be viewed with caution. However, the symmetry axes of q can be accurately determined and such studies have proved a valuable supplement to x-ray studies²⁶. Phase transitions in solids involving changes in site symmetry around the quadrupolar nucleus can be followed by NQR. A good example of this type of work is the investigation of Nb^{93} NQR in KNbO_3 by Cotts and Knight²⁷.

The temperature and pressure dependence of NQR frequency has been studied theoretically and experimentally in several solids²³. The increase in amplitude of molecular motion in solids with temperature affects the NQR frequency due to the averaging of the electric field gradients which may be of both intra and intermolecular origin. The theories of Bayer, Kushida et. al and others relate the various modes of torsional vibrations of the molecule in solids to the change in NQR frequency. Similarly, volume changes due to temperature and pressure changes can also contribute to the observed frequency variation.

XI. ELECTRON PARAMAGNETIC RESONANCE

Although EPR is based on principles very similar to those of NMR a fundamental aspect in which it differs is in the orbital contribution to electron paramagnetism. The total magnetic moment of the electron is due to its orbital and spin parts. We shall refer to the total effective spin of the electron as \vec{S} . The interaction between the electron moment and the externally applied uniform field H_0 in a solid can be then written as

$$H_{\text{Zeeman}} = \beta \vec{H}_0 \cdot \vec{g} \cdot \vec{S} \quad (89)$$

where \vec{H}_o is a row vector, g is a matrix and \vec{S} is a column vector and β is the Bohr magneton. g is a tensor, in general, and after suitable transformation has the principal components g_{xx} , g_{yy} and g_{zz} . In case $g_{xx} = g_{yy}$ it is referred to as g_{\perp} and $g_{zz} = g_{\parallel}$. In solids one finds the g tensor to be anisotropic unlike solutions where $g = 1/3 (g_{xx} + g_{yy} + g_{zz})$. The EPR spectrum of a solid can be obtained in principle in the same manner as for NMR. Due to the larger magnetic moment of the electron as compared to nuclei the resonance frequency at a field of 3500 gauss for the electron is 10 kHz. It is therefore convenient to use microwave techniques. Wave guides, klystrons and sample cavities are commonly employed in EPR experimental arrangements. The observed spectra in solids frequently reveal rather narrow lines due to exchange narrowing. In some cases, one finds complex multiplets due to the interaction between the electron moment and various nuclear moments nearby. There are two principal types of magnetic interactions to be considered in the EPR of solids: (i) the "Fermi contact" interaction and (ii) the dipolar interaction. The former was discussed in connection with the Knight shift in metals. The latter is similar in form to the dipolar term for NMR in solids.

$$\underline{H}_{\text{Fermi}} = \gamma_N g \beta \vec{I} \cdot \vec{S} \delta(\vec{r}) \quad (90)$$

$$\underline{H}_{\text{e.d.}} = -\gamma_N g \beta \left[\frac{\vec{I} \cdot \vec{S}}{r^3} - \frac{3(\vec{I} \cdot \vec{r})(\vec{S} \cdot \vec{r})}{r^5} \right] \quad (91)$$

The transitions between the various energy levels can be interpreted in terms of the total Hamiltonian

$$\underline{H} = \underline{H}_{\text{Zeeman}} + \underline{H}_{\text{Fermi}} + \underline{H}_{\text{e.d.}} \quad (92)$$

Other terms such as nuclear Zeeman and quadrupole interactions may also be included, if necessary. In triplet state studies electron-electron interactions may also have to be included. Normal EPR transitions correspond to $\Delta m_s = \pm 1$. EPR studies on a variety of paramagnetic solids have been undertaken. To mention a few: Stable organic free radicals, inorganic complexes, irradiated solids etc. Both single crystals and powder specimen have been employed. The data obtained for single crystal studies can be better interpreted in terms of orientations of the g tensor and hyperfine interaction tensor axes. A form of the Hamiltonian that is often employed is

$$\underline{H} = \beta \vec{H}_o \cdot \underline{g} \cdot \vec{S} + \vec{S} \cdot \underline{T} \cdot \underline{I} \quad (93)$$

\underline{T} is the hyperfine interaction tensor. This tensor consists of both isotropic and anisotropic parts. The principal axes of the g tensor need not coincide with the principal axes of the T tensor although in many cases it does so. Referring to the isotropic

part of T as a and the anisotropic part as T' we have in the common axes case

$$\vec{S} \cdot \vec{T} \cdot \vec{I} = A \hat{S}_x I_x + B \hat{S}_y I_y + C \hat{S}_z I_z \quad (94)$$

here,

$$A = a + T'_1; \quad B = a + T'_2 \text{ and } C = a + T'_3 \quad (95)$$

Further $T'_1 + T'_2 + T'_3 = 0$ since T' is a traceless tensor. The analysis of the experimental spectrum is carried out in terms of the principal components of g and the hyperfine coupling constants A , B and C . A number of γ -irradiated carboxylic acid radicals have been studied by McConnell and coworkers²⁸ and the orientation of the radicals in the solid determined. In the EPR study of transition metal ions and complexes ligand field theory provides an insight into the energy level diagrams for the complex and depending upon the electron configuration of the transition metal ion and the symmetry and strength of the ligand field the value of g varies. Spin-orbit coupling modifies the g value and it is often convenient to use the effective spin Hamiltonian described in terms of \vec{S} . Detailed discussions on this aspect are available in the literature²⁹. For triplet states the interaction term of the form

$$H_D = \vec{S} \cdot D \cdot \vec{S}. \quad (96)$$

has to be considered. D refers to the electron-electron dipole interaction tensor. Transitions with $\Delta m_s = \pm 2$ are now possible besides the usual ones with $\Delta m_s = \pm 1$. The EPR studies of Hutchinson and Mangum³⁰ on the photo-excited triplet state of naphthalene deserve to be mentioned here.

EPR of paramagnetic ions in diamagnetic host crystals has been studies extensively. The creation of vacancies due to charge compensation leads to complicated spectra. Fe^{3+} and Mn^{2+} have been especially popular in this field. The original work of Watkins³¹ may be referred here. Paramagnetic centres such as V_k , F etc. have been extensively studied by EPR with a view to determine their orientations. The availability of higher resolution with electron-nuclear double resonance (ENDOR) techniques has led to more accurate sets of data on these centres. A detailed picture of these centres has thus emerged³². The study of weak unresolved hyperfine lines due to ligand magnetic nuclei have also been observed by ENDOR techniques in recent years. The interpretation of these data in terms of the electronic structure of these systems shows clearly that theory and experiment have been blended harmoniously.

XII. CONCLUDING REMARKS

There are several other areas of applications of magnetic resonance to solids such as NMR in paramagnetic substances, anti-ferromagnetic and ferromagnetic substances, dynamic polarization, to name a few, which have not been discussed here. It is, however, clear that magnetic resonance methods offer a valuable approach to the study of solids in general and solution of problems in solid state chemistry in particular.

REFERENCES

1. E.R. Andrew, "Nuclear Magnetic Resonance", Cambridge University Press, 1955.
2. A. Abragam, "The Principles of Nuclear Magnetism", Oxford, 1961.
3. C.P. Slichter, "Principles of Magnetic Resonance", Harper and Row, 1963.
4. A. Carrington and A.D. McLachlan, "Introduction to Magnetic Resonance", Harper and Row, 1967.
5. H.G. Hecht, "Magnetic Resonance Spectroscopy", John Wiley and Sons, 1967.
6. D. Ibers and W.C. Hamilton, "Hydrogen Bonding in Solids", Benjamin, 1968.
7. F. Jona and G. Shirane, "Ferroelectric Crystals", Pergamon Press, 1962.
8. R. Blinc, "Magnetic Resonance in Hydrogen Bonded Ferroelectrics", in Advances in Magnetic Resonance, Ed. J.S. Waugh, Vol. 3, pp 141-204. Academic Press, 1968.
9. E.R. Andrew and R.G. Eades, Proc. Roy. Soc. (Lond), A218, 537, 1953.
10. G. Benedek, "Magnetic Resonance at High Pressures", Interscience, 1963.
11. M.H. Cohen and F. Reif, Solid State Physics, 5, 321 (1957).
12. R.V. Pound, Phys. Rev., 79, 685 (1950).
13. "Magnetic Resonance and Relaxation", Ed. R. Blinc, North Holland, 1967; see especially pp. 708-774.

14. W.D. Knight, Solid State Physics, 2, 93 (1956).
15. "Nuclear Magnetic Resonance and Relaxation in Solids", Ed. L. Van Gerven, North Holland, 1965.
16. T.J. Rowland, "Nuclear Magnetic Resonance in Metals", Vol. 9 of "Progress in Materials Science", Ed. B. Chalmers, Pergamon, (1961).
17. N. Bloembergen and T.J. Rowland, Phys. Rev., 97, 1679, (1955); P.C. Lauterber, Phys. Rev. Letters, 25, 1289, (1956); F.K. Kneubuhl, J. Chem. Phys., 33, 1074, (1960); J.L. Ragle, J. Chem. Phys., 35, 753 (1961); E.R. Andrew and D.P. Tunstall, Proc. Phys. Soc. (Lond), 81, 986 (1963); E.A.C. Lucken and D.F. Williams, Molec. Phys., 16, 17 (1969).
18. For references to the early literature, see p. 388 of Ref. 13.
19. J.S. Waugh, C.H. Wang, L.M. Huber and R.L. Vold, J. Chem. Phys., 48, 662 (1968).
20. M.A. Ruderman and C. Kittel, Phys. Rev., 96, 99 (1954).
21. T. Kasuya, Progr. Theoret. Phys., 16, 45 (1956).
22. K. Yosida, Phys. Rev., 106, 893 (1957).
23. T.P. Das and E.L. Hahn, "Nuclear Quadrupole Resonance Spectroscopy", Solid State Physics, Suppl. 1, Academic Press (1958).
24. Y. Morino and M. Toyama, J. Chem. Phys., 35, 1289 (1961).
25. E.A.C. Lucken, "Nuclear Quadrupole Coupling Constants", Academic Press (1969).
26. G. A. Jeffrey and T. Sakurai, "Progress in Solid State Chemistry", Ed. H. Reiss, Pergamon Press, pp. 380-416 (1964).
27. R.M. Cotts and W.D. Knight, Phys. Rev., 96, 1285 (1954).
28. See Chapt. 7 of Ref. 4.
29. J.S. Griffith, "The Theory of Transition Metal Ions", Cambridge (1961).
30. C.A. Hutchinson and B.W. Mangum, J. Chem. Phys., 34, 908 (1961).

31. G.C. Watkins, Phys. Rev., 113, 79 (1959).
32. H. Seidel in "Magnetic Resonance and Radio Frequency Spectroscopy", Ed. P. Averbuch, North Holland, 1969 page 141.

NUCLEAR MAGNETIC RESONANCE IN SUPERCONDUCTORS AND DILUTE ALLOYS

R. Vijayaraghavan

Tata Institute of Fundamental Research

Bombay, India

Nuclear magnetic resonance (NMR) of an isotope occurs at a fixed frequency for a fixed externally applied magnetic field. To be exact, the angular frequency ω is equal to the product of gyromagnetic ratio γ and the magnetic field H_0 . γ is simply the ratio of the magnetic moment of the isotope to the spin of the same. In this paper we will be reviewing the changes in the resonance frequency from a metal to a superconductor and the changes in the resonance line width in dilute alloys of transition element impurities in nonmagnetic metals. The principles of magnetic resonance have been dealt with in the previous chapter by P.T. Narasimhan. In this chapter we shall be concerned with the NMR in superconductors and dilute alloys.

I. NMR IN SUPERCONDUCTORS

The nucleus cannot be studied "bare" in an NMR experiment and is always embedded within an atom which forms part of a substance either as a solid, liquid or gas. If the sample, containing the nucleus, exhibits additional magnetic field, this would add up to the external field, thereby shifting the resonance frequency. In metals, for example, this shift occurs due to the conduction electron paramagnetism. When a metallic specimen is placed in a field H_0 , the electron spins are polarised to give a net field ΔH at the position of nuclei due to s-state hyperfine interaction. The shift, called the Knight shift¹, is therefore proportional to Pauli spin susceptibility.

$$K_s = \frac{\Delta H}{H_0} = \alpha \chi_s \quad (1)$$

The shift, which is large for s-electrons and high z metals, is largely independent of temperature. This has been verified in the case of normal metals¹. The transition metals are characterised by large values of temperature dependent paramagnetic susceptibility and high electronic specific heats. These are attributed to the high density of d electron states which form a narrow band intersecting the Fermi surface. The d spin susceptibility does not give rise to NMR shift directly. However, due to core polarisation, a negative shift occurs², which is proportional to the d spin paramagnetism and is consequently dependent on temperature.

$$K_d = \frac{\Delta H}{H_0} = \beta \chi_d \quad (2)$$

The net shift is the sum of equations (1) and (2). α and β represent the hyperfine fields per s-spin and d-spin respectively. Usually α is very much greater than β and is opposite in sign. The case of platinum is a typical example where both the contributions are important. Platinum, which has an unfilled d band, has an atomic configuration $5d^9 6s$.

So far we have seen that the spin susceptibility, which arises from unpaired spins, is responsible for the observed Knight shift in metals. Let us examine the case of a metal when it becomes a superconductor. According to the BCS theory of superconductivity,³ the spin susceptibility of the conduction electrons and hence the Knight shift will vanish exponentially as the reduced temperature T/T_c approaches zero (Fig. 1). This follows from the fact that all states are chosen to be eigen states of the total spin S , and the ground state expectation value of the spin is zero. The higher states are removed in energy by Δ (Energy gap) and hence 2Δ is required to excite a Cooper pair to states for which $\langle |S_z| \rangle \neq 0$. So the measurement of the Knight shift in a superconductor would give us an idea about the temperature dependence of the spin paramagnetism.

Experiment and Results. NMR studies below T_c require the use of particles, small enough to permit complete penetration of the magnetic field. Transition metal alloys are brittle which makes possible the preparation of fine particles of micron size by use of an automatic agate mortar and pestle. Particles of V₃Si and V₃Ga, prepared in this manner, are evidently small enough to permit complete flux penetration at fields of 14 kilogauss. The susceptibility and the shifts in V₃Ga at various temperatures⁴ are given in Table 1 ($T_c = 16.5^\circ K$).

The table reveals certain striking facts. (i) The change in the Knight shift, below T_c , corresponds to a larger internal field, in striking contrast to the behaviour in Sn, Hg and Al where a small reduction (about 30%) has been observed⁴ in the superconduct-

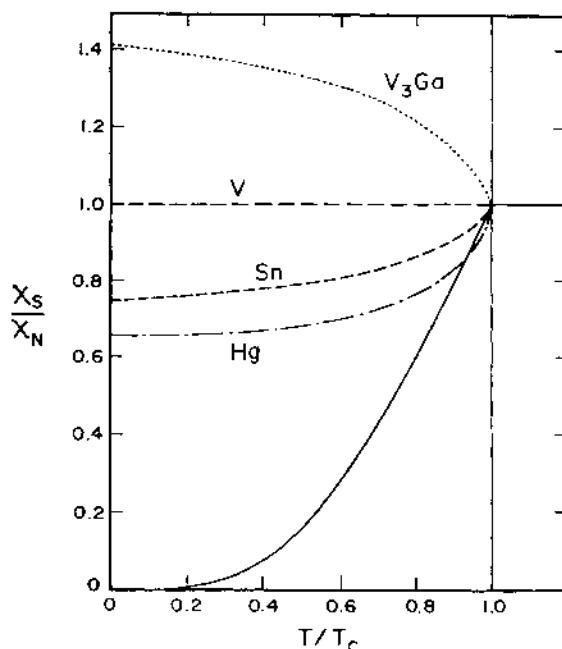


Fig. 1. Ratio of the Spin Susceptibility in superconducting and normal regions as revealed by the Knight shift measurements in the two regions. Thick line is according to Yosida's theory.

Table 1

Temp.	Susceptibility $\times 10^4$ emu/mole	V^{51} Shift in %	Ga^{71} Shift in %	Remarks
300°K	9.8	0.58	-0.36	Normal state
200°K	10.2	0.55	-0.45	"
100°K	12.8	0.50	-0.60	"
20°K	15.6	0.45	-0.80	"
1.8°K	-	0.65	-0.20	Superconducting state

ing region. (ii) The change is in the positive direction, opposite to diamagnetic shielding resulting from partial flux exclusion.

(iii) The magnitude at V⁵¹ and Ga⁷¹ sites are different and hence the result is microscopic. (iv) The shift is more positive at T→0 than it is at any temperature in the normal state, contrary to the BCS theory.

Theory and Discussion. Many suggestions have been made to explain these disturbing results. One of them⁵ is that there is a contribution to the Knight shift, orbital in origin, which does not change in the superconducting state. The other is spin-orbit scattering⁶ by an impurity or surface irregularities in small specimens. The latter has been found to be specifically valid for heavy metals like tin and lead and recent experiments^{7,8} have conclusively proved that these are, in fact, the dominant mechanism for the observation of the residual Knight shifts in Sn¹¹⁹ and Pb²⁰⁷. Because of the spin-orbit scattering by boundary atoms or impurity, the wave functions are no longer eigen functions of S. If ξ_0 is the correlation distance,

$$\frac{\xi_0}{l_{s.o.}} \gg 1 \text{ then } \frac{\chi(0)}{\chi(T_c)} = \frac{K(0)}{K(T_c)} \approx 1, \quad (3)$$

$$\frac{\xi_0}{l_{s.o.}} \ll 1 \text{ then } \frac{\chi(0)}{\chi(T_c)} = \frac{K(0)}{K(T_c)} \approx 0. \quad (4)$$

Reasonable values of $l_{s.o.}$ give results satisfying equation (3) in the case of Sn and Pb, while, in the case of Al, where equation (4) is applicable, the residual Knight shift⁹ indeed goes to zero in accordance with the BCS theory.

We shall now examine the orbital contribution to the susceptibility and hence to the Knight shift K_{orb} . Kubo and Obata gave general theoretical treatment of the magnetic susceptibilities of metals. They treat the case of conduction electrons in an external field H. Apart from the usual Landau (χ_L) and Pauli (χ_p) terms, they get

$$\chi_{\text{orb}} = \beta^2 \sum_{n \neq n'} \int \frac{d^3 k}{(2\pi)^3} \frac{f(E_n) - f(E_{n'})}{E_n - E_{n'}} \langle n k | L | n' k \rangle \langle n' k | L | n k \rangle. \quad (5)$$

Here k is the conduction electron wave vector, n represents the other quantum numbers of the electronic state, E_n is its energy, and $f(E)$ is the Fermi function. This happens to be the orbital susceptibility. The points to be noted are (i) the matrix element exists only for $k = k'$, since the wave functions are Bloch functions and L, when properly defined, has the periodicity of the lattice. (ii) In tight binding approximation, L has matrix elements between n and n' which differ only in m_l, S states, of course, give vanishing matrix elements. Contributions from filled to filled states or from

unfilled to unfilled states are zero. Therefore only contribution to χ_{orb} will come from filled to unfilled states i.e., in our case, from the vanadium 3d band. This is the metallic analog of the familiar Van Vleck second order paramagnetic susceptibility in crystals. An order of magnitude estimate for $\chi_{\text{orb}} = 2 \beta^2 N A (0|L_z|u)^2 / \Delta E$ where N is the number of d electrons for 3 V atoms and ΔE the mean energy separation of filled levels ($0|$ from unfilled levels $|U\rangle$) for which matrix elements of L_z exist, gives a value of $\chi_{\text{orb}} = 6 \times 10^{-4}$ emu/mole for $N = 12$, $(0|L_z|u)^2 = 2$ and $\Delta E = 2.5$ eV. In principle, detailed knowledge of band structure would be necessary to determine χ_{orb} which at first sight, may look prohibitive. However, in practice, it is possible to disentangle the various contributions to susceptibility using the model given by Clogston et al.⁴ This susceptibility is proportional to the product $N_0 N_u$ i.e. number of electrons occupied multiplied by number unoccupied. Hence it would be very large for vanadium and small for platinum.

$$\frac{\chi_{\text{orb}}(\text{V})}{\chi_{\text{orb}}(\text{Pt})} = \frac{(N_0 N_u) \text{V}}{(N_0 N_u) \text{Pt}} = \frac{24}{2} = 12 ;$$

this is in good agreement with the experimentally determined ratio of 8. In platinum, d band spin susceptibility is overwhelming. We write now for the case of $V_3\text{Ga}$,

$$\chi = \chi_s + \chi_d + \chi_{\text{orb}} + \chi_{\text{dia}}, \quad (6)$$

$$K = \alpha \chi_s + \beta \chi_d + \gamma \chi_{\text{orb}}. \quad (7)$$

An estimated upper limit on χ_s is $\chi_s = 0.30 \times 10^{-4}$ emu/mole, corresponding to $K_s = 0.1\%$. In the tight binding approximation, $K_{\text{orb}} = 2 \chi_{\text{orb}} A^{-1} \langle r^{-3} \rangle$ where r is the single radial wave function belonging to n and n' . K_{orb} turns out to be 0.70% and K_d is -0.18% . If the contributions from K_s and K_d were to diminish due to spin pairing as a consequence of BCS theory, the Knight shift will at most change by 0.1% , when $V_3\text{Ga}$ becomes superconducting. The positive increase in the Knight shift shows that the contribution from the d spin susceptibility greatly diminishes, resulting in the negative shift going down. Therefore it would appear that d electrons are the superconducting electrons.

The experimental evidence for the existence of large χ_{orb} in $V_3\text{Ga}$ and $V_3\text{Si}$ resolves an apparent anomaly in the relation between the specific heat and the susceptibility of these metals. In the presence of a net attractive electron-electron interaction, a simple theory⁴, neglecting retardation effects, gives for the spin susceptibility.

$$\chi_{\text{spin}} = \chi_{\text{sp. ht}} [1 + N(0) V]$$

where V is the interaction parameter, appearing in the BCS theory. Consequently, for superconductors $\chi_{\text{spin}} < \chi_{\text{sp.ht}}$ whereas experimentally $\chi_{\text{meas}} > \chi_{\text{sp.ht}}$ for $V_3 X$ alloys. This anomaly is removed

Table 2

$R = \chi_d(0)/\chi_d(T_c)$	$\chi_{\text{meas}} \times 10^4$	χ_{orb}	$\chi_{\text{sp.ht}}$	$N(0)V$ from NMR	$N(0)V$ from BCS
$V_3\text{Ga}$ $0 \leq R \leq 0.25$	17.9	5.2-7.5	14.0	0.2-0.4	0.35
$V_3\text{Si}$ $0 \leq R \leq 0.25$	15.7	"	10.7	0.1-0.35	0.35

by subtracting χ_{orb} from total susceptibility (see Table 2). The value of $N(0)V$ thus obtained, is in reasonable agreement with that deduced from BCS theory⁴.

Conclusions. In transition metals, there is an orbital contribution to the susceptibility which is nearly equal in magnitude to the spin susceptibility. This arises from states far removed from the Fermi surface so that its value does not change appreciably with temperature. Hence there is no reduction in the orbital contribution when the metal becomes superconducting. The Knight shift arising from the Pauli spin susceptibility does decrease in the superconducting region and actually becomes zero in very pure aluminium as $T \rightarrow 0$. In this light metal, both orbital paramagnetism and spin-orbit effects are negligible. This experimental result vindicates the BCS theory. In transition metals, the residual Knight shift is entirely due to the orbital paramagnetism. In heavy nontransition metals, the spin-orbit effects are responsible for nonvanishing Knight shift as in the cases of tin and lead.

II. NMR IN DILUTE ALLOYS

Introduction. We shall be discussing here the experimental observation obtained from NMR on the effects of transition metal impurities in non-transition metal matrices. One area of interest is whether the magnetic localised states are formed or not. The alloys which exhibit a local moment show a minimum in the resistivity at low temperature, the depth of which and whose position as a function of temperature, is known to depend on solute concentration. In 1964, Kondo¹⁰ carried the perturbation theory of the exchange scattering ($\sim J$ S.s) of the conduction electrons by a single localised spin S beyond the first Born approximation. In the third order perturbation, $\log T$ term appears in the resistivity due to spin-

dependent scattering, which is linear both in the concentration c of localised spins and the exchange coupling J . When J is negative (~ 0.2 eV) corresponding to an antiferromagnetic coupling between localised and conduction electron spins, the resistivity minimum occurs at

$$T_{\min} = c^{1/5} (|\rho_i| / 5a)^{1/5}$$

where ρ_i is spin dependent resistivity due to the impurity and 'a' is a constant depending on the host lattice.

It has been suggested that at sufficiently low temperature, the Kondo effect may lead to the spin of the localised moment of the magnetic impurity being totally compensated by the spin cloud of conduction electrons which are polarised near the impurity atom, leading to a zero spin-state of the impurity plus conduction electron polarisation cloud as a whole. From this view, there would be a phase transition, rather a broad one, at some critical temperature T_k below which the net localised moment is either zero or greatly reduced from its high temperature value. High magnetic fields and higher temperatures can gradually break up this quasi bound state. Rough estimates show that the transition temperature T_k is approximately given as¹¹

$$T_k \approx T_0 \exp - (1/N(0) |J|) \quad (8)$$

where T_0 is of the order of Fermi temperature of the conduction electrons and $N(0)$ is the density of bound states of one spin orientation at the Fermi surface. Since the exponent varies so rapidly for small changes in J , modest variations in $N(0) |J|$ could sweep T_k from milli degrees to thousands of degrees. Hence some of the impurities which do not show moment at room temperatures may have their spin compensated at a higher temperature. We will not discuss here the relevant theory and experimental data but refer the reader to an excellent review article¹² where its important consequences in resistivity, specific heat, susceptibility and thermoelectric power measurements have been described.

From the NMR point of view, we will be interested in seeing the changes in the line width and the shift when one measures the resonance at temperatures lower than T_k . This technique is well suited for studies of dilute alloys since those nuclei (host lattice) in the vicinity of a given impurity atom are sensitive to the local conduction electron polarisation, caused by the scattering from the spin dependent impurity potential. The resulting spin oscillations can be probed by the nuclear spins I of the host metal via the hyperfine interaction $A I.S$, resulting in shift and the width of the resonance line. We will discuss here the NMR data of Cu^{63} in dilute alloys of Mn, Fe, Cr and Co in copper metal¹³, and V^{51}

resonance in Au V alloys.¹⁴ The amount of impurity in the case of copper has been less than 0.1 at.% to prevent impurity-impurity interaction and the copper metal was 99.99% pure.

Experimental Details. The alloys are melted and annealed at 800°K for 2 or 3 days followed by quenching into water. Metallographic examinations made on the most concentrated alloys showed no indication of precipitation or cluster formation on the grain boundaries. The low temperature apparatus can be varied from 0.03°K onwards. The line widths due to impurity are calculated by the method described by Heeger et al.¹⁵ Normally the width should be directly proportional to external field H_0 and inversely to temperature as in susceptibility.

Discussion. The Kondo temperatures¹³ for Fe in Cu are $\sim 15^{\circ}\text{K}$, Cr in Cu $\sim 1.4^{\circ}\text{K}$, Co in Cu $\sim 1000^{\circ}\text{K}$, and Mn in Cu $\sim 0.05^{\circ}\text{K}$. We will discuss here mainly the results of NMR in Cu Fe alloy. The impurity contribution to the line width at temperatures less than T_k are shown in Fig. 2. If no quenching or spin compensation took place, the static broadening would be proportional to $B_s (\mu_{\text{eff}} H_0/kT)$. With $B_s, s = 2$, as Brillouin function, B_2 would saturate in fields of less than 1 kilogauss at $T = 0.02^{\circ}\text{K}$. From high temperature data ($T > T_k$), the Cu⁶³ NMR line width is approximately given by $120 B_2$ g which would be about 120 gauss at $T = 0.1^{\circ}\text{K}$ for fields higher than 1 kilogauss. Experimentally the most important features are:
(i) The width is narrower by one to two orders of magnitude. (ii) A field dependent line width in the temperature region where free spin behaviour would be independent of field. (iii) A temperature-independent line width at all fields. The non-zero intercept at $H_0 = 0$ is due to impurity-impurity effects even in such dilute alloys. The impurity contribution to line width is proportional to the impurity susceptibility. Therefore, the impurity contribution to the width has the same temperature dependences as $\langle S_z \rangle$. Hence the spin polarisation around the impurity persists except that the magnitude scales with the true $\langle S_z \rangle$, rather than its free spin counterpart. The ratio $\langle S_z \rangle / S$ can be as low as 20% at 10 kilogauss at 0.5°K , illustrating clearly the gross reduction in the Fe moments below Kondo temperature.

Due to small T_k ($\sim 1.4^{\circ}\text{K}$), the Cr impurity is only 47 ppm in copper and smaller fields reproduce the same results as Fe in copper¹³.

No field dependence was observed in the 0.040 at.% Cu Co alloy. This would imply a T_k higher than 1000°K, consistent with other data.

However, for Mn in Cu, T_k is around 0.05°K . With such low

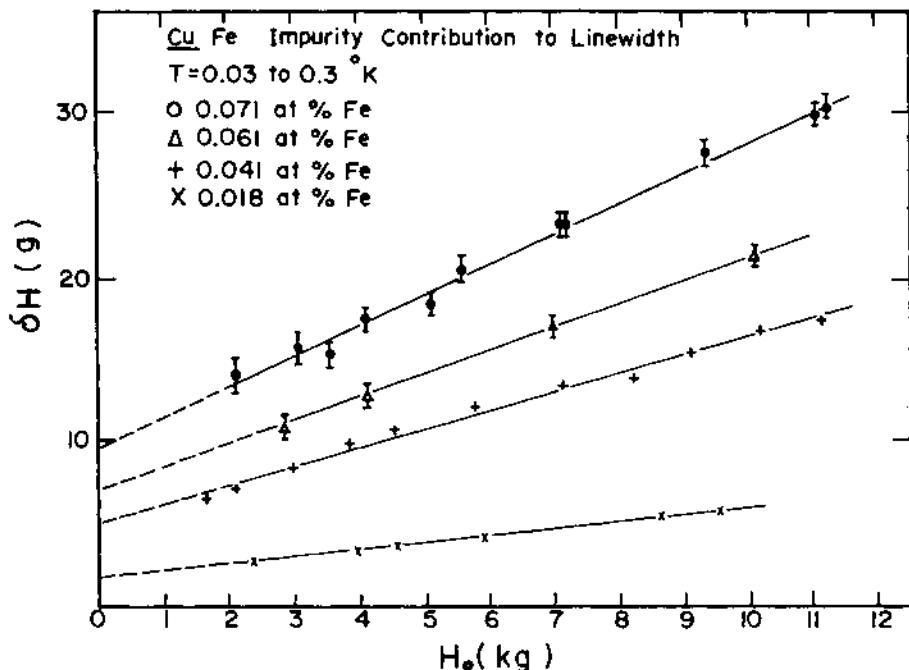


Fig. 2. The impurity contribution to the Cu^{63} linewidth.

temperature, magnetic fields greater than 1 kilogauss can disrupt the condensed state ($\mu H > K T_k$). Because of Mn-Mn interactions, self quenching of the spin compensated state is also to be expected, leading to short range magnetic order. As the temperature is lowered, each Mn moment will saturate along a local axis of quantisation, determined by the local magnetic field, i.e., by the external field and the local field due to other Mn spins in the vicinity. The Mn spins will thus saturate ($\langle S_z \rangle = S$) along a local axis which varies in direction randomly throughout the alloy. For very low temperatures, the magnitude of spin density oscillations will saturate at a maximum value independent of external field giving about 55 gauss in width for 3 to 10 kilogauss fields¹³.

Knight Shift of V⁵¹ in Au V Alloy. The magnetic properties of dilute solid solutions of vanadium in gold occupy a middle ground between the localised magnetic moments of chromium in gold and the non-magnetic impurities of titanium in gold. From resistivity data¹⁵ it has been interpreted as the spin compensated system with a high $T_k \sim 290^\circ\text{K}$ and this has been supported by the susceptibility. Another remarkable feature is the high amount of vanadium concentration (0.1 to 10 at.%) that can be safely alloyed and studied in this high T_k system. We report here the NMR in the impurity itself, V⁵¹, as direct evidence for spin compensation.

The V⁵¹ NMR was detected¹⁴ at very low temperatures (range 1 to 4°K) using very high fields (~40 kilogauss). In the 0.1 at.% sample, the Knight shift was - 1.5% and had a width of 0.7% of the applied field. This negative shift arises from the partially spin compensated d - spin magnetisation. Vanadium d spins would produce a shift of -10% in the uncompensated state. The disparity between the observed shift and that due to V²⁺ d-spin, is attributed by Narath *et al.*¹⁴ and Heeger *et al.*¹³ to a partially spin compensated state whose value is necessarily smaller than the ionic value.

On the other hand, (20 at.% V) Au₄V is ferromagnetic¹³. We have seen above, that upto 10 at.% V, there is no magnetic ordering in Au V alloy. These seemingly diverse phenomena may be understood from the study of the magnetic impurity problem in relation to magnetism in transition metals.

REFERENCES

1. C.H. Townes, C. Herring and W.D. Knight, Phys. Rev. 77, 852 (1950).
2. R.E. Watson and A.J. Freeman, Phys. Rev. 123, 2027 (1961).
3. J. Bardeen, L.N. Cooper and J.R. Schrieffer, Phys. Rev. 108, 1175 (1957); K. Yosida, Phys. Rev. 110, 769 (1958).
4. A.M. Clogston, A.C. Gossard, V. Jaccarino and Y. Yafet, Rev. Mod. Phys. 36, 170 (1964) and references therein.
5. R. Kubo and Y. Obata, J. Phys. Soc. Japan, 11, 547 (1956).
6. A.A. Abrikosov and L.P. Gorkov, JETP 15, 752 (1962) and reference therein.
7. F. Wright, W.A. Hines and W.D. Knight, Phys. Rev. Lett. 18, 115 (1967).
8. W.A. Hines and W.D. Knight, Phys. Rev. Lett. 18, 341 (1967).
9. R.H. Hammond and G.M. Kelley, Phys. Rev. Lett. 18, 156 (1967).
10. J. Kondo, Progr. Theoret. Phys. (Kyoto) 32, 37 (1964).
11. J.R. Schrieffer, J. Appl. Phys. 38, 1143 (1967).
12. M.D. Daybell and W.A. Steyert, Rev. Mod. Phys. 40, 380 (1968).

13. A.J. Heeger, L.B. Welsh, M.A. Jensen and G. Gladstone, Phys. Rev. 172, 302 (1968).
14. A. Narath, A.C. Gossard and J.H. Wernick, Phys. Rev. Lett. 20, 795 (1968).
15. K. Kume, Phys. Letters 24A, 743 (1967).

MAGNETOSTRICTION IN MATERIALS

R. K. Ray

Department of Physics

Indian Institute of Technology, Kanpur, India

I. INTRODUCTION

When a magnetic material is under a magnetic field, H , a stress is generated into it, which gives rise to a resulting strain, i.e. change in dimension. We call it magnetostriction. Usually, materials in bulk form exhibit any property as a sum of that property for a number of single crystals; however, if single crystals are studied one could get a clear indication of the intrinsic property - this may reveal important anisotropies for a particular physical property for different crystalline directions. For example, there are 'easy' and 'hard' directions of magnetizations. This 'magnetocrystalline anisotropy' is quite different in the cases of Fe (bcc), Ni (fcc) and Co (hexagonal); the easy direction is along [100], [111] and [0001] and the hard direction is along [111], [100] and [1010] for Fe, Ni and Co, respectively. The reason, at least partly, is in the asymmetry of the overlap of the electron distribution on neighbouring ions. Because of spin-orbit interaction, the charge distribution is not spherical. The asymmetry is tied to the direction of magnetization by the spin orbit coupling; changing the spin direction changes the overlap energy.

The total energy of a ferromagnet depends on the state of strain in the substance. The body will deform spontaneously if that reduces the total energy. The easiest to observe and measure is the linear magnetostriction, $\lambda = \Delta l/l$, i.e. the change in length per unit length. It is also called the Joule effect and is commonly $\sim 10^{-6}$ in magnitude. Throughout this chapter we shall express λ in terms of 10^{-6} units only. Both longitudinal ($\lambda \parallel H$) and transverse ($\lambda \perp H$) coefficients have been measured. These

are found to be significant enough towards recognizing characteristic processes of magnetic alignment that occur in several systems of interest. Their role will be discussed in later sections. In comparison to λ , the volume magnetostriiction, $\Delta v/v$, is a quantity of the second order. We would restrict our discussions mostly to linear magnetostriiction. There is also the inverse effect that strain affects the magnetization, the so called Villari effect. For example, longitudinal deformation leads to changes in permeability in the direction of the applied strain. If λ is positive, saturation magnetization, M_s , increases under tension; if λ is negative, M_s decreases under tension.

The study of magnetostriiction has gained importance, for the following reasons: (i) It may throw light on the nature of the internal forces (ii) It may help in investigating magnetization and analyzing magnetization curves. Hopefully, it can give a clear indication of the onset of a new phase or order in a complicated ordering process which otherwise would be revealed only by elaborate investigations, such as by neutron diffraction. Of course, samples have to be suitably chosen and studied for variation of temperature, magnetic field, stress, etc. (iii) It may help in practical aspects. For example, it converts electric pulse to mechanical and vice versa which principle has been utilized in the production of ultrasonic oscillations. Further, the noise in a transformer is initiated by magnetostriuctive vibration of the core and can be greatly reduced by suitable choice of material and design (iv) It can easily indicate, while 'quality controlling' batches of samples of, say, ferrites for a standard requirement of their magnetic property, whether any undesirable characteristic has crept in or not.

II. BEHAVIOUR OF COMMON MATERIALS

Linear magnetostriiction, λ , measured for crystalline directions of Fe, Ni and Co as indicated are shown in Fig. 1. Note that for iron λ is positive for [100], negative for [111] and for the [110] direction it is positive at low fields, next passes through zero becoming negative at high fields. For Ni and Co, λ is negative throughout. These behaviours are associated with the presence of domains, their direction of preferred magnetization and related domain energy. We may correlate magnetization in different crystalline directions in terms of the measurable constants of the materials. In crystals with cubic symmetry, the relation between domain energy and the position of the magnetic vector is given by an expression

$$E = K_0 + K_1 (s_1^2 s_2^2 + s_2^2 s_3^2 + s_3^2 s_1^2) + K_2 (s_1^2 s_2^2 s_3^2)$$

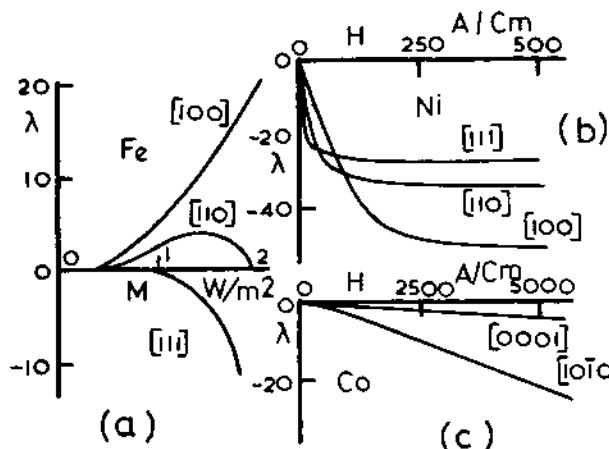


Fig. 1. Magnetostriiction of single crystal iron, nickel and cobalt.
Webster Proc. Roy. Soc., A 109, 570, 1925; Masiyama Sci.
Rep. Tohoku Univ., 17, 925, 1928; Nishiyama Sci. Rep.
Tohoku Univ., 18, 341, 1929.

where S's are the direction cosines for the direction of saturation magnetization, M_s , in the domain relative to the rectangular coordinates axes. One can obtain the values of K_0 , K_1 and K_2 from the experimental M-H curves in terms of the areas corresponding to the stored energies for magnetization along the chosen directions, such as E_{100} , E_{110} and E_{111} . Standard but pioneering work in this connection was done by Honda and Kaya¹, later supplemented by other workers². One can obtain $K_0 = E_{100}$, $K_1 = 4(E_{110} - E_{100})$, $K_2 = 27(E_{111} - E_{100}) - 36(E_{110} - E_{100})$. For Co, hexagonal below 477°C, $E = K_0 + K_1 \sin^2 \theta + K_2 \sin \theta$; $\cos \theta$ is the direction cosine of M_s relative to the principal axis of the crystal. We indicate the variation of K_1 and K_2 with temperature in Fig. 2. Note that the changes are remarkable. Fig. 3 gives the magnetostriction for polycrystalline material. Here again λ is negative for Co and Ni for all values of H whereas for Fe it registers a positive value for low H , goes through a maximum to zero, to a negative saturation value. Further, for Fe, tension makes λ go more negative for any value of H . For Ni, tension lowers magnetostriction and compression increases it, for any value of H .

Magnetostriction also depends on composition. If heat treated in hydrogen, highly pure Fe, free from O, C, N and S, gave very low hysteresis ($19 \text{ J/m}^3/\text{cycle}$), 5% of that found in commercial unalloyed iron sheet material. It had very high initial relative permeability, $\mu_r \sim 19000$ and maximum relative permeability, $\mu' \sim 280,000$. Becker³ has given a theoretical expression for initial relative permeability in terms of saturation magnetization,

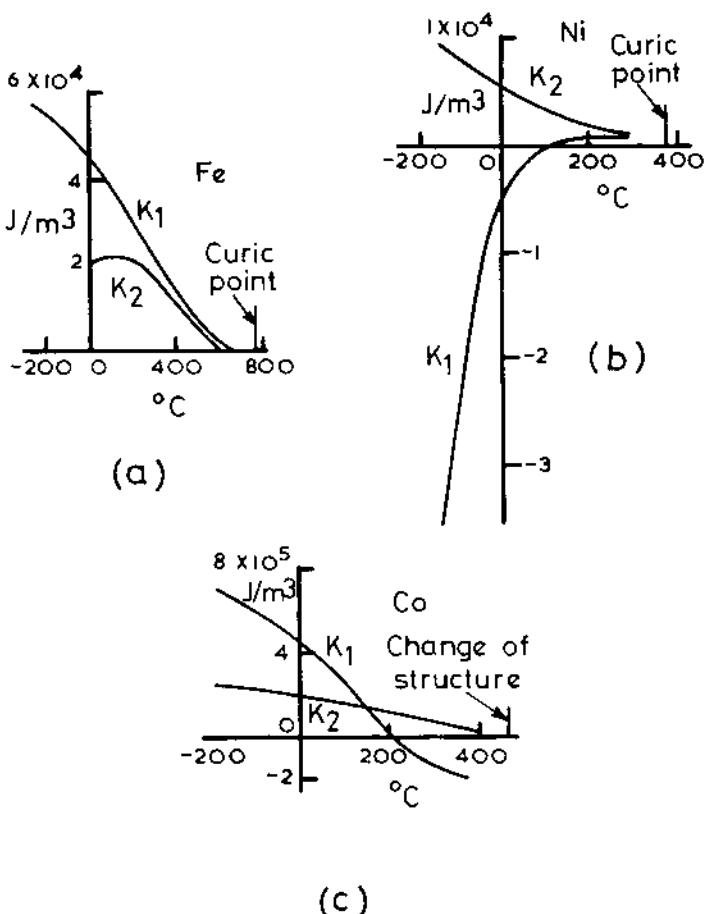


Fig. 2. Magnetic anisotropy constants of iron, nickel and cobalt in relation to temperature.

saturation magnetostriction and internal stresses. The maximum relative permeability is estimated as $\mu' = (\mu_r)_{\max} = (2M_s^2/3\pi\mu_0)\lambda_s^2 E_e$ where E_e is the elastic constant. At room temperature $M_s = 2.158 \text{ W/m}^2$, $\lambda_s = 20 \times 10^{-6}$ and $E_e = 20 \times 10^{10} \text{ N/m}^2$. Experimental value of the maximum relative permeability is $\sim 14,000$ for annealed crystals compared to the calculated estimate of 10,000. So, the quantitative result is of the right order of magnitude. For the [100] direction the highest reported μ' was 1.43×10^6 for magnetic induction $B = 1.75 \text{ W/m}^2$. For Fe with 6.5% Si, μ' rose to as high as 3.8×10^6 when the specimen was heat treated in hydrogen. Further, as temperature rises K_1 and K_2 decrease thus leading to increase in permeability since, $K \rightarrow 0$, all the directions tend to become 'easy'

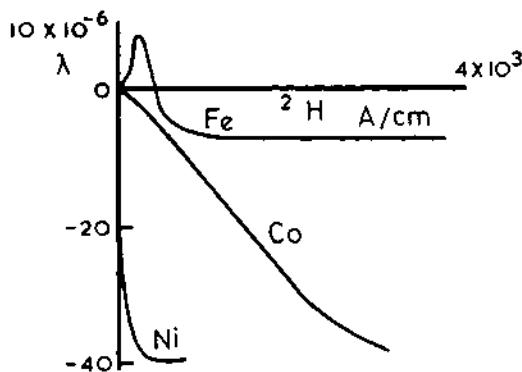


Fig. 3. Magnetostriiction of polycrystalline iron, nickel and cobalt.
Kornetzki Zeits. f. Phys. 87, 560, 1933; Masiyama Sci.
Rep. Tohoku Univ. 20, 544, 1931; Nishiyama Sci. Rep.
Tohoku Univ. 18, 341, 1929.

direction for magnetization.

Alloys of Iron. If we add 40% Co in Fe we get magnetization to increase a great deal. Adding Si ~ 6%, the resistivity increases by a factor of 8 compared to that in pure iron ($10\mu\text{ohm cm}$). Silicon steel, for transformer core use has still now no competitor at normal power frequency. But K_1 falls as Si concentration rises and there is also considerable effect of Si on the magnetostriiction as shown in Fig. 4. At about 7% Si, λ is very low and at 8.37%, λ is substantially negative. Explanation is as yet not satisfactory.¹

On account of magnetostriiction, a transformer core changes dimension with double the frequency of supply to the core. This is

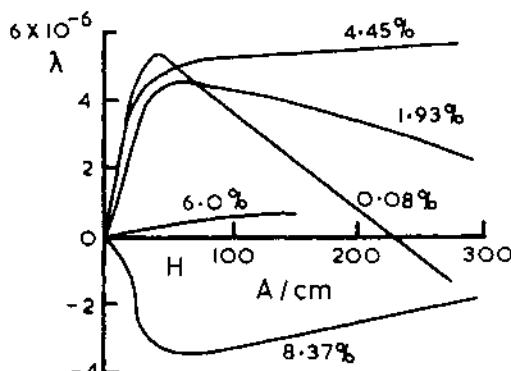


Fig. 4. Magnetostriiction of the iron silicon alloys. Schulze
Zeits. f. Phys. 50, 448, 1928.

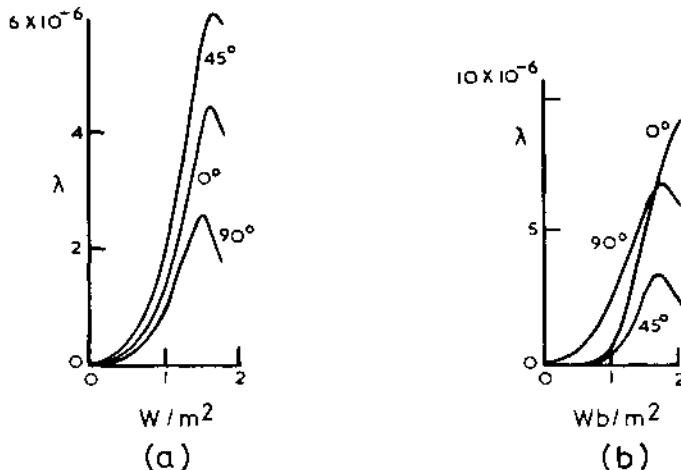


Fig. 5. (a) Magnetostriiction for a sample of 1.88% silicon sheet steel showing variation with different direction relative to the rolling direction (b) Magnetostriiction of cold rolled 3% silicon steel sheet showing directional variation with angle relative to rolling direction, Brailsford and Martindale Jour. I.E.E. 89, Part I, 225, 1942.

an initiating cause of the noise emitted by transformers and may set a limit to the working induction in the cores. So, λ of sheet materials is of considerable practical importance. In 1.88% Si sheet steel (Fig. 5a), the hot rolled material has preferred direction for magnetization relative to the rolling direction. In sharp contrast, cold rolled 3% Si sheet steel gives a different behaviour for the dependence of preferred direction relative to the rolling direction (Fig. 5b). It is further interesting to conclude, from the absence of magnetostriiction observed for values of H below a certain value ($0.4 W/m^2$), that for the direction of rolling itself, at first, the magnetization is by 180° reversal of the domain vectors and thus does not contribute to λ . At higher magnetization

λ is not equal to zero and we infer that 90° turning of domain vectors is occurring. We may thus point out here that a rather important fact about the physical 'micro' process may be sometimes learnt from a careful analysis of the λ curve in favourable circumstances.

In nickel alloys, 80:20 Ni-Fe has very low λ . Fig. 6 gives λ for polycrystalline Ni-Fe. The temperature coefficient of expansion of alloy of 36% Ni in Fe at room temperature is found to be extremely low; this material, called Invar, has practical importance. The electrical resistance is found to be maximum for 30% Ni in Fe. For polycrystalline Ni-Fe alloys, λ changes from positive to negative for Ni beyond 82%. Elman found that suitable heat treatment gave very high initial permeability and saturation

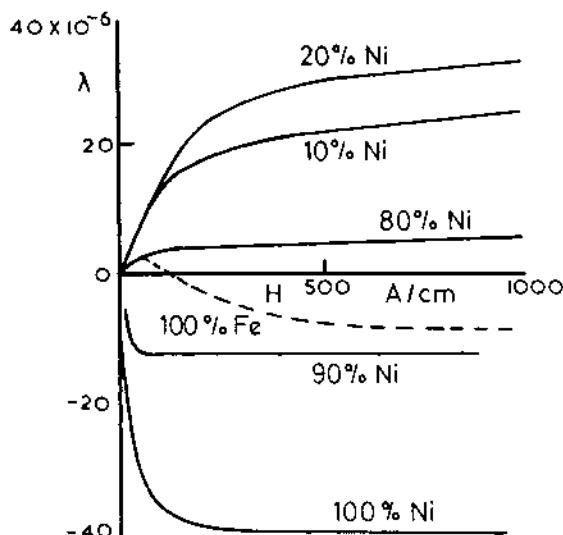


Fig. 6. Longitudinal magnetostriiction of the nickel iron alloys (polycrystals) Masiyama Sci. Rep. Tohoku Univ. 20, 574, 1931.

magnetization for samples which gave very low values of the magnetostriiction. For best magnetic properties, rapid cooling is necessary. For 78.5% Ni, the permeability is reported to be highest and such alloys are called 'permalloys'; λ decreases at such compositions (Fig. 7). The initial permeability is very high for air-quenched samples for Ni-Fe polycrystals and is much higher than that found for slow-cooled samples. We may note⁴ that high permeability is linked with λ going down to zero. $\mu' \sim M_s^2 / \lambda_s^2 E_e$. Bozorth⁵ concludes that high μ' is associated with very low λ and

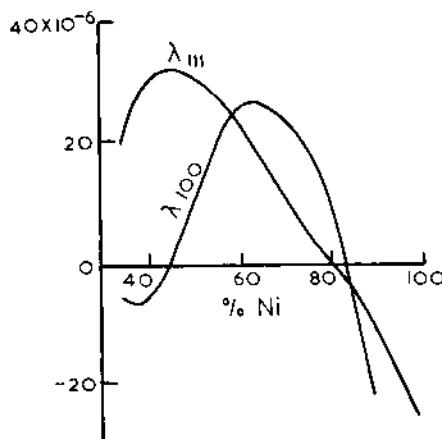


Fig. 7. Saturation magnetostriiction of single crystals of nickel iron alloys Arnold and Elmen Jour. Franklin Inst. 195, 621, 1923; Bozorth and Walker Phys. Rev. 89, 624, 1953.

also low values of the magnetic anisotropy constants.

Ferrites. These commercially important materials have cubic spinel structure or hexagonal structure. Since metallic alloys have at best a resistivity, $\rho \sim 10^{-4}$ ohm cm, large eddy current losses are faced. But the ferrites with useful magnetic properties have $\rho \sim 10^{-10}$ ohm cm and are thus important as cores both in crystal form and ordinary shape without any laminating or powdering. Pure magnetite has $\mu \sim 70$ and $\rho \sim 10^{-2}$ ohm cm. Snoek put divalent metal ions in place of divalent iron ions and synthesized a series of useful magnetic oxides, $M^{2+} (2Fe^{3+})O_4^{2-}$, where M is in 'A' site and Fe is in 'B' sites. We may note that the ferrites of normal structure do not have ferromagnetic properties. However, the inverse spinel structures have magnetic ferrite behaviour ($Fe^{3+}(Fe^{3+}M^{2+})O_4^{2-}$ with M = Mn, Fe, Co, Ni in MFe_2O_4). As in ferromagnetic metals and paramagnetic solids, the magnetic moments due to orbital motion of electrons are quenched by internal fields; i.e. these are not oriented by externally applied fields. In order to understand the picture, let us see the role of exchange interaction between the spins at A and B sites (i.e. AA, BB and AB). The negative interaction AB predominates and spin of A is anti-parallel to that of B site. So the saturation value of spontaneous magnetization of the domains is due to the difference of ionic moments on A and B sites and finally, therefore, only due to the unbalanced metal ion. For Mn ferrite the formula unit has $5\mu_B$ and for Fe, $4\mu_B$ only, where μ_B stands for Bohr magneton. In most ferrites the magnetic anisotropy constant K_1 is negative resulting in the easy direction of magnetization being [111] as in Ni. The Zn ferrite has normal structure and the Zn ions, on A sites, have no magnetic moments; therefore there are no AA or AB interactions. The presence of negative BB interaction alone may give it an antiferromagnetic alignment, but it has not been observed for certain.

Table I gives the properties of some binary ferrites and it is seen that saturation magnetostriction for polycrystalline specimens is negative in all except magnetite. However, in Co ferrite K_1 is strongly positive and [100] is the easy direction of magnetization, as for Fe. The saturation magnetostriction along [100] is negative and large in magnitude. Both Mn and Cu ferrites have large initial permeability and correlates well with low values of magnetostriction.

Mixed ferrites. Useful ferrites are obtained by combining different single ferrites together. Solid solution of different ferrites make possible a wide variation in chemical composition. Specially important are the double ferrites involving Ni and Mn ferrites and oddly enough Zn ferrite which is not even ferrimagnetic.

Table II.1*

Ferrite	$K_1 \times 10^{-4}$ (J/m ³)	Saturation Magnetostriiction ($\times 10^6$)			M_s Wb/m ²	Curie Temp. T_c	Initial Permeability μ_r
		λ_s	λ_{111}	λ_{100}			
Mn Fe ₂ O ₄	-0.28	-5			0.50	300	250
Fe Fe ₂ O ₄	-1.1	+41	+81	-19	0.60	585	70
Co Fe ₂ O ₄	+38	-110	+120	-590	0.53	520	
Ni Fe ₂ O ₄	-0.62	-26			0.34	585	10
Cu Fe ₂ O ₄	-0.60	-10			0.17	455	70
Mg Fe ₂ O ₄	-0.25	-6			0.15	440	10

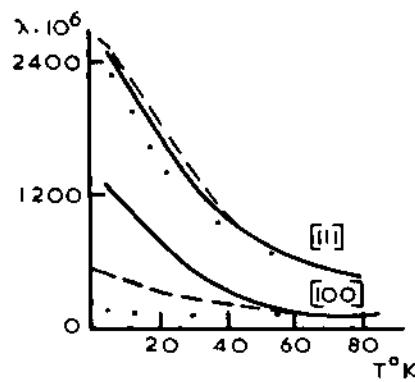
* E.W. Gorter, Proc. I.R.E., 43 (2), 1945, 1955.

Ferrites are sintered at 1200-1400°C and the oxygen content controlled e.g. $(\text{Zn}_x^{2+} \text{Fe}_{1-x}^{3+})_A (\text{M}_{1-x}^{2+} \text{Fe}_{1+x}^{3+})_B \text{O}_4^{2-}$, $0 < x < 1$. The Zn ions of zero magnetic moment increase saturation magnetization value of these mixed ferrites. There exists a need for square shaped hysteresis loop and high resistivity for devices such as magnetic cores and magnetic memory stores in computers and switching devices. Also of great importance are these materials in the case of high frequency systems. The flux density at high frequency in materials is low and the initial permeability is therefore important in controlling suitable amount of flux entering the bulk of the material.

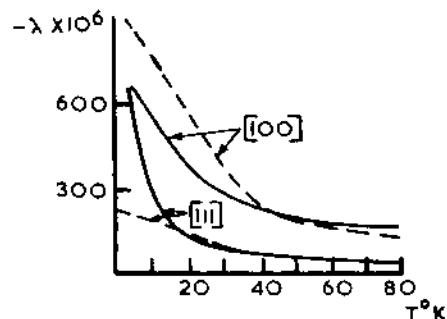
Garnets. The close relation of magnetic properties such as initial permeability or saturation magnetization with magnetostriiction has been indicated for alloys and ferrites. Next we shall discuss some of the recent experiments on magnetostriiction which could be cited as 'test' for various theories of magnetization, as stated in the Introduction. Let us mention the system $\text{Tb}_x \text{Y}_{3-x} \text{Fe}_5 \text{O}_{12}$ (garnet). Saturation magnetization σ and λ are measured down to 4.2° K. x can vary from 0 to 3. It was found that at low temperature, λ rises enormously and sharply with concentration of Tb ions. For Tb concentration, $x = 0.8$, $\lambda = 62$; for $x = 3$, $\lambda = 2180$. These data also may give estimate of magneto-elastic energy provided Young's modulus is known. The $\lambda(T)$ curve indicates a minimum at the compensation temperature. Lower the compensation temperature; i.e. higher the magnetization, sharper

are the minima. For $x > 1$, clear correlation is absent between λ and ϵ in their temperature dependence. For $x = 3$, in 4 - 30°K range, λ increases by 2.5 times while ϵ increases by only 6%. These results do not agree well with the 'single ion model' phenomenological theory of magnetoelastic interaction developed for the Néel ferromagnets by Callen⁷. For further investigation $Tb_3Fe_5O_{12}$ was taken up and longitudinal magnetostriction measured⁸ as a function of H and T in single crystal samples between 80 and 4°K; H varied from 0 to 25 KOe. Again one notes a sharp rise in λ at low temperatures as the magnetoelastic contribution to the anisotropy energy tends to increase sharply with lowered temperature. For [111], the easy direction of magnetization $\lambda \approx -2420 \times 10^{-6}$ and for [100], the hard direction, it is $\sim 1200 \times 10^{-6}$ and does not saturate (Fig. 8a). This figure also compares the data with theory and shows satisfactory agreement for the [111] direction; the data for [100] does not agree so well, perhaps because single ion model is not fully satisfactory. The anisotropy constants cannot be calculated for the lowest temperature region. However, for 80°K, the estimated values are $K_1 = -164$ and $K_2 = -1010$ in units of 10^6 ergs/cc. In contrast, measured data of $\lambda_{111}(H)$ for the samples of $Ho_3Fe_5O_{12}$ is anomalous, perhaps due to the noncollinearity of Ho magnetic moments in the garnet structure. This is not taken into account in the single ion theory. A complementary study of neutron diffraction at 4.2°K established that the magnetic moment of Ho makes $\sim 60^\circ$ with [111] axis. So, perhaps, the increase in λ_{111} is due to disturbance of angular configuration of magnetic moment of Ho; i.e. an exchange magnetostriction occurs. The non-collinearity of the magnetic moments arises apparently only below 15°K. Fig. 8b confirms this, from comparison of data on thermal dependence of λ with theory. The values above 15°K agree. The discrepancy below 15°K is obviously due to the fact that phenomenological theory of magnetoelastic interaction could not apply to noncollinear moment configuration in the ferrimagnet. We may conclude that an enormous rise in λ may mean the onset of a special process of spin alignment.

Helicoidal Spin Structures and Their Effect on λ . One of the most interesting results of the measurements of λ , is obtained from the study of helicoidally ordered spin systems. Neutron diffraction indicates that metals like Dy, and metamagnetic alloys like $MnAu_2$ have helicoidal spin arrangement i.e. associated with magnetization there are layers with ferromagnetic spin ordering and each layer has resultant magnetic moment; these are then related by $\theta \neq 180^\circ$, forming a helicoid (Fig. 9). Nearest neighbour atoms are coupled by ferromagnetic exchange, J. The next nearest neighbour atoms are coupled by antiferromagnetic exchange interaction $-E J$, where E is a positive constant. A comparative study of any such system should go a long way to explain the nature of the helicoid structure. One of the interesting intermetallic alloys is



(a)



(b)

Fig. 8. (a) Temperature dependence of magnetostriiction of terbium iron garnet --- theoretical data ($H=0$); solid line - experimental data ($H=23$ KOe); \bullet experimental data extrapolated to $H=0$. (b) Temperature dependence of magnetostriiction of single crystal holmium iron garnet; --- theoretical data ($H=0$); — experimental data ($H=23$ KOe); Sokolov JETP 25, 986, 1967.

MnAu_2 . The Néel temperature, T_N is 90°C and threshold field, H_{th} is 18 KOe. A study of transverse and longitudinal magnetostriiction may lead to the understanding of the nature of the spin ordering, since both exchange and magnetic interaction forces appear directly in λ . Measurements of λ_\perp and λ_\parallel were made below and above the Néel temperature⁹ and it was observed that λ was small in the anti-ferromagnetic state. There was, however, a sharp contraction of the sample accompanying the transition to ferromagnetic ordering. For λ_\perp , $H > H_{th}$, no singularity was seen near T_N , unlike the behaviour of normal ferromagnets near their T_c . This is obviously because in MnAu_2 , λ arises from the disruption of the helicoid structure. On approaching T_N , ordered distribution of spin is gradually disturbed which reduces λ . Such decrease in λ due to destruction of

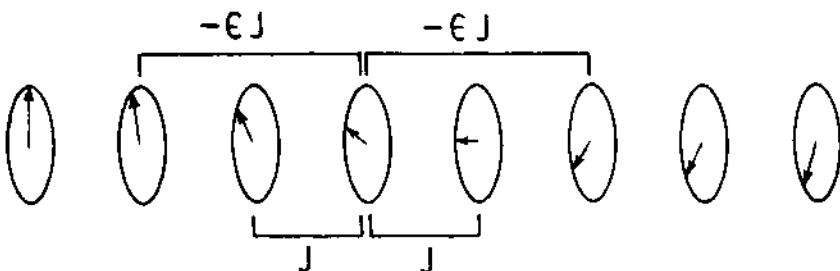


Fig. 9. Helical spin array. Nearest neighbour atoms coupled by ferromagnetic exchange J ; next nearest neighbour atoms are coupled by antiferromagnetic exchange $-\epsilon J$ where ϵ is a positive constant.

helicoid structure was observed at first in Dy samples¹⁰. Further, accompanying the helicoid structure disruption (under high H) of the alloys $MnAu_2$, $\Delta v/v$ is very large, being $\sim 500 \times 10^6$ in 19 KOe and at 20°C. This is also caused by the destruction of helicoid structure under the external field which alters the exchange energy of interaction between the spin layers. This large $\Delta v/v$ indicates that exchange energy depends strongly on the interatomic spacing. In the ferromagnetically ordered state, spontaneous magnetization is related to spontaneous ordering of the lattice. Simple theory suggests that $\lambda = a(T) \sigma_{mag}^2$. So if $\sigma_{sp} + \sigma_H = \sigma_{total}$ due to external field, H , then λ should be $\lambda_{sp} + \lambda_H$ (external field). So, if we plot λ_{\perp} vs. σ^2 for different temperatures near T_N , the curves give linear portions in the plot, and can be extended to give λ_{sp} at various temperatures (Fig. 10). Finally, a plot of λ_{sp} as a function of σ_{sp}^2 is found experimentally to be linear for $MnAu_2$, same as that for ferromagnets. Also the compression data on the same material are available to be compared with the volume magnetostriction. From the linear and volume magnetostriction study, it becomes clear that in ferromagnetic ordered region one has an anomalously large $\Delta v/v$ accompanying the destruction of the helicoid order.

III. CONCLUSIONS

We have seen that magnetostriction has sufficient practical and theoretical interest. The materials that can be profitably

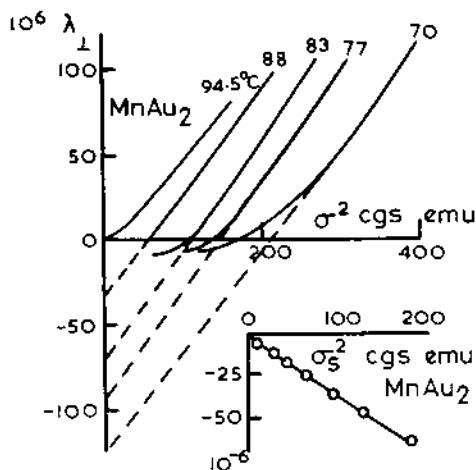


Fig. 10. (a) Dependence of transverse magnetostriiction λ_1 of MnAu₂ on square of magnetization σ^2 at temperatures close to Néel point; (b) dependence of the spontaneous deformation λ_{s1} on square of spontaneous magnetization σ_s^2 for MnAu₂. Nikolaeve JETP 18, 331, 1964.

looked into, range from the conventional engineering grade Iron and Ni alloys, to Ferrites, Garnets and the extremely fascinating ordered structures of the helicoidal and other spin arrangements¹¹. The measurements are usually done by two techniques: (i) By attaching strain gauges which are precalibrated to the samples, magnetostriiction under external field can be measured. However, this needs elaborate precautions. (ii) The sample is attached to one plate of a condenser where the other plate can be regulated with a screw. This whole system forms a part of a cryostat with suitable arrangements to control the temperature. This process eliminates the need to standardise the measurements, since the change in dimensions is directly measurable in terms of the variation of frequency in a tuned electronic circuit. We may conclude that a rather straightforward type of experiment¹² such as magnetostriiction, could be of consequence if suitably chosen systems are studied.

REFERENCES

1. K. Honda and S. Kaya, Sci. Rep. Tohoku Univ. 15, 721, (1926) and 17, 639, 1157 (1928).
2. H.J. Williams, Phys. Rev. 52, 747 (1937).
3. R. Becker, Phys. Zeits. 33, 905 (1932).

4. G.W. Elmen, Elect. Engg. 54, 1292 (1935); C.W. McKeehan, Phys. Rev., 28, 158 (1926).
5. R.M. Bozorth, Rev. Mod. Phys. 25, 42 (1953).
6. V.P. Kiriukhin, JETP, 24, 287 (1967).
7. E.R. Callen, et. al, Phys. Rev. 130, 1735 (1963).
8. V.I. Sokolov, JETP, 25, 986 (1967).
9. V.E. Nikolaev, JETP, 18, 331 (1964).
10. V. Belov, JETP, 15, 279 (1962).
11. A.W. Lee, Phys. Rev. Lett. 22, 1436 (1969).
12. General Reference - 'Magnetic Materials' by F. Brailsford, 3rd Edition, 1960, London, Methuen and Co. Ltd. This monograph contains a good deal of information on ordinary materials and gives many of the older references.

MÖSSBAUER EFFECT AND DYNAMICS OF ATOMIC MOTION IN SOLIDS

V. G. Bhide

National Physical Laboratory
New Delhi, India

Of the many fundamental discoveries of recent years, the discovery of Mössbauer effect¹ has been one of the most fruitful. It has found applications in such diverse domains as coordination chemistry, solid state chemistry, nuclear physics, geology and biology. The elegance of the technique, the simplicity of the equipment and the remarkable accuracy it yields in the determination of the energy of the nucleus has made it possible to investigate a variety of hyperfine interactions of the nucleus with the surrounding changes. Mössbauer effect is the resonance fluorescence or the recoilless emission and absorption of γ -rays. We shall first discuss the difficulties encountered in observing resonance fluorescence of γ -ray and then briefly describe the mechanism of Mössbauer effect. We shall end with a brief description of some of the applications of Mössbauer effect to the study of condensed state of matter.

I. RESONANCE FLUORESCENCE

The story of resonance fluorescence starting with the early suggestion of Lord Rayleigh, at the end of last century, regarding the feasibility in atomic systems and ending at least for the present with the discovery of Mössbauer effect in 1957 is ever so fascinating and typifies the development of physics over the last two centuries. We are all familiar with mechanical resonance. Thus, if we have two simple pendulums identical in all respects and if one of them is oscillated, then it is found that the second picks up the oscillations. The transfer of energy back and forth between these two identical systems is a consequence of resonance

between the two systems. Similarly, acoustic resonance can also be easily demonstrated.

With the emission of characteristic line spectra by atoms it was but natural to expect resonance in atomic systems and indeed Lord Rayleigh predicted that resonance scattering should occur in atomic systems. In 1904, Wood² demonstrated the phenomenon of resonance fluorescence in atomic systems. He used the familiar yellow light emitted by sodium atoms which can be excited by any one of the several processes, the simplest of which is to keep table salt in a Bunsen flame. The absorber consisted of an evacuated glass bulb containing sodium vapour in equilibrium with some sodium metal. The yellow light from the sodium atoms in the flame is focussed with a lense system to give a bright yellow patch AB on a screen. On introducing the bulb containing sodium vapour between the flame and the screen, the intensity of the yellow patch diminishes strikingly indicating that the sodium atoms in the absorber have absorbed preferentially the yellow light emanating from the flame. If we perform this experiment in the scattering geometry, then it is found that the same yellow light is now scattered by atoms in the bulb in all directions. These experiments clearly demonstrate that atoms absorb preferentially those radiations which they emit when they are excited. a principle which is embodied in the Kirchoff's law.

With the advent of Bohr's theory, the phenomenon of resonance fluorescence found easy explanation, for, according to this theory, the emission and absorption of light is governed by the well-known relation, $E_1 - E_0 = h\nu$. It is easy to see that when a sodium atom is excited, it emits photons corresponding to definite frequencies and hence energies. When these photons strike absorbing atoms they are absorbed raising the atoms from the ground state to an excited state where they remain for a very short time after which the absorbing atoms return to the ground state with an emission of radiation of the same frequency in all directions. Thus the intensity of the incoming radiation is decreased along the direction joining the emitter and the absorber, but increased along any other direction. This mechanism predicted that the primary and secondary radiation from the same atomic system should have the same wavelength and that absorption and remission are independent processes such that no phase relation should exist between the primary and the scattered radiation.

With the explanation of the salient features of the phenomenon of resonance fluorescence, finer aspects of the characteristics of radiation emitted by an atom such as the line width, coherence, radiation damping, etc., attracted attention. These aspects could not be easily explained on the basis of simple Bohr theory but the Dirac theory of radiation as developed by Heitler³, Weisskopff⁴,

Wigner⁵ and others provided a complete description of the emission and absorption processes.

The observation of γ -rays with sharply defined frequencies emitted by nuclei either spontaneously or after excitation by one of the several processes, revealed that nuclei can exist in excited states, the energies of which are also determined from Bohr's relation, $E_1 - E_2 = h\nu$. It was, therefore, merely an extension of the argument given for the observation of resonance fluorescence in atomic systems to ask whether it would be possible to observe resonance fluorescence of γ -rays. Indeed a systematic search started with the experiments of Kuhn⁶ in 1929. Although there were several unsuccessful attempts, it was only Mössbauer who could demonstrate the feasibility of observing resonance fluorescence of γ -rays by embedding the absorbing and emitting nuclei in well bound crystalline solids. In order to appreciate the difficulties in observing resonance fluorescence of γ -rays and at the same time the importance of Mössbauer's discovery it is necessary to look more closely into the process of resonance fluorescence.

Consider an emitting system, be it a mechanical system such as a gun shooting a bullet or an atomic system emitting a photon with frequency in the visible part of the spectrum or a nuclear system emitting a γ -ray. Let the emitted particle or photon have a momentum along the direction of emission equal to p . The law of conservation of momentum requires that the emitting system should suffer a recoil such that the recoil momentum is equal to p and is in direction opposite to the direction in which the particle or the photon is emitted. The recoil of a gun on shooting a bullet is a familiar phenomenon and essentially the same thing happens during the emission of a photon during the deexcitation of the atomic or the nuclear system. If mass of the emitting system which was supposed to be at rest before emission is m , then the recoil momentum necessitates that the emitting system should receive an energy which may be called the recoil energy R given by, $R = p^2/2m$. This energy is obviously obtained from the deexcitation energy of the emitting system. Thus it is easy to see from the point of view of conservation of energy that the emitted particle or a photon should have energy less than the deexcitation energy of the emitting system. This means that the bullet moves slower than what it would have, had the gun not recoiled. In the same way, the photon energy will be less or, in other words, there will be a shift towards the red end of the spectrum when the emitting system (either the atomic or the nuclear) recoils.

Confining our attention to the quantum mechanical system such as an atom or a nucleus emitting an optical or a γ -ray photon, let us determine the change in the energy of the photon due to the recoil of the emitting system. Such a quantum mechanical system

deexcites itself from an excited state E_1 to the ground state E_0 by the emission of a photon. If all the transition energy were to go into the photon, then its frequency would be, $(E_1 - E_0) = h\nu$. The emitted line in this case will be centered at $(E_1 - E_0)$ on the energy scale. However, it is known that a photon has a momentum $p = h\nu/c$ along the direction of propagation. The emitting system will recoil in a direction opposite to the direction of the emission of a photon and get a recoil momentum p equal to the momentum of the photon. The kinetic energy acquired by the emitting system as a consequence of this recoil is given by, $R = p^2/2m = h^2\nu^2/2mc^2$, where m is the mass of the emitting system.

The recoil energy, R , is proportional to the square of the photon energy and is, therefore, much larger for γ -ray emission (with energies ranging from a few KeV to a few MeV) as compared to the emission of an optical photon which has energies ranging between 1-10 eV. Indeed, the recoil suffered by the nucleus during γ -emission is some 10^6 to 10^{12} times the recoil suffered by an atom during the emission of photon in the visible range.

The recoil energy of the emitting system is obviously obtained from the transition energy and consequently we have, $(E_1 - E_0) = h\nu + R$. In this situation the photon has an energy slightly less than the transition energy. The emission line is no more centered at $(E_1 - E_0)$ but is centered at $(E_1 - E_0) - R$ (i.e., slightly shifted towards the red end of the spectrum).

When a photon corresponding to the transition energy $(E_1 - E_0)$ is incident on an absorbing atom or the nucleus, then the incoming photon is absorbed. During the process of absorption, the photon momentum is reduced to zero as it were and hence to conserve the momentum the absorbing system recoils and obtains a momentum p , along the direction of the incoming photon in much the same way as a hollow sphere will acquire a momentum if we shoot a bullet into it. Consequent on this recoil, the absorbing system acquires an energy R . This kinetic energy has necessarily to be supplied by the incoming photon. Thus, to excite the absorbing system from the ground state E_0 to the excited E_1 it is not enough to have an energy $(E_1 - E_0)$ but the photon must have an energy $(E_1 - E_0) + R$. In other words, the absorption line is not centered at $(E_1 - E_0)$ but is located at $E_1 - E_0 + R$. The shift of the absorption line towards the violet end of the spectrum is equal to the shift of the emission line towards the red end of the spectrum. Indeed, the two are separated on the energy scale by $2R$. This implies that the system has maximum emission corresponding to the energy $(E_1 - E_0 - R)$ and maximum absorption corresponding to the energy $(E_1 - E_0 + R)$ apparently violating Kirchoff's law (Fig. 1).

It is thus seen that the absorption line does not overlap the

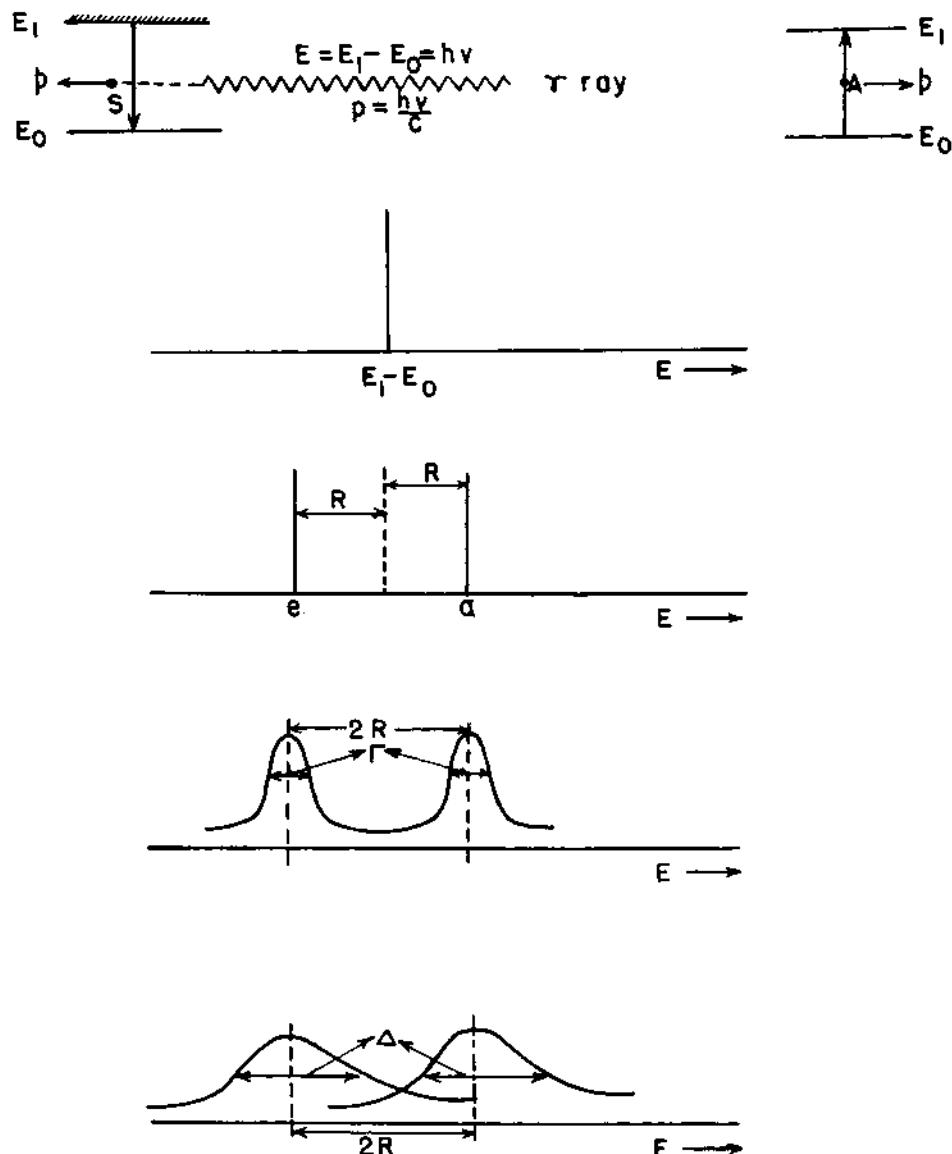


Fig. 1. The position of the emission and absorption lines on the energy scale, (a) position showing the transition energy, (b) position of the emission and absorption lines affected by the recoil, (c) broadening of the emission and absorption lines due to finite life time, (d) broadening due to Doppler effect.

emission line because of the recoil obtained by the emitting and absorbing processes respectively. Indeed, no resonance fluorescence should be observed even in the visible part of the spectrum emitted by atomic systems unless some other fortunate circumstance makes it possible for the emission line and absorption line to overlap at least partially. One of the factors which makes it possible is the natural broadening of the spectral lines.

In the classical description, the emission line has a finite width arising out of the reaction force of the radiation on the oscillator. In the quantum mechanical formalism radiation damping is included just as fully as in the classical theory. We may, therefore, expect that the quantum theory will also account for the natural breadth of a spectral line without difficulty. The relation between the breadth of spectral line and the transition probability can be understood from the uncertainty relation,

$\Delta E \cdot \Delta t \approx \hbar$. In our case, the excited state of an atom has a lifetime, $\tau = 1/\gamma$, due to radiation transition probability (or lifetime of the oscillator). The energy of the excited state is only defined with an uncertainty $\Delta E = \hbar \gamma$. If we represent the lifetime of the system in the excited state by Γ , then the natural breadth of a spectral line is given by, $\Gamma = \hbar/\tau$.

Several complications may arise and broaden the line further. In a gas of finite density an excited state atom undergoes collision with neighbouring atoms which may induce a transition to the ground state. The effect of these collisions on the breadth can be discussed as follows: If the number of effective collisions per second is N , then the lifetime of the excited state will be shortened. Indeed, due to collisions the system is being forced to deexcite itself faster. Consequently, the total number of transitions (normal radiative and collision induced) is now equal to $\gamma + N$. This reduces the lifetime of this system and broadens the spectral line. In the γ -ray emission process, the deexcitation of the system may take place through several mechanisms such as radiative transition, internal conversion, change in the valence state of the atom, etc. All these act in parallel channels and decrease the lifetime consequently broadening the γ -ray line. If the excited state can decay in several different ways, the total energy spread of the excited state level Γ is equal to the sum of all partial widths Γ_i related by the uncertainty principle to the lifetime τ_i against the i^{th} mode of decay. As the number of modes of decay increases, the mean lifetime of the excited state decreases, $1/\tau = \sum (1/\tau_i)$. Correspondingly, $\Gamma = \sum \Gamma_i = \hbar \sum (1/\tau_i)$.

Let us assume that our emitting system has a lifetime τ in the excited state E_1 from which it decays to a ground state E_0 .

The spectral line that would be emitted would have a Lorentzian shape having a frequency distribution

$$I(\nu) d\nu = \frac{I_0}{2\pi} \frac{\Gamma^2/4}{(\nu - \nu_0)^2 + \Gamma^2/4}$$

such that, $\Gamma = \hbar/\tau$ and $\nu_0 = (E_1 - E_0)/\hbar$.

Similarly for an identical system, the absorption line would also be a Lorentzian with the same width and similar frequency distribution. But, as we have seen, both the emission and absorption lines are displaced due to the recoil suffered during the process of emission and absorption. Consequently, the emission and absorption line has a frequency distribution $I(\nu)$ corresponding to

$$\frac{I_0}{2\pi} \frac{\Gamma^2/4}{(\nu - \nu_0 + R/h)^2 + \Gamma^2/4} \text{ and } \frac{I_0}{2\pi} \frac{\Gamma^2/4}{(\nu - \nu_0 - R/h)^2 + \Gamma^2/4}$$

respectively. It is obvious that the two lines will partially overlap if and only if, $\Gamma > 2R$. In most atomic systems, this condition is usually satisfied and hence resonance fluorescence is observed; in the nuclear systems emitting γ -rays, this condition is generally not satisfied and hence resonance fluorescence cannot be observed. Besides natural broadening there is yet another important process which tends to broaden the spectral lines. This broadening is essentially due to Doppler effect and arises out of the movement of the emitting and absorbing systems as a result of temperature. In our foregoing discussion, we assumed that the emitting and absorbing systems were at rest. This is seldom the case and indeed we find that the emitting and absorbing systems are in a constant state of agitation as a consequence of which spectral lines are broadened; the broadening would be dependent on temperature. If we observe the light emitted in the x direction, the line will be shifted because of the Doppler effect by an amount, $\Delta\nu = \nu_0 v_x/c$, where v_x is the velocity of the emitting atom or a nucleus. Using Maxwell-distribution of velocities, $\exp(-mv_x^2/2kT)$, we obtain a broad line with an intensity distribution

$$I(\nu) = \text{const} \times \exp - \frac{mc^2(\nu - \nu_0)^2}{2 \nu_0^2 kT}$$

with a breadth at half maximum Δ given by

$$\Delta = \nu_0 \sqrt{\frac{2kT}{mc^2} \log 2}$$

In general, Doppler broadening Δ is much greater than the natural breadth. The intensity distribution is, however, exponential and, therefore, decreases rapidly with the distance from the maximum in contrast with the natural broadening which has a very long tail. The intensity observed at large distances from the maximum is, therefore, due to the natural broadening. The spectral line due to Doppler broadening is no more a Lorentzian but becomes a Gaussian following the relation, $I(\nu) = C \exp(-B(\nu - \nu_0)^2/kT)$, where C and B are constants. Quantum mechanically, Doppler effect results from the recoil momentum given to the radiating system by the emitted photon as first suggested by Fermi⁷.

Several interesting cases arise which modify the expression for the intensity distribution of Doppler broadened line. Indeed, the Doppler broadened line shape will be determined by the velocity distribution of the emitting atoms or nuclei. In a gas at pressures normally encountered, the time for which the atom is actually in collision is negligibly small. Doppler breadth, within limits, would be expected to be pressure independent. Actually, this assumption is not at all correct. Collisions which do not affect the internal state of the radiating system have a large effect upon the Doppler breadth. If the emitting system is embedded in a crystalline solid, the Doppler broadened spectra is determined by the lattice vibrations and we then have a possibility of having the emission without a recoil. These and similar interesting cases would be discussed in later sections.

The Doppler breadth is proportional to the energy of the emitted photon and square root of absolute temperature. Fig. 2 shows the variation of recoil energy, natural broadening for spectral lines with energies ranging from few eV to few MeV. It is obvious that the emission and absorption lines would overlap even partially making resonance fluorescence possible if the Doppler broadening Δ is greater than the recoil energy, $\Delta > 2R$. In general, for spectral lines in the visible region, $\Delta > 2R$, making the observation of resonance fluorescence possible. However, for γ -rays emitted by nuclei the recoil energy R is comparable or at times even greater than the Doppler broadening making it difficult to observe resonance fluorescence. The difficulty in the observation of resonance fluorescence of γ -rays is further accentuated by the additional consideration of cross section for resonance fluorescence.

Cross Section for Resonance Fluorescence. Resonance experiments are usually performed either in transmission geometry or in the scattering geometry. In the former, we determine the attenuation of the incoming beam because of the intervening absorber. In the latter, we determine the radiation emitted by the absorbing system in a direction other than the direction of the incident

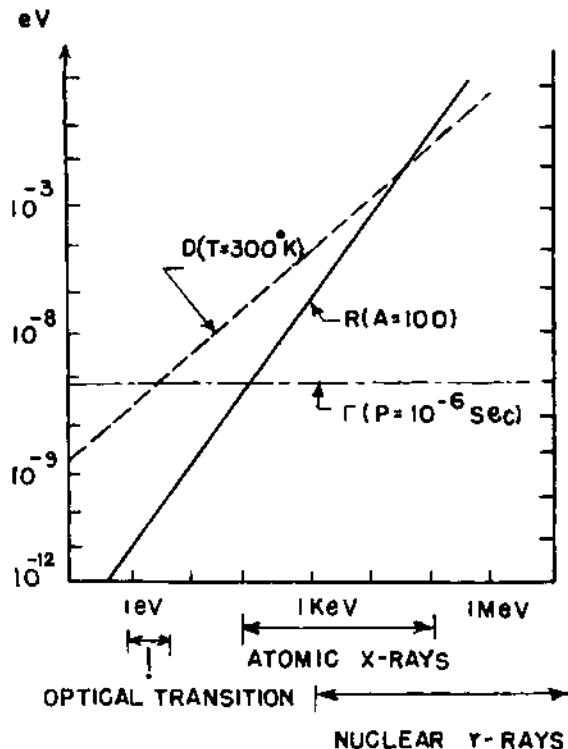


Fig. 2. Variation of recoil energy, Doppler broadening and natural width for spectral lines with energies ranging from few electron volts to few million electron volts.

radiation. Both these processes have a certain cross section which is determined for a thin absorber (so that multiple scattering is not possible) by the wavelength of the incident radiation. It is possible to determine the cross section for scattering as well as for absorption in the simplest cases using both the classical or the quantum theory as shown in detail by Heitler.

The scattering cross section and the absorption cross section assume the form⁸,

$$\sigma_{\text{scatt}}(E) = \frac{\pi \lambda^2}{2} \frac{2I_1 + 1}{2I_o + 1} \frac{\Gamma^2}{(E - E_o)^2 + \Gamma^2/4} \frac{1}{(1 + \alpha)^2}$$

$$\sigma_{\text{abs}}(E) = \frac{\pi \lambda^2}{2} \frac{2I_1 + 1}{2I_o + 1} \frac{\Gamma_o \Gamma}{(E - E_o)^2 + \Gamma^2/4} \frac{1}{1 + \alpha}$$

where I_1 and I_0 are the total angular momenta of the excited and ground states respectively, E_0 is the resonance energy, λ the corresponding wavelength divided by 2π , Γ the natural width of the level, Γ_0 the partial width of the direct γ -ray transition to the ground state and α is the internal conversion coefficient. This means that scattering cross section is reduced by $(1/(1+\alpha))^2$ and the absorption cross section is reduced by $(1/(1+\alpha))$. Taking the example of the Mössbauer isotope Fe⁵⁷, emitting 14.4 keV γ -ray with a corresponding internal conversion coefficient of 9, we have the scattering cross section reduced by a factor of 100. In Ge⁷⁵ emitting 13.5 keV γ -ray, the internal conversion coefficient is 3600 reducing the scattering cross section by a factor of about 7×10^{-8} .

It may be pointed out that the above expressions for scattering and absorption cross sections are dependent on the following assumptions: (i) The incident radiation is a delta function or the emission line is very sharp line. (ii) Only one scattering or the absorption level exists. If more than one level exists in close neighbourhood, the expression has to be modified as shown by Margulius.

If the incident radiation has a certain intensity distribution $I(E)$ which may be Lorentzian for a naturally broadened line or a Gaussian for a Doppler broadened line then the effective scattering and absorption cross section are given by similarly

$$\sigma_{\text{eff.scat}} = \sigma_0 = 4 \int \pi \lambda^2 \frac{2I_1+1}{2I_0+1} \frac{\Gamma^2}{(E-E_0)^2 + \Gamma^2/4} \frac{I(E)dE}{1+\alpha}.$$

If $I(E)$ is described by a delta function and $x = 0$, then

$$\sigma_{\text{eff.scat}} = \sigma_0 = 4\pi \lambda^2 \frac{2I_1+1}{2(2I_0+1)} \frac{1}{(1+\alpha)^2}.$$

The actual lines are in general wider, than indicated above, because of Doppler broadening. If for a nucleus at rest, the gamma radiation has the energy E , the energy E' for a nucleus moving towards the source of radiation with a velocity v is shifted to, $E' = E_0(1+(v/c))/\sqrt{1-(v^2/c^2)}$, c being the velocity of light. If the velocities of the nuclei in the absorber are distributed according to Maxwellian distribution we get the normal broadened line whose width is given by, $\Delta = (E/c)\sqrt{2kT/m}$. In order to obtain effective cross section for gamma rays of energy E , expression for scattering cross section has the form,

$$\sigma(E) = \sigma_0 / \left\{ \left[\frac{2(E-E_0)}{\Gamma} \right]^2 + 1 \right\}$$

which has to be averaged over all possible values of E' of the incoming gamma ray energy. Thus,

$$\sigma(E, t) = \int \sigma_0(E) \omega(E') dE' = \sigma_0 \Psi(x, t) \quad (*)$$

where $x = \frac{2(E-E_0)}{\Gamma}$; $t = (\frac{\Delta}{\Gamma})^2$

$$\Psi(x, t) = \frac{1}{2(\pi t)^{\frac{1}{2}}} \int_{-\infty}^{\infty} \frac{\exp[-(x-y)^2/4t]}{1+y^2} dy$$

and $y = 2(E' - E)/\Gamma$

$\Psi(x, t)$ is tabulated for values of x upto 300 and of t upto 2500 (available in literature).

For large values of t beyond those tabulated, $\Psi(x, t)$ is well approximated by

$$\Psi(x, t) = \pi^{\frac{1}{2}} (2t)^{\frac{1}{2}} \exp(-\frac{x^2}{4t})$$

provided that $x \ll t$. This means that for $\Delta \gg \Gamma$ the effective cross section has the Doppler form

$$\sigma_D(E) = \sigma_0 \left(\frac{\pi^{\frac{1}{2}}}{2\Delta} \right) \exp \left[- \left(\frac{E-E_r}{\Delta} \right)^2 \right].$$

Considering the above integral equation marked (*) and the fact that $\int_{-\infty}^{\infty} \Psi(x, t) dx$ has the value π independent of t , it is clear that the integrated cross section

$$\int \sigma(E, t) dt = \sigma_0 \Gamma \pi / 2$$

does not depend upon the relative size of the natural width and the Doppler width. However, the maximum cross section is reduced,

$$\sigma_{\text{max. scat}}^D = \frac{\sigma_0 \Gamma}{D} \frac{1}{1+\alpha}.$$

Correspondingly the max. absorption cross section becomes

$$\sigma_{\text{max. abs.}}^D = \sigma_0 \frac{\Gamma^2}{\Delta^2} \cdot \frac{1}{(1+\alpha)^2}$$

It is thus clear that because of thermal motion of the emitting and absorbing systems the scattering cross section is reduced tremendously.

To summarize the difficulties in observing resonance fluorescence we have, (i) the recoil energy, being proportional to the photon energy is nearly 10^6 to 10^{12} times large for γ -rays with energies ranging from few keV to few MeV than for the radiation in the visible region of the spectrum. The recoil suffered by the emitting and absorbing systems consequent on the emission and absorption of radiation, separates the emission and absorption lines by an energy $2R$ making it very difficult for the γ -ray emission line to even partially overlap the absorption line. It is thus not possible to observe resonance fluorescence of γ -rays whereas in the optical part of the spectrum since the recoil energy is small, the emission line overlaps the absorption line making resonance fluorescence a rule. (ii) Since the maximum absorption and scattering cross sections are inversely proportional to the square of the photon energy, even if there is a small overlap of the emission and absorption line in the γ -ray region, the small cross section for both resonance absorption and scattering makes it very difficult to observe resonance fluorescence of γ -rays.

Most nuclear transitions are forbidden transitions in the atomic sense (i.e., they are not electric dipoles). Moreover, the nuclear matrix element for a given multipole order are smaller than the atomic matrix elements because of the smaller dimensions of nuclei¹⁰. As a result, the gamma lines are very narrow and the recoil energy losses prevent the observation of resonance fluorescence. This difficulty explains to a large extent the slow development of the resonance fluorescence of γ -rays. The unsuccessful attempts which covered a period of twenty years following the first suggestion and experiment by Kuhn can all be blamed on the sharpness of the lines due to electric quadrupole transitions.

Once the hope of finding wide γ -ray lines was abandoned, efforts were concentrated on restoring the resonance condition by the creation of special source conditions; then it was that nuclear resonance fluorescence was observed¹⁶ (Moon 1951). Moon performed the first successful experiment by plating a source of Au¹⁹⁸ on to the tips of a steel rotor and spinning the rotor upto tip speeds of 800 m/sec. Gamma rays emitted tangentially thus gained additional energy sufficient to compensate for the recoil loss. Other possibilities for compensating for the recoil energy loss, for instance the use of the recoil from a transition proceeding the gamma ray to be

investigated or the heating of the source and absorber to increase the average kinetic energy have also been used with some success.

Despite the difficulties associated with classical nuclear resonance fluorescence experiments due to Doppler broadening of the emission and absorption lines and small cross sections, several experiments have been performed¹⁷⁻¹⁹.

II. MÖSSBAUER'S EXPERIMENT

Mössbauer^{20,21} was investigating the nuclear resonance scattering of the 129 keV γ -rays from Ir¹⁹¹. For this transition, the free recoil energy R is 0.05 eV whereas Doppler broadening at room temperature is about 0.1 eV. The lifetime of the excited state is of the order of 10^{-10} sec giving the natural width of 6.5×10^{-8} eV. At room temperature, because of the fact $\Delta > R$, there is an overlap of the absorption and the emission lines yielding some resonance absorption. In order to determine the background in his experiment, Mössbauer cooled the system with the hope of reducing the Doppler broadening and eliminating the overlap of the emission and absorption line. Contrary to his expectation, the resonance fluorescence increased considerably. With the help of a theory developed by Lamb²², this effect was attributed to the fact that in solids the recoil momentum does not always produce a change in the vibrational state of the crystal lattice. Instead, for a fraction of the gamma transitions, the solid as a whole can take up the recoil momentum. Thus according to this theory, the emission and absorption spectra contain very strong lines of natural width superposed over a broad distribution resulting from the thermal motion of the atoms bound in a crystal lattice. Because of the vanishingly small recoil energy losses, these lines appear undisplaced at the resonance energy position. Mössbauer demonstrated the existence of these unshifted resonance lines by means of a centrifuge method, employing velocities of only a few centimeters per second. The experimental arrangement used by Mössbauer is shown in Fig. 3, while Fig. 4 shows the results obtained for the 129 keV transition in Ir¹⁹¹. Thus Mössbauer for the first time showed dramatically and yet accidentally the feasibility of observing γ -ray resonance fluorescence by embedding the emitting and absorbing nuclei in well bound solid. Under these conditions, he showed that a fraction of γ -ray transitions take place without changing the vibrational state of the crystal lattice and these γ -rays have the natural width and are not broadened by the phonon spectrum of the lattice. Mössbauer thus in one single experiment achieved the compensation of the recoil and also eliminated the Doppler broadening. Consequently, Mössbauer effect is sometimes called nuclear gamma ray resonance or zero phonon emission and absorption of gamma ray or the recoilless emission and absorption of gamma rays.

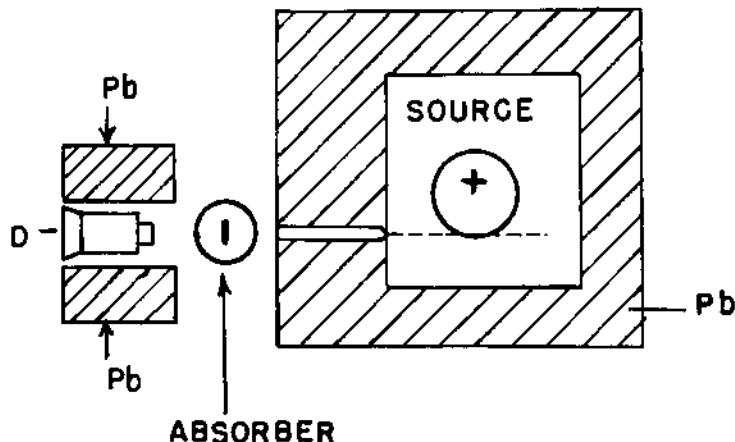


Fig. 3. Experimental set-up used by Mössbauer to demonstrate resonance absorption of gamma-rays emitted by Ir^{191} .

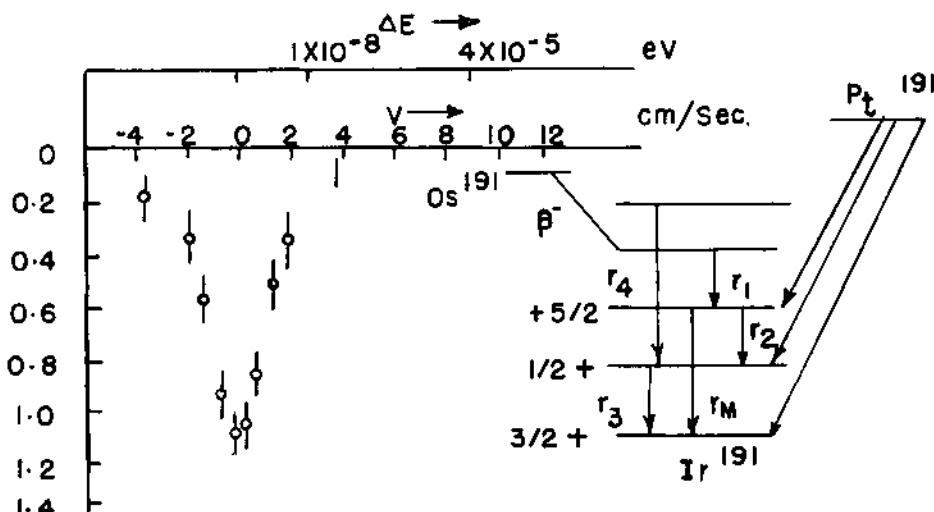


Fig. 4. The first Mössbauer spectrum obtained with 129 keV gamma-rays from Ir^{191} .

Taking the case of popular Mössbauer isotope Fe^{57} which gives 14.4 keV γ -ray radiation, one finds that under conditions of Mössbauer effect, one can observe this line with a natural line width of 4.19×10^{-9} eV (corresponding to a lifetime of the excited state $\tau = 1.4 \times 10^{-7}$ sec). The experiment immediately brought out the fact that with this effect one could investigate any changes in the nuclear energy levels greater than 10^{-9} eV. Significantly, the variety of hyperfine interactions which cause either the shift or the splitting of the levels are greater than this value. Mössbauer effect thus provided a powerful tool to investigate variety of hyperfine interactions. In the following sections we will try to understand the mechanism of Mössbauer effect which will also point out the conditions which need to be satisfied to observe it.

Mechanism of Mössbauer Effect²³. Let us consider a crystal lattice and assume for simplicity that it conforms to the Einstein model. Assume that the nucleus of one of the atoms in this crystalline solid decays by γ -emission. Let the energy of the gamma ray be $E = h\nu$ and its momentum be $h\nu/c = p$. Let us inquire as to how the emitting system takes the recoil momentum and the recoil energy. Since, the atom whose nucleus decays is a part of the crystal lattice, we have to consider the following possibilities for accounting the conservation of momentum. The possibilities are: (i) The atom alone may recoil and get displaced from its equilibrium position in the lattice. (ii) The crystal lattice as a whole recoils so that every atom in the lattice moves and that the lattice structure is still preserved. (iii) The lattice vibrations or any other collective oscillations of the lattice may take up the recoil momentum. It is well known that the energy required to displace a lattice point from its equilibrium position in a well bound crystalline lattice is of the order of 10 eV. On the other hand, the recoil energy is some 10^{-2} eV. Thus, it looks improbable that the individual atom will take the recoil and get displaced from its equilibrium position. Lattice vibrations similarly cannot take up the momentum since, to each wave with its momentum pointing in one direction, there will be a corresponding one pointing in the opposite direction. Thus it appears that as a consequence of the γ -emission, the entire crystal recoils. The mass of the recoiling system is, therefore, the mass of the crystal which is some 10^{21} times or more than the mass of a single atom (assuming that the crystal has a size of 1 cc). The energy taken up during this recoil is insignificant because the mass of the recoiling system is very large indeed. Thus, although, the entire crystal takes the momentum, the energy transfer is negligibly small, making recoilless emission possible. In fact, the momentum p is transferred to a single nucleus in a short time compared to the period of lattice vibrations. Under these conditions, the momentum is entirely taken up by a single nucleus and then distributed throughout the whole crystal with the velocity of sound via the binding forces. The centre of mass of the crystal

responds immediately to the momentum transfer.

Now let us consider the energy conservation. We have again to consider the same three possibilities mentioned earlier, but we would expect the recoil energy R to be dissipated in the lattice vibrations. We have two energy states of the lattice, namely, the ground state G and the excited state E , separated by the energy $\hbar\Omega$, where Ω is the frequency with which each of the atoms vibrate when excited. We shall consider more realistic situations later, but let us assume that the lattice before the emission of γ -rays is in the ground state. We may now ask a question, "what happens to the lattice as the nucleus of one of the atoms emits a γ -ray"? There are two possibilities, viz., the lattice may continue to remain in the ground state or it may get excited. In the latter possibility, the lattice gains energy from the γ -ray with the result, the γ -ray is shifted on the energy scale by $\hbar\Omega$ towards the low energy side. The lattice, on the other hand, will get heated as it were. Let us assume that the probability for the lattice to remain in the same state before and after emission of γ -ray is f_0 and the probability of excitation to the excited state is f_1 , then, since either of the two must happen, we have, $f_0 + f_1 = 1$. Considering the energetics, we get, $f_0 \cdot 0 + f_1 \cdot \hbar\Omega = R$, or, $f_1 = R/\hbar\Omega$ and $f_0 = 1 - R/\hbar\Omega$. Thus, we find that the probability for the lattice to remain in the same state before and after emission or the probability that the gamma ray emission takes place without exchange of energy with the lattice (zero phonon emission) depends upon the recoil energy R and the lattice vibration energy $\hbar\Omega = k\theta_E$ where $\hbar\Omega/k$ is the Einstein temperature. We see that as $k\theta_E$ or the binding of the lattice increases f_0 becomes large. We find that $f_0 = 1/2$ if $k\theta_E = 2R$ and $f_0 = 3/4$ when $k\theta_E = 4R$. Thus the probability of zero phonon emission increases as the binding energy of the lattice is increased. We may note that all the gamma rays are not emitted without interaction with the lattice. Indeed, there is a definite probability for γ -emission in which lattice takes a given energy from the γ -rays. Thus, we have, instead of one single line centered at ν ($\hbar\nu = E_1 - E_0$), two more lines on either side of it centered at $\nu + \Omega$ and $\nu - \Omega$. In the zero phonon case, the γ -ray carries with it the full transition energy and in these processes we have the full overlap of the emission and absorption lines, making resonance fluorescence possible.

We also see that if the zero phonon line is to be distinctly seen or is to be unaffected by the phonon components, we must have the natural width of the γ -ray line less than the phonon energy, or $\Gamma \ll \hbar\Omega$ since $\Gamma \propto h$ where Γ is the lifetime of the nucleus in the excited state. If this condition is satisfied, we will have the γ -ray lines with natural line width centered at ν_0 . This

condition can be rewritten in the form, $1/\tau_{\text{life}} \ll \Omega$, or $1/\tau_{\text{life}} \ll 1/\tau_{\text{L.V.}}$ where $\tau_{\text{L.V.}}$ is the time for one lattice vibration. It is easy to see that the above condition states that if the lattice vibration time is very much less than the lifetime, then the γ -ray line has a natural width. Normally, in well bound solids $\tau_{\text{L.V.}} \approx 10^{-12}$ sec whereas $\tau_{\text{life}} \approx 10^{-6}$ sec. We thus see that if $\tau_{\text{L.V.}} \ll \tau_{\text{life}}$ then the γ -ray line is unaffected by lattice vibration or is not Doppler broadened. We thus see that the conditions to observe Mössbauer effect, i.e., to have γ -ray emission without recoil and the γ -ray line to have natural width are: (1) $h\Omega = k\theta_E \gg R$; (2) $h\Omega \gg \Gamma$. Once these conditions are satisfied, we can get γ -ray lines with natural width. In the following section, we shall try to obtain these very conditions using the theory based on atomic motions.

The total energy transfer depends upon the momentum P possessed by the emitting nucleus at the time of emission. In the case of freely recoiling nucleus this momentum could be determined thus providing a well defined relationship between the momentum and energy transfer to the nucleus. In the case of a bound nucleus, classical mechanics presents no barriers at least in principle to the determination of the initial momentum of the emitting nucleus at the time of γ -emission. In sharp contrast to the classical mechanics, quantum mechanics does not permit a determination of the momentum of the emitting nucleus at the time of emission. This is an immediate consequence of the fact that the nucleus, being bound in the crystal is limited in its movement to a region of the order of interatomic distances. As a result, the momentum has an uncertainty $\Delta p = \hbar/a \approx 10^{-19}$ g cm/sec where 'a' is the interatomic distance. On the other hand, in order to be able to distinguish between recoilless and non-recoilless processes, we require the knowledge of the initial momentum of the emitting nucleus $\Delta p \leq \Gamma/\hbar K$ where Γ is the natural line width and $\hbar K$ is equal to E_0/c where E_0 is the energy of γ -ray. Substituting the typical values for E_0 , Γ and m , we obtain $\Delta p \approx 10^{-24}$ g cm/sec. These two conditions on Δp are incompatible. The uncertainty in the initial momentum Δp of the emitting nucleus leads to an uncertainty in the energy transferred to the lattice, which is so large that it becomes impossible to predict whether a single emission process is recoilless or with recoil. Quantum mechanics only permits the determination of the probabilities for the process in which specific emission takes place with zero phonon or with emission or absorption of phonons.

III. RESONANCE ABSORPTION OF GAMMA RAYS AND THE DYNAMICS OF ATOMIC MOTION

In the foregoing section we have discussed the mechanism of Mössbauer effect from very elementary considerations. Various sophisticated theories can be given to explain Mössbauer effect but out of all these the most physically revealing one is based on the dynamics of atomic motion. In this section we will briefly give this theory and outline the power of the Mössbauer technique to the investigation of atomic motion in any interacting system.

Let us consider a general situation in which there is an impurity atom in a crystal giving rise to a localized mode. The expression for cross section for the absorption of a γ -ray of energy E by a nucleus in the crystal is given by

$$\sigma_a(E) = \frac{1}{4} \sigma_0 \Gamma^2 \frac{\sum g_n \langle m | e^{-i\vec{K} \cdot \vec{R}_j(o)} | n \rangle \langle n | e^{-i\vec{K} \cdot \vec{R}_i(t)} | m \rangle}{(E_o - E + E_n - E_m)^2 + \Gamma^2/4}$$

where E_o is the energy difference between the final and the initial state of the absorbing nucleus, E_m and E_n are the energies of the eigen states $|m\rangle$ and $|n\rangle$ of the crystal, Γ is the natural width, K is the γ -ray wave vector and g_n is the weightage factor. With the aid of integral representation for the denominator of this expression, we can reduce this expression to the form

$$\sigma_a(\omega) = \frac{1}{2} \sigma_0 Y \int_{-\infty}^{\infty} e^{i\omega t} e^{-\gamma t} F_s(\vec{K}, t) dt$$

where $F_s(K, t)$ is the intermediate scattering factor²⁴ and $\hbar Y = \Gamma$.

In a crystalline solid we can show that

$$F_s(\vec{K}, t) = e^{-i\vec{K} \cdot \vec{R}_j} e^{i\vec{K} \cdot \vec{R}_i} \langle e^{i\vec{K} \cdot (\vec{u}_j(o) - \vec{u}_i(t))} \rangle e^{-\frac{1}{2} [i\vec{K} \cdot \vec{u}_j(o), i\vec{K} \cdot \vec{u}_i(t)]}$$

where \vec{R}_j and \vec{R}_i are equilibrium positions of nuclei j and i and are time-independent c numbers and $\vec{u}_j(o)$ and $\vec{u}_i(t)$ are displacement vectors. The first and third factors are c numbers and are time-independent. Thus, $F_s(\vec{K}, t)$ reduces to

$$F_s(\vec{K}, t) = c \langle e^{-i\vec{K} \cdot (\vec{u}_j(o) - \vec{u}_i(t))} \rangle_T$$

using Ott's theorem, we get

$$F_s(\vec{K}, t) = c \langle e^{-\frac{1}{2}((\vec{K} \cdot \vec{u}_j(o))^2 + (\vec{K} \cdot \vec{u}_i(t))^2 - 2\vec{K} \cdot \vec{u}_j(o)\vec{K} \cdot \vec{u}_i(t))} \rangle_T.$$

For Mössbauer effect we have essentially to consider the same nucleus and thus we have $i = j$ giving

$$F_s(\vec{K}, t) = c \langle e^{-\frac{1}{2}(2(\vec{K} \cdot \vec{u}_i(o))^2 - 2\vec{K} \cdot \vec{u}_i(o)\vec{K} \cdot \vec{u}_i(t))} \rangle_T.$$

Substituting for $F_s(\vec{K}, t)$ in the expression for σ_a we get,

$$\sigma_a(\omega) = \frac{1}{2} \sigma_0 \gamma \int_{-\infty}^{\infty} e^{i\omega t} e^{-\gamma t} e^{-2M} \langle \vec{K} \cdot \vec{u}_i(o) \vec{K} \cdot \vec{u}_i(t) \rangle_T$$

where $2M = (\vec{K} \cdot \vec{u}_i(o))^2$. $\vec{u}_i(o)$ is the displacement of the i^{th} particle at $t = o$ and $\vec{u}_i(t)$ is the displacement of the same particle at $t = t$. We may now write the displacement operator in terms of the normal modes of the system

$$\vec{u}_i(t)$$

$$= \sum_s \left(\frac{\hbar}{2MN\omega_s} \right)^{\frac{1}{2}} \hat{\epsilon}_s \left\{ a_s e^{i(\vec{q}_s \cdot \vec{R}_s - \omega_s t)} + a_s^+ e^{-i(\vec{q}_s \cdot \vec{R}_s - \omega_s t)} \right\}$$

$$\vec{u}_i(o) = \sum_s \left(\frac{\hbar}{2MN\omega_s} \right)^{\frac{1}{2}} \hat{\epsilon}_s \left\{ a_s e^{i\vec{q}_s \cdot \vec{R}_s} + a_s^+ e^{-i\vec{q}_s \cdot \vec{R}_s} \right\}$$

where $(\hbar/2MN\omega_s)^{\frac{1}{2}}$ is the normalization factor, $\hat{\epsilon}_s$ is the polarization vector of the s^{th} mode, \vec{q}_s is the wave vector of the phonon whose creation and annihilation parameters are a_s and a_s^+ respectively. We may now consider two distinct cases, viz., when localized mode is absent and when the localized mode is present.

No Localized Mode is Present. In this case, the calculation of the absorption cross section is relatively easy. Going back to the equation for $\sigma_a(\omega)$,

$$\sigma_a(\omega) = \frac{1}{2} \sigma_0 \gamma \int_{-\infty}^{\infty} e^{i\omega t} e^{-\gamma t} e^{-2M} e^{-\epsilon_s} \langle \vec{k} \cdot \vec{u}_j(0) \vec{k} \cdot \vec{u}_i(t) \rangle_T dt$$

we calculate the value of $2M$,

$$2M = \sum_s (\vec{k} \cdot \hat{\epsilon}_s)^2 (\hbar / 2MN\omega_s) \langle a_s^+ a_s + a_s a_s^+ \rangle$$

$a_s^+ a_s$ is just the number operator, such that $a_s^+ a_s = \bar{n}_s =$

$$\frac{1}{\left(\exp \frac{\hbar \omega_s}{kT} - 1 \right)}.$$

We thus have,

$$2M = \sum_s (\vec{k} \cdot \hat{\epsilon}_s)^2 \frac{\hbar \omega_s}{2MN\omega_s} \coth \frac{\hbar \omega_s}{2kT}$$

which is the exponent of the usual Debye Waller factor. We are interested in the value of this factor under various conditions. But before we do that, we have to calculate the value of $\langle \vec{k} \cdot \vec{u}_j(0) \vec{k} \cdot \vec{u}_i(t) \rangle_T$; after substituting for $\vec{u}_i(0)$ and $\vec{u}_i(t)$ we get,

$$\sigma_a(\omega) = \frac{1}{2} \sigma_0 \gamma \int_{-\infty}^{\infty} e^{i\omega t} e^{-\gamma t} e^{-2M} \exp \left[\sum_s (\vec{k} \cdot \hat{\epsilon}_s)^2 \frac{\hbar}{2MN\omega_s} \right] \left[(\bar{n}_s + 1) e^{i\omega_s t} + n_s e^{-i\omega_s t} \right].$$

The shape of the absorption line turns out to be,

$$\sigma_a(\omega) = \frac{\sigma_0 e^{-2M} \gamma^2}{(\omega - \omega_0)^2 + \gamma^2} + \sigma_0 e^{-2M} \left[\sum_s (\vec{k} \cdot \hat{\epsilon}_s)^2 \frac{\hbar}{2MN\omega_s} \left[\frac{(n_s + 1) \gamma^2}{(\omega - \omega_0 - \omega_s)^2 + \gamma^2} + \frac{n_s \gamma^2}{(\omega - \omega_0 + \omega_s)^2 + \gamma^2} \right] \right].$$

We thus see that in addition to the line centred at ω_0 , we have additional lines centred at $\omega \pm \omega_s$. These are termed as one phonon lines. It is of interest to mention that both the zero phonon and the one phonon lines, under harmonic approximation that we have assumed, have natural width. In the Mössbauer effect we

are essentially concerned with the zero phonon line centred at ω_0 . The intensity of this line is determined by the factor e^{-2M} which is sometimes called the Lamb Mössbauer factor and is similar to the Debye Waller factor in x-ray scattering. We may now calculate the value of this factor in a special case.

For simplicity let us take a cubic lattice and also use the Debye model to evaluate $2M$. In the Debye approximation this works out to be

$$2M = \frac{3\hbar^2 K^2}{M(k_B \Theta_D)} \left[\frac{1}{4} + \left(\frac{T}{\Theta_D} \right)^2 \int_0^{\Theta_D/T} \frac{x}{e^x - 1} dx \right]$$

$$\text{where } x = \hbar\omega / k_B T .$$

Let us now consider the two limiting cases, (i) $T/\Theta_D \ll 1$ and (ii) $T/\Theta_D \gg 1$. For case (i) we obtain,

$$2M = \frac{3\hbar^2 K^2}{4M(k_B \Theta_D)} = \frac{3}{2} \frac{R}{k_B \Theta_D} .$$

For case (ii) we get,

$$2M = \frac{3\hbar^2 K^2}{M k_B \Theta_D} (T/\Theta_D) .$$

In the general case,

$$2M = \frac{3\hbar^2 K^2}{M k_B \Theta_D} \left(1 + \frac{2}{3} - \frac{\pi^2 T^2}{\Theta_D^2} \right)$$

where R is the recoil energy (Fig. 5). Thus the zero phonon absorption cross section is

$$\sigma_a(\omega) = \sigma_0 \left(e^{-\frac{3}{2} \frac{R}{k_B \Theta_D}} \left[1 + \frac{2}{3} - \frac{\pi^2 T^2}{\Theta_D^2} \right] \right) \frac{\gamma^2}{(\omega - \omega_0)^2 + \gamma^2} .$$

The above equation immediately gives the conditions for observing Mössbauer effect. These are (1) R the recoil energy should be small. The energy of γ -rays that can be used for Mössbauer effect have thus been limited to 100 KeV. (2) Θ_D should be large. It implies the binding energy of the solid should be as large as possible. (3) The ambient temperature should be small. Before we discuss the applications of Mössbauer effect, let us

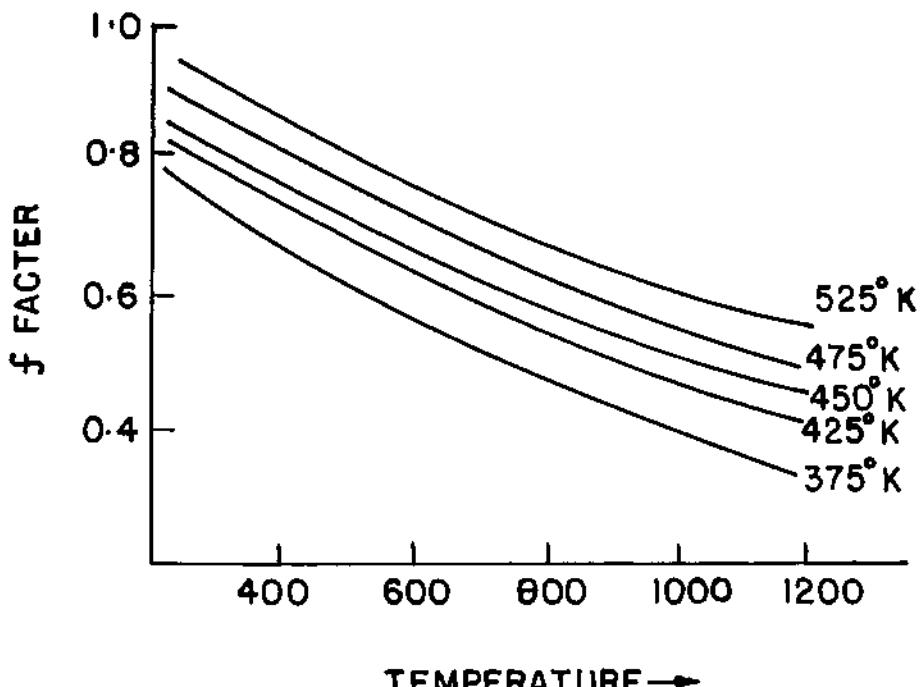


Fig. 5. Variation of Lamb Mössbauer factor for various values of the Debye temperatures.

consider the case when localized mode is present. In the presence of localized mode^{25,26}, the zero phonon cross section gets slightly modified. Here, we get,

$$\sigma(\omega) = \frac{\sigma_0 \gamma e^{-2M}}{2} \left(\frac{I_0(2c_s^2)}{(\omega - \omega_0)^2 + \gamma^2} \right) + \sigma_0^1(\omega) + \sigma_0^2(\omega).$$

We obtain $\sigma_a^0(\omega)$ as

$$\sigma_a^0(\omega) = \frac{1}{2} \sigma_0 \gamma e^{-2M} \prod_s I_0(2c_s^2) \frac{2\gamma}{(\omega - \omega_0)^2 + \gamma^2}$$

where $I_0(2c_s^2) = 1 + \frac{2c_s^2}{L^2} + \frac{6c_s^4}{L^4} + \dots$

The argument of the Bessel function $2c_s^2$ is of $O(N^{-1})$ for the wave like or in band modes of the perturbed crystal where N

is the number of atoms in the crystal. This argument is of $O(1)$ for any of the localized modes that may be present. If N becomes very large, all the factors in the product become equal to one. In such a case the product has to be taken over the localized modes alone. If there are no localized modes, the Bessel function is absent as we have already seen. In the presence of localized modes, we get an additional term $\prod_s I_0(2c_s^2)$. This enhancement is sometimes referred to as the Bessel function enhancement. The Bessel function enhancement is essentially the contribution to the cross section from processes in which the local mode phonons are excited and deexcited in all possible ways. It may be interesting to note that although the local mode phonons are emitted and absorbed, the line shape continues to be Lorentzian with a line width γ . This is really a consequence of harmonic approximation which we have assumed. The process in which local mode phonons are emitted without change in the energy of the lattice before and after the emission or absorption of γ -rays are sometimes called "net zero phonon"²⁷ processes.

It is possible to write the corresponding expression for the one phonon cross section. In this process, one phonon is either emitted or absorbed. Here it is necessary to classify the entire situation in two distinct classes viz., (1) when ω_L , the maximum frequency of the perfect host lattice is greater than the local mode and (2) when the local mode frequency is greater than ω_L . Under these conditions, we may write the one phonon cross section as

$$\sigma_a'(\omega) = \frac{\sigma_0 \gamma}{2} e^{-2M} \sum_s \prod' I_0(2c_s^2) I_1(2c_p^2) \times \\ \left\{ \frac{1}{p^2} \frac{2\gamma}{(\omega - \omega_o - \omega_s)^2 + \gamma^2} + \frac{\frac{1}{p^2} 2\gamma}{(\omega - \omega_o + \omega_s)^2 + \gamma^2} \right\}$$

for $|\omega| > \omega_L$.

In the latter case, we get Lorentzian peaks of the same width as the Mössbauer pip at energies which correspond to $E = E_0 \pm \hbar\omega_s$. The former case gives an exact expression for one phonon absorption cross section due to all modes lying within $0 - \omega_L$. Maradudin²⁵ has given a simplified form for the Debye-Waller factor, zero phonon cross section and the one phonon cross section by making a reasonable assumption that $\gamma = 0$. The lifetime of Fe⁵⁷ in the first excited state is 9.6×10^{-7} sec giving $\hbar\gamma = 4.19 \times 10^{-9}$ eV. The typical value of lattice vibration time is 10^{-13} sec giving vibration energy of the order of 10^{-3} eV. Thus except for frequencies in the bottom of the one phonon

absorption spectrum $\gamma \ll \omega$ is well satisfied. With this simplifying assumption we get

$$\begin{aligned}\sigma_a'(\omega) &= \frac{k^2 \pi \hbar}{m} \sigma_0 e^{-2M} \gamma I_0(2b_0^2) \left\{ \frac{\sin \omega}{-\beta \hbar \omega_0} \right. \\ &\quad \left. \frac{G_0(\omega^2)}{[1 - \epsilon \omega^2 G(\omega^2)]^2 + \pi^2 \epsilon^2 \omega^4 G_0^2(\omega^2)} \right\} \\ &\quad 0 < \omega^2 < \omega_L^2 \\ \sigma_a'(\omega) &= \frac{1}{2} \sigma_0 \gamma e^{-2M} I_1(2b_0^2) \\ &\quad \left\{ e^{-\frac{1}{2} \beta \hbar \omega_s} \frac{2\gamma}{\gamma^2 + (\omega - \omega_0 - \omega_s)^2} + \frac{1}{2} e^{\beta \hbar \omega_s} \frac{2\gamma}{\gamma^2 + (\omega - \omega_0 + \omega_s)^2} \right\} \\ &\quad \omega_s^2 > \omega_L^2\end{aligned}$$

where ϵ is the mass defect parameter defined as $(M' - M)/M$ where M' is the mass of the impurity atom and M is the mass of the host atom and

$$2b_0^2 = \frac{k^2 \hbar}{m} \frac{1}{\epsilon^2 \omega_0^2 \beta(\omega_s^2)} - \frac{e^{\frac{1}{2} \beta \hbar \omega_s}}{\beta \hbar \omega_s - 1}$$

$$G_0(\omega^2) = \frac{1}{3N} \sum_{\vec{k}, j} \frac{1}{[\omega^2 - \omega_j^2(\vec{k})]} .$$

It may be noted that the first expression for $\sigma_a'(\omega)$ is valid when there is either no localized mode or when the localized mode has a frequency less than ω_L . Such a situation arises when the mass of the impurity atom M' is greater than M . In Fig. 6 $\sigma_a'(\omega)$ is plotted against ω^2/ω_L^2 for different values of ϵ when the impurity is in the f.c.c. lattice. $\epsilon = 0$, corresponds

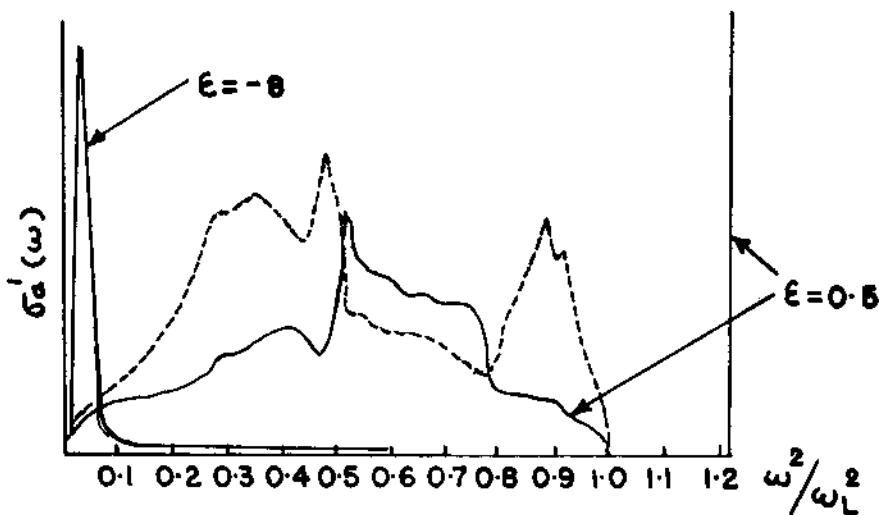


Fig. 6. The shape of the one phonon spectrum for the resonance absorption of gamma-rays by nuclei bound in a crystal for three different values of the mass of the resonance nucleus relative to that of the atoms in the host crystal. The curves are plotted on the bases of nearest neighbour, central force metal of F.C.C. crystal at 0°K.

to a case when no impurity is present in the lattice. In such a case, it is seen from the above expression that the figure gives essentially the plot of $G_o(\omega^2)$ and in this case one phonon absorption cross section merely maps out the frequency spectrum of the host lattice. This property of the Mössbauer effect, namely, one phonon absorption cross section plotting the frequency spectrum, was pointed out by Visscher²⁸. For $\epsilon = 0.5$, we get a localized mode out of the band. However, for $\epsilon = -8$, we have a situation in which the mass of the impurity is some nine times the mass of the host atom. In this case, the one phonon absorption cross section can give very relevant information regarding the perturbation caused by the impurity atom.

The above discussion has brought out the basic fact that σ_a^0 , $2M$ and σ_a' give fund of information regarding the dynamical behaviour of the host lattice and the impurity probe nucleus. σ_a^0 and $2M$ give essentially the same result except when there are localized modes. There is yet another parameter of a Mössbauer

spectrum which essentially gives information on the mean square velocity of the probe nucleus. This affords a check on the lattice dynamical information yielded through σ_a^0 and $2M$ which essentially measure mean square displacement.

Second Order Doppler Effect. The existence of the second order Doppler effect in the Mössbauer spectra was pointed out by Rebka and Pound²⁹ and Josephson³⁰. Mössbauer pip is essentially due to zero phonon process in which the phonon occupation number of the lattice remains unchanged during γ -emission or absorption. It is consequently expected that the γ -ray will carry the full transition energy and the lattice energy will remain unchanged. This is, however, not accurately true. When a nucleus of one of the atoms in a crystal lattice emits a γ -ray photon, then its mass changes by a magnitude $h\nu/c^2$ where ν is the frequency of the γ -ray. The change in the mass of the atom changes the energy of the interacting system to which the emitting atom belongs. Regarding the solid as a system of interacting particles,

the Hamiltonian H^2 is given by

$$\underline{H} = \sum \frac{p_i^2}{2m_i} + V(r_1, r_2, r_3, \dots)$$

As the nucleus of one of the atoms emits a γ -ray, the mass changes by $\delta m_i = h\nu/c^2$. The change in the energy dE of the solid as a consequence is given by $dE = \langle \Delta H \rangle = \delta \sum p_i^2/2m_i = -\delta m_i \langle p_i^2/2m_i \rangle$. $\langle p_i^2/2m_i \rangle$ is the average kinetic energy of the i^{th} atom whose nucleus has emitted a γ -ray. We thus have

$$dE = - \frac{h\nu}{c^2} \langle \frac{1}{2} m_i v_i^2 \rangle$$

The change in the energy of the lattice is reflected in the corresponding change in the energy of the gamma ray photon. Thus, the energy of the γ -ray photon is increased by

$$\Delta E = h\Delta\nu = - dE = (h\nu/m_i c^2) \langle \frac{1}{2} m_i v_i^2 \rangle$$

or $\Delta\nu/\nu = (1/2 c^2) \langle v_i^2 \rangle$. One can observe the change in the frequency of γ -ray and, therefore, obtain information about the mean square velocity. However, there is also the chemical shift which alters the energy of the γ -ray photon because of the electrostatic interaction between the nuclear density and the surrounding electron cloud. The isomer shift is not temperature dependent whereas because of the fact $\langle v^2 \rangle$ changes with temperature, the second order Doppler effect is temperature dependent. If one takes the temperature coefficient of the second order Doppler shift we get,

$$\frac{\partial \Delta\nu}{\partial T} = -\frac{\nu}{2mc^2} \frac{\partial}{\partial T} \left(\langle \frac{p^2}{2m} \rangle \right) = -\frac{\nu}{2mc^2} C_L$$

where C_L is the specific heat of the lattice. In the high temperature approximation $C_L = 3R$ and consequently

$$\frac{\partial \Delta\nu}{\partial T} = -\frac{\nu}{2mc^2} \times 3R = 2.44 \times 10^{-15} \times \nu \text{ per } ^\circ\text{K.}$$

Pound and Rebka²⁹ were the first to demonstrate the existence of second order Doppler effect. We reproduce the results obtained by them which bring out essential validity of the above expression. Fig. 7 shows the fractional shift in the energy $\Delta\nu/\nu$ as a function of temperature of the source for 14.4 keV γ -rays emitted by Fe⁵⁷. The solid line is derived assuming Debye temperature of 420°K. The agreement is indeed very good.

At ordinary temperatures, we have

$$\frac{\Delta\nu}{\nu} \frac{2c^2}{\nu} = \langle v^2 \rangle = \frac{3kT}{M} \left[1 + \frac{1}{20} \left(\frac{\theta_D}{T} \right)^2 - \frac{1}{1680} \left(\frac{\theta_D}{T} \right)^4 \right].$$

At high temperatures, $\langle v^2 \rangle = 3kT/M$ given essentially by equipartition theorem. It is independent of θ_D and consequently independent of crystalline forces or the nature of the crystal.

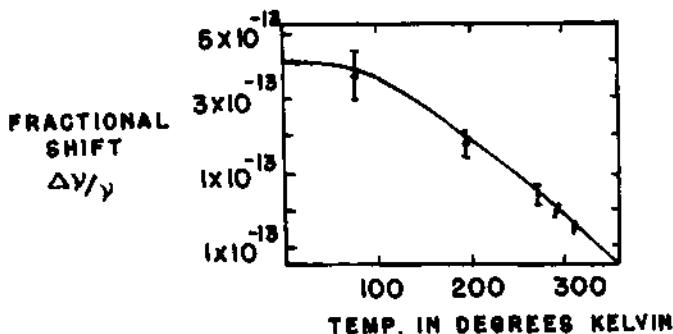


Fig. 7. Fractional shift in the energy of 14.4 keV gamma-ray emitted by Fe⁵⁷ as a function of the temperature of the source.

The above expression for the mean square velocity is true for an isotropic harmonic lattice described by the Debye model. However, if the Mössbauer probe is an impurity in the lattice then there is a possibility that the force constants remain the same, following Flinn and Ruby³¹, we may write

$$\langle v^2 \rangle = \frac{3kT}{M} \left[1 + \frac{1}{20} \frac{m}{M} \left(\frac{\theta_D}{T} \right)^2 - \frac{1}{1680} \left(\frac{\theta_D}{T} \right)^4 \left(\frac{m}{M} \right)^2 + \dots \right].$$

This merely implies that the effective Debye temperature as seen by the impurity probe has changed to a value $\theta_D' = (m/M)^{1/2} \theta_D$, where m is the mass of the impurity and M is the mass of the host atom. If the presence of an impurity alters the force constants, then for a f.c.c. lattice we get

$$\langle v^2 \rangle = \frac{3kT}{m} \left[1 + \frac{1}{12} \frac{\hbar}{kT} \left(\frac{m}{M} \right) \frac{12(\alpha'_1 + \beta'_1) + 6(\alpha'_2 + \beta'_2)}{3M} + \dots \right]$$

where α'_1 , β'_1 , α'_2 and β'_2 are two force constants between nearest neighbours. This gives

$$\frac{\Delta\nu}{\nu} = \frac{3kT}{2mc^2} \left[1 + \frac{1}{12} \left(\frac{\hbar}{kT} \right)^2 \frac{m}{M} \left(\frac{12(\alpha'_1 + \beta'_1) + 6(\alpha'_2 + \beta'_2)}{3M} \right) + \dots \right].$$

If we now plot $(\Delta\nu/\nu) - (3kT/mc^2)$ against $1/T$ at high temperature limit, we will get a straight line whose slope will give

$$\frac{3k}{2mc^2} \quad \frac{1}{12} \quad \frac{\hbar^2}{k} \quad \frac{12(\alpha'_1 + \beta'_1) + 6(\alpha'_2 + \beta'_2)}{3M}.$$

Fig. 8 shows the variation of $(\Delta\nu/\nu) - (3kT/2mc^2)$ against $1/T$ for Fe⁵⁷ in Be and Fe⁵⁷ in stainless steel³². The results show that an iron atom in Be lattice is somewhat more tightly bound than the Be atoms themselves.

Of all the parameters of a Mössbauer spectrum, at the present time there appear to be three dynamical quantities of interest in connection with the resonance absorption of gamma rays. These are the zero phonon absorption cross sections giving essentially the Lamb Mössbauer factors, the one phonon absorption cross section yielding essentially the same information of the localized modes if any and through it the information of the force constant between the impurity and the host atom and the second order Doppler effect yielding information about the mean square velocity of the Mössbauer probe.

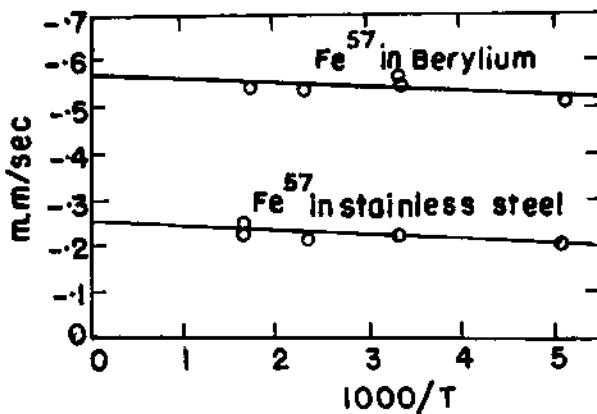


Fig. 8. The variation with temperature of the quantity $(S - 3kT/2mc^2)$ where S is a total observed centre shift at high temperatures and which is the sum of the Isomer shift and the second order Doppler shift for Fe^{57} in beryllium and Fe^{57} in stainless steel.

In addition to Mössbauer spectroscopy, there are several methods which give essentially the same information. These are thermal diffusion scattering of x-rays, neutron spectroscopy, and infrared absorption experiments. From the standpoint of obtaining fundamental lattice dynamical information, no method at present surpasses the coherent inelastic one phonon neutron scattering experiments. These experiments permit one to measure the energies of the individual normal modes of a crystal. The coherent scattering of neutrons from crystals essentially probes the dynamics of an individual normal mode of a crystal. In contrast, Mössbauer effect probes the dynamics of the probe nucleus. In this respect, the Mössbauer effect is similar to the incoherent inelastic scattering of neutrons from crystals and yields the frequency spectra of the host lattice. Since this method gives information regarding the atomic motion of the probe nucleus, it can give fund of information about the dynamical properties of the impurity atoms in the crystal. There does not seem to be any other method at the present time yielding information about the mean square amplitude and velocity of the impurity atom in the crystal. It is equally interesting to point out that Mössbauer effect provides a bench tool to study the dynamical properties of an impurity atom in a crystal. We will now describe some of the experimental results.

IV. EXPERIMENTAL RESULTS

In spite of the considerable development in the techniques of Mössbauer spectroscopy and also the fund of information which one phonon absorption cross section can yield regarding localized modes, no meaningful investigations have so far been reported on the one phonon cross section. The experimental difficulties involved in the measurement of one phonon cross section lie essentially in the large relative velocities to map one phonon component of the localized mode as also the very low value of cross section. As pointed out earlier, if experiments succeed at a later date, the one phonon cross section would yield information about the frequency of the host crystal and about localized modes if they exist. So far experiments have in this field been concentrated in either determining the Lamb Mössbauer factor or the second order Doppler shift.

Lamb Mössbauer Factor. Large number of measurements have been reported on the actual values of the Lamb Mössbauer factor (LM), its variation with temperature for essentially two Mössbauer isotopes Fe^{57} and Sn^{119} . Since investigations have also been reported for the variation of the L.M. for Fe^{57} in iron and Sn^{119} in metallic tin, it has been possible to determine the Debye temperature of these two lattices. In addition to these studies, several investigations have been reported on measurement of L.M. and its variation with temperature when both Fe^{57} and Sn^{119} were impurity atoms in different host lattices.

Careful measurements have been made for Sn^{119} in various host lattices such as vanadium, gold, platinum and thallium by Bryukhanov et al.³³ All these host lattices are cubic Bravais lattices and, therefore, it is possible to use the expression for the L.M. discussed earlier. This expression assumes that Sn^{119} acts as a mere mass defect in the crystal. Knowing the frequency spectrum of vanadium determined experimentally by Chernoplekov et al.³⁴, it is possible to determine the values of $G_0(\omega^2)$ and $\tilde{G}_0(\omega^2)$. Using these computed values, one can determine the L.M. and compare these computed values with the experimentally determined ones. The analysis has shown that the agreement is extremely good. In the absence of lattice vibration spectrum of gold, platinum and thallium, the authors reasonably assumed the Debye model and calculated the L.M. The agreement in these lattices is also equally good. Although, Sn^{119} can give rise to localized mode in gold, platinum and thallium, yet the mean square amplitude of the impurity probe can be satisfactorily explained by assuming that the Debye temperature is modified in such a way that the effective Debye temperature is $\Theta_{\text{eff}} = (M/M')^{\frac{1}{2}} \Theta_{\text{host}}$ where M is the mass of the host atom and M' that of the impurity atom. Similar investigations have been reported for Fe^{57} in various crystal lattices.

Such measurements have been made in beryllium, copper, tungsten, palladium, platinum, rubidium and titanium by Steyert and Taylor³⁵, in beryllium and copper by Housely *et al*³⁶ and in copper, vanadium and titanium by Moyzer *et al*³⁷. Schiffer *et al*³⁸ studied the recoil free fraction at different temperatures. Instead of measuring the absolute values of L.M. at these temperatures they determined the ratios of L.M.'s at 298 °K and 425 °K to its value at 82 °K. The experimental results as in the case of Sn¹¹⁹ showed that the L.M. is generally given by the Debye model in which the effective Debye temperature is given by $\theta_{\text{eff}} = (\frac{M_{\text{host}}}{M_{\text{imp}}})^{\frac{1}{2}} \theta_{\text{host}}$.

Steyert and Taylor³⁵ measured the actual values of 2M for Fe⁵⁷ in Ir, Rh, Pt, Cu, Pd, Au and Ti at 4 °K, 213 °K and 773 °K. Although their values differ from the values reported by Nikovaev and Yakimov³⁹ for gold and from the results of Schiffer *et al*³⁸ for Cu and Pt, the general agreement between theory and experiment is good. Steyert and Taylor³⁵ have further found that the changes in force constants associated with Fe⁵⁷ as seen experimentally agree with the theoretical prediction of Maradudin and Flin⁴⁰. The coupling between Fe⁵⁷ and the host atoms was seen to be loose in Kr, Rh, Pt, Pd whereas it was more tight in Cu and Au.

In contrast to these results, Steyert and Craig⁴¹ have reported that for Fe⁵⁷ in indium, the L.M. has a very weak temperature dependence. Fig. 9 shows the variation of L.M. for Fe⁵⁷ in indium as a function of temperature. Steyert and Craig tried to explain these results on the suggestion that iron atoms in the lattice behave as if they are in a box.

Boyle⁴² and Perlow⁴³ and Hafemeister *et al*⁴⁴ determined the L.M. for Cs¹³³ and I¹²⁹ in various diatomic lattices in which these isotopes were a part of the lattice. Thus Boyle and Perlow determined the L.M. for Cs¹³³ in CsF, CsCl, CsBr and CsI whereas Hafemeister *et al* determined the L.M. for I¹²⁹ in LiI, RbI, NaI and CsI. These alkali halides are simplest of the diatomic lattices for which detailed theoretical study of lattice dynamics is available. Using the expression $2M = \exp(-3/2(R/k\theta_D))$, they calculated θ_D from the knowledge of 2M. Their values are given in table

	2M	θ_D calculated from L.M.	θ_D
CsF	1.59 ± 0.06	109 ± 1	-
CsCl	1.45 ± 0.07	106 ± 1	-
CsBr	1.35 ± 0.007	104 ± 1	149
CsI	0.94 ± 0.08	96 ± 2	125

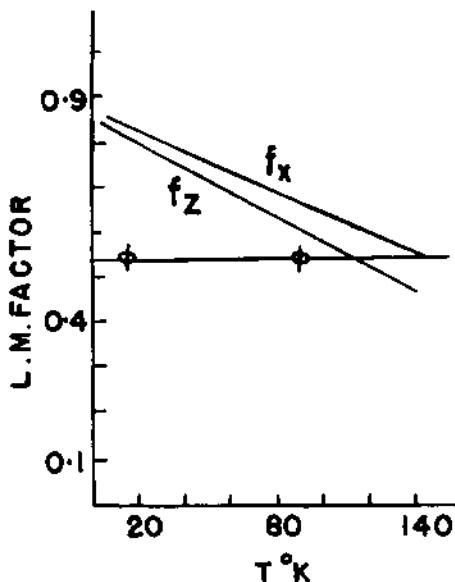


Fig. 9. Variation of Lamb Mössbauer factor for Fe^{57} in indium as a function of temperature. f_z and f_x are the Debye Waller factors for gamma-rays with nucleus moving along Z and X axis respectively. Z axis is the fourfold axis of tetragonal Indium lattice.

	$2M/2M$ for NaI	L.M. determined from line width method	$(c_{11} \times a)$	θ_D
LiI	1.0 ± 0.03	0.23 ± 0.05	8.6×10^3	311
NaI	1.0 ± 0.03	0.29 ± 0.05	9.8×10^3	185
KI	-	-	9.8×10^3	157
RbI	0.8 ± 0.02	-	9.5×10^3	120
CsI	0.9 ± 0.2	0.24 ± 0.05	-	96

a is the lattice spacing.

L.M. involves the mean square displacement. The contribution to the mean square displacement by each of the phonon modes will be inversely proportional to the frequency w of the mode. At 0°K, the modes are equally populated and since the frequency

distribution increases roughly as w^2 , the major contribution to the mean square displacement will come from the region just inside the zone boundary. For a diatomic lattice with a significant energy gap at the zone boundary, the polarization of the modes is such that in the optical branch, the heavier mass is virtually at rest, while in the acoustic branch it is the lighter mass that is at rest. Thus the optical branch makes practically no contribution to the L.M. if the Mössbauer probe is the heavier atom. The major contribution arises from the motion of this atom in the potential of its almost stationary neighbours. In alkali iodides, the mass of the emitters Cs¹¹³ is constant. On this picture any variation in the L.M. could arise only through the difference in the force constants. This picture obviously loses validity as the two masses become equal. In this case, there will be some contribution to L.M. because of the optical branch.

The low temperature specific heat is characterised by quite a different portion of the phonon spectrum and depends only on the slope of the dispersion relation near the centre of the zone. In the long wavelength acoustic modes, both masses move more or less in phase and the Debye temperature will depend upon the sum of the masses as also on the non central force constants which determine the velocity of the low frequency transverse wave. Thus Θ_D calculated from specific heat is expected to decrease more rapidly than Θ_D calculated from L.M. with increase in the mass of the lighter component.

Kagan and Maslov⁴⁵ have derived relations between L.M. and Θ_D for monoatomic and diatomic cubic lattices and have made some numerical calculations for different values of the two masses and for the ratio $\xi = C_{11}/C_{44}$. Hafemeister *et al* found excellent agreement between Kagan's calculations and their experimental values for NaI, RbI but for LiI the experimental values are considerably lower.

Second Order Doppler Effect. So far, experiments designed to obtain information regarding lattice dynamics from mean square velocity have been done on simple lattices for obvious reasons. The second order Doppler effect can be expressed in terms of the relative velocity imparted to either the source or the absorber as follows: If the relative velocities required to cause resonance at two temperatures T_1 and T_2 are

$$\begin{aligned} v_1 - v_2 &= \frac{1}{2c} \left[\langle v_s^2(T_1) \rangle - \langle v_s^2(T_2) \rangle \right] \\ &= \frac{3k}{2Mc^2} (T_1 - T_2) + \frac{3}{40} \frac{k}{Mc} \Theta_D^2 \left(\frac{1}{T_1} - \frac{1}{T_2} \right) + \dots \end{aligned}$$

The value of Θ_D can thus be obtained knowing V_1 and V_2 and the two temperatures. Maradudin *et al.*⁴⁶ determined Θ_D for Fe⁵⁷ in Be. Similar measurements on the variation of second order Doppler shift with temperature have been made for Co⁵⁷, in Be, Cu, W and platinum by Shiffer *et al.*⁴⁷ The Debye temperatures obtained are in qualitative agreement with the value of the Debye temperature determined using L.M.'s. The Debye temperature determined by both these methods agree with $\Theta_{\text{eff}} = (M_{\text{host}}/M_{\text{imp}})^{\frac{1}{2}} \Theta_{\text{host}}$. Steyert and Taylor³⁵ determined second order Doppler shift at various temperatures for Fe⁵⁷ in Au, Cu, Ir, Pd, Pt, Rh, Ti. The value of the Debye temperature determined by these authors agree with the above expression.

Study of L.M. in Ferroelectrics and Detection of the Soft Mode. Ferroelectrics similar to ferromagnetics exhibit reversible spontaneous polarization and show dielectric hysteresis. Various theories have been proposed to explain ferroelectricity in crystals but the one which has shown some promise is the one due to Cochran^{48,49}. The theory is based on lattice dynamical considerations. We will, in this section, assume the results of Cochran. He suggested that in ferroelectrics there exists a temperature dependent optical mode which disappears at the Curie temperature. The crystal on becoming unstable against this mode undergoes a phase transition at the critical temperature. Using Lyddane-Sachs-Teller relation⁵⁰, $\Pi(\omega^2/\omega_T^2) = \epsilon_s/\epsilon_e$ where ϵ_s is the static dielectric constant, Cochran was able to show that at the transition temperature the static dielectric constant will become infinite. if one of the transverse optical mode becomes zero. He further assumed that this mode has a temperature dependence of the type $\omega_T^2 = C(T-T_c)$ where T_c is the temperature at which the crystal becomes unstable. Using this temperature dependence he could explain another dominant characteristic of ferroelectrics, namely, in the temperature region $T > T_c$, Curie Weiss Law is satisfied. In the following section we will analyse the effect of this temperature dependent optical mode on the Lamb Mössbauer factor and examine the possibility of verifying the conclusions through Mössbauer spectroscopy. Indeed the connection of the anomalous optical mode in ferroelectrics to the recoilless fraction was first suggested by Muzikar *et al.*⁵¹. Unfortunately they did not carry the derivation through to its conclusion to get the analytical form of the variation of L.M. with temperature. The first experimental evidence of the anomalous change in the L.M. at the transition temperature was afforded by Bhide and Multani⁵². Since then several investigations have been reported. We will analyse these after we compute an analytical form of the L.M. in these class of compounds.

We recall that Lamb Mössbauer factor is given by

$$f(T) = \exp \left[- \sum_s \frac{\hbar}{2M\omega_s} (\vec{K} \cdot \hat{\epsilon}_s)^2 \coth \frac{\hbar\omega_s}{2kT} \right]$$

$$= \exp \left[- \sum_s \frac{\hbar}{2M\omega_s} (\vec{K} \cdot \hat{\epsilon}_s)^2 (2n_s + 1) \right]$$

Let us now assume that one of the s modes is anomalous and designate it by 'a', then we will have $f(T)$ in the form

$$f(T) = \exp \left[- \sum_{s \neq a} \frac{\hbar}{2M\omega_s} (\vec{K} \cdot \hat{\epsilon}_s)^2 (2n_s + 1) \right] \times$$

$$\exp \left[- (2n_a + 1) \frac{\hbar}{2M\omega_a} (\vec{K} \cdot \hat{\epsilon}_a)^2 \right].$$

Removal of one mode 'a' from the sum should not have much effect, as we have shown previously that the other modes are not affected by the anomalous behaviour of the temperature dependent mode. We, therefore, use the same expression for the first factor as we normally have for a crystal. Taking the Debye model, we get

$$f(T) = \exp \left[- \frac{(\hbar k)^2}{2M} \frac{6}{K_B \Theta_D} \left[\frac{1}{4} + \left(\frac{T}{\Theta_D} \right)^2 \int_0^{\infty} \frac{x^2 dx}{e^x - 1} \right] \right]$$

$$\times \exp \left\{ -(2n_a + 1) \left[\frac{(\hbar k)^2}{2M\hbar\omega_a} \epsilon_a^2 \right] \right\}.$$

The second exponent in the above equation is

$$Q(\omega_a, T) = (2n_a + 1) \left[\frac{(\hbar k)^2}{2M\hbar\omega_a} \epsilon_a^2 \right].$$

To evaluate $Q(\omega_a, T)$, we substitute the value of n_a and ω_a ,

$$Q(T) = \left[2 / \left[e^{\frac{(\hbar/K_B T) \sqrt{G(T-T_o)}}{-1}} + 1 \right] \right] \left[\frac{(\hbar k)^2 \epsilon_a^2}{2M\hbar} \frac{1}{\sqrt{G(T-T_o)}} \right]$$

$$\text{when } T = T_o \text{ both the terms } 2 / \left\{ e^{\frac{\hbar}{K_B T} \sqrt{G(T-T_o)}} - 1 \right\} \text{ and } 1 / \sqrt{G(T-T_o)}$$

go to infinity and consequently $f(T)$ reduces to zero. Thus it is expected that in the second order transition the Lamb Mössbauer factor will decrease to zero as T approaches T_o . In the first order transition T never equals T_o because instability sets in and the transition takes place even before T_o or even before ω^2 becomes zero. In this case we may write $G(T - T_o) = G(T - T_c)^2 + G(T_c - T_o)$; now the factor $G(T_c - T_o)$ is constant.

$$Q(T) = \left[\frac{2}{\left\{ \frac{\hbar}{K_B T} (\sqrt{G(T-T_c)} + G(T_c - T)) \right\}} + 1 \right] \left[\frac{(\hbar k)^2 \epsilon_a^2}{2M\hbar} \frac{1}{\sqrt{G(T-T_c)} + G(T_c - T_o)} \right]$$

This has two fold effects. One of them is as before, namely, as temperature approaches the Curie temperature from the higher temperature side, the f factor will decrease anomalously giving a minimum at $T = T_c$ and furthermore there will be broadening of the resonance line.

There are large number of parameters in the above equation which determine the analytical variation of $f(T)$ with T . A complete specification of the variation of $f(T)$ would necessitate the knowledge of Θ_p , ϵ_a^2 , G , T_o . These parameters may have different values below and above the Curie temperature. It is equally necessary to know T_c . There are thus nine parameters whose knowledge is essential to determine the function $f(T)$. To add to this confusion if ϵ_a^2 is temperature dependent, then the problem becomes really messy.

Bhide and Multani⁵² demonstrated for the first time the anomalous variation in the Lamb Mössbauer factor for Co^{57} doped BaTiO_3 , which has a Curie temperature of 120°C . Fig. 10 shows the plot of normalized area under resonance against temperature. The variation is in qualitative agreement with what has been inferred earlier. Bhide and Hegde⁵³ showed by incorporating Co^{57} as well as Sn^{119} in PbTiO_3 that such anomalous changes in the f factor occur in ferroelectric lead titanate at its Curie temperature (Fig. 11). Svyarevskii⁵⁴ et al observed such changes in iron doped lead niobate. Bokov et al⁵⁵ doped BaTiO_3 with Sn^{119} and observed a dip in the f factor-temperature curve at the Curie temperature. In the earlier experiments particularly carried out with Fe^{57} , we had Mössbauer probe as an impurity which was both a mass defect as well as a charge defect. This would give rise to considerable changes in theoretical formulation of the problem. In the case of Sn^{119} doped experiments, the Mössbauer probe was

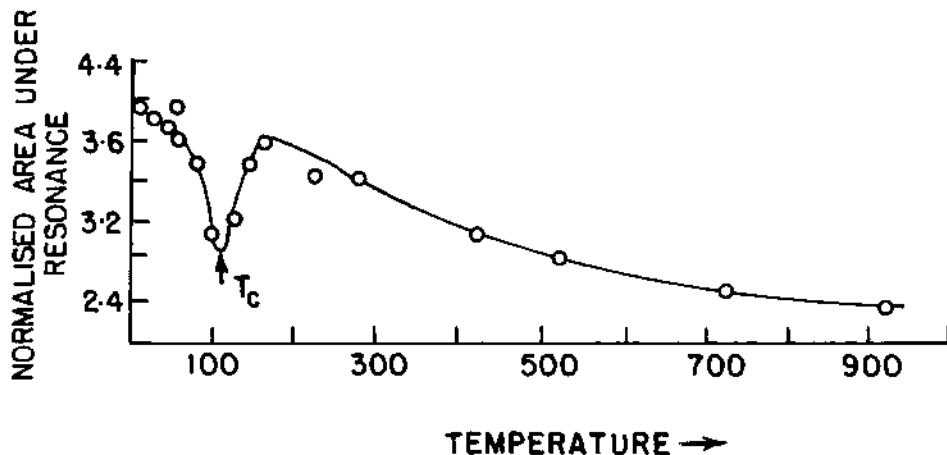


Fig. 10. Variation of the area under resonance as a function of temperature for Fe^{57} : BaTiO_3 . The anomalous dip in the area under the resonance curve is observed at the Curie temperature.

only a mass defect in the lattice and the theory outlined above is applicable. Gleason⁵⁶ studied some ferroelectric lattices in which iron was a part of the host lattice. Thus he investigated the variation of Lamb Mössbauer factor in ferrous ammonium sulphate and observed a dip in the f-T curve near its Curie temperature.

Implicit in the above analysis for the f anomaly at T_c is our ability to describe the lattice with an effective harmonic density matrix. This may be valid for many crystals near 0°K but it is questionable for BaTiO_3 at its transition, where the motion is highly anharmonic and the renormalized phonons strongly damped. It is felt that the situation should be analysed theoretically in greater detail before meaningful interpretation of the experimental results can be attempted.

V. CONCLUDING REMARKS

This very selective survey of applications of Mössbauer effect to atomic motion in solids was intended to highlight the power and sophistication of the tool. With such a limited goal,

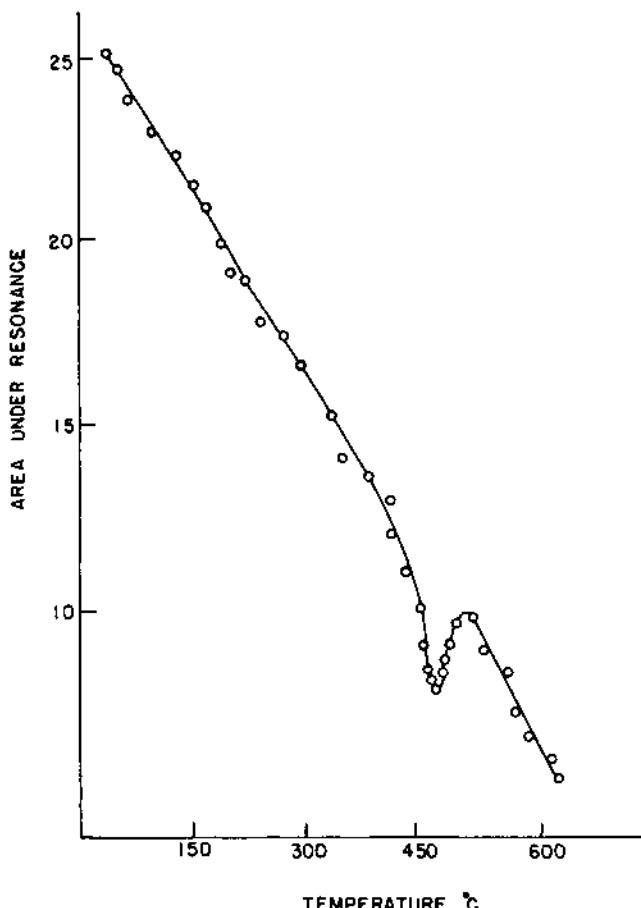


Fig. 11. Variation of area under the resonance with temperature for $\text{PbTiO}_3 : \text{Sn}^{119\text{m}}$ vs. BaSnO_3 .

much relevant material had to be omitted⁵⁷. The large body of excluded research is no less meritorious than the work described here. As is typical for a rapidly expanding field, many experimental results have a rather short lifetime; both the theoretical interpretation and the evaluations are thus continually in need of updating.

REFERENCES

1. R.L. Mössbauer, *Naturwissenschaften* 45, 538 (1958); *Z. Physik*, 151, 124 (1958).

2. R.W. Wood, 'Physical Optics', MacMillan, New York, 1934.
3. W. Heitler, 'Quantum theory of radiation', Clarendon Press, Oxford, 1949.
4. V. Weisskopf, Ann. Physik, 2, 23 (1931).
5. V. Weisskopf and E. Wigner, A. Physik, 63, 54 (1930); 65, 18 (1930).
6. W. Kuhn, Phil. Mag., 8, 625 (1929).
7. E. Fermi, Rev. Modern Phys., 4, 105 (1932).
8. J.D. Jackson, Can. J. Phys., 33, 575 (1955).
9. H.A. Bethe, Rev. Modern Physics, 2, 69 (1937); Phys. Rev., 51, 450 (1937).
10. J.M. Blatt and V. Weisskopf, Theoretical Nuclear Physics, New York, 1952.
11. L. Meitner and H.H. Hupfield, Z. Phys., 67, 147 (1931).
12. K. Zubev, Helv. Phys. Acta, 16, 407 (1943); 16, 429 (1943).
13. E. Pollard and D.E. Alburger, Phys. Rev., 74, 926 (1948).
14. F.R. Metzger, Phys. Rev., 83, 842 (1951).
15. F.R. Metzger and W.B. Todd, Phys. Rev., 95, 858 (1954).
16. P.B. Moon, Proc. Phy. Soc., A64, 76 (1951).
17. F.R. Metzger, Progress in Nuclear Physics, 1959.
18. S. Devons, Nucl. Spectroscopy, Part A, 1960.
19. K.G. Malnofors, in "Beta and Gamma Ray Spectroscopy", K. Siegbahn ed., Amsterdam, North-Holland Pub. Co., 1965.
20. R.L. Mössbauer, Z. Physik, 151, 124 (1958).
21. R.L. Mössbauer, Naturwissenschaften, 45, 538 (1958).
22. W.E. Lamb (Jr.), Phys. Rev., 55, 194 (1960).
23. H. Fraunfelder, The Mössbauer Effect, Benjamin, New York, 1962.

24. K.S. Singwi and A. Sjolander, Phys. Rev., 120, 1093 (1960).
25. A.A. Maradudin, Reviews of Modern Physics, 36, 420 (1964).
26. A.A. Maradudin, Solid State Physics, 18, 367 (1966).
27. B. Kaufman and H.J. Kiptun, Ann. Phys., 18, 294 (1962).
28. W.M. Visscher, Ann. Phys., 2, 194 (1960).
29. R.V. Pound and G.A. Rebka, Phys. Rev. Letters, 4, 274 (1960).
30. B.D. Josephson, Phys. Rev. Letters, 4, 341 (1960).
31. A.A. Maradudin, P.A. Flinn and S.L. Ruby, 'Mössbauer Effect', John Wiley, 1962.
32. A.A. Maradudin, P.A. Flinn and S.L. Ruby, Phys. Rev., 126, 9 (1962).
33. V.A. Bryukhanov, N.N. Delyagin and Yu.M. Kagan, Z. Eks. i. Theor. Fiz., 46, 825 (1964).
34. N.A. Chernoplekov, M.G. Zemlyanov and A.G. Chicherin, J.E.T.P., 16, 1472 (1963).
35. W.A. Steyert and R.D. Taylor, Phys. Rev., 134, 716 (1964).
36. J.G. Dash, R.M. Housley and N.E. Erichson, Rev. Modern Phys., 36, 439 (1964).
37. B. Mozer and G.H. Vineyard, Bull. Am. Phys. Soc., 6, 135 (1961).
38. J.P. Schiffer, P.N. Parks and J. Heberle, Phys. Rev., 133, 1553 (1964).
39. V.I. Nikolaev and S.S. Yakinov, Soviet Phys. JETP, 19, 264 (1964).
40. A.A. Maradudin and P.A. Flinn, Phys. 126, 2059 (1962).
41. W.A. Steyert and P.P. Craig, Phys. Lett., 2, 165 (1962).
42. A.J.F. Boyle and C.J. Perlow, Phys. Rev., 149, 1, 165 (1966).
43. A.J.F. Boyle and C.J. Perlow, Phys. Rev., 151, 1, 211 (1966).

44. D.W. Hafemeister, De Pasquilli and H. De Waard, Phys. Rev., 132B, 1089 (1964).
45. Yu.M. Kagan and V.A. Meslov, Soviet Phys., JETP, 14, 922 (1962).
46. A.A. Maradudin, P.A. Flinn and S.L. Ruby, Phys. Rev. 126, 9 (1962).
47. J.P. Schiffer, P.N. Parks and J. Heberle, Phys. Rev., 133, 1533 (1964).
48. W. Cochran, Phys. Rev. Letters, 3, 412 (1959); Adv. Physics, 10, 401 (1961).
49. W. Cochran and R.A. Cowley, J. Phys. Chem. Solids 23, 447 (1962); Phys. Rev. 134, 981 (1964).
50. R.H. Lyddane, R.C. Sachs and E. Teller, Phys. Rev., 59, 673 (1941).
51. C. Muzikar, J.V. Janovee and V. Drovak, Phys. Stat. Soli., 3, K9 (1963).
52. V.G. Bhide and M.S. Multani, Phys. Rev., 139, 1983 (1965).
53. V.G. Bhide and M.S. Hegde, (in press).
54. V.V. Swayashevshii et al., JETP Letters, 135, 3 (1966).
55. V.A. Bokov, V.P. Romonov and V.V. Chekin, Soviet Phys. - Solid State, 7, 1521 (1965).
56. T.D. Gleason, Ph.D. Thesis, John Hopkins Univ., U.S.A., 1968.
57. For an elementary treatment of Mössbauer effect and chemical applications (including spinels etc.) see
 - (a) R.S. Drago, "Physical Methods in Inorganic Chemistry", Reinhold Pub. Corp., New York, 1965.
 - (b) H.B. Mathur in "Spectroscopy in Inorganic Chemistry", Vol. I, Edited by C.N.R. Rao and J.R. Ferraro, Academic Press, New York, 1970.

POSITRON ANNIHILATION IN SOLIDS

R. M. Singru

Department of Physics

Indian Institute of Technology, Kanpur, India

I. INTRODUCTION

Positron is the antiparticle of electron and its existence was predicted by Dirac's relativistic theory of electron. Positrons were experimentally discovered by Anderson in cloud chamber photographs of cosmic rays. In a material medium, positrons do not live long because they can either annihilate with electrons in the medium, or can form a metastable system of a positronium atom which is short-lived. In the first type of decay, called prompt or free annihilation, a positron, after thermalisation, annihilates with an electron in the medium either through a singlet collision (spins antiparallel) or through a triplet collision (spins parallel). Selection rules governing the annihilation show that a singlet collision results in emission of two photons in exactly opposite direction and each one having an energy of about 511 ($= m_0 c^2$) KeV. This annihilation has a lifetime $\sim 10^{-10}$ s. On the other hand, a triplet collision gives rise to an annihilation into three photons with a lifetime $\sim 10^{-7}$ secs. In this case the total rest energy ($= 2 m_0 c^2$) of the electron-positron pair is shared between the three photons which are emitted in one plane. The ratio of singlet (2 photon) to triplet (3-photon) annihilation rates was calculated by Ore and Powell² to be

$$\lambda_s / \lambda_t = 1115 \quad (1)$$

A positron can be bound to an electron forming a bound state ($e^+ e^-$) called "positronium atom". This lightest atom, positronium, is denoted by a symbol Ps and has properties similar to the hydrogen atom except for quantum electrodynamic details. For example, the

reduced mass of the positron-electron pair being $1/2 m_0$ (where m_0 is the rest mass of the electron) the energy levels of Ps atom are given by

$$E_n = - \frac{m_0 e^4}{4n^2 h^2} \quad (2)$$

Thus the Ps - atom has the following properties:

- i) the ionisation potential is 6.8 eV,
- ii) the Bohr radii are about twice the corresponding radii for hydrogen atom,
- iii) the first excited state has an energy of 5.1 eV,
- iv) the wavelength of Lyman α line is 2430 Å.

The ground state of Ps-atom consists of a lower 1S_0 singlet (para - Ps) state and higher 3S_1 triplet (ortho - Ps) state depending upon the relative spin-directions of the electron and positron. The energy splitting between these two states is $\sim 10^{-3}$ eV and the radiative transitions between them occur at a very slow rate. If Ps - atom is formed in an excited state ($l > 0$), this results in a radiative transition to the ground state followed by annihilation.

The singlet state 1S_0 decays through two photons with a singlet lifetime $\tau_s = 1.25 \times 10^{-10}$ s, while the triplet state decays with a triplet lifetime $\tau_t = 1.4 \times 10^{-7}$ s by emission of three photons.

Earlier experiments with positron annihilation were concerned with the verification of predictions made by quantum electrodynamics. However, present experimental research in positron annihilation uses positron as a probe to investigate two fields: chemistry and solid state. Positron annihilation gives rise to gamma rays and usually two kinds of measurements are made with the annihilation radiation. In the first method, one measures the angular distribution of the gamma rays while in the second, one measures the lifetimes of positrons in the material under study. There is also a third kind of measurement, three-photon coincidence measurement, which determines the rate of three-photon decay, but this experiment is tedious and is not performed so frequently.

II. METHODS AND RESULTS

Angular Distribution Measurements. If the annihilating electron-positron pair has zero momentum, then the pair would annihilate into two gamma rays emitted exactly 180° apart. The positron being thermalised in the material before annihilation,

has negligible momentum. However, electrons will have a momentum distribution upto the Fermi momentum. This will give rise to a finite momentum for the electron-positron pair and as a result the two annihilation gamma rays will come out at a certain angle ($180 \pm \theta$), θ being decided by the momentum of the pair. Usually θ is of the order of milliradians.

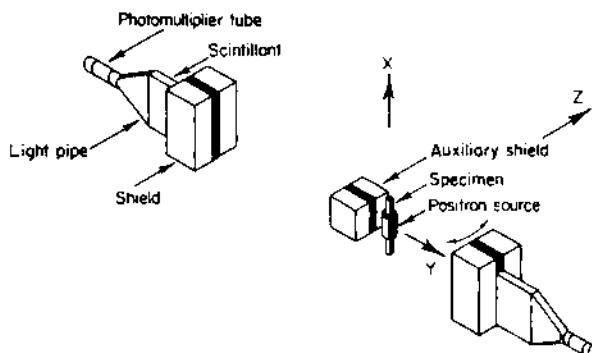


Fig. 1. Experimental set-up⁵ for measuring the angular distribution of the photons resulting from positron annihilation in a solid specimen. (J. GREEN and J. LEE, Positronium Chemistry, Academic Press, 1964).

A typical experimental arrangement to observe the angular distribution of annihilation radiation is shown in Fig. 1. In the experiment usually a projection of angle θ on a plane (say YZ plane) formed by the axis joining the two detectors and the direction of the movement of the source and the sample is measured. In the geometry shown in Fig. 1 the detectors are sensitive only to the Z-component ($\hbar k_z$) of the electron momentum and the measured angle $\theta = \hbar k_z/mc$. Elementary considerations³ show that the observed curve showing coincidence counting rate as a function of measured angle θ should be given by

$$\begin{aligned} I(\theta) &= \beta (k_F^2 - k_z^2) \quad \text{for } k_z \leq k_F \\ &= 0 \quad \quad \quad k_z \geq k_F \end{aligned}$$

where k_F is the Fermi momentum of the electrons in the metal. The free electron theory, therefore, predicts the angular distribution curve as an inverted parabola.

Observed angular distribution curve⁴ indeed shows this parabolic curve confirming above considerations but also shows a 'tail' at higher angles corresponding to momenta higher than k_F . These tails are generally ascribed to annihilation of positrons with core electrons. Angular distribution curve for simple metals like Li,

Na, K, Cs, Ca and Al consist of an inverted parabola with a sharp cut-off corresponding to Fermi momentum and a weak tail persisting

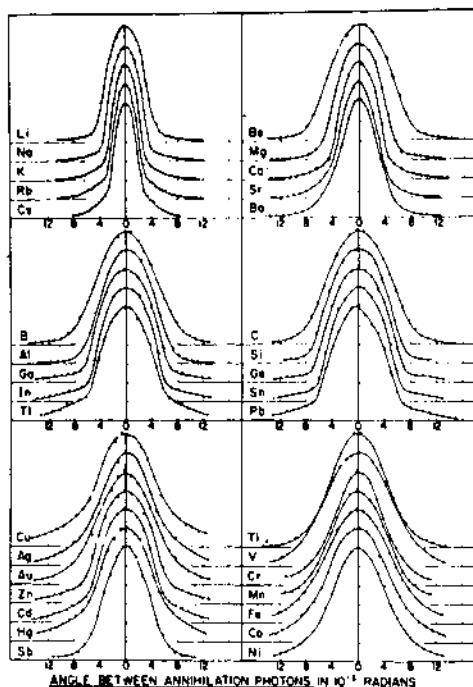


Fig. 2. Angular distribution of photons from positrons annihilating in various metals. (After A.T. STEWART, Canad. J. Phys., 35, 168, 1957).

beyond it (Fig. 2). On the other hand, for metals like Cu, Ag, Au, the transition and rare earth metals, the angular distribution curve is like an inverted parabola but possesses a larger tail. Lang and De Benedetti⁵ have discussed the deviations of the angular correlation curve from the parabolic shape and suggest following effects responsible for such deviations: (i) non-thermalisation of positrons, (ii) non-spherical Fermi surface, (iii) annihilation of positrons with valence as well as "core" electrons and (iv) the "excluded volume" effect arising out of the zero-point motion of the positrons.

The interaction of a positron with the electron gas in metal can be better understood if one knows the effective mass of positron. An estimate of the effective mass m^*/m_0 , in units of bare mass m_0 , can be estimated by measuring the angular distribution curve of the annihilation photons at different temperatures and examining the thermal smearing of the curve. Such measurements by Stewart et al.⁶ have indicated that $m^*/m_0 \sim 2$.

The measurement of two-photon angular distribution, as seen above, can offer valuable information about momentum distribution of electrons in metals. In a way, this method is superior to X-ray spectroscopic methods because it helps to determine the shape of Fermi surfaces and zones in the k-space. Measurement of angular distribution of photons in positron annihilation has become a powerful tool in the study of metal physics and such measurements have been made in metals, alloys, ferromagnetics, liquid metals, etc. Experimental and theoretical work in this field has been reviewed in detail by Ferrell⁵, Stewart⁶, Majumdar⁷ and Dekhtyar⁸.

Lifetime Measurements. We shall consider the energetics of Ps formation before discussing the experimental results on lifetime measurements. Ore⁹ was the first to point out that only positrons having an energy in a certain range can form positronium. This can be easily understood if we look at the formation of Ps in gases where an electron has first to be removed from a gas molecule to bind a Ps-atom. This involves supplying an energy equal to I eV, where I is the ionisation potential of the gas molecule. However, an energy equal to 6.8 eV, the binding energy of Ps atom, is released once Ps atom is formed. Thus the energy threshold for Ps-formation is $E_t = (I - 6.8)$ eV. Gas molecules can be excited and these excitation levels can lie between E_t and I . If the lowest excitation energy is E_1 then for positron energies E such that $I > E > E_1$ excitation collisions will compete with positron capture. The energy range from E_t to E_1 is called Ore gap, and Ps-formation is most probable when the positron energy falls in this gap. As long as positron energy is higher than I , positronium formation is ruled out on energetic considerations. In this energy range, positrons lose their energy by inelastic collisions. Fig. 3 indicates how the probability of Ps formation is expected to depend on positron energy. The curve is imaginary and is shown to indicate the qualitative features of the process.

Fig. 4 shows a flow diagram for positron annihilation processes. Positrons emitted by a radioactive source usually have energies of the order of 100 KeV. As the positrons travel in any material medium, their energy is rapidly reduced by collisions. These thermalised positrons can either undergo a free annihilation, or a bound annihilation. In the latter case depending upon the energy, it can form a chemical bond with a molecule or form Ps in a para or ortho state. Usually the para (singlet) state decays into 2 photons by a mean lifetime of $\tau_s \approx 1.25 \times 10^{-10}$ s and the ortho (triplet) state decays into 3 photons by a lifetime $\tau_t \approx 10^{-7}$ s. However, in most cases the ortho state does not have to wait for 10^{-7} s before decaying. Usually the ortho-state is "quenched" and decays within 10^{-9} s. Any process which allows the ortho-Ps to decay by emission of two photons, thus lowering the lifetime τ_t down to $\sim 10^{-9}$ s, is called a quenching process.

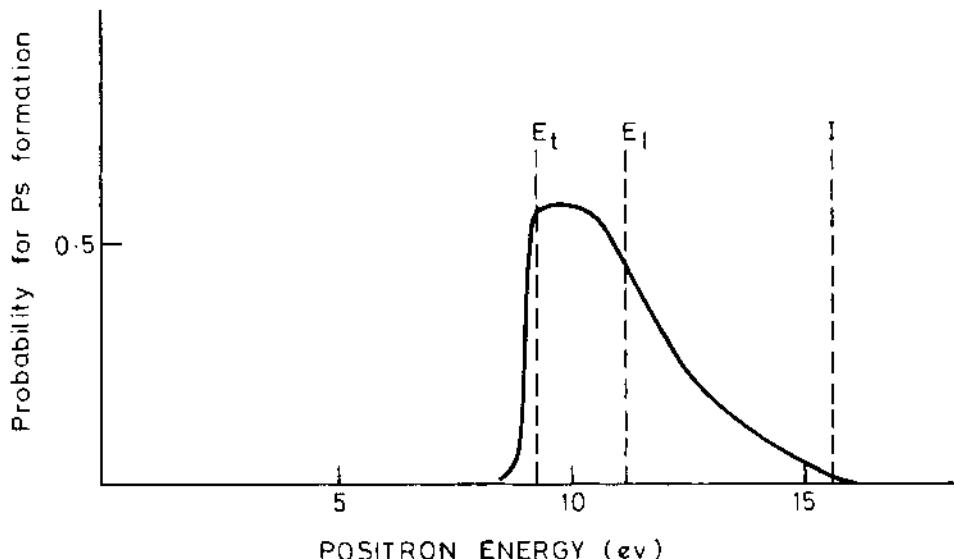


Fig. 3. Hypothetical dependence of Ps formation probability in argon gas as a function of positron energy (J. GREEN and J. LEE, Positronium Chemistry, Academic Press, 1964).

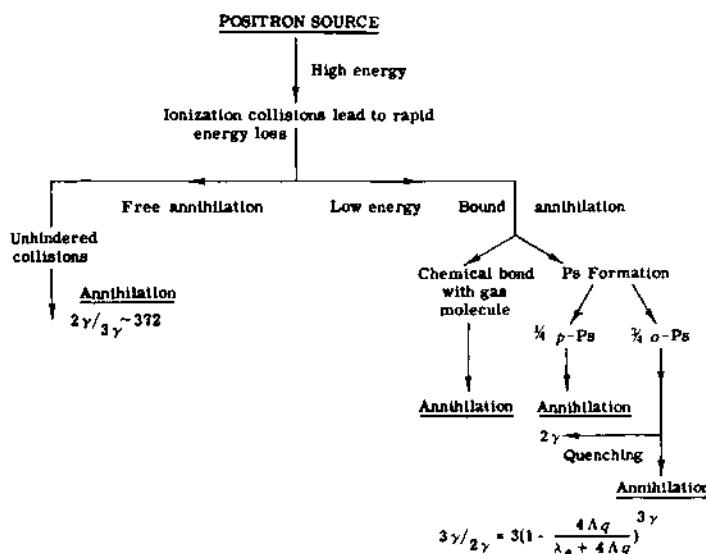


Fig. 4. Flow diagram for positron annihilation. Positronium annihilation results in $3\gamma/2\gamma = 0$, if all positrons form Ps and $\lambda_q = 0$. (J. GREEN and J. LEE, Positronium Chemistry, Academic Press, 1964).

There are several mechanisms which can contribute to the quenching process: for example: unpaired electron exchange, spin flip, chemical compound formation, pick-off and spin reversal etc. In most cases quenching is induced by the pick-off process where ortho-positronium collides with atoms and molecules in the material, and positron of this Ps atom annihilates with an electron of the surrounding medium in a relative singlet state.

It is thus clear that the lifetimes with which positrons annihilate in matter are characterised by the various possible interactions of positrons shown in Fig. 4.

Experimental Method. These lifetimes being in the sub-nanosecond region can be easily measured by nuclear techniques developed to measure nuclear lifetimes. A typical experimental arrangement for lifetime measurement is shown in Fig. 5a. The source of positrons is usually the radioactive material Na^{22} . In the radioactive decay of Na^{22} , positron emission is almost simultaneous with a 1274 - KeV gamma ray (Fig. 5b), while the positron annihilation is accompanied by 511 - KeV gamma ray. Thus, measurement of time interval between the emission of these two gamma rays can yield the lifetime of positrons.

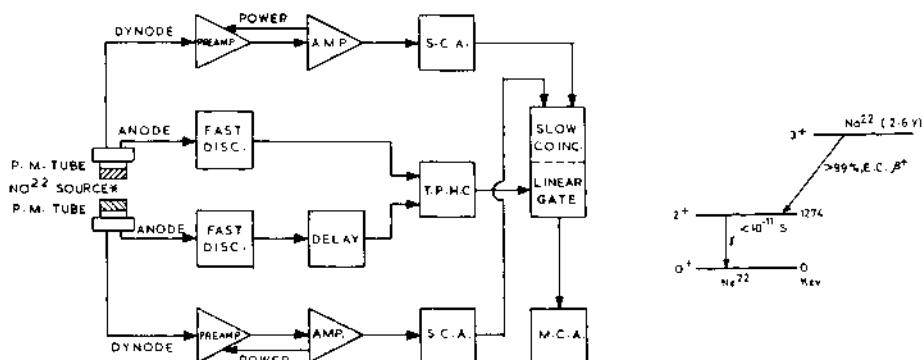


Fig. 5. a) Block diagram of the electronic set-up for the measurement of positron lifetimes, (After ORTEC manual)
b) Decay scheme of Na^{22} .

The gamma rays are usually detected by plastic scintillators optically coupled to fast photomultiplier tubes. A thin source of Na^{22} deposited on a thin film is sandwiched between the material to be studied. The pulses from the photomultiplier are analysed for time and energy information in the usual way¹⁰ and the time distribution spectrum is obtained in a multichannel analyzer. A typical time spectrum obtained for Teflon is shown in Fig. 6. The time-spectrum is analysed so that each lifetime and its relative intensity is determined.

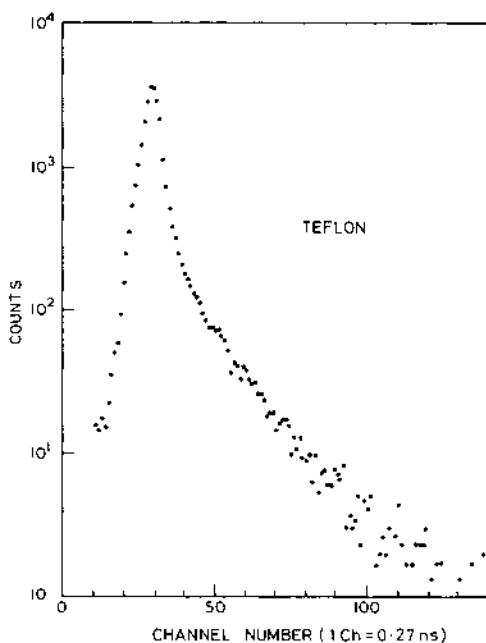


Fig. 6. Typical lifetime spectrum of positrons annihilating in Teflon³³.

The lifetime corresponding to free and para-Ps annihilation is usually denoted by τ_1 , having an intensity I_1 . The quenched lifetime of the ortho-Ps state is usually denoted by τ_2 with intensity I_2 .

Results of Lifetime Measurements: Gases. Lifetimes of positrons in various gases have been extensively studied. During the initial stages the studies were aimed at checking quantum electrodynamics of the process. Later on the focus shifted to positronium formation and quenching. It was soon realized that simple Ore gap theory, though useful for energetic considerations, is not adequate for quantitative predictions of positronium formation. Available

data¹¹ on positronium formation fractions, electric and magnetic field effects, and quenching cross sections of oxygen in Argon, is not completely understood yet¹².

Development of lifetime measurement techniques with better resolution and longer range allowed more interesting problems to be studied. Free positron lifetimes in argon¹³, positron lifetime dependence on gas pressure¹⁴ and other processes¹⁵ have been critically examined.

Recently considerable interest is focussed on the study of reactions involving slow positrons in gases, the so-called "gaseous positronics"¹⁶. One expects to obtain new information about the atomic collision theory by applying it to collisions between positrons (or positronium) and atoms and molecules. There is a lot of scope for both theoretical and experimental work in this area. Recently Brandt¹⁷ has pointed out that studies with subexcitation positrons can yield valuable information about electronic interaction in gases.

Metals, Semiconductors and Oxides. Experimental data¹⁸ show that positrons annihilate in metals with a single lifetime of the order of $\tau_1 \sim 2 \times 10^{-10}$ s. This is understood from the fact that positronium formation is extremely unlikely in metals and all the annihilations must occur via free annihilations. Some results indicate a second lifetime with low intensity but it has been ascribed to spurious effects arising from surface layers or lattice defects. Experimental annihilation rates in metals have been compared¹⁹ with the theoretical rates estimated from many-body calculations considering one-positron-many-electron system and there is reasonably good agreement between the two. The results on lifetime measurements in metals are well summarised by Majumdar⁷.

Positron lifetime spectra in semiconductors and metal oxides usually show two components²⁰. Gianotti *et al.*²⁰ have measured the positron lifetime spectra in V_2O_3 in the temperature range (80 ~ 300)°K in which the transition from semiconductor to metal state occurs. They find that the lifetimes remain constant in this range but the intensities change with temperature. This effect has been ascribed to the screening effect of the conduction electrons on the repulsive action of the V_3^+ lattice ions. With the improvement in the time resolution of the apparatus, these studies have taken a very interesting turn²¹ and it is hoped that the mechanism of positron annihilation in metals, semiconductors and metal oxides would be more clear in near future.

Ionic Crystals. The lifetime spectrum of positrons in alkali halides is complex. It is mainly composed of two components (i) a short lifetime $\tau_1 \sim 2 \times 10^{-10}$ s having an intensity of about 75%,

(ii) a longer lifetime $\tau_2 \sim 5 \times 10^{-10}$ s, with an intensity of about 25%. Sometimes a third longer lifetime appears at the tail of the spectrum with an intensity less than 1%. Ferrell³ ruled out formation of Ps-atom in alkali halides on energetic grounds. Other way to look at this is to realise that classically Ps-atom has a size of 1.08 Å and there is not enough room for Ps atom in an ionic crystal with regular structure. However, there is a possibility that positron can be captured by a negative ion. Indeed, angular correlation experiments have already established the important role played by negative ion in this process. The mechanism responsible for positron annihilation in alkali halides is not completely understood yet. Goldanskii²² has suggested three possible formations of positrons in alkali halides (i) positron bound to the negative ion (ii) positron-polaron state (iii) positron bound to crystal defect.

Bussolati and coworkers²³ interpreted the lifetime spectra in alkali halides in terms of the e^+ - anion state. In support of this view Singh *et al.*²⁴ find that the lifetime of the long-lived component (τ_2) in alkali halides depends on the ratio of the anion and interstitial volumes. Evidence for the effect of color centers and other crystal defects on lifetime spectra^{25,26} and angular distribution curves²⁷ of alkali halides has been reported. Singh *et al.*²⁸ have observed that presence of cation vacancies, anion vacancies and color centers influence the lifetime spectra in KCl crystals. It thus appears that the systems (i) e^+ - anion (ii) e^+ - crystal defect play an important role in the annihilation of positrons in ionic crystals. More experimental work is necessary before the role played by each process is defined.

Positron annihilation has also been studied by Colombino *et al.*²⁹ in some ionic compounds such as hydroxides and sulphates by measuring the angular distribution curves. The results indicate that a fraction of positrons do not form positronium but annihilate almost exclusively in the neighbourhood of the negative ion oxygen.

Molecular Materials. Many measurements of lifetimes of positrons annihilating in molecular materials (i.e. organic compounds, polymers and other condensed materials) have been reported in literature. A typical lifetime spectrum of positrons annihilating in Teflon was shown in Fig. 6. It is characterised by two lifetimes, the faster component $\tau_1 = 4 \times 10^{-10}$ s with ($I_1 = 80\%$), arising due to free annihilation and p-Ps state while the slower component $\tau_2 = 2 \times 10^{-9}$ s with $I_2 = 20\%$, arising out of quenching of the ortho-Ps state. The study of positron annihilation lifetimes in molecular materials provides useful data in understanding the formation and quenching of orthopositronium in these materials.

The process of Ps-formation in molecular materials could perhaps be described by an Ore gap picture with proper modifications. Another quantity that should determine the formation and decay of positronium will be the free volume available in the condensed medium.

Such free volume is necessary for a positronium atom whose classical diameter is 1.08 Å. A triplet Ps atom lodged in such a 'site' (free volume) is expected to be quenched by a pick-off process with the surrounding material. The observables are (i) the longer lifetime τ_2 describing the quenched rate of annihilation of ortho-Ps and (ii) the intensity I_2 of this component which describes the formation probability of Ps-atoms. In molecular materials, therefore, the quantities τ_2 and I_2 should depend on the free volume available as well as the width of Ore gap. Such an attempt to correlate τ_2 and I_2 on the basis of a free volume model has been made by Thosar *et al.*³⁰. However, the model is semiempirical in nature and does not take the dependence on the details of the width of Ore gap into account. Brandt³¹ has also discussed the quantitative aspects of this process.

It is expected that a more systematic (theoretical and experimental) investigation of τ_2 and I_2 in molecular materials will yield a better understanding of the processes involved.

Lifetimes of positron annihilation have been measured in many organic compounds by Gray *et al.*³² who found that I_2 - values depend upon specific atoms or groups in the molecule rather than upon any property of the molecule as a whole. The annihilation quenching cross section has been correlated with electron polarizabilities. Singh *et al.*³³ have also studied positron lifetimes in condensed aromatic hydrocarbons, substituted benzoic acids, nitro-aromatic compounds and heteroaromatic compounds. The results show an interesting behaviour of I_2 with the Hammett's constant σ and ionisation potential I_p , lending support to Ore gap picture.

Interesting measurements on lifetimes of positronium in molecular material irradiated by gamma rays or neutrons have been reported^{34,35}. Such experiments can yield valuable information about free volume concept and about the location of Ps-formation in the material.

Positronium Chemistry. Positronium being the 'lightest atom' can be considered as a "model" free radical. Furthermore, the annihilation gamma rays can allow us to 'label' this atom, and these gamma rays carry the information about how Ps-atom reacts with the surrounding medium. As pointed by Goldanskii³⁶ the observation of positronium quenching and the inhibition of its formation can lead us to a study of (i) elementary processes of

radiation chemistry, (ii) mechanism of oxidative reactions, and (iii) the role of quantum effects in chemical kinetics. Usually, positron lifetimes are measured for studying positronium chemistry.

Positronium can form compounds like Ps Cl having bond energy of 1.55 eV and these compounds can be considered as chemically stable at room temperature. Positronium has some special features³⁷ as compared to the hydrogen radical. For example, positronium can react with the medium before it thermalises. Secondly, the short lifetime of Ps never allows chemical reaction involving Ps to reach an equilibrium state. The field of positronium chemistry is very well reviewed by Goldanskii³⁶, Tao and Green³⁷, Green and Lee³⁸ and McGervey³⁹.

In conclusion, one can say that the study of positron annihilation is now an established research tool in the study of physics and chemistry of solids.

REFERENCES*

1. M. Deutsch, Progr. Nuc. Phys., 3, 131 (1953).
2. A. Ore and J.L. Powell, Phys. Rev., 75, 1696 (1949).
3. R.A. Ferrell, Rev. Mod. Phys., 28, 308 (1956).
4. A.T. Stewart, Canad. J. Phys., 35, 168 (1957).
5. L.G. Lang and S. DeBenedetti, Phys. Rev., 108, 914 (1957).
6. A.T. Stewart, in "Positron Annihilation", edited by A.T. Stewart and L.O. Roellig, Academic Press, New York, 1967.
7. C.K. Majumdar, in "Theory of Condensed Matter", International Atomic Energy Agency, Vienna, 1968.
8. I. Ya. Dekhtyar, Czech. J. Phys., 18B, 1 (1968).
9. A. Ore, Univ. i Bergen Naturvit, rekke 9, (1949).
10. A.Z. Schwarzschild and E.K. Warburton, Ann. Rev. Nuclear Sci., 18, 265 (1968).
11. G.J. Celitans, S.J. Tao and J.H. Green, Proc. Phys. Soc. (London), 83, 833 (1964).

* A comprehensive bibliography on positron annihilation has been compiled by J.A. Merrigan, Eastman Kodak Company, Rochester, N.Y., U.S.A.

12. J.H. Green, in "Positron Annihilation", edited by A.T. Stewart and L.O. Roellig, Academic Press, New York, 1967.
13. S.J. Tao, J.H. Green and G.J. Celitans, Proc. Phys. Soc. (London), 81, 1091 (1963).
14. J.H. Green and S.J. Tao, J. Chem. Phys., 39, 3160 (1953).
15. C.Y. Leung and D.A.L. Paul, J. Phys. (Lond.), B 2, 1278 (1969).
16. H.S.W. Massey, in "Positron Annihilation" edited by A.T. Stewart and L.O. Roellig, Academic Press, New York, 1967.
17. W. Brandt and H. Feibus, Phys. Rev., 184, 277 (1969).
18. H. Weisberg and S. Berko, Phys. Rev., 154, 249 (1967).
19. S. Kahana, Phys. Rev., 129, 1622 (1963).
20. A. Gianotti, C. Ghessi and M. Manbredi, Nuovo Cimento, 62 B, 121 (1969).
21. P. Sen and A.P. Patro, Phys. Lett., 28 A, 414 (1963).
22. V.I. Goldanskii and E.P. Prokop'ev, Sov. Phys. Solid State, 6, 2641 (1964).
23. C. Bussolati, A. Dupasquier and L. Zappa, Nuovo Cimento, 52 B, 529 (1967).
24. K.P. Singh, R.M. Singru and C.N.R. Rao, Chem. Phys. Letts., 3, 247 (1969).
25. H.F. Waung, W. Brandt and P.W. Levy, Bull. Am. Phys. Soc., 13, 1473 (1968).
26. J.L. Williams and H.J. Ache, J. Chem. Phys., 51, 3536 (1969).
27. D. Herlach and F. Heinrich, Phys. Letts., 31 A, 47 (1970).
28. K.P. Singh, R.M. Singru, M.S. Tomar and C.N.R. Rao (To be published).
29. P. Colombino and B. Fiscella, Nuovo Cimento, 58 B, 413 (1968).
30. B.V. Thosar, V.G. Kulkarni, R.G. Lagu and G. Chandra, Phys. Letts., 28 A, 760 (1969).

31. W. Brandt, in "Positron Annihilation", edited by A.T. Stewart and L.O. Roellig, Academic Press, New York, 1967.
32. P.R. Gray, C.F. Cook and G.P. Strum, Jr., J. Chem. Phys., 48, 1145 (1968).
33. K.P. Singh, R.M. Singru and C.N.R. Rao (To be published).
34. G. Chandra, V.G. Kulkarni, R.G. Lagu and B.V. Thosar, Phys. Letts., 19, 201 (1965).
35. B.V. Thosar, V.G. Kulkarni, R.G. Lagu and G. Chandra, Phys. Letts., 21, 647 (1966).
36. V.I. Goldanskii, in "Positron Annihilation", edited by A.T. Stewart and L.O. Roellig, Academic Press, New York, 1967.
37. S.J. Tao and J.H. Green, J. Chem. Soc. A., 408 (1968).
38. J.H. Green and J. Lee, "Positronium Chemistry", Academic Press, New York, 1964.
39. J.D. McGervey, in "Positron Annihilation", edited by A.T. Stewart and L.O. Roellig, Academic Press, New York, 1967.

ELEMENTS OF ORDER-DISORDER THEORY AND DIVERSE APPLICATIONS

J. M. Honig

Department of Chemistry, Purdue University

West Lafayette, Indiana, 47907, U.S.A.

In the spirit of the lectures delivered at the Winter School, the methodology of order-disorder theory will only be outlined, and the mathematical details will be skipped. The aim is to present the reader with a survey of this area and to stimulate him to do additional reading on his own.

I. STATEMENT OF THE PROBLEM

In order to be definitive we shall consider here the case of a two-dimensional, essentially infinite, hexagonal array of sites which are capable of existing in one of two configurations: they can be empty or filled. Filling can take place by exposing the set of sites to a gas at fixed temperature (T) and pressure (p). A certain number of gaseous atoms then are transferred spontaneously to the array and localized on the sites with a concomitant liberation of energy ε . The question to be settled is the following: Corresponding to a given T and p , what fraction of the available sites is filled?

Obviously, a theory which deals with this problem in full generality becomes overwhelmingly complex; for, it is necessary to take into account the lateral interaction of all atoms absorbed on the lattice array. Thus, to deal properly with the configuration of one site, the population of all the remaining ones must be specified. Therefore, we must look for a systematic method of simplification. The method of Huijmans and de Boer² consists in replacing the lattice by a set of simpler arrays, the number of sites in each unit of the array being sufficiently small, so that

only a small number of microvariables is required for a complete description of the statistical properties. The problem then resolves itself to a decomposition of the lattice into simpler figures in such a manner that all lattice properties are faithfully mirrored by those of the simpler arrays. Furthermore, the procedure must be so set up that, as these arrays become increasingly complex, one will be able to approximate the properties of the lattice to any desired degree.

II. THE ZERO ORDER APPROXIMATION (LANGMUIR ISOTHERM EQUATION)

We illustrate the fundamentals of our approach by considering the zero order case where lateral interactions are ignored altogether. Then the lattice may be regarded simply as a collection of L points of which a fraction α_0 is empty and a fraction α_1 is filled; the associated energies will be designated as ϵ_A and ϵ_B respectively. This information is shown in tabular form in Table II.1.

TABLE II.1

Configurations for the Point Representation of a Lattice of Sites

Figure	(1) Designa- tion of state	(2) Probabil- ity of occur- rence	(3) Number of sites in state A or B	(4) Energy of configu- rations	(5) Energy of each set
0	A	α_0	$L \alpha_0$	ϵ_A	$L \alpha_0 \epsilon_A$
0	B	α_1	$L \alpha_1$	ϵ_B	$L \alpha_1 \epsilon_B$

In this approximation the configurational energy of the lattice is given by

$$E_c = L (\alpha_0 \epsilon_A + \alpha_1 \epsilon_B) \quad (1)$$

and the configurational entropy for the lattice points is given by the well-known Boltzmann relation

$$S_c = k \ln \Omega = k \ln \frac{L!}{(L \alpha_0)! (L \alpha_1)!} \quad (2)$$

which involves the number of equivalent ways of distributing $L \alpha_1$

particles among L equivalent sites, one to a site.

Invoking Stirling's approximation, valid for large n, namely $\ln n! \approx n \ln n - n$, we can simplify eqn. (2) to obtain

$$S_c \approx -k (\alpha_0 \ln \alpha_0 + \alpha_1 \ln \alpha_1) \quad (3)$$

where use was made of the normalizing condition

$$\alpha_0 + \alpha_1 = 1 \quad (4)$$

We may combine (1) and (3) to obtain an expression for the configurational force energy $F_c = E_c - TS_c$ and adjoin to this the thermal free energy, F_T , arising from the internal degrees of freedom of the atoms localized at the various lattice sites. Thus we obtain the total free energy as

$$F = F_T + L \{ (1-\alpha_1) \varepsilon_A + \alpha_1 \varepsilon_B + kT [(1-\alpha_1) \ln (1-\alpha_1) + \alpha_1 \ln \alpha_1] \} \quad (5)$$

From this one can determine the chemical potentials

$$\mu_s = (1/L) (\partial F / \partial \alpha_1)_{L,T} = \mu_T - \varepsilon_A + \varepsilon_B + kT [\ln \alpha_1 - \ln(1-\alpha_1)] \quad (6)$$

where μ_T is the chemical potential due to the internal degrees of freedom. At equilibrium, eqn. (6) may be equated with the chemical potential of the atoms in the gas phase

$$\mu_g = \mu_o + kT \ln p = \mu_s \quad (7)$$

Eqn. (7) provides an expression relating $\ln p$ to $\ln [\alpha_1 / (1-\alpha_1)]$. On solving for p one obtains

$$p = K \alpha_1 / (1-\alpha_1) \quad (8)$$

with

$$K \equiv \exp \{ [(\varepsilon_B - \varepsilon_A) + (\mu_T - \mu_o)] / kT \} \quad (9)$$

In the literature on gas absorption it is customary to denote the fraction of occupied surface sites by $\theta \equiv \alpha_1$. On solving eqn. (8) for this quantity one obtains

$$\theta = \frac{p/K}{1+p/K} = \frac{1}{1+K/p} \quad (10)$$

which is the well-known Langmuir isotherm equation.

III. THE FERMI-DIRAC DISTRIBUTION FUNCTION

As a second application of the above formulation we turn to the electron gas in a metal. There is, after all, no formal distinction between the statement that a given surface site is empty or filled by an atom and the statement that a given energy level is empty or occupied by an electron. Let us designate by $\epsilon \equiv \epsilon_B - \epsilon_A$ the energy of the L quantum states under consideration, and recall the fact that the Fermi level is the chemical potential of the electrons when there is no externally applied electric field. We can then immediately adapt eqn. (6) to the problem under discussion: μ now represents the Fermi level μ ; $\mu_T = 0$ since the electron has no internal degrees of freedom; $\epsilon_B - \epsilon_A \equiv \epsilon$, as stated earlier; finally, it is customary to set $\alpha_1 \equiv f$ in the context of the present problem. On solving eqn. (6), with the above notational changes, for f one obtains

$$f = \frac{1}{1 + e^{(\epsilon - \mu)/kT}} \quad (11)$$

which is the Fermi-Dirac distribution function.

IV. THE FIRST ORDER CASE; DECOMPOSITION OF THE LATTICE INTO REPRESENTATIVE SUBFIGURES.

We next take up the first order case, in which lateral interactions among atoms absorbed on nearest neighboring sites is taken into account. This requires that one consider the presence of paired units on the array. Accordingly, to simplify the problem at hand, we now represent the lattice by a collection of independent site pairs, termed 'bonds', and second collection of individual sites, termed 'points'.

For simplicity, let us consider first the two dimensional

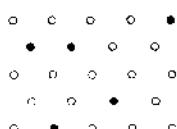


Fig. 1. Part of a regular, infinite, two-dimensional, hexagonal array of adsorption sites. Full circles indicate sites covered by an adsorbed atom.

triangular lattice shown in Fig. 1. The first task is to determine the number of bonds and points obtained by decomposition of the lattice containing L sites. Referring to Fig. 1, it is seen that each point is shared by six radial lines connecting it to the nearest neighbors, and that each line terminates in two points. Consequently, the lattice contains a total of $3L$ bonds which will be collected into an array known as a figure assembly. This immediately introduces a complication; for, in setting up the bond assembly one also automatically includes the $6L$ points contained in the $3L$ bonds. Now since the correct number of points in the lattice is only L , one must introduce another figure assembly which corrects for the overcounting of points. This is achieved by setting up a point array consisting of $-5L$ units. The fact that the second assembly contains a negative number of members need be of no great concern; we shall see later that this formalism leads to correct and self-consistent results, since the total number of points included in both the bond and point assemblies totals up correctly to L .

Next we shall consider the possible configuration or occupation states of members of the figure assemblies. It was mentioned earlier that each lattice site is either empty or filled. These states will be distinguished by assigning a site the configuration A or B. Correspondingly, the termini of the bonds or the points in the figure assemblies may be in the configurations A or B. One then encounters the particular occupation states, sketched in Table IV.1. To each configuration we assign a certain probability of occurrence, which is designated by the Greek symbol on the right. More precisely, a quantity such as β_0 indicates the fraction of the $3L$ bonds in the figure assemblies that is in the configuration AA. The α_j or β_j are termed distribution numbers.

The preceding concepts may now be generalized by considering any regular lattice in which a given site is surrounded by Z nearest neighbors. The number of members in the bond figure assembly is $(Z/2)L$ since each point has Z nearest neighbors and every bond involves two points. The number of units in the point figure assembly is $(1-Z)L$. Of the total members in the bond figure assembly, $\beta_0(Z/2)L$ are in the configuration AA, $2\beta_1(Z/2)L$ are in configuration AB, $\beta_2(Z/2)L$ are in the configuration BB, furthermore, of the members in the point figure assemblies, $\alpha_0(1-Z)L$ are in the occupation state A, and $\alpha_1(1-Z)L$ are in the occupation state B. This information is summarized in Table IV.1. One should note that the various distribution numbers β_j ($j = 0, 1, 2$) and α_j ($j = 0, 1$) are not independent, as is shown by the following equations:

$$\alpha_0 + \alpha_1 = 1 \quad (12)$$

TABLE IV.1
Configurations, Energies, and Probabilities for the Members of the Bond and Point Figure Assemblies

$$\beta_0 + 2\beta_1 + \beta_2 = 1 \quad (13)$$

$$\alpha_1 = \beta_1 + \beta_2 \quad (14)$$

Eqn. (12) and (13) represent normalization conditions; the sum over all the fractional parts of the system must be equivalent to the entire system. The factor 2 in conjunction with β_1 arises from the equivalence of the configurations AB and BA. Eqn. (14) is called a consistency relation; it arises from the fact that in the lattice proper there are $L\alpha_1$ occupied sites while, according to the tabulation of Table IV.1 the number of occupied sites is given by, $(ZL/2)(2\beta_1 + 2\beta_2) + (1-Z)L\alpha_1$. From the requirement that these two tallies must match we obtain eqn. (14).

The next step consists in finding the energy of the various figure assemblies. Let us associate with every unit in the configuration A or B in the energies ϵ_A or ϵ_B respectively, and let us associate with the bonds AA, AB, or BB the energies ϵ_{AA} , ϵ_{AB} and ϵ_{BB} respectively. Then, as is also evident from Table IV.1, the total configurational energy for the various members of the figure assemblies is given by the expression

$$E_c = L \left[\frac{Z}{2} (\beta_0 \epsilon_{AA} + 2\beta_1 \epsilon_{AB} + \beta_2 \epsilon_{BB}) + (1-Z) (\alpha_0 \epsilon_A + \alpha_1 \epsilon_B) \right] \quad (15)$$

That is, each energy contribution is multiplied by the number of units of that type occurring in the figure assemblies, and a summation over the various types is then carried out.

We now seek an expression for the configurational entropy of the system under consideration. This is found from the Boltzmann relation between entropy and the statistical state, W, of the system,

$$S_c = k \ln W \quad (16)$$

We shall consider the various members of the figure assemblies as mutually independent. Then, insertion of the standard combinatorial expression for W yields

$$W = \frac{\left(\frac{Z}{2}L\right)!}{(\beta_0 \frac{Z}{2}L)! (\beta_1 \frac{Z}{2}L)!^2 (\beta_2 \frac{Z}{2}L)!} \cdot \frac{[(1-Z)L]!}{[(1-Z)\alpha_0 L]! [(1-Z)\alpha_1 L]!} \quad (17)$$

Upon taking the logarithm of both sides and again introducing Stirling's approximation we have

$$\begin{aligned} \ln W &= \frac{Z}{2} L \ln \left(\frac{Z}{2} L \right) - \beta_0 \frac{Z}{2} L \ln (\beta_0 \frac{Z}{2} L) - 2\beta_1 \frac{Z}{2} L \ln (\beta_1 \frac{Z}{2} L) \\ &\quad - \beta_2 \frac{Z}{2} L \ln (\beta_2 \frac{Z}{2} L) + (1-Z) L \ln [(1-Z) L] \\ &\quad - (1-Z) \alpha_0 L \ln [(1-Z) \alpha_0 L] - (1-Z) \alpha_1 L \ln [(1-Z) \alpha_1 L] \end{aligned} \quad (18)$$

The nonlogarithmic terms cancelled each other because of the normalization requirements (12) and (13). If terms such as $\ln \{\beta_j(Z/2)L\}$ are written out as $\ln(Z/2)L + \ln \beta_j$, further cancellations occur, and (18) simplifies to

$$\begin{aligned} S_c &= -k L \left\{ \frac{Z}{2} (\beta_0 \ln \beta_0 + 2\beta_1 \ln \beta_1 + \beta_2 \ln \beta_2) \right. \\ &\quad \left. + (1-Z) (\alpha_0 \ln \alpha_0 + \alpha_1 \ln \alpha_1) \right\} \end{aligned} \quad (19)$$

At this stage we introduce the fundamental assumption that eqns. (15) and (19), which hold for independent figure assemblies, are also applicable to the lattice proper. This represents a fairly serious assumption, but it reduces the mathematical complexities of the subsequent treatment enormously. The ultimate justification for this step is that the results based on it are reasonable. Eqns. (15) and (19) will be combined with the corresponding terms accounting for the internal degrees of freedom to obtain the free energy expression; when properly minimized, this yields the mathematical conditions to be met by the system at equilibrium. Before carrying out the differentiation, it is necessary to decide on a choice of independent variables. Since there exist three conditions (12), (13) and (14), interconnecting the five unknowns α_0 , α_1 , β_0 , β_1 , and β_2 , only two of the distribution numbers can be considered as independent. In conformity with standard practice we shall select α_1 and β_2 as the independent quantities. From (12), (13) and (14) the following relations between the dependent and independent variables are then obtained:

$$\alpha_0 = 1 - \alpha_1 \quad (20)$$

$$\beta_1 = \alpha_1 - \beta_2 \quad (21)$$

$$\beta_0 = 1 - 2\alpha_1 + \beta_2 \quad (22)$$

We now write, as before, $F = F_T + E_c - TS_c$ with $F = F(\alpha_1, \beta_2)$. We can then determine the chemical potential by setting

$\mu_s = (\partial F / \partial \alpha_1)_{\beta_2, T, L}$, and we obtain an equilibrium condition from the requirement $(\partial F / \partial \beta_2)_{\alpha_1, T, L} = 0$. Thus,

$$\begin{aligned}\mu_s &= \frac{\partial (E_c - TS_c)}{\partial \alpha_1} + \frac{\partial (E_c - TS_c)}{\partial \alpha_o} \frac{\partial \alpha_o}{\partial \alpha_1} + \frac{\partial (E_c - TS_c)}{\partial \beta_o} \frac{\partial \beta_o}{\partial \alpha_1} + \\ &\quad \frac{\partial (E_c - TS_c)}{\partial \beta_1} \frac{\partial \beta_1}{\partial \alpha_1} + \frac{\partial F}{\partial \alpha_1} \quad (23)\end{aligned}$$

$$0 = \frac{\partial (E_c - TS_c)}{\partial \beta_2} + \frac{\partial (E_c - TS_c)}{\partial \beta_o} \frac{\partial \beta_o}{\partial \beta_2} + \frac{\partial (E_c - TS_c)}{\partial \beta_1} \frac{\partial \beta_1}{\partial \beta_2} \quad (24)$$

Now substitute from eqn. (15) and (19) into (23) and (24) and evaluate partial differentials such as $\partial \beta_o / \partial \alpha_1$ from (20), (21) and (22). The final result is

$$\begin{aligned}\mu_s &= \mu_T + (1-Z) \left[(\epsilon_B - \epsilon_A) + kT (\ln \alpha_1 - \ln \alpha_o) \right] \\ &\quad + Z \left[(\epsilon_{AB} - \epsilon_{AA}) + kT (\ln \beta_1 - \ln \beta_o) \right] \quad (25)\end{aligned}$$

Proceeding similarly with (24), this leads to the equation

$$0 = \frac{Z}{2} \left[\epsilon_{AA} + kT \ln \beta_o - 2(\epsilon_{AB} + kT \ln \beta_1) + \epsilon_{BB} + kT \ln \beta_2 \right] = 0 \quad (26)$$

We may rewrite eqn. (26) as the following, so-called quasi-chemical equation:

$$\frac{\beta_o \beta_2}{\beta_1^2} = e^{(-\epsilon_{BB} - \epsilon_{AA} + 2\epsilon_{AB})/kT} \equiv e^{-w/kT} \equiv C \quad (27)$$

$$\text{where } w \equiv \epsilon_{BB} + \epsilon_{AA} - 2\epsilon_{AB} \quad (28)$$

represents the lateral interaction among atoms absorbed on nearest neighbor sites. The reason for the terminology becomes clear upon regarding the quantity C in (27) as an "equilibrium constant" for the "reaction":



The physical justification for (27) is the following: Having decomposed the lattice into representative figure assemblies we

distributed species A and B among the sites in these collections. In the point figure assemblies there was no choice; the number of sites in the configurations A or B, $(1-Z)L\alpha_j$ ($j = 0,1$), is determined by the composition. In the bond figure assembly, however, the $ZL\alpha_j$ units of type A or B can be distributed among the bond termini in many different ways. One extreme is that in which as many bonds as possible contain termini of the same configuration (i.e., in which as many bonds as possible are in the configuration AA or BB); the other extreme occurs when as many bonds as possible are in the configuration AB. Any nontrivial rearrangement (which leads to changes in the numerical values of the β_j) must occur in accordance with the "chemical" equation (29), if the total number of units in the configurations A and B is to remain invariant. Thus, we could take (29) as our point of departure and immediately write out (27); however, statistical theory is then still required for the numerical evaluation of the constant C.

Eqn. (27) may now be solved for β_1 as a function of α_0 and β_1 , by substitution for β_0 and β_2 from (22) and (14) respectively. The final result is

$$\beta_1 = \{ -1 \pm \sqrt{1 + 4(C-1)\alpha_0\alpha_1} \} / 2(C-1) \quad (30)$$

It is seen from (27) that C is restricted to the range $(0, \infty)$; if β_1 is to remain in the range $(0,1)$ for all allowed values of α_0 , α_1 , and C, then it is necessary to adopt the positive sign in the above equation. Once β_1 has been determined, β_0 and β_2 are found from the relations

$$\beta_0 = \alpha_0 - \beta_1 \quad \beta_2 = \alpha_1 - \beta_1 \quad (31)$$

The case $C \rightarrow 1$ can be handled by expanding the square root in (30) to first order terms in $(C-1)$, and then taking the limit. This yields

$$\beta_1 = \alpha_0\alpha_1 ; \beta_0 = \alpha_0^2 ; \beta_2 = \alpha_1^2 \quad (32)$$

The above shows that the β_j are now comparable in magnitude; the probabilities β_0 , β_1 , β_2 are roughly equal whenever the interaction energies ϵ_{AA} , ϵ_{AB} , and ϵ_{BB} are comparable, rendering $C \approx 1$. Under these conditions a purely random distribution of occupation states A and B among sites is possible.

V. THE FOWLER-CUGGENHEIM ADSORPTION ISOTHERM

In our first application of the preceding development we turn again to the case of gas adsorption. We note that there exist five relations involving α_0 , α_1 , β_0 , β_1 , and β_2 , namely

eqns. (25), (26), (12), (13) and (14). Moreover, we have the equilibrium condition that the chemical potential specified by (25) must be equal to the chemical potential of the gas phase $\mu_g = \mu_0 + kT \ln p$. One can therefore first determine μ_s in (25) solely as a function of $\alpha_1 \equiv \theta$ and then from the relation $\mu_s = \mu_g$ relate this function of θ to $\ln p$. The execution of these steps is a tedious though straightforward exercise in algebra which ultimately leads to the result

$$p = K \frac{\theta}{1-\theta} \left(\frac{B-1+2\theta}{2\theta} \right)^Z e^{zw/kT} \quad (33)$$

where

$$\begin{aligned} B &\equiv \sqrt{1+4\theta(1-\theta)(C-1)} \\ C &\equiv e^{-w/kT} \\ w &\equiv \epsilon_{BB} + \epsilon_{AA} - 2\epsilon_{AB} \end{aligned} \quad (34)$$

The above is the well known Fowler-Guggenheim adsorption isotherm equation.

The reader should note that θ is specified implicitly through eqns. (33) and (34) in terms of the variable p , and the parameters T , w , and Z . We have thus been able to take into account explicitly the effect of lateral interactions among nearest neighbors adsorbed on adjacent surface sites.

VI. THERMODYNAMIC PROPERTIES OF BINARY MIXTURES IN THE BRAGG-WILLIAMS APPROXIMATION

We next adapt the results of section IV to binary mixtures. The model here involves a three-dimensional lattice array in which any lattice site is occupied either by species A or by species B. In taking over the earlier formation we simply assign Z a numerical value consistent with a three-dimensional array, and we let ϵ_{AA} , ϵ_{BB} , and ϵ_{AB} represent the interaction energies among among units situated on the termini of bonds in the corresponding configurations.

Since the first order approximation is still rather complex we shall ultimately utilize a simpler variant, based on eqn. (32) rather than on the exact solutions derived from (30) and (31). We begin with eqns. (15) and (19), and substitute $\alpha_1 - \beta_1$ for β_2 and $\alpha_0 - \beta_1$ for β_0 whenever these quantities occur in (15) and as coefficient of the logarithmic terms in (19); we then

eliminate the arguments of the logarithmic terms utilizing eqn.(26). This yields

$$\begin{aligned} F = & F_T + (Z/2) L \{ \alpha_0 \epsilon_{AA} + \alpha_1 \epsilon_{BB} \} + (1-Z) L \{ \alpha_0 \epsilon_A + \alpha_1 \epsilon_B \} \\ & + kT (Z/2) L \{ \alpha_0 \ln \beta_0 + \alpha_1 \ln \beta_2 \} + (1-Z) L \{ \alpha_0 \ln \alpha_0 + \alpha_1 \ln \alpha_1 \} \end{aligned} \quad (35)$$

When eqn. (30) is expanded to first order terms in (C-1) one finds

$$\beta_1 = \alpha_0 \alpha_1 - \alpha_0^2 \alpha_1^2 (C-1) \quad (36)$$

Then, $\beta_0 = \alpha_0 - \beta_1 = \alpha_0^2 [1 + \alpha_1^2 (C-1)] \quad (37)$

$$\beta_2 = \alpha_1 - \beta_1 = \alpha_1^2 [1 + \alpha_0^2 (C-1)]$$

and to first order in (C-1) we have

$$\ln \beta_0 = 2 \ln \alpha_0 + \alpha_1^2 (C-1)$$

$$\ln \beta_2 = 2 \ln \alpha_1 + \alpha_0^2 (C-1) \quad (38)$$

Introducing (38) in (36) and setting (C-1) $\alpha_0 \alpha_1 \approx -\alpha_0 \alpha_1 w/kT$ for small w/kT we finally obtain

$$\begin{aligned} F = & F_T + (Z/2) L \{ \alpha_0 \epsilon_{AA} + \alpha_1 \epsilon_{BB} \} + (1-Z) L \{ \alpha_0 \epsilon_A + \alpha_1 \epsilon_B \} \\ & + kT L \{ \alpha_0 \ln \alpha_0 + \alpha_1 \ln \alpha_1 - \alpha_0 \alpha_1 Z w / 2kT \} \end{aligned} \quad (39)$$

Relations such as (39), based on equations (36) - (38), are said to be specified in the Bragg-Williams approximation. For further reference it is important to set $\alpha_0 L = N_A$, $\alpha_1 L = N_B$, $\alpha_0 = N_A / (N_A + N_B)$, and $\alpha_1 = N_B / (N_A + N_B)$; further, it is expedient to write $X_A \equiv (Z/2)(\epsilon_{AA} - 2\epsilon_B)$, $X_B \equiv (Z/2)(\epsilon_{BB} - 2\epsilon_B)$. Thus, finally

$$\begin{aligned} F = & F_T + N_A \epsilon_A + N_B \epsilon_B + N_A X_A + N_B X_B \\ & + kT \left\{ N_A \ln \left(\frac{N_A}{N_A + N_B} \right) + N_B \ln \left(\frac{N_B}{N_A + N_B} \right) \right\} - \frac{N_A N_B}{N_A + N_B} \frac{Zw}{2} \end{aligned} \quad (40)$$

We can now compute the chemical potential. Note that in contrast to section V we do not keep L constant, because we add species A

to the mixture of A + B while keeping B fixed, thus altering the total number of units, L, in the mixture. By contrast, in dealing with the case of gas absorption, we converted an empty into a filled site, keeping L fixed. For this reason it was necessary to eliminate L in favor of N_A and N_B when eqn. (40) was set up. Accordingly, the chemical potential is now given by

$$\begin{aligned} \mu_A &= \left(\frac{\partial F}{\partial N_A} \right)_{T, N_B} = \mu_T^0 + \epsilon_A + X_A + kT \ln \left(\frac{N_A}{N_A + N_B} \right) \\ &\quad - \left(\frac{N_B}{N_A + N_B} \right)^2 \frac{Zw}{2} \equiv \mu_A^0 + kT \ln \alpha_0 - \alpha_1^2 Zw/2 \quad (41) \end{aligned}$$

We may compare the above to the thermodynamic expression for the chemical potential of a binary mixture, namely $\mu_A = \mu_A^0 + kT \ln X_A + RT \ln \gamma_A$ where X_A is the mole fraction of species A and γ_A is the activity coefficient. Comparison with (41) shows that in the Bragg-Williams approximation

$$\ln \gamma_A = - (1 - \alpha_0)^2 Zw/kT \quad (42)$$

Thus we see that order-disorder theory has enabled us to specify the activity coefficient in terms of the parameter Zw/kT . Insofar as a liquid may be regarded as a loosely coupled solid, with interactions among neighboring atoms conforming to the Bragg-Williams approximation, we have thus evolved a scheme for determining activity coefficients in binary solutions. We have further identified, in the present approximation, the standard chemical potential as $\mu_T^0 + \epsilon_A + X_A$.

VII. ORDER-DISORDER THEORY

Our final application of the formalism deals with a 50-50 binary alloy such as the Zn-Cu brass, where there are two extreme possibilities. At one extreme all Zn atoms settle on one sublattice and all Cu atoms on the second sublattice of the two simple cubic inter-penetrating structures making up the crystal structure of brass, the alloy is then said to be completely ordered. At the other extreme, the Cu and Zn atoms are located at random on the two sublattices; the crystal is then said to be completely disordered.

Again, there is no fundamental distinction between 'filled' or 'empty' sites in gas adsorption theory and sites occupied by the 'right' or 'wrong' atom in a 50-50 binary alloy. Accordingly, we introduce here the quantity

$$s \equiv \alpha_1 - \alpha_0 = 2\alpha_1 - 1 = 1 - 2\alpha_0 \quad (43)$$

as a measure of the degree of order; α_1 now represents the fraction of all sites occupied by the 'right' type of atom.

Since we are interested only in that part of the free energy expression dealing with the mixing process we abstract from eqn. (39) the quantity,

$$F_m = kT L (\alpha_0 \ln \alpha_1 + \alpha_1 \ln \alpha_1) - L \alpha_0 \alpha_1 Zw/2 \quad (44)$$

and replace α_0 and α_1 in terms of s . This yields

$$\begin{aligned} F_m &= kT L \left\{ \left(\frac{1-s}{2} \right) \ln \left(\frac{1-s}{2} \right) + \left(\frac{1-s}{2} \right) \ln \left(\frac{1+s}{2} \right) \right\} \\ &\quad - \frac{L(1-s^2)Zw}{8} \end{aligned} \quad (45)$$

Equilibrium conditions are characterized by the requirement $(\partial F / \partial s)_{T,L} = 0$. When the appropriate manipulations are carried out one obtains an expression for the s^* prevailing under equilibrium conditions, namely

$$\ln \left(\frac{1+s^*}{1-s^*} \right) = -s^* \frac{Zw}{2kT} \equiv 2s^* \frac{T_c}{T} \quad (46)$$

where we have set $T_c \equiv (-w) Z/4k$. A plot of s^* vs T/T_c is shown in Fig. 2. As expected, when $T \rightarrow 0$, the system is completely ordered; $s^* \rightarrow 1$. When $T \rightarrow T_c$ the system becomes completely disordered, with $s^* \equiv \alpha_1 - \alpha_0 = 0$.

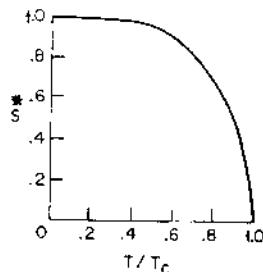


Fig. 2. The degree of order s as a function of the ratio T/T_c in the range where order-disorder transitions may occur; s is a root of equation (46).

One experimental technique for detecting the order-disorder transition in the range $0 \leq T/T_c \leq 1$, is to measure the heat capacity. It should be clear from eqn. (45) that the mixing energy is given by

$$E_m(s) = -L(1-s^2) Z_w/8 \quad (47)$$

From Fig. 2 one obtains s^* as a function of T/T_c , so that $E_m(T/T_c)$ is now known. A sketch of E_m vs T/T_c is shown in Fig. 3(a). Finally, one may determine $C_m = \partial E_m / \partial T$ as the heat capacity for the mixing process. This quantity is sketched in Fig. 3(b). It is seen that this duplicates the well-known Λ -type anomaly in the heat capacity which is observed in second-order phase changes.

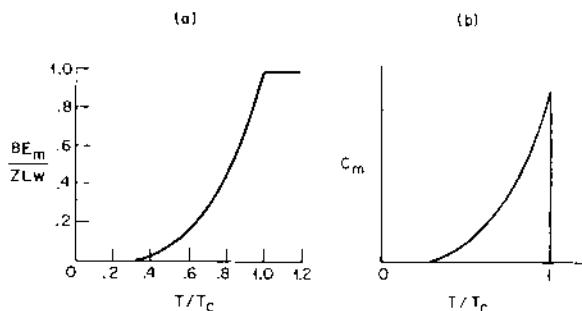


Fig. 3. (a) The variation of E_m with T/T_c computed according to eqn. (47), where s is the root of eqn. (46).
 (b) Schematic diagram of the configurational heat capacity of mixing vs T/T_c , obtained from Fig. 3(a).

VIII. SUMMARIZING REMARKS

We have seen how a single formalism originally specified in this form by Hijmans and de Boer² can be applied consistently to describe a very large variety of physical phenomena, namely all those where a given unit can be in either of two mutually exclusive states. Further applications of the formalism have been discussed in the literature where cases such as ferromagnetism, antiferromagnetism¹, and kinetics of diffusionless phase transformations³⁻⁵ have been treated.

REFERENCES

1. J.M. Honig, J. Chem. Ed., 38, 538 (1961).

2. J. Hijmans and J. de Boer, *Physica*, 21, 471, 485, 499 (1955).
3. J.M. Honig in "Kinetics of High Temperature Processes", Ed. by W.D. Kingery, John Wiley and Sons Inc., New York, 1959.
4. J.M. Honig, *J. Chem. Phys.*, 28, 723 (1958).
5. C.N.R. Rao, S.R. Yoganarasimhan and P.A. Faeth, *Trans. Faraday Soc.*, 57, 504 (1961).

DIFFUSION IN SOLIDS

V. Ramakrishna

Department of Chemistry

Indian Institute of Technology, New Delhi-29, India

During the last three to four decades two factors have spurred the interest of scientists enormously in this subject: one is the availability of radioactive isotopes for diffusion studies; and the second, the greater preoccupation of metallurgists with the scientific foundations of what has been till recently a semi-empirical activity. Quantitative measurements of diffusion have been carried out extensively and excellent monographs and reviews of this work are available (1-12). These quantitative data are usually expressed in terms of a "diffusion coefficient". Unfortunately, however, in the earlier experiments conditions of measurements have not been defined scrupulously. Hence it is not an exaggeration to state that the data reported until about 1950 or so are of doubtful dependability. In the last two decades, our understanding of the diffusion phenomena has improved considerably. At the same time considerable advances have been made in the experimental techniques using a variety of radio-active isotopes of high specific activity. Also, some novel methods of diffusion measurements have been introduced. In the present article, a discussion about the nature of diffusion coefficient D will first be presented, followed by a brief outline of experimental methods.

I. THE DIFFUSION COEFFICIENT

Confining our attention to one dimensional transfer of matter, the diffusion coefficient, D, is defined by the following equation, which is Fick's first law:

$$J_i = - D_i \frac{\partial n_i}{\partial x} \quad (1)$$

where J_i = the flux of the component i , n_i = concentration in any convenient units e.g., g-moles or number of atoms/volume, and x the space coordinate. We will consider the diffusion phenomena in a two component system only, which is the simplest system of practical importance. The diffusion of isotopes may be regarded as a special case of a two-component system.

The diffusion equation (1) can be generalised, using the vector notation for a three-dimensional space; thus

$$\vec{J}_i = - \vec{D}_i \text{ grad } n_i. \quad (2)$$

Since the total number of atoms does not change due to diffusion, the continuity condition may be applied:

$$\text{Div } \vec{J}_i + \frac{\partial n_i}{\partial t} = 0. \quad (3)$$

Combining eqns. (2) and (3) we obtain the following relationship, usually known as the Fick's second law:

$$\frac{\partial n_i}{\partial t} = \text{div } \vec{D}_i \text{ grad } n_i \quad (4)$$

or for one dimensional diffusion

$$\frac{\partial n_i}{\partial t} = \frac{\partial}{\partial x} (D_i \frac{\partial n_i}{\partial x}). \quad (5)$$

As stated earlier, the application of eqn. (1) to actual experimental data requires a clear definition of the frame of reference used in the measurements. Consider the interdiffusion of two solids A and B such that there is no change of volume upon mixing. Two diffusion coefficients, one for each substance, can be defined according to eqn. (1):

$$J_A = - D_A^V \frac{\partial n_A}{\partial x} \quad (6)$$

$$J_B = - D_B^V \frac{\partial n_B}{\partial x} \quad (7)$$

The fluxes in the above equation are measured across a section which is defined by the condition that the total volume on either side of it remains constant as the diffusion proceeds. Let V_A and V_B be the molar volumes of A and B respectively, for the situation when n_i 's in eqns. (6) and (7) are moles/volume. We will further assume that $V_A \approx V_A$ and $V_B \approx V_B$. The volume transfer of A per unit time

across unit area of the section defined earlier is given by

$$-D_A^V V_A \frac{\partial n_A}{\partial X}$$

and that of B

$$-D_B^V V_B \frac{\partial n_B}{\partial X} .$$

According to the condition of constant volume, upon mixing therefore

$$D_A^V V_A \frac{\partial n_A}{\partial X} + D_B^V V_B \frac{\partial n_B}{\partial X} = 0 \quad (8)$$

$$\text{Also, } n_A V_A + n_B V_B = \text{Constant} \quad (9)$$

$$\text{i.e., } V_A \frac{\partial n_A}{\partial X} + V_B \frac{\partial n_B}{\partial X} = 0 \quad (10)$$

Comparing equations (8) and (10) it follows that

$$D_A^V \equiv D_B^V \quad (11)$$

$$\text{or else that } V_A = 0 \text{ or } V_B = 0 \quad (12)$$

In either case, the behaviour of a two component system satisfying the condition of no volume change upon mixing, can be described by a single diffusion coefficient. It is referred to as the "mutual" diffusion coefficient, D^V , the superscript-V denoting the condition of constant volume. This diffusion coefficient is quite familiar in the inter-diffusion of gases.

When the system chosen or measurements made do not conform to the condition of constant volume, alternative frames of reference must be defined to make the use of diffusion coefficient meaningful. There are several possible alternatives. The most obvious is based on the fact that the total mass of the system is always conserved, even if the volume does not remain constant. A section can therefore be defined such that the mass of the system on either side of the section remains constant during diffusion. When a convention other than that of constant volume on either side is used in defining a section, the second order differential equation describing diffusion may not take the standard form of eqn. (5). Crank (6) considers it better to maintain the standard form by departing from the orthodox linear scale, e.g. cm, for the spatial coordinate X, and by measuring concentration in a slightly different way. Let

the modified scale for length be ξ and consider two sections at ξ and $\xi + d\xi$. The rate of accumulation of species A in the volume enclosed between these sections is, on the basis of the same convention as before,

$$-(\partial J_A / \partial \xi) d\xi$$

and this is always true independent of how ξ and J_A are measured. This can only be equated to $(\partial n_A / \partial t) d\xi$ when n_A and ξ are measured in consistent units. Thus, for example, if the sections are fixed with respect to total mass, then ξ must be measured so that equal increments of ξ always include equal increment of total mass, and n_A must be defined as the amount of A per unit total mass. In general, for all values of ξ and t the element of unit length in terms of ξ , and of unit cross sectional area, is that which contains an amount of A equal to the unit used in defining the concentration n_A . When the quantities n_A and ξ satisfy this condition, the usual relationship

$$\frac{\partial n_A}{\partial t} d\xi = - \frac{\partial J_A}{\partial \xi} d\xi \quad (13)$$

holds good; and by using

$$J_A = - D_A \frac{\partial n_A}{\partial \xi} \quad (14)$$

the second law of Fick takes the standard form,

$$\frac{\partial n_A}{\partial t} = - \frac{\partial}{\partial \xi} (D_A \frac{\partial n_A}{\partial \xi}) \quad (15)$$

It is convenient that ξ should have the dimension of length and D the usual dimension of cm^2/sec . This can be done by multiplying the total mass by an arbitrary constant, specific volume. The volume represented by the product of the total mass and the arbitrary constant specific volume is referred to by Crank⁶ as "basic volume" of that total mass. The use of this particular basic volume has the convenience that all the concentrations measured in the different frames of reference tend to the same value in dilute solutions. If the arbitrary specific volume is denoted by V^o , then

$$d\xi_M = V^o (n_A^V + n_B^V) dx \quad (16)$$

which relates ξ to x formally. Here the subscript M refers to the fact that the unit ξ contains per unit area, the unit basic volume of A and B together.

Referring back to the system where volume changes occur on diffusion, which is now defined with respect to a mass-fixed section it can be shown that across this section

$$D_A^M \frac{\partial n_A}{\partial \xi_M} + D_B^M \frac{\partial n_B}{\partial \xi_M} = 0 \quad (17)$$

and that

$$n_A^M + n_B^M = 1/V^0 \quad (18)$$

Differentiating eqn.(18) and comparing with eqn.(17) leads to the condition

$$D_A^M \equiv D_B^M \quad (19)$$

The superscript M denotes constancy of mass. Thus the statement that the diffusion behaviour of a two component system can be described in terms of a single diffusion coefficient is still valid, even when volume changes occur, provided that an appropriate frame of reference is used in defining diffusion coefficient. The frame of reference could be so chosen that the two coefficients are not identical and neither is zero, but they would be related through some function of the partial volumes and would not be independent measures of two separate diffusion processes. The possibility of measuring the diffusion coefficient of two species independently will be discussed later.

In some cases when volume changes occur, it may be convenient to use a frame of reference with fixed volume with respect to one component, say component B. With this frame of reference, the diffusion for component B is necessarily zero, while that of the other component is denoted by D_A^B . It is still true that only one diffusion coefficient is required to describe the process of diffusion.

Applying simple considerations, Crank⁶ has obtained relationships between the various diffusion coefficients of two component systems. They are

$$\begin{aligned} D^M &= D^V (n_B^V/n_B^M)^2 \\ &= D^V \left(\frac{\text{basic total volume}}{\text{true total volume}} \right)^2 \end{aligned} \quad (20)$$

$$D_A^B = D^M (V_B^0 n_B^M)^2 = D^V (V_B n_B^V)^2 = D^V \phi_B^2 \quad (21)$$

where ϕ_B is the volume fraction of B.

II. INTRINSIC DIFFUSION COEFFICIENTS

While it is true that a system of two components can be described

by a single of "mutual" diffusion coefficient, it is now well established that the values so obtained are rendered ambiguous owing to the mass-flow of the solution accompanying diffusion. This effect is striking when the mass and size of the molecules A and B, are greatly different. In such a case, the random motions of molecules A result in its transfer being greater or less than that of B. Consequently a "hydrostatic" pressure builds up in the region of the solution which contributes least to the volume rate of transfer. The obvious result is that there is a compensating mass-flow to relieve the pressure. The existence of this mass-flow has been demonstrated in the case of gases diffusing across a rigid porous plate. In the case of solids, this phenomenon has been a more recent discovery. It has been demonstrated by placing inert markers at an initial interface between two specimens of differing chemical composition and observing their apparent movement with respect to the fixed ends of the diffusion couple as the diffusion proceeds. Although first developed by Hartley¹³, it is now known as the Kirkendall effect¹⁴ and illustrated in Fig. 1

In systems where the Kirkendall movement takes place, the use of a single mutual diffusion coefficient, D^V , appears to be unrealistic

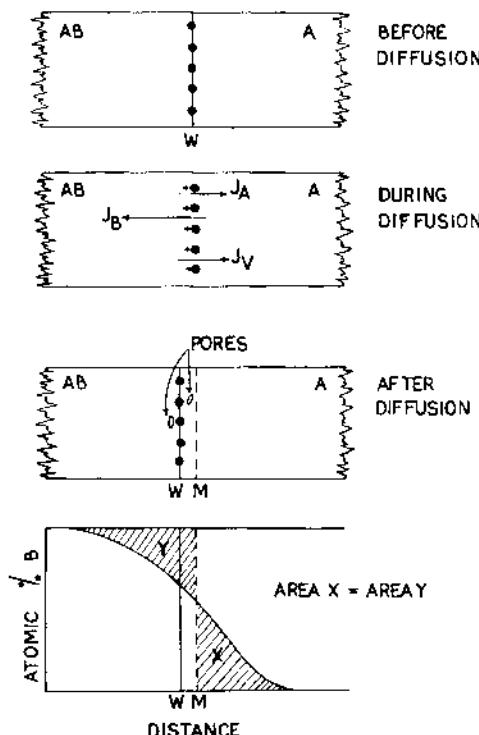


Fig. 1. Schematic Diffusion Process in an Infinite Diffusion Couple with an Initial Concentration Gradient: W, Markers; M, Matano surface.

in view of the presence of mass-flow. It is desirable to define two diffusion coefficients, D_A and D_B , in terms of the rate of transfer of A and B, respectively, across a section such that no mass-flow occurs through it. Evidently, this section has to be located on the markers and it must follow the mass-flow, even though the latter may not always be observable. It will be recognised that this frame of reference is completely different from all the others discussed previously. Hartley and Crank¹⁵ term these diffusion coefficients as "intrinsic" diffusion coefficients. When the partial volumes of the components remain constant, these intrinsic diffusion coefficients can be shown to be related to the mutual diffusion coefficients in a simple way.⁶

$$D^V = V_A n_A^V (D_B - D_A) + D_A \quad (22)$$

If the partial volumes are not constant, then D^V has no significance; in such cases the intrinsic diffusion coefficients (D_i) can be related to D_A^B . Thus

$$D_A^B = V_B n_B^V (V_A n_A^V)^2 \left\{ D_A + D_B \frac{\frac{V}{n_A} \frac{V}{n_A} + \frac{n_A}{V} \frac{\partial V_A / \partial n_A}{V}}{\frac{V}{n_B} \frac{V}{n_B} + \frac{n_B}{V} \frac{\partial V_B / \partial n_B}{V}} \right\} \quad (23)$$

Eqn. (23) can be shown to reduce to eqn. (22) when V_A and V_B are constant.

It will be observed from eqns. (22) and (23) that the values of D_A and D_B can not be obtained separately unless some information additional to D^V or D_A^B is available. This additional information can be obtained by studying the rate of mass-flow, namely, the movement of the markers. It can be shown that this rate (u) relative to the Matano surface (a term which is explained later) is,

$$u = (D_A - D_B) \frac{\partial n_A}{\partial X} \quad (24)$$

It has been observed empirically that u varies as the square root of time, whence the Kirkendall shift X_M , relative to the Matano surface can be shown to be

$$X_M = 2 ut = 2t (D_A - D_B) \frac{\partial n_A}{\partial X} \quad (25)$$

It is thus possible to evaluate the values of D_A and D_B separately.

Another possibility which occurs in the case of solvent-polymer systems is that the intrinsic diffusion for one component, *viz.* the polymer, is so much smaller compared to the other that it can be approximated to zero. If then $D_B \approx 0$ eqn.(22) yields,

$$D_A = \frac{D^V}{1 - V_A n_A^V} = \frac{D^V}{V_B n_B^V} = \frac{D_V}{\phi_B} = \frac{D_A^B}{\phi_B^3} \quad (26)$$

where ϕ_B = vol. fraction of B. It can be shown further that

$$D^V = D_A N_B + D_B N_A \quad (27)$$

where N_A and N_B are the mole or atom fractions of the components. These ideas have been discussed at considerable length by Crank⁶, Bardeen⁷, Lazarus⁸ and Darken¹⁶.

III. SELF-DIFFUSION COEFFICIENTS

Lastly, there remains to consider diffusion phenomena under conditions when there is no chemical concentration gradient. Consider a two-component system, in which a few of the atoms/molecules of one component A are labelled radioactively. The process of diffusion of the labelled A* atoms/molecules which are identical in all other respects to the unlabelled A, will not involve any change of composition and hence there will be no mass-flow. The diffusion coefficient measured under these conditions is known as the "self"-diffusion coefficient D_A^* and is the true mobility of the labelled A* with respect to the stationary solution. It has been observed, however, that the self diffusion coefficient is not equal to the corresponding intrinsic diffusion coefficient at the same composition. The relationship between the two has been obtained by Darken¹⁶ and Prager¹⁷ based on the thermodynamic properties of the system.

Consider a two-component system comprising molecules A and B and let the gradient of concentration n_A of A be maintained in equilibrium by applying a force F_A per g. mole of A in the direction of increasing X. The generalised form of the condition for this thermodynamic equilibrium is

$$F_A = \frac{\partial \mu_A}{\partial X} \quad (28)$$

where μ_A = chemical potential of A. The rate of transfer of A due to force F_A is

$$\frac{F_A n_A}{\rho} = \frac{n_A}{\rho} \frac{\partial \mu_A}{\partial X} \quad (29)$$

where ρ = resistance coefficient. At equilibrium, this must equal the rate of transfer by diffusion; and so we have,

$$D_A \frac{\partial n_A}{\partial X} = \frac{n_A}{\rho} \frac{\partial \mu_A}{\partial n_A} \frac{\partial n_A}{\partial X} \quad (30)$$

Hence $D_A = \frac{n_A}{\rho} \frac{\partial \mu_A}{\partial n_A}$ (31)

Applying the same criteria to the labelled atoms/molecules,

$$F_A^* = \frac{\partial \mu_A^*}{\partial X} = \frac{RT}{n_A^*} \frac{\partial n_A^*}{\partial X} \quad (32)$$

because of the ideality of the system with respect to the labelled species. Instead of eqn. (31), therefore, we have assuming that

$$D_A^* = \frac{RT}{\bar{c}} \quad (33)$$

the same resistance coefficient holds for the two species at the same chemical composition. Thus,

$$D_A = \frac{D_A^* n_A}{RT} = \frac{\partial \mu_A}{\partial n_A} \quad (34)$$

$$= D_A^* \left(1 + \frac{\partial \ln \gamma_A}{\partial \ln n_A} \right) \quad (35)$$

IV. EXPERIMENTAL PROCEDURES

Having discussed the nature of the term D , it is worthwhile to consider briefly the principles of several experimental procedures employed for its determination. There are two most general and direct methods for the measurement of diffusion coefficients.

In the first method, which may be called the permeation method, the flux J_i due to diffusion across a membrane of the solid sample of a definite geometry (plane sheet, hollow cylinder or sphere) is determined under conditions where the surface concentrations, $n_i(s)$ are well defined. Under steady conditions, for constant D_i eqn. (1) is solved as follows:

$$\text{For a plane membrane: } n_{ix} = \frac{n_{i1} - n_{i2}}{x_1 - x_2} (x - x_2) + n_{i1} \quad (36)$$

$$J_i = D_i (n_1 - n_2)/\delta$$

$$\text{where } \delta = (x_2 - x_1) \quad (37)$$

$$\text{For a hollow cylinder: } n_{ir} = \frac{n_{i1} \ln(r_1/r) + n_{i2} \ln(r/r_2)}{\ln(r_2/r_1)} \quad (38)$$

Flux per unit length of cylinder:

$$J_i = \frac{2\pi D_i (n_2 - n_1)}{\ln(r_2/r_1)} \quad (39)$$

For a hollow sphere :

$$n_{ir} = \frac{n_{i1}r_1(r_2-r) + n_{i2}r_2(r-r_1)}{r(r_2-r_1)} \quad (40)$$

$$J_i = 4\pi D_i \frac{r_1 r_2}{r_2 - r_1} (n_{i2} - n_{i1}) \quad (41)$$

From eqns. (37), (38) and (39) it is clear that a measurement of J_i , with known n_{i1} , n_{i2} and δ or r_1 and r_2 is sufficient for the calculation of D_i . In systems where gases or vapours diffuse through a membrane, it is customary to introduce pressure terms in place of the surface concentrations, $n_i(s)$. Further it is possible to relate the equilibrium pressure of gases or vapours and the corresponding surface concentrations by suitable adsorption or solubility measurements. For example, in the linear portion of the adsorption isotherms of vapours on polymers, $n_i = sp_i$; or in the case of solubility of permanent gases in metals at high temperatures $n_i = s(p_i)^{1/2}$, according to Sieverts Law. The proportionality constant s is the solubility (in appropriate units) of the diffusant in the solid. Whence the flux in such cases for plane membranes can be written as

$$J_i = D_i s \left[\frac{(p_1 - p_2)}{\delta} \right] \quad (42)$$

or

$$J_i = D_i s \left[\frac{(p_1)^{1/2} - (p_2)^{1/2}}{\delta} \right] \quad (43)$$

The flux per unit area per unit value of the terms in the square brackets in (42) and (43) is called the permeability coefficient, P . Similar considerations may be extended to diffusion from solution, the term 's' having the significance of the partition ratio between the membrane material and the solution.

Unless the concentration gradients are extremely small, it is rare to find D constant, either in metal or polymer systems. For example, in the now well known experiment of Smith¹⁸ for the diffusion of carbon in γ -iron in the form of a hollow cylinder at 1000°C, the concentration profile of C through the cylinder wall was determined after a steady state had been attained. It was found that n_c was not a linear function of $\log r$ as required by eqn.(38)

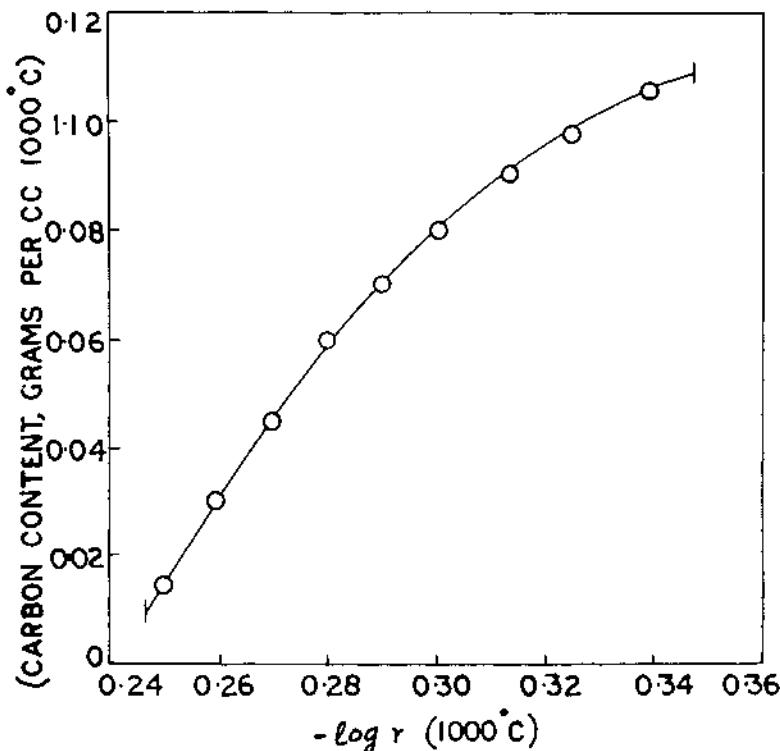


Fig. 2. Concentration-log Y plot for the diffusion of carbon through a hollow iron cylinder at 1000°C (ref: 18)

in which D was held constant. This is shown in Fig. 2. Even if D is a function of composition, eqn. (1) namely,

$$J_i = - D_i \frac{\partial n_i}{\partial X} = \text{constant}$$

holds good under steady state conditions. However, what is measured is an average value of D_i . Because, upon integration

$$J_i = - \frac{1}{\delta} \int D_i \frac{dn_i}{dx} = \frac{\bar{D}_i (n_1 - n_2)}{\delta} \quad (44)$$

where

$$\bar{D}_i = \frac{1}{n_1 - n_2} \int D_i \frac{dn_i}{dx} \quad (45)$$

If the concentration profile through the thickness of the membrane as well as the flux are determined in a single experiment under

steady state conditions, it is possible to evaluate the concentration dependence of D^{19} .

A variation of the permeation method, especially preferred for the diffusion of gases in high polymers, is to measure the quantity of diffusant, (Q_t), diffusing in time t across a membrane of the solid material when steady state has been attained. The graph of Q_t against t , gives an intercept L on the t -axis, which is called the "time-lag"¹². It can be shown that L has the following values depending on the geometry of the membrane, whence D can be calculated:

$$\text{For plane sheet : } L = \delta^2/6D \quad (46)$$

$$\text{For hollow cylinder : } L = \frac{r_2^2 - r_1^2 + (r_2^2 + r_1^2)}{4 D} \frac{\ln(r_1/r_2)}{\ln(r_1/r_2)} \quad (47)$$

$$\text{For a sphere : } L = (r_2 - r_1)^2/6D \quad (48)$$

When D is a function of concentration, the solution of the differential equations for the time-lag method are more complicated. Frisch²⁰ has discussed such situations.

Several precautions have to be taken in the preparation of membranes for the permeation experiments. The influence of inhomogeneities of structure and chemical composition of the membrane and grain size distribution in the case of metals, end and surface diffusion effects, and experimental cell assemblies have been discussed at considerable length^{8,12}. For polymer membranes, Rosen and Singleton²¹ have described an automatic recording apparatus, while Huldy²² has developed multiple cell assemblies.

The other direct method of determining diffusion coefficients is to analyse the concentration-distance curves, also known as the penetration curves. For this purpose, two effectively infinite media are brought together at $t = 0$, placed in a thermostat or furnace at constant temperature, and diffusion is allowed to take place. After a duration of time, the diffusion couples are taken out and the concentration profile in the diffusion zone is analysed. When the conditions of the experiment conform to the boundary conditions:

$$n_i = n_i(s), \quad X < 0, \quad t = 0$$

$$n_i = 0, \quad X > 0, \quad t = 0$$

$$n_i = n_i(s), \quad X = 0, \quad t > 0 \quad (49)$$

the partial differential eqn. (5) has a fairly simple solution, assuming D is constant and no change takes place in volume upon mixing. Thus,

$$n_i = \frac{1}{2} n_{i0} \operatorname{erfc} \frac{x}{2(D_i t)^{1/2}} \quad (50)$$

D_i can be computed readily by comparing the experimental penetration curve with the theoretical given by eqn. (50). Tables and graphs for the error function complement, erfc , are available^{16a,23}. Similarly, solutions of Fick's eqn. (5) for unsteady state diffusion into solid objects of definite geometric shape (like sphere, finite and infinite cylinder, the infinite slab, the infinite rectangular rod, etc.) are available in terms of $D_i t / l^2$, where l is a characteristic length of the object^{16a}.

When D is not constant, but a function of concentration, Fick's diffusion eqn. (5) cannot be solved even for a two-dimensional flow. However, by using the Boltzmann transformation, $\lambda = x/2t^{1/2}$, the original differential equation is converted to an ordinary differential equation. Thus

$$-2\lambda \frac{dn_i}{d\lambda} = \frac{d}{d\lambda} (D_i \frac{dn_i}{d\lambda}) \quad (51)$$

or $d(D_i \frac{dn_i}{d\lambda}) = -2\lambda dn_i$

Integrating from $n_i = n_{i0}$ (when $dn_i/d\lambda = 0$) to $n_i = n'_i$

$$D_i \frac{dn_i}{d\lambda} = - \int_{n_{i0}}^{n'_i} \lambda dn_i \quad (52)$$

By rearranging and introducing X and t again,

$$D_i n_i = n'_i = - \frac{1}{2t} \frac{dX}{dn_i} \int_{n_{i0}}^{n_i} X dn_i \quad (53)$$

Since $(D_i \frac{dn_i}{d\lambda}) n_i = n_{i\infty} = 0$ also, it follows from 52 that

$$\int_0^\infty X dn_i = \int_0^\infty \lambda dn_i = 0 \quad (54)$$

Eqn. 54 signifies that the boundary conditions will be satisfied completely when plane 'X' in the diffusion couple is chosen to make this equation applicable. Such a plane 'X' is called the Matano surface²⁴ and is shown in Fig. 1. In several metal-alloy diffusion couples, development of porosity at the diffusion zone has rendered the 'X'- measurements considerably uncertain, making it difficult to locate the Matano surface²⁵. Also, when intermediate alloy compositions predominate, Matano analysis of diffusion couples has not been considered satisfactory²⁵. Winslow²⁶ has given computer programs for the solution of Fick's equations. Nebauer²⁷ has shown that in several cases the penetration curves cannot be analysed on the basis of a single diffusion phenomena. More than one diffusion mechanism may be expected to occur simultaneously. Nebauer has considered the resolution of cases where dual diffusion processes are superimposed on each other such as: (i) the combination of two error function, X- dependences; (ii) the combination of an error function, X- dependence at relatively small X- values and an exponential dependence at the tail end; (iii) non-Fickian X-profiles which cannot be resolved into error function or exponential function components²⁷.

Extreme care has to be taken in the preparation and bonding of diffusion couples. The samples in the case of metals and alloys are well annealed to remove mechanical and chemical inhomogeneities, and prepared in the form of short cylindrical specimens with flat ends by careful cutting and polishing techniques. Cold worked layers are removed by electropolishing or etching. If Kirkendall movements are to be studied, fine wire or powder markers are inserted at the interface before the bonding of the couple elements. Bonding is achieved by subjecting the interface to high pressures in a high temperature furnace, or by electrical resistance bonding in a special apparatus²⁸ or by careful welding in inert atmosphere. Where only a thin layer (few hundred Angstroms) of a highly specific radioactive material is to be deposited on the surface, electro-deposition or vacuum evaporation methods have been employed. The diffusion couples are then sealed off in a close fitting quartz capsule under vacuum, in order to avoid oxidation during the diffusion experiments, which are conducted in a constant temperature furnace for a known period of time. In high temperature work, corrections may have to be made for the warmup time of the specimens, if the diffusion times are short.

After the diffusion experiment, the specimen is removed for the determination of the concentration profile. Several ingenious methods are available. For metal and alloys, accurate sectioning techniques are now known, yielding several thin uniform slices of few micron thickness of material for analysis²⁹. In the case of polymers, the material is often previously sliced, and the layers put together, to enable a quick splitting of the layers into

composite parts³⁰. When the diffusion of radioactive isotopes is studied, which is by far the most accurate method and has the advantage of minimising concentration gradients, the concentration of the isotope is obtained by standard counting techniques. Electron micro beam probes have also been employed successfully²⁵. Microchemical analysis has occasionally been applied to the slices from diffusion couples⁸, as well as X-ray fluorescence and microhardness measurement methods. Refractive index based on interference patterns and the "Lamm scale" technique have both been used¹² for polymer systems. In the case of diffusion of coloured substances, a microdensitometer method or X-ray microradiography has also been employed¹².

Some investigators have attempted to avoid the sectioning of the specimens. Instead, by cutting the sample at a small angle with respect to the initial plane, the tracer profile has been determined autoradiographically³¹. Alternatively, the penetration of the tracer has been estimated by determining the rate of decrease in the activity at the initial surface³². In polymer systems, a variant of this method has been to observe the rate of advance of sharp boundaries into the polymer, e.g. water-polymer interface or dye-polymer interface³³. There are other interesting indirect methods of determining the concentration profiles; however, none of these can be said to achieve the same degree of precision as the radioactive tracer analysis method following direct sectioning of the diffusion couples.

Besides the direct methods analysed above, a number of interesting indirect methods are known. Only a few will be mentioned here. The first of a class of such methods is based on relaxation phenomena, and has received considerable attention in recent years^{34,35}. For cubic systems it can be shown that

$$D = d^2/6\gamma \quad (55)$$

where d = nearest - neighbour distance in the crystal, and γ the "mean jump time". When d is known and γ determined experimentally, D can be calculated without a macroscopic mass flow of the diffusing atoms. Mechanical, dielectric and magnetic relaxation methods are techniques for measuring γ directly. These methods are most frequently employed for the determination of diffusion coefficients of interstitial impurity atoms in metals. Since the method depends on the movement of dissimilar atoms under applied stress, it is not useful for self-diffusion studies in pure metals. Another technique is based on the use of nuclear magnetic resonance. It depends on the measurement of the mode of decay of the transverse (spin-spin) and the longitudinal (spin-lattice) magnetisation and the corresponding relaxation times. A modification which depends on a train of magnetic pulses given to the sample, instead of a single pulse,

at definite intervals of time, known as the spin-echo technique, is said to be more appropriate for diffusion studies³⁶. Self diffusion in alkali metals³⁷ and diffusion in polyisobutene-solvent systems³⁸ have been determined by this method. Although at present its accuracy is not high and diffusion coefficients less than $\sim 10^{-7} \text{ cm}^2/\text{sec}$ cannot be measured satisfactorily, this technique is particularly suitable for low temperature measurements and studies under non-equilibrium conditions.

For semiconducting materials, the movement of the p-n junction under definite boundary conditions has been followed in order to obtain the diffusion coefficients. The method, therefore, is necessarily limited to the diffusion of acceptors in n-type and donors in p-type semi-conductors. The location of the p-n junction is determined by lapping the specimen at a small angle; the surface is then probed with a thin wire, the resulting polarity being determined by a cold probe (rectifier) or hot probe (thermo-e.m.f.) technique. The measurement of electrical conductivity or the junction capacity have also been found to yield satisfactory data for the concentration of the impurity atoms in the diffusion zone^{10,39}.

In the case of diffusion of vapours or gases in polymers, a common indirect technique is to determine the adsorption - desorption kinetics under specified conditions¹². This technique assumes that the rate of uptake, for example, is a diffusion controlled process. A measure of the ratio of the uptake at time t to that at infinite time (equilibrium concentration) gives a measure of the diffusion coefficient. If D is constant, then the solution is simple; for example,

$$D = \frac{0.04919}{(t/\delta^2)^{1/2}} \quad (56)$$

where $(t/\delta^2)^{1/2}$ stands for the value at half-time of the adsorption or desorption process. If D is concentration dependent, then a series of experiments at different initial concentrations are conducted. The initial slope of each experimental curve yields a mean value, \bar{D}_i . Thence using the approximate relationship

$$\bar{D}_i = \left(\frac{1}{n_{io}}\right) \int_0^{n_{io}} D_i dn_i \quad (57)$$

a graph of $\bar{D}_i n_{io}$ vs n_{io} is drawn; and its numerical or graphical differentiation yields an approximate function for D_i in terms of n_i . Several variations of this method have been adopted in practice¹².

REFERENCES

1. R.F. Mehl, J. Appl. Phys., 8, 174 (1937).
2. R.M. Barrer, Diffusion in & through Solids, Cambridge University Press, London (1941).
3. C.J. Smithells, Metals Reference Book, Butterworths, London (1949).
4. W. Jost, Diffusion in Solids, Liquids & Gases, Academic Press, N.Y. (1952).
5. W. Seith, Diffusion in Metallen, Springer Verlag, Berlin (1955).
6. J. Crank, "Mathematics of Diffusion" Clarendon Press, Oxford (1956).
7. J. Bardeen, Phys. Rev., 76, 1403 (1949).
8. D. Lazarus, Solid State Physics, Vol. 10, 71 (1960).
9. J. Manning, Phys. Rev., 116, 69 (1959).
10. B.I. Bolkaks, "Diffusion in Semiconductors", Translated by J.I. Carasso, Infosearch Ltd., London (1963).
11. B.I. Bolkaks, "Diffusion in B.C.C. Metals", Am. Soc. Metals, Metals Park, Ohio (1965).
12. J. Crank and G.S. Park, "Diffusion in Polymers", Academic Press, London (1968).
13. G.S. Hartley, Trans. Farad. Soc., 42, 6 (1946).
14. A. Smigelskas and E. Kirkendall, Trans. AIME, 171, 130 (1947).
15. G.S. Hartley and J. Crank, Trans. Farad. Soc., 45, 801 (1949).
16. (a) L.S. Darken, Trans. AIME, 174, 184 (1948).
(b) L.S. Darken and Gurry, Physical Chemistry of Metals, McGraw Hill (1963).
17. S. Prager, J. Chem. Phys., 21, 1844 (1953).
18. R.P. Smith, Acta Met., 1: 578 (1953).
19. R.M. Barrer, Proc. Phys. Soc., 58, 32 (1946).

20. H.L. Frisch, J. Phys. Chem., Wash., 61, 93 (1957).
21. B. Rosen and J.H. Singleton, J. Polym. Sci., 25, 225 (1957).
22. J.H. Huldy, J. Appl. Polym. Sci., 8, 2883 (1964).
23. H.S. Carslaw and J.C. Jaeger, "Conduction of heat in Solids", Oxford Univ. Press, London (1947).
24. C. Matano, Jap. J. Phys., 8, 109 (1932-3).
25. G.S. Hartley, J.E. Steedly and L.D. Parsons, (p. 51-75 of ref. 11).
26. F.R. Winslow, A Fortran Program for Calculating Diffusion Coefficients and Plotting Penetration Curves, ORNL - TM - 726 (1963).
27. E. Nebauer, phys. stat. sol., 36, K63 (1969).
28. A Bolk, Acta. Met., 9, 632 (1961).
29. G.T. Tomizuka, "Methods in experimental physics", Vol. 6, Acad. Press, New York, pages 364 - 73 (1959).
30. T. Gillespie and B.M. Williams, J. Polym. Sci., A-1, 4, 933 (1966).
31. H.C. Gatos and A.D. Kurtz, J. Metals, 6, 616 (1954).
32. A.A. Zhukhovitsky, Conf. Acad. Sci., USSR on "Peaceful Uses of Atomic Energy", Washington D.C. 1956.
33. a) G.C. Hartley, Trans. Farad. Soc., 45, 820 (1949).
b) R.E. Barker Jr, J. Polym. Sci., 58, 583 (1962).
34. A.S. Nowick, "Progress in Metal Physics", Vol. 4, Chapter 1 (1953).
35. R. Gibala and G.A. Wert, (pp: 131-149 in ref. 11).
36. E.O. Stejskal and co-workers, J. Chem. Phys., 43, 288 (1965)
ibid, 43, 1068 (1965).
37. R.E. Norberg and C.P. Slichter, Phys. Rev., 83, 1074 (1951).
38. D.W. McCall and co-workers, J. Polym. Sci., A1, 1709 (1963).
39. W.C. Dunlap Jr, Phys. Rev., 94, 1531 (1954).

ELECTRICAL TRANSPORT IN IONIC CRYSTALS

C. Ramasastry, Y. V. G. S. Murty, and B. S. V. S. R. Acharyulu

Department of Physics

Indian Institute of Technology, Madras-36, India

I. CONDUCTION MECHANISMS IN IONIC CRYSTALS

Electrical conduction in solids is caused by the directed movement of electrons, holes or ions under the action of electric fields applied from outside. The electrical conductivity σ is equal to $\sum n_i q_i \mu_i$ for a solid obeying Ohm's law. Here, n_i is the number per unit volume of the mobile charge carriers of the i th type, q_i the charge on each such carrier and μ_i the mobility. There is no transport of matter when electrical conduction is by electrons or holes as in metals and semi conductors. Such a conduction can be recognised by Hall effect measurements. On the other hand, in ionic conduction, charge transport is accompanied by mass transport. Crystals like KCl are appropriately called solid electrolytes. Hall effect, is too small to be observed in this case. However, when a d.c. current is passed for a sufficiently long time through three plates of the crystal in series, the end plates should show a change of mass if the conduction is by the movement of the ions. In fact, it should be possible to deduce the transport numbers of ions. If the measurements are sufficiently accurate¹⁻⁷. Even otherwise, qualitative information can be obtained as to the more mobile carriers. In the case of mixed conduction, partly due to electrons and partly due to ions, the situation would be more complex; Hall effect and mass transport measurements should be able to separate out the two effects. Measurement of electrochemical potentials may also be of assistance in this respect.

A crystal may be referred to as ionic if fifty percent or more of its binding energy is due to Coulomb forces between its constituents, the rest of the binding being partly due to van der Waals

and covalent interactions. At all temperatures, the ions will be

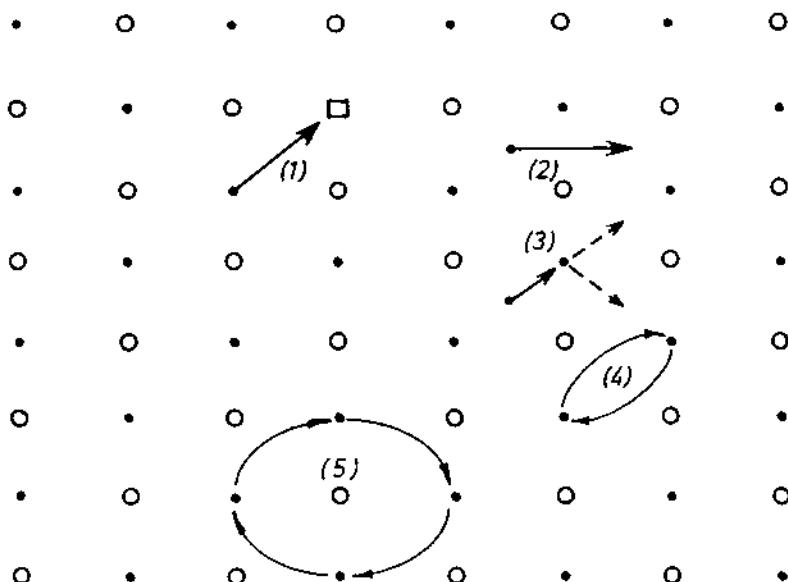


Fig. 1. Jump Mechanism in Crystals (1) Vacancy Mechanism; (2) Interstitial Mechanism; (3) Two possible interstitialacy Mechanism; (4) Two Atom Ring and (5) Four Atom Ring. The Ring Mechanisms 4 and 5 do not contribute to mass or charge transport but they are effective in tracer diffusion.

vibrating about their equilibrium positions but they may leave their minimum energy positions in the lattice if the thermal energy is sufficiently high; Schottky or Frenkel defect pairs are then created. Because of the ordered way in which the atoms or ions are arranged in crystals, there are only a few ways in which the ions can jump about. These are illustrated in Fig. 1 and the possible jumps of a positive ion vacancy in an NaCl type lattice are indicated in Fig. 2.

Normally, the motions of the ions are random but there will be a net flow in a particular direction under the influence of an electric field. This directed motion causes electrical conduction. Fig. 3 shows how the energy of a potential maximum occurring down the field is lowered with respect to a minimum, thereby increasing the probability of the positive ion jumps in the direction of the field. Atoms or ions can also move under concentration gradients which cause diffusion. Mass transport occurs in both the processes. If the jump mechanisms are the same, the electrical conductivity

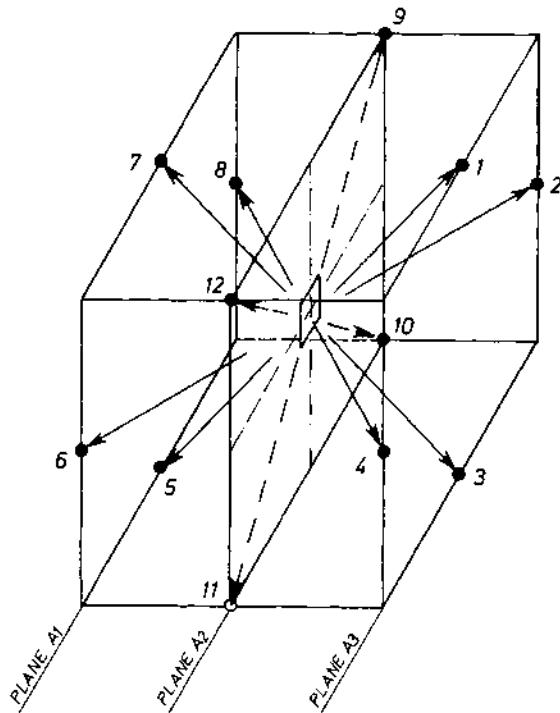


Fig. 2. Unit cell of an f.c.c. ionic lattice: The full circles represent the positive ion sites. The square at the centre is the vacancy. Positions of the negative ions are not shown. The positive ion vacancy carries an effective negative charge in ionic crystals. Sites 1- 12 are nearest like ion positions to which the vacancy can jump.

is related to the diffusion coefficient D by the Nernst-Einstein-Bardeen relation,

$$\frac{\sigma}{D} = f \frac{Nq^2}{KT} \quad (1)$$

where N is the number per unit volume of the sites where defects can be formed, q the effective charge on each defect, k the Boltzmann constant and T the absolute temperature. In (1) f is the correction factor (slightly less than unity) due to Bardeen and Herring⁸ who took into account the possibility of the atom jumping back into the place vacated by it. The factor 'f' is different for different jump mechanisms (see the chapter by V. Ramakrishna on Diffusion in Solids).

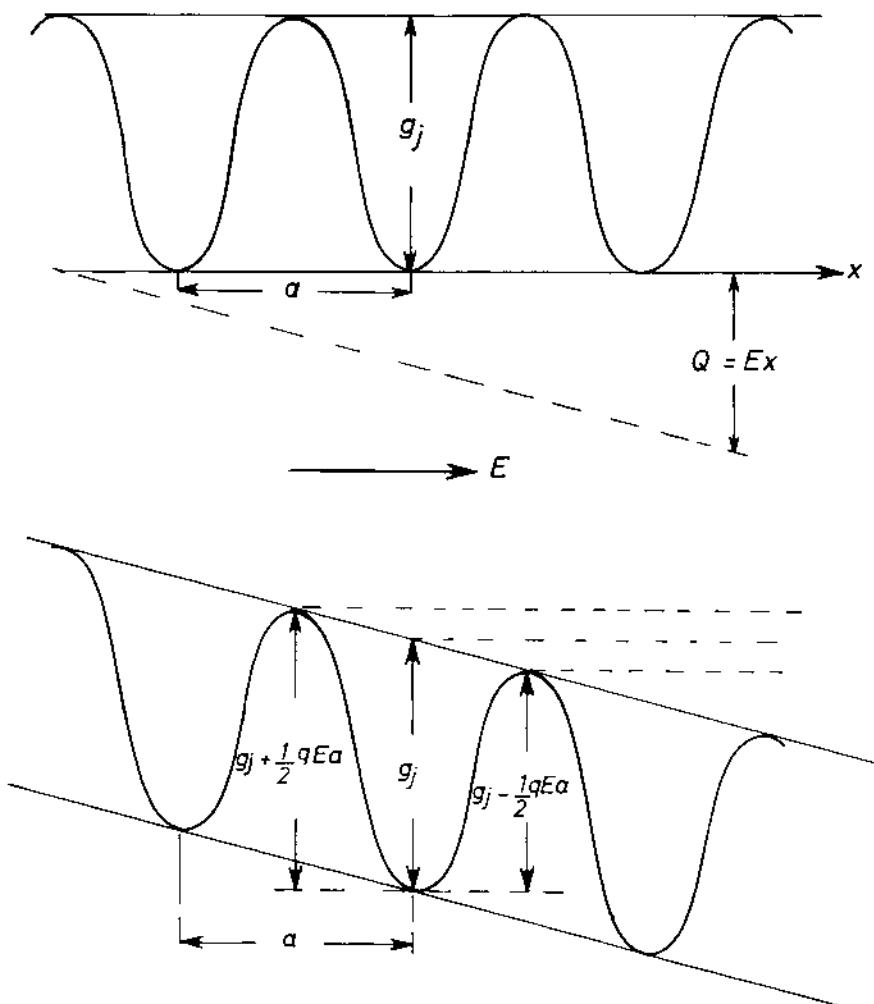


Fig. 3. Potential in a periodic lattice : (a) without an external electric field; (b) with an external electric field E . Jumps of positive ions are favoured to the right and that of positive ion vacancies to the left because of the altered potentials due to the applied field.

The ring mechanisms shown in Fig. 1 can cause diffusion of atoms without the accompanying transport of mass or charge⁹. In the interstitialcy mechanism, the charge jumps through one lattice spacing while the tracer atom moves only through one half the distance¹⁰. A positive ion vacancy and a negative ion vacancy may move together causing mass transport, but not charge transport.

In all these cases, the above relation between σ and D does not obviously hold good. However, the determination of both σ and D will help in elucidating the type of jumps involved and the nature of the current carriers. In the classic example of silver sulphide, self-diffusion studies have played a crucial role in deciding in favour of electronic conduction rather than ionic conduction^{11,12}.

The concentration of defects can be altered in ionic crystals at the lower temperatures by the addition of suitable impurities, because of the electrical neutrality condition. A positive ion vacancy appears to carry a negative charge and the negative ion vacancy a positive charge. The excess charge of a divalent cation impurity added to monovalent ionic crystal can be compensated for either by positive ion vacancies or negative ion interstitials. If ions like Ca^{2+} are introduced into crystals like KCl, the positive ion concentration increases for purposes of charge compensation. As a result, the conductivity of the crystal would increase if the conduction is by the vacancy jumps. This holds true for alkali halide crystals. On the other hand, if the conduction is by the motion of the interstitials, the vacancies would suck-in the interstitials hereby reducing the conductivity. Such a decrease in conductivity has been observed in silver halides doped with cadmium¹³ and in sodium nitrate doped with barium¹⁴. If the impurity doping is excessive, it may be precipitated as a separate phase or the interstitials may be reduced to such a low concentration that the vacancy mechanism may become prominent^{3,7,13}. The effect of changing the concentration of one type of defects on the concentration of other defects in the crystal can be quantitatively understood with the help of the law of mass action. Association reactions are possible between defects of opposite charge: divalent positive ion impurities, Z^{2+} , with positive ion vacancies, V_p^+ , or negative ion interstitials, X^- , and of positive ion vacancies with negative ion vacancies. At the lower temperatures, so long as these associations are effective, the concentration of 'free' defects will not be constant but depends on temperature and impurity concentration (see the chapter by J.S. Anderson on Defect Chemistry and the chapter by G.N.R. Rao and S. Ramdas on Theory of Point Defects for details).

We shall now examine how these ideas about energies of association, jump and formation of defects will help in understanding the experimental data on the electrical conductivity of ionic crystals. A.C. conductivity measurements are generally preferred to avoid polarisation effects, but at low temperatures (when the crystal resistance is rather high) one needs to use high voltages upto 500 volts. D.C. methods are more sensitive in such cases¹⁵. It is preferable to have the meter input resistance R_g variable so that its value can be chosen to have about 50% of the applied voltage across the crystal. Leakage paths along the edges of the crystal should be checked by the guard ring method¹⁶.

It is recommended that the resistance values at different temperatures be taken to correspond to the same voltage across the crystal. Thermolectric effects should be taken into account while calculating the conductivity or may be made negligible by having a sufficient voltage drop across the crystal. Circulation of pure dry inert gases like nitrogen or argon through the furnace is recommended to reduce temperature gradients near the crystal and also to avoid oxidation effects. Thin aluminium foils may be used as electrodes for low melting crystals while platinum foils should be used for high temperatures (above 350°C). It is good to use graphite or silver conducting paints if there is no danger of contamination¹⁷. It may be advisable to take the measurements first without paint and then with paint and compare the results. The $\log (\sigma T)$ vs $1000/T$ plots obtained for the two cases may be fused together in the high temperature region by a small vertical shift of the curves to take into account any differences in the areas of contact of the electrode foils with the crystal faces. Then, the extrinsic region will indicate if any impurities have diffused from the conducting paint into the crystal.

The conductivity data obtained on two KBr crystals is shown in Fig. 4. The dots refer to a Harshaw crystal and the circles to a copper doped crystal. That copper has entered the crystal is verified from its optical absorption around 265 m⁻¹¹⁸. It is in a dispersed state but could not change the conductivity significantly. This implies that the vacancy concentration in the crystal is not affected in the lower temperature regions; copper has entered the lattice substitutionally as Cu⁺, but not as Cu²⁺.

The conductivity of a crystal is expressed as a sum of exponentials,

$$\sigma - T = A_1 e^{-E_1/KT} + A_2 e^{-E_2/KT} + A_3 e^{-E_3/KT}$$

The activation energies are different and often each exponential may be operative in a specific temperature range. The $\log (\sigma T)$ vs $1000/T$ plot generally exhibits breaks with two or three straight line regions with different slopes; these slopes yield activation energies effective in the different regions. It may be not however, be easy to readily draw such straight lines since the experimental curve may show smooth bends (rather than sharp breaks) over an extended temperature region. Unless the crystal is relatively pure and has a high melting point, the change from extrinsic to intrinsic condition is not clear cut and sharp. The situation is all the more difficult in the case of low melting crystals like NaNO₃ or NaClO₃. The method developed by Mr. Viswanatha Reddy of this laboratory is illustrated in fig. 4. A few arbitrary lines P, Q, R are drawn to represent the exponential effective in the lowest temperature region

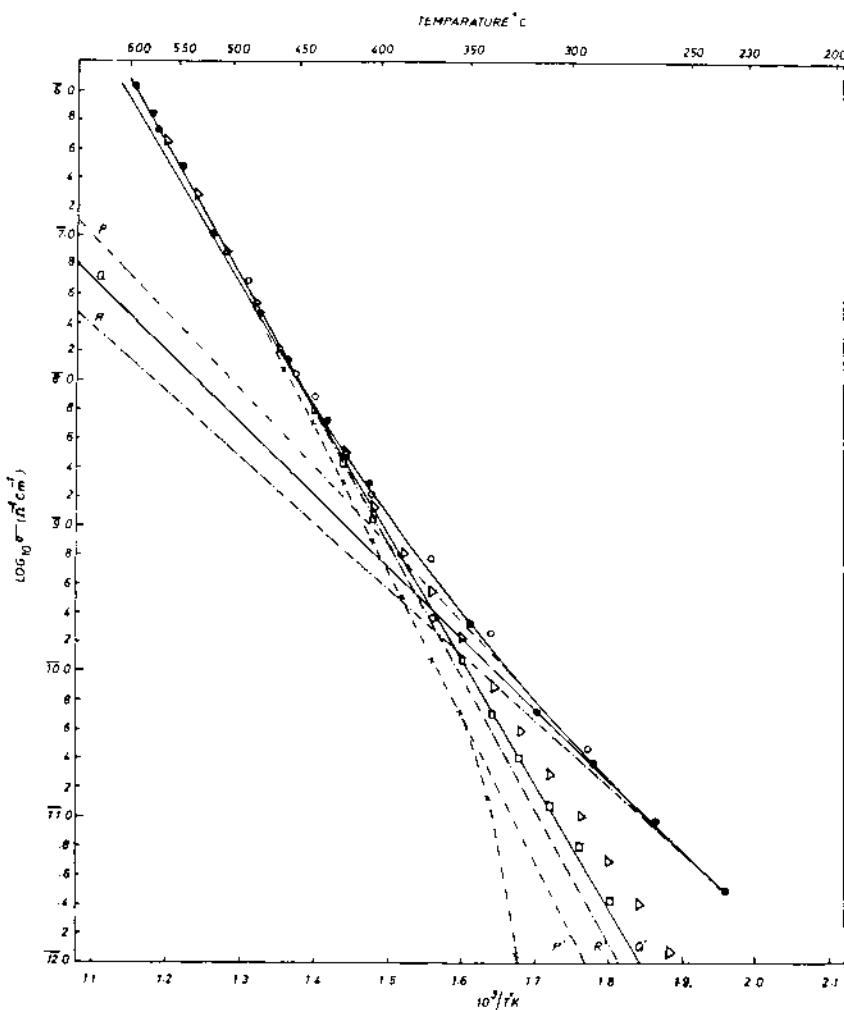


Fig. 4. Electrical conductivity of potassium bromide crystals: Dots represent the data on a Harshaw crystal and open circles, the data on a copper doped KBr crystal with the starting material being of Merck guaranteed reagent purity. Copper enters the lattice as Cu^+ .

of the curve. Antilogs of the various points on these lines are subtracted from the antilogs of the corresponding points on the main experimental curve and then the logarithms of the differences are plotted again to give the residual conductivity still to be accounted for. The line Q is satisfactory as all the residual points (squares) fall on the second line Q' from the mid temperatures down to low temperatures studied. For the lines P and R, the

residual points lie below and above the extended parts of the lines P' and R'. The activation energies (slopes multiplied by 0.2) corresponding to the lines Q and Q' are 1.01 and 1.77 eV respectively. The slopes and the intercepts so obtained can be fed back into the expression for N_d and the calculated values compared with the experimental values. One should then attempt a least square fit.

Interpretation of activation energies is beset with difficulties. In the very low temperature region, defects will be released from their association with the impurities as the temperature is raised. The jump probability also increases. Hence the activation energy in this region should correspond to $(E_j + E_a/2)$ where E_a is the energy of association. At somewhat higher temperatures, all the charge compensating defects would get detached and become free to move about. Rise in temperature increases the jump probability $\exp(-E_j/kT)$, but not the number of current carrying mobile defects to any significant extent. The slope in this temperature range corresponds to the jump activation energy E_j only. But it is very difficult to know where in the curve the dissociation of the impurity-defect complex is complete, unless one gets a lower slope for intermediate temperatures compared to the slope for low temperatures, $E_j < E_j + (E_a/2)$. At still higher temperatures, thermal energy becomes adequate to toss up the ions or atoms from their lattice sites to produce the so called intrinsic defects, namely, Schottky or Frenkel pairs. From now on, the concentration of defects increases exponentially besides the increase in the jump probability. The activation energy in this region represents $(E_j + E_f/2)$ where E_f is the formation energy of a defect pair.

At temperatures near the melting point, ions of both charges may contribute to electrical conduction. Further complications may arise if there is electronic conduction also, the temperature being sufficiently high to excite the electrons from the valence band to the conduction band. However, electron hole conduction is generally negligible in pure ionic crystals due to the large band gap. If electronic conduction is present at all, it can be detected in principle by Hall effect measurements, but there are difficulties in such measurements at high temperatures. The activation energy in this region should give $(E_{2j} + E_{2f}/2)$ after deducting the conduction due to the first type of ions which are effective from mid-temperature upwards (the subscript 2 refers to the second ion or to any other jump mechanism). For Schottky defects, $E_{2f} = E_f$, but $E_{2j} \neq E_j$. If this comes out to be pretty large and near about the forbidden energy gap, one may associate it with electronic conduction. In not too pure crystals, the concentration of mobile defects may be affected by the presence of impurities and electron or hole centres like colour centres. Electrons or holes may then be released from these defects at comparatively lower temperatures as in thermolu-

minescence. There may be some bursts of current but they will be in narrow temperature regions and should not cause much difficulty in studies on ionic conduction. It is common practice to attribute the excess conduction at temperatures nearing the melting point to the mobility of the second type of ions; experimental data on alkali halides are analysed to include the contribution of halogen ion motions in addition to those of the alkali ions¹⁹⁻²³.

II. ANISOTROPY OF ELECTRICAL CONDUCTION IN IONIC CRYSTALS

Electrical conductivity of a non-cubic crystal is not an isotropic property. It is to be represented by a second rank tensor as it relates the electric current and electric field which are two vector quantities: $j_i = \sigma_{ik} E_k$ where the repeated index k indicates summation over all its possible values ($J = \sigma E$ is Ohm's law). The detailed relationship between the components of the current density and electric field is

$$j_1 = \sigma_{11} E_1 + \sigma_{12} E_2 + \sigma_{13} E_3$$

$$j_2 = \sigma_{21} E_1 + \sigma_{22} E_2 + \sigma_{23} E_3$$

$$j_3 = \sigma_{31} E_1 + \sigma_{32} E_2 + \sigma_{33} E_3$$

The conductivity is thus to be specified by nine coefficients σ_{ij} ($i = 1, 2, 3$ and $j = 1, 2, 3$). An interesting feature of this tensorial property is that when an electric field E is applied along the direction X_1 , so that $E_1 = E$, we have currents in the perpendicular directions X_2 and X_3 : $j_2 = \sigma_{21} E_1$ and $j_3 = \sigma_{31} E_1$ besides the current $j_1 = \sigma_{11} E_1$ along the applied field E . According to Onsager's theorem²⁴, $\sigma_{ij} = \sigma_{ji}$ for $i = j$ and we, therefore, need to know only six coefficients. Crystal symmetry further simplifies the tensor according to Neumann's law^{25,26} which states that every physical property of a crystal would possess at least the symmetry exhibited by the morphology of the crystal if not more. A physical property represented by a second rank tensor like the electrical conductivity exhibits spherical symmetry (isotropy) for a crystal which has only the cubic symmetry.

A second rank tensor may appear simple if the reference coordinate frame is rotated through a suitable angle so that the tensor has diagonal elements only. If a is the transformation matrix, the diagonalisation procedure requires, $a^{-1} \sigma \cdot a = \sigma_{\text{diagonal}}$. In the diagonalised form, we have only j_1 for E_1 , j_2 for E_2 and j_3 for E_3 , so that the current and electric field are in the same direction. These

special directions in the crystal are called the principal directions of the tensorial physical property. The principal directions of the conductivity tensor and the minimum number of non-vanishing tensor coefficients arrived at from the application of the Neumann's principle for the different crystal systems are given below.

<u>Crystal System</u>	(a) <u>Non-Vanishing tensor coefficients</u> (b) <u>Principal directions</u>
Cubic	(a) One only, σ (b) Any three mutually perpendicular directions.
Tetragonal	(a) Two; $\sigma_{11} = \sigma_{22}$, and σ_{33} (b) The unique axis as X_3 and any two mutually perpendicular directions in the plane perpendicular to X_3 .
Trigonal Hexagonal	Same as for tetragonal
Orthorhombic	(a) Three; σ_{11} , σ_{22} , σ_{33} . (b) The three crystallographic axes.
Monoclinic	(a) Four; σ_{11} , σ_{22} , σ_{33} , σ_{12} (b) The two fold or 2 as X_3 and any two mutually perpendicular directions that would diagonalise the tensor.
Triclinic	(a) Six; σ_{11} , σ_{22} , σ_{33} , σ_{12} , σ_{13} , σ_{23} . (b) Three mutually perpendicular directions such that the tensor is diagonal.

Neumann's principle does not include the Onsager theorem. The former, for example predicts that in a trigonal system, $\sigma_{21} = -\sigma_{12}$ and the latter that $\sigma_{21} = \sigma_{12}$ so that in between them they make $\sigma_{21} = \sigma_{-21} = 0$.

Anisotropy causes inconvenience in the experimental determination of electrical conductivity. Let a crystal plate be cut in an arbitrary direction and the coordinate axes X'_1, X'_2, X'_3 be chosen with X'_1 normal to, and X'_2, X'_3 in the plane of, the faces of the crystal plate. If electrodes are attached to its faces and a known voltage applied, an ammeter in series with the crystal would measure the sum of three current components j'_1, j'_2 and j'_3 for the field E'_1 along X'_1 . The resultant current in the crystal is not along the electric field. The conductivity, if calculated, does not represent the physical property of the material since it depends on the dimensions of the plate. If we choose a needle-like crystal with a rectangular

cross section $t_2 \times t_3$ and its length along X'_1 , the current is constrained to flow along X'_1 because of the geometry²⁷ and $j'_2 = j'_3 = 0$. Then $\sigma'_{11} = j'_{11} = j'_1/E'_1$. Further, the current j'_1 will produce cross voltages across t_2 and t_3 even though the applied voltage is along X'_1 only. One can perhaps determine the various coefficients of the resistivity tensor, ρ , from the measurement of such cross voltages but till now, no such voltages are reported for any crystal. The only alternative then is to choose, as far as possible, crystal plates cut perpendicular to the principal directions.

Conductivity anisotropy for ionic crystals has been determined till now with two crystals. The results of Ramasastry and Murti¹⁴ on NaNO_3 which belongs to the trigonal system are: (i) Conductivity along the triagonal axis, [111], (σ'_{11}) is higher than in the plane perpendicular to it (σ'_1) both in the intrinsic and extrinsic regions. (ii) The slopes of $\log (\sigma T)$ vs $1000/T$ plots are not different for the two directions both in the intrinsic and extrinsic regions. (iii) The conduction mechanism is by the jumps of the interstitial positive ions; whether the mechanism is interstitialcy or interstitial is not yet clear. (iv) The ratios of the conductivities $\sigma'_{11}:\sigma'_1$ are approximately 1.95 : 1.0 in the intrinsic range and 1.48 : 1.0 in the extrinsic range.

The authors concluded that the type of defects and the jump mechanisms are the same for both the directions and that the cause for the anisotropy should be looked for in the preexponential factors but not in the activation energies.

The second example is potassium sulphate which belongs to the orthorhombic system. Ramasastry and Acharyulu^{28,29} studied the conductivities of these crystals upto 400°C and the anisotropy observed for the three principal directions (crystallographic axes) is shown in Fig. 5. The results are believed to correspond to the extrinsic range. The ratio of the conductivities $\sigma'_{11} : \sigma'_{22} : \sigma'_{33} = 3.3 : 1.6 : 1$ which agrees well with the squares of the sides of the unit cell ($a = 10.06$, $b = 7.485$, $c = 5.726 \text{ \AA}$) as required by the theoretical expressions for the conductivity in the three directions if the same jumps are involved. This encouraged these authors to find an explanation of the anisotropy of electrical conduction in ionic crystals in terms of the ion jumps. This will now be illustrated for a plane lattice. The explanation should be rather simple because the isotropic conduction in a square lattice should become anisotropic theoretically even if the square is slightly distorted to become a rhombus. For slight changes, the atomic jumps should not be changed materially, but now, because of the distortion, the distances of the first like neighbours will be changed.

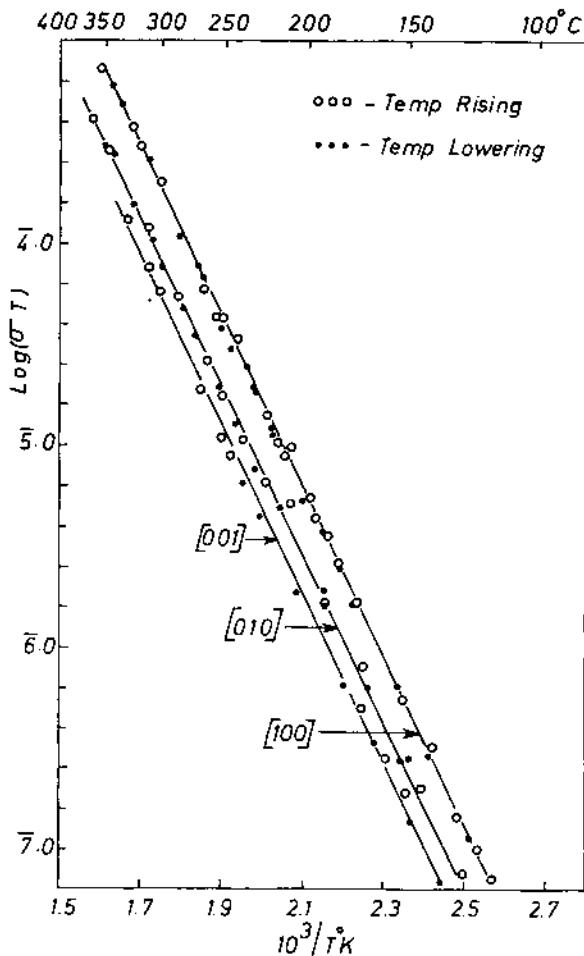


Fig. 5. Anisotropic electrical conduction in potassium sulphate crystals (orthorhombic system): The conductivities in the three principal directions are 3.3 : 1.6 : 1 which agrees with the ratio of the sides of the unit cell (10.06, 7.485 and 5.726 Å).

REFERENCES

1. F. Seitz, The Modern Theory of Solids, McGraw Hill Book Co. Inc., New York, 1940, Pages 65 - 72.
2. N.F. Mott and R.W. Gurney, The Electronic Processes in ionic crystals, Oxford University Press, Oxford, 1946.

3. A.B. Lidiard, *Handbuch der Physik*, Springer-Verlag., Gottingen, 1957, Pages 250 - 255.
4. O. Stasiw, *Electron and Ion Processes in Crystals*, Springer-verlog., Berlin, 1959.
5. A.F. Ioffe, *Physics of Semi-conductors*, Infosearch Ltd., London, 1960, Pages 2-5.
6. A.J. Dekker, *Solid State Physics*, Macmillan and Co., London 1965, Pages 175 - 176.
7. Frederick C. Brown, *The Physics of Solids*, W.A. Benjamin Inc., New York, 1967, Page 300.
8. J. Bardeen and C. Herring, *Imperfections in nearly perfect crystals*, John Wiley and Sons, New York, 1952, page 261.
9. C. Zener, *Acta Cryst.*, 3, 346 (1950).
10. W.D. Compton and R.J. Maurer, *Journal of Phys. Chem. Solids*, 1, 191 (1956).
11. C. Tubandt, *Handbuch der Experimental Physik*, Vol. 12 (1) (1932) Springer-Verlag, Gottingen.
12. C. Wagner, *Zeits fur Physik Chem.* 1, 23, 467, 1933.
13. I. Ebert and J. Teltow, *Ann Phys.*, 15, 268 (1955).
14. C. Ramasastry and Y.V.G.S. Murty, *Proc. Roy. Soc., A* 305, 441, - 445 (1968).
15. R.W. Dreyfus and A.S. Nowick, *Phy. Rev.*, 126, 4, 1962.
16. W. Crawford Dunlop. Jr. *Methods of Experimental Physics* Vol. 6, Part B. Academic Press, New York, 1959, Page 39.
17. H. Grundig, *Z. Physik.*, 182, 477 - 486 (1965).
18. Kratzig, *et al*, *Physica Stat. Solidi.*, 10, 709 (1965).
19. R.G. Fuller, *Phys. Rev.*, 142, 524 (1966).
20. R.G. Fuller and M.H. Reilly, *Physical Review Letters*, 19, 113 - 115 (1967).
21. D.K. Dawson and L.W. Barr, *Physical Review Letters*, 19, 344 (1967).

22. R.J. Friauf, Phys. Rev., 105, 843 (1957).
23. R.J. Friauf, J. Phys. Chem. Solids., 18, 205 (1961).
24. L. Onsager, Phys. Rev., 37, 405 (1931).
25. N.F. Nye, Physical Properties of Crystals Oxford University Press, Oxford, 1957.
26. S. Bagavantham, Crystal Symmetry and Physical Property, Academic Press, London 1966.
27. Smith, C.S., Solid State Physics. Vol. 6 Academic Press New York, 1958.
28. C. Ramasastry and B.S.V.S.R. Acharyulu, Proc. International Conference on Science and Technology of Non-metallic Crystals, Delhi, 1969, Robert Maclehose and Company, University Press, Glasgow, 1970.
29. C. Ramasastry and B.S.V.S.R. Acharyulu, Proc. Nuclear Physics and Solid State Physics Symposium, Roorkee, India Dec. 1969.

INORGANIC GLASSES

D. Chakravorty

Department of Metallurgical Engineering

Indian Institute of Technology, Kanpur, India

I. CONSTITUTION OF GLASS

The American Society for Testing Materials have proposed¹ the definition of "glass" which runs as follows: "Glass is an inorganic product of fusion which has been cooled to a rigid condition without crystallizing". The definition is restrictive by the use of the word "inorganic", since many organic substances form glasses. However, for our present purpose this is adequate as we will be concerned only with inorganic glasses.

Volume-Temperature Relationship. The interrelationship between the liquid, glassy and crystalline states of material can best be brought out in terms of a Volume-Temperature plot as shown in Fig. 1. If a liquid of fixed mass is cooled from the original state represented by 'a' in the diagram, its volume decreases steadily. If the rate of cooling is sufficiently low a discontinuous change occurs at some point 'b' (temperature T_f) which corresponds to crystallization taking place in the melt. As the temperature is lowered further the volume decreases but with a slope different from that obtained prior to crystallization. The temperature T_f is usually referred to as the Transformation temperature. If, on the other hand, the rate of cooling is very fast, crystallization does not take place at T_f , and the volume decreases steadily along the line 'be' until at a certain temperature 'e' (temperature T_g) the volume-temperature curve undergoes a sharp change in direction and continues almost parallel to the line 'cd'. T_g is usually referred to as the Glass-transition temperature and its value depends on the rate of cooling employed - thus, for instance, T_g would be shifted towards T_f if the rate is higher and vice versa. Between T_g and T_f the

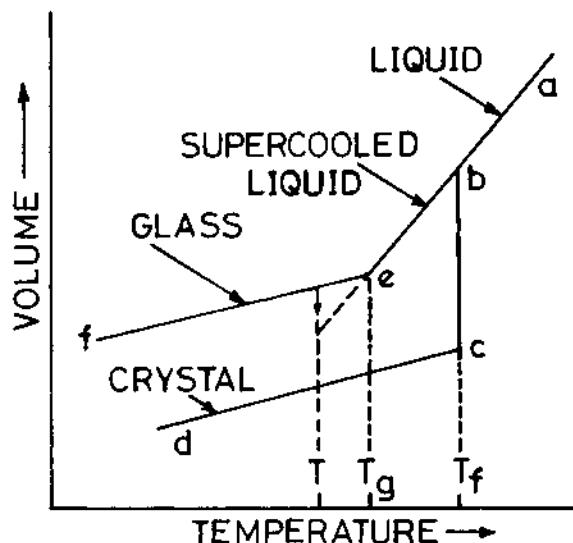


Fig. 1. Relation between the glassy, liquid and solid states (Jones, 1956).

material is a supercooled liquid and below T_g it is glass. Since the volume of the material in the glassy state is larger than that in the crystalline state for a particular temperature (lines 'ef' and 'cd' in the figure) it is obvious that thermodynamically glassy state is a metastable one.

Atomic Structure of Glass. From the discussion in the last section it is apparent that the transition from the liquid to the glassy state is a gradual one. Hence, a similarity in their structures could be expected. This is, in fact, confirmed by their x-ray diffraction patterns *viz.*, broad diffuse diffraction rings are obtained from both these forms² in contrast to the sharp line spectra of crystalline solids. Thus, there is a lack of long range order in glasses.

Chemical Nature of Glasses. While it is true that some inorganic materials form glasses more easily than others the point must be emphasized that whether a substance forms a glass or not depends also on the rate of cooling it is subjected to from its molten state. Some suitable criterion regarding the cooling rate must therefore be used in compiling a list of possible glass-forming systems. In the list presented in the following sections only those materials have been included which form glasses when melted on a scale of not less than about 20 milligrams and allowed to cool freely³.

A. Elements. Some Group V and VI elements in the periodic table form glasses e.g., sulphur and selenium. A vitreous form of phosphorus has recently been reported⁴.

B. Oxide Glasses. SiO_2 , GeO_2 , B_2O_3 , P_2O_5 and As_2O_3 readily form glasses. When cooled very rapidly Sb_2O_3 also does so. Each of the above oxides will form glasses when melted with a second oxide or a mixture of oxides within certain limits of the percentage of components added.

TeO_2 , SeO_2 , MoO_2 , WO_3 , Bi_2O_3 , Al_2O_3 , Ga_2O_3 and V_2O_5 are the so-called "conditional glass formers" viz., by themselves they do not form glasses but will do so when melted with the required amount of a second oxide or mixture of oxides. For example, BaO - TeO_2 system containing 11-18 weight per cent BaO can form fairly stable glasses.

C. Hydrogen-bond Glasses. The chemical bonds existing in the oxide glasses mentioned earlier are known to have partly ionic and partly covalent nature. However, some oxides having hydrogen bonds also form glasses. Examples are, H_2O_2 , HCl , HClO_4 , NH_4OH , KOH and LiCl .

D. Halide Glasses. Glass forming halides are BeF_2 and ZnCl_2 .

E. Chalcogenide Glasses. The binary systems As-S, As-Se, P-Se and Ge-Se form glasses over fairly wide regions of composition.

II. GLASS FORMATION

Structural Theory: Radius Ratio Criterion. Goldschmidt⁵ tried to correlate the ability of an oxide A_mO_n to form glass with its ionic radius ratio $\rho = R_A/R_O$. The importance of ρ in determining the co-ordination number of an ion in ionic solids is well-known⁶. Most crystals with a radius ratio lying between 0.2 and 0.4 have a structure such that each cation is surrounded by four anions - the latter being situated at the corners of a tetrahedron. It has also been found that the glassy oxides, for example SiO_2 , GeO_2 and P_2O_5 have ρ values of 0.2 - 0.4. This led Goldschmidt to conclude that a tetrahedral arrangement of the oxygen ions around the cation A is necessary for glass formation. However, this approach has some inconsistencies in it. The glass-forming oxides under consideration do not have purely ionic bonds; these are partly covalent. Hence the co-ordination number for atoms A as deduced on the basis of non-directional ionic bonds will not hold true in case of the partly covalent A - O bonds which have directional characteristics. Also, the Goldschmidt criterion is not satisfactory even as an empirical rule, since all oxides having radius ratio within the specified

range do not form glasses e.g., BeO.

Zachariasen - Warren Model. Zachariasen based his random network hypothesis⁷ on the following analysis of the problem. The interatomic forces in glass must be essentially the same as in crystals, since the strength of glass is of the same order as that of crystals. Hence, as in crystals, the atoms in glass must form an extended three-dimensional network. However, the diffuseness of x-ray diffraction patterns shows that this network is not a periodic one. Thus Zachariasen proposed the condition for glass formation as the ability of a substance to "form extended three-dimensional networks lacking periodicity, with an energy content comparable with that of the corresponding crystal network".

From the condition that the energy content should be comparable in the two types of networks, one can conclude that the co-ordination number for the cation A in glass should be approximately the same as that in the corresponding crystal. Considering as an example the crystalline SiO_2 and silicates where the silicon atom is surrounded by four oxygen atoms forming SiO_4 tetrahedra one could expect that silica and silicate glasses would also be built up of SiO_4 tetrahedra. The difference between the crystalline and glassy forms is shown schematically in Fig. 2 where the structures of a hypothetical oxide A_2O_3 in both these forms are compared.

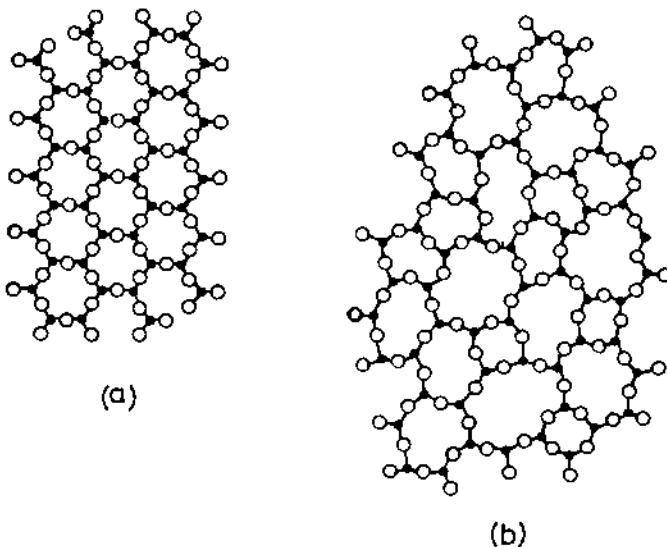


Fig. 2. Schematic two-dimensional representation of the structure of (a) a hypothetical crystalline compound A_2O_3 .

In the oxides A_0 on the other hand, the edges of the structural units must be shared thus leading to the formation of a regular, rigid and periodic structure and no glassy form could be expected of such compound.

The above considerations led Zachariasen to formulate a number of rules for glass formation. They are: (i) No oxygen atom may be linked to more than two atoms A. (ii) The number of oxygen atoms surrounding atoms A must be small, (iii) The oxygen polyhedra share corners with each other, not edges or faces, and form three-dimensional networks. (iv) At least three corners in each oxygen polyhedron must be shared.

Most of the common glass forming oxides are found to obey these rules e.g., SiO_2 , GeO_2 , P_2O_5 , B_2O_3 and As_2O_3 ; the first three having structures based on AO_4 tetrahedra and the last two on AO_3 triangles.

Zachariasen model of glass structure was supported by detailed x-ray diffraction spectra of glasses investigated by Warren and co-workers^{2,8,9}. Also, Warren confirmed the structure of glasses containing other oxides viz., Na_2O , K_2O , CaO etc. in addition to the glass-forming oxides SiO_2 , GeO_2 etc. as proposed by Zachariasen. Fig. 3 is a two-dimensional representation of the structure of a

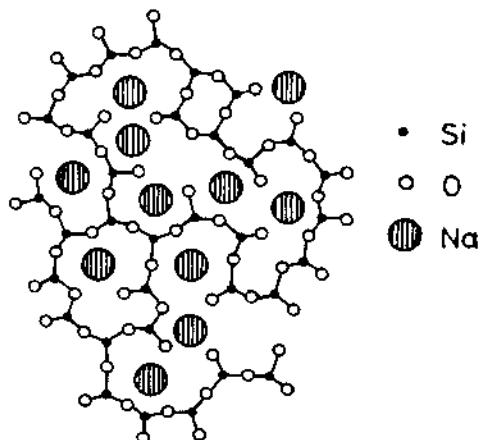


Fig. 3. Two-dimensional representation of the structure of a sodium-silicate glass (Warren, 1938).

sodium silicate glass based on these studies. It is evident from this figure that Na_2O destroys the rigidity of the three-dimensional silica network and the structure becomes more open. Such oxides are therefore referred to as "network-modifying oxides" as opposed

to SiO₂, GeO₂, etc. which are "network forming oxides". Also, in the modified glass network the oxygen atoms bonded to only one silicon atom are usually termed "non-bridging" oxygens and those linked to two silicons are termed "bridging oxygens". According to Warren the distribution of network modifying ions in the glass structure is random which however has not been supported by later workers^{10,11}. In recent years, there has been increasing evidence of cluster formation by the modifying ions⁹.

The validity of the present model has come under serious doubt with the discovery of glasses not obeying Zachariasen's rules. Nevertheless, it has made a significant contribution in this field by introducing, for the first time, the concepts and principles of crystal chemistry into the study of inorganic glasses.

Glass formation and Bond type. The idea of the presence of "mixed" chemical bonds in a material as a necessary requirement for it to form glass was propounded by Smekal¹². According to this theory a random arrangement of atoms that can be maintained on cooling is inconsistent with sharply defined bond lengths and bond angles. In other words, materials having purely covalent bonds are ruled out. On the other hand, purely ionic or metallic bond with complete absence of directional characteristics will ensure an easy rearrangement of ions or atoms at the freezing temperature to form a crystalline structure.

Stanworth¹³⁻¹⁵ tried to put the correlation between the glass forming ability and bond type on a quantitative basis by making use of Pauling's¹⁶ electronegativity values for elements. He grouped different elements along with their electronegativity values in a tabular form as shown in Table II.1. Here, Group I elements forms network forming oxides, Group III elements the network modifying oxides whilst those in Group II occupy an intermediate position. Some definite trend can be observed in these values viz., all elements in a particular group have electronegativity values very close to each other

Stanworth¹⁷ predicted the existence of an entirely new group of glasses on the above considerations. Thus Tellurium has the same electronegativity (2.1) as Phosphorus. Since P₂O₅ readily forms glass. Stanworth looked into the possibility of TeO₂ also being a glass former. He found that TeO₂ does not form glass when melted alone but that stable glasses could be made in many binary systems such as TeO₂-Al₂O₃, TeO₂-BaO, TeO₂-PbO etc. However, Stanworth's criterion failed to give correct prediction in case of Sb₂O₃ and SnO₂ even though Sb and Sn have electronegativity values very close to that of Si.

TABLE II.1

Electronegativity Values of Elements Found in Oxide Glasses¹⁴

$$\text{O} = 3.5$$

Group I		Group II		Group III	
B	2.0	Be	1.5	Mg	1.2
Si	1.8	Al	1.5	Ca	1.0
P	2.1	Ti	1.6 (1.5)*	Sr	1.0
Ge	1.8	Zr	1.6 (1.4)*	Ba	0.9
As	2.0	Sn	1.7 (1.8)*	Li	1.0
Sb	1.8 (1.9)*			Na	0.9
				K	0.8
				Rb	0.8
				Cs	0.7

* Values in brackets are revised values according to Pauling¹⁶.

Glass Formation and Bond Strength. Sun¹⁸ put forward the idea that, since the process of crystallization involves the breaking and reforming of interatomic bonds to bring about a regular arrangement of the atoms, some correlation between the strength of these bonds and the glass-forming ability of a material could be expected. Thus, for instance, the stronger the bonds, the more sluggish will be the process of rearrangement and hence glass will be formed more readily. Sun calculated the bond strengths in different oxides from the values of dissociation energies derived by Sun and Huggins¹⁹. His results are shown in Table II.2. It is seen from this table that bond strengths in glass forming oxides are quite high (greater than 90 Kcal/mole) as compared to those in modifiers (less than 60 Kcal/mole).

Rawson²⁰ modified Sun's approach by combining the effect of thermal energy available to break bonds with the bond strengths. The melting temperature, in case of a simple compound, or the liquidus temperature for a multicomponent system is a measure of the amount of energy available. Rawson therefore used the ratio of the bond strength to the melting temperature in °K as a criterion for glass formation. Table II.3 shows such values. From this table it is seen that B₂O₃ has a very large value for this ratio and on the basis of Rawson criterion should form very stable glasses. In fact, it is almost impossible to devitrify boric oxide glass. However both Sun as well as Rawson criteria fail in the case of V₂O₅ and CO₂ neither of which form glasses.

TABLE II.2
Calculated Bond Strengths of Oxides¹⁸

M in MO _x	Valence	Dissociation Energy per MO _x E _d (Kcal.)	Co-ordination Number	Single bond strength B _{M-O} (Kcal.)
Glass formers				
B	3	356	3	119
Si	4	424	4	106
Ge	4	431	4	108
Al	3	402-317	4	101-79
B	3	356	4	89
P	5	442	4	88-111
V	5	449	4	90-112
As	5	349	4	70-87
Sb	5	339	4	68-85
Zr	4	485	6	81
Intermediates				
Ti	4	435	6	73
Zn	2	144	2	72
Pb	2	145	2	73
Al	3	317-402	6	53-67
Th	4	516	8	64
Be	2	250	4	63
Zr	4	485	8	61
Cd	2	119	2	60
Modifiers				
Sc	3	362	6	60
La	3	406	7	58
Y	3	399	8	50
Sn	4	278	6	46
Ga	3	267	6	45
In	3	259	6	43
Th	4	516	12	43
Pb	4	232	6	39
Mg	2	222	6	37
Li	1	144	4	36
Pb	2	145	4	36
Zn	2	144	4	36
Ba	2	260	8	33
Ca	2	257	8	32
Sr	2	256	8	32

Contd.

Table II.2 Contd.

M in MO _x	Valence	Dissociation Energy per MO _x E _d (Kcal.)	Co-ordination Number	Single bond strength B _{M-O} (Kcal.)
Modifiers				
Cd	2	119	4	30
Na	1	120	6	20
Cd	2	119	6	20
K	1	115	9	13
Rb	1	115	10	12
Mg	2	68	6	11
Cs	1	114	12	10

TABLE II.3

Ratio of Single Bond Strength to Melting Point (°K) for oxides²⁰

Oxide	Single Bond Strength B _{M-O} (Kcal.)	Melting Point T _m (°K)	$\frac{B_{M-O}}{T_m}$
B ₂ O ₃	119 or 89	723	0.164 or 0.122
SiO ₂	106	1993	0.053
GeO ₂	108	1388	0.078
P ₂ O ₅	88 - 111	843	0.104 - 0.131
V ₂ O ₅	90 - 112	943	0.095 - 0.119
TiO ₂	73	2123	0.034
ZrO ₂	81	2923	0.023
MoO ₃	92	1068	0.086
WO ₃	103	1748	0.059
TeO ₂	68	1006	0.067
MgO	37	2913	0.013
CaO	32	2773	0.011
BaO	33	2193	0.015

Rawson's "liquidus temperature effect" helps explain the existence of the "conditional" glass formers (e.g., V₂O₅) which

form glasses only when melted with a second oxide. In such cases, the second oxide has the effect of lowering the liquidus temperature thereby reducing the amount of thermal energy available for breaking bonds. Fig. 4 shows the region of glass formation for the binary system $\text{PbO-V}_2\text{O}_5$.

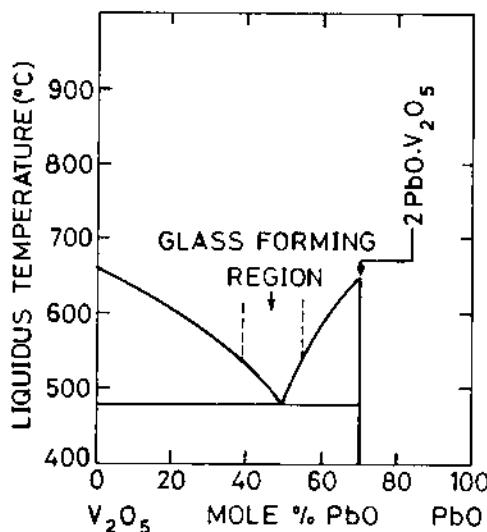


Fig. 4. The glass-forming region in relation to the phase diagram for the system $\text{PbO-V}_2\text{O}_5$ (Rawson, 1956).

Kinetic Theory. The driving force for crystallization to occur in a melt when the latter is cooled below its freezing point is a lowering of the thermodynamic free energy associated with such transformation. A finite barrier of activation energy, however, has to be overcome to bring about the change as represented schematically in Fig. 5. Here A and B correspond to the parent and the product

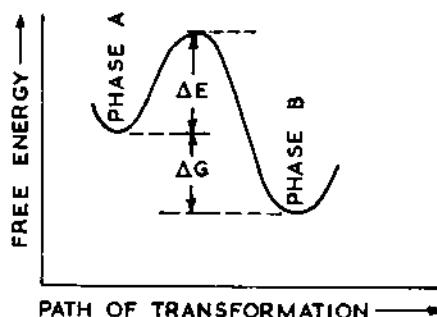


Fig. 5. Schematic representation of the free energy change associated with a transformation.

phases respectively. ΔG represents the driving force and ΔE the activation energy for the transformation.

The crystallization does not occur by an instantaneous and homogeneous transformation of the material from liquid to solid. Firstly, nuclei of the new phase appear, perhaps only a few hundred atoms in size. Secondly, these nuclei then grow in size until the entire melt is transformed. Thus the rate of crystallization will depend on both the nucleation as well as the growth rates.

It can be shown²¹⁻²⁵ that the nucleation rate \dot{n} is given by

$$\dot{n} = \frac{dn}{dt} = N \nu s^* \exp \left\{ -(\Delta f^* + \Delta f_D)/kT \right\} \quad (1)$$

where, N is the number of potential nucleation sites, ν the frequency of vibration of the atoms, s^* the number of atoms of parent phase which are immediate neighbours of the nucleus. Δf^* the free energy of activation barrier to nucleation, Δf_D the free energy of activation barrier for diffusion of migrating species. k the Boltzmann constant and T the temperature in degrees Kelvin.

Also, it is shown²¹⁻²⁵ that

$$\Delta f^* = \frac{K \sigma^3 T_e^2}{(\Delta H_v \Delta T)^2} \quad (2)$$

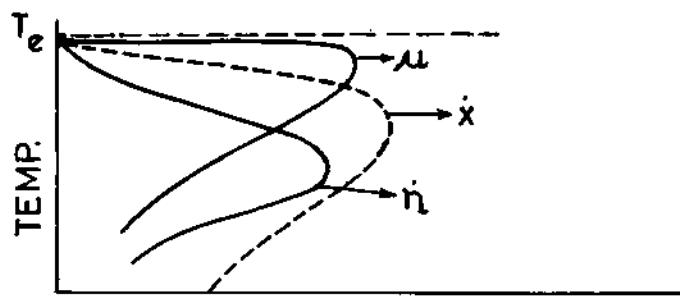
where, K is the numerical factor depending on the shape of the nucleus (for a spherical nucleus $K = 16 \pi/3$), σ the free energy per unit area of the crystal-liquid interface. T_e the freezing temperature of the liquid, ΔH_v the enthalpy change associated with transformation and ΔT the supercooling.

On the other hand, the temperature variation of growth rate u is given by

$$u = C \exp \left(-\frac{\Delta f_D}{kT} \right) \left[1 - \exp \left(-\frac{\Delta g}{kT} \right) \right] \quad (3)$$

where C is a constant and the significance of the other symbols have already been explained.

The temperature variations of \dot{n} and u as predicted by equations (1), (2) and (3) are shown schematically in Fig. 6. The overall transformation rate \dot{x} which is the product of \dot{n} and u is also shown schematically in the same figure. It is evident from this figure that the crystallization rate attains a maximum value at a temperature lower than the freezing point. It is usual to plot the



NUCLEATION RATE, GROWTH RATE, OVERALL TRANS FORMATION RATE

Fig. 6. Schematic temperature variation of nucleation (n), growth (u) and overall transformation (x) rates.

time for different amounts of transformation, e.g. 1%, 50% and 99% at various temperatures to yield a set of C curves as shown in Fig. 7. The diagram is called a T - T - T (Time - Temperature - Transformation) diagram. The lines A and C represent schematically

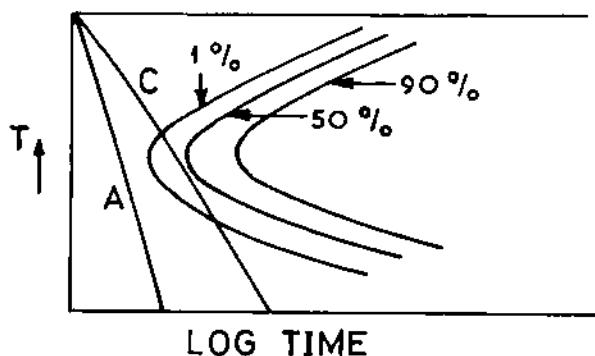


Fig. 7. T - T - T (time-temperature-transformation) diagram. Each curve gives the time-temperature relation for the indicated percentage of overall transformation.

two different rates of cooling, one (A) faster than the other (C). It is obvious that the rate of cooling C will induce some crystallization in the melt because it intersects the "nose" of some of the C-curves whereas A will not be able to do so. In the latter

case, a glass will result. Hence, the position of "nose" of the C-curves associated with a particular material will determine whether it would form glass or not when cooled freely from its liquid state. In general, the tip of the curves will be very close to the temperature axis for materials that crystallize easily and will be far away from the same for glass-forming materials.

III. INORGANIC GLASS-FORMING SYSTEMS

Silicate Glasses. Vitreous silica is the only one-component oxide glass manufactured commercially because of its useful properties. Fused silica however, gets devitrified depending on the impurities present in glass and on the nature of the surrounding atmosphere^{26,28}. It is observed that devitrification usually results in the formation of cristobalite in spite of the fact that the latter is not the stable phase below 1470°C. Stevens²⁹ has explained this by suggestion that the vitreous silica is "cristobalite like" so that the formation of their nuclei requires the least drastic rearrangement of the structure.

The growth of crystals in silica glass is dependent on the amount of second oxide present³⁰. Thus, in $\text{Na}_2\text{O}-\text{SiO}_2$ system by the addition of even 0.32 weight percent of Na_2O , the maximum devitrification rate is increased twenty to thirty times as shown in Fig. 8. In $\text{Al}_2\text{O}_3-\text{SiO}_2$ glasses having Al/Si ratio between 0 and 1.1×10^{-2} , the Al_2O_3 -containing glasses devitrify more rapidly than the pure silica glass³¹.

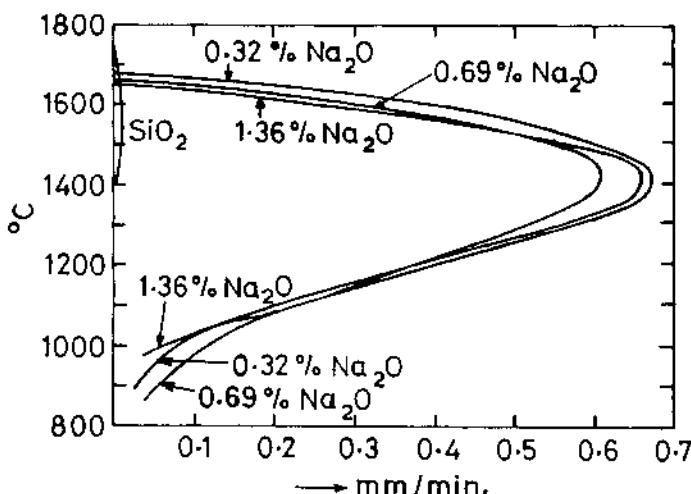


Fig. 8. Rate of crystallization as a function of temperature of vitreous silica and three soda-silica glasses of low Na_2O content (Dietzel and Wickert, 1956).

Commercially the most important glasses are the so-called soda-lime-silica glasses. Besides Na_2O , CaO and SiO_2 they contain some MgO to improve their resistance to devitrification and also Al_2O_3 which makes them chemically more durable.

As discussed in Section I the alkali silicate glasses produce some "non-bridging" oxygen atoms which make the structure weaker than that of vitreous silica. This is confirmed by the higher expansion coefficient and lower viscosity which results when the alkali oxide is added to SiO_2 . The ranges of glass formation in $\text{Li}_2\text{O}-\text{SiO}_2$, $\text{Na}_2\text{O}-\text{SiO}_2$ and $\text{K}_2\text{O}-\text{SiO}_2$ systems have been shown³² to be continuous from SiO_2 up to 35.5 mole percent Li_2O , 57.8 mole percent Na_2O and 54.5 mole percent K_2O respectively.

CaO can also be introduced into an alkali silicate glass and the effect on the structure is believed to be qualitatively the same as that of introducing Na_2O . However, such additions improve the durability of the glass, its thermal coefficient is reduced and the viscosity of the melt increased. These show that the coherence of the structure has improved. The glass forming regions in various such ternary silicate systems are summarized³³ in Fig. 9.

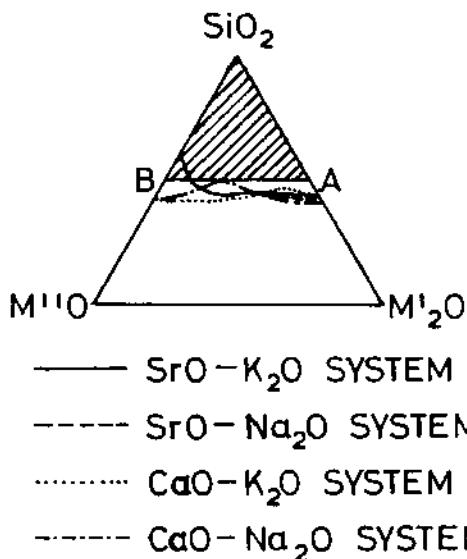


Fig. 9. Glass-formation regions in various ternary silicate systems $\text{SiO}_2-\text{M}''\text{O}-\text{M}'_2\text{O}$. M'' is a divalent and M' a monovalent element (Imaoka, 1962).

In alkali aluminosilicate glasses some of the silicon ions are replaced isomorphously by aluminium ions, the local neutralization

of charge being maintained by the presence of the univalent alkali ions. The latter are held in voids in the framework formed by the SiO_4 and AlO_4 tetrahedra. It has been suggested³⁴ that when $\text{R}_2\text{O}/\text{Al}_2\text{O}_3 > 1$, all the aluminium ions are in four-fold co-ordination and for $\text{R}_2\text{O}/\text{Al}_2\text{O}_3 < 1$, some of the aluminium ions are in six fold co-ordination. The alkali aluminosilicate glasses have high viscosities and it is difficult to devitrify them.

Glasses of lower SiO_2 content (as low as 40 mole percent) can be made in some multicomponent glasses (e.g., $\text{Na}_2\text{O}-\text{MgO}-\text{SiO}_2$) than in the binary $\text{Na}_2\text{O}-\text{SiO}_2$ system³⁵⁻³⁷. Trap and Stevels suggest an inversion in the structural role of modifying ions at the 50 mole percent SiO_2 composition. Thus, according to these authors above 50 mole percent SiO_2 , the modifying ions destroy the continuity of the silica network whereas below 50 percent the random structure of glass is stabilised by the ionic bonds acting between the modifier cations and the small groups of SiO_4 tetrahedra. Such glasses are therefore, referred to as "Invert glasses".

Borate Glasses. Boric oxide glass does not have much technological importance. Its main use is in the manufacture of Borosilicate glasses which form excellent heat resisting materials. As mentioned in Section I, it is very difficult to devitrify boric oxide glasses. Boric oxide form glasses with several other oxides. The range of glass formation in different binary borate systems, is summarized in Table III.1. It is seen from this table that only in systems containing K_2O , Na_2O , Li_2O and Tl_2O can glasses be made over a composition range which is continuous from 100% B_2O_3 to that containing maximum possible concentration of the modifying oxide. In all the other systems the difficulty arises due to the existence of two immiscible liquids of differing composition in the melt at temperatures just above the liquids temperature³³.

Alkaline earth aluminoborate glasses have excellent electrical properties^{38,39}. Extensive regions of glass formation have been reported in systems containing an oxide of a divalent metal, B_2O_3 and Al_2O_3 ⁴⁰. Maximum glass formation region is found in system $\text{BaO}-\text{Al}_2\text{O}_3-\text{B}_2\text{O}_3$, the results being shown in Fig. 10. Glass forming region shrinks with decreasing size of the alkaline earth cation and is very limited in the MgO system.

Phosphate Glasses. Glasses based on P_2O_5 do not have much technical importance. P_2O_5 forms glasses with a number of other oxides³³. Table III.2 shows the glass-formation regions in binary phosphate systems. X-ray diffraction work on phosphate glasses⁴¹ shows their structures to be similar to those of binary silicate glasses viz., both are based on tetrahedral units (PO_4 and SiO_4) and the addition of modifying oxides results in an increasing number of "non-bridging" oxygen atoms.

TABLE III.1
Ranges of Glass Formation in Binary Borate Systems³³

System	Range (Mole per cent B_2O_3)
K_2O	100 - 62.5
Na_2O	100 - 62.0 33.5 - 28.5
Li_2O	100 - 57.3
BaO	83.0 - 60.2
SrO	75.8 - 57.0
CaO	72.9 - 58.9
MgO	57.0 - 55.8
La_2O_3	81.0 - 71.8
Tl_2O	100.0 - 55.5
ZnO	56.0 - 36.4
CdO	60.9 - 45.0
PbO	80.0 - 23.5
Bi_2O_3	78.0 - 34.7

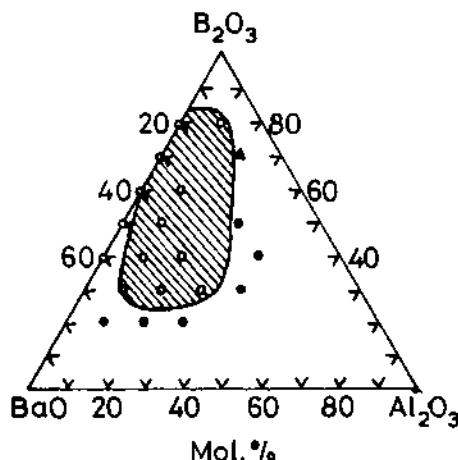


Fig. 10. Region of glass formation in system
 $BaO - Al_2O_3 - B_2O_3$, 0 clear glass;
immiscible; devitrified (Hirayama, 1961).

TABLE III.2
Regions of Glass Formation in Binary Phosphate Systems³³

Modifier	Maximum percentage of modifier (mole per cent)
K ₂ O	47
Na ₂ O	60
Li ₂ O	60
BaO	58
SrO	56
CaO	56
MgO	60
BeO	66
Ag ₂ O	66
Tl ₂ O	50
ZnO	64
CdO	57
PbO	62

Germanate Glasses: Glass formation regions in different binary systems are summarized³³ in Table III.5.

Tellurite and Vanadate Glasses: As mentioned in Section I TeO₂ and V₂O₅ are "conditional glass formers". Tellurite glasses have high refractive indices because of the high polarizability of the tellurium ion and hence they are suitable for use in optical components. Table III.4 gives the glass forming regions in the binary tellurite systems³³. X-ray diffraction^{42,43} as well as infrared studies⁴⁴ on some tellurite glasses indicate that tellurium atoms are surrounded by six oxygens. It has been suggested^{20,42,43} that the addition of a modifying oxide reduces the fraction of octahedra which share edges, thus reducing the rigidity of the structure.

Table III.5 shows the glass forming regions of binary vanadate systems⁴⁵. They have unusually low electrical resistivity for glasses having the highest V₂O₅ content.

Nitrate and Halide Glasses: Glass formation occurs only in systems containing two or more nitrates. Thilo et al.⁴⁶ have carried out the most comprehensive survey of glass formation in

TABLE III.3

Regions of Glass Formation in Binary Germanate Systems³³

System	Region of Glass Formation (Mole per cent)
K ₂ O - GeO ₂	0 - 59.5% K ₂ O
Na ₂ O - GeO ₂	0 - 38% Na ₂ O
Li ₂ O - GeO ₂	0 - 23.8% Li ₂ O
BaO - GeO ₂	0 - 10% and 17.5 - 29.6% BaO
SrO - GeO ₂	14 - 39% SrO
CaO - GeO ₂	15.5 - 35.5% CaO
Tl ₂ O - GeO ₂	0 - 47.5% Tl ₂ O
ZnO - GeO ₂	0 - 48% ZnO
PbO - GeO ₂	0 - 57% PbO
Bi ₂ O ₃ - GeO ₂	0 - 34% Bi ₂ O ₃

TABLE III.4

Regions of Glass Formation in Binary Tellurite Systems³³

System	Range of Glass Formation (Mole per cent of Second Oxide)
K ₂ O - TeO ₂	4.5 - 23.0
Na ₂ O - TeO ₂	8.5 - 40.5
Li ₂ O - TeO ₂	13.0 - 30.5
BaO - TeO ₂	7.0 - 20.0 and 36.0 - 36.5
SrO - TeO ₂	10.0 - 13.0
CaO - TeO ₂	-
MgO - TeO ₂	12.0 - 34.5
BeO - TeO ₂	11.0 - 23.0
La ₂ O ₃ - TeO ₂	4.5 - 9.5
Al ₂ O ₃ - TeO ₂	4.5 - 15.5

Table III.4 Contd.

System	Range of Glass Formation (Mole per cent of Second Oxide)
$\text{ThO}_2 - \text{TeO}_2$	8.5 - 10.0
$\text{TiO}_2 - \text{TeO}_2$	-
$\text{Ta}_{2\text{O}}_5 - \text{TeO}_2$	1.5 - 2.0
$\text{Nb}_{2\text{O}}_5 - \text{TeO}_2$	2.0 - 25.0
$\text{WO}_3 - \text{TeO}_2$	6.0 - 35.0
$\text{Tl}_{2\text{O}} - \text{TeO}_2$	13.5 - 37.5
$\text{In}_{2\text{O}}_3 - \text{TeO}_2$	6.5 - 14.0
$\text{ZnO} - \text{TeO}_2$	9.5 - 39.5
$\text{CdO} - \text{TeO}_2$	-
$\text{PbO} - \text{TeO}_2$	10.5 - 22.0

TABLE III.5

Regions of Glass Formation in Binary Vanadate Systems⁴⁵

System	Mole per cent $\text{V}_{2\text{O}}_5$	Crucible Used
$\text{P}_{2\text{O}}_5 - \text{V}_{2\text{O}}_5$	94	Platinum
$\text{GeO}_2 - \text{V}_{2\text{O}}_5$	6 - 63.5	Platinum
$\text{TeO}_2 - \text{V}_{2\text{O}}_5$	9 - 57	Silica and Alumina
$\text{As}_{2\text{O}}_3 - \text{V}_{2\text{O}}_5$	52.5	Silica and Alumina
$\text{BaO} - \text{V}_{2\text{O}}_5$	58.5 - 69.5	Platinum
$\text{PbO} - \text{V}_{2\text{O}}_5$	51 - 66.5	Silica and Alumina

binary nitrate systems. The results are given in Table III.6. Most of the investigations have however centered around the system $\text{KNO}_3 - \text{Ca}(\text{NO}_3)_2$. It has been suggested⁴⁷ using the ideas of the Zachariasen network theory that $\text{Ca}(\text{NO}_3)_2$ is the network former having a tetrahedral co-ordination of $(\text{NO}_3)^-$ ions and KNO_3 is the network modifier.

TABLE III.6

Regions of Glass Formation in Binary Nitrate Systems⁴⁶

System	Glass-forming Region (Mole per cent Alkali Nitrate)
Mg(NO ₃) ₂ - LiNO ₃	36 - 52
Mg(NO ₃) ₂ - NaNO ₃	21 - 57
Mg(NO ₃) ₂ - KNO ₃	24 - 61
Mg(NO ₃) ₂ - RbNO ₃	40 - 67
Mg(NO ₃) ₂ - CsNO ₃	47 - 75
Ca(NO ₃) ₂ - LiNO ₃	-
Ca(NO ₃) ₂ - NaNO ₃	55 - 62
Ca(NO ₃) ₂ - KNO ₃	40 - 70
Ca(NO ₃) ₂ - RbNO ₃	32 - 75
Ca(NO ₃) ₂ - CsNO ₃	64 - 75
Sr(NO ₃) ₂ - LiNO ₃	-
Sr(NO ₃) ₂ - NaNO ₃	-
Sr(NO ₃) ₂ - KNO ₃	-
Sr(NO ₃) ₂ - RbNO ₃	69 - 83
Sr(NO ₃) ₂ - CsNO ₃	68 - 77
Ba(NO ₃) ₂ - CsNO ₃	78 - 79
Cd(NO ₃) ₂ - LiNO ₃	36 - 63
Cd(NO ₃) ₂ - NaNO ₃	28 - 64
Cd(NO ₃) ₂ - KNO ₃	21 - 67
Cd(NO ₃) ₂ - RbNO ₃	34 - 71
Cd(NO ₃) ₂ - CsNO ₃	27 - 75

Of the glass forming halides BeF₂ and ZnCl₂, the former give more stable glasses. The glass formation is explained in terms of the low melting point of BeF₂ (540°C) in relation to the bond strength³. The crystalline forms of BeF₂ have three-dimensional network structures in which BeF₄ tetrahedra are linked at the

corners. By x-ray work it has been shown⁴⁸ that glass has a similar but disordered structure. Glass forming ability of $ZnCl_2$ is explained on identifical basis³.

Chalcogenide Glasses: Sulphur and selenium form glasses when they are cooled from the molten state^{49,50}. These are essentially long chain molecules held together by predominantly Van der Waals' bonds. Glasses having useful properties may be made by combining the chalcogen elements sulphur, selenium and tellurium with some of the Group IV and Group V elements of the periodic table e.g., Ge, Si, As, Sb, Bi³. Vaipolin and Porai-Koshits^{51,52} have carried out x-ray studies of vitreous As_2S_3 , As_2Se_3 and As_2Te_3 and a number of binary glass compositions based on these compounds. Their results indicate corrugated or puckered layer structures for these glasses.

IV. TRANSPORT PROPERTIES OF GLASS

Diffusion in Glasses: Diffusion studies in glasses are quite important in understanding some of their physical properties. Thus, electrical properties, resistance to corrosion, permeability to gases, rates of devitrification etc. are markedly dependent upon the diffusion of substances in the glass.

Diffusion of matter through any medium is governed by the following equations (Fick's Laws):

$$J = -D \text{ grad } C \quad (4)$$

$$\text{and} \quad \frac{\partial C}{\partial t} = \text{div} (D \text{ grad } C) \quad (5)$$

where J is flux of a certain species flowing through unit area per unit time, and D the diffusion coefficient.

Solutions of these equations will depend on the boundary conditions present in a particular problem⁵³. Study of diffusion phenomena in glasses like in other solids, involve the determination of the diffusion coefficient D . The different experimental techniques used in the case of glasses for such investigations are listed below⁵⁴. (i) Gas permeation: This involves measurement of gas transport through a thin glass membrane (in the form of tubes or bulbs). (ii) Gaseous exchange: The rate of exchange of a substance between the glass and gas is studied by measuring the concentration of diffusing material in either phase. (iii) Analysis of diffusing material: The concentration profile of the diffusing species in glass is measured as a function of time and distance within the solid. (iv) Resistance change during diffusion: If the diffusing species changes the electrical

resistance of glass, the progress of diffusion can be followed by measuring the electrical resistance. (v) Surface reaction: This is based on the fact that a glass layer might form on the surface of a material by a chemical reaction the rate of which is controlled by the diffusion of a component through the layer itself. (vi) Moving boundary: If by the diffusion of a substance in glass the transparency or colour of the latter is changed, the progress of diffusion can be studied by following visually the boundary between the altered and unaltered regions. (vii) Diffusion-controlled precipitation: Here the precipitation of a substance from glass is controlled by diffusion. (viii) Measurements on liquid Silicates: For glasses in the non-rigid condition, techniques developed for liquids are employed viz., capillary method, growth or contraction of a gas bubble in a liquid etc.⁵⁴

The temperature variation of diffusion coefficient D is usually assumed to be of the form⁵⁴:

$$D = D_0 \exp \left(-\frac{Q}{RT} \right) \quad (6)$$

where, D_0 is the pre-exponential factor, Q the activation energy for diffusion, R the gas constant, and T the absolute temperature. The results of diffusion studies are generally reported in terms of D_0 and Q. Depending on the nature of the substance diffusing in glass, the diffusion could be molecular or ionic⁵⁴.

Table IV.1 gives some data for molecular diffusion in glass. These results show that the activation energy for such diffusion increases with an increase in the molecular size. This is to be expected since the mechanical energy of distortion of the glass network associated with the migration of molecules should increase as the size of the molecule goes up.

Table IV.2 gives data for ionic diffusion in glass. The diffusion coefficients of ions fit equation (6) reasonably well just like those of molecules. As the amount of alkali in a glass is increased the diffusion coefficient of the alkali ions increases. This is because of the more open structure in glass brought about by addition of the univalent and/or divalent ions.

Electrical Conductivity of Glasses. Ionic Conductivity: In glasses containing modifier alkali ions the d.c. conductivity has been shown to be ionic owing to the movement of the alkali ions through the rigid oxide network⁵⁶. The dependence of the resistivity ρ on temperature up to about 600°C can be represented by⁵⁷

$$\log_e \rho = A + \frac{B}{T} \quad (7)$$

TABLE IV.1
Data for Molecular Diffusion in Glass

Element, its molecular diameter, and glass	Q Kcal./mole	D_0 $\text{cm}^2/\text{second}$	Temperature range °C	Measure- ment Technique	Refer- ences
Helium 2.0 Å Fused Silica	6.6	7.4×10^{-4}	300°-1034°	Gas permeation	55
Pyrex	6.5	4.8×10^{-4}	25°-350°		56
Neon 2.4 Å Fused Silica	11.4	2.2×10^{-4}	440°-980°		57
Hydrogen 2.5 Å Fused Silica	11.3	6.5×10^{-4}	200°-1000°		58, 59
Gold 2.9 Å Soda-alumina	56	3.5×10^{-5}	400°-530°	Diffusion-controlled precipitation	54
Oxygen 3.2 Å Borosilicate	32	0.3	1000°-1300°	Capillary method	60
Soda-lime	53	42	1000°-1300°		60
Alkali-alumina-Baria	24	5.4×10^{-3}	1000°-1300°		60
Argon 3.2 Å Potash-lime	42	350	308°-470°	Radioactive isotope	61

TABLE IV.2
Data for Ionic Diffusion in Glass

Ion and Glass	Q Kcal./mole	D_0 $\text{cm}^2/\text{second}$	Temperature Range °C	Measure- ment Technique	Refer- ences
Silver, Soda-lime-silica	24.3	7×10^{-2}	238°-422°	Weight change method of analysis of diffusing material	62

Table IV.2 contd.

Ion and Glass	Q Kcal/mole	D_0 $\text{cm}^2/\text{second}$	Temperature Range °C	Measure- ment Technique	Refe- rences
Sodium, Sodium Silicate					
(i) 11.4 wt.% Na_2O	14 22	5.5×10^{-5} 2.8×10^{-2}	350°-500° 350°-800°	Radioactive isotope	63
(ii) 16.6 wt.% Na_2O	14 22	1.1×10^{-4} 6×10^{-2}	350°-500° 550°-800°		
(iii) 25.4 wt.% Na_2O	14 22	1.5×10^{-4} 1.2×10^{-1}	300°-400° 450°-650°		
Lithium, Soda-lime-silica	25.9	1.1	150°-300°	Analysis of diffusing material by flame photo- meter	64
Potassium, Potassium Silicate					
(i) 20.4 wt.% K_2O	14 23	1.3×10^{-5} 8×10^{-3}	425°-550° 550°-800°	Radioactive isotope	63
(ii) 27.7 wt.% K_2O	14 23	2.2×10^{-5} 2.1×10^{-2}	425°-550° 550°-800°		
Calcium, Calcium-Alumina- Silicate					
10 wt.% Al_2O_3					
(i) 27.5 wt.% CaO	28	10^{-3}	1350°-1550°	Diffusion couple method	65

where A and B are constants and T is the temperature in °K. The constant B is usually interpreted in terms of an activation energy ϕ , where $B = \phi/K$, K being Boltzmann's constant. ϕ varies from 0.6 to 1.5 eV depending on the composition⁶⁷.

Stevels⁶⁷ has proposed a model for ionic conductivity in glasses in which the mobile ion is assumed to move between interstices in the network. On this model the expression for resistivity is of the form,

$$\rho = \frac{6 k T}{n \nu s d^2 e^2} \exp(\phi/kT) \quad (8)$$

where n is the number of mobile ions/c.c., e the charge of the ion, k Boltzmann constant, ν the frequency of vibration of ion in the interstice, d the average distance between interstices and s the number of neighbouring interstices.

Isard⁶⁸ has given a model for the activation energy of charge migration in soda glasses which is closely akin to the migration of interstitial ions in crystals (Frenkel defects). Here, the activation energy is given by,

$$\phi = \frac{1}{2} W + V \quad (9)$$

where W is the formation energy of the defect and V is the energy barrier to migration. V can be assumed to be small compared to W in soda glass because of the openness of the glass network. The value of W is derived in this model by assuming that the alkali ion is surrounded by a shell of oxygen ions and that the charge of the nonbridging oxygen ion is uniformly distributed over a spherical shell of the same radius as the ion shell. Thus, the activation energy is given by,

$$\phi = \frac{e^2}{2 r \epsilon} \quad (10)$$

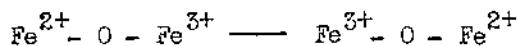
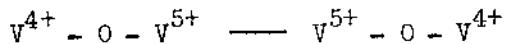
where r is the radius of the shell and ϵ is the dielectric constant (taken equal to 4 for silica glass). Study of dielectric properties of glasses under very high pressure⁶⁹ tends to confirm the validity of the above model.

Electronic Conductivity: Electronic conduction in glass was first reported⁷⁰ in the system $P_2O_5 - V_2O_5$. Most of the "semiconducting" glasses observed to date are either V_2O_5 or Fe_3O_4 - based⁷¹⁻⁷⁹.

To establish the fact that conductivity in any glass system is due to electronic migration the following types of measurements are usually carried out⁸⁰. (i) Electrolysis: The electrical resistance is measured as a function of time. Resistance of an ionically conducting glass increases whereas that of a

"semiconducting" glass remains constant. (ii) E.M.F. Measurement: For a pure ionic conductor an e.m.f. will be developed across the sample if its ends are in contact with atmospheres of different oxygen partial pressures. No such e.m.f. is generated for a pure electronic conductor. (iii) Magnitude of Activation Energy: The activation energies of charge migration in semiconducting glasses are less than those in the corresponding ionically conducting glass systems. Thus, Table IV.3 gives some data where such values are compared in GeO_2 and B_2O_3 based glasses with and without the transition metal oxides which bring about the electronic conduction. (iv) Effect of Pressure: The pressure coefficient of resistivity is negative in semiconducting glasses⁸⁰ whereas opposite is true in ionically conducting systems⁶⁹.

Semiconducting oxide glasses can be regarded as the concentrated solution of one or more semiconducting oxide in an insulating glassy solvent. Electrical conduction in such media can therefore be visualized as arising due to transfer of an electron or hole between ions of the same transition metal in different valence states. e.g.,



The possibility of a Polaron mechanism of charge transport in semiconducting glasses has been investigated recently⁸².

A large amount of work has been done in U.S.S.R. on Chalcogenide semiconducting glasses⁶³. The systems which have been studied are As_2Se_3 - As_2Te_3 ⁸³, As-Te-I⁸³, As-Ge-Pb-Se⁸⁴. The conduction in these glasses is found to be intrinsic and not impurity controlled.

Dielectric Properties of Glasses: Like any other insulating material, the dielectric characteristics of glass can also be represented by a complex permittivity⁸⁵, , such that

$$\epsilon = \epsilon' - j \epsilon'' \quad (11)$$

where, ϵ' and ϵ'' are the real and imaginary parts respectively and are related to the dissipation factor $\tan \delta$ as

$$\tan \delta = \frac{\epsilon''}{\epsilon'} \quad (12)$$

Stevens⁶⁷ has classified the different mechanisms of dielectric loss in glass as follows: (i) Conduction loss, (ii) Dipole relaxation loss, (iii) Deformation loss and (iv) Vibration loss.

TABLE IV.3

Comparison of Activation Energies for Conduction of Semiconducting Oxide Glasses (E^*) and that of the Corresponding Glasses without the Semiconducting Component (E')⁸¹

Composition in mole per cent	E^* Kcal/mole	E' Kcal/mole
20 BaO - 60 GeO ₂ - 30 V ₂ O ₅	12.5	
40 BaO - 60 GeO ₂		38.0
31 CaO - 61 B ₂ O ₃ - 8 Fe ₂ O ₃	15.7	
40 CaO - 60 B ₂ O ₃		40.0
30 CaO - 60 B ₂ O ₃ - 10 CoO	24.0	
40 CaO - 60 B ₂ O ₃		40.0

(i) & (iii) arise due to movement of the alkali ions through the glass network, (iii) is due to small deformations in the network itself and (iv) arises out of a resonance phenomenon between the electric field applied and vibrations of all ions about their rest positions. For an ordinary soda glass at room temperature the conduction and dipole relaxation losses occur in the audio-frequency range whereas the deformation and vibration losses predominate at frequencies of the order of 10^{12} c/s.

Taylor⁸⁶⁻⁸⁸ measured the components of dielectric permittivity in the audio-frequency range and the d.c. resistivity at different temperatures for some soda-lime-silica glasses. He analysed the data on the basis of Debye equations⁸⁹.

$$\epsilon' = \epsilon_{\infty} + (\epsilon_s - \epsilon_{\infty}) \int_0^{\infty} \frac{y(\tau) d\tau}{1 + \omega^2 \tau^2} \quad (10)$$

$$\epsilon'' = (\epsilon_s - \epsilon_{\infty}) \int_0^{\infty} \frac{y(\tau) \omega \tau d\tau}{1 + \omega^2 \tau^2} \quad (11)$$

where $y(\tau)$ is the normalized distribution of relaxation times, ϵ_s , ϵ_{∞} the asymptotic values of ϵ' at low and high frequencies

respectively and ω is the angular frequency of the applied field. Taylor, however, found that the activation energy for the relaxation mechanism was in close agreement with that of d.c. resistivity.

Isard⁹⁰ has studied the dielectric loss spectra of soda glasses by using non-blocking electrodes and analysed the data on the basis of Maxwell-Wagner-Sillars heterogeneous dielectric model⁹¹⁻⁹³. Such analyses have been used extensively in studying the phase-separation characteristics in Lithium Silicate and sodium boro-silicate glasses⁹⁴⁻⁹⁶.

REFERENCES

1. Glass Industry, 26, 417 (1945).
2. B. E. Warren, J. Appl. Phys., 8, 645 (1937).
3. H. Rawson, Inorganic Glass-forming Systems, Academic Press, London & New York 1967.
4. R.C. Ellis, Inorg. Chem., 2, 22-24 (1963).
5. V.M. Goldschmidt, Skrifter Norske Videnskaps Akad (Oslo), I. Mathnaturwiss. Kl. No. 8, 7-156 (1926).
6. R.C. Evans, An Introduction to Crystal Chemistry, Cambridge University Press, Lond. (1948).
7. W.H. Zachariasen, J. Am. Chem. Soc., 54, 3841-3851 (1932).
8. J.E. Stanworth, Physical Properties of Glass, Oxford, 1950.
9. S. Urnes, In "Modern Aspects of the Vitreous State, Vol. I", (J.D. Mackenzie, ed.) pp. 10-37. Butterworths, London (1960).
10. N. Valenkov, and E.A. Porai-Koshits, Z. Kristallogr., 95, 195-229 (1936).
11. G. Hartlief, Z. Anorg. allg. Chem., 238, 353-384 (1958).
12. A. Smekal, J. Soc. Glass Technol., 35, 411-420 T (1951).
13. J.E. Stanworth, J. Soc. Glass Techn., 30, 54-64 (1946).
14. J.E. Stanworth, J. Soc. Glass Techn., 32, 154-172 (1948).
15. J.E. Stanworth, J. Soc. Glass Techn., 36, 217-241 (1952).

16. L. Pauling, "The Nature of the Chemical Bond", 3rd, Ed., 644 pp. Cornell Univ. Press, Ithaca (1960).
17. J.E. Stanworth, J. Soc. Glass Technol., 36, 217-241T (1952).
18. K.H. Sun, J. Am. Ceram. Soc., 30, 277-281 (1947).
19. K.H. Sun and M.L. Huggins, J. Phys. Colloid Chem., 51, 438-443 (1947).
20. H. Rawson, In "Proceedings IV International Congress on Glass, Paris", pp. 62-69, Imprimerie Chaix, Paris (1956).
21. J. Frenkel, "The Kinetic theory of liquids", Clarendon Press, Oxford, 1946.
22. L.A.K. Staveley, In "The Vitreous State", pp. 85-90, Glass Delegacy of the University of Sheffield (1955).
23. D. Turnbull and M.H. Cohen, J. Chem. Phys., 29, 1049-1059 (1958).
24. D. Turnbull and M.H. Cohen, in "Modern Aspects of Vitreous State", Vol. I (ed. by J.D. Mackenzie) (1960).
25. D. Turnbull and M.H. Cohen, Nature, 189, 131-132 (1961).
26. S.D. Brown and S.S. Kistler, J. Am. Ceram. Soc., 42, 263-270 (1959).
27. N.G. Ainslie, C.R. Morelock and D. Turnbull, In "Symposium on Nucleation and Crystallization in Glasses and melts," (M.K. Reser, G. Smith and H. Insley, eds.) pp. 97-107, American Ceramic Society, Columbus (1962).
28. F.E. Wagstaff, S.D. Brown and I.B. Cutler, Phys. Chem. Glasses, 5, 76-81 (1964).
29. J.M. Stevels, J. Soc. Glass Technol., 35, 284-288T (1951).
30. A. Dietzel and H. Wickert, Glastech. Ber., 29, 1-4 (1956).
31. S.D. Brown and S.S. Kistler, J. Am. Ceram. Soc., 42, 263-270 (1959).
32. M. Imaoka and T. Yamazaki, J. Ceram. Assn. Japan, 71, 215-223 (1963).
33. M. Imaoka, In "Advances in Glass Technology Part I", pp. 149-164, Plenum Press, New York (1962).

34. J.O. Isard, *J. Soc. Glass Technol.*, 43, 113-123T (1959).
35. H.L. Trap and J.M. Stevels, *Glastechn. Ber.*, 32K (VI), 51-52 (1959).
36. H.L. Trap and J.M. Stevels, *Phys. Chem. Glasses*, 1, 107-118 (1960).
37. H.L. Trap and J.M. Stevels, *Phys. Chem. Glasses*, 1, 181-188 (1960).
38. A.E. Owen, *Phys. Chem. Glasses*, 2, 87-98, 152-162 (1961).
39. A.E. Owen, *Phys. Chem. Glasses*, 3, 134-138 (1962).
40. C. Hirayama, *J. Am. Ceram. Soc.*, 44, 602-606 (1961).
41. J. Biscoe, A.G. Pincus, C.S. Smith and B.E. Warren, *J. Am. Ceram. Soc.*, 24, 116-119 (1941).
42. G.W. Brady, *J. Chem. Phys.*, 24, 477 (1956).
43. G.W. Brady, *J. Chem. Phys.*, 27, 300-303 (1957).
44. R.V. Adams, *Phys. Chem. Glasses*, 2, 101-110 (1961).
45. E.P. Denton, H. Rawson and J.E. Stanworth, *Nature*, 173, 1030-1032 (1954).
46. E. Thilo, C. Wiecker and W. Wiecker, *Silikattechnik*, 15, 109-111 (1964).
47. A. Dietzel and H.J. Poegel, In "Proc. 3rd. Intern. Congress on Glass, Venice", 219-243 (1953).
48. B.E. Warren and C.F. Hill, *Z. Kristallogr., Abt A.* 89, 481-486 (1934).
49. A.V. Tobolsky, G.D.T. Owen and A. Eisenberg, *J. Colloid Sci.*, 17, 717-725 (1962).
50. A. Eisenberg, *Polym. Letters*, 1, 177-179 (1963).
51. A. Vaipolin and E.A. Porai-Koshits, *Sov. Phys. Solid St.*, 2, 1500-1508 (1961).
52. A.A. Vaipolin and E.A. Porai-Koshits, *Sov. Phys. Solid St.*, 5, 178-185, 186-190, 497-500 (1963).

53. J. Crank, *The Mathematics of Diffusion*, Oxford University Press, London (1956).
54. R.H. Doremus, In "Modern Aspects of the Vitreous State Vol.2" (J.D. Mackenzie, ed.) pp. 2-10. Butterworths, London (1962).
55. D.E. Swets, R.W. Lee and R.C. Frank, *J. Chem. Phys.*, 34, 17 (1961).
56. W.A. Rogers, R.S. Buritz and D. Alpert, *J. Appl. Phys.*, 25, 868 (1954).
57. R.C. Frank, D.E. Swets and R.W. Lee, *J. Chem. Phys.*, 35, 1451 (1961).
58. G.A. Williams and J.B. Ferguson, *J. Am. Chem. Soc.*, 44, 2160 (1922).
59. R.M. Barrer, *J. Chem. Soc.*, 378 (1934).
60. R.H. Doremus, *J. Amer. Ceram. Soc.*, 43, 655 (1960).
61. M.B. Reynolds, *J. Am. Ceram. Soc.*, 40, 395 (1957).
62. G. Schulze, *Ann. Phys. Lpz.*, 40, 335 (1913).
63. J.R. Johnson, Thesis, Ohio State University (1950).
64. C.M. Hollabangh and F.M. Ernsberger, Presented at the October 1960 Meeting of the Am. Ceram. Soc.
65. N. McCallum and L.R. Barrett, *Trans. Brit. Ceram. Soc.*, 51, 523 (1952).
66. P.M. Sutton, In "Progress in Dielectrics, Vol. II" Ed. by J.B. Birks and J.H. Schulman (1960).
67. J.M. Stevels, In "Handbuch der Physik (Encyclopedia of Physics), Vol. XX", Ed. by S.W. Flugge, Springer-Verlag, Berlin, 1957.
68. J.O. Isard, *J. Soc. Glass Technol.*, 43 211 (1959).
69. D. Chakravorty and L.E. Cross, *J. Am. Ceram. Soc.*, 47 8 , 370 (1964).
70. E.P. Denton, H. Rawson and J.E. Stanworth, *Nature*, 173 (1954).
71. P.L. Baynton, H. Rawson and J.E. Stanworth, *J. Electrochem. Soc.*, 104, 237 (1957).

72. V.A. Ioffe, J.B. Patrina and I.S. Poberovshaya, Sov. Phys. Solid State, 2, 609 (1960).
73. M. Munakata, Solid State Electronics, 1, 159 (1960).
74. M. Munakata and M. Iwamoto, J. Ceram. Ass. Japan, 68, 125 (1960).
75. T.N. Kennedy and J.D. Mackenzie, Phys. Chem. Glasses, 8, 5 (1967).
76. Bh. V. Janakiramarao, J. Am. Ceram. Soc., 48 6 , 311 (1965).
77. J.D. MacKenzie, In "Physics of Non-crystalline Solids", (Proc. Intern. Conf., Deft. 1964) Ed. by J.A. Prins, John Wiley and Sons, Inc., N.Y. (1965).
78. K.W. Hansen, J. Electrochem. Soc., 112 10 , 994 (1965).
79. H.J.L. Trap, In "Physics of Non-Crystalline Solids", (Proc. Intern. Conf., Deft, 1964) Ed. by J.A. Prins, John Wiley and Sons, Inc., N.Y. (1965).
80. J.D. Mackenzie, In "Modern Aspects of the Vitreous State Vol. 3", Ed. by J.D. Mackenzie, Butterworths (Washington) (1964).
81. J.D. Mackenzie, J. Am. Ceram. Soc., 47, 211 (1964).
82. A.P. Schmid, J. Appl. Physics, 39 , 7 , 3140 (1968).
83. A.D. Pearson, In "Modern Aspects of the Vitreous State Vol.3", Ed. by J.D. Mackenzie, Butterworths (Washington) (1964).
84. V.A. Kharyuzov and A.M. Efimov, In "The Structure of Glass Vol. 4", Ed. by O.V. Mazurin (in English), Consultants Bureau, N.Y. (1965).
85. A.R. Von Hippel, "Dielectrics and Waves", John Wiley and Sons, Inc. N.Y. (1959).
86. H.E. Taylor, Trans. Farad. Soc., 52, 873 (1956).
87. H.E. Taylor, J. Soc. Glass Technol., 41, 350 (1957).
88. H.E. Taylor, J. Soc. Glass Technol., 43, 124 (1959).
89. A.J. Dekker, Solid State Physics, Prentice-Hall, Inc. N.J. (1961).

90. J.O. Isard, Proc. Inst. Elec. Engrs. (London), 109B, Suppl. No. 22 , 440 (1962).
91. J.C. Maxwell, Electricity and Magnetism, Vol. 1, p. 328, Clarendon Press, London, 1892.
92. K.W. Wagner, Arch. Elektrotech., 2 , 9 , 371 (1914).
93. R.W. Sillars, J. Inst. Elec. Engrs. (London), 80, 378 (1937).
94. R.J. Charles, J. Am. Ceram. Soc., 46 , 5 , 235 (1963).
95. R.J. Charles, J. Am. Ceram. Soc., 47 , 11 , 559 (1964).
96. D.L. Kinser and L.L. Hench, J. Am. Ceram. Soc., 51 , 8 , 445 (1968).

ELECTRICAL PROPERTIES OF SOLID CATALYSTS

M. V. C. Sastri, V. Srinivasan, and B. Viswanathan

Department of Chemistry

Indian Institute of Technology, Madras-36, India

I. HISTORICAL INTRODUCTION

Widely different kinds of solid materials, ranging from the noble metals to activated earths, have been used to catalyse a variety of chemical reactions. Thus, the use of nickel for hydrogenation, of iron for ammonia synthesis, of vanadium pentoxide for oxidation reactions, of chemically processed clay and synthetic silica-alumina matrix for cracking of petroleum, is well known. Specificity of catalytic action is one of the most outstanding properties of a catalyst, particularly a contact catalyst. For each reaction or reaction-type, there is a restricted group of substances which can catalyse it. As a result, catalysts came to be classified application-wise¹ as hydrogenation catalysts, oxidation catalysts, dehydration catalysts, cracking catalysts and so on. Within each group of catalysts, the activity (as defined by the accelerating effect on a given reaction) varied from substance to substance, so that the choice of the optimum catalyst for a given reaction still rested on empirical trial and error methods. Many attempts have been made, particularly during the last three decades, to rationalise the method of catalyst selection. Though we are still far from achieving the goal of a priori prescription of the optimum catalyst for a given application, it cannot be denied that the enormous research efforts of the past 30 years have resulted in a much deeper understanding of contact catalysis.

The first step in this comprehension was the realization that the catalytic action occurs entirely at the interface between the contact mass and the reaction medium and that it may invariably be analysed into three elementary steps (Fig. 1): First, adsorption

ADSORPTION - CONVERSION - DESORPTION

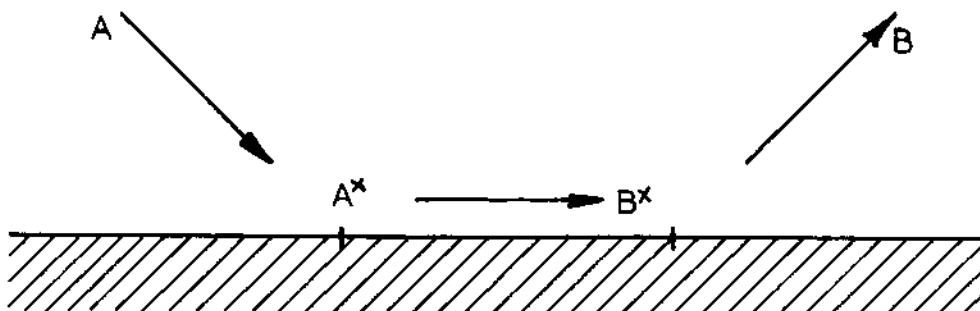


Fig. 1. Schematic diagram of successive processes in surface catalysed reaction.

of reactants, followed by reaction or transformation of the adsorbed species and finally the desorption of the products formed.

The next step was the identification of two kinds of adsorption, designated respectively as physical adsorption and chemical adsorption or chemisorption. The distinction between the two types lies in the nature of the forces of interaction involved. Physical adsorption is a phenomenon more or less akin to condensation of vapours and caused by van der Waals' forces of electrostatic origin. The forces responsible for chemisorption are of chemical nature, i.e. exchange forces causing the virtual formation of a chemical bond, ionic or covalent, between the adsorbed species and the surface atom or atoms of the solid. Wolkenstein² has expressed the theoretical distinction between the two types of adsorption very elegantly in the following words: "In the case of physical adsorption, the adsorbed molecule and the lattice of the adsorbent can be regarded as two independent systems. The action of the adsorbent on the adsorbate can be treated in this case as a slight perturbation and the problem can be solved by the methods of the perturbation theory. In the case of chemical adsorption, the molecule and the lattice form a single quantum-mechanical system and must be considered as a united whole. In this case,

the adsorption, as has already been pointed out, is a chemical combination of the molecule with the adsorbent."

In principle, chemisorption may be distinguished from physical adsorption by any of the experimental criteria that can be used to differentiate between chemical bonding and physical, van der Waals' cohesion. However, the heat of adsorption is the most common criterion employed. The heat liberated (reckoned per mole of gas adsorbed) in physical adsorption is of nearly the same order of magnitude as the heat of liquefaction, seldom exceeding the latter by more than a factor of two or three. During chemisorption, however, heats of considerably greater magnitude, close to the heats of formation of the corresponding compounds, are evolved.

It has been necessary to refer to the phenomenon of adsorption at some length, because it constitutes the essential first step in contact catalytic action and virtually holds the key to it. As a result of this adsorption, the reactant molecule is transformed into a chemically more reactive state. This can happen only with chemisorption and so, obviously, it is chemisorption that is directly concerned in the action sequence of contact catalysis. Chemical interaction between the reactant molecule and the solid catalyst implicates the electronic properties of both partners. The relevance of the electronic properties of the reactant molecule with regard to its reactivity is well known and lies outside the scope of the present discussion. It is our concern here to discuss the electronic properties of catalytic solids in so far as they are relevant to chemisorption and catalysis. Throughout this discussion it should be borne in mind that contact catalysis is a surface phenomenon and the electronic properties that are really relevant are those of the surface layers of the solid catalysts and not so much those of the interior of the lattice.

II. CLASSIFICATION OF SOLID CATALYSTS BASED ON ELECTRONIC PROPERTIES

Although the importance of the electronic factor in contact catalysis was recognised much earlier, it was only in 1950 that a classification of solid catalysts based on electronic properties was formally announced by Dowden². This classification sorted out catalysts into three distinct groups: (a) Conductors (metals); (b) Semiconductors (metal oxides and sulphides) and (c) Insulators (refractory oxides). The broad distinction in the electronic properties of the three classes of solids is best illustrated by means of the energy band diagrams shown in Fig. 2. This classification was the first concerted step to focus the lights of physics

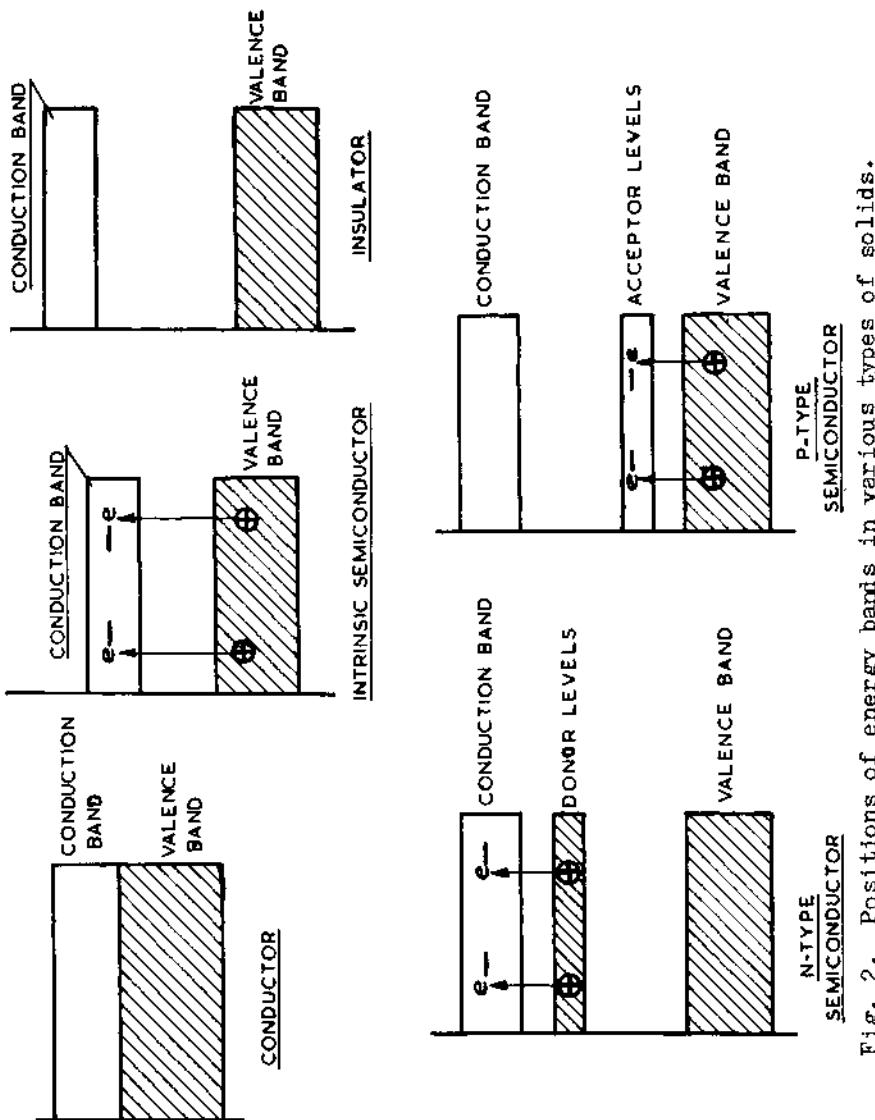


Fig. 2. Positions of energy bands in various types of solids.

on the chemical problem of catalysis.

It is obvious that electron mobility or electron transfer between the solid and the substrate can be expected only in the case of metals and semiconductors. Insulators catalyse certain types of reactions, such as dehydration of enols and isomerisation of hydrocarbons by proton-transfer or acid catalysis mechanism and will not be discussed further because no electronic properties of the solid are involved in their catalytic action. The following discussion will therefore be confined to catalytically active metals and semiconductors.

Metal Catalysts. Before entering into a discussion of the catalytic behaviour of metals, it is pertinent to refer briefly to the theory of metals. The two current theories are (1) the Band Theory due to Mott and Jones⁴ and (2) Pauling's Valence Bond Theory.⁵ The Band Theory is based essentially on the following assumptions. (i) The cohesive forces in a metal crystal are due to the interaction of the outermost s electrons alone. (ii) The electrons are referred to collectively without reference to individual atoms so that the electron wave function is confined simply to the metal crystal as a whole and not to the neighbourhood of individual atoms. However, the electrons in the crystal retain much of the character that they possess in the isolated metal atom in the sense that it is still permissible to speak of s, p, d, f, etc. electrons in a metal crystal as in the isolated atom.

The basic difference between the isolated atom and the metal crystal is that the allowed energy values in an atom are discrete and single valued, whereas in the crystal we have energy bands of permitted values. It is to be noted, however, that the number of electrons per atom in a band of the crystal could vary from that in the corresponding shell of the isolated atom. For example, the electronic configurations of the 3d and 4s orbitals of the isolated Fe and Ni atoms are 3d⁶ 4s² and 3d⁸ 4s² respectively, whereas in the metallic state, their band structures, on an average, are 3d^{7.8} 4s^{0.2} and 3d^{9.4} 4s^{0.6} respectively. It is thus clear that there is an incomplete d-band in the case of metallic crystals of both iron and nickel. Knowing that d⁸ and d¹⁰ configurations of Fe and Ni respectively to be discrete, in the terminology of the band theory, one refers to the existence of 0.2 and 0.6 'holes' respectively in the d-band of iron and nickel crystals. It has been observed that one of the important distinguishing features of transition metals is that their crystals possess such incomplete d-bands which in turn reflects on the high activity of these metals as catalysts.

In contrast to the Band Theory's notion of the electron being left at large, so to speak, in the metal crystal, the Valence Bond

Theory of Pauling visualises the electron as being localised around particular atoms or pairs of atoms in the crystal. Cohesion is ascribed to resonance between all possible structures in which definite one-electron bonds and two-electron bonds are formed between the atoms constituting the crystal. With regard to the transition metals, which are important catalytically, the theory states that the s, p and d electrons of these metals are considered to take part in bonding, the overlapping orbitals resulting in the bonding being spd or dsp hybrids. An important feature of the theory is "the percentage d character" (δ)⁶ which is taken as an index of the bond strength and depends upon the number of available bonding orbitals in the various possible valence states for the metal in question. A decrease in percentage d-character indicates increasing availability of atomic d-orbitals for surface bonding. Value of percentage d character, calculated by Pauling for some of the transition metals, are given in Table II.1. In addition to

Table II.1

Percentage d-Character in the Metallic Bond of Transition Metals

Metal	Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu
Percentage <u>d</u> -character	20	27	35	39	40.1	39.5	39.7	40	36
Metal	Y	Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag
Percentage <u>d</u> -character	19	31	39	43	46	50	50	46	36
Metal	La	Hf	Ta	W	Rh	Os	Ir	Pt	Au
Percentage <u>d</u> -character	19	29	39	43	46	49	49	44	--

these "bonding orbitals", "non-bonding atomic d-orbitals", responsible for the magnetic properties of the metal, and "metallic d-orbitals", responsible for electrical conduction, have also been postulated in Pauling's theory.

Mechanism of Chemisorption by Metals. As stated earlier, chemisorption occurs by electron-sharing or electron-transfer between the metal surface and the adsorbed molecule or atom. An

electron transfer from the adsorbate to the metal or vice-versa would be expected to occur depending upon whether the highest occupied electron level of the adsorbate lies above or below the Fermi level of the metal. This is schematically represented in Fig. 3. A high work function (Φ) of the metal and a low ionisation potential (I) of the adsorbate favours the formation of a positively charged (cationic) chemisorbed species on the metal (Fig. 3a). The converse, i.e. negative ion formation would be expected to occur if the metal possesses a low work function and the adsorbate has a high electron affinity (α) as shown in Fig. 3b.

Covalent chemisorption bonding, where the chemisorbed phase is stabilised by electron exchange, is explained on the basis of the Band Theory as follows (Fig. 3b): A metal electron in a filled level, the energy of which is the same as that of the electron of the adsorbate, is promoted to the Fermi level; the vacancy thus created in the metal levels interacts with the electron of the adsorbate to give rise to a homopolar band. Such covalent bond formation will therefore be favoured for a given adsorbate if the metal work function is large or, alternatively, if the ionisation potential of the adsorbate is low. The situation

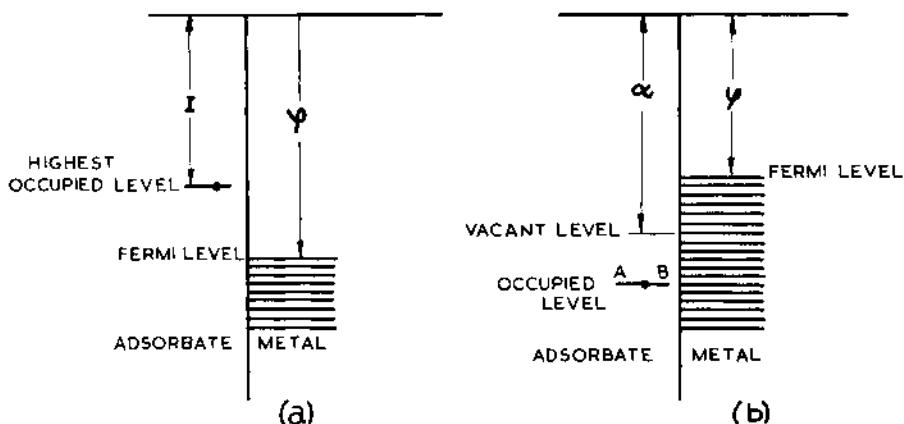


Fig. 3. (a) Chemisorption on a metal with positive ion formation (cationic chemisorption); I, ionisation potential; Φ , Work function of the metal
 (b) Chemisorption on a metal with negative ion formation (anionic chemisorption); α , electron affinity of the adsorbate.

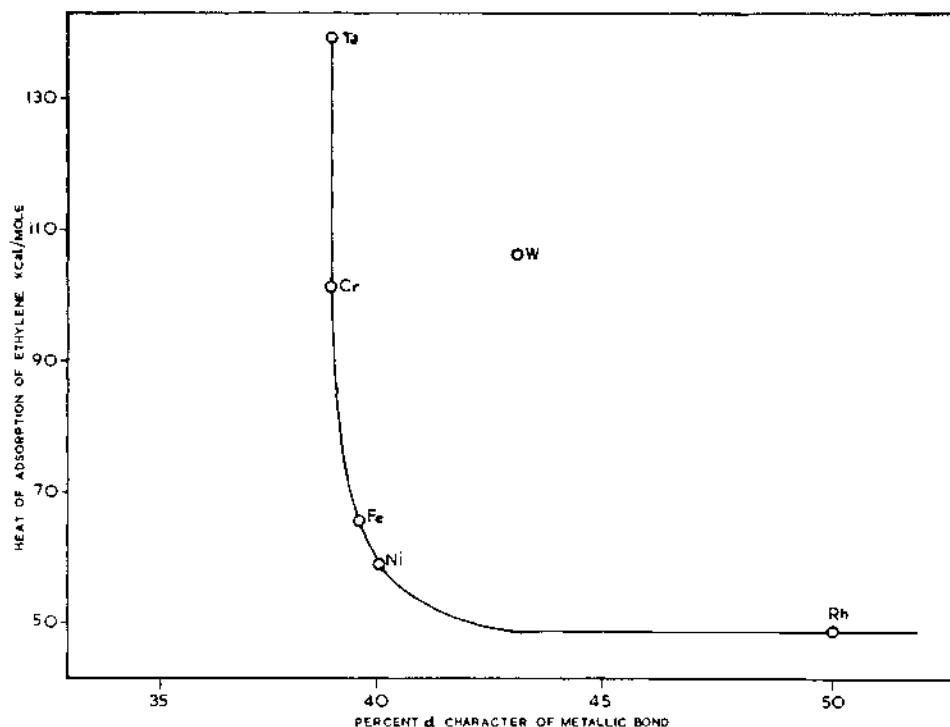


Fig. 4. Heat of chemisorption of ethylene as a function of the % d-character of the metal (according to Beeck⁷).

is explained in a simpler manner on Valence Bond model, according to which covalent bond formation occurs by the sharing of the electron of the adsorbed molecule with an available unpaired electron in the atomic d-orbital of the metal. The foregoing explains why high activity for chemisorption and catalysis is prevalent almost exclusively among the transition metals, because it is these metals that possess the vacancies or 'holes' in the d-band or the vacant atomic d-orbitals so essential for chemisorption bonding.

These ideas are supported by the observed correlation between the "percentage d-character" and the heat of chemisorption⁷. The greater the d-character, the smaller will be the availability of atomic d-orbitals for chemical bonding and hence weaker the chemisorption bond, resulting in a low heat of chemisorption. This is clear from Fig. 4, showing the heats of chemisorption of ethylene on a series of transition metals.

It should be noted in this context that though the chemisorp-

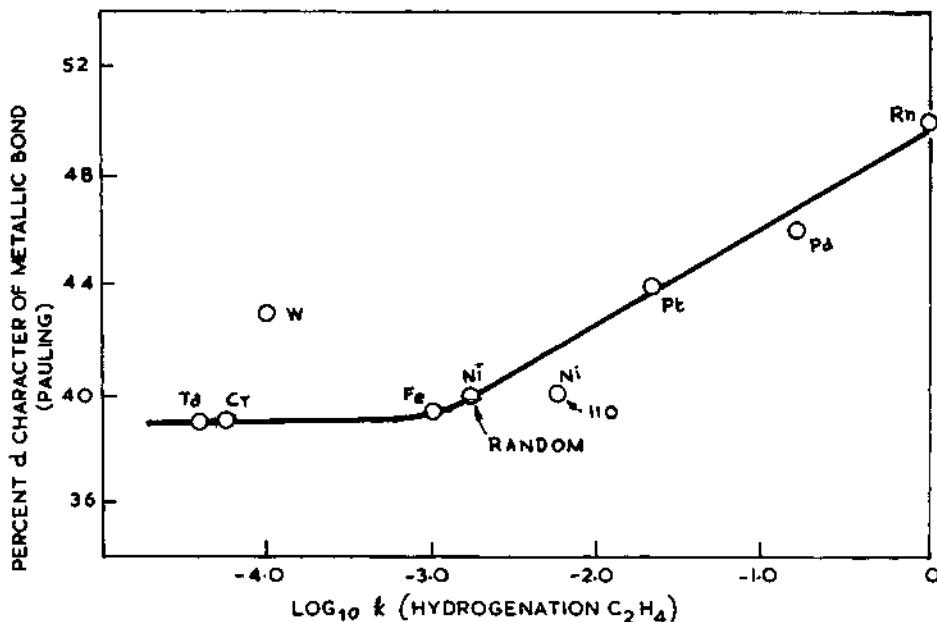


Fig. 5. Catalytic activity as a function of % d-character (according to Beeck⁷).

tion of a reactant is vital to catalysis, the chemisorbed species should not be bound too strongly to the catalyst surface. In fact, there is an inverse correlation between the heat of chemisorption and catalytic activity. What is expected of the catalyst is that it should hold the adsorbed reactant in a reactive and labile condition. Consequently, the catalytic activity would increase with increasing d-character, as shown in Fig. 5 for the case of ethylene hydrogenation; a reaction which has been well studied on clean metal surfaces⁷. An even more spectacular correlation (Fig. 6) was obtained by Schuit⁸ by plotting the logarithm of the reaction rate against the product of the valence of the metal (according to the Pauling Theory) and the percentage d-character.

Experimental Methods for the Study of Chemisorption of Gases by Metals. A variety of novel methods have been developed in recent years to study the electronic consequences of the chemisorption of a gas on a metal surface. Two types of measurements usually made are: (a) work function measurements and (b) measurements of electrical resistance. In both cases, the surfaces of the metals are specially prepared and cleaned by high vacuum techniques to ensure absolute freedom from surface contaminations.

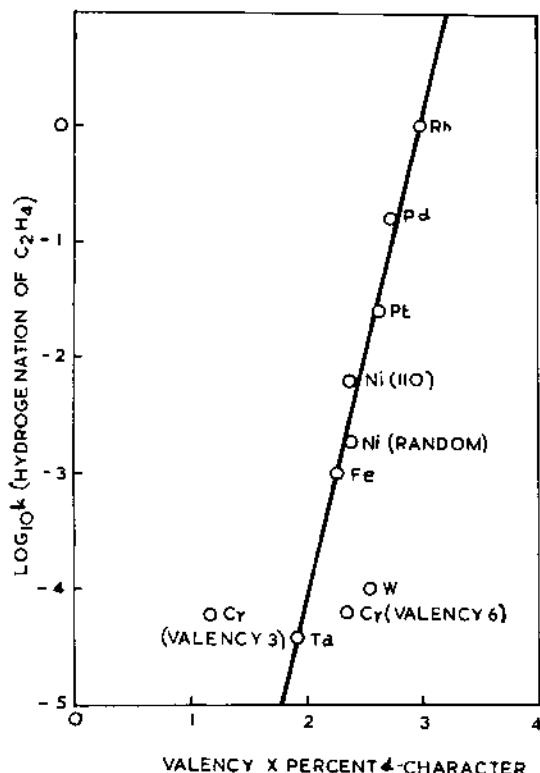


Fig. 6. Catalytic activity as a function of the product of valence and % d -character (according to Schuit⁸).

Work function measurements: The work function ($e\Phi$) is defined as the energy required to remove an electron from its highest occupied level within the solid into vacuum outside. A species having an ionic character in the adsorbed state will obviously modify the value of the work function. The work function of the metal would be either decreased or increased depending upon whether the adsorbed species is charged positively or negatively with respect to the metal surface. One can visualise the chemisorption bonds as a series of dipoles on the metal surface. Thus the adsorbed layer could be regarded as a parallel plate condenser. The apparent dipole moment (μ) of the metal-adsorbate bond is then related to the potential energy drop (ΔV) across this layer by the Helmholtz equation, $\Delta V = 4\pi C_s \theta \mu$, where C_s is the number of elementary sites per cm^2 of the surface and θ is the fractional coverage. This potential drop across the adsorbed layer is equal to $\Delta\Phi$, the change in work function of the metal produced by adsorption. The greatest significance of the above equation is that it enables one to calculate the dipole moment (μ) of the

adsorption bond from measurements of surface potential. The work function and changes occurring therein can be measured directly by photoelectric, thermionic or field emission methods or indirectly by the determination of the contact potential of the metal under study with reference to an uncharged electrode surface. Details of these methods are available in a number of texts^{9,10} that have appeared recently on the subject.

Electrical resistance measurements: Measurement of the electrical resistance of the metal surface gives more direct information concerning the transfer of electrons from the metal surface to the adsorbed gas or vice versa. To be meaningful, such studies should be carried out only with thin films of the metal (prepared by evaporation in high vacuum) so as to minimise interfering bulk conductivity effects. For details, reference may be made to a review by Suhrmann¹¹ who has made excellent use of this technique to elucidate the nature of the bond formed between gases like oxygen, hydrogen, nitrogen, carbon monoxide, nitrous oxide etc. and a variety of metal surfaces like nickel, platinum, etc.

Semiconductor Catalysts. Semiconductors may be classified as intrinsic semiconductors and extrinsic (or defect) semiconductors. Intrinsic semiconductors could again be the elemental type such as germanium and silicon or stoichiometric oxides like Fe_3O_4 , Co_2O_3 and Cr_2O_3 . Elemental semiconductors have rarely been used as catalysts. Only a limited study on a typical reaction like the decomposition of nitrous oxide on elemental germanium surfaces has been made¹². From the standpoint of widespread application as catalysts, extrinsic semiconductors, which owe their electrical conductance to defect lattice structure or non-stoichiometry or both, are of absorbing interest. Non-stoichiometry may arise out of the incorporation of extra atoms into the crystal at interstitial sites or due to the vacancies caused by the absence of atoms from normal sites. For example, the excess Zn atoms in ZnO are placed mostly in interstitial positions and act as typical donors giving rise to n-type conductivity of the oxide whereas in the case of NiO , the Ni^{+2} vacancies, each associated with two Ni^{+3} to preserve electrical neutrality, provide typical acceptor levels leading to p-type conductivity. While the absence of an atom from the normal structural site is termed a vacancy, the absence of an electron from a normal band or bond orbital is denoted as a 'hole'.

Chemisorption by Semiconductors. In most cases, adsorption at the surface of a semiconductor leads to the formation of a positively or negatively charged adsorbed layer. The term 'lonosorption' has therefore been suggested to describe such adsorption. The process of donation or acceptance of electrons at a gas-semiconductor interface resembles in many respects that occurring at a metal-semi-

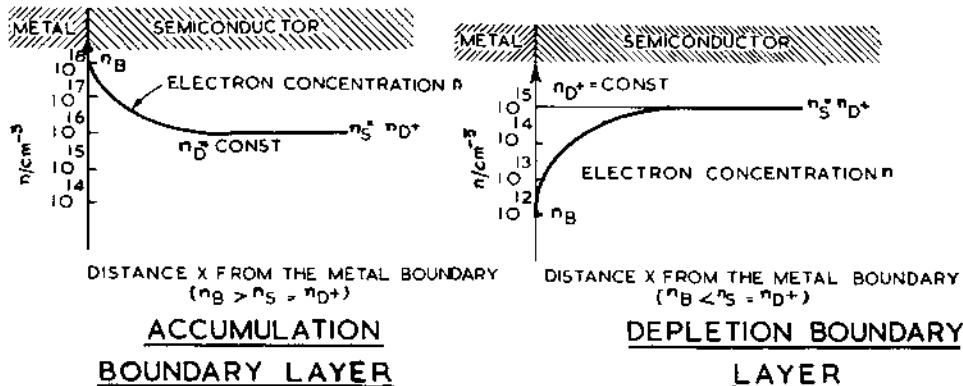


Fig. 7. Boundary layers at metal-semiconductor interfaces: n_D , the concentration of donors for instance on an n type semiconductor; n_S , the concentration of electrons at the interior of the semiconductor; n_B , the concentration of electrons at the boundary.

conductor interface. The latter therefore serves as a convenient model for understanding the mechanism of gas adsorption on semiconductors.

When a metal is placed in contact with a semiconductor, flow of electrons takes place across the interface as long as there exists a difference between the free energies of the electrons in the two substances. The charge gained thereby by the semiconductor (in contrast to the metal) is not spread over its surface only but extends to considerable depths (several atom layers) beneath the surface, since the impurity levels in the semiconductor are usually low in concentration. Thus, a boundary layer is set up to a finite depth in the semiconductor and it may be either an 'accumulation layer' or a 'depletion layer' as shown in Fig. 7. Typical cases of anionic adsorption on n -type and cationic adsorption on p -type semiconductors result in a decrease in the density of charge carriers in the surface region compared to the density in the bulk of the semiconductor. Such a region of depleted charge carrier concentration is known as 'depletion' region. On the other hand, in cationic adsorption on an n -type semiconductor, the surface is positively charged, and in this case the potential well collects the electrons from the conduction band and a large concentration of negative charge carriers are found close to the surface as a

part of the space charge region. This layer is said to constitute the 'accumulation' layer.

Turning our attention to the chemisorption of gases on semiconductor surfaces, it will be recognised that they are, in all, four possibilities, two each on n-type and p-type semiconductors respectively, as follows: Case 1: Cationic chemisorption on n-type semiconductors (e.g. H₂ on ZnO). Electrons transferred from adsorbate to semiconductor. Case 2: Anionic chemisorption on n-type semiconductors (e.g. O₂ on ZnO). Electrons transferred from semiconductor to adsorbate. Case 3: Cationic chemisorption on p-type semiconductor (e.g. H₂ on NiO). Direction of electron transfer same as in case 1, leading to depletion of holes. Case 4: Anionic chemisorption on p-type semiconductor (e.g. O₂ on NiO). Direction of electron transfer same as in case 2, leading to increase in hole concentration.

Thus, it is seen that the electronic properties of both types of semiconductors are profoundly modified by chemisorption of gases at their surfaces. Cationic chemisorption on n-type semiconductors (case 1) and anionic chemisorption on p-type semiconductors (case 4) give rise to increase in the concentration of charge carriers (namely, electrons and holes respectively) and thereby to an increase in overall electrical conductivity. These are therefore designated as "cumulative chemisorption". On the contrary, anionic chemisorption on n-type (case 2) and cationic chemisorption on p-type (case 3) semiconductors are cases of "depletive chemisorption", because in both cases the process is attended with depletion of charge carriers and hence decrease of electrical conductivity.

Apart from the changes in the electrical conductivity of the solid, there are quantitative aspects of the adsorption process itself in which sharp differences are manifested between cumulative and depletive chemisorptions. There being ample room in the conduction band of n-type semiconductors and for holes in the valence band of p-type semiconductors, cumulative adsorption will proceed to almost to the monolayer saturation limit without serious difficulty, as has been observed¹³ in the case of hydrogen chemisorption on zinc oxide and oxygen chemisorption on nickel oxide. On the contrary, depletive chemisorption will be severely restricted in surface coverage mainly because the 'space charge' that builds up in the boundary layer opposes the flow of electrons across the interface. This, very briefly, is the sum and substance of the boundary layer theory of depletive chemisorption, which was put forward in 1952 independently by Aigrain and Dugas¹⁴, and Hauffe and Engell¹⁵ and by Weiss¹⁶. A clear and succinct review of the

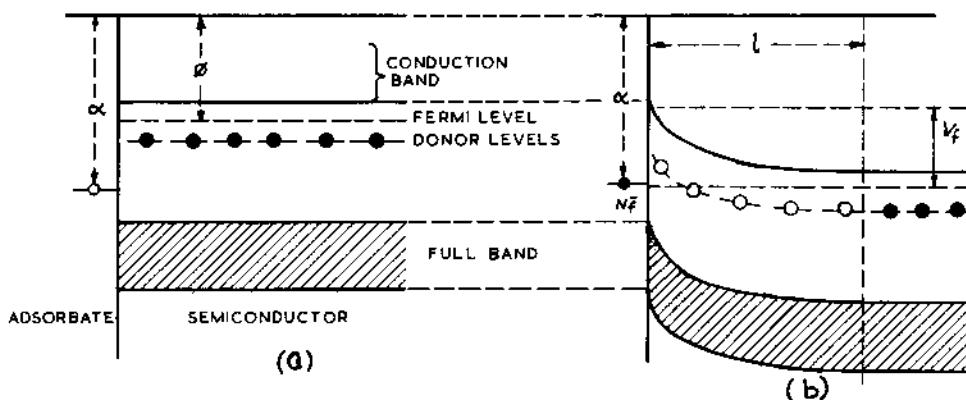


Fig. 8. Diagrammatic representation of anionic chemisorption on an n-type semiconductor (a) Before chemisorption and (b) after chemisorption: α , Electron affinity of the adsorbate; ϕ , work function of the semiconductor; V_f , $(\alpha - \phi)$.

theory is given by Stone¹³. Fig. 8a shows the relative energy levels in the adsorbate and in the solid prior to anionic chemisorption on an n-type semiconductor.

α is the electron affinity of the adsorbed atom and ϕ the work function of the semiconductor. The energy of chemisorption of the first atom on the n-type semiconductor will be $(\alpha - \phi) e$ where e is the electronic charge. As more atoms are adsorbed and more electrons are transferred, a space-charge builds up in the boundary layer, since the donor levels deeper in the solid will have to be tapped for electrons. Consequently, the potential energy of electrons in the solid becomes modified, and in going from the solid to the adsorbate they have to surmount a potential energy barrier V . The height of this barrier increases with each atom that is adsorbed, so that the Fermi level in the semi-conductor is progressively depressed. Ultimately, as soon as the potential energy of the electrons in the solid (i.e. the Fermi level) equals that in the adsorbate (as shown in Fig. 8b), equilibrium would be established and no further net adsorption takes place. If, at equilibrium, the barrier has a height V_f , and the number of

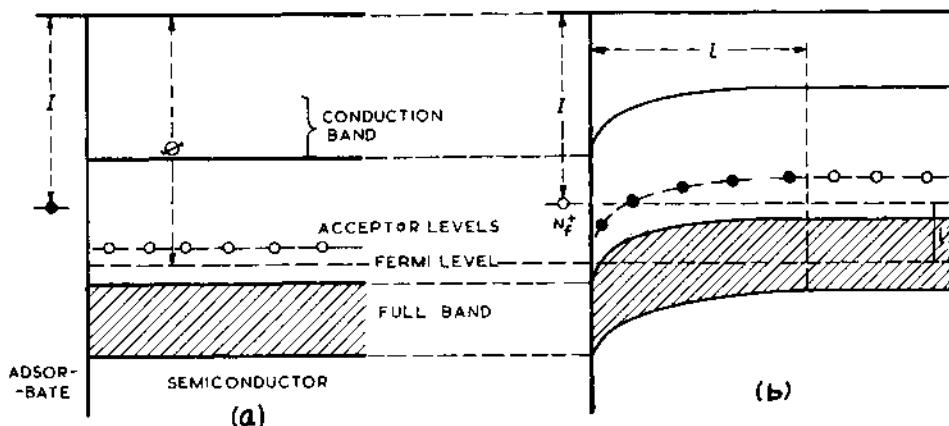


Fig. 9. Diagrammatic representation of cationic chemisorption on a p-type semiconductor (a) before chemisorption and (b) after chemisorption: I, Ionisation potential of adsorbate; ϕ , work function of the semiconductor; V_s , ($\phi - I$)

atoms adsorbed as anions = N_f , it follows that N_f electrons have been transferred from the semi-conductor, there being a depletion of carriers in a boundary layer of depth l .

Fig. 9 illustrates the case of a cationic chemisorption on a p-type semi-conductor. Again, Fig. 9a refers to the situation before chemisorption. The energy of chemisorption of the first adatom will be $(\phi - I)e$, where I is the ionisation potential of the adsorbate. As chemisorption progresses, the Fermi level of the solid, in this case, gradually rises and, at equilibrium (Fig. 9b) equals the potential energy of the electrons in the adsorbate. The height of the potential energy barrier at equilibrium, when N_f atoms are adsorbed cationically, is again denoted by V_f . The surface layers of the semiconductor have been depleted of carriers (holes in this case) to a depth l .

By assuming that the space charge density is constant in a

layer of thickness l , it can be shown¹³ that

$$N_f = \left[\frac{K}{2\pi e} n_0 V_f \right]^{\frac{1}{2}}$$

where K is the dielectric constant and n_0 is the concentration of carriers in the interior of the semiconductor. Using typical values for the parameters of the above equation, it has been estimated that equilibrium coverage in depletive chemisorption is attained at about 1% of the monolayer coverage. Although many systems seem to conform to this estimate, it is felt by critics of the boundary layer theory that too much emphasis cannot be placed on a purely physical model, as such an approach tends to minimize important chemical characteristics associated with semiconductor and metal surfaces.

Experimental Aspects of Chemisorption by Semiconductors. Measurements of electrical conductivity, Hall coefficients and thermoelectric potential have been used to a considerable extent to obtain information on the changes occurring in semiconductor catalysts under varying ambient conditions. Measurement of conductivity during adsorption is particularly useful for ascertaining the type of chemisorption that occurs on the semiconductor. Details of the technique are described in a recent review by Gray¹⁷.

Effect of Impurity Cations on Catalytic Properties of Semiconductors. The catalytic properties of a semiconductor, such as its specificity and activity for a particular reaction, are strongly dependent upon its electronic properties. This is strikingly demonstrated by the effect of 'doping' (i.e. the deliberate addition of small amounts of cations of valency different from that of the cation of the host lattice) on the electrical and catalytic properties of the semiconductor.

The systematic studies of Schwab and Block¹⁸ on the oxidation of CO on Li- and Cr-doped NiO and on Li- and Ga-doped ZnO provide suitable examples to illustrate this correlation. It was seen that the addition of Li^+ ions increases the conductivity and lowers the activation energy because of the increase in acceptor levels in NiO. Increased catalytic activity could be inferred from the decrease of activation energy. Doping with trivalent chromium ions had the opposite effect. These observations are consistent with the conclusion that the reaction involves transfer of electrons from the adsorbed CO to the surface, resulting in the formation of positive ions and hence would proceed much faster on a p-type semiconductor. These results are shown in Fig. 10.

With ZnO, addition of Li^+ ions caused decrease in conductivity due to the decrease in the number of electrons and addition of

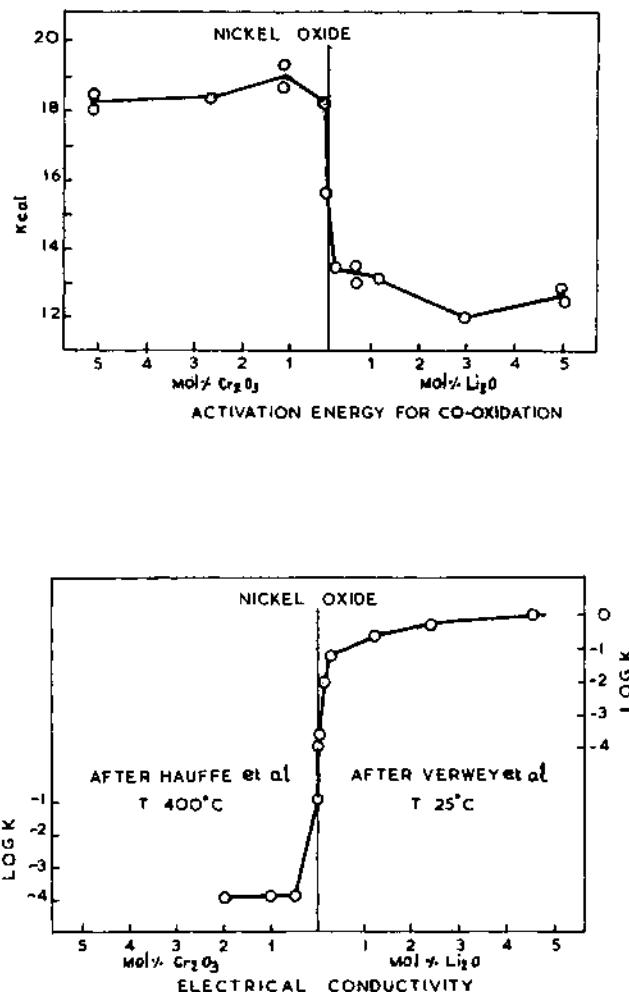


Fig. 10. Variation of activation energy with doping of nickel oxide with Li^+ and Cr^{3+} and the corresponding changes in the electrical conductivity¹⁸.

tervalent gallium increased the conductivity (by increasing the number of free electrons) and decreased the activation energy. This was consistent with the view that in the oxidation of CO on ZnO, chemisorption of oxygen which is an electron acceptor step, must be the slow one.

Some recent studies¹⁹ made in the authors' laboratory on the

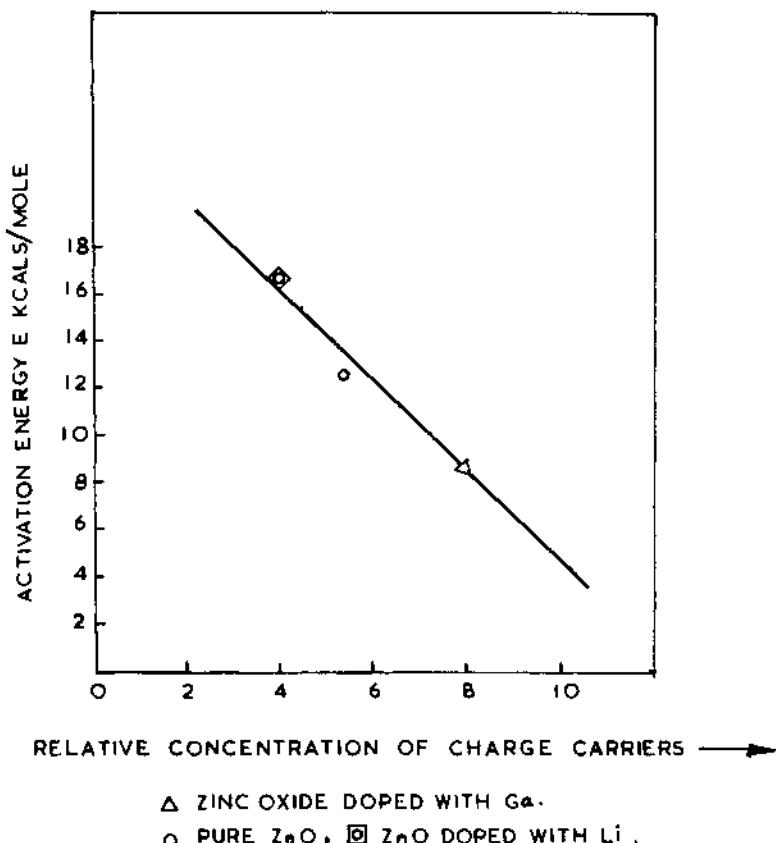


Fig. 11. Variation of activation energy with charge carrier concentration¹⁹.

dehydrogenation of isopropyl alcohol on zinc oxide catalysts, also illustrate the importance of the electronic factor in heterogeneous catalysis. The results of the experiments with 'doped' catalysts are particularly interesting. It was found that addition of Li^+ to ZnO increased the activation energy for this reaction while doping with Ga^{3+} decreased it. Further, a linear relationship was found to exist between the activation energy for the dehydrogenation and the relative concentration of the charge carriers as found from Seebeck potential measurements (Fig. 11). It was also observed that the electrical conductivity of the catalyst increased during reaction thus suggesting the possibility of simultaneous donor and acceptor type electron transfers in the rate determining step - which appears to be the adsorption of the alcohol through both the hydrogen and oxygen of the hydroxyl group.

III. INVERSE MIXED CATALYSTS

From the foregoing discussion it should be evident that the activity of a catalyst is closely related to the energy level of the electrons in the solid (i.e. the Fermi level) and the latter may be altered by adsorption of gases or incorporation of impurities. Schwab²⁰ has recently demonstrated the modification of the Fermi level by placing the semiconductor in contact with a metal support. The electron concentration per unit volume is several orders of magnitude greater in a metal than in a semiconductor and therefore a controlled deposition of a semiconductor on a metallic support, which by itself is catalytically inactive, could be expected to result in an active catalyst, because of the suitable manipulation of the energy levels in the semiconductor. This is the reverse of the usual practice of supporting a metal catalyst on a semiconductor base and has therefore been termed by Schwab as 'inverse mixed catalyst'.

Schwab has reported interesting results with three such catalyst-reaction systems, namely, (i) The oxidation of CO on NiO supported on silver - a donor type reaction favoured by p-type catalyst, (ii) The oxidation of SO₂ on Fe₂O₃ supported on silver - an acceptor type reaction favoured by n-type catalysts, and (iii) The oxidation of methanol vapour by air on ZnO mixed intimately with silver powder.

In the first of these catalysts, nickel was electro-deposited on silver foil and subsequently oxidised by heating with oxygen. Oxide layers of different thicknesses ranging from 80 Å to 15,000 Å were prepared and studied. On oxide layers thicker than 500 Å, the activation energy for CO - oxidation was nearly the same as on bulk (unsupported) nickel oxide, (16 kcal/mole) and the effect of the metal support was not felt. On oxide layers thinner than 500 Å however, the activation energy increased with decreasing thickness, and reached a value of 46 kcal/mole at 150 Å thickness. In the second case, with 60 Å thick layer of Fe₂O₃ on silver, the activation energy for SO₂ oxidation was lowered by as much as 18 kcal/mole. (The activation energy with unsupported bulk oxide was 31 kcal/mole and that with the silver supported oxide layer was 13 kcal/mole). In the third case also, the indications were that the overall activation energy was lowered.

These results have been explained as follows: The work function of silver is lower than that of the three semiconductors used. Hence, when contact is accomplished, the Fermi energies of both phases will level out through an injection of electrons from the metal into the semiconductor. The Fermi level of the semiconductor is thereby pushed up in the boundary layer. This causes the activation energy of the donor type reaction (i) on a p-type

semiconductor to be increased and, conversely, the activation energies of the acceptor type reactions (ii) and (iii) on n-type semiconductors to be decreased. Further studies in this new area of the effect of manipulation of the energy levels of semiconductors on catalytic activity could open up a new vista for catalyst selection.

IV. CONCLUSIONS

In the foregoing an attempt has been made to correlate catalyst performance with the semiconductor type based on electronic theory. However other correlations have also been proposed such as (a) the one connecting the d shell configuration of the transition metal ion with the activity of these systems towards H₂-D₂ exchange and other reactions involving hydrogen²¹ and (b) the relationship between catalytic activity of the transition metal oxides and the oxidation-reduction potentials of the oxides or the binding energy of oxygen in the bulk and on the surface of the oxide²².

REFERENCES

1. W.B. Innes, 'Catalysis' (Ed. P.H. Emmett) Vol. II, Reinhold, New York, Chapter 1 (1955).
2. F.F. Wolkenstein, 'The Electron Theory of Catalysis on Semiconductors' (Transl. by E.J.H. Birch) Pergamon Press, Oxford, p. 10-11 (1963).
3. D.A. Dowden, J. Chem. Soc., 242 (1950).
4. N.F. Mott and H. Jones, 'The Theory of Metals and Alloys', Oxford University Press (1940).
5. L. Pauling, Phys. Rev., 54, 899 (1938).
6. L. Pauling, Proc. Roy. Soc., A196, 343 (1949).
7. O. Beeck, Disc. Faraday Soc., 8, 118 (1950).
8. G.C.A. Schuit, ibid., 8, 205 (1950).
9. J.M. Thomas and W.J. Thomas, 'Introduction to the Principles of Heterogeneous Catalysis', Academic Press, New York, p. 134 (1967).

10. P.M. Gundry and F.C. Tompkins, 'Experimental Methods in Catalytic Research' (Ed. R.B. Anderson), Academic Press, New York, p. 100 (1968).
11. R. Suhrmann, *Advan. Catalysis*, 7, 316 (1955).
12. A.C. Zettlemoyer and R.D. Iyengar, 'Gas-Solid Interface' (Ed. E.A. Flood), Marcel Dekker Inc., New York., Vol. 2, p. 787 (1967).
13. F.S. Stone, 'Chemisorption' (Ed. W.E. Garner) Butterworths, London, p. 379 (1957).
14. P. Aigrain and C. Dugas, *Z. Electrochem.*, 56, 366 (1952).
15. K. Hauffe and H.J. Engell, *Z. Electrochem.*, 56, 366 (1952).
16. P.B. Weisz, *J. Chem. Phys.*, 20, 1483 (1952).
17. T.J. Gray, 'Experimental Methods in Catalytic Research' (Ed. R.B. Anderson) Academic Press, New York, p. 293 (1968).
18. G.M. Schwab and J. Block., *Z. Physik. Chem. N.F.*, 1, 42 (1954); *Z. Electrochem.*, 58, 756 (1954).
19. B. Viswanathan, Ph.D. Thesis, Indian Institute of Technology, Madras (1969).
20. G.M. Schwab, *Surface Science*, 13, 198 (1969).
21. G.M. Dixon, D. Nichols and H. Steiner, Proc. 3rd Int. Congr. Catalysis, North Holland Publishing Co., Amsterdam, 815 (1965).
22. K. Klier, *J. Catalysis*, 8, 14 (1967).

DISLOCATIONS AND SOLID STATE REACTIONS

A. S. Parashnis

Department of Physics

Indian Institute of Technology, Kanpur, India

The concept of a dislocation was evolved in order to explain the observed mechanical properties of metals and alloys. These properties include plastic flow, work-hardening, creep, etc. During the last twenty years or so it has been increasingly recognized that dislocations are involved in a much wider range of properties of materials than just mechanical. To appreciate this extensive influence it is necessary to understand the geometrical models and properties of dislocations. Section I is concerned with developing these and Section II with the role of dislocations in typical solid state chemical phenomena. We shall in fact not discuss mechanical phenomena though they are of physical and metallurgical importance. However, models of these atomic configurations called dislocations and their motions are more easily understood if we begin at where the subject began in the 1920's.

I. DISLOCATIONS, THEIR STRUCTURE AND PROPERTIES

A solid is in equilibrium when the forces acting on its constituents add to give zero. In addition to interatomic and intermolecular attractive and repulsive forces there can be external forces*. When the external forces are varied equilibrium is disturbed, but is re-established by suitable atomic movements. Starting from zero external forces and keeping them sufficiently small we have the experimental observation, viz., Hooke's law,

*We are here not concerned with what are called 'body forces' such as gravity.

that the stresses developed in the solid are proportional to the strains of deformation. Each of the two physical quantities, strain and stress, is a symmetric tensor of rank two, so that each has six components. The components of stress are linear functions of those of strain, the coefficients being called 'stiffness constants'. Clearly, there are 36 of them. In the case of crystalline solids (with which we are concerned here exclusively) the symmetry reduces this number. For example, the elastic properties of cubic crystals are completely described by 3 constants. It turns out, however, that the broad features of dislocation theory which is applicable to all crystals can be established under the assumption of isotropy over most of the solid. For the description of an isotropic, homogeneous, continuous medium two constants are generally sufficient. These are the so-called Lame' constants, denoted by λ and G . The latter is the familiar 'shear modulus'.

Experimental Observations on the Plastic Flow of Crystals.

As long as external forces are small enough the crystal is only elastically deformed. Beyond a certain value of the applied stress it deforms plastically, i.e., a permanent strain is developed. Plastic deformation takes place by a shearing movement of parallel atomic planes over one another, not by processes such as elongation on an atomic scale. It is, therefore, only the shear stresses which need be considered. The shearing movement takes place invariably on crystallographic planes and along crystallographic directions in those planes, mostly on the closest packed planes and along closest packed rows. The reason is obvious: the interplanar force for the closest packed planes must be weakest because the interplanar spacing is the greatest. Similarly for the closest packed rows. Before and after the atomic movements involved have taken place the solid is crystalline.

The shearing motion, called 'slip', has thus a slip plane and a slip direction associated with it as shown in Fig. 1. The following table gives these for the most common structures.

<u>Crystal</u>	<u>Slip plane</u>	<u>Slip direction</u>
fcc metals	{111}	<110>
hcp metals	{0001}	<1120>
bcc metals	{110}, {123}	<111>
NaCl, MgO	{110}	<110>
PbS	{100}	<110>
AgCl, AgBr	not well defined	<110>
Diamond	{111}	<110>
CsCl	{110}	<111>

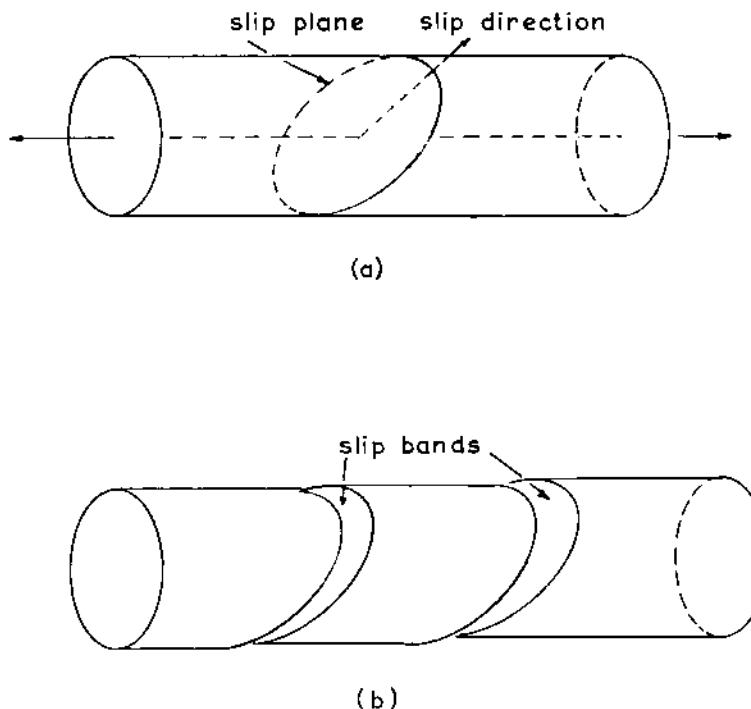


Fig. 1. To show schematically how a crystal in the form of a cylinder deforms plastically according to the slip system it has. (a) Undefomed, (b) Deformed.

It is clear that the actual slip plane is not determined by packing density alone. For example, NaCl and PbS have identical crystals structures but different slip planes, while for the silver halides the slip plane is not even defined. Other factors, notably the kind of chemical bonding, must be operative. In general, salts of rocksalt structure in which the sum of ionic polarizabilities is high slip, like PbS, on $\{100\}$ planes. In the case of very high values of polarizabilities, as in the silver halides, pencil glide may occur. In crystals with directional bonds other factors may be important.

Any mechanism that may be evolved to explain plastic flow must take all such experimental observations into account.

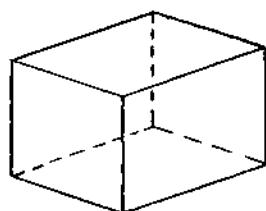
Dislocations. The first attempt (by Frenkel, in 1926) at calculating the theoretical shear strength of a crystal gave the value $G/2\pi$, which is typically of the order of $10^{11} \text{ dy cm}^{-2}$. This calculation was based on the assumption of a simultaneous slip all across a slip plane. Real crystals deform plastically

at much lower levels of stress, about $10^5 - 10^8 \text{ dy cm}^{-2}$. It was to explain this discrepancy that Taylor (in 1934) evolved the concept of consecutive slip. Slip is supposed to begin at one end of a crystal, with only a fraction of the slip plane being involved, and to proceed step by step across the slip plane, with only a few atoms participating in shearing movements at any one time. This kind of movements can beautifully be seen in the soap bubble movie made by Bragg and Nye.

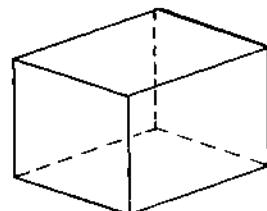
Consecutive slip involves at any one time the breaking and re-making of only a few atomic bonds. Consequently, this process requires less stress than does Frenkel's process.

If the process of consecutive slip is imagined to be stopped midway in the crystal the boundary between the slipped and unslipped parts of the slip plane will be a region of severe distortion. Atoms nearabout this boundary will not have, for example, the correct number of nearest neighbours. In the present-day model of this region the distortion is thought to be severe in a long cylinder with the boundary line as its axis and is assumed to decrease rapidly with distance. The lateral dimensions of the cylinder of distortion are a few interatomic distances, which is why it is called a linear imperfection; the name given to the imperfection is 'dislocation'.

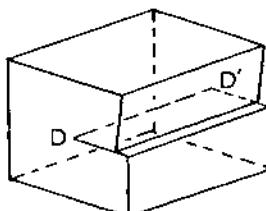
What is the atomic arrangement at a dislocation? This depends upon the crystal structure and on the way the boundary is created. The boundary may be along the direction of the applied force or perpendicular to it or at some other angle. We consider here the two simple cases of the boundary being normal to and parallel to the force applied to a simple cubic crystal. Fig. 2(a) and 3(a) show the undistorted crystal supposed to be of the shape of a cube; Fig. 2(b) and 3(b) show the appearance of the crystal after the external stress has been applied for some time and then removed; Fig. 2(c) and 3(c) show the atomic arrangements in the two cases. The dislocation of Fig. 2(c) is called an 'edge dislocation' because it is the edge of a 'half-plane' which may notionally be employed to describe the situation. The dislocation of Fig. 3(c) is called a 'screw dislocation' because its creation has converted the stack of parallel planes that the crystal was into a continuous helicoidal surface. A dislocation can be straight, curved, or a closed loop. A given dislocation can, depending upon circumstances, transform from one to another suitable form, with a corresponding change in the atomic arrangement. The invariant quantity that characterises a dislocation is its Burgers vector. For finding it we start with some atom in the undislocated crystal and take vectorial steps, a_i , from atom to atom, finally returning to the starting point. Clearly then $\sum_i a_i = 0$. We now take a similar circuit, the Burgers circuit, in the dislocated crystal,



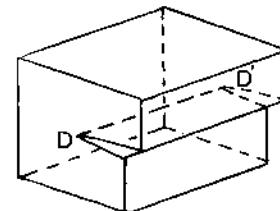
(a)



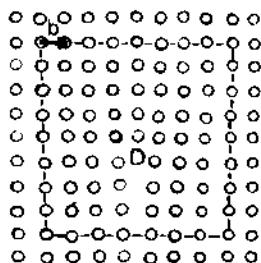
(a)



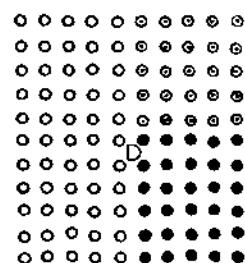
(b)



(b)



(c)



(c)

FIG. 2

FIG. 3

Fig. 2 and 3. Fig. 2(a) and 3(a) show a cube-shaped crystal before deformation while Fig. 2(b) and 3(b) show the deformation produced by applying an external stress. Fig. 2(c) and 3(c) show the atomic models of edge and screw dislocations respectively in a plane normal to the line of dislocation. Open circles represent atoms in the plane of the paper, full circles those in front and centred circles those behind. In Fig. 3(c) atoms have not yet relaxed to their new positions. It is clearly seen why precipitation tends to occur more easily along edge dislocations than along screw.

and find that (assuming that we are sufficiently far away from the severely distorted regions of the crystal) in this case the summation is non-zero, if the circuit encloses the dislocation. (See Fig. 2(c) and 3(c).) If the sum is $\sum z_i = -b$, we call b the Burgers vector. No matter where the circuit is taken the result is the same as long as it encloses the dislocation. In fcc crystals, for example, $b = a/2 \langle 110 \rangle$, meaning that half the cube edge, ($a/2$), when projected along a $\langle 110 \rangle$ direction gives b . (More formally, in an isotropic solid, the components of the Burgers vector are given by $-b_i = \phi (\partial u_i / \partial x_j) dx_j$, where u_i are the elastic displacements and x_j the space co-ordinates). A consequence of this is that a dislocation line cannot end within a crystal except at a node of dislocations or an internal surface such as a grain boundary. Clearly, it can form a closed loop within the crystal, for a loop is endless.

Elastic Fields of Dislocations. By virtue of the fact that there is a severe distortion at dislocations they possess stress and strain fields. The strain at a dislocation in a continuous medium turns out to be infinite. This mathematical difficulty cannot arise for dislocations in crystals because atomic displacements are discrete. As, however, we do not have a detailed knowledge of the atomic displacements of a dislocation we must treat the problem quasi-classically. For this purpose we think of the core of the dislocation, i.e., the region of severe distortion, as being only a few atomic distances in diameter. This is a perfectly reasonable assumption. Outside this core we treat the crystal as though it were a classical continuum.

Let a screw dislocation be produced in a cylindrical crystal as shown in Fig. 4(a,b). A cylindrical shell of radius r and thickness dr when opened out before deformation would be a rectangle (Fig. 5a); after deformation it would become a parallelogram (Fig. 5b). Thus shear strain at a distance r from the axis is

$$\epsilon = \theta = b/2\pi r, \quad (r > r_o)$$

and the stress is

$$\tau = G\epsilon = Gb/2\pi r \quad (r > r_o).$$

They go to zero as $r \rightarrow \infty$. There is thus throughout the crystal a strain energy. The strain energy density is given by $u = \frac{1}{2} (\text{stress} \times \text{strain})$. If we multiply this by the volume $2\pi r L dr$ of the shell we get the energy in the shell, viz.,

$$dU = udV = \frac{1}{2} \frac{Gb}{2\pi r} \cdot \frac{b}{2\pi r} \cdot 2\pi r L dr = \frac{Gb^2 L}{4\pi} \frac{dr}{r}$$

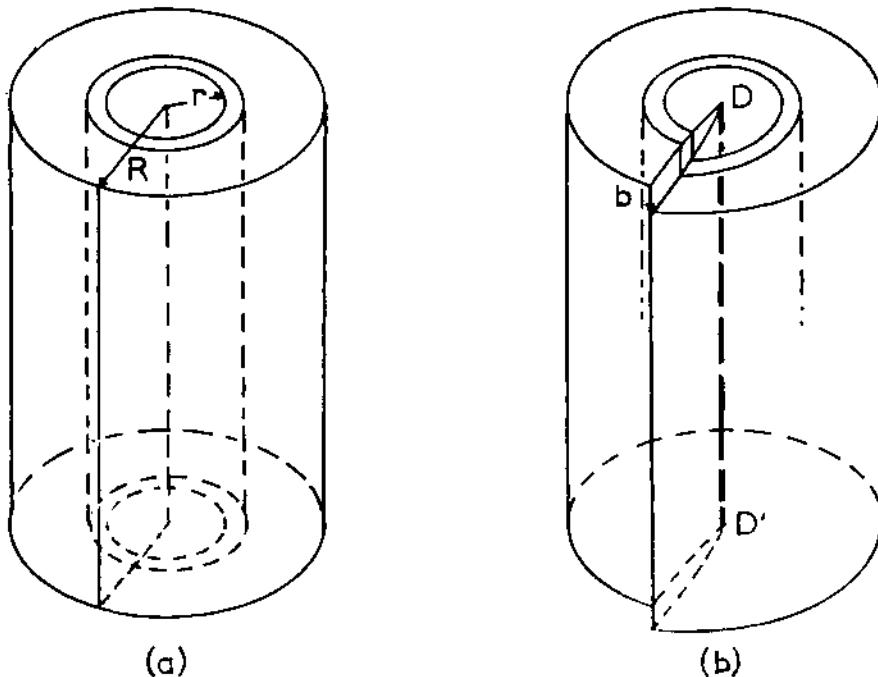


Fig. 4. To show how a cylindrical crystal is deformed when a screw dislocation is produced along its axis. The same state of stress could be produced by a stress acting along any axial plane. In a continuous isotropic medium a screw dislocation thus has no slip plane as such. In crystalline media, of course, the situation is different.

The total energy outside the core in the cylindrical crystal is obtained by integrating this:

$$U = \int dU = \int_{r_0}^R \frac{Gb^2 L}{4\pi} \frac{dr}{r} = \frac{Gb^2 L}{4\pi} \log \frac{R}{r_0}$$

The energy per unit length is thus

$$\xi \approx \frac{Gb^2}{4} \log \frac{R}{r_0} .$$

For a crystal of copper, for example, of $R = 1$ cm we have $b = 2.5$ Å, $r_0 \approx 5b$, $G = 4 \times 10^{11}$ dy cm $^{-2}$ so that ξ is about 3×10^{-4} erg cm $^{-1}$ or about 5 eV per atom plane. The core energy is estimated to be about $\xi_0 = \xi/20$, and may usually be neglected

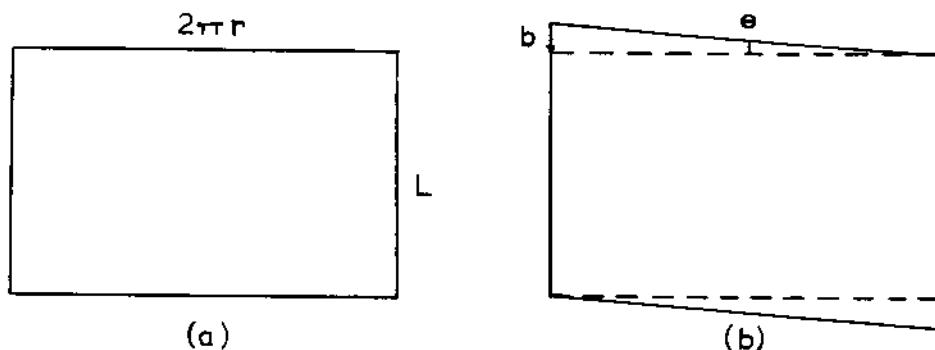


Fig. 5. To show how a co-axial shell is deformed after the screw dislocation of Fig. 4 is produced. (a) Undeformed, (b) deformed.

for phenomena occurring at a distance from the dislocation. The calculation for an edge or a mixed dislocation proceeds on similar lines but is more complicated because of what are called normal stresses. The energy is slightly more, about 8 eV per atom plane.

The quantity ξ is a contribution to internal energy. The contribution of entropy has been estimated to be about 10^{-6} kT, i.e., it is entirely negligible. As ξ is positive, dislocations cannot exist as thermodynamically stable imperfections.

As the energy U is proportional to the length of the dislocation the latter tends to minimize energy by contracting as much as possible (within, of course, the constraints of the crystal structure). This tendency gives rise to the 'line tension' of a dislocation. It is given by $T \approx Gb^2$ and thus has the dimensions of force.

It can be shown, further, that if a crystal containing a dislocation is subjected to a shear stress τ a force τb per unit length acts on the dislocation making it glide in its slip plane. This force is not on the crystal or individual atoms but on the atomic configuration that a dislocation is. The shear stress τ could be external or internal; the latter can be the shear stress due to another dislocation or a particle of a second phase in the crystal. Dislocations thus interact with one another through their elastic fields. This interaction leads to configura-

tions of dislocations called small angle grain boundaries and is also of importance in mechanical properties of materials.

Interactions with Point Imperfections. If point imperfections (vacancies, interstitials, impurity atoms) are present in the crystal, very interesting effects occur. Consider the edge dislocation of Fig. 2(c) as an example. It is easy to see that the average interatomic distance around the dislocation is, in the upper half of the crystal, smaller than the equilibrium distance, while in the lower half it is greater. In other words, the upper half is in a state of compression while the lower half is in a state of tension. Vacancies, therefore, tend to be around the dislocation in the upper half and interstitial atoms in the lower one. The atom marked D in Fig. 2(c), for example, is particularly vulnerable to attack by a vacancy. If a vacancy takes away this (or similar atoms) the dislocation develops a 'jog', i.e., in this case, the edge of the half plane does not remain straight; it is rather jagged. An interstitial atom adsorbing below the atom D would cause a jog of the opposite kind to develop. In ionic crystals there is one more effect. The point defects corresponding to cations and anions have different sizes, and different formation and activation energies. So, more of one kind may be there around a dislocation. Consequently, dislocations can have an electrostatic charge.

Motion of Dislocations. Thus a dislocation (or its segments) can move out of the slip plane into another parallel slip plane. A screw dislocation develops jogs as a result of interaction with vacancies; these jogs clearly have an edge character. This motion, called 'climb', which involves the transport of matter from or to a dislocation is described as non-conservative motion as against the conservative motion of glide in the slip plane. Point defects are essentially chemical defects. So climb can be said to be activated by chemical forces just as glide is activated by mechanical forces.

II. THE ROLE OF DISLOCATIONS IN CHEMICAL REACTIONS

The first recorded observation of the influence of mechanical forces on chemical phenomena was by Faraday (in 1834) who noticed that efflorescence of sodium carbonate was activated by mechanical violence. In modern times one of the earlier observations was by Pohl (in 1934) that the photolysis of silver halides was more in plastically deformed crystals. Today a multitude of such reactions on the surface or within the bulk of crystals are known and known to be closely connected with dislocations. We shall study a few of them.

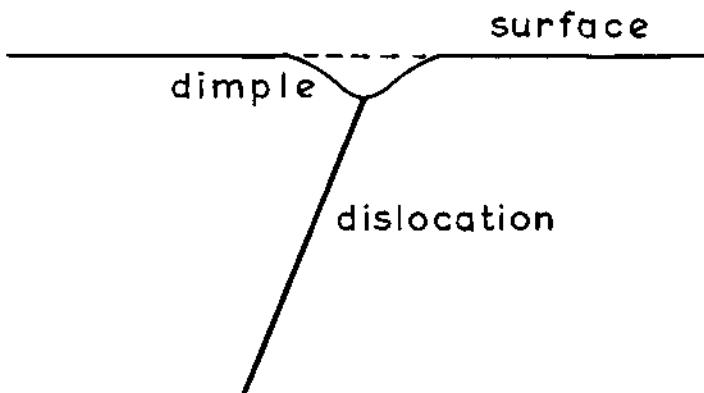


Fig. 6. To show the 'dimple' produced at the site of the emergence of a dislocation on the surface of a crystal.

Surface Effects. There are several reasons why dislocations must play an important role in solid state reactions. Let us consider surface reactions. In the first place, the surface of a solid has, in addition to its surface energy (γ erg cm $^{-2}$) a surface tension (g erg cm $^{-2}$) because the interatomic spacing at the surface is different from that in the bulk. The surface tension is compressive or tensile depending upon which spacing is greater. The core of a dislocation as explained in Section I has similar features. Dislocations, therefore, have a tendency to lower their energy by being near the surface. Dislocations attracted to the surface due to this effect will have their length and Burgers vectors parallel to the surface. Second, if a dislocation emerges at the surface and has a component of b normal to the surface there is a ridge on the surface. (Recall Fig. 3(b).) Finally, as a result of the line tension there is a dimple at the emergence point (Fig. 6).

For these and similar reasons it is not difficult to see that adsorption of atoms (from solutions or vapours) is preferentially at dislocations. In a similar way, evaporation or similar processes are more rapid at dislocations. These two effects lead to chemical and thermal etching which are powerful methods of studying dislocations.

Hollow Dislocations. Consider the equilibrium of a crystal surface with an emerging dislocation. Let the dislocation have a hollow core of radius r_0 . Since the elastic field of the dislocation varies as $1/r$ the energy is reduced as r_0 is increased.

But thereby the surface energy associated with the hollow core is increased. For a unit length of the dislocation the latter is $2\pi r_0 \gamma$. The self-energy per unit length of the dislocation is $\xi = (Gb^2/4\pi) \log(R/r_0)$. The total free energy is to be minimum, that is

$$\frac{d}{dr_0} \left\{ 2\pi r_0 \gamma + \frac{Gb^2}{4\pi} \log \frac{R}{r_0} \right\} = \left\{ 2\pi \gamma - \frac{Gb^2}{4\pi} \frac{1}{r_0} \right\} = 0.$$

This gives the equilibrium value $r_0 = Gb^2/8\pi\gamma$. For metals and other ordinary crystals, a typical value of γ is 10^3 erg cm $^{-2}$, which gives $r_0 \approx 0.5 \times 10^{-8}$ cm. That is, in such cases a dislocation will not be hollow. If however, b is large, as in complicated structures like silicates, in molecular crystals of high molecular weight and in protein crystals, the dislocation core will be hollow (Fig. 7). This will greatly enhance diffusion of atoms along the dislocation, and hence chemical activity.

Crystal Growth. Chemico-physical phenomena such as crystal growth from solution and from vapour, etching and evaporation are thought to occur by the movement of molecular steps across the surface. If the surfaces of a crystal are assumed to be ideally smooth or with unviable steps and kinks the probability of spontaneous nucleation for further growth is very small. It would take 10^{40} yr for a crystal to grow from solution or vapour under very low supersaturation. Yet crystals grow to observable sizes in a matter of hours. Looking for a mechanism to explain

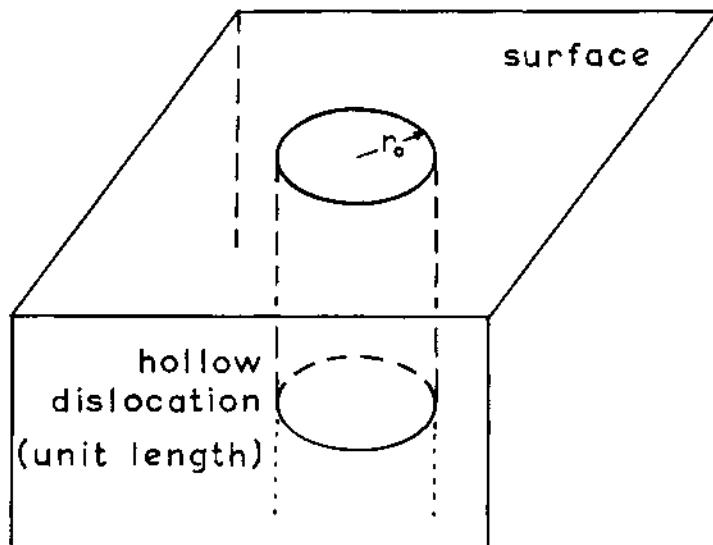


Fig. 7. To show a 'hollow dislocation'.

this, Frank developed his theory of 'spiral growth' which invokes a self-perpetuating step on the surface. This theory has been amply confirmed in innumerable cases but there are others where it fails.

Solution and Evaporation. There are several observations like Faraday's. Silver whiskers (which presumably contain no dislocations) are much more resistant to nitric acid than large crystals. This must be connected with their comparatively smoother surfaces. Hydroquinone evaporates far less rapidly when it is in whisker form, for the same reason. If a whisker is bent, it dissolves rapidly at the bend.

The theory of dissolution and evaporation is not entirely analogous to that of growth, since corners and edges of crystals provide effective nuclei for dissolution and evaporation. If corners and edges are far away dissolution and evaporation commonly occur by the formation of pits at emergence points of dislocations. If ΔP is the excess of energy of unit volume of the crystal over the same material in solution or vapour, the contribution to energy is $-\pi r_o^2 \Delta P$. The free energy then is,

$$F = 2\pi\gamma r_o - \frac{Gb^2}{4\pi} \log \frac{R}{r_o} - \pi r_o^2 \Delta P.$$

Minimizing this we have,

$$\frac{dF}{dr_o} = 2\pi\gamma - \frac{Gb^2}{4\pi r_o} - 2\pi r_o \Delta P = 0$$

$$\text{or } (2\pi \Delta P) r_o^2 - (2\pi\gamma) r_o + \frac{Gb^2}{4\pi} = 0,$$

which is a quadratic in r_o . The roots are real if $4\pi^2\gamma^2 > 2Gb^2(\Delta P)$, and equilibrium is possible so that no visible etch pit is formed. If $4\pi^2\gamma^2 < 2Gb^2(\Delta P)$ or $(\Delta P) > Gb^2/2\pi^2\gamma^2$, the roots are complex and the hollow at the surface goes on expanding till a visible pit formed. This then is the condition for etch-pitting. Even if a pit is formed it will not be clearly identified as such unless its sides are steep. Here we come to the question of the kinematics of the travel of molecular steps. They will travel such that the sides are steep if the solution contains a suitable 'poison'. Frank has given an elegant theory on these lines. Segregation of impurities in dislocations changes the picture considerably. Here the evidence is conflicting. In many ionic solids new dislocations etch faster since the old dislocations have less energy because of impurity segregation but in metals and semiconductors the opposite is true.

Formation of etch pits at dislocation sites gives a very

powerful method of studying dislocations as agents of plastic deformation. Once the right etchant is known it is also the simplest and most direct method. However, surface irregularities other than those connected with dislocations (for example, nuclei of impurities, 'dust' particles, features incorporated during growth, etc.) can also give rise to etch pits and other etch figures, so that it is necessary first to establish a one-to-one correspondence between dislocations and etch figures. This is not always easy and there is an undesirable tendency to ascribe all dissolution to dislocations.

Precipitation. One of the earliest methods of making dislocations visible in the optical microscope depends on enhanced precipitation of impurities along dislocations. As dislocations are regions of increased chemical potential nuclei will grow to visible particles preferentially at dislocations. In such a chemical reaction pipe diffusion along the dislocation also plays a part. The theory proceeds along similar lines. Now, however, ΔP is the release of free energy when unit volume of precipitate is formed and γ is the interfacial energy (per unit area) between the precipitate and the matrix. The free energy per unit length of a dislocation is given by,

$$F = 2\pi\gamma r - (Gb^2/4\pi) \log(r/r_0) - \pi r^2 \Delta P,$$

which gives as the equilibrium condition

$$\frac{dF}{dr} = 2\pi\gamma - Gb^2/4\pi r - 2\pi r \Delta P$$

just as in the case of dissolution. For sufficiently large values of ΔP , dF/dr is negative and F is a decreasing function of r . The precipitate will then grow indefinitely as long as there are impurity atoms available and they are mobile. For small ΔP , dF/dr vanishes for two values of r , which are the roots of

$$(8\pi^2 \Delta P) r^2 - (8\pi^2 \gamma) - Gb^2 = 0,$$

a small value giving rise to a stable precipitate and a larger one corresponding to a precipitate that will grow indefinitely. If ΔP is such that the discriminant of the quadratic equation vanishes we have

$$(\Delta P)_c = \frac{2\pi^2 \gamma^2}{Gb^2} = \frac{\gamma}{4r_h},$$

r_h being the radius of the stable hollow dislocation. We may define the ratio, α , of the actual value of ΔP with this critical value:

$$\alpha = (\Delta P) / (\Delta P)_c = Gb^2 \Delta P / 2\pi^2 \gamma^2 .$$

Expressing the smaller of the two roots in terms of this we get for the radius, r_s , of the stable precipitate

$$r_s = 2\alpha^{-1} \left\{ 1 - (1-\alpha)^{\frac{1}{2}} \right\} r_h .$$

It is clear that a stable precipitate will be formed only if α lies between zero and unity, which two limits give the values r_h and $2r_h$ for r_s .

It was mentioned earlier that precipitation of impurities on dislocations leads to the possibility of actually seeing dislocations in an optical microscope. This kind of decoration of dislocations has been observed in a wide variety of crystals, ionic crystals and transparent minerals in particular. It frequently turns out that in decorated crystal there are random precipitates too. There seems to be a competition between precipitation on dislocations and in regions where there are no discernable dislocations.

If a decorated crystal is chemically etched etch pits are formed at the ends of decorated dislocations. The literature contains several observations establishing this correspondence.

Photochemical reactions such as the photolysis of silver halides are known to be very closely related to dislocations. On the other hand the photochemical reduction of lead azide does not seem to involve dislocations. The interaction between dislocations and precipitation of photolytic silver in crystals of silver halides is a very interesting phenomenon. When these crystals are exposed to actinic light silver particles start growing along dislocations. As they grow they strain the matrix and by the process of prismatic punching generate sequences of dislocation loops. These loops in turn are sites for further photolytic precipitation.

Precipitation, even more than adsorption of individual solute atoms, along dislocations has the effect of 'locking' them, i.e., making them immobile. On this effect depends profoundly the technologically important phenomenon of age-hardening.

Oxidation, Sulphidization, Decomposition, Etc. Oxidation, sulphidization and similar reactions are of interest from the point of view of not only chemistry but technology of corrosion also. Not much work has been done on these problems in so far as dislocations play a part, and this remains a virgin field.

It is at once clear, at least in principle, that chemical transformations in solid systems can be greatly facilitated at dislocation sites. This is because not only does chemical potential differ but also does stereochemical environment in the vicinity of dislocations. The nucleation of a transformation may therefore be reasonably expected to start at dislocation sites. Further, chemical kinetic measurements should reveal this fact. These measurements seem to be difficult but they have been made and the connection with dislocations fairly well established in some cases.

As long as oxygen pressure is low and the oxide film thin, oxide nuclei seem to form on iron at dislocation sites. There is indication that iron oxide nucleates at slip lines also in the case of iron whiskers. The oxidation of copper does not seem to prefer dislocation sites.

Thomas and co-workers have established the role of dislocations in solid state reactions in specific but quite diverse cases. We shall consider some of them as examples, though others also have made similar observations.

(i) Calcite crystals decompose beyond about 500°C. By measuring the rates of decomposition at different elevated temperatures the activation energy for decomposition can be determined in the usual way. The following data were obtained.

Samples	Activation Energy 'linear' rate	(Kcal mole ⁻¹) 'initial' rate
Untreated	51 ± 4	38 ± 6
Quenched	48 ± 4	27 ± 6

Etch pit counts showed that, as expected, quenching introduced a number of dislocations; the dislocation density increased by 10³-fold. The activation energy during nucleation and the initial decomposition is, as seen in the table above, 30% less for the crystals containing a large number of dislocations. This means that the intrinsic energy (see Section I) of dislocations contributes to the activation of ions to form a transition complex. Once the reaction proceeds, i.e. in the linear region, the reaction products are bound to influence the further reaction, even more than dislocations would. That is why the activation energy is the same in this part. (See Thomas and Renshaw 1967a).

(ii) Sodium nitrate (NaNO₃) undergoes a λ -transition at 277°C. The resulting lattice distortion is not very severe; the new structure is only slightly different from the original calcite-

structure and still is rhombohedral. Melt-grown crystals have a dislocation density of 10^7 cm^{-2} while solution-grown crystals have 10^4 cm^{-2} , as determined by etch pit counts. There is marked difference between two differently grown samples which (difference) is ascribable to dislocations. Further, dislocation multiplication occurs as soon as the λ -transition commences. In fact, this latter kind of involvement of dislocations should exist in most transformations. (See Bahl and Thomas 1967b.)

(iii) There are very few organic crystals the dislocations in which have been studied. Williams and Thomas (1967c) chose to study dislocations in anthracene because divergent results had been obtained by different workers making measurements on self-diffusion, fluorescence spectra, photoconductivity, radio-thermoluminescence, etc. and these discrepancies were being ascribed to dislocations or other defects. Of all known etching methods for anthracene, alternate washing with hot and cold water was found to be the most reliable method for studying dislocations.

Anthracene is monoclinic, has two molecules per unit cell and the unit cell volume is about 580 \AA^3 . It undergoes photodimerization in the solid state to yield orthorhombic dianthracene. This is accompanied by a volume reduction of 20%. Experimental evidence is that appreciable distortion of the anthracene lattice is required to permit photodimerization. Taking into account the nature of the distortion required it had already been suggested that the dimerization reaction occurs at the surface rather than in the bulk. It is clear from the two structures that at some stage of conversion to the dimer it is necessary to have columns of parallel anthracene molecules. Williams and Thomas investigated the structure of edge dislocations emerging on the surface and found that there is around them such a parallelism though only rough. It is, therefore, not surprising that the dimerization reaction begins at the surface.

Another organic crystal investigated by Thomas and Williams (1967d) is ordinary sucrose which undergoes decomposition at 120°C at pressures of $1 \mu\text{torr}$. Apparently a similar decomposition in air has some technological importance in the sugar industry. Thomas and Williams evolved a suitable etchant, established the Burgers vector and investigated whether there is a connection between the decomposition nuclei and dislocations. Results were inconclusive.

Biochemical Phenomena. An interesting application of the model of a dislocation in a crystal is to muscular movement (Nabarro 1967). A muscle is composed of a regular two-dimensional array of two types of filament: one type is the protein actin and the other is one containing the protein myosin. An actin

filament consists of two strands twisted together while the myosin-containing protein has regularly spaced projections, the repeat distance being 435 Å. The twisted strands of actin make a contact with the myosin-containing filament at every 406 Å. The latter is attached by its projections to six surrounding action filaments. The contact between two neighbouring unlike filaments forms a vernier with a repeat distance of $1/(1/406 - 1/435) = 6100$ Å. If the two spacings were exactly equal it would be very difficult for one to slide continuously on the other. Still the vernier model does not make clear how the sliding which constitutes muscular movement takes place. It becomes clear if the system is thought of as a crystal containing dislocations spaced at 6100 Å. Muscular movement then consists simply of the passage of these dislocations.

In conclusion it may be said that there are now known innumerable instances where the connection between solid state reactions and dislocations is well established. For want of time many interesting cases (Thomas, 1970) have not been treated in this article. But it seems that much more work is required and can be done on systems of chemical and technological importance.

REFERENCES

1. J.M. Thomas and G.C. Renshaw, J. Chem. Soc. (A), p. 2058 (1967).
2. O.P. Bahl and J.M. Thomas, J. Mat. Sci. 2, 500 (1967).
3. J.O. Williams and J.M. Thomas, Trans. Faraday Soc., 63, 1720 (1967).
4. J.M. Thomas and J.O. Williams, Trans. Faraday Soc., 63, 1922 (1967).
5. F.R.N. Nabarro, "Theory of Crystal Dislocations" (Oxford University Press, London), p. 794 (1967).
6. J.M. Thomas, Chemistry in Britain, 6, 66 (1970).

BIBLIOGRAPHY

1. S. Amelinckx, "The Direct Observation of Dislocations", Supplement 6 of Solid State Physics, Academic Press, N.Y., pp. 1-108 (1964).

2. J. C. Fisher et al. (Ed.), 'Dislocations and Mechanical Properties of Crystals', Proceedings of an International Conference held in 1956 at Lake Placid, John Wiley & Sons, N.Y., pp 1-175 (1957).
3. F.R.N. Nabarro, 'Theory of Crystal Dislocations', Oxford University Press, London (1967).
4. J. Weertman & J.R. Weertman, 'Elementary Dislocation Theory', The Macmillan Company, N.Y. (1964).
5. Recent Volumes of 'Reactivity of Solids', Proceedings of various international conferences of the same title. (a) Amsterdam (1960). (b) Munich (1964). (c) Schenectady (1968).

SOLID STATE REACTIONS

S. R. Yoganarasimhan

Department of Chemistry

Indian Institute of Technology, New Delhi, India

I. ROLE OF DEFECTS IN THE REACTIVITY OF SOLIDS

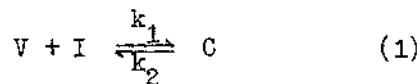
The growth of our knowledge in the properties of solids, namely, electronic spectra, electrical conductance, thermal and photochemical decompositions, elastic properties etc., has been closely associated with the growth of information on the defect state of solids. During the last forty years sufficient data have been accumulated to prove an apparently simple statement "the more perfect a crystal is, the smaller is its reactivity. The seats of reactivity are found at those places where there is a defect or fault in the perfect order"¹. The defects can be point defects, dislocations or stacking faults of atomic order or bulk defects originating from the misalignment of crystallites, the high angle grain boundaries or networks of dislocations in the crystallites. The defects of a atomic order are of extreme importance in controlling the mechanism of many reactions in solids. Mass transport which is the main process in most reactions is greatly controlled by the nature and distribution of these defects.

Point Defects: The following, amongst others, have been listed by Hedvall² as the factors influencing the reactivity of solids. (i) Deviations from normal crystallographic structure of solids; (ii) Imperfections in crystals; (iii) Differences in phases; and (iv) Changes in electronic and magnetic states.

The present discussion will be limited to the role of imperfections on the reactivity of solids. The importance of the role played by lattice vacancies and other point defects has been discussed at length by Dienes and Damask³. Since almost all solid

state reactions involve some kind of migration of atoms, the migration of vacancies could determine the mechanism of reactions. The vacancies migrate either to the sinks such as dislocations or to the surface. Such a straightforward diffusion of vacancies can only be a qualitative description of the diffusion process. The diffusion becomes considerably more complex because of interactions between vacancies and between vacancies and impurity atoms. These interactions have been described by Dienes and Damask³ in terms of a chemical rate theory.

The diffusion can take place by various modes with the possible exception of a direct exchange of atoms without the intervention of a lattice defect, which is energetically unfavourable. The vacancy diffusion, interstitial diffusion are the other possible modes. If impurity atoms are also present in the lattice, they are more likely to be accommodated very near the vacancies, so that the strain accompanying the introduction of the impurity atom can be relieved to some extent. Dienes and Damask³ conclude that the strain interaction between an oversized atom and a vacancy can be an important part of the binding energy. So one expects the vacancies to be bound in the compressed regions, and interstitials in the regions of dilation. If V is the free vacancy concentration in the crystal and I, the concentration of unbound impurities and C, the concentration of valancy-impurity complex, then



$$\frac{k_1}{k_2} = Z e^{-B/kT} \quad (2)$$

where Z is the co-ordination number of the impurity atom. B is an energy term corresponding to the binding energy. However, they have not made specific calculations for the binding of a vacancy to an impurity atom by strain relief. The vacancy impurity binding energies when both electrostatic and elastic interactions are taken into account are of the order of a few tenths of an electron volt.

Dienes and Damask have also obtained suitable rate expressions for annealing when (i) the vacancy-impurity complex is mobile and (ii) the complex is immobile. Annealing of aluminium containing magnesium indicate that the vacancies are bound to magnesium atoms⁴ and that this annealing is speeded up by the above mechanism⁵. The migration and trapping of interstitial iron in Fe-C alloys have been studied⁶. Interstitial iron is supposed to be trapped at immobile carbon quenched into solution.

Dislocations: Two types of dislocations are known; the Edge and Screw dislocations. Both can be described in terms of a suitable Burger's Vector. The potential energy of a solute atom around a dislocation depends upon the distance x from the dislocation and its angular coordinate θ from the slip plane,

$$V(x, \theta) = \frac{A \sin \theta}{x} \quad (3)$$

the constant A is a function of the relative sizes of atoms, the rigidity modulus and Poisson's ratio of the crystal⁷. The foreign atoms form clusters around dislocation similar to the Debye-Hückel cloud in electrolytic solutions. The equilibrium concentration $C(x, \theta)$ of the impurities around an edge dislocation is⁸,

$$C(x, \theta) = C_0 \exp \left[- V(x, \theta)/kT \right] \quad (4)$$

where C_0 is the average concentration at large distances from the dislocation. Any change in the average concentration due to a chemical reaction or due to any changes in temperature, this atmosphere of clusters tends to condense as a line close to the dislocation resulting in nucleus formation.

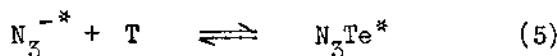
The dislocations can have jogs or kinks which can be electrically charged. Tompkins⁹ suggests that interactions with vacancies and interstitials and with electrons and holes are preferred at charged jogs than at uncharged ones or at unjogged dislocations. At the surface, however, a kink site takes over the function of a jog.

Surface Defects. The surface itself is a major defect in the periodicity of the lattice because of the discontinuity produced by the surface in the periodic field. Tamm¹⁰ suggested that band levels can exist at such surfaces and that these levels could trap electrons. These trapping levels called Tamm levels can exist both at the surface and also in the forbidden energy gap. Further investigations on the surface states of a semi infinite crystal assuming a perturbing potential due to the surface¹¹, a semi infinite perfect crystal, a semi infinite perfect crystal with adsorbed foreign atoms etc.^{12,13} were carried out. In all cases surface states were found to depend upon the lattice parameters, the height of the surface barrier and the potential strength. Though different conclusions have been drawn as to the origin of these surface levels Shockley's treatment concludes that surface states are levels arising from the valence band as well as the conduction band provided that the broadening of the original atomic levels is such that they cross. Kouticky has postulated

that the surface levels could be subsurface states perturbed out of valance band.

In the case of ionic crystals it was shown that a potential difference between surface and bulk was created as a consequence of the difference in the energies of formation of vacancies of different signs¹⁴. An extention of this calculation to NaCl showed a potential difference between surface and bulk of -0.28 V meaning that an excess of +ve ion vacancies will be present at the surface¹⁵.

The role of these surface defects in controlling the mechanisms of decompositions of solids has been stressed in the photolysis of BaN₆¹⁶. Barium azide samples of different surface areas were prepared, (i) by crystallization from aqueous solutions above 52.5°C and (ii) and (iii) by dehydration of BaN₆.H₂O and BaN₆.1.5H₂O respectively. The various processes during the photolysis of barium azide may be formulated as,



where T is a trap and S is a deep surface level. Since in all the cases the degree of decomposition, $\alpha < 10^{-4}$ catalytic influence of metallic specks may be ignored. On a reasoning similar to that employed by Jacobs, Tompkins and Young¹⁷,

$$\text{Rate of photolysis} = \frac{K_4 K_2 K_1 I^2}{K_{-1} K_2 (K_{-1} + K_2 + K_3)} \quad (9)$$

where I is the intensity of exciting radiation. What interests us most is the concentration of surface imperfections, S and the rate of reaction (6), which amounts to a deactivation of excitons. The greater the concentration of the surface imperfections and deeper the surface levels, the less chance of an interaction between the two excitons, trapped or otherwise. The progressive decrease in the rate of photolysis under similar conditions from BaN₆ (i) to BaN₆ (iii) through BaN₆ (ii) (Fig.1) could thus be

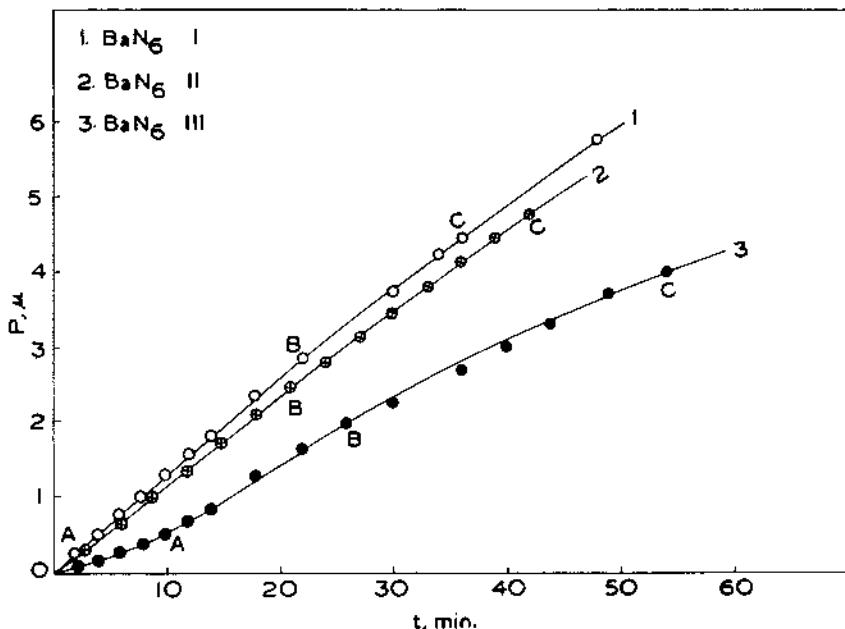


Fig. 1. Photolysis of barium azide samples (i), (ii) and (iii) at a constant intensity of u.v. light.

ascribed to an increase in number of surface impurities which act as deep traps. In the case of BaN_6 (iii) deactivation reaction is probably predominant during the initial stages and continues to be so until most of the traps are filled. Thus K_2 gradually decreases with progress in photolysis, with the result, we find an acceleratory period.

Examples of Reactions: Even a cursory study of the crystalline state is enough to demonstrate the heterogeneity of a "pure" crystalline solid. The heterogeneity is imparted first by virtue of entirely different characteristics of the impurity sites or the defects rendering them virtually different chemical species and secondly the non-uniform distribution of these defects. It would be advisable to illustrate the influence of this heterogeneity in some chemical-reactions by a discussion of the reactions rather than a detailed account of the examples of each type.

Sintering: Simplest chemical reaction that one could visualize is the sintering reaction. Geometric changes which take place when powder compacts are isothermally heated over a period of time can be classified as sintering. Compacting pressure, particle-size, particle size distribution, particle-shape, sintering temperature,

time, atmosphere etc. play an important role in determining the sintering phenomenon. The present understanding of the sintering phenomenon stems from the work of Pines¹⁸ who suggested atomic-diffusion as responsible for sintering. Viscous flow, surface diffusion, bulk diffusion and evaporation condensation mechanisms were all considered by Kuczynski¹⁹ as possible mechanisms of material transport. However, a controversy whether plastic flow or bulk diffusion was responsible for the neck growth between particles persisted until Nabarro²⁰ and Herring²¹ proposed an assumption that lattice vacancies can be discharged at grain-boundaries. The experiments of Alexander of Balluffi²² which showed that pores which are isolated from grain boundaries do not shrink paved the way for the application of grain boundary-vacancy sink concept to sintering²². One can arbitrarily visualize several stages of sintering. The initial stage consists in the elimination of the surface roughness followed by the second stage involving the adhesion of particles. Finally there is a densification period. Coble and Burke²³ who have made detailed studies of sintering in metals and oxides suggest the following stages: (a) A significant change in the shape as the necks grow between the contacting particles. Considerable change in shape occurs and grain growth is completely inhibited. (b) The start of the grain growth indicates the beginning of the second stage. A meta stable phase formed during the second stage has a completely continuous pore phase and maintains an essentially constant shape. (c) Discontinuous grain growth marks the third stage. (d) The pores are nearly spherical in this stage and those near or intersected by grain boundaries continue to shrink without change of shape. In other words this is the final densification period.

Coble²³ made an extensive study of the initial stage of the sintering for the two models: (a) Between two spheres; (b) a sphere and a plate.

The rate of increase in the area of contact was measured. The rate law to be expected for the viscous plastic flow is

$$\frac{x^2}{r} = k_1 t \quad (10)$$

for evaporation condensation

$$\frac{x^3}{r} = k_2 t \quad (11)$$

and for diffusion

$$\frac{x^5}{r^2} = k_3 t \quad (12)$$

where x = radius of contact; r = radius of the sphere or cylinder. In most cases Coble found that the rates of neck growth could be represented by

$$\left(\frac{x}{r}\right)^5 = \left(\frac{BD\gamma a_0^3}{l^3 KT}\right) t \quad (13)$$

B is a numerical constant, D lattice diffusion co-efficient.
 γ = surface-energy, a_0 = lattice spacing.

The expression has been derived by Kuczynski¹⁹ by assuming that neck surface is a vacancy source, the grain boundary is the sink and the bulk diffusion is the predominant-mechanism. Glass²⁴ is reported to sinter by viscous flow mechanism, while evaporation-condensation mechanism is predominant in the sintering of NaCl²⁵.

Vacancy or interstitial diffusion control the rate of sintering in many cases like ZnO, Al₂O₃, MgO. The effect of incorporation of alervalent ions which effect the density of the vacancies has been studied in these cases²⁶. The change in stoichiometry effected by heating in different atmospheres like O₂, H₂, Ar and vacuum can also affect the sintering rate.

In most of the studies on defect in ionic lattices, the concentration of the defects is generally small- << 1% the concentration, normal to semi-conductors. Rustum Roy has investigated the formation of crystalline solutions containing very large concentration of the defects upto 50%. Rustum Roy has introduced a novel method of defining the vacancy concentration in crystalline solids in terms of vacuum. He claims that TiO at 1:1 ratio has a concentration of upto 15% Schottky defects. Since the diffusion rates of these defects in the crystalline solutions are slow, he has prepared solutions with 15% quenched in "vacuum".

By a measurement of density interstitial anions²⁷ in fluorite lattice CaF₂-YF₃, interstitial anions and substituted cations CaF₂-NaF-LnF₃²⁸ and interstitial cations say Ti⁴⁺ in TiO₂ have been characterised. In this work on TiO₂ a large number of magnetic phases have been found to exist. Though reduced rutile in the region TiO_{1.998}-TiO_{2.00} contain predominantly Ti⁴⁺ interstitials, the formation of appreciable concentration of vacancies cannot be ruled out.

Now the general nature of most of the defects are fairly well known as also their influence on mechanism of solid state reactions. However, there are still certain ambiguities, like the one on decomposition of BaN₆. There is a need of a few well chosen systems with new techniques like the esr, x-ray analysis, nmr, field ion microscopy and so on for a clearer understanding of the role of imperfections.

II. SOLID-GAS REACTIONS: OXIDATION OF METALS

The most important difference between homogeneous reactions in the gas phase and heterogeneous reactions in solids is the non availability of reactant molecules according to relatively simple statistical laws. In gas or solution reaction molecules of the product merely exert a "mass action" effect on further reaction between the unchanged reactions. In solid state reactions a relatively immobile product layer usually builds up at the initial point of contact between reactants and the continued interaction depends on a diffusion mechanism by which the reacting species can pass through the product layer²⁹.

The metals in general are reactive towards gases or vapours in the atmosphere around them because of the non equilibrium conditions that exist between most metals and their ambient atmospheres. Extensive investigations have been carried out on the oxidation of metals since oxygen is one of the main components of our atmosphere, and the oxides of essentially all the metals are thermodynamically more stable than the metals themselves.

Imagine a very simple reaction of the type, $2M + O_2 \rightarrow 2MO$, where M is a metal atom. If the oxidation has already progressed to some extent, we have three phases, the metal and the gaseous oxygen phase separated by the oxide phase (Fig.2).

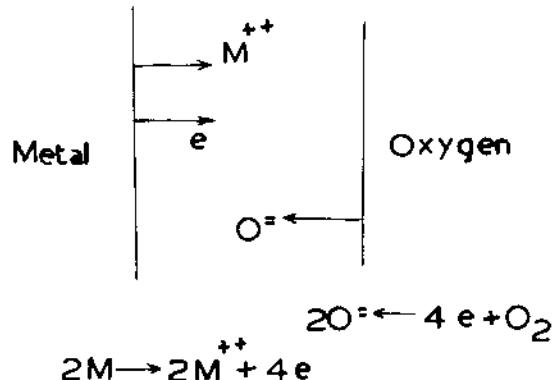


Fig. 2. Interphase and diffusion processes during oxidation of metals.

At the metal-oxide interface we can write, $M \rightarrow M^{++} + 2e^-$, and at the oxide-gas interface, $O_2 + 4e^- \rightarrow 2O^{\ddagger}$. The first reaction yields an excess of electrons and metal ions at the metal-oxide interface. While the second yields an excess of oxide ions at the oxide-gas interface. The concentration gradient thus created helps counter diffusions of the ions through the oxide layer. The

relative rates of diffusion of the cations, anions and electrons determine the relative position and the rate of formation of the new oxide layer. The oxidation process thus comprises of two interface reactions taking place at the metal-oxide and oxide-gas interfaces and one diffusion of electrons and ions. Invariably the interface reactions are assumed to be rapid so that the rate of growth of the oxide layer depends upon the slowest step, i.e. the diffusion of ions and electrons.

Two types of growth have been observed namely thin layer growth and thick layer growth, the demarcation between the two being at a thickness of about 1000 \AA^{30} . The thickness of the oxide layer increases with time in both cases with progress in the reaction. Different laws express the growth rate in the two regions. Kroger³⁰ has summarized the laws as follows:

Thin Layer Growth

1. A cubic law $x^3 = K_c t$

where a third power of the layer thickness is proportional to the time, t .

2. The logarithmic law $x = K_1 - K_2 \ln t$ (14)

3. The reciprocal logarithmic law $\frac{1}{x} = K'_1 - K'_2 \ln t$ (15)

where x is the thickness of the oxide layer at time ' t ' and the K 's are constants. Cabrera and Mott³¹ have explained these laws in terms of electrical charges in the metal-oxide and oxide-vapour interfaces. While Hauffe³² makes use of non neutrality of the oxide phase. Recently, Hauffe³³ has found that the oxidation of nickel between 400°C and 500°C is governed by a fourth power law.

$$x^4 = K_3 t \quad (16)$$

Thick Layer Growth

Two laws have been observed for thick layer growth

1. Linear law $x = bt,$ (17)

as a result of constant rate for layer growth.

2. Parabolic law $x = \sqrt{2At}$ (18)

as a result of decreasing rate for layer growth.

The linear law is observed when diffusion of the reactants

through the oxide layer is very fast, and the interface reaction is rate controlling while parabolic law is found to hold good if diffusion is rate controlling. The interfacial reaction is seldom found to be rate controlling if gaseous oxygen is the tarnishing agent. Only if the source of oxygen for the oxidation process is not molecular oxygen but gases like CO or CO₂, the interface reaction becomes prominent. In all other cases diffusion of the ions through the oxide layer would be rate determining. However, the same system can show both types of behaviour depending upon the conditions and the degree of oxidation. There are instances of non-applicability of any of these rate laws as in the oxidation of beryllium metal in dry oxygen at 0.1 atmosphere and 800°C³⁴.

The theories of thick layer growth particularly for parabolic law have been worked out in detail by Wagner³⁵. His theories are essentially concerned with the bulk properties of the oxide layer. Wagner treated this problem with the help of general thermodynamic properties such as thermodynamic potentials of lattice components. He related the transport properties to the observed specific conductivity, transport numbers etc.

Mott and Gurney³⁶ relate the transport properties to the imperfections and crystalline disorders in the oxide layer and arrive at parabolic law.

Electrons have higher mobilities than ions in oxide lattices. These electrons set up a diffusion potential which takes the charged atomic imperfections along. Since the number of the charged atomic imperfections is relatively high, even with a small mobility an ion current sufficient to keep pace with electron current can be set up³⁰. It is the diffusion of atomic imperfections that determines the rate of diffusion of both neutral and ionic species. The diffusion of the ionised species is more commonly encountered in practice.

Very many tarnishing reactions show the parabolic law either in its simple form,

$$\frac{dx}{dt} = \frac{1}{x} \quad (19)$$

or in the form,

$$\frac{dx}{dt} = \frac{v}{x} \quad (20)$$

Beryllium oxidation³⁴ in oxygen or air in the temperature range 600-950°C follows parabolic law for a period but accelerates thereon. While accelerated oxidations have been attributed to preferential attack at metal grain boundaries, such catastrophic

oxidations are observed even in strain free single crystals of beryllium. (Fig. 3). Elvin and Mackay³⁴ attribute the acceleration in the rate of oxidation of beryllium to the structural defects in the metal surface, presumably the sites of dislocations. Hondros (see discussion in ref. 34) has, however, criticized this explanation as exotic. Condit and Holt³⁷ have reviewed the experimental studies on oxidations of nickel, beryllium, aluminum, zirconium and magnesium making use of radioactive tracers as markers. They conclude that the high temperature oxidation of Ni, Be and probably Al proceeds via cation diffusion through the oxide. Zr may oxidise either by cation or by anion diffusion depending on circumstances. The possibility of oxygen diffusion along grain boundaries in the oxidation of Zr by gaseous oxygen in the region 600–900°C has been indicated³⁸. Under conditions of linear oxidation, Mg is found to oxidise via oxygen diffusion.

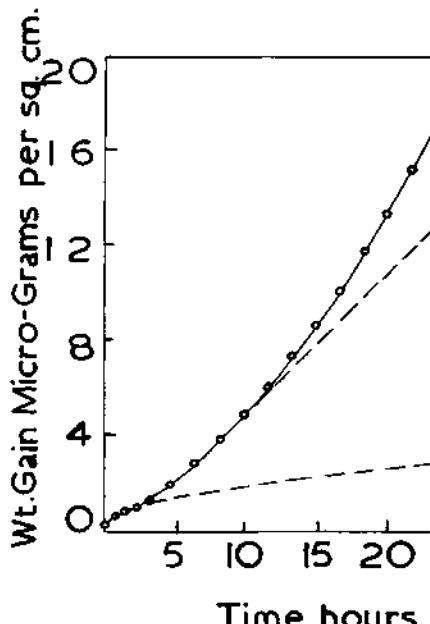


Fig. 3. Oxidation of high purity Be in dry oxygen at 0.1 atmosphere and 800°C

Oxidation of aluminum has posed a serious problem. The oxidation of Al in pure oxygen at 10 mm pressure (Fig. 4) shows two branches OA and AB representing the growth of amorphous alumina, and nucleation and growth of crystalline alumina respectively³⁹. The branch AB shows approximately linear kinetics. The rate of oxidation in this region is assumed to be controlled by ionic diffusion through a preformed amorphous layer. The low energy of

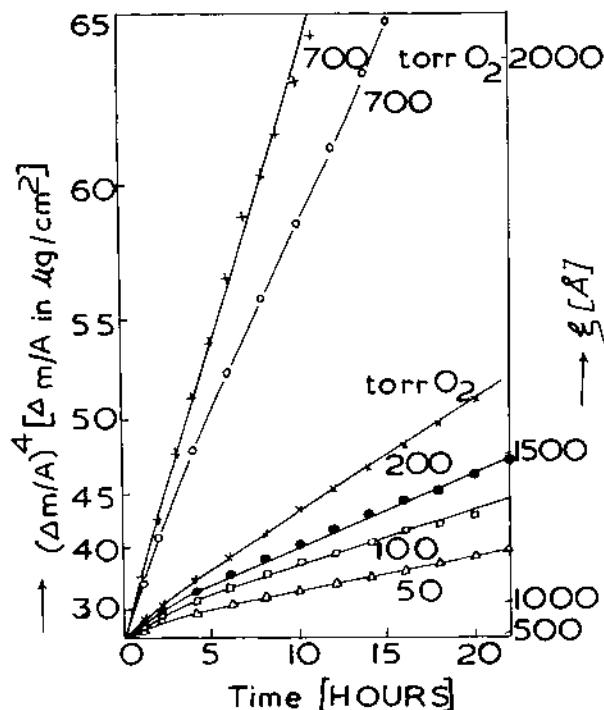


Fig. 4. Oxidation of Nickel in Oxygen

activation for oxidation, 17-19 Kcal mole⁻¹ relative to that for the diffusion of oxygen through Al₂O₃⁴⁰ suggests grain boundary diffusion. Serious doubts have been expressed regarding the species during the oxidation of Al.^{39,41-44}

Silicon and germanium belong to an entirely different class because of the extreme care taken with the preparation of the samples. Si of semiconductor purity is oxidised via oxygen diffusion³⁷. Large concentration of donor complexes of SiO₄ are postulated to account for abnormal conductivities during the oxidation of Si at 400-450°C⁴⁵. Bennett and Tompkins⁴⁶ have studied the low temperature oxidation of Ge by measuring the oxygen uptake. They find that the surface germanium are comparatively mobile as evidenced by the ease of sintering during oxidation, so that place-exchange with underlying Ge atoms is comparatively easy. They believe that incorporation of oxygen takes place at singularities (terraces and kinks) only.

Thin Layer Growth. Parabolic law of growth is not expected during the early stages of oxidation. The applicability of equations 14 to 16 to describe thin layer growth is ascribed to field controlling processes in thin oxide layers. Since electro-

neutrality does not necessarily hold good near the surface. deviations may occur resulting in the build up of a space charge⁴⁷. The growth may depend on the transfer processes at the surfaces as much on diffusion through the layer. The fourth power law⁴⁸ will be discussed in some detail to illustrate the processes responsible for thin layer growth. The rate of oxidation of nickel at different oxygen pressure can be expressed by the fourth power law (Fig. 4). Assuming equilibria at the interfaces, transport of nickel ions through vacancies can be assumed to be rate determining. Assuming,

$$\left| D \frac{dc}{dx} \right| \ll \left| c \mu \frac{dv}{dx} \right|$$

where D and μ are the diffusion coefficient and mobilities of cation vacancies, C the concentration of vacancies and $\frac{dv}{dx}$ the potential gradient, Hauffe⁴⁸ obtains

$$\frac{dx}{dt} = \frac{k}{4} \frac{1}{x^3} \quad (21)$$

which is the fourth power law.

For very thin films, strong field theory of oxidation yields,

$$\frac{dx}{dt} = k' \exp(v/x) \quad (22)$$

where k' contains the probability that the ionic defect necessary for transport is created in the oxide, and $\exp(v/x)$ expresses the effect of potential created by oxygen ions adsorbed at the surface giving a field strong enough to draw the ions across the oxide.

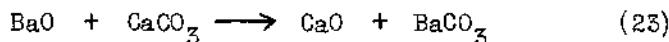
The oxidation of pure metals can be summarized as follows⁴⁹: (i) The first stage is a general chemisorption over the whole surface upto a few layers thickness. (ii) A weakly chemisorbed layer of oxygen is then ultimately formed, which is mobile over the surface and these atoms only suffer localization at sites such as the intersection of dislocations with the surface where metal atoms are most readily ejected. Once nucleation sets in like this growth process proceeds depending upon whether cation diffusion, anion diffusion or phase boundary reaction is rate determining.

III. SOLID-SOLID REACTIONS

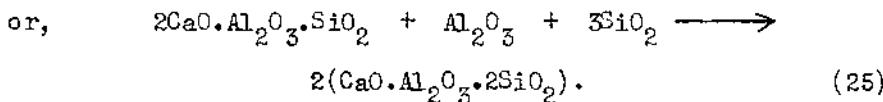
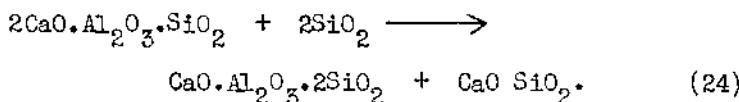
A solid-solid reaction is generally visualized as follows. The two reactant phases are initially in close contact with only a single phase boundary separating them. Once the reaction starts a layer of product is formed between the two reactant phases giving

rise to two different phase boundaries. The product layer progressively becomes thicker until the original reactant crystal lattices are completely consumed. The two important processes involved in this reaction are (i) diffusion of the reactant molecules through the product layer, and (ii) the phase boundary reaction. The relative rates, whatever their mechanisms, determine the overall rate of the reaction.

The reactions between solids apparently are extremely complex. Some of the simpler ones are the exchange reactions of the type



at 345°C. The same reactants yield a mixed oxide at 900°C. The technically important solid-solid reactions, particularly reactions producing ceramics, are extremely complex. For example,



It will be often necessary to draw some generalizations in the simple as well as complex solid state reactions from reasonably correct and accurate experimentation.

Experimental Methods. The experimental methods employed for the study of solid-solid reactions shall be such as to give informations on the structure of the phases. The development of X-ray diffraction analysis of polycrystalline substances has given a great fillip to the understanding of solid-state reactions; e.g. X-ray studies have been employed for finding out the "optimal formation temperature" of $\text{Mg}_2\text{Ba}_2\text{Fe}_{12}\text{O}_{22}$ and $\text{Ni}_2\text{Ba}_2\text{Fe}_{12}\text{O}_{22}$ by Tolksdorf⁵⁰. Another classical sample is the reaction between CaO , SiO_2 and Al_2O_3 to yield $2\text{CaO}\cdot\text{Al}_2\text{O}_5\cdot\text{SiO}_2$ (gehlenite) and/or $\text{CaO}\cdot\text{Al}_2\text{O}_5\cdot 2\text{SiO}_2$ (anorthite). The method requires that the sample be crystalline and in appreciable quantity since X-ray diffraction analysis fails to show small concentrations of less crystalline substances, with the result that negative results from X-ray methods should be treated cautiously.

Use of diffuse reflectance spectra can be helpful to follow the kinetics of solid state reactions. Spinel formations from suitable combinations of MgO , ZnO , NiO and CuO with Al_2O_3 , Ga_2O_3

and In_2O_3 have been investigated⁵¹ by means of diffuse reflectance spectra (Fig. 5). This technique could be a useful tool if the two component oxides and the resulting spinel have clearly distinguishable absorption bands.

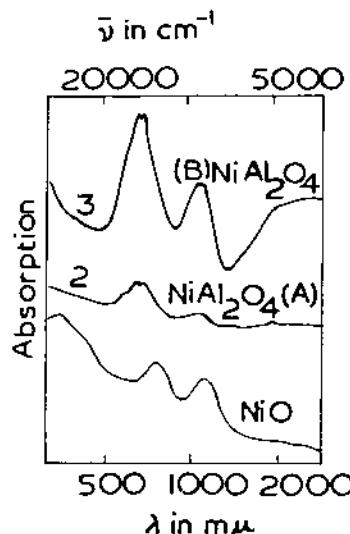
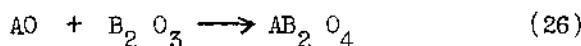


Fig. 5. Diffuse reflectance spectra of NiO , NiAl_2O_4 (A) and NiAl_2O_4 (B)

Mechanism. The mechanism of a solid-state reaction is often discussed with the aid of a diffusion couple arrangement as shown in Fig. 6 for the spinel formation



where A and B are divalent and tervalent metal atoms, respectively⁵²⁻⁵⁴. The original theory of solid-solid reactions in terms of the diffusion of cations through the product layer by a vacancy mechanism was given by Wagner⁵⁴. The basis mechanism is counter diffusion of the cations (Fig. 6c) through the product layer. The two other possibilities involving the diffusion of the anion and one of the two cations are illustrated in Fig. 6a and b. Schmalzried⁵² has discussed the formation of spinels as a function of chemical potentials of the two component oxides and the defect structure of the spinel phase. The structure sensitive diffusion effects are very important because the ionic diffusion, by whatever mechanism, occurs through a product layer which has been freshly formed often under such conditions as are not conducive to effective crystallization. Such lattices are bound to be highly imperfect, the defect concentration being much in excess of

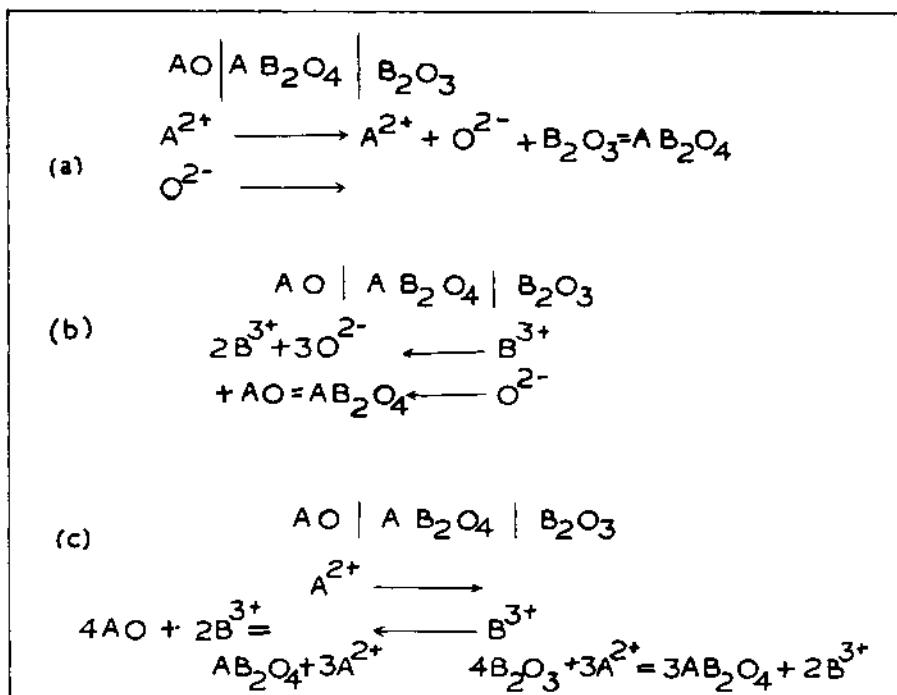
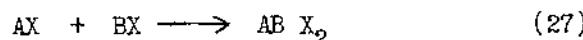


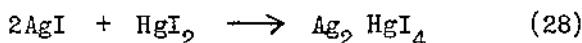
Fig. 6. Diffusion couple to illustrate the mechanism of reaction of spinel formation

equilibrium concentration. The increased defect concentration results in enhanced or accelerated diffusion of the ions through the defect lattice. Although the exact nature of diffusion in such highly defective lattices is not clearly understood, experimental evidence points to considerable mobility of lattice units²⁹. Mobility along crystal surfaces and grain boundaries should be considerable at lower temperatures where bulk diffusion would be small. Therefore, any kinetic analysis of solid state reactions should take into account particle size, porosity and surface structure.

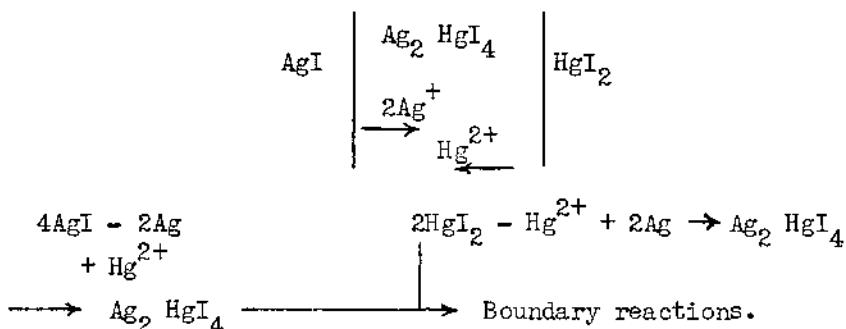
Kinetic Equations: Wagner^{30, 54} gave a quantitative thermodynamic theory similar to the one he proposed for the oxidation of metals. He considered a simple addition reaction of the type



He made the following assumptions: (i) the rate of reaction is determined by diffusion through the product layer (and not by surface effects); (ii) only the cations are mobile with no change in their valency; (iii) the mobility of the cations corresponds to their transport fraction in an electrolysis experiment; and (iv) there is no appreciable electronic conduction. On this basis, for the reaction,



the total reaction rate can be determined as the sum of the two individual particle currents.



The total rate

$$f = - \int \frac{[A^+] + [B^+]}{1 + \frac{Z_B}{Z_A} \frac{\nu_B}{\nu_A}} \times \frac{Z_B}{l} \nu_B [B^+] d\mu \quad (29)$$

where $[A^+]$ and $[B^+]$ are the concentrations of ions, Z_A , Z_B and ν_A , ν_B are the corresponding valencies and mobilities respectively, l is the diffusion length. On integration under two extreme cases $\nu_B \gg \nu_A$ and $\nu_A \gg \nu_B$ the reaction rate comes out to be proportional to transport number t_A and t_B respectively and the difference in chemical potentials of the Ax phase and Bx phase respectively. In the special case of $\nu_A = \nu_B$

$$f = \frac{1}{l e^2} \frac{[A^+] + [B^+]}{[B^+]} t_A t_B \sigma \left[\mu(\text{Ax})_{\text{II}} - \mu(\text{Ax})_{\text{I}} \right] \quad (30)$$

Koch and Wagner obtained a value of 2.1×10^{-11} equivalents per cm per sec for $k (= f l)$ for the formation of Ag_2HgI_4 .

Wagner's theory has been extended to cover also cases in

which anions are also mobile^{52,53,55} (Fig. 6a & b). In the case of a & b the anions also take part in diffusion. In the extreme case the diffusion of one cation is compensated completely by the diffusion of anions. Attempts have been made to distinguish between these three reaction mechanisms by studying the location of the phase boundary. Since diffusion is the rate controlling step in solid-solid reactions, according to Wagner's theory.

$$\text{the rate of reaction} = \frac{dx}{dt} \approx \frac{C}{x} \quad (31)$$

or a parabolic law must be observed. A linear law should be applicable for a phase boundary reaction controlled process.

Among the other factors which affect the reaction kinetics particle size and distribution comes foremost. The first analysis was made by Jander⁵⁶ who assumed that the surface of the minor component is completely and continuously covered with the particles of the excess component as if the former particles were immersed in a sea of the latter. If the two reactants consist of fine powders this assumption will be invalid. Komatsu⁵⁷ has obtained an equation

$$1 - (1 - \alpha)^{2/3} = kt \quad (32)$$

where α is the fraction of reaction completed at time t . k is a rate constant and $k = K(T, a, x)$. a is the ratio of the radius of the two components in powder form, and x is the mixing ratio. The model assumed here is that the reaction starts at the contact points between particles and that the number of contact points play a very important role. Komatsu has verified his equation by applying to systems of $\text{CaCO}_3\text{-MoO}_3$, $\text{BaCO}_3\text{-SiO}_2$, Si-CuCl and Si-PbCl_2 .

Solid Solutions. The formation of solid solutions presents an entirely different problem. Schmalzried⁵⁸ has discussed in detail the reactions between two compounds forming either extended solid solutions or one or several product phases. He has considered the problem from two angles; (i) the experimental determination of the reaction rate and morphology as a function of all independent variables and (ii) the calculation of reaction rates and a prediction of the morphology under a given set of independent variables in terms of known thermodynamic and transport properties. The simplest cases of solid solution formation are KCl-RbCl , KBr-AgBr , NiO-MgO , $\text{CoAl}_2\text{O}_4\text{-MgAl}_2\text{O}_4$ etc. The main problem in calculations for reactions of this kind is the determination of interdiffusion coefficient as function of composition. The interdiffusion of ions manifests experimentally as the Kirkendall effect (see the chapter on Diffusion by V. Ramakrishna). If the solid solution is 'ideal', and the partial molal volumes constant, there

will be no shift of the markus and so Kirkendall effect in the usually defined sense of the term should be absent. He redefines Kirkendall effect under these conditions as the local production or annihilation of lattice defects. This concept has been applied to the NiO-MgO system. The interdiffusion coefficient for hole conduction

$$\bar{D} = (1 - x_{\text{MgO}}) \cdot D_{\text{Mg}} + x_{\text{MgO}} \cdot D_{\text{Ni}} \quad (33)$$

x_{MgO} = mole fraction of MgO and D_i is the self diffusion coefficient. On comparison with (a) analogous system MgO-CoO and NiO-CoO, where the mole fraction of the cation vacancies depend on the composition as

$$x_v = \frac{x_h}{2} = x^0 \left[\exp \left\{ - \beta (1 - x_{\text{CoO}}) \right\} \right] \quad (34)$$

and (b) experimental results the following conclusion is made that there is an exponential relation between defect concentration and the mole fraction of MgO in the solid solution NiO-MgO. This conclusion is confirmed by a comparison of the self diffusion coefficient of Ni in NiO and the extrapolated interdiffusion coefficient. Fig. 7 shows the calculated and measured⁵⁸ diffusion profiles. Such studies have been carried out on NiO-CoO⁵⁹ and MgO-FeO⁶⁰.

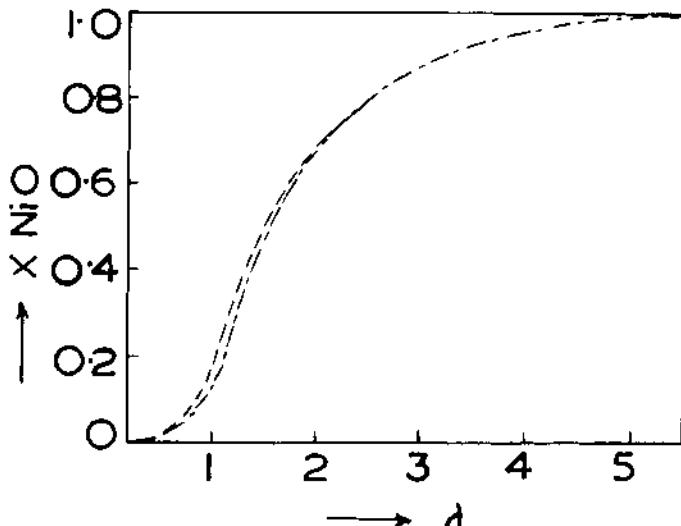


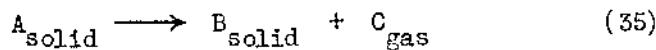
Fig. 7. Calculated and measured diffusion profiles⁵⁸ for Ni in NiO.

Schmalzried has extended his consideration to reactions giving rise to several product phases of negligible homogeneity range.

IV. THERMAL DECOMPOSITION OF SOLIDS

A tendency for many solid state chemists studying decomposition of solids is to extend some of the concepts of gas phase kinetics to solid state kinetics without realizing the full gravity of innumerable difficulties encountered during such investigations. Some of these difficulties are connected with the assumption of the equilibrium between the activated and non activated molecules⁶¹. The assumption, however, justified in gas phase reactions, becomes untenable in solid state kinetics.

In this review of solid state decompositions, I shall concern myself with isothermal decomposition kinetics and avoid as far as possible to non isothermal methods. To simplify the treatment further, the reactions of the type



are discussed. Excellent treatises on solid decompositions have been published^{62,63}. The proceedings of the conferences on the "Reactivity of Solids" supply valuable additional information on specific decompositions.

The experimental results obtained in isothermal decompositions of solids are (a) the fractional decomposition, α as a function of time at different constant temperatures, and (b) direct observations of the form and distribution of solid reaction product supported by X-ray and electron diffraction studies. The shape of α -t curves encountered commonly in solid state decompositions is shown in Fig. 8 in a generalized way. Decompositions of particular solids may show some or all of the fractions shown in the figure.

Extensive theoretical as well as experimental studies of the acceleratory period have been carried out. The kinetic results are generally fitted into either exponential functions or power functions of time.

$$\alpha = f(e^{kt}) \quad (36)$$

$$\text{or } \alpha = f(kt^n) \quad (37)$$

In order to get a good fit of the data into these equations,

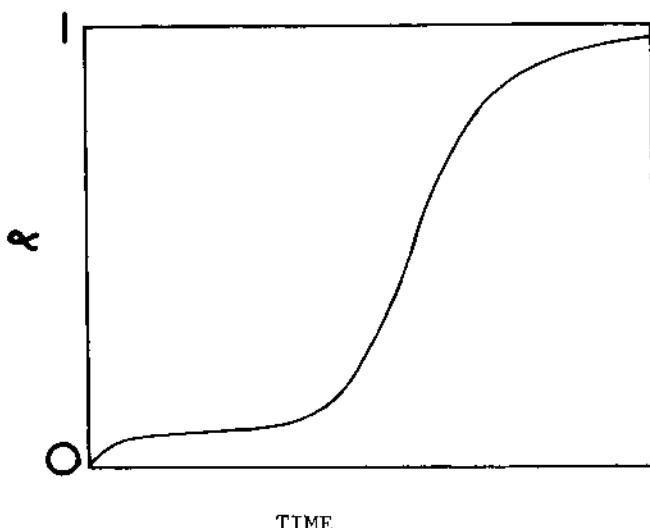


Fig. 8. A generalized α -t curve for thermal decomposition of solids

a t_0 term is introduced. However, a justification for introduction of this adjustable parameter is often hard to give. The decay period can be analyzed less satisfactorily than the acceleratory period. A contracting volume expression is found to fit the data in many cases. An example showing all these features is the decomposition of nickel oxalate⁶⁴. Decomposition of finely ground silver azide⁶⁵ shows no acceleratory period. The only feature that BaN_6 decomposition does not show is the initial rapid reaction⁶⁶.

All the theories for the decomposition of a solid to yield a second product phase is to assume that the reaction starts at certain discrete and energetically favoured positions in the reactant matrix. These positions are called nucleus forming sites.

The Laws of Nucleus Formation. A number of sites having abnormal environment such as crystal imperfections or sites of highly localized decomposition may favour the further decomposition of the solid due to the relatively low energy of activation required at those sites. The low energy of activation may be the consequence of the enormous strain energy associated with such sites. These sites called "germ nuclei" grow to become "growth nuclei" as decomposition progresses. The three laws of nucleation are: (a) A constant number of nuclei exist in the lattice

and they grow as the reaction progress. (b) The number of nuclei at time, t ,

$$N_t = N_0 [1 - e^{-kt}] \quad (38)$$

where N is number of germ nuclei at $t = 0$ and k is a rate constant. (c) In addition to the exponential law there is a power law. The theory has been worked out quite satisfactorily by Bagdassarian⁶⁷. Powers greater than unity are obtained if several successive decompositions are required to form a stable nucleus. If β such steps are involved,

$$N_t = \frac{N_0 (k_1 t)^\beta}{\beta!} \quad (39)$$

or in general terms we get the expression

$$N_i = \frac{N_0 (K_1 T)^i}{i!} \exp (-k_1 t) \quad (40)$$

where K_1 is the rate constant for nucleation during induction period and N_i is the number of nuclei of size ' i ' at time t .

Another approach involving the reaction between active intermediaries at potential nucleus forming sites has been proposed by Thomas and Tompkins⁶⁶. The minimum requirement for the motion of the active intermediaries is transfer of energy localized in an excited state of the crystal. Young⁶³ has criticized this mechanism on the ground that this "mechanism requires that the undecomposed fraction of the crystal shall continue to store energy in mobile excited species at a constant rate per unit volume throughout the region of validity of nucleation law. In a real crystal this possibility is not an attractive one".

Nucleation has been assumed in almost all thermal decompositions and experimentally proved in many to be a surface phenomenon associated essentially with surface defects. Catalytic effect of adsorbed ions and oxygen on the decomposition of silver oxalate has been adduced as evidence to stress that nucleation is a surface event⁶⁸.

The concentration of germ nuclei and hence the rate of nucleus formation can be remarkably altered by changing the conditions of preparation of the material or by irradiating the sample with ultraviolet light or high energy radiations. Acceleration of subsequent decomposition due to prior irradiation has been observed

in a number of cases such as alkaline earth azides⁶⁹, silver oxalate⁷⁰, potassium bromate⁷¹ and lead azide⁷². Preirradiation does not seem to have any appreciable effect on subsequent thermal decomposition of alkali azides⁷³ and silver permanganate⁷⁴. Experiments on mercury fulminate have indicated an increase in the rate constants for decomposition due to irradiation⁷⁵.

Laws of Nucleus Growth. The laws governing the growth of the nuclei have been discussed in detail by Jacobs and Tompkins⁶² and Young⁶³. The growth process though essentially an interface phenomenon, differs appreciably between diffusion controlled reactions and interface reactions.

Polanyi and Wigner⁷⁶ provided one of the first equations for the kinetics of decomposition of solids. Assuming that the energy of activation for the growth process is less than that for nucleation, the probability of any ion or radical acquiring this energy is

$$\gamma = e^{-E/kT} \quad (41)$$

where E is the energy of activation and k is the Boltzmann constant. The rate of interface reaction is

$$-\frac{dN}{dt} \cdot N \cdot \gamma e^{-E/kT} \quad (42)$$

where N is the total number of radicals at the interface. The interface grows by this process of the radicals reacting at the interface and consequently transferring an electron to the nucleus while evolving the gaseous product. The rate of growth is assumed to be constant once the nucleus attains the critical size. It is, however, more reasonable to expect that the rate of linear growth of the nucleus will be vectorial in the case of polyhedral nuclei though a constant growth rate could be justified for spherical nuclei. There is some evidence of very slow growth in BaN₆⁷⁷ though the evidence is only inferential. The question of slow growth of small nuclei is complex not merely because it is very difficult to observe them experimentally but it is also very difficult to separate it from time dependent nucleation. An unusual case of nucleus growth is the dehydration of barium styphnate hydrate. Dehydration is highly sensitive to crystal size and the kinetic laws are $\alpha = kT$ for large crystals and $\alpha = k_o^2(t + t_o)^2$ for smaller crystals⁷⁸. Slow growth of nuclei appears to be a phenomenon frequently encountered in the decomposition of metal azides. Young has summarized the minimum requirements for slow growth in azides in his excellent treatise⁶³, in terms of the effect of microstrain on the thermal properties of the reactant. The slow growing nucleus itself can be a cluster

of slowly growing nuclei. Once these nuclei touch, the cluster undergoes an abrupt increase in effective size becoming large enough to grow normally^{63,79}. This process is probably responsible for cessation of slow growth and the onset of normal growth.

General Rate Equations. The rate equations proposed for solid decompositions assume a constant rate of growth which is so in most cases once the nuclei attain the critical size. Though several rate equations for discrete nucleus formation and growth have been developed by various people (refer to 63) equations of greater consequence are the Avrami-Erofeev^{80,81} and Prout-Tompkins⁸² equations.

Avrami assumes a random distribution of germ nuclei which are consumed by activation of the terms into growth nuclei and by impingement of germs into the advancing phase of growth nuclei. This model leads to

$$-\ln(1-\alpha) = (kt)^n \quad (43)$$

An expression of a similar form has been given by Erofeev⁸¹. Erofeev has further extended this equation to cover reactions of solids with different specific surfaces to obtain⁸³,

$$\frac{d\alpha}{dt} = k\alpha^2(1-\alpha)^y \quad (44)$$

where x and y depend upon the value of n in the equation. $x = \frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \dots$ and $y = 0.77, 0.70, 0.66, \dots$ correspond to values of $n = 2, 3, 4, \dots$. The validity of equation (44) is verified for the kinetics of polymorphic transformation of NH_4NO_3 . Further studies on the decomposition of KMnO_4 and $(\text{NH}_4)_2\text{Cr}_2\text{O}_7$ by the same author shows that the surface plays different roles in different substances. Since equation (44) is only an approximation of the more general equation

$$\frac{d\alpha}{dt} = n k^{1/n} \left[-\ln(1-\alpha)^{(n-1)/n} \right] (1-\alpha) \quad (45)$$

the former has limited applicability. It may not be suitable for initial stages of decomposition.

This conventional theory of decomposition of solids assumes discrete formation and growth of nuclei in a very real way. Each process is associated with a different rate constant. Allnatt and Jacobs⁸⁴ feel that this is a realistic procedure if a nucleus can be formed as a result of a single reaction step. If a nucleus is formed by a single step and the reaction is auto-catalytic in

character, $k_{\text{growth}} > k_{\text{formation}}$. If the nucleus is formed, on other hand, at special or unique sites whose geometry or the special feature is destroyed during the decomposition $k_{\text{formation}} < k_{\text{growth}}$. The reaction follows a simple exponential law in the latter case. Allnatt and Jacobs have extended Avrami's theory to incorporate the multistage nucleation process as envisaged by Bagdassarian⁶⁷. By postulating different rates for each step of nucleation process until the nucleus contains p -atoms and constant rate thereafter they obtain,

$$-\log(1-\alpha) = 6\sigma \left[V_\infty^{-1} \right] n_0(\alpha) G^3$$

$$\sum_{l=0}^{p-1} B_{pl} \frac{(-k_1 t)^{p+3}}{(p+3)!} \quad \frac{1}{k_1}$$

Thus $-\log(1-\alpha)$ behaves like t^{p+3} when $k_1 t \ll 1$. Apparently this extension appears to be the only significant advance on the Avrami-Erofeev equation.

A simple power law is used to describe the kinetics of decomposition of metal azides.

$$\alpha = k (1 + t_0)^n \quad (46)$$

Depending upon the time dependence of the rate of nucleus formation n can be equal to 3, 4 or 6 for three dimensional nucleus growth. In the case of silver oxalate⁷⁰ the power law is applicable with $3.5 < n < 4$. n decreases to 3 as a result of preirradiation.

Decomposition of Azides. The best studied metal azides are the alkaline earth azides, of which barium azide is the most thoroughly investigated. Wisclln⁸⁵ was the first to make a systematic study of the decomposition of barium azide. She found that the overall reaction could be expressed by equation (46) with $6 < n < 8$ and $t_0 = 0$. Mott⁸⁶ gave a theoretical interpretation in terms of nucleation by bimolecular aggregation of mobile interstitial barium ions assuming that barium azide is an ionic as well as a photoconductor. The rate determining step was assumed to be the thermal excitation of an electron from the full band at the metal azide interface to the conduction band of the metal. Thomas and Tompkins⁶⁶ observed a million fold discrepancy between the observed and calculated values of ionic conductivities for Mott's mechanism to be applicable. They proposed an interfacial mechanism for nucleus formation. Tompkins⁹ has proposed a revised model for

nucleation involving a process of enhanced diffusion along specific channels associated with imperfections. The effect of the degree of imperfection on barium azide decomposition has been studied⁷⁷. Circular nuclei are formed evenly over the surface according to a cubic law in the case of most imperfect crystals while time dependent nucleation did not occur in the case of most perfect crystals while time dependent nucleation did not occur in the case of most perfect crystals. In the latter circular nuclei became diamond shaped nuclei and grew at twice the rate of circular nuclei. Under certain conditions the acceleratory period obeyed cubic law $\alpha = k(t - t_0)^3$.

Spath and Torkar⁸⁷ confirmed that the nuclei grow as diamonds in the decomposition of single crystals of barium azide according to an exponential law. One of their findings is that 75% of the azide decomposes to a nitride. They have proposed a chain mechanism for the formation of nitride. Prout and Moore⁹¹ found that for the same decomposition Avrami-Erofeev equation with $n = 4$ was applicable for the acceleratory period and contracting sphere equation could describe the decay period.

Decomposition of calcium and strontium azides have also been studied. The unusual feature of decomposition of calcium azide⁶⁹ is that the Arrhenius plot shows a break at about 97°C for annealed samples. This lead Tompkins and Young to believe that annealing occurs rapidly within the temperature range of decomposition. The cubic law is found to describe the decomposition of both calcium and strontium azides⁸⁸.

Decomposition of Carbonates. Many carbonates decompose reversibly to oxides and carbon dioxide. Several carbonate produce oxides which do not readily absorb carbon dioxide or do so very slowly. Certain other produce oxides which are described as "active" because of their unusually high activity. Calcium carbonate decomposition in carbon dioxide is a classic case⁸⁹.

The mechanism of decomposition is pictured as a two stage phenomenon. Initially carbon dioxide leaves the calcium carbonate unit cell and the residue of calcium oxide assumes a metastable rhombohedral configuration similar to the original carbonate. The second step of the reaction is the formation of a highly crystallized cubic calcium oxide from the "active" calcium oxide. The active calcium oxide thus acts as a bridge to this reversible reaction. This intermediate metastable oxide which forms at the surface during decomposition can reform into the ordered cubic oxide lattice or can combine with CO₂ to form calcium carbonate depending upon the partial pressure of carbon dioxide. However, the velocity of decomposition of CaCO₃, or the recombination of calcium oxide and carbon dioxide does not bear a simple relation

to $|p - p_0|$ where p_0 is the equilibrium pressure of carbon dioxide at any given temperature. This non ideality is again ascribed to the highly active metal oxide nuclei.

Decomposition of cadmium carbonate shows similar behaviour for repeated runs on the same sample. Very early work on decomposition of zinc carbonate shows that it is characterized by a slow reverse reaction. The decomposition is described by

$$\frac{d\alpha}{dt} = k_1 \alpha^{2/3} - k_2 \alpha^{2/3} p - x \quad (47)$$

where p is the pressure of carbon dioxide and x is the retardation due to the presence of an "active" zinc oxide. Relatively more recent studies⁹⁰ on decomposition of calcite, magnesite and related compounds are described by a contracting envelope equation

$$(1 - \alpha)^{1-1/n} = 1 - k(t + \Delta t_0) \quad (48)$$

where n is 0.58 ± 0.06 and Δt_0 is a zero time correction. One characteristic feature of most decomposition of carbonates is that the energy of activation is of the same order of the standard heat of decomposition.

Summary. In summary we can conclude that the basic mechanisms of decomposition of almost all solids involves nucleation and propagation of the nuclei. Tompkins⁹ feels that "this general type of mechanism is rarely in dispute, and, at least, in many solid decompositions no alternative analysis which has sufficient flexibility to account for the variable kinetics and for rates ranging from times of half change of a few seconds to many days has been advanced".

REFERENCES

1. J.H. de Boer, "Reactivity of Solids", Elsevier, Amsterdam, 1961, p. 1.
2. J.A. Hedvall, "Solid State Chemistry", Elsevier, Amsterdam, 1966.
3. G.J. Dienes and A.C. Damask, "Reactivity of Solids", Ed. by J.H. de Boer, Elsevier, Amsterdam, 1965, p. 227.
4. J. Takamura, K. Okazaki and I.G. Greenfield, J. Phys. Soc. Japan, 18, 78 (1963).

5. T. Kino, S. Kabemato and S. Yoshida, *J. Phys. Soc. Japan*, 18, 85 (1963).
6. A.C. Damask and A.R. Arndt, "Reactivity of Solids", Ed. by G-M. Schwab, Elsevier, Amsterdam, 1965, p. 238.
7. W.J. Dunning, "Chemistry of the Solid State", Ed. by W.E. Garner, Butterworths, London 1956.
8. A.H. Cottrell and B.A. Bilby, *Proc. Phys. Soc. London*, A62, 49 (1949).
9. F.C. Tompkins, *Pure Appl. Chem.*, 9, 387 (1964).
10. I. Tamm, *Phys. Z. S.S.R.*, 1, 733 (1932).
11. J. Koutecky, *Phys. Rev.*, 108, 13 (1957).
12. E. Aerts, "Solid State Physics in Electronics and Telecommunications", Academic Press, London, 1960, p. 628.
13. W. Shockley, *Phys. Rev.*, 56, 317 (1939).
14. K. Lehovec, *J. Chem. Phys.*, 21, 1123 (1946).
15. J.D. Eshelby, C.W.A. Newey, P.L. Fratt and A.B. Liddiard, *Phil. Mag.*, 3, 76 (1958).
16. F. C. Tompkins and S.R. Yoganarasimhan, Unpublished results.
17. P.W.M. Jacobs, F.C. Tompkins and D.A. Young, *Disc. Faraday Soc.*, 28, 234 (1959).
18. E.Y. Pines, *J. Tech. Phys.*, *S.S.S.R.*, 16, 737 (1946).
19. G.C. Kuczynski, *J. Metals*, 1, 2 (1949); *Trans. A.M.I.E.*, 185, 165 (1949); *J. Appl. Phys.*, 21, 632 (1950).
20. F.R.N. Nabarro, "Reformation of Crystals by the Motion of Single Ions", *Phys. Soc. London*, 1948.
21. C. Herring, *J. Appl. Phys.*, 21, 436 (1950).
22. B.H. Alexander and R.W. Balluffi, *J. Metals*, 2, 2119 (1960).
23. R.L. Coble, *J. Am. Ceram. Soc.*, 41, 55 (1958).
24. J.H. Cannon and J. White, *Trans. British Ceram. Soc.*, 55, 82 (1956).

25. J.B. Moser and D.H. Whitmore, *J. Appl. Phys.*, 31, 488 (1960).
26. P.A. Marshall, D.P. Emight and W.A. Weyl, "Reactivity of Solids", Gothenberg, 1954, p. 273; J.T. Jones, P.K. Maitra and I.B. Cutler, *J. Am. Ceram. Soc.*, 41, 353 (1958); G.K. Layden and M.C. Macquire, *ibid*, 42, 89 (1959).
27. J.H. Short and R. Roy, *J. Phys. Chem.*, 67, 1860 (1969).
28. D.M. Roy and R. Roy, *J. Electrochem. Soc.*, 111, 424 (1964).
29. A.J.E. Welch, "Chemistry of the Solid State", Ed. by W.E. Garner, Butterworths, London, 1956, p. 297.
30. F.A. Kroger, "The Chemistry of Imperfect Crystals", North Holland, Amsterdam (1964).
31. N. Cabrera and N.F. Mott, *Rept. Progr. Phys.*, 12, 163 (1949).
32. H.J. Engel, K. Hauffe and B. Ilschner, *Z. Electrochem.*, 58, 382, 478 (1954).
33. K. Hauffe, L.D. Pethe, R. Schmidt and S.R. Morrison, *J. Electrochem. Soc.*, 115, 456 (1968).
34. G. Erven Jr. and T.L. Mackay, "Recativity of Solids", Ed. by G.M. Schwab, Elsevier, Amsterdam, 1965, p. 290.
35. C. Wagner, *Z. Phys. Chem.*, B21, 25 (1932); B32, 447 (1936).
36. N.F. Mott and R.W. Gurney, "Electronic Processes in Ionic Crystals", Dover, New York, 1964.
37. R.H. Condit and J.B. Holt, "Reactivity of Solids", Ed. by G.M. Schwab, Elsevier, Amsterdam, 1965, p. 334.
38. J.B. Lightstone and J.P. Femsler, "Reactivity of Solids", Johnwiley, New York, N.Y. 1969, p. 615.
39. R.J. Breakspere, *ibid*, p. 363.
40. Y. Oishi and W.D. Kingery, *J. Chem. Phys.*, 33, 480 (1960).
41. M.J. Dignam and W.R. Fawcett, *J. Electrochem. Soc.*, 113, 663 (1966).

42. M.J. Dignam, W.R. Fawcett and H. Bohne, *J. Electrochem. Soc.*, 113, 656 (1966).
43. J.J. Randall and W.J. Bernard, *J. Appl. Phys.*, 35, 1317 (1964).
44. P.E. Doherty and R.S. Davies, *ibid*, 34, 619 (1963).
45. H. Reiss, *ibid*, 30, 141 (1959).
46. M.J. Bennett and F.C. Tompkins, "Reactivity of Solids", Ed. by J.H. de Boer, Elsevier, Amsterdam, 1961, p. 154.
47. K. Hauffe and B. Ilschner, *Z. Electrochem.*, 58, 467 (1964).
48. K. Hauffe, "Reactivity of Solids", John Wiley, New York, N.Y., 1969, p. 311.
49. F.S. Stone, "Reactivity of Solids", Ed. by J.H. de Boer, Elsevier, Amsterdam, 1961, p. 14.
50. W. Tolksdorf, "Reactivity of Solids", Ed. by G-M. Schwab, Elsevier, Amsterdam, 1965, p. 297.
51. F.S. Stone and R.J.D. Tilley, *ibid*, p. 587.
52. H. Schmalzried, *ibid*, p. 204.
53. C. Kooy, *ibid*, p. 21.
54. C. Wagner, *Z. Physik. Chem.*, B34, 309 (1936).
55. H. Schmalzried, *Z. Phys. Chem. (Frankfurt)*, 33, 111 (1962).
56. W. Jander, *Z. anorg. allgem. Chem.*, 163, 1 (1927).
57. W. Komatsu, "Reactivity of Solids", Ed. by G-M. Schwab, Elsevier, Amsterdam, 1965, p. 182.
58. H. Schmalzried, "Reactivity of Solids", John Wiley, New York, N.Y., 1969, p. 551.
59. D.H. Whitmore, *Am. Ceram. Soc. Bull.*, 43, 337 (1964).
60. I.B. Cuttler, *J. Am. Ceram. Soc.*, 48, 95 (1965).
61. W. Gomes, *Nature*, 192, 865 (1961).
62. P.W.M. Jacobs and F.C. Tompkins, "Chemistry of the Solid State", Ed. by W.E. Garner, Butterworths, London, 1956.

63. D.A. Young, "Decomposition of Solids", Pergamon Press, Oxford, 1966.
64. P.W.M. Jacobs and T. Khureishy, "Reactivity of Solids", Ed. by J.H. de Boer, Elsevier, Amsterdam, 1961, p. 353.
65. B.E. Bartlett, F.C. Tompkins and D.A. Young, Proc. Royal Soc., A246, 206 (1958).
66. J.G.N. Thomas and F.C. Tompkins, ibid, A210, 111 (1951).
67. C. Bagdassarian, Acta Phys. Chem., U.S.S.R., 20, 441 (1945).
68. Z.G. Szabo and E. Biro-sugar, Z. Electrochem., 50, 869 (1956).
69. F.C. Tompkins and D.A. Young, Trans. Faraday Soc., 61, 1470 (1965).
70. F.C. Tompkins, ibid, 44, 206 (1948).
71. J. Jach, "Reactivity of Solids", Ed. by J.H. de Boer, Elsevier, Amsterdam, 1961.
72. E.G. Prout and F.C. Tompkins, Trans. Faraday Soc., 43, 148 (1947).
73. P.W.M. Jacobs and F.C. Tompkins, Proc. Royal Soc., A215, 265 (1952).
74. E.G. Prout and F.C. Tompkins, Trans. Faraday Soc., 42, 482 (1946).
75. B.E. Bartlett, F.C. Tompkins and D.A. Young, J. Chem. Soc., 3323 (1956).
76. M. Polanyi and E. Wigner, Z. physik. Chemie, A139, 439 (1928).
77. B.E. Bartlett, F.C. Tompkins and D.A. Young, Nature, 179, 365 (1957).
78. F.C. Tompkins and D.A. Young, Trans. Faraday Soc., 52, 1245 (1956).
79. D.A. Young, Nature, 204, 281 (1964).
80. M. Avrami, J. Chem. Phys., 7, 1103 (1939); 8, 212 (1940); 9, 177 (1941).

81. B.V. Erofeev, Compt. Rend. Akad. Sci. U.R.S.S., 52, 511 (1946).
82. E.G. Prout and F.C. Tompkins, Trans. Faraday Soc., 40, 488 (1944).
83. B.V. Erofeev, "Reactivity of Solids", Ed. by J.H. de Boer, Elsevier, Amsterdam, 1961, p. 273.
84. A.R. Allnatt and P.W.M. Jacobs, Canadian J. Chem., 46, 111 (1968).
85. A. Wischin, Proc. Royal Soc., A172, 314 (1939).
86. N.F. Mott, ibid, A172, 325 (1939).
87. K. Torkar and H.E. Spath, Monatsch., 99, ... (1968).
88. W.E. Garner and L.E. Reeves, Trans. Farady Soc., 51, 694 (1955).
89. E.P. Hyatt, I.B. Cutler and M.E. Wadsworth, J. Am. Ceram. Soc., 41, 76 (1958).
90. H.T.S. Britton, S.J. Gregg and G.W. Winsor, Trans. Faraday Soc., 48, 63 (1952) and early references quoted in 63.
91. E.G. Prout and D.J. Moore, Effects of High Energy Radiation on Inorganic Substances. Special Technical Publication No. 400, A.S.T.M., 1966.

BAND AND TRANSPORT THEORIES IN SOLIDS

J. M. Honig

Department of Chemistry

Purdue University, Lafayette, Indiana, U.S.A.

Within the necessarily limited lecture schedule at the Winter School the author attempted to cover a few fundamental topics rather than to provide a rapid-fire survey of the entire area. The coverage is highly selective and the approach extremely elementary. It is hoped that this material will stimulate the reader to study the subject matter again on his own with the degree of rigor which it deserves, and to proceed to other areas not covered in these notes.

I. INTRODUCTION TO BAND STRUCTURE CONCEPTS: THE PARTICLE-IN-THE-BOX MODEL

Prior to a consideration of transport phenomena it is necessary to study the energy states appropriate to electrons in a crystal. In zero order approximation we can treat an electron in a solid as a particle in a box. To simplify matters as much as possible, we restrict ourselves to the case of one dimension. Within the box the potential is a constant and outside the box it approaches infinity. The Schrödinger equation in such a case takes the form,

$$\left\{ \frac{-\hbar^2}{2m} \frac{\partial^2}{\partial x^2} + V - \epsilon \right\} \Psi = 0 \quad (1)$$

with $V = V_0$ when $0 < x < L_x$ and $V \rightarrow \infty$ otherwise; here L_x is the length of the box. The remaining symbols in (1) are assigned the standard interpretation. It is readily verified by substitution that the solution of Eqn. (1) is of the form

$$\Psi = e^{\frac{ik_x x}{\hbar}}. \quad (2)$$

Substituting Eqn. (2) in Eqn. (1) and taking $V_0 = 0$ (This is justified, since ϵ is known only to within an additive constant; the above choice simply shifts all the energy levels by a constant quantity) we obtain for the energy,

$$\epsilon = \frac{\hbar^2 k_x^2}{2m}. \quad (3)$$

To determine the significance of k_x , let us consider a free particle with momentum p . The kinetic energy of this particle is given by

$$\epsilon = p_x^2 / 2m. \quad (4)$$

On comparing Eqns. (3) and (4) we can set $p_x = \hbar k_x$. Moreover, this is consistent with the de Broglie relationship $p = h/\lambda$ if we define $\lambda = 2\pi/\lambda$. Accordingly, k_x is to be interpreted as the x component of the wave number vector \underline{k} .

We may now write the energy of the particle as

$$\epsilon = \hbar^2/2m \lambda^2. \quad (5)$$

The values taken by k or λ are not arbitrary but are severely restricted by the requirement that the particle be confined to the box. Using standard quantum mechanical procedures it may be shown that λ must satisfy the condition

$$L_x = n_x \lambda / 2 \quad n_x = 1, 2, 3, \dots \quad (6)$$

On substituting Eqn. (6) in (5) we obtain the energy of a particle in a box

$$\epsilon = \frac{n_x^2 \hbar^2}{8m L_x^2}. \quad (7)$$

Comparison of Eqn. (3) with Eqn. (7) shows that k_x is restricted to the set of values

$$k_x = \frac{n_x \pi}{L_x}, \quad n_x = 1, 2, 3, \dots \quad (8)$$

Hence, according to this approximation, the particle can be assigned to only one of the following quantized set of energy levels,

$$\epsilon = \frac{\hbar^2 k_x^2}{2m}, \text{ where } k_x = \frac{n_x \pi}{L_x}, n_x = 1, 2, 3, \dots \quad (9)$$

Though this crude approximation points out the essential properties of an electron in a solid, it does not take into account the effect of the lattice potential on the electron, and hence is not really satisfactory.

II. THE KRONIG-PENNEY MODEL

A modification which brings out the effect of atomic periodicity in the lattice on the electrons to a reasonable extent, has been worked out by Kronig and Penney. Here the periodicity properties are simulated by taking the square well potential described in Fig. 1, where we set $V = 0$ for intervals such as $0 < x < b$ and $V = V_0$ for intervals such as $b < x < c$. We shall show later that the solution of the Schrödinger equation in such a case must have the general form ($k \equiv k_x$)

$$\Psi = e^{ikx} u(x) \quad (10)$$

where $u(x)$ is a function which must be periodic with lattice periodicity $a = b + c$. This type of wave function is called a Bloch function. On now substituting for Ψ and V in the Schrödinger equation we obtain a differential equation in $u(x)$ because the exponential terms cancel out:

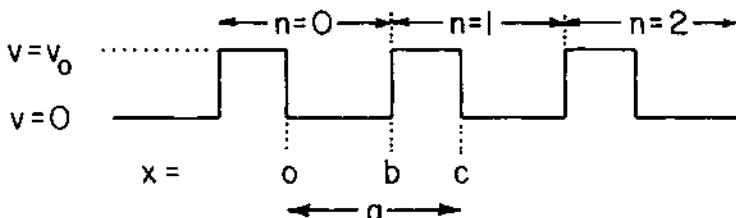


Fig. 1 Kronig - Penney lattice potential.

$$u'' + 2iku' + \underline{k} [\varepsilon - v - k^2/\underline{k}] u = 0 \quad (11)$$

where $\underline{k} \equiv 2m/h^2$. The energy ε will fall into one of the two ranges, namely (a) $0 < \varepsilon < V_o$, or (b) $\varepsilon > V_o$. Taking up case (a) first we define two parameters,

$$\alpha \equiv \sqrt{\underline{k}\varepsilon} \quad , \quad \beta \equiv \sqrt{\underline{k}(V_o - \varepsilon)} . \quad (12)$$

[Notice that β is imaginary in the energy range (b)] and we rewrite the Schrödinger equation in terms of these parameters for the regions (i) $0 < x < b$ and (ii) $b < x < c$ as follows:

$$\begin{aligned} u'' + 2iku' + (\alpha^2 - k^2) u &= 0, & (i) \\ u'' + 2iku' - (\beta^2 + k^2) u &= 0. & (ii) \end{aligned} \quad (13)$$

By direct substitution it may be verified that these equations have the following solutions:

$$u_k(x) = A_n e^{i(\alpha - k)x} + B_n e^{-i(\alpha + k)x} \quad (i) \quad (14)$$

and $u_k(x) = C_n e^{(\beta - ik)x} + D_n e^{-(\beta + ik)x} . \quad (ii)$

This is an infinite set of equations, since the index which labels the particular hill and valley in the periodic potential of Fig. 1 can range from $-\infty$ to ∞ . The solutions for a given k differ only in the coefficients A_n , B_n , C_n , and D_n . However, we must take account of the periodicity requirement associated with the translational symmetry. This imposes the condition that the solution at every translationally equivalent point in the lattice must be the same. Hence the various constants are related as follows:

$$\begin{aligned} A_n &= A_0 e^{-i(\alpha - k)na} & C_n &= C_0 e^{-(\beta - ik)na} \\ B_n &= B_0 e^{i(\alpha + k)na} & D_n &= D_0 e^{(\beta + ik)na} \end{aligned} \quad (15)$$

In the above a is the repeating distance and n is the integer discussed earlier. Eqn. (15) reduces the infinite set of equations to just four. Next we must couple the solutions of region (i) with those of region (ii). This can be done by imposing the conditions,

$$\begin{aligned} u_k(0^+) &= u_k(0^-) & u_k(b^+) &= u_k(b^-) \\ u'_k(0^+) &= u'_k(0^-) & u'_k(b^+) &= u'_k(b^-), \end{aligned} \quad (16)$$

where 0 and b are points of discontinuity in the potential.

These conditions insure that the solution and its first derivative are continuous across the boundaries of the potential. The above four conditions yield four relations involving the four coefficients. They are:

$$\begin{aligned} A_0 + B_0 - C_0 - D_0 &= 0, & (i) \\ i(\alpha - k) A_0 - i(\alpha + k) B_0 - (\beta - ik) C_0 + (\beta + ik) D_0 &= 0, & (ii) \\ e^{i(\alpha - k)b} A_0 + e^{-i(\alpha + k)b} B_0 - e^{-(\beta - ik)c} C_0 - e^{(\beta + ik)c} D_0 &= 0 & (iii) \\ i(\alpha - k)e^{i(\alpha - k)b} A_0 - i(\alpha + k)e^{-i(\alpha + k)b} B_0 \\ -(\beta - ik)e^{-(\beta - ik)c} C_0 + (\beta + ik)e^{(\beta + ik)c} D_0 &= 0. & (iv) \end{aligned}$$

(17)

These equations in the four unknowns, A_0 , B_0 , C_0 and D_0 have no solution other than the trivial one $A_0 = B_0 = C_0 = D_0 = 0$ unless the determinant of coefficients vanishes. On setting up this determinant, expanding it, and collecting terms one obtains the following relation:

$$\cosh(\beta c) \cos(\alpha b) + \left(\frac{\beta^2 - \alpha^2}{\alpha \beta} \right) \sinh(\beta c) \sin(\alpha b) = \cos ak. \quad (18a)$$

Defining the left hand side as $F(\alpha, \beta)$ the above may be rewritten as

$$F(\alpha, \beta) = \cos ak. \quad (18b)$$

For real values of k the right hand side of Eqn. (18) remains within the limits -1 and +1. Hence we require that

$$-1 < F(\alpha, \beta) < +1. \quad (18c)$$

Since by Eqn. (12), α and β are functions of the energy ε ,

$F(\alpha, \beta)$ is also a function of ϵ . A plot of $F(\alpha, \beta)$ vs. ϵ is shown in Fig. 2 for the indicated values of the parameters. We note that there are ranges of ϵ where F satisfies (18c) and other ranges of energy for which $|F(\alpha, \beta)| > 1$.

Clearly, the solutions in these latter domains must be discarded as incompatible with the boundary conditions imposed earlier.

We turn next to case (b) where $\epsilon < V_0$. Here we define the parameters

$$\alpha \equiv \sqrt{k\epsilon} \quad \text{and} \quad \delta \equiv \sqrt{k(\epsilon - V_0)} . \quad (19)$$

By a procedure quite similar to the above, one obtains an equation analogous to (18a), namely

$$\begin{aligned} G(\alpha, \beta) \equiv & \cos(\delta c) \cos(\alpha b) - \left(\frac{\alpha^2 + \delta^2}{2\alpha\delta} \right) \sin(\delta c) \sin(\alpha b) \\ & \approx \cos ak \end{aligned} \quad (20)$$

Again, for real k , we have to require that $-1 < G(\alpha, \beta) < 1$, and discard that portion which violates the later requirement. A plot of $G(\alpha, \delta)$ vs ϵ is shown in Fig. 2 for the indicated values of the parameters. The two portions of the curve corresponding to $F(\alpha, \beta)$ and $G(\alpha, \delta)$ merge smoothly.

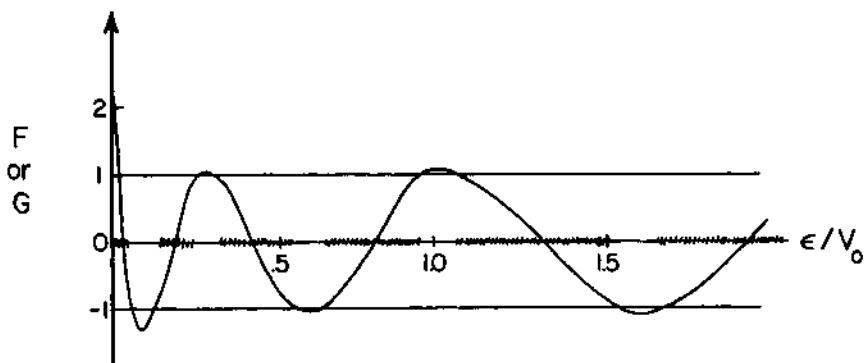


Fig. 2. Plot of $F(\alpha, \beta)$ and $G(\alpha, \delta)$ vs. ϵ / V_0 for $V_0 b^2 k = 144$, and $c/b = 1/24$.

From the above discussion it emerges that acceptable solutions must be restricted to energy values ϵ for which $-1 < F(\alpha, \beta)$, $G(\alpha, \delta) < +1$. These various ϵ ranges are shown by cross-hatched markings in the diagram. Accordingly, we may distinguish between regions of permitted energies, termed bands, and regions of forbidden energies, termed gaps. Thus, we have demonstrated for a rather primitive model that a periodic lattice potential leads to band formation. Note that as ϵ increases the bands become wider and the gaps become narrower; this is a quite general feature encountered in the more sophisticated derivations of the band structures in solids.

III. BLOCH FUNCTIONS FOR THE CIRCULAR CHAIN

We now refine the above model by taking into account more realistically the presence of N atoms spaced regularly along a linear chain. In order to avoid edge effects, the chain is to be bent in a ring, with the initial ($n = 0$) and terminal ($n = N-1$) atoms separated by the normal separation distance a , as shown in Fig. 3. All the various atoms in the chain are now indistinguishable.

Designate by x the distance along the circular chain and let $n = 0, 1, 2, \dots, N-1$ designate the various atoms, each separated from its neighbors by the distance a . First, the fact that the point x is physically indistinguishable from the equivalent point $x' = x - na$ along the chain, must be taken into consideration in setting up the wave functions. Thus, we must require that

$$|\Psi(x)|^2 = |\Psi(x - na)|^2. \quad (21)$$

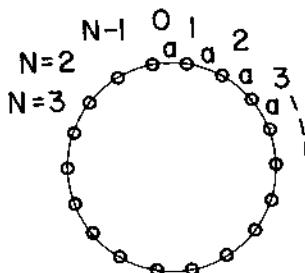


Fig. 3. One-dimensional circular lattice chain.

It is easily seen that the above requirement is met by setting

$$\Psi(x - na) = e^{i\phi} \Psi(x) \quad (22)$$

where ϕ is a phase angle. As a corollary of this we write, $\Psi(x - na - n'a) = e^{i(\phi_n + \phi_{n'})} \Psi(x)$, which shows that the phase angle must be advanced proportionally with na . Hence we must set ϕ proportional to na , or $\phi = kna$, where k is now a proportionality constant. It follows that

$$\Psi(x - na) = e^{ikna} \Psi(x). \quad (23)$$

A special situation arises when we have completely gone around the chain once; for, having left point x and returned to the identical point, now labelled $x + Na$, the wave functions themselves (not just their norms squared) must be identical. Therefore, we require that

$$\Psi(x) = \Psi(x + Na) = e^{-ikNa} \Psi(x) \quad (24)$$

which in turn means that $e^{-ikNa} = 1$. This requirement is satisfied whenever

$$kNa = 2\pi s$$

$$\text{or } k = \frac{2\pi s}{Na}, \text{ where } s = 0, \pm 1, \pm 2, \dots \quad (25)$$

$$\text{or } k = \frac{2\pi s}{L}$$

where L is the total chain length.

We see that as a result of the above, the proportionality constant k is quantized. As will be shown later, the quantity $\hbar k$ may again be interpreted as the momentum p of an electron subjected to the potential of the lattice under study.

According to de Broglie's relation, $p = \hbar/\lambda$; hence, $k \equiv k_x = 2\pi/\lambda$ is once again to be interpreted as a wave number vector component, quantized according to the prescription of Eqn. (25).

The requirement $\Psi(x - na) = e^{ikna} \Psi(x)$ leads to the following general form for $\Psi(x)$:

$$\Psi(x) = e^{ikx} u_k(x) \text{ with } u_k(x - na) = u_k(x). \quad (26)$$

To prove this assertion, substitute (26) into (23) and note that

$$\begin{aligned}\psi(x - na) &= e^{ik(x - na)} u_k(x - na) = e^{ikx} e^{-ikna} u_k(x) \\ &= e^{-ikna} \psi(x) \text{ as required.}\end{aligned}$$

Observe that in setting up (26) we have just demonstrated Bloch's theorem concerning wave functions for electrons subjected to periodic potentials.

If in (26) we set u_k equal to a constant then we recover the particle-in-the-box problem; in Section I we had already demonstrated that the momentum of the free particle is given by the relation $p = \hbar k$.

IV. BAND STRUCTURE FOR A CIRCULAR CHAIN IN THE TIGHT-BINDING APPROXIMATION

In the preceding section we set up the general form of the wave function for an electron moving in the periodic potential of a one-dimensional lattice, and we showed how the quantization of the wave number vector arises. We now take up the related problem: What can one say about the energy of the system studied in the preceding section, and what quantities are needed to specify this energy? These matters can only be answered by indirect methods.

Let $\Phi(x - qa)$ be an atomic orbital centered about the q^{th} atom in the chain. We now postulate, by analogy to molecular orbital theory, that the wave function for the system may be taken to be a linear combination of the atomic orbitals; thus, we set

$$\Psi = \sum_{q=1}^N c_q \Phi(x - qa). \quad (27)$$

Since all atoms in the chain are equivalent, the various combination coefficients c_q in (27) can differ at most by a phase factor whose magnitude is unity. Hence we must set

$$c_q = C e^{i\theta_q} = C e^{ikqa} \quad (28)$$

where C is a quantity yet to be determined. The identification of θ_q with kqa was considered in the last section. Thus,

$$\Psi = C \sum_{q=1}^N e^{ikqa} \Phi(x - qa). \quad (29)$$

Substituting (29) in the Schrödinger wave equation

$$\left\{ -\frac{\hbar^2}{2m} \nabla^2 + V - \varepsilon \right\} \Psi = 0 \quad (30)$$

and setting $V = -e U(x)$ where U is the periodic lattice potential, we find

$$\sum_{q=1}^N C e^{ikqa} \left\{ -\frac{\hbar^2}{2m} \nabla^2 - e U(x) - \varepsilon \right\} \Phi(x - qa) = 0 \quad (31)$$

Here $U(x)$ must be a function which is periodic in x with the periodicity of the lattice.

Since $\Phi(x - qa)$ must itself satisfy the Schrödinger equation for an isolated atom we also must be able to write

$$\left\{ -\frac{\hbar^2}{2m} \nabla^2 - e V(x - qa) - \varepsilon_0 \right\} \Phi(x - qa) = 0 \quad (32)$$

where $-e V(x - qa)$ is the potential of the q^{th} atom when isolated, and ε_0 is the corresponding energy. On now eliminating the second derivatives between (31) and (32) we obtain the relation

$$C \sum_{q=1}^N e^{ikqa} \left\{ (\varepsilon - \varepsilon_0) + e [U(x) - V(x - qa)] \right\} \Phi(x - qa) = 0. \quad (33)$$

Next, operate on both sides of the above equation with the quantity

$$e^{-ikq'a} \int_{-Na/2}^{Na/2} dx \Phi^*(x - q'a).$$

This yields a relation of the form

$$C \sum_{q=1}^N e^{ika(q - q')} \left[(\varepsilon - \varepsilon_0) S_{q'q} - Q_{q'q} \right] = 0 \quad (34)$$

$$\text{where } S_{q'q} = \int \Phi^*(x - q'a) \Phi(x - qa) dx, \quad (35)$$

is an overlap integral, and $Q_{q'q}$ is called a transfer or generalized exchange integral, given by

$$Q_{q',q} \equiv -e \int \phi^*(x - q'a) [U(x) - V(x - qa)] \phi(x - qa) dx. \quad (36)$$

Eqn. (34) represents a set of simultaneous linear equations. In order to solve these we now introduce the following simplifying assumption: (i) We neglect the overlap term completely except when $q = q'$. This assumption is the essence of the tight binding scheme in zero order approximation. Mathematically we state this condition by setting $S_{q',q} = \delta_{q',q}$ where $\delta_{q',q}$ is equal to zero for $q \neq q'$ and 1 for $q = q'$. The latter condition presupposes that the atomic wave functions have already been normalized.

(ii) We further assume that, $Q_{q',q} = 0$ unless $q' = q + 1, q, q - 1$. Accordingly, Eqn. (36) reduces to the following cases:

$$Q_{qq} = -e \int \phi^*(x - qa) [U(x) - V(x - qa)] \phi(x - qa) dx.$$

Using a change in variable $x' = qa$ the above may be written as

$$Q_{qq} = -e \int \phi^*(x') [U(x') - V(x')] \phi(x') dx' \equiv C \quad (37)$$

where C denotes the coulomb integral involving the difference in potential energy ($U - V$) weighted by the function $|\phi(x)|^2$. Also, we need the integrals,

$$Q_{q\pm 1,q} = -e \int \phi^*(x - (q \pm 1)a) [U(x) - V(x - qa)] \phi(x - qa) dx$$

which may be rewritten as

$$Q_{q\pm 1,q} = -e \int \phi^*(x \pm a) [U(x) - V(x)] \phi(x) dx \equiv A_{\pm} \quad (38)$$

where A_{\pm} denotes the exchange integral.

In light of the above, equation (34) can now be written as

$$(\varepsilon - \varepsilon_0 - C) - A_+ e^{-ika} - A_- e^{ika} = 0 \quad (39)$$

From the symmetry of the chain we see that, $A_+ = A_- \equiv A$. Hence, Eqn. (39) becomes

$$\varepsilon = \varepsilon_0 + C + 2A \cos ka \quad (40)$$

which solves the problem at hand; we have now specified the energy ε in terms of the parameters ε_0 , C , A , and a , and functionally through the dependence on k .

The wave function Ψ given by (29) can be rewritten as

$$\Psi = N^{-\frac{1}{2}} \sum_q e^{ikqa} \phi(x - qa) \quad (41)$$

where the term $N^{-\frac{1}{2}}$ ensures normalization in the approximation of setting $S_{q'q} = \delta_{q'q}$.

Several points should be carefully noted here: (a) As argued earlier, k is itself quantized, being restricted to the values $k = k_s = 2\pi s/L$, where s is an integer. However, when L is very large, successive values of k are spaced so closely that k may be regarded as essentially continuous. Thus, strictly speaking ϵ is also quantized, but in the case of large L it may be regarded as a quasi-continuous function of k . (b) From (40) we note that ϵ must fall between the limits (ϵ_0 , C , and A turn out to be negative):

$$\epsilon = -|\epsilon_0| - |C| + 2|A|. \quad (42)$$

The ϵ values falling in this range constitute a band. We obtain only one band because only one kind of atomic orbital was initially included in the overall wave function Ψ . Had more atomic wave functions been utilized, further bands would have been found with intervening gaps. The more complete wave function is given by

$$\Psi = \sum_{q=1}^N \sum_{l=1}^n c_{ql} \phi_l(x - qa) \quad (43)$$

where l indexes the different types of wave functions that are to be taken into account. (c) Returning to Eqn. (40), it is seen that ϵ is a sinusoidal function of k . With such duplication there is no need to consider k values outside a fundamental interval within which all possible ϵ values are specified. Thus we restrict ourselves to that set of k values for which $-\pi < ka \leq \pi$. This region is called the central or first Brillouin Zone. The analogous restrictions become more complicated in three dimensions. (d) Observe that the permitted k values serve as the appropriate quantum numbers for the quantization of ϵ . The atomic quantum numbers are of no direct relevance here except as they enter Eqn. (43) through the index l .

It is of interest to determine what happens in the neighborhood of $k = 0$, where, on expansion of the cosine function we obtain,

$$\epsilon = -|\epsilon_0| - |C| - 2|A| [1 - k^2 a^2 / 2] = \epsilon_c + |A| k^2 a^2 \quad (44)$$

with $\varepsilon_0 \equiv -|\varepsilon_0| - |C| - 2|A|$. On comparing this result with Eqn. (9) we see that in the neighborhood of $k = 0$ the electron behaves as a free particle with a kinetic energy $\varepsilon - \varepsilon_0 = \hbar^2 k^2 / 2m$. Thus, we may identify the effective mass of the electron with the quantity, $m \equiv \hbar^2 / [2|A|a^2]$. We see that the effective mass is determined by the characteristics of the crystal through the parameters a and A , the latter being specified by Eqn. (38).

At the top edge of the band $ka \rightarrow \pm\pi$. When we define $k' \equiv (\pi/a) - k$, which measures the separation between ka and the end point π , the energy becomes

$$\begin{aligned}\varepsilon &= -|\varepsilon_0| - |C| - 2|A| \cos(\pi - k'a) = -|\varepsilon_0| - |C| + 2|A| \cos(k'a) \\ &= -|\varepsilon_0| - |C| + 2|A| [1 - k'^2 a^2 / 2] = -|\varepsilon_0| - |C| + 2|A| - |A| k'^2 a^2.\end{aligned}\quad (45)$$

Again we see that $\varepsilon \propto k'^2$ so that once more the particle appears to be free at the top edge of the band. However, a conceptual problem arises here; for, on comparison with Eqn. (9) we note that $m = -\hbar^2 / [2|A|a^2]$, meaning that this free particle has a negative effective mass. This gives rise to the notion of a hole. We shall see later on why this situation arises and how to circumvent these conceptual difficulties.

It should be noted that while these results have all been developed on the basis of a very simple model they can be shown to hold much more generally. One of the most important conclusions reached here is that when there is appreciable overlap between wave functions so as to render $|A|$ large, the one electron states for the individual atoms coalesce into an equal number of band states for the solid considered as a unit. The larger $|A|$ is the wider the bands become, as is seen from Eqn. (42), and the smaller is the effective mass of the carriers at the band edges.

V. THE PAULI EXCLUSION PRINCIPLE AND FERMI-DIRAC STATISTICS

We base our description of transport properties in solids on the free-electron model, according to which the kinetic energy of the electron is given by

$$\varepsilon = \hbar^2 k^2 / 2m = \hbar^2 (k_x^2 + k_y^2 + k_z^2) / 2m \quad (46)$$

together with the quantization conditions

$$k_\lambda = 2\pi n_\lambda / L_\lambda \quad \lambda = x, y, z \quad n_\lambda = 0, \pm 1, \pm 2, \dots \quad (47)$$

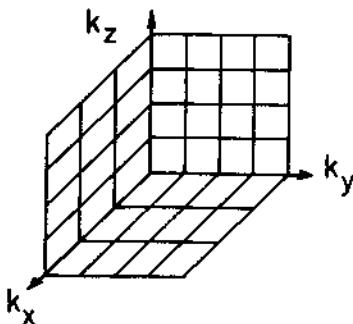


Fig. 4. k -space with projections of orthogonal sets of grids projected onto planes $k_x = k_y = k_z = 0$.

Each permitted triplet of integers (k_x, k_y, k_z) defines a point in reciprocal (k) space spanned by the cartesian coordinate axes shown in Fig. 4. All such points may be joined by an orthogonal set of grids defining a set of rectangular parallelepipeds (or cubes when $L_x = L_y = L_z \equiv L$, as in Fig. 4) which fill up all of k -space. Each unit is of volume $(2\pi)^3/L_x L_y L_z = 8\pi^3/V$, where V is the volume of the crystal under study. Neglecting edge effects, the entire assembly of rectangular parallelepipeds may now be shifted uniformly, so that every point originally at a vertex is now in the center of one of the parallelepipeds. Hence each permitted quantum state as specified by (47) may be considered as 'occupying' a space of volume $8\pi^3/V$, termed the exclusion volume, in reciprocal space.

When the $\epsilon(k)$ dependence is spherically symmetric, as in (46), what is of interest is the number of exclusion volumes located between two concentric spherical shells of radius k and $k + dk$. This corresponds to the number of quantized states whose energy varies in the range ϵ to $\epsilon + d\epsilon$, as specified by Eqn. (46). The above quantity, multiplied by two to allow for spin degeneracy, is the number of electrons that can have energies falling in this range. Denoting the latter quantity by $h(k)dk$ we find that

$$h(k)dk = 2(4\pi k^2 dk) / (8\pi^3/V). \quad (48)$$

Per unit volume of crystal, we obtain a density of states $g(k) = h(k)/V$ given by,

$$g(k)dk = (k^2/\pi^2)dk. \quad (49)$$

It is customary to specify this quantity per unit of energy. Setting $g(\epsilon)d\epsilon \equiv g(k)dk$ and employing Eqn. (46) we obtain

$$g(\epsilon) = g(k)/(d\epsilon/dk) \approx (2\pi^2)^{-1} (2m/h^2)^{3/2} \epsilon^{1/2}, \quad (50)$$

showing that the density of electron states which have energies in a range $d\epsilon$ about a value ϵ increases with energy as $\epsilon^{1/2}$.

Eqn. (50) represents the maximum density of available states. The fraction of these actually occupied by electrons is given by the well known Fermi-Dirac distribution function:

$$f(\epsilon) = \frac{1}{1 + e^{(\epsilon - \mu)/kT}} \quad (51)$$

which is derived by the author in the Chapter on Order-Disorder Theory. Here, μ is a parameter, known as the Fermi energy or Fermi level whose significance will be clarified as we proceed. In (51) k is the Boltzmann constant.

A plot of $f(\epsilon)$ vs. ϵ is shown in Fig. 5 for various temperatures, where we set $\mu = 3$ ev. It is seen that at 0 K $f(\epsilon)$ is a step function, with $f(\epsilon) = 1$ for $\epsilon < \mu$ and $f(\epsilon) = 0$ for $\epsilon > \mu$. The effect of raising the temperature is to round off the corners of the distribution function slightly. Thus, at 300 K the change in f from 1 to 0 occurs over an energy interval in the range of approximately 0.2 ev centered about the value of 3 ev.

This distribution function is a reflection of the Pauli exclusion principle. Since no more than two electrons with

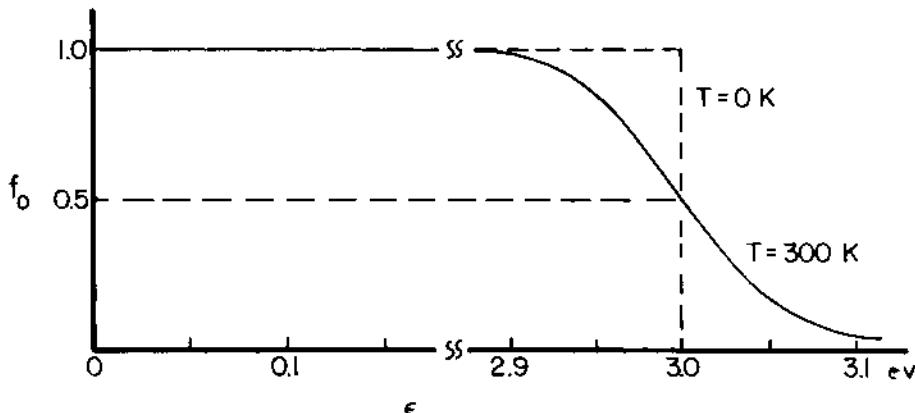


Fig. 5. Fermi-Dirac distribution corresponding to $\mu = 3$ ev and $T = 0, 300$ K.

opposite spin may be placed in a given quantum state, $N/2$ exclusion volumes are needed to accommodate N electrons; for large N the energy of the states required for placement of the last electrons may become very high.

Fig. 5 furnishes us with two interpretations for μ . The mathematical one involves the statement that for $\epsilon = \mu$, f assumes the value of $1/2$. Physically, μ is in the nature of a cutoff energy for metals. All quantum states for which $\epsilon < \mu$ are filled at 0 K; all states for which $\epsilon > \mu$ are empty.

The value of μ is related to the charge carrier density as follows: Since $g(\epsilon)d\epsilon$ is the density of available states of energy ϵ and since $f(\epsilon)$ is the probability that these states are actually occupied, the density of carriers having energies in the range ϵ to $\epsilon + d\epsilon$ is $f g d\epsilon$, and the total carrier density is given by

$$n = \int_0^\infty f(\epsilon)g(\epsilon)d\epsilon = \frac{1}{2\pi^2} \left(\frac{2m}{\hbar^2}\right)^{3/2} \int_0^\infty \frac{\epsilon^{1/2}d\epsilon}{1+e^{(\epsilon-\mu)/kT}}. \quad (52)$$

It is possible to simplify the above by the expedient of carrying out the integration at 0 K; then $f(\epsilon) = 1$ so long as $\epsilon < \mu_0$ (here the subscript denotes the value of μ determined at $T = 0$). Accordingly,

$$n = \frac{1}{2\pi^2} \left(\frac{2m}{\hbar^2}\right)^{3/2} \int_0^{\mu_0} \epsilon^{1/2} d\epsilon = \frac{1}{3\pi} \left(\frac{2m}{\hbar^2}\right)^{3/2} \mu_0^{3/2} \quad (53)$$

which may be solved for

$$\mu_0 = (3\pi^2 n)^{2/3} (\hbar^2/2m). \quad (54)$$

VI. HEAT CAPACITY OF A METAL

We now derive an expression for the heat capacity of a degenerate metal, using the above model. First it is necessary to obtain the average electronic kinetic energy by summing the product of (a) the energy ϵ , with (b) the density of states, and (c) the probability of occupancy of these states lying in the range ϵ to $\epsilon + d\epsilon$. At 0 K, the states in the region between $\epsilon = 0$ and $\epsilon = \mu_0$ are fully occupied, all others being empty. Hence the total kinetic energy of an electron at 0 K is equal to

$$U_0 = \int_0^{\mu_0} \epsilon g(\epsilon) d\epsilon. \quad (55)$$

Substituting for $g(\epsilon)$ from Eqn. (50) we obtain

$$U_0 = \int_0^{\mu_0} \frac{1}{2 \pi^2} \left(\frac{2m}{\hbar^2} \right)^{3/2} \epsilon^{3/2} d\epsilon \quad (56)$$

and on evaluating the integral one finds that

$$U_0 = \left(\frac{2m}{\hbar^2} \right)^{3/2} \mu_0^{5/2} / \pi^2; \quad (57)$$

comparison of (57) and (53) allows us to write

$$U_0 = 3/5 n \mu_0. \quad (58)$$

So much for the situation at 0 K. At $T > 0$, the total kinetic energy of an electron is equal to

$$U = \int_0^{\infty} \epsilon g(\epsilon) f(\epsilon) d\epsilon. \quad (59)$$

Substituting for $g(\epsilon)$ and $f(\epsilon)$ from Eqns. (50) and (51) respectively yields,

$$U = \frac{1}{2 \pi^2} \left(\frac{2m}{\hbar^2} \right)^{3/2} \int_0^{\infty} \frac{\epsilon^{3/2}}{1 + e^{(\epsilon - \mu)/kT}} d\epsilon. \quad (60)$$

The above integral is an example of the so-called Fermi-Dirac integral which has the general form

$$F_{\frac{1}{2}} \equiv \int_0^{\infty} \frac{\epsilon^{\frac{1}{2}} d\epsilon}{1 + e^{(\epsilon - \mu)/kT}}. \quad (61)$$

For highly degenerate materials, the Fermi distribution function deviates only slightly from the step function shown in Fig. 5. In these circumstances it may be shown that $F_{\frac{1}{2}}$ is in first approximation given by the following relation, $\frac{1}{2}$ valid for $\mu/kT \gg 1$:

$$F_{\frac{1}{2}} \approx \frac{1}{(1 + 1)} \mu^{(1 + 1)} \left\{ 1 + \frac{(\pi kT)^2}{6 \mu^2} \frac{1(1 + 1)}{2} + \dots \right\}. \quad (62)$$

It is clear that for use in Eqn. (52) we must set $\underline{l} = 1/2$ in (62); then $n = (1/2 \pi^2) (2m/\hbar^2)^{3/2} F_{1/2}$. Expanding $F_{1/2}$ as in (62) we find that

$$n = \frac{1}{2\pi^2} \left(\frac{2m}{\hbar^2} \right)^{3/2} (2/3) \mu^{3/2} \left\{ 1 + \frac{(\pi kT)^2}{8\mu^2} + \dots \right\} \quad (63)$$

This specifies the density of electrons at nonzero temperatures. Comparing (53) and (63) we note that $\mu = \mu_0 [1 + (\pi kT)^2/8\mu^2]^{-2/3}$. Since $\mu/kT \gg 1$ by assumption we may expand the above in a binomial series for which the first two terms read

$$\mu = \mu_0 \left[1 - (\pi kT)^2 / 12\mu^2 \right]. \quad (64)$$

Similarly from (60) and (62), with $\underline{l} = 3/2$, U can be evaluated as

$$U = \frac{1}{5\pi^2} \left(\frac{2m}{\hbar^2} \right)^{3/2} \mu^{5/2} \left\{ 1 + \frac{5}{8} \frac{\pi kT}{\mu} + \dots \right\}. \quad (65)$$

If we ignore the small difference between μ and μ_0 we may utilize (48) and (53) or (54) to write

$$U = U_0 \left[1 + \frac{5}{12} \left(\frac{\pi kT}{\mu_0} \right)^2 \right] = U_0 + \frac{n(\pi kT)^2}{4\mu_0}. \quad (66)$$

The heat capacity of free electron, C_e , may now be found by determining (dU/dT) . From (66) one obtains

$$C_e = \frac{n\pi^2 k^2 T}{2\mu_0} \quad (67)$$

which is smaller by a factor of $\pi^2 kT/3\mu_0$ (≈ 0.025 for typical metals) than the classical value of $(3/2)kn$.

The above theory shows why the heat capacity for electrons in a metal is abnormally low. Physically, this arises because an electron which is to respond to thermal excitation must have available empty energy levels which it can occupy. As is clear from Fig. 5, such a situation does not prevail for the majority of electrons with energies ϵ less than μ . Only those within a range $\sim kT$ in the vicinity of the Fermi level have empty energy states thermally accessible to them; hence it is only this small fraction that provides a net response to a change in temperature. Correspondingly, C_e is exceedingly small. This restriction no

longer applies to semiconductors where the band is filled to only a small extent; in this case virtually all electrons have empty energy levels available to them.

VII. CLASSIFICATION OF MATERIALS

The above sets the stage for the classification of solids, according to their electrical properties; we shall have to use the scheme extensively later on. We consider first that class of solids where overlap between orbitals in the valence shells of the constituent atoms is sufficiently extensive to allow bands to be formed. With reference to Fig. 6 the following cases may then be recognized: (a) The number of electrons is such that at 0 K all bands are either totally filled or completely empty; further, $kT \ll \epsilon_g$ and impurity effects may be neglected. One is then dealing with an insulator. Obviously empty bands do not contribute to electrical properties, while in filled bands the electrons have no available empty states which they can occupy; there is no net transport of charge. This matter will be discussed more extensively in Section IX.

(b) The conditions in (a) are fulfilled, but now $kT \approx \epsilon_g$; in this case electrons can be thermally promoted across the gap from the top of the valence band to the bottom of the conduction band in limited numbers. As we shall see shortly, the absence of carriers in the valence band may be treated as if these lacunae were positively charged particles, or holes, in their own right. A material satisfying criterion (b) is said to be an intrinsic semiconductor; here, the contribution of both the electrons in the conduction bands and of the holes in the valence band must be taken into consideration in studying its electrical characteristics. An extreme case occurs when kT becomes actually comparable to or greater than ϵ_g . In this case the charge carrier density becomes so large in each band that the material must be classified as a degenerate semiconductor.

(c) The conditions specified in (a) apply, but the material contains an appreciable number of impurities, whether added intentionally or not. The situation may best be discussed with some specific examples: Suppose one is dealing with Ge containing As as impurities. The As enters the lattice substitutionally but relative to Ge possesses one additional electron that cannot be utilized in the tetrahedral coordination prevalent in the host material. At very low temperatures this extra electron remains in the vicinity of the As, somewhat as an electron which surrounds the proton in the elemental hydrogen. However, since the As-e unit is imbedded in a lattice of dielectric constant 16, the ionization energy required to detach the electron is smaller by a factor of

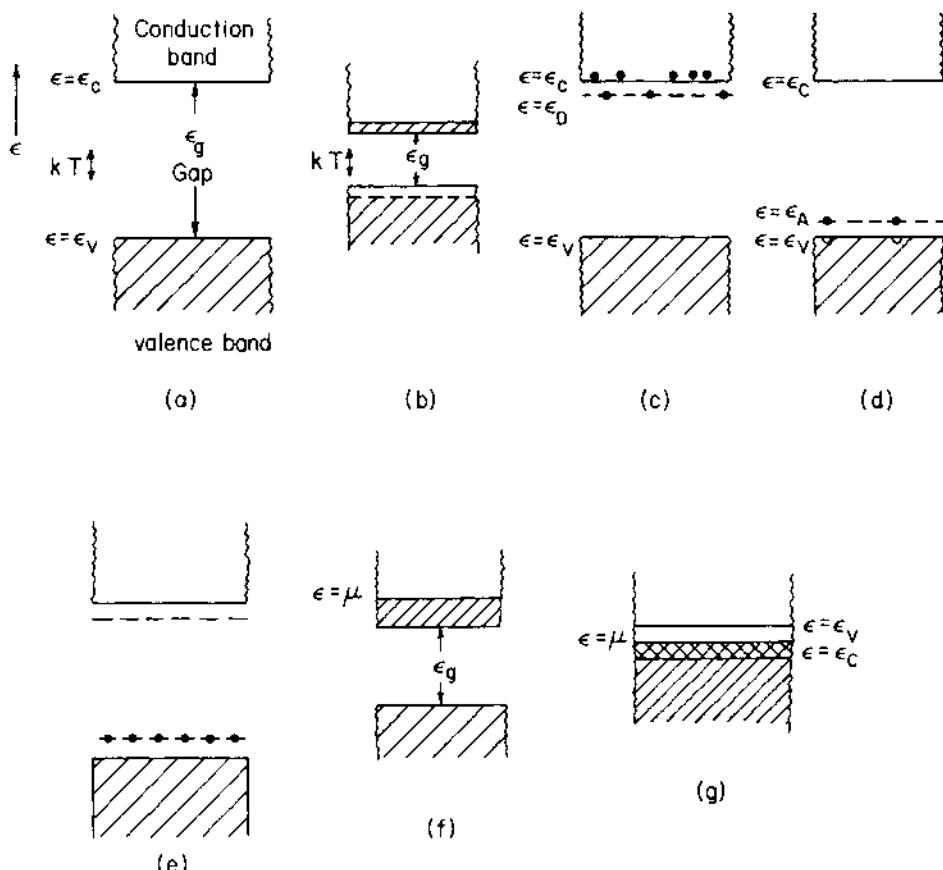


Fig. 6. Simplified band structure schemes for solids. (a) Insulator: $kT \ll \epsilon_g$; all band states below $\epsilon = \epsilon_v$ are occupied, all others are empty. (b) Intrinsic semiconductor: kT comparable to ϵ_g ; electrons are thermally promoted from the top of the valence to the bottom of the conduction band. (c) and (d) Extrinsic semiconductors: impurity levels lie within the gap; $kT \ll \epsilon_g$, but comparable to either $\epsilon_c - \epsilon_d$ or $\epsilon_a - \epsilon_v$. (e) Compensated semiconductor. (f) Metal: One band is partially though extensively filled with electrons. (g) Semimetal: The top of the valence band lies below the bottom of the conduction band.

256 than that in hydrogen atom proper. Hence, very small thermal energies suffice to "dissociate" the As-e unit into an As^+ site and a charge carrier e^- which, moving freely about the lattice, appears as an electron in the conduction band. Impurities that act as As does in Ge are termed donors. On the band diagram in Fig. 6 (c) the presence of such impurities is indicated by small horizontal lines placed below the bottom of the conduction band at a distance E_d corresponding to the ionization energy for the impurity in question. Sites corresponding to the As-e unit are denoted by a heavy dot. As T rises all sites are ionized; one then has reached the exhaustion range in impurity ionization.

The corresponding situation for the valence band depicted in Fig. 6 (d), may be illustrated with the case of Ga as an impurity in the Ge lattice. Since Ga contains one less electron in the valence shell than Ge, incorporation of the impurity causes incomplete saturation of the bonds, so that at 0 K voids in the electron configuration exist in the vicinity of the Ga atoms. At higher temperatures, electrons from neighboring Ge atoms can move into the voids as illustrated schematically in the sequence of diagrams in Fig. 7. Corresponding to every movement of an electron, the void is transferred in the opposite direction. Thus at higher temperatures each void will tend to be found at a considerable distance from the Ga. The process by which this occurs may be regarded as an 'ionization' of the lacuna from the Ga in accord with the schematic relation $\text{Ga} \cdot p \rightarrow \text{Ga}^+ + p^+$, where p represents the void;

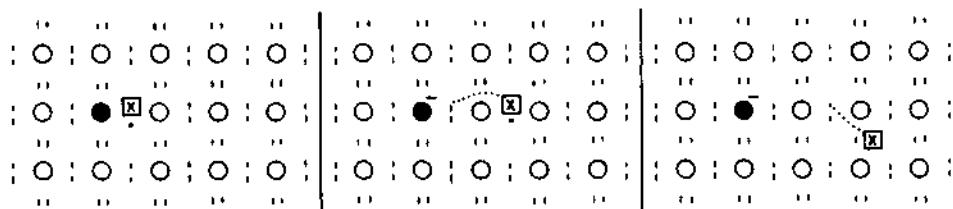


Fig. 7. Schematic diagram illustrating the motion of a lacuna through the valence banding structure in Ge. Ge atoms denoted by open circles, As impurity represented by filled circle, the lacuna indicated by cross. Note that the direction of motion of lacuna is opposite to that of the electron with which it interchanges position.

the electron which takes the place of the void originally centered about the Ga atom imparts a negative charge to the impurity. Though the correspondence is incomplete, the absence of an electron from a structure where it belongs may be regarded as an entity with a positive charge termed a hole; once ionized, it acquires a drift velocity under an applied electric field in the manner of a positively charged particle. The Ga + p combination may be regarded as an anti-hydrogen atom imbedded in the Ge lattice. This is indicated by placing a short vertical bar above the top of the valence band at a distance ε_a corresponding to the ionization energy for the hole in the Ge lattice. At 0 K all these levels are empty, but for $T > 0$ they acquire electrons as indicated by the heavy dots in these bars, and a corresponding void, representing the hole, appears at the top of the valence band. When all the impurity levels have been filled the exhaustion range has been reached; no more holes can now be generated thermally until kT becomes comparable to ε_g . Impurities such as Ga in Ge are termed acceptors. Materials in which electrons or holes are generated in this manner are termed extrinsic semiconductors. Generally, donors and acceptors tend to be present simultaneously; if it happens that their total concentrations are equal, the material is said to be a compensated semiconductor; this case is depicted in Fig. 6 (e). One should note that in those circumstances all electrons in the lower levels drop into the acceptor levels at 0 K, thus blocking thermal ionization of carriers until kT again becomes nearly comparable to ε_g . It is difficult to distinguish between compensated and intrinsic semiconductors: the former tend to exhibit considerably lower electron and hole mobilities because of the presence of the impurities.

(d) Whenever it happens that the number of electrons is insufficient to fill the highest lying band, the Fermi level falls within this band and the material is said to be a metal. The properties of this class of materials is intimately linked to the Pauli exclusion principle as manifested by Fermi-Dirac statistics. This matter was already discussed earlier at some length; the corresponding situation is depicted in Fig. 6 (f).

(e) There remains the case of band overlap, depicted in Fig. 6 (g); a more adequate representation in terms of an energy band diagram in reciprocal space for two dimensions is shown in Fig. 8. This situation arises if it so happens that the bottom of the $(n+1)$ th band lies energetically lower than the top of the n th band. As is illustrated in Fig. 8, this type of overlap involves different regions of k -space. The net result is as follows: Suppose the n th band were completely filled if no overlap existed; then under the conditions depicted in Fig. 6 (g) and 8, the electrons at the top of the valence band will 'spill over' into the bottom of the conduction band until the Fermi level

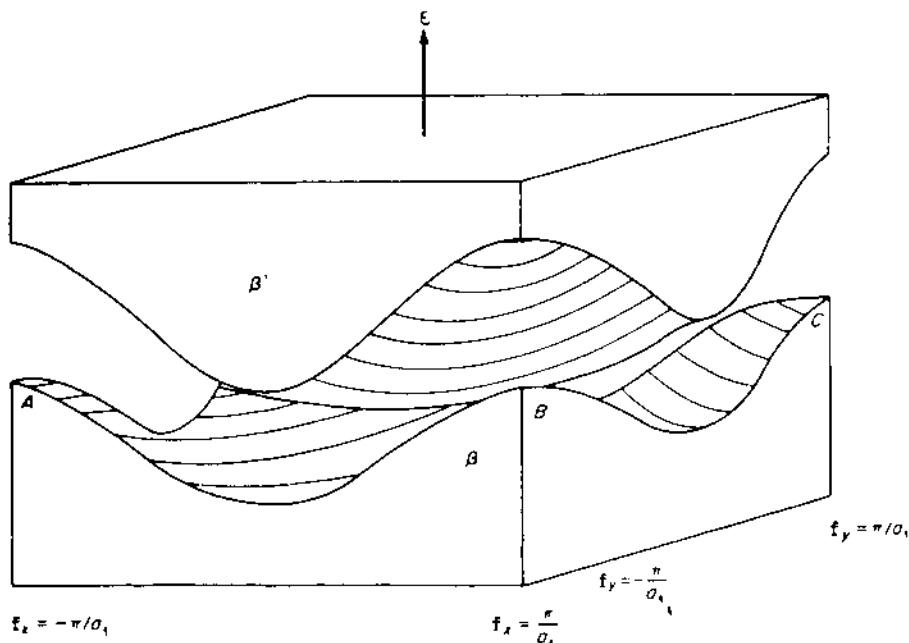


Fig. 8. Band overlap in reciprocal space. Note that the 'overlap' occurs in different regions of \tilde{k} -space.

properly intersects both sets of bands. Holes and electrons now coexist even at 0 K; a material of this type is said to be semimetal.

So far, we have discussed the cases where bands are formed. We now consider the opposite extreme; here, orbital overlap is so small that in first approximation the charge carriers are localized about their atomic centers. A jump of an electron from one atomic site to an equivalent adjacent one can occur (i) if the latter happens to be empty, and (ii) if the electron acquires the requisite activation energy for the diffusive type of motion. Materials where electron transport occurs via an activated process are known as hopping type materials, or colloquially, as 'hoppers'.

There is a third general classification of narrow-band materials, symptomatic of considerable interactions between charge carriers and the lattice; here the electron is neither nearly free nor nearly bound. When electronic movement is only slightly impeded, due to the relatively weak interactions with the lattice,

one describes the carrier and its interaction as being a large polaron. The interaction takes the form of a small lattice polarization about the electron that moves with the charge carrier; the latter is then said to be 'dressed'. This is manifested in an increase in the apparent effective mass of the electron relative to the mass it would have in the absence of such interactions. On the other hand, if the interaction is quite strong, one is dealing with a small polaron, consisting of the charge carrier surrounded locally by a strongly deformed lattice. The deformation tends to move with the carrier, whose freedom of maneuver in consequence is considerably impeded. At low temperatures the small polarons exhibit band-like properties. As T rises, the band narrows down until it becomes indistinguishable from the localized levels discussed in connection with the hopping-type materials.

VIII. ELEMENTARY DISCUSSIONS OF VARIOUS TRANSPORT PHENOMENA

Within the framework developed earlier we now discuss some of the common transport phenomena in solids. The proper derivations of the results discussed below turn out to be quite lengthy, as may be seen by reference to several books and review articles¹⁻³. Hence, we attempt here only to capture the essence of the derivations. As a result, the final relations are usually not quite correct; this approach seems to be better suited here than the rather involved derivations needed to obtain the strictly correct relations.

The Electrical Conductivity. Let us consider a set of freely moving electrons in a band. According to Newton's second law of motion, the rate of change of momentum is equal to the force, \mathcal{F} exerted on the particles. Since in the free electron approximation the momentum is given by $\hbar k$ it follows that, $\hbar(\partial k/\partial t) = \mathcal{F}$. The drift velocity v_d is the extra velocity acquired by the particle in the force field; it is related to the extra momentum acquired by the particle through the expression, $v_d = \hbar \Delta k/m$.

In an ideal solid the free carriers subjected to an external electric field are accelerated without interruption. However, scattering occurs because of the lattice vibrations (phonons) or other imperfections in the crystal; this event interrupts the acceleration and produces the electrical resistance. If τ , the relaxation time, is the average time between scattering events, the increase in the momentum of the particle is then given by, $\hbar \Delta k = \mathcal{F} \tau$. It follows that, $v_d = \hbar \Delta k/m = \mathcal{F} \tau/m$. In an electric field $\mathcal{F} = -eE$; we therefore find that the drift velocity is

$$v_d = \frac{-eE\tau}{m} = -u_d E \quad (68)$$

where $u_d \equiv |v_d| / |E|$ is the drift mobility of the carrier, i.e. its drift velocity in unit electric field.

Now the current passing through unit cross section, J , is the rate of passage of charge ($-ne$) past unit cross section: Using Eqn. (68) we can write,

$$\begin{aligned} J &= -nev_d = + \frac{ne^2\tau}{m} E \\ &= (neu_d) E \equiv \sigma E \end{aligned} \quad (69)$$

where we have introduced the conductivity σ through the relation

$$\sigma = neu_d = ne^2\tau/m. \quad (70)$$

Equation (69) is the usual formulation for Ohm's law in terms of current density and electric field. Eqn. (70) holds strictly only for an isotropic solid where σ is a scalar quantity. Eqn. (68) shows that the mobility may also be written in terms of τ as,

$$u_d = \frac{e\tau}{m}. \quad (71)$$

Let l represent the mean free path, i.e., the distance travelled by a charge carrier between collisions; then the relaxation time is related to l by $\tau = l/v$ where v is the total velocity, i.e., the sum of the velocities due to the thermal and drift components. From (69) we then obtain

$$\sigma = ne^2 l/mv. \quad (72)$$

We must now distinguish between two types of electrons in the band under consideration: (i) Free electrons which comprise the totality of all electrons present, and (ii) conduction electrons which are those that actually provide a net contribution to the electrical conductivity in a solid. When the field is applied every electron gains in energy; however, the preponderant majority of carriers is shifted from the levels they occupied initially to those previously occupied by other carriers. These electrons simply take each other's place and provide no net contribution to charge transport. Only carriers present near the Fermi level can move to available empty levels in responding to the applied electric field, and it is this small grouping in a metal which provides a net response to the field. Thus, a shift in the distribution takes

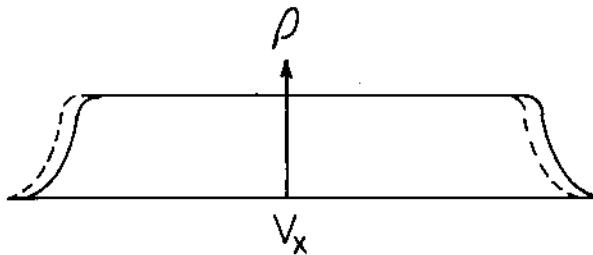


Fig. 9. Plot of probability P vs. total velocity v_x under equilibrium conditions (dotted line) and in the presence of an applied field along the $-x$ direction (solid line).

place; all the electrons respond to the field, even though an overall effect is seen only for the electrons near the Fermi level. However, for semiconductors, with a very limited population of a band, the situation is different. Here, virtually all carriers have empty levels available to them, so that all electrons can provide a net response to the electric field, i.e., all free electrons are also conduction electrons.

Suppose we plot the probability density function (P) against velocity (v_x) in the x direction as in Fig. 9. In equilibrium the net current passing through the solid is zero because there are as many carriers with velocities $+v_x$ as with velocities $-v_x$ for all v_x . However, under the influence of an external electric field, $\tilde{E} = -\frac{1}{2} E_x$, there will be a shift in $P(v_x, \tau)$ opposite to the direction of E , as shown in Fig. 9; all electrons have now acquired a net drift velocity along the positive x direction. The displacement shown in Fig. 9 holds steady after a lapse of time $\Delta t = \tau$, when the steady state has set in. For a small perturbation arising from \tilde{E} one finds to first order approximation,

$$P(v_x, \tau) = P(v_x, 0) - \frac{\partial P(v_x, 0)}{\partial v_x} \cdot \Delta v_x, \quad (73)$$

where Δv_x is the additional (drift) velocity ($\equiv -eE_x\tau/m$, from (68)). $P(v_x, \tau)$ is assumed to be normalized to unity. Hence, the average velocity of the charge carrier in the direction of the field is given by

$$\begin{aligned}\bar{v}_x &= \int_{-\infty}^{+\infty} v_x P(v_x, \tau) dv_x \\ &= \left(\frac{e \tau E_x}{m} \right) \int_{-\infty}^{\infty} v_x \frac{\partial P(v_x, 0)}{\partial v_x} \cdot dv_x .\end{aligned}\quad (74)$$

Note that the term $\int_{-\infty}^{\infty} v_x P(v_x, 0) dv_x$ vanishes since $P(v_x, 0)$ is even and v_x is odd under a sign inversion. After integrating Eqn. (74) by parts,

$$\begin{aligned}\bar{v}_x &= \frac{e \tau E_x}{m} \left[\{v_x P(v_x, 0)\} \Big|_0^+ - \int_{-\infty}^{+\infty} P(v_x, 0) \cdot dv_x \right] \\ &= - \frac{e \tau E_x}{m} .\end{aligned}\quad (75)$$

In the above the first term vanishes identically, since $P(v_x, 0)$ decreases exponentially to zero as $v_x \rightarrow \pm\infty$. Since $P(v_x, 0)$ is normalized over the interval $-\infty < v_x < \infty$, the second integral converges to unity.

Let us examine the results: Whereas at equilibrium $\bar{v}_x \equiv 0$, we see now that the imposition of an electric field has led to a nonvanishing value for \bar{v}_x which must clearly be identified with the drift velocity acquired by the carriers under the influence of the field. Eqn. (75) is then seen to agree with Eqn. (68). The above analysis is a more sophisticated version of the oversimplified arguments used at the beginning of this section. Henceforth it will be taken for granted that we do not have to worry further about averaging procedures, since in the present context they do not significantly alter the derivations based on unaveraged quantities.

We now proceed by extremely crude arguments to reformulate Eqn. (72). For a band which is only slightly filled, as in extrinsic semiconductors, the average kinetic energy ($= 1/2 mv^2$) of a carrier may be equated with $3/2 kT$; thus from (72),

$$\sigma = ne^2 l / (3mkT)^{1/2} .\quad (76)$$

The correct expression using appropriate statistics and assuming that the main scattering process is due to lattice vibrations, reads

$$\sigma = ne^2 l / (9\pi mkT/8)^{1/2} . \quad (77)$$

For metals it is plausible to concentrate on the conduction (as against the free) electrons and to substitute for v the Fermi velocity, v_f , so that $\frac{1}{2}mv_f^2 = \mu_0$. Since μ_0 is related to the charge carrier concentration by Eqn. (54), we can write for the conductivity of a metal

$$\sigma = ne l / mv_f = \left[\frac{8\pi}{3n} \right]^{1/3} \frac{ne^2 l}{m} . \quad (78)$$

Measurement of the Electrical Conductivity of Solids. The usual two-probe method is associated with uncertainties because of contact resistance and grain boundary effects. The contact resistance will usually be of the same order of magnitude as the resistance of the material. The generally accepted procedure for conductivity measurements involves a four-probe method. A schematic diagram is shown in Fig. 10. Current is passed through the sample via junctions a and b. The voltage drop across the sample is measured on a null basis across junctions c and d. Since no current passes through the secondary circuit, contact resistance effects do not affect the voltage measurements. It is advisable to use point probes for voltage measurement and these should be collinear with the long axis of the sample to avoid errors due to nonlinearity of the current paths through the sample. Care should also be taken to see that the ohmic behavior of the sample is verified by making measurements at various currents.

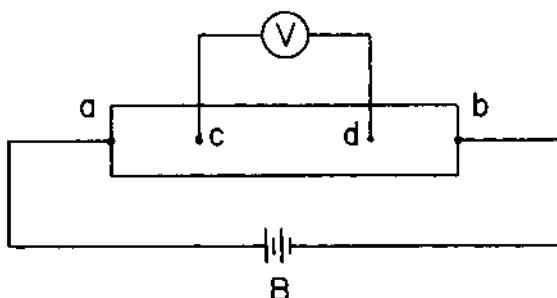


Fig. 10. Four-probe method of measuring resistivities in samples with rectangular parallelepiped geometry.

The Thermal Conductivity. We now determine that part of the total thermal conductivity in an isotropic solid which is due to the motion of the free electrons.

Consider a solid in which a thermal gradient in temperature dT/dx exists, giving rise to a temperature difference ΔT over a distance Δx along the x direction. Let c_e be the heat capacity of the charge carrier; then, in moving from a region of higher temperature to one of lower temperature, the heat exchanged is $c_e \Delta T$. Here one must choose $\Delta x = l (\equiv v_x)$ as the mean free path length, since it is only during collisions that the particle is able to exchange heat with the surroundings. Thus, the heat exchanged is $q = c_e T = -c_e (\partial T/\partial x) \tau v_x$ for one particle. The minus sign arises because in our convention T diminishes in the direction of increasing x . If $\frac{1}{2} n$ is the number of particles moving on the average in the $+x$ direction, then the flux of particles is $\frac{1}{2} nv_x$ and hence the thermal energy transported in this direction is,

$$Q_{+x} = q(\frac{1}{2}nv_x) = -c_e(\frac{1}{2}nv_x^2)\tau (\partial T/\partial x). \quad (79)$$

In the steady state there is an equal flux in the opposite direction, viz., Q_{-x} . Here, relative to the expression for Q_{+x} the signs in v_x and dT/dx are both interchanged. Hence, the total rate of transport of heat is given by

$$Q = Q_{+x} + Q_{-x} = -nv_x^2 c_e \tau (\partial T/\partial x). \quad (80)$$

Comparing this with Fourier's law for heat conduction, $Q = -k (\partial T/\partial x)$, where k is the thermal conductivity, we write

$$k_e = nv_x^2 c_e \tau. \quad (81)$$

The subscript e in (81) serves as a reminder that only the contribution of electrons to the total thermal conductivity is being considered. Further, since $v^2 = v_x^2 + v_y^2 + v_z^2$ and since on the average $v_x^2 = v_y^2 = v_z^2$, we can write $v_x^2 = (1/3)v^2$. Finally, we substitute $1/v$ for τ . This yields

$$k_e = (1/3) c_e n l v = (1/3) C_e l v, \quad (82)$$

where C_e is the heat capacity per unit volume by the motion of electrons in the crystal. Eqn. (82) is the fundamental formula; we now try to evaluate some of the fundamental parameters in this equation.

Taking up the case of semiconductors, where all electrons in

the band can respond to the thermal gradient, we can in first approximation set $C_e = 3/2 k n$ and $\frac{1}{2}mv^2 = 3/2 kT$. This allows us to write (82) as

$$\begin{aligned} \kappa_e &= 1/3 C_e l v \equiv (k/e^2) (1/2 mv^2) \left(\frac{ne^2 l}{mv} \right) \\ &= \frac{3}{2} \left[\frac{k^2}{e^2} T \right] \sigma. \end{aligned} \quad (83)$$

The correct expression for κ_e is

$$\kappa_e = 2 \left[\frac{k^2}{e^2} T \right] \sigma. \quad (84)$$

Eqn. (84) represents an example of the Wiedemann-Franz law, according to which κ_e is proportional to σ . The proportionality coefficient, $\kappa_e/\sigma T = 2(k^2/e^2)$ in this case, is known as the Lorenz number, and is a universal constant.

For metals, it is more to the point to invoke the properties of the electrons near the Fermi level. We first replace v in (82) by v_f . Then

$$\kappa_e = \frac{1}{3} C_e n l v_f = \frac{2}{3} \left(\frac{C_e}{e^2} \right) \left(\frac{1}{2} m v_f^2 \right) \left(\frac{ne^2 l}{m v_f} \right);$$

next we utilize the relation, $C_e = \pi^2 k^2 T / 2 \mu_0$ derived in Eqn. (67), as well as $\mu_0 = \frac{1}{2} m v_f^2$, and Eqn. (72). This yields

$$\kappa_e = \frac{\pi^2}{3} \left[\frac{k^2}{e^2} T \right] \sigma \quad (85)$$

which again conforms to the Wiedemann-Franz law, with a Lorenz number, $\kappa_e/\sigma T = (\pi^2/3)(k^2/e^2)$ which is another universal constant.

Thermoelectric Properties. We have seen that current can flow through a crystal when carriers in a partially filled band move in response to an externally applied electric field $E = -\nabla \phi_s$, where ϕ_s is the electrostatic potential. However, any other directed force can serve equally well to propel charge carriers through the lattice. We consider here two additional types of force; one arising from any inequality of charge carrier density which is manifested by the establishment of a gradient in the chemical potential $\mu_n = \mu_n^0 + kT \ln a_n$; the other, due to the presence of a temperature gradient in the sample. We will see in more detail

later on how a concentration gradient in the sample affects electrical properties; here we simply note that this effect can be grouped with the electric field effect by defining the electrochemical potential $\xi \equiv \mu_n - e\phi_s$. Then at constant temperature we should observe that the current density J is proportional to $\nabla \xi$. We find it convenient to express this proportionality by writing $J = L_{22} \nabla (\xi/e)$, where the compensating factor e has been absorbed in the proportionality constant. Again, this formulation describes the response of charge carriers in a solid to the joint action of electric fields and effects arising out of concentration gradients. We next assume that the effects due to any temperature gradient which may also be present is properly described by a linear superposition of the new force on the forces already mentioned. A physical mechanism explaining how carriers respond to the imposition of a temperature gradient is provided later in this Section. The assumption concerning the linear superposition of the effects due to $\nabla (\xi/e)$ and to ∇T has no fundamental justification, but is experimentally justified so long as one does not subject the solid to drastic departures from equilibrium conditions. In accord with the above discussion we now write $J = L_{21} \nabla T + L_{22} \nabla (\xi/e)$, where L_{21} and L_{22} are known as phenomenological or macroscopic coefficients.

In a solid where electrons are mobile, the imposition of a temperature gradient results not only in a directed flow of heat but also in a net movement of the charge carriers, as we shall see later. Both these effects combine to alter the entropy balance, as manifested by an entropy flux $J_s = L_{11} \nabla T$ under the effect of a temperature gradient. Again, the distribution of particles, the entropy balance, and hence J_s , may be affected by any gradients in electrochemical potential. Assuming that the principle of linear superposition applies here as well, we write for the total entropy flux $J_s = L_{11} \nabla T + L_{12} \nabla (\xi/e)$. Thus we arrive at a set of linear phenomenological or macroscopic equations of the type

$$J_s = L_{11} \nabla T + L_{12} \nabla (\xi/e), \quad (86)$$

$$J = L_{21} \nabla T + L_{22} \nabla (\xi/e). \quad (87)$$

On account of the Onsager Reciprocity Conditions the two off-diagonal coefficients are connected through the relation $L_{12} = -L_{21}$ in the particular formulation of the phenomenological equations shown above. If one were to replace ∇T with $\nabla(1/T)$ as the driving force due to the temperature field the resulting macroscopic equations would then be reformulated so that the off-diagonal coefficients are equal to each other. This is a refinement which is not essential to our further discussion.

To identify the various L_{ij} shown above let us consider several special cases: (i) Suppose $\nabla T = 0$; then by elimination of $\nabla(\xi/e)$ between (86) and (87), we obtain

$$\frac{J_s}{J_n} = \frac{L_{12}}{L_{22}} J = -e \frac{L_{12}}{L_{22}} J_n ,$$

where we have introduced the electron particle flux vector J_n through the relation $J = -eJ_n$.

It follows that one may define the entropy transport per particle by the relation

$$S^* \equiv \frac{|J_s|}{|J_n|} = -e \frac{L_{12}}{L_{22}} = +e \frac{L_{21}}{L_{22}} . \quad (88)$$

Further, from Eqn. (87) one obtains

$$\frac{J}{J_n} = L_{22} \nabla(\xi/e), \quad (89)$$

which is the generalized form of Ohm's law. It follows that $L_{22} \equiv \sigma$.

Let us next consider what happens for the case (ii) $J = 0$; from Eqns. (86) and (88) we then obtain

$$\nabla(\xi/e) = - \left[\frac{L_{21}}{L_{22}} \right] \nabla T \equiv \alpha \nabla T = - \frac{S^*}{e} \nabla T \quad (90)$$

which states that when no current flows in a sample, the imposition of a temperature gradient results in the establishment of a gradient in ξ . Equation (90) is then a manifestation of the thermoelectric effect; α is called the Seebeck Coefficient, which is related to the entropy transported per particle by $\alpha = -S^*/e$.

The origin of the thermoelectric effect can be understood by reference to Fig. 11. If we have an n-type semiconductor or metal across which there exists a temperature gradient the electrons at the hot end on the average will have a larger velocity component in the x direction than the electrons at the cold end have along the $-x$ direction. Consequently, there occurs a net movement of electrons such that an excess accumulates at the cold end, and deficiency of electrons develops at the hot end. Thus, a gradient

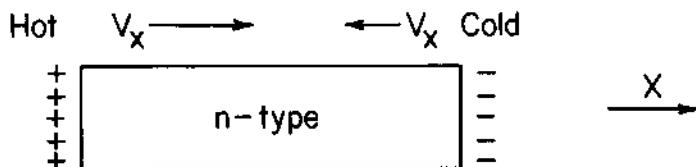


Fig. 11. Origin of the Seebeck thermoelectric potential for n-type materials. Note accumulation of excess electrons at cold, and deficit at hot end.

in electrochemical potential is set up. This process cannot continue indefinitely; for, as a result of the movement of carriers a growing internal electric field is created; under steady state conditions the net flow is ultimately stopped by the internal field. The gradient in electrochemical potential is termed the thermal emf.

If the electric circuit in Fig. 11 is completed by insertion of an external load resistor, current can be made to flow through the circuit; electric power is thus generated by direct conversion of heat energy into electric power in earth satellites. These units can be made very compact since there are no moving parts in the system. However, the efficiency generally remains well below 10%.

We now attempt to evaluate the Seebeck coefficient in terms of fundamental parameters. For this purpose, return to Eqn. (90) and set $\xi = E^* - TS^*$ which is the appropriate defining relation for the electrochemical potential. Then, for electrons

$$\alpha_n = \frac{-S^*}{e} = - \frac{E^* - \xi}{Te} ; \quad (91)$$

similarly, for holes,

$$\alpha_p = \frac{+S^*}{e} = - \frac{\xi - E^*}{Te} . \quad (92)$$

If we are dealing with extrinsic semiconductors, where the density of carriers is low we can assume that classical statistics holds and that all the electrons respond to the external fields. Then the total energy carried and the electrochemical potential

per particle is

$$E^* = \frac{3}{2} kT + \varepsilon_c \quad \text{and} \quad \tau = \mu_c + \varepsilon_c \quad (93)$$

where ε_c is the conduction band edge energy. Thus,

$$\alpha_n = -\frac{(3/2)kT - \mu_c}{T_e} = -\frac{k}{e} \left[\frac{3}{2} - \frac{\mu_c}{kT} \right] \quad (94)$$

which is the expression for Seebeck coefficient in terms of the Fermi level. Again, the arguments used here are too crude: Assuming that scattering by lattice thermal vibrations predominates, the correct treatment leads to the result

$$\alpha_n = -\frac{k}{e} \left[2 - \frac{\mu_c}{kT} \right] ; \quad (95)$$

similarly, for holes,

$$\alpha_p = +\frac{k}{e} \left[2 - \frac{\mu_v}{kT} \right] \quad (96)$$

where $\mu_v \equiv \varepsilon_v - \tau$.

For semiconductors α is in the range from 0.1 to 1 mV/deg and decreases with rising temperature. However, as shown later, μ_c or μ_v also change with T, so that the dependence of α with T is more complex than appears from Eqn. (95) or (96).

For metals, we may proceed as follows: Returning to Eqn. (90) and (67) we write

$$\alpha_n = -\frac{S^*}{e}$$

and $S^* = \int_0^T \frac{c_e dT}{T} = \int_0^T \left[\frac{\pi^2 k^2 T}{2 \mu_0} \right] \frac{dT}{T}$ (97)

where c_e is the heat capacity per electron. It follows that

$$\alpha_n = -\frac{\pi^2 k^2 T}{2 \mu_0 e} . \quad (98)$$

The correct expression reads

$$\alpha = \frac{\pi^2}{3} \left(\frac{k^2 T}{e} \right) \frac{1}{\mu_0} . \quad (99)$$

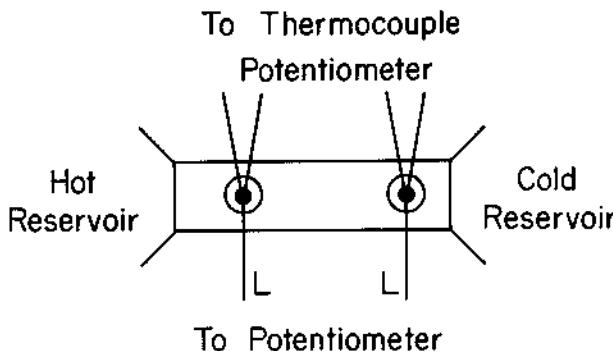


Fig. 12. Experimental arrangement for the measurement of Seebeck coefficients.

It turns out that α_n is small ($\sim 1 - 10 \mu\text{V}/\text{deg}$) since the factor kT/μ_0 (99) is very small. We also see that α_n tends to increase with T , unlike the case of semiconductors. Knowing μ_0 through a determination of α_n one can then determine n via Eqn. (53).

The Measurement of Seebeck Coefficient. The measurement of α can be performed on ceramic samples or on single crystals since no net current is caused to flow under open circuit conditions. To obtain accurate results one should carry out the measurements as shown in Fig. 12; this avoids boundary effects by having the thermocouples and voltage probes attached in the interior of the sample. One then usually measures $\alpha = \alpha_S - \alpha_L$ where the second term represents the Seebeck coefficient for the lead-wires. The junctions must be isothermal so that the electrochemical potential remains constant across them. It should be evident that from the sign of the potentiometer terminal connected to the cold end of the sample one deduces the sign of the predominant charge carriers in the specimen.

The Hall Effect. Let us consider a bar-shaped specimen subjected to a longitudinal electric field $E = \hat{\mathbf{E}}_x$ and to a transverse magnetic field $H = \hat{\mathbf{H}}_z$ as shown in Fig. 13.

The force component on the sample in the x direction is specified through the Lorentz force acting on the n electrons making up the current:

$$F_y = Zen E_y - I_x H_z/c \quad (100)$$

where c is the velocity of light; this quantity reminds us that

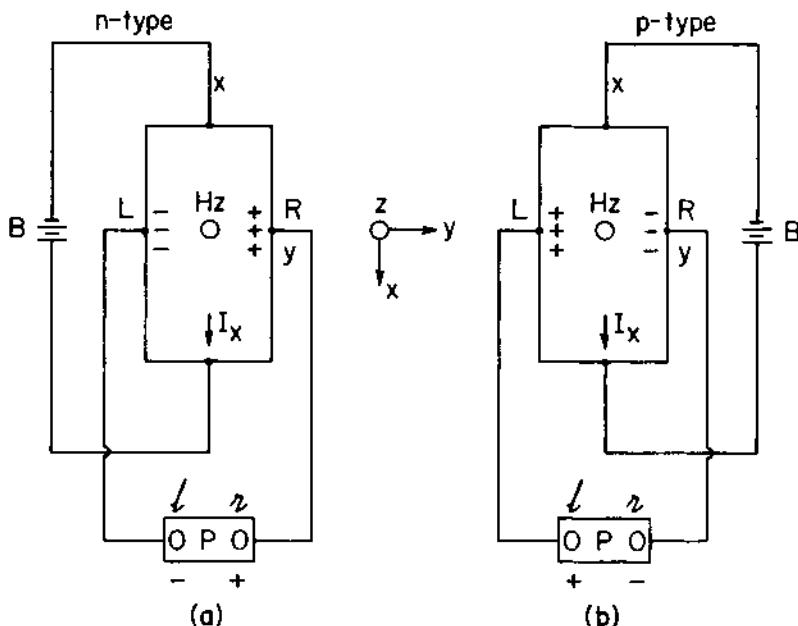


Fig. 13. Experimental arrangement for measuring the Hall coefficient in p- and n-type samples.

proper units be used in numerical calculations, as will be detailed later. Note regardless of the sign of the charge carriers, the magnetic force field always imparts a force $(I \times H)/c$ to the mobile species; under the arrangement of Fig. 13 excess carriers accumulate on face L and a deficiency develops on face R. In n- (p-) type materials side L will become negative (positive) relative to side R. From the polarity of the potentiometer terminal connected to side L one can at once determine whether a given material is n- or p-type.

As in the thermoelectric effect, the building of a charge differential cannot continue indefinitely, but is at some point counteracted by the resultant internal field of sufficient intensity to stop any further net flow along the y direction. Since the internal force now involves not only the electrostatic field but also an imbalance of particle density, it is appropriate to substitute in place of $E_x = -\nabla_x \phi_s$ the quantity $\nabla_x(\zeta/e)$ where $\zeta = \mu_n - e\phi_s$. Moreover, under steady-state conditions the net force in the y direction vanishes; Eqn. (100) thus reads

$$Zen \quad \nabla_y(\zeta/e) = I_x H_z/c. \quad (101)$$

We now define the Hall coefficient R as shown below in terms of

measurable quantities

$$\frac{\nabla_y(\tau/e)}{I_x H_z/c} \equiv R = \frac{1}{Zen} . \quad (102)$$

As before, $\nabla_y(\tau/e)$ is the difference in electrochemical potential as read off on the voltmeter, in going from face L to face R in Fig. 13, divided by the distance. Since I_x and H_z are also known, R may be calculated. This quantity is inversely proportional to the charge carrier density. For reasons shown earlier the sign of the carriers enters through the quantity Z.

The correct expression is given by

$$R = A/Zen \quad (103)$$

where, under conditions when lattice scattering predominates, $A = 3\pi/8$ for semiconductors and $A = 1 + (\pi kT)^2/12 \mu_o^2$ for metals.

Note that if σ and R are both determined then their product

$$\sigma |R| = neu_d \left[\frac{A}{ne} \right] = Au_d \quad (104)$$

is a measure of the drift mobility. The consistent use of the emu system requires that in terms of practical units, $u_d H/c = 1$ when $u_d = 1 \text{ cm}^2/\text{v-sec}$ and $H = 10^8 \text{ gauss}$.

The Ettingshausen Effect and Ettingshausen-Nernst Effect. These particular effects are best explained in terms of a model involving the simultaneous presence of both holes and electrons in a solid. Suppose as in Fig. 14 a current I is passing in y direction, and a transverse magnetic field H_z is applied in the z direction; then a drift of both types of charge carriers (electrons and holes) also occurs along the positive x direction in accord with the Lorentz law. This leads to a concentration of charge carriers on face A and to a deficit thereof on face B. According to Le Châtelier's principle, the system compensates for the oversupply on face A by electron-hole annihilation which releases the heat of combination to the surroundings near A. On side B, in an attempt to overcome the deficit of carriers, electron-hole pairs are generated, the requisite energy for this purpose being withdrawn from the surroundings near face B. This setup can thus be used as a refrigerator or heat pump without any moving parts; unfortunately, the efficiency in commonly available magnetic fields is exceedingly low. The phenomenon itself is termed here the Ettingshausen effect.

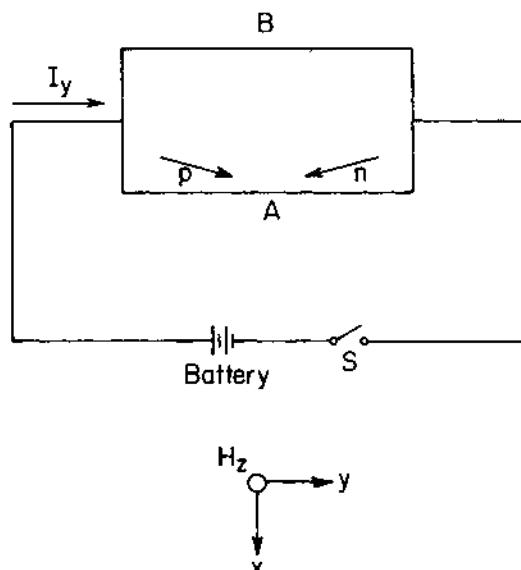


Fig. 14. Diagram illustrating the Ettingshausen effect.

The inverse of this phenomenon is also observed. Here the solid is subjected to a temperature difference between face B(hot) and A(cold) as in Fig. 15. As already discussed for the Seebeck effect, under the influence of the temperature gradient, electrons and holes will acquire drift velocities v_x^- and v_x^+ from B to A, as shown by arrows a and b. The corresponding electron and hole currents point along the negative and positive directions as shown by arrows c and d (note that $I_x^+ + I_x^- \equiv 0$). Under a magnetic field pointing out of the plane, the electrons and holes

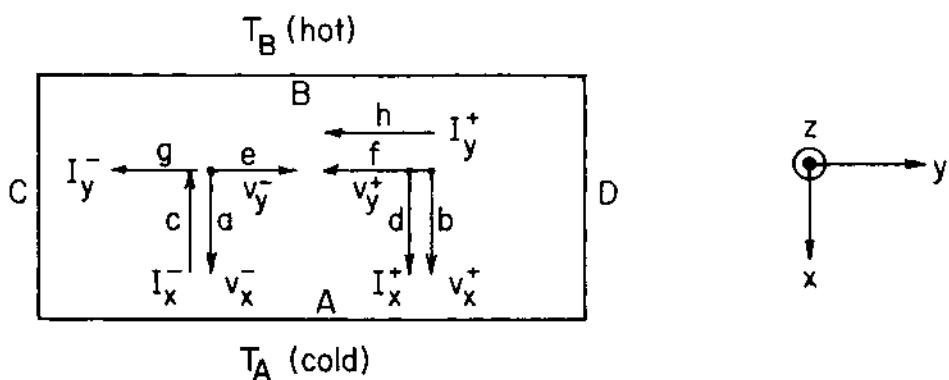


Fig. 15. Diagram illustrating the transverse Nernst effect.

acquire a component of velocity v_y^- to the right and v_y^+ to the left, in accord with the $\mathbf{I} \times \mathbf{H}/c$ (right-hand) rule of Lorentz; this movement of particles is represented by arrows e and f. From the earlier discussion of Hall effect it follows that there is initiated a flow of current towards the left due to holes (I_y^+) as well as a current due to electrons (I_y^-); this is represented by the arrows g and h. The flow can be made continuous if the external circuit is completed across faces C and D; alternatively, a gradient in electrochemical potential is developed across faces C and D under open circuit conditions. This field ultimately stops any net charge transfer. The latter phenomenon is known as the transverse Nernst effect.

Both the Ettingshausen as well as the transverse Nernst effects are very much smaller for one-band than for two-band semiconductors or metals, because in the one-band case one must rely on differences in trajectory between slow and fast moving carriers, with concomitant differences in thermal energies, to achieve the phenomena described above.

Righi-Leduc Effect. Here, a temperature gradient along x, $\nabla_x T$, when coupled with a magnetic field H_z along z, results in a temperature gradient $\nabla_y T$ along y. The explanation of this phenomenon is quite similar to that of the Ettingshausen effect, except that $\nabla_x T$ itself rather than an external field is used as a means of impelling charge carriers to move along the x direction.

IX. FERMI LEVELS AND CONDUCTIVITY IN INTRINSIC AND EXTRINSIC SEMICONDUCTORS

We examine here the case of semiconductor materials doped with donors and acceptors. A model for this case was discussed earlier in Section VII. We assume that N_d , the number of donors per unit volume, is in excess of N_a , the number of acceptors per unit volume: $N_d > N_a$. The discussion will be broken into three parts, described in conjunction with the energy level diagram shown in Fig. 6.

Low Temperature Region. In the low temperature limit, $kT \ll \varepsilon_g$. Let $\Delta \equiv N_d - N_a$ be the net number of electrons that are localized about donor sites in unit volume: this number is maximal at 0 K. The thermally activated process by which an electron is promoted from the donor levels into the conduction band will be represented by the chemical equation, $D \rightarrow D^+ + e^-$. Now let Δ_{occ} represent the density of the occupied and Δ_+ be the density of the ionized impurity levels. The ratio Δ_{occ}/Δ is governed by the Fermi-Dirac distribution function

$$\left(\frac{\Delta_{\text{occ}}}{\Delta} \right) = \frac{1}{1 + g_e e^{(\varepsilon - \xi)/kT}} \quad (105)$$

where g is a statistical weight factor which has a value of $\frac{1}{2}$ for electrons and 2 for holes. Since $\varepsilon_d - \xi \equiv (\varepsilon_d - \varepsilon_c) - (\xi - \varepsilon_c) \equiv -\varepsilon_d - \mu_c$, equation (105) takes the form,

$$\frac{\Delta_{\text{occ}}}{\Delta} = \frac{1}{1 + g_n e^{(-\varepsilon_d - \mu_c)/kT}} \quad (106)$$

Further, $\frac{\Delta_+}{\Delta} = 1 - \left(\frac{\Delta_{\text{occ}}}{\Delta} \right)$.

Let us now compute the quantity shown on the left side of Eqn. (107) below. After setting $n_n = \frac{\Delta_+}{\Delta}$ and $\frac{\Delta_{\text{occ}}}{\Delta} = \frac{\Delta - n_n}{\Delta}$ one obtains the term in the center involving n_n^2 . On the right we have substituted for n_n from Eqn. (111) derived below, and Δ_+/ Δ and $\Delta_{\text{occ}} / \Delta$, from Eqns. (105) and (106);

$$\frac{\Delta_+}{\Delta_{\text{occ}}} n_n \equiv \frac{\left(\frac{\Delta_+}{\Delta} \right)}{\left(\frac{\Delta_{\text{occ}}}{\Delta} \right)} n_n = \frac{n_n^2}{\Delta - n_n} = 2g_n \frac{(2\pi m_n kT)^{3/2}}{h^3} e^{-\varepsilon_d/kT} \quad (107)$$

At low temperature ($kT \ll \varepsilon_g$), $\Delta - n_n \approx \Delta = N_d - N_a$; hence we can solve Eqn. (107) on the right for

$$n_n = \left[2g_n (N_d - N_a) \right]^{\frac{1}{2}} \left[\frac{2\pi m_n kT}{h^2} \right]^{3/4} e^{-\varepsilon_d/2kT}. \quad (108)$$

According to Eqn. (111), one may write

$$\mu_c = kT \ln \left[\frac{n_n h^3}{2} \left(2\pi m_n kT \right)^{3/2} \right].$$

Using this relation to eliminate n from (108) one finds

$$\mu_c = \frac{-\varepsilon_d}{2} + kT \ln \left\{ \frac{\left[h^3 g_n (N_d - N_a)/2 \right]^{1/2}}{(2\pi m_n kT)^{3/4}} \right\}. \quad (109)$$

The above equation shows that at $T = 0$ K, where the second term vanishes, the Fermi level lies at $-\varepsilon_d/2$, i.e., half-way between

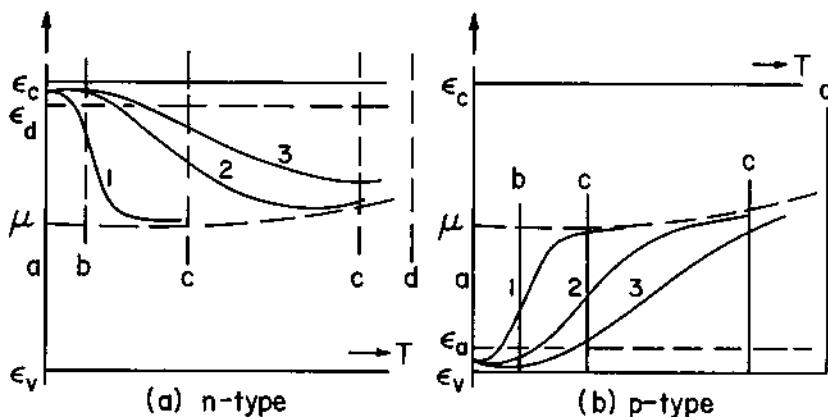


Fig. 16. Variation of Fermi level with temperature for a variety of cases of interest in semiconductors. Part (a): n-type material. Part (b): p-type material. 1,2,3 mark curves with progressively greater doping.

the donor level and the bottom of the conduction band. As T rises, the Fermi level will begin to move to higher energies as shown in part a-b of Fig. 16(a).

We now return to Eqn. (77);

$$\sigma = \frac{ne^2}{mv} = \frac{4e^2 \ln}{3(2\pi mkT)^{\frac{1}{2}}} .$$

After substituting for n from (108), we obtain the final result valid for low temperatures

$$\sigma = 4\sqrt{\frac{2g_n(N_d - N_a)}{3\hbar^{3/2}}} \cdot 1e^2 (2\pi m_n kT)^{\frac{1}{4}} e^{-\epsilon_d/2kT} . \quad (110)$$

Note that essentially $\sigma \sim \exp(-\epsilon_d/2kT)$ since $T^{\frac{1}{4}}$ in the preexponential factor has very little influence. Hence, a plot of $\ln \sigma$ against $1/T$ should yield a straight line whose slope is proportional to ϵ_d .

High Temperature Regions. We examine here the situation that prevails when kT becomes nearly comparable with ϵ_g in a semi-

conductor. As we discussed in Section VII, electrons are then thermally excited across the gap. We assume that the number of carriers so promoted is large compared to the number of electrons contributed by the donors or acceptors, yet small enough so that classical statistics may be employed as will be done below.

We had shown in Section V that the density of states is given by $g(\epsilon) = (1/2\pi^2)(2m/h^2)^{3/2}\epsilon^{1/2}$ where $m \equiv m_n$, $\epsilon = \epsilon - \epsilon_c$ for electrons, and $m \equiv m_p$, $\epsilon = \epsilon_v - \epsilon$ for holes. We now evaluate the carrier density for the case where $-\mu/kT \gg 1$, which requires that the Fermi level lie in the gap, so that $\epsilon - \epsilon_c \equiv \mu_c$ or $\epsilon_v - \epsilon \equiv \mu_v$ remains negative. When the above condition is met, $\exp(\epsilon - \mu)/kT \gg 1$ and the Fermi-Dirac distribution function goes over into the Maxwell-Boltzmann distribution function $f = \exp(\mu/kT) \cdot \exp(-\epsilon/kT)$. Accordingly, in this range we find

$$\begin{aligned} n_n &= \int_0^\infty g(\epsilon) f(\epsilon) d\epsilon \approx \frac{1}{2\pi^2} \left[\frac{2m}{h^2} \right]^{3/2} e^{\mu_c/kT} \int_0^\infty \epsilon^{1/2} e^{-\epsilon/kT} d\epsilon \\ &= 2 \left[\frac{2\pi m_n kT}{h^2} \right]^{3/2} e^{\mu_c/kT} \equiv N_c e^{\mu_c/kT} \end{aligned} \quad (111)$$

where N_c is called effective density of states for the conduction band. We may solve this expression for μ in terms of n as follows:

$$\mu_c = kT \ln \left[nh^3 / 2(2\pi m_n kT)^{3/2} \right]. \quad (112)$$

A corresponding relation is found for holes in the valence band:

$$n_p = 2 \left[\frac{2\pi m_p kT}{h^2} \right]^{3/2} e^{\mu_v/kT} \equiv N_v e^{\mu_v/kT}. \quad (113)$$

Now under the conditions discussed earlier the intrinsic semiconductor can be characterized by the relation obtained from (111) and (113) namely,

$$n_n n_p = 4(2\pi(m_n m_p)^{1/2} kT/h^2)^3 e^{-\epsilon/kT} \equiv n_i^2 = N_c N_v e^{-\epsilon/kT} \quad (114)$$

where the quantity characterized by n_i^2 is independent of the Fermi level and constant at constant T.

To determine the Fermi level in the intrinsic material we set

$$n_n = n_p = N_c e^{(\epsilon_c - \epsilon_c)/kT} = N_v e^{(\epsilon_v - \epsilon_c)/kT}$$

and solve for $e^{\epsilon_c/kT}$, giving

$$\begin{aligned} &= \frac{1}{2} (\epsilon_v + \epsilon_c) + \frac{1}{2} kT \ln(N_c/N_v) \\ &= \frac{1}{2} (\epsilon_v + \epsilon_c) + kT \ln(m_p/m_n)^{3/4}. \end{aligned} \quad (115)$$

We note that if the material remained intrinsic down to 0 K, i.e., if the impurity donor and acceptor levels were absent, then the Fermi level would fall midway in the gap. As the temperature rises the Fermi level increases if $m_p > m_n$ (as is usually the case); this situation is depicted by the dotted line in Fig. 16. Actually, impurities are always present in any material; these dominate the properties of the solid in the range $kT \ll \epsilon_g$; hence one would always expect the Fermi level to start off at $-\epsilon_d/2$ or $-\epsilon_a/2$ and to follow the curve a-b in Fig. 16. As discussed in Section VII, on the other hand, for $kT \sim \epsilon_g$, the regime described here dominates and the Fermi level will then follow the dotted curves c-d in Fig. 16. Clearly, there will have to be an intermediate temperature range, discussed below, where a transition from b to c occurs, as sketched in Fig. 16, which smoothly joins segments ab with cd.

To determine the conductivity of the intrinsic material we must add the partial currents due to the directed movement of carriers in the valence and conduction bands, $I_+ = n_p e v_p$ and $I_- = -n_n e v_n$; thus, $I = e(n_n u_n + n_p u_p) E = \sigma E$. It follows that for intrinsic material where $n_n = n_p \equiv n_i$,

$$\sigma = e n_i (u_n + u_p) = e n_i u_p (b + 1) \quad (116)$$

where $b \equiv u_n/u_p$. Finally, we substitute for n_i from (114); this yields

$$\sigma = 2e u_p (b + 1) [2\pi(m_n m_p)^{1/2} kT/h^2]^{3/2} e^{-\epsilon_g/2kT}. \quad (117)$$

Generally the variation of $T^{3/2}$, $(m_n m_p)^{1/2}$, b , and u_p with temperature is small compared to that of the exponential factor. Hence, a plot of $\ln \sigma$ vs $1/T$ should yield a straight line whose slope is proportional to ϵ_g .

Intermediate Temperature Range. We examine here the temperature range in which the extrinsic regime gradually goes over into the intrinsic regime. The derivations are based on the law of

electroneutrality according to which

$$n_p - n_n + N_d - N_a + n'_p - n'_n = 0, \quad (118)$$

where n is the free carrier density, n' is the density of carriers bound to donor or acceptor levels, and N is the density of impurities. Under the conditions examined here we assume $n' \ll n$ but we do not necessarily assume that the intrinsic regime dominates the conduction properties. In this event, the last two terms may be dropped and Eqns. (111) and (113) utilized to substitute in the first two. This yields

$$\frac{N_v e^{-(\epsilon_v - \epsilon)/kT}}{N_c e^{-\epsilon_c/kT}} + N_d - N_a = 0.$$

The above is a quadratic equation in $e^{+\epsilon/kT}$ which may be solved for this variable and then for ϵ . One obtains

$$\frac{\epsilon}{kT} = \ln \left\{ \frac{N_d - N_a}{2N_c e^{-\epsilon_c/kT}} + \left[\left(\frac{N_d - N_a}{2N_c e^{-\epsilon_c/kT}} \right)^2 + \frac{N_v e^{\epsilon/kT}}{N_c e^{-\epsilon_c/kT}} \right]^{\frac{1}{2}} \right\}. \quad (119)$$

Now making use of the identity

$$\ln(a + \sqrt{a^2 + x^2}) \equiv \ln x + \sinh^{-1}(a/x),$$

Eqn. (119) may be rewritten as

$$\epsilon = \frac{1}{2}(\epsilon_v + \epsilon_c) + kT \ln(\frac{m_p/m_n}{N_d - N_a})^{3/4} + kT \sinh^{-1} \left[\frac{(N_d - N_a)/2n_i}{N_d - N_a} \right] \quad (120)$$

where (111) and (113) were introduced to obtain the above result.

It is clear that Eqn. (120) is a generalization of (115). Several limiting cases are now of interest: (a) If $N_d = N_a$, the last term in (120) drops out and the material behaves exactly as a true intrinsic semiconductor, except in the very lowest temperature ranges where extrinsic properties may become noticeable. The material is said to be a compensated semiconductor. (b) If $N_i \ll |N_d - N_a|$, one may apply the relation $\sinh^{-1} x \approx \pm \ln |2x|$, valid for large x , to find

$$\epsilon = \frac{1}{2}(\epsilon_v + \epsilon_c) + kT \ln(\frac{m_p/m_n}{N_d - N_a})^{3/4} \pm kT \ln(|N_d - N_a|/n_i) \quad (121)$$

where the upper and lower signs apply to n and p type material respectively. This formula corresponds to the changes in Fermi level in regions b-c on the diagram of Fig. 16.

The conductivity corresponding to this range is found after determining the carrier densities according to Eqn. (114); $n_n n_p = n_i^2$. This relation holds regardless whether n_n and n_p are equal or not; n_i^2 is simply an equilibrium constant with a fixed value at a given temperature. On substituting this relation in the electroneutrality condition (118) and eliminating n_p we obtain, $n_n^2 + (N_d - N_a) n_n - n_i^2 = 0$, with the solution

$$n_n = \frac{1}{2}(N_d - N_a) + \sqrt{\left[\frac{1}{4}(N_d - N_a)^2 + n_i^2\right]} . \quad (122)$$

Similarly, we find

$$n_p = \frac{1}{2}(N_d - N_a) + \sqrt{\left[\frac{1}{4}(N_d - N_a)^2 + n_i^2\right]} . \quad (123)$$

We now introduce the above expressions in the relation (115) for $\sigma = e u_p (b n_n + n_p)$ to obtain

$$\sigma = e u_p \left\{ \frac{1}{2}(b - 1)(N_d - N_a) + (b + 1)\sqrt{\left[\frac{1}{4}(N_d - N_a)^2 + n_i^2\right]} \right\} . \quad (124)$$

Again, several limiting cases are of interest here. (a) When $n_i \ll |N_d - N_a|$ the square root may be expanded to first order terms and the above reduces to

$$\sigma = (N_d - N_a) b e u_p = (N_d - N_a) e u_n . \quad (125)$$

One observes that the charge carrier density does not enter this expression and that σ varies with T only as u_n does. We thus deal here with the exhaustion region mentioned in Section VII. (b) If, on the contrary, $n_i \ll |N_d - N_a|$ then expansion of the square root leads to the result $\sigma \approx e u_p n_i (b + 1)$, in agreement with Eqn. (116).

X. TRANSPORT EFFECTS IN MATERIALS CHARACTERIZED BY 'HOPPING' ELECTRONS

The Electrical Conductivity. So far, we have examined the transport phenomena encountered when a charge carrier moves through

a band. We turn now to the opposite case described earlier, in which the electron is essentially localized about atomic centers in the lattice but able to move from one site to an adjacent one by an activated hopping process.

We begin the study of conduction processes by turning to Eqn. (89), according to which the current density is given by $J = \sigma \nabla (\tau/e)$, which can be rewritten as

$$\mathbf{J} = \sigma \nabla \frac{\mu_n}{e} - \nabla \phi_s \quad (126)$$

since $\tau \equiv \mu_n - e\phi_s$. The chemical potential itself is specified through the relation

$$\mu_n = \mu_0 + kT \ln a_n \quad (127)$$

where μ_0 is the chemical potential in the standard state and a_n is the activity of the charged species. Thus, (126) may be rewritten as

$$\mathbf{J} = \sigma \left(\frac{kT}{ea_n} \right) \nabla a_n - \nabla \phi_s = -eJ_n \quad (128)$$

where $J_n = -J/e$ is the particle flux vector. Consider now the case where $\nabla \phi_s \equiv 0$; then

$$J_n = - \left[\frac{\sigma kT}{e^2 a_n} \right] \nabla a_n \equiv -D \nabla a_n. \quad (129)$$

Here we have introduced the diffusion coefficient D on the right since J_n has been found to be proportional to the gradient in activity (i.e., essentially $J_n \sim \nabla a_n$, in accord with Fick's law). We may thus identify

$$\sigma = \frac{e^2 a_n D}{kT} \quad (130)$$

which is the Einstein law of diffusion. An expression for the diffusion coefficient can be obtained as follows: Consider the crystal lattice as parallel plane array of atoms separated by a lattice constant, and suppose electrons at their cation lattice positions can jump from one site to another on adjacent lattice planes as shown in Fig. 17. The jump processes will take place, for example, if we have the same cation in different valence states (as, for example Pr^{3+} and Pr^{4+} in PrO_x or Ni^{2+} and Ni^{3+} in

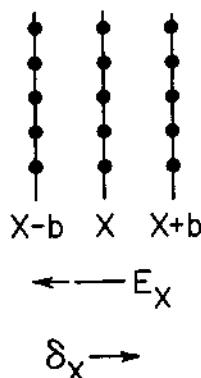


Fig. 17. Parallel lattice planes containing lattice atoms.
Planes are oriented perpendicular to applied field.

$\text{NiO} \cdot (\text{Li})$), and if the requisite activation energy is available. Schematically, the electron transfer can be represented by the reaction, $M_a^{2+} + M_b^{3+} \rightarrow M_a^{2+} + M_b^{2+}$ where M_1^{n+} is a cation. Now, if θ is the fraction of occupied sites in the lattice, $(1 - \theta)$ is the fraction unoccupied. Also, if v is the attempt frequency for a jump process and δ_x the probability of a jump process being successful along the field direction $-E_x$, then $v \delta_x (1 - \theta)$ specifies the probability of successful jumps in unit time along the negative direction of the external field. Let $n(x)$ be the density of carriers on plane x at time t per unit volume. Then $n(x + b)$ and $n(x - b)$ will be the corresponding density on adjacent planes at time t . Expanding $n(x + b)$ and $n(x - b)$ for small b in Taylor's Series, we have

$$n(x + b) = n(x) + bn'(x) + \frac{1}{2}b^2n''(x) + \dots \quad (131)$$

$$n(x - b) = n(x) - bn'(x) + \frac{1}{2}b^2n''(x) - \dots \quad (132)$$

Now, the change in the carrier density in the given plane $dn(x)$ in time dt is the difference in the net 'inflow' and net 'outflow' of the charge carriers, which involves the number of successful jumps. Therefore,

$$dn(x) = dt [\{n(x + b) + n(x - b)\} - \{2n(x)\}] \delta_x (1 - \theta)v \quad (133)$$

From (131), (132) and (133),

$$\frac{\partial n}{\partial t} = \delta_x (1 - \theta) v b^2 n''(x) \equiv D \frac{\partial^2 n}{\partial x^2}. \quad (134)$$

Here again we have introduced the diffusion coefficient D since the equation which connects $\partial n / \partial t$ to $\partial^2 n / \partial x^2$ is a diffusion-type differential equation. Accordingly,

$$D = \delta_x (1 - \theta) v b^2. \quad (135)$$

Next, v will be expressed as

$$v = v^+ e^{-\Delta G^\ddagger/kT} \quad (136)$$

where we have adopted Eyring's theory of rate processes, v^+ being the preexponential factor, and ΔG^\ddagger is the free energy of activation. Eqn. (136) now becomes

$$\sigma = \frac{(1 - \theta) v^+ a_n e^2 \delta_x b^2}{kT} e^{-\Delta G^\ddagger/kT}. \quad (137)$$

If C sites per unit volume are present in the crystal and V is the total volume then, $(C/V)\theta = n_n = a_n/\gamma_n$ where γ_n is the activity coefficient of the charge carriers. Now $C/V = c/v_0$ where c is the number of centers available per unit cell whose volume is v_0 ($\equiv b^3 \delta_2$)

$$a_n = \frac{c\theta}{b^3} \cdot \frac{\gamma_n}{\delta_2}. \quad (138)$$

Here δ_2 is a conversion factor to cubic symmetry for the crystal ($\delta_2 = 1$ for cubic crystals). On introducing (138) in (137) we obtain

$$\sigma = K \frac{\theta(1 - \theta)e^2 v^+}{kT} e^{-\Delta G^\ddagger/kT} \quad (139)$$

where K is a collection of constants. Eqn. (139) is the conductivity equation applicable for 'hopping' type materials.

Some interesting conclusion can be drawn from Eqn. (139):
 (1) We note that if the concentration of charge carriers is constant (this is true when the composition of the solid is fixed, e.g. PrO_x) $\log(\sigma T)$ is proportional to $1/T$; thus, from the slope of the plot the free energy of activation for conduction can be obtained. This shows explicitly how the increase in conductivity is due to the thermally activated mobility. (ii) If θ is zero or

unity, $\sigma \rightarrow 0$, indicating that the material becomes an insulator as the stoichiometric composition is reached. When $\theta = \frac{1}{2}$, σ tends to be a maximum. This is entirely reasonable: for $\theta = 0$ or $\theta = 1$ there are either no carriers or no positions to which they can jump. For $\theta = \frac{1}{2}$ both the numbers of carriers and the number of empty positions to which they can move is optimized.

Thermoelectric Power of Materials Characterized by 'Hopping' Electrons. We have seen earlier that the general equation for the thermoelectric power or Seebeck coefficient, α , is of the form

$$\alpha = -\frac{s^*}{e} = -\frac{1}{e} [S_c^* + S_T^*] \quad (140)$$

where S_c^* and S_T^* are contributions to the entropy due to the electronic configuration and due to the thermal energy of the electron in its site.

If in first approximation we consider the carriers as localized then the total configurational entropy, S_c can be calculated as

$$S_c = k \ln \Omega = k \ln \frac{c!}{(c-n)! n!} \quad (141)$$

where Ω is the number of ways of arranging the different particles, n , among various available sites, c . Using Stirling's approximation, $\ln n! \approx n \ln n - n$, valid for large n , one finds that

$$S_c = k [c \ln c - n \ln n - (c - n) \ln (c - n)] . \quad (142)$$

Hence

$$S_c^* \equiv \frac{\partial S_c}{\partial n} = k \ln \left(\frac{c-n}{n} \right) = k \ln \left(\frac{1-\theta}{\theta} \right) \quad (143)$$

since $n/c = \theta$. From (142) and (143), it follows that

$$\alpha = -\frac{1}{e} \left[k \ln \left(\frac{1-\theta}{\theta} \right) + S_T^* \right] = -\frac{k}{e} \left[\ln \left(\frac{1-\theta}{\theta} \right) + \frac{S_T^*}{k} \right]. \quad (144)$$

From Eqn. (144) we see that α is large and negative if $\theta = 0$ and large and positive if $\theta = 1$; also, α changes sign as θ varies from 0 to 1, since in general $S_T^*/k < \ln [(1-\theta)/\theta]$. Further we note that for a given value of θ , α is roughly independent of temperature and any temperature variation should only involve the S_T^* term. The physical interpretation of Eqn. (144) is offered in Section XI.

Thermal Conductivity for Materials Characterized by 'Hopping' Charge Carriers. According to the Wiedemann-Franz law, $k_e \propto \sigma$. Since σ is very small for 'hopper-type' materials, at not too high temperatures, we can assume that the electronic part of the thermal conductivity, k_e , is small compared to the lattice contribution to k .

XI. BRIEF SURVEY OF EXPERIMENTAL RESULTS

We present below a highly selective survey of experimental results that illustrate some of the general principles described in the earlier sections.

Figs. 18 and 19 illustrate the resistivity and Hall coefficients⁴ of several Ge samples doped with As in varying degrees. In the conductivity curves one must distinguish between three regions: a) The low temperature range ($T \leq 20K$), where for the purest samples the plot of $\ln \sigma$ vs $1/T$ yields a straight line with a slope proportional to ϵ_d , in conformity with Eqn. (110). Here the conductivity rises with temperature as progressively more charge carriers are thermally excited into the conduction band. b) The exhaustion range ($300 \geq T \geq 20K$), where the number of charge carriers remains constant, since all donor levels are ionized; the conductivity as given by Eqn. (128) now diminishes with rising temperature, since u_n is found to vary as T^{-n} ($n \approx 3/2$). c) The intrinsic region ($T \geq 300K$), where the plot of $\ln \sigma$ vs. $1/T$ yields a straight line with a slope proportional to ϵ_g in accordance with Eqn. (117). As the impurity content is raised, region (a) becomes progressively less distinct. Ultimately, for $N_d - N_a \approx 10^{18} \text{ cm}^{-3}$ the average separation distance between impurity centers falls below a critical value, at which point impurity band formation sets in; this explains the property of the topmost curve in Fig. 18. The reader can readily distinguish the same three regions in Fig. 19, where the charge carrier density as computed from the Hall coefficient is shown in its dependence on T .

The remainder of this section is devoted to the electrical properties of oxides. In this connection we introduce one operational definition; we regard materials as characterized by band-like or collective electron properties if the carrier mobility is nonactivated; that is, if u varies as T^{-n} , rather than showing the activated T dependence $u \sim e^{-\epsilon/kT}$. As discussed in Section VII, this grouping comprises the metals, overlap metals, and the

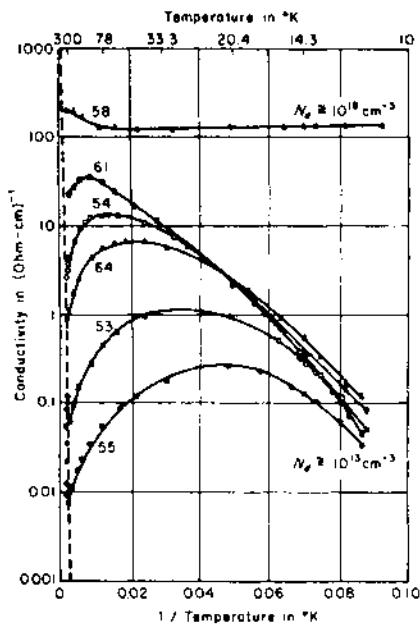


Fig. 18. Conductivity of various n-type Ge samples doped with As. After ref. 4.

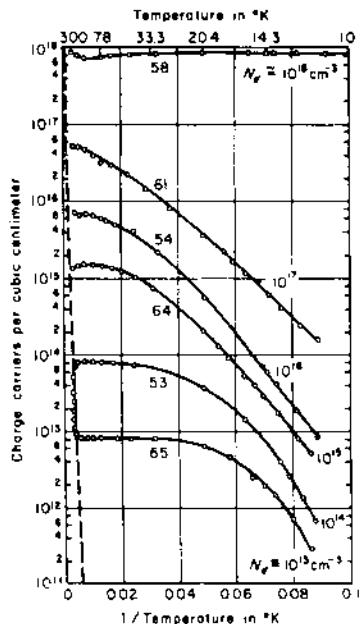


Fig. 19. Charge carrier density of n-type Ge for various As dopings. After ref. 4.

band-type semiconductors and insulators. Note that in the latter two categories the conductivity (though not the mobility) rises exponentially with T , because the thermal promotion of carriers into band states is an activated process ($n \sim e^{-E/kT}$). If, on the contrary, the mobility itself is activated, then the material is classified as exhibiting 'hopping' or 'localized' electronic properties. Such materials often are also classified as semiconductors or insulators, but clearly they differ in their properties from their band-like counterparts, as already discussed in Section VII. Next, there is a category termed 'small polaron' materials, which exhibits band-like properties at low temperature ($T \leq \theta/2$), and hopping properties at high temperature ($T \geq \theta/2$), θ being the Debye temperature. Here, as discussed in detail by Holstein⁵, the

band progressively narrows with rising T , until near $\theta/2$ a transition to the localized regime takes place; these matters are also described qualitatively in some detail in an article by Schmid⁶.

Finally, there is an intermediate variant of the above, namely the case of 'large' polarons, where carriers interact sufficiently with the lattice to cause significant polarization distortions and fluctuations as the electrons move through the lattice. The charge carriers are then said to be 'dressed' by the polarization; they exhibit relatively large effective masses and low mobilities.

The oxides run the entire gamut of electrical characteristics described above*. The room temperature resistivities of oxides range from 10^{-5} to 10^{+15} ohm-cm, spanning some twenty orders of magnitude. There is no question that there exist oxides such as firebrick or zirconia which at 300 K are excellent insulators. On the other hand, there also exist oxides such as RuO_2 or ReO_3 which are more highly conducting than the metals from which they are derived.

Specifically, at 1.1K, RuO_2 has a resistivity of 10^{-10} ohm-cm⁷, quite comparable with that of the best obtainable copper metal. However, RuO_2 exhibits an extremely large temperature coefficient of resistivity, so that at 300 K, $\rho \approx 10^{-5}$ ohm-cm, which exceeds that of Cu by one order of magnitude. The variation of ρ with T for RuO_2 ⁸ is shown in Fig. 20.

The roughly linear increase of ρ with T for metals is in accord with predictions based on Eqn. (78), which shows that $\rho \sim l^{-1}$. As temperature rises there is increasing thermal disorder in the solid; this shortens the mean free pathlength l and thereby increases the resistivity of the sample. In many cases the increase is a nearly linear function of T .

The resistivity of several ReO_3 single crystal specimens is shown and compared to that of copper⁹ in Fig. 21; again, the metallic character of this material is self-evident from an examination of the data. These findings are corroborated by reflectivity studies¹⁰ of plasma edge as well as interband transitions in the

*For recent reviews on electrical transport and related properties of oxide materials, see the following references:

1. D. Adler, Solid State Phys., 21, 1 (1968).
2. A.T. Howe and P.J. Fensham, Quart. Rev. (Lond.), 21, 507 (1967).
3. C.N.R. Rao and G.V. Subba Rao, Phys. Stat. Sol., (a), 1, 597 (1970).

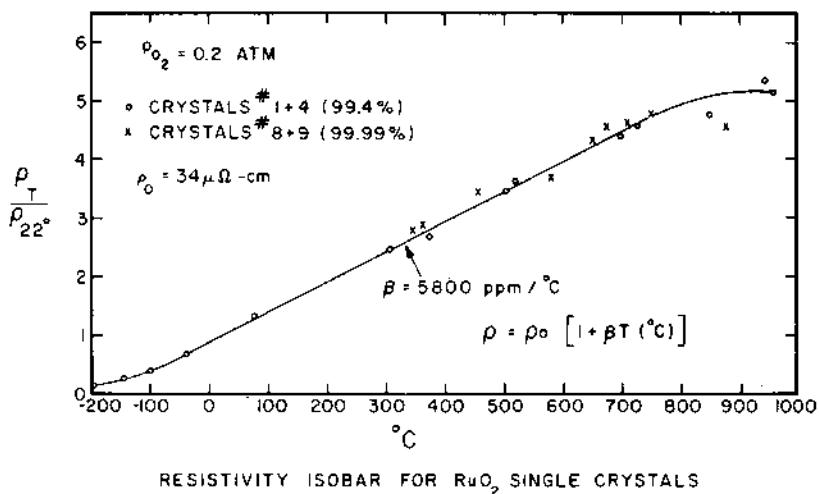


Fig. 20. Variation of resistivity with temperature for RuO_2 . After ref. 8.

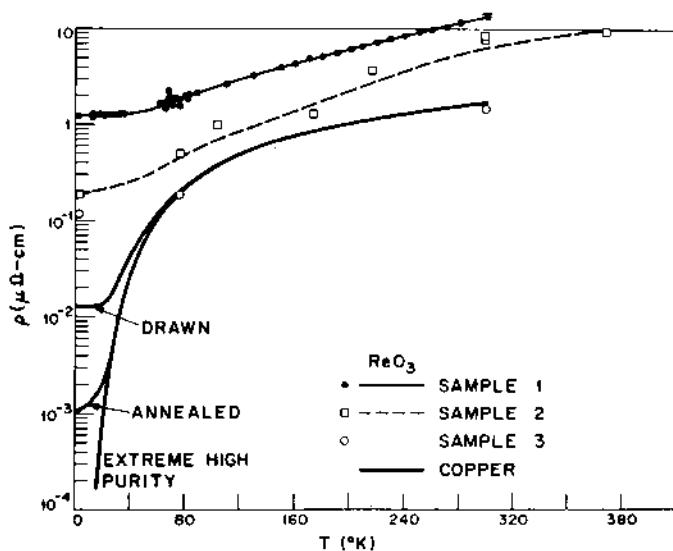


Fig. 21. Electrical characteristics of ReO_3 as a function of temperature. After ref. 9.

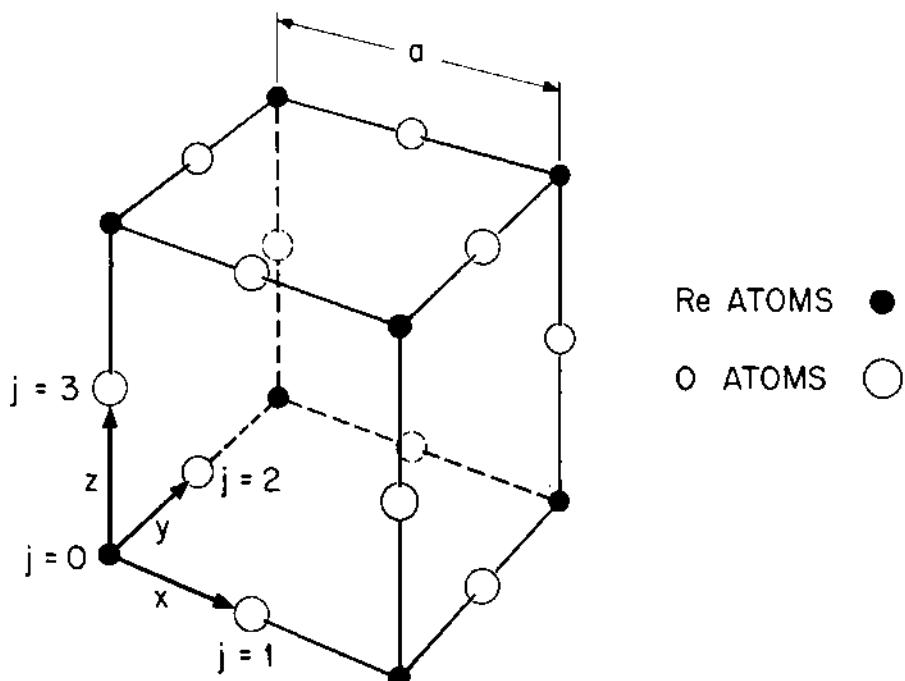


Fig. 22. Arrangement of atoms in perovskite lattices.

visible and ultraviolet region of the spectrum; in fact, ReO_3 looks like pinkish copper metal.

These observations may be rationalized by a scheme proposed by Goodenough^{11,12}. It is evident from the perovskite-type symmetry displayed by ReO_3 (see Fig. 22) that each cation is octahedrally surrounded by anions and that each anion is linearly surrounded by cations as nearest neighbors. Focussing attention to Pauling's rules, the $6s$, $6p_x$, $6p_y$, $6p_z$, $5d_{x^2-y^2}$ and $5d_{z^2}$

orbitals of Re to obtain six σ -type wave functions, with large lobes pointing directly towards the nearest neighbor oxygen atoms. Similarly, for each of the oxygens, one hybridizes the $2s$ and either $2p_x$ or $2p_y$ or $2p_z$, as appropriate, to obtain a σ -type wave function with large lobes pointing directly at the two nearest neighbor cations. As a result, one forms six bonding σ -type and six antibonding σ^* -type states per ReO_3 unit, produced in analogy to molecular orbital theory. Note that orbital overlap occurs continuously throughout the crystal. Thus, instead of molecular orbital levels one obtains band states labelled σ and σ^* , with a

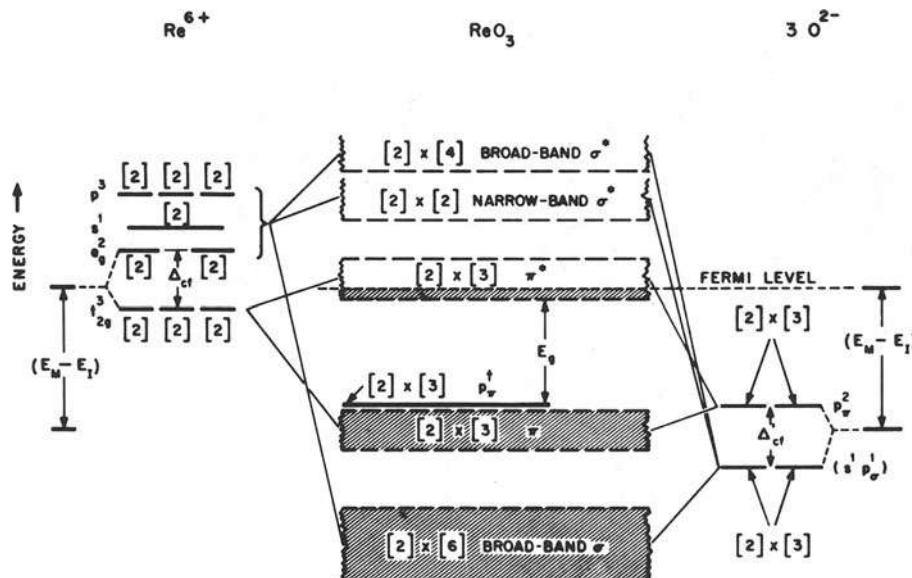


Fig. 23. Band structure scheme for ReO_3 . Energy levels on the right and left are atomic levels; band formation is shown in center. After ref. 11.

capacity each of 12 electrons per ReO_3 formula unit. This is displayed schematically in Fig. 23.

Examination of Fig. 24 shows further that the three unused cationic 5d_{xy} , 5d_{xz} , and 5d_{yz} orbitals can be overlapped in π -type fashion with the appropriate p_π type orbital of the oxygen to form three bonding π states and three antibonding π^* states per ReO_3 unit, again with continuous orbital overlap through the entire crystal. Correspondingly, one forms the band states labelled π and π^* in Fig. 23; each has a capacity for six electrons per ReO_3 formula unit. These bands are narrower than their σ -type counterparts because of the smaller overlap between wave functions, linked in π -type bonding. Finally, there remain three unused p_π orbitals associated with the oxygen; these point at each other across the cube faces but are separated too far for appreciable overlap to take place: in consequence, these orbitals remain nonbonding; they represent three atomic states with a total capacity for six electrons per ReO_3 formula unit. In now accommodating the 25 electrons in the outer valence shell of ReO_3 , the σ , π , and nonbonding states are filled by 24 electrons, and the remaining electron must therefore be accommodated in the π^* band, which

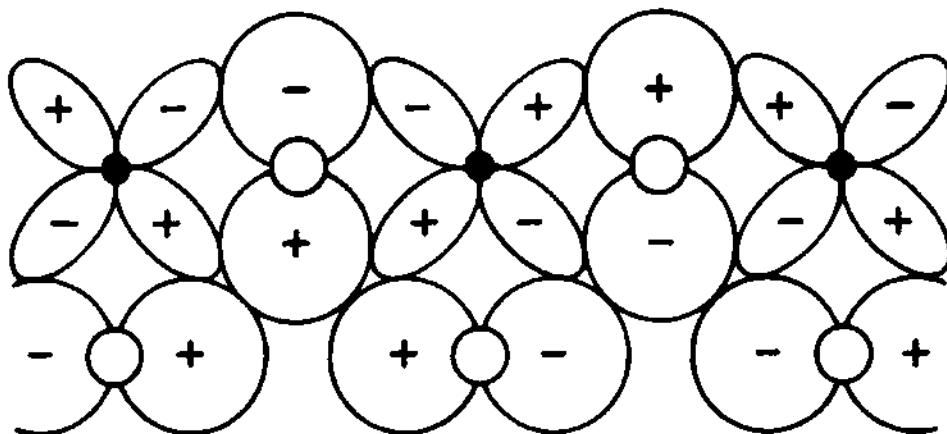


Fig. 24. Overlap of W-type in ReO_3 by the APW approximation. After ref. 14.

thereby is only partly filled. The observed metallic characteristics of this material are thus immediately rationalized.

The essential correctness of this model has recently been established by solving the Schrödinger wave equation in the tight-binding approximation¹³ for ReO_3 . The separation of bands into σ , π , and nonbonding types does obtain throughout the entire Brillouin zone under a set of very stringent though not unphysical assumptions invoked in a zero-order scheme. The dispersion relations $\epsilon(k)$ appropriate to the triply degenerate π^* band were also obtained in this approximation. However, in a less restrictive treatment, the energy band structure becomes much more complex and the simple separation of band states into various categories can no longer be maintained. This fact was also noted by Mattheiss¹⁴, who made numerical calculations, coupled with some fitting to experimental data. His final results are shown in Fig. 25 in a plot of ϵ vs. k for the s, t_{2g} , and e_g -like band states along the principal symmetry directions in reciprocal space. It further turns out to be necessary to take spin-orbit splitting of the various band states into account. This results in a removal of various degeneracies. However, these complications, considerable

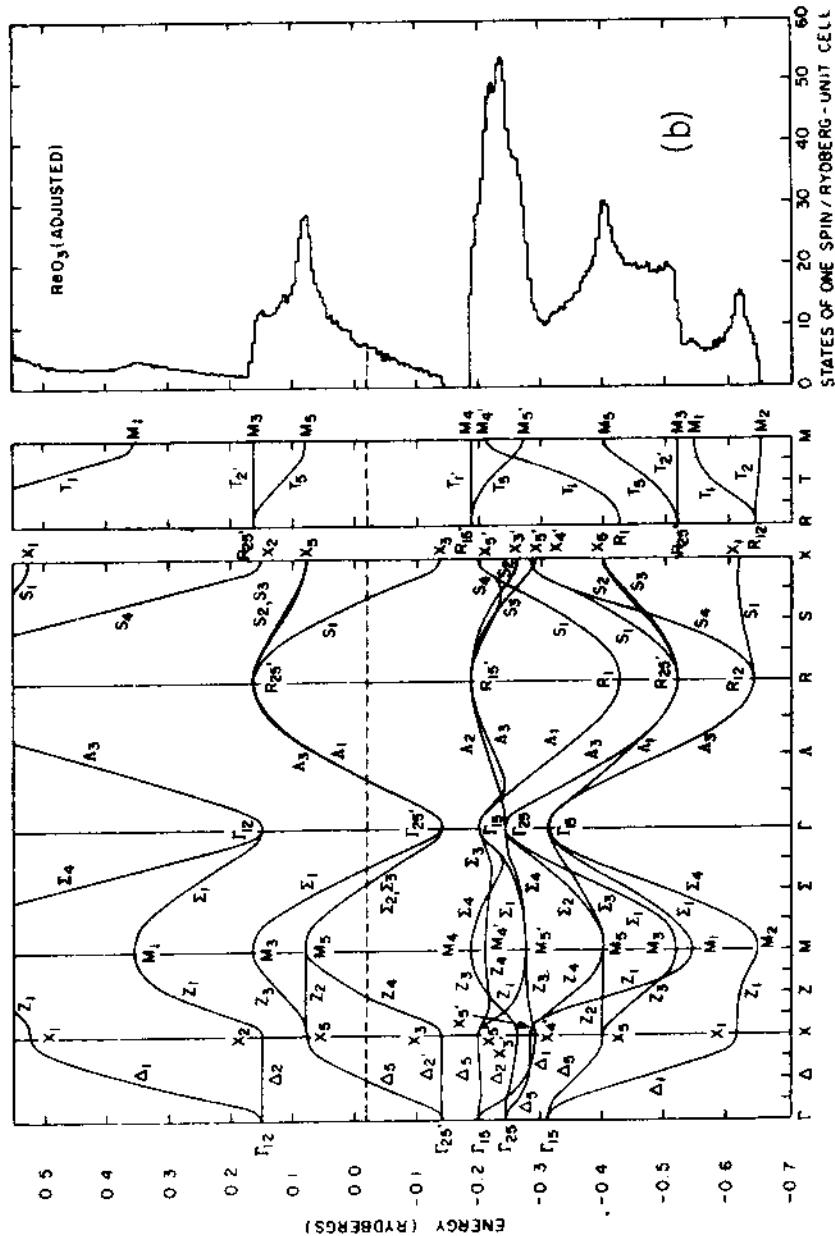


Fig. 25. Energy band calculations for ReO_3 by the APW approximation. After ref. 14.

though they are, should not obscure the essential correctness of the elegant qualitative approach pioneered by Goodenough.

We turn now briefly to several other classes of metallic oxides. TiO_x can exist over a wide range of nonstoichiometric compositions ($0.75 \leq x \leq 1.25$), with up to 15% of vacancies in both the cationic and the anionic sublattices. As x increases the appearance of the material changes from silvery to brass, to gold. Various physical properties of this compound have recently been studied in considerable detail¹⁵. According to very recent studies, NbO_x ($0.98 < x < 1.02$) appears to be an overlap metal with a plasma edge at ≈ 4 ev, which causes it to display a silvery luster^{16,17}.

CrO_2 is a metal which is simultaneously ferromagnetic; it has a magnetic moment of two Bohr magnetons and a Curie temperature of $T_c \approx 190^\circ C$. Controversies concerning the electrical properties of this compound seem to have been settled in favor of its metallic characteristics. In Fig. 26 is shown a plot of α/T vs. T for highly compressed CrO_2 pellets¹⁸. According to the model of Eqn. (99) this should yield a straight horizontal line for $T > T_c$, as is indeed the case. In the range $T < T_c$, the effect of rising T is to increase the spin disorder and thereby, to raise the entropy, S^* , and hence, $-\alpha_n$. Complementary results have been

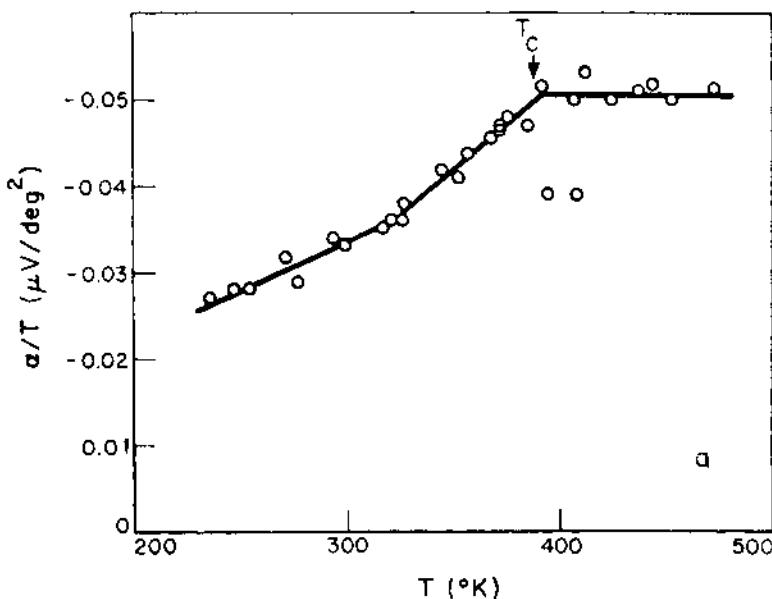


Fig. 26. Seebeck coefficient of CrO_2 as a function of temperature. After ref. 18.

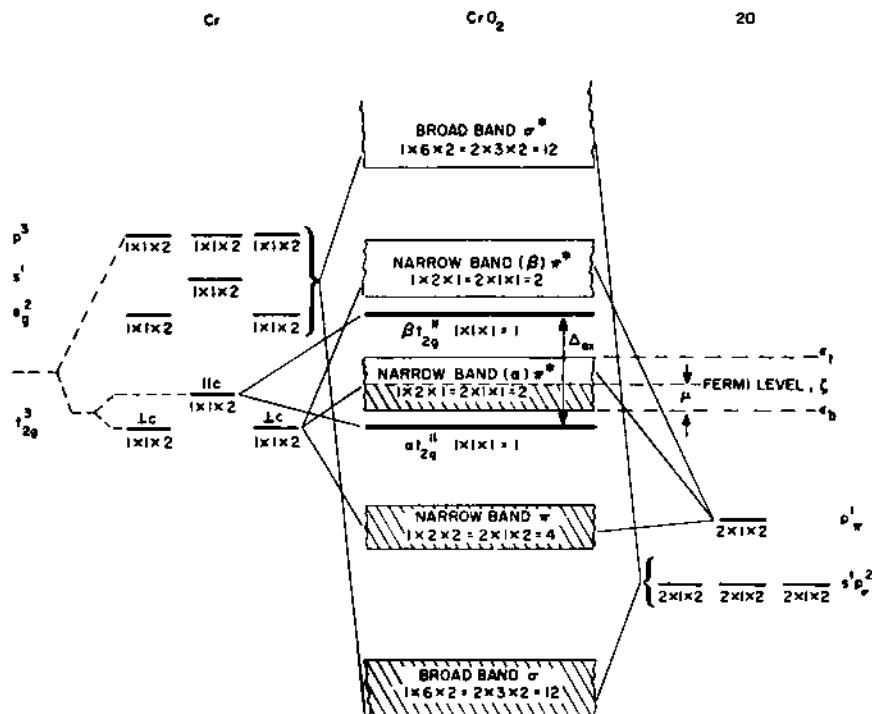


Fig. 27. Energy level diagram for CrO_2 . After ref. 11.

reported for the resistivity of epitaxially grown CrO_2 single crystals¹⁹. Again, the normal rise of ρ with T is observed only for $T > T_c$; in the lower T range the effect of increasing spin disorder is detectable in terms of an abnormal, sigmoidal increase of ρ with T . The various properties of CrO_2 may be rationalized with Fig. 27, proposed originally by Goodenough¹¹. In first approximation, the Cr are octahedrally surrounded by O and the O are trigonally surrounded by Cr in the rutile structure. After appropriate hybridizations, one obtains the σ , σ^* , π , π^* and nonbonding states schematically depicted in the figure. Sixteen of eighteen valence electrons per CrO_2 formula unit are accommodated in the σ and π bands. One then has a choice: The two remaining electrons may now be accommodated with opposing spins ($S = 0$) in the nonbonding t_{2g} level; however, Hund's rule would then be violated. Alternatively, the t_{2g} nonbonding and π^* antibonding states may be split by the exchange energy into α - and β -substates. The two electrons are then accommodated with anti-parallel spin ($S = 1$) into the α -substates. This would mean that the localized electrons in the t_{2g} states are ferromagnetically coupled by a like number of itinerant electrons in the π^* band.

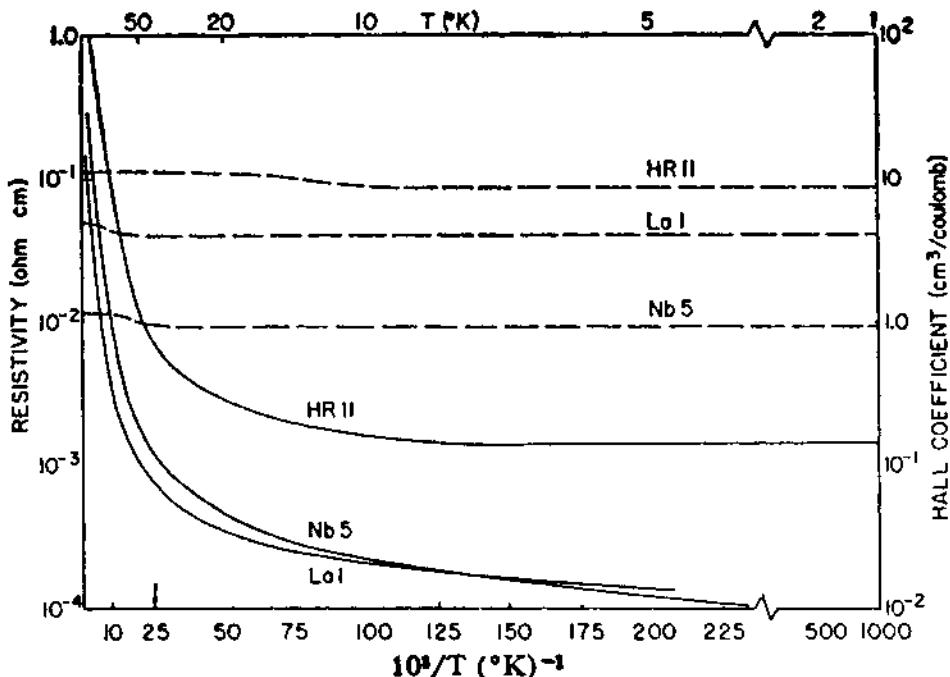


Fig. 28. Resistivity and Hall coefficient of SrTiO_3 as a function of temperature. After ref. 20.

which is half-filled. Until better models become available, Goodenough's rationalization serves very well in a qualitative explanation of the data.

We turn next to a second class of materials, namely the perovskites of type ABO_3 which become extrinsic semiconductors when doped with altervalent cations that take their place in the A position, or when the material is rendered deficient in oxygen. SrTiO_3 has been studied in considerable detail²⁰⁻²². A plot of σ and R vs. $1/T$ is shown in Fig. 28: by virtue of the extremely high dielectric constant $\epsilon \sim 1000$, all donor impurities are essentially ionized at 1 K, so that R remains independent of T ; the enormous variation of ρ with T must thus be ascribed to a strong variation of mobility with temperature. This is shown explicitly in Fig. 29 for $200 < T < 1200$ K. For KTaO_3 mobilities of the order of $250,000 \text{ cm/V-sec}$ have been reported at 4.2 K, - a stupefying value, considering that one is dealing with an oxide. On the other hand, since u varies as $\sim T^{-3}$, the mobility of SrTiO_3 reaches a value of 1 cm/V-sec at ~ 500 K, in which range the material switches over to the hopping regime. Above 900 K, Walters and Grace²² successfully interpreted the properties

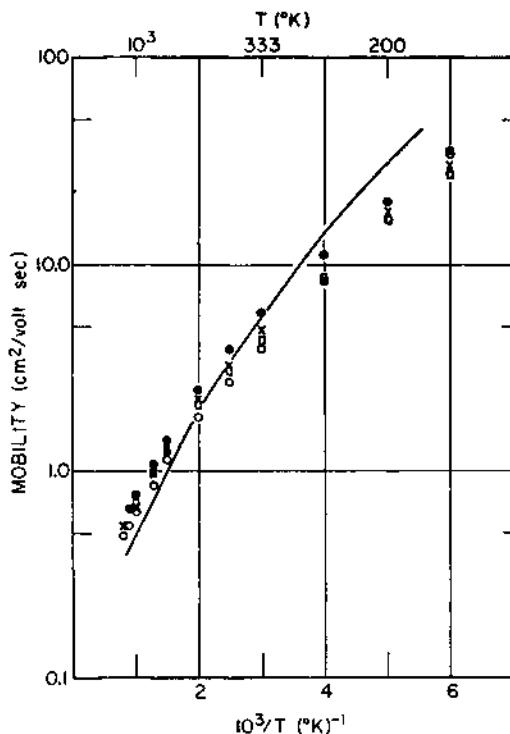


Fig. 29. Variation of mobility with temperature for $T > 200\text{K}$ in SrTiO_3 . After ref. 20.

of SrTiO_3 in terms of a hopping model. SrTiO_3 thus is an example of a compound that bridges the range between activated and nonactivated mobilities.

The system $\text{Pr}_2\text{O}_3 - \text{PrO}_2$ is representative of a class of oxides that is characterized by an activated mobility at all temperatures. Its electrical properties have been interpreted in terms of the very elementary model described in Section X^{23,24}. Because the hopping model is employed here which relies only on correlations between nearest neighbors one may, in lowest approximation, ignore the formation of the large set of intermediate ordered phases $\text{Pr}_n\text{O}_{2n-2}$ ($4 \leq n \leq 12$), all of which involve a relatively small shift in the positions of the cations. One may therefore adduce the following qualitative arguments: In order to produce slightly nonstoichiometric PrO_{2-y} it is necessary to remove some O° from the lattice. However, it is O rather than O° that appears as half of an oxygen molecule in the gas phase. The electrons left behind in the oxygen extraction process are distributed among the

cations, a corresponding number of which are converted from the Pr^{IV} state appropriate to PrO_2 to the Pr^{III} state. These extra electrons can then be transferred in an activated process from a Pr^{III} to a neighboring Pr^{IV} ion, thus converting the near insulating PrO_2 into an n-type hopping material, PrO_{2-y} . Consider next the slightly nonstoichiometric $\text{PrO}_{1.5+y}$ which is produced by incorporating oxygen from the gas phase as O²⁻ ions in the lattice. The requisite electrons are transferred to the oxygens at the expense of Pr by converting the necessary number of trivalent cations into quadrivalent cations. Thus in $\text{PrO}_{1.5+y}$ there is a small number of Pr^{IV} units of relative positive charge in a 'sea' of Pr^{III} units. The material should now be a p-type hopping material. One would expect the conductivity to be a maximum in the middle of the stoichiometry range ($x = 1.75$ in PrO_x) since at that point the number of carriers as well as the number of vacant positions is optimized. The passage in composition from $x < 1.75$ to $x > 1.75$ should also be accompanied by a transition from n to p-type properties. These various predictions are reasonably well satisfied by the data of Fig. 30 for the resistance of compressed

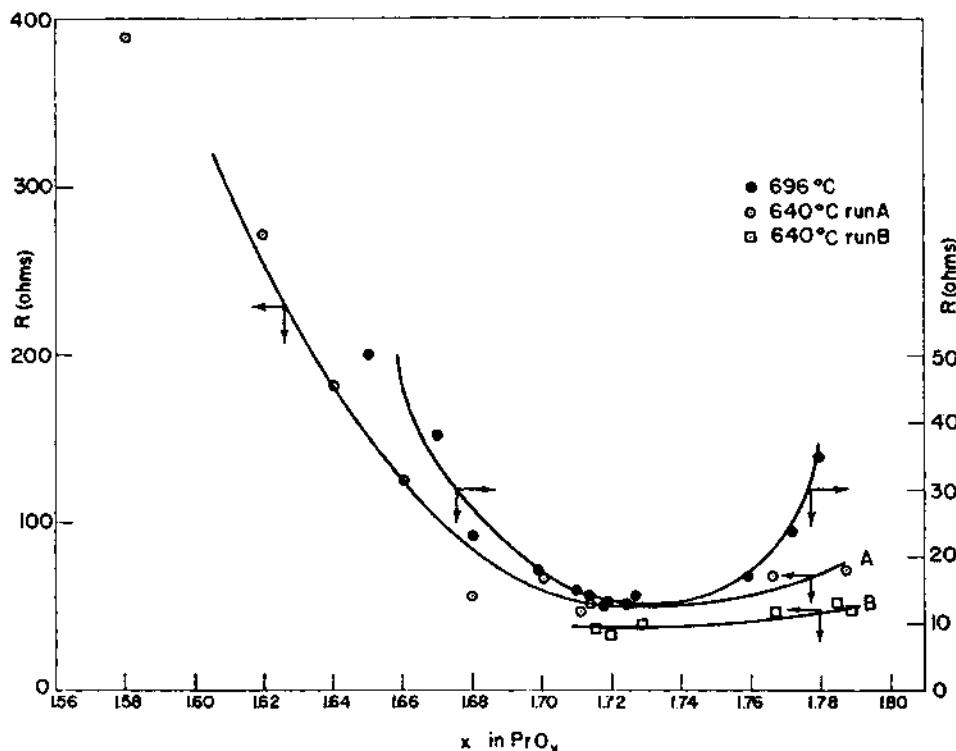


Fig. 30. Resistivity of PrO_x as a function of x for compressed, sintered pellets. After ref. 24.

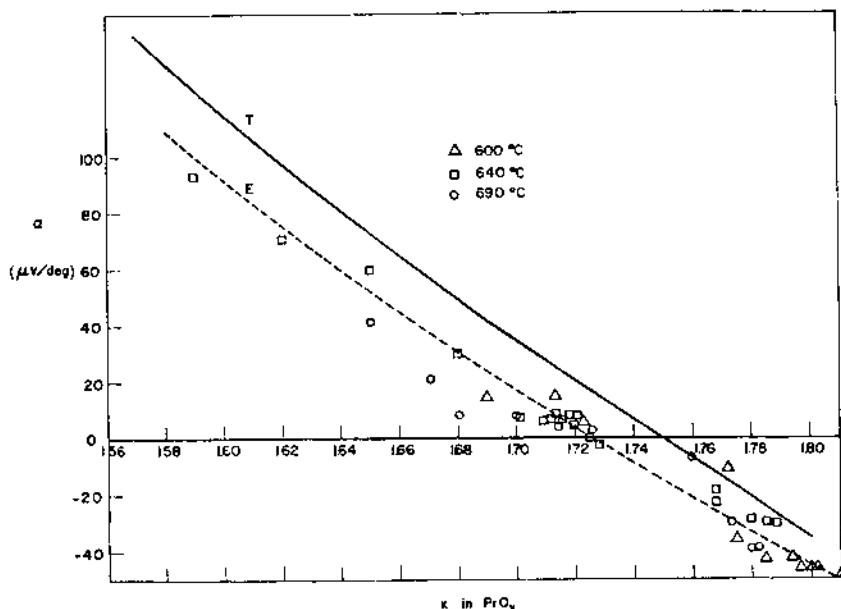


Fig. 31. Seebeck coefficient of PrO_x as a function of x for compressed, sintered pellets. After ref. 24.

pellets of PrO_x . The data, taken at constant temperature, conform to the theoretical prediction as required by Eqn. (139). Moreover, it is found for single crystal specimens of $\text{PrO}_{1.515}$ that α increases exponentially with T as specified by Eqn. (139). The Seebeck coefficient varies with composition as shown by the dotted line in Fig. 31; the solid curve is that portion of Eqn. (144) which is due to the configurational effects: $\alpha \sim -(k/e) \ln [\theta/(1-\theta)]$. The constant difference between the theoretical and experimental curves is a measure of the thermal entropy term S_T^* in Eqn. (144). Moreover, α is found to be very nearly independent of temperature, as is typical of this class of materials.

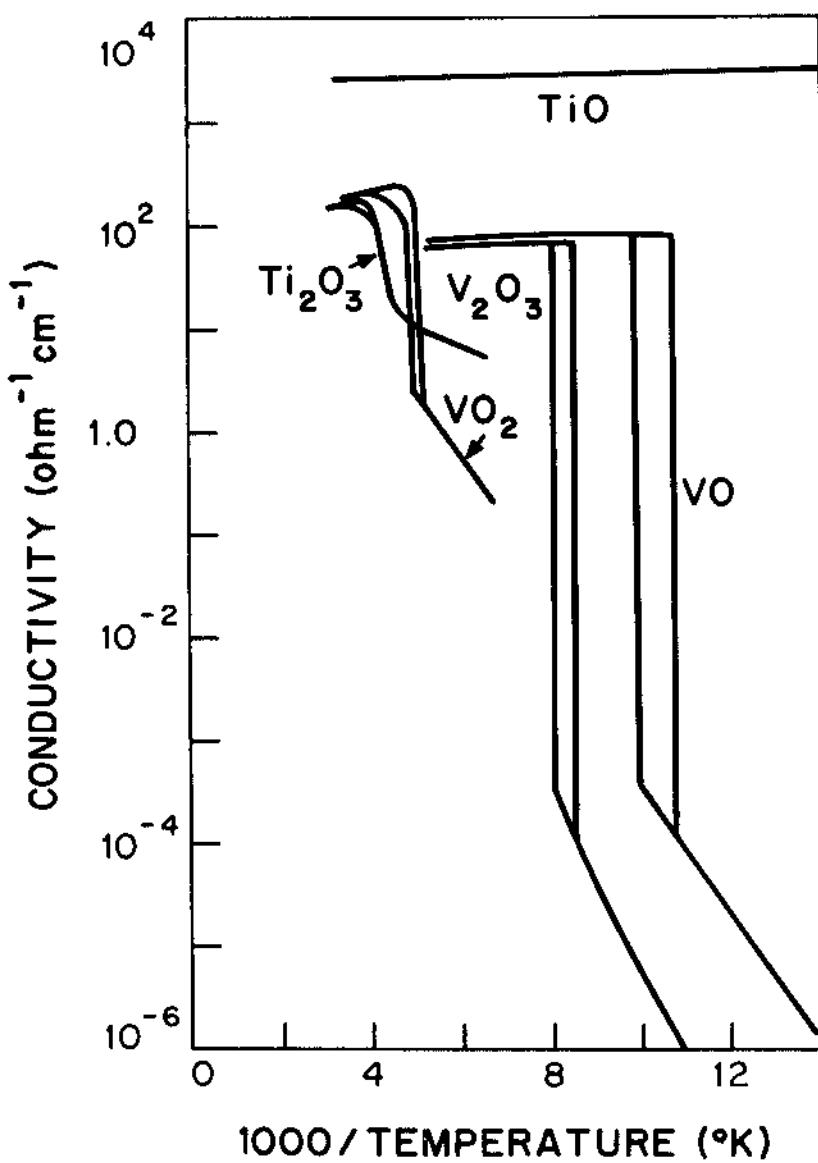
A considerable variety of other compounds is thought to display similar characteristics; nonstoichiometric FeO , ZnO , and UO_2 among them. The reader must be warned, however, that there is considerable controversy in some cases. NiO is an excellent example; there is tremendous uncertainty as to whether the mobility is activated or not. The definitive experiments on this material are yet to be made, since impurities in the 20-50 ppm range have a very appreciable effect on the electrical properties of this material²⁵. Similar uncertainties exist in the case of TiO_2 , where different investigators have invoked either an overlap band model or the small polaron model to explain the data. In this case the failure of workers to specify the exact chemical composition and physical characteristics

of the materials with which they are working obscures the resolution of the problem.

Oxides such as MgO, BeO, ZrO₂, or ThO₂ which remain nearly insulating up to quite high temperatures are not of interest in this survey and will not be discussed further.

There remains another category of oxides of relevance, namely those which undergo a transition in electrical characteristics. Two examples in this category are shown in Fig. 32. These results, reported by Morin²⁶, are by now outdated, since it has been shown repeatedly now that VO does not in fact suffer a semiconductor-metal transition. The original work seems to have been done on a material closer to V₂O₃ in composition.

There is still considerable doubt concerning the precise mechanism responsible for the observed changes in V₂O₃ and VO₂. Both transitions are clearly linked to the changes in crystal structure that occur simultaneously. The situation is not clearcut, however, since the structural rearrangement may itself be determined by the electronic properties of the material. Thus it may be argued (a) that distortions occur at the critical temperature T₀ in such a manner as to open up a gap at the Fermi level. This could lower all occupied electron states in energy while raising the unoccupied ones. If the band is initially half filled and then split in two as a concomitant to the crystalline distortion, there will be transition from metallic characteristics corresponding to the more symmetric phase (rhombohedral in V₂O₃) to semiconducting or insulating properties for the less symmetric phase (monoclinic in V₂O₃). At least six other mechanisms have been proposed to account for the transition in V₂O₃; these invoke (b) the onset of antiferromagnetism, whose existence has recently been verified by polarized neutron scattering experiments²⁷. Here the potential of the electrons differs for the spin-up and spin-down alignments of V ions of the antiferromagnetic phase; hence, there is a doubling of the unit cell in the magnetic superlattice, which is accompanied by a splitting of the band in two portions corresponding to the α and β spin alignments. (c) A combination of the two preceding mechanisms has also been invoked^{28,29}. The band gap opening up under the joint action of these two effects is then sufficiently large (~ 0.3 ev) to cause a corresponding drop in carrier density sufficient to account for the increase in resistivity. This would mean that the mobility remains nonactivated and essentially unchanged. On the other hand, if only mechanisms (a) or (b) were operative then a band gap closer to ~ 0.1 ev would be anticipated, which could correspond to a drop in carrier density of only two orders of magnitude. In that event, a change in mobility of five orders of magnitude is required to account for the observations. In these circumstances the mobility in the



(after Morin)

Fig. 32. Variation of resistivity with $1/T$ for selected oxides.
After ref. 26. Note that according to later work VO
does not suffer a metal-semiconductor transition.

monoclinic phase would be so low as to correspond to an activated process. (d) The possibility that band overlap may occur in V_2O_3 has recently been suggested, based on Hall and Seebeck coefficient measurements³⁰⁻³³. In this event, the transition could be thought of as being driven by the Coulomb interaction between the holes and electrons that are simultaneously present in the overlapping bands. This would result in a distortion that lifts the band overlap. The insulating phase resulting from this mechanism is known as an 'excitonic insulator'; for details the reader is referred to work by McWhan and Rice³³ who have shown in the course of their studies that the transition can be entirely suppressed by application of hydrostatic pressures in excess of 26 k bars. Resistivity studies of the metallic phase under compression down to liquid helium temperatures also strongly indicate the presence of overlapping bands. (e) It has been suggested that V_2O_3 may be an example of a 'Mott insulator'³⁴. This would imply that small variations in interatomic distances in the lattice cause sufficient alterations in band overlap to change the system from collective to localized behavior. This is not an unreasonable scheme to invoke here since the vanadium oxides tend to be at the very edge of the oxides with sufficient orbital overlap to render them metallic. (f) The effect of electron correlations have also been deemed important, since the bands in V_2O_3 are relatively narrow³⁵. Correlation theory predicts that if the band width lies in a certain narrow range and is further reduced (e.g., by temperature changes) then a splitting into two subbands takes place; the remainder of the argument proceeds as in cases (a) and (b). Finally, (g) models have been proposed in which localized states separate out from the conduction band, but the mechanism for this process remains to be elucidated.

In surveying these various models one gets the impression that it should be possible to devise a single, comprehensive theory, in which all the above possibilities arise as special cases. To sort out the various alternatives it would be helpful to obtain further data that show (i) whether overlap does or does not occur and (ii) whether the motion of carriers in the low temperature range is activated or not.

The case of VO_2 is similar to that discussed above except that this oxide does not exhibit magnetic order.

Ti_2O_3 differs from the cases discussed earlier in that this oxide undergoes a transition over a rather wide temperature range of 200 degrees. Contrary to early reports, it exhibits no magnetic order³⁶, neither does it undergo a crystallographic transition. On the basis of magnetoresistance and resistivity measurements, coupled with the fact that there is a distension of the unit cell (though no symmetry change) precisely in the temperature range where

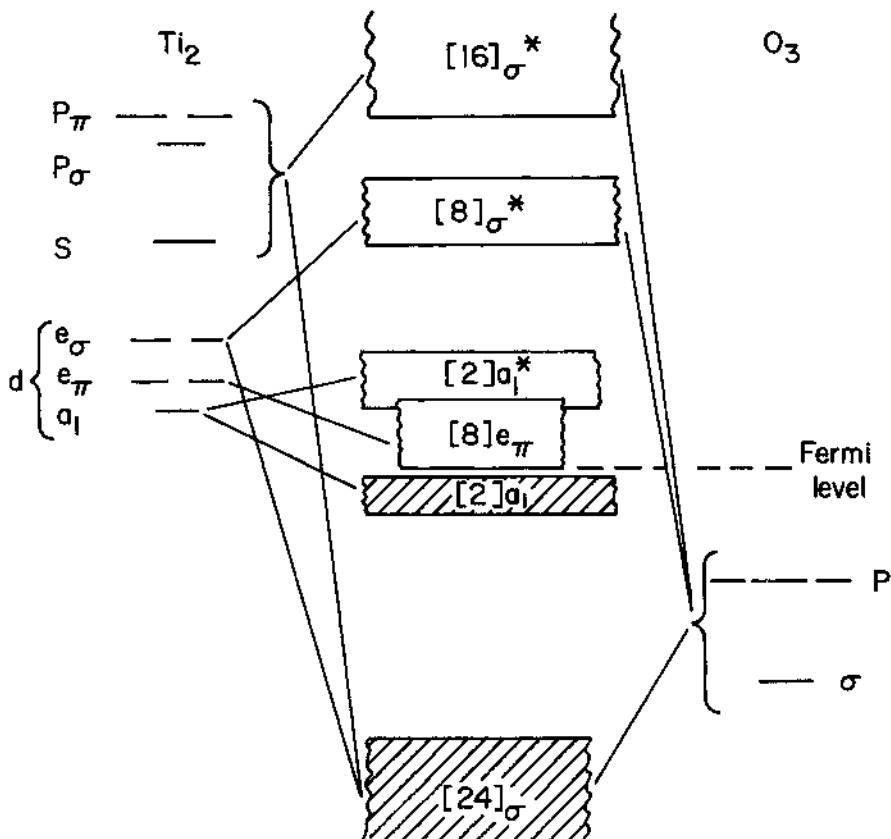


Fig. 33. Band structure for Ti_2O_3 . After ref. 37.

the transition occurs, a new model has been developed:³⁷⁻⁴⁰. This model is based on the band structure shown in Fig. 33 which is consistent with all known experimental facts. At low temperatures the completely filled band is separated by a gap of ~ 0.06 ev from the next higher one; the pure material is thus an intrinsic semiconductor in the liquid helium temperature range. With rising temperature the distension of the unit cell is in such a direction as to widen the bands and to eliminate the gap, thus gradually converting the compound to an overlap material.

This extremely cursory survey is designed to illustrate the wide variety of electrical properties which are encountered in solids and to show how studies in this area provide interesting information concerning the band structure of these materials.

The author is thankful to Dr. G.V. Subba Rao, Mr. S. Ramdas and Mr. S. Ramasesha for taking the lecture notes on which the present chapter is based.

REFERENCES

1. J.M. Ziman, "Electrons and Phonons: The Theory of Transport Phenomena in Solids" (Oxford Press, London 1960).
2. A.H. Wilson, "The Theory of Metals" (Cambridge Press, London 1958).
3. T.C. Harman and J.M. Honig, "Thermoelectric and Thermomagnetic Effects and Applications" (McGraw-Hill, New York, 1967).
4. E.M. Conwell, Proc. I.R.E. 40, 1327 (1952).
5. T. Holstein, Ann. Phys. (N.Y.) 8, 343 (1959).
6. A.P. Schmid, J. Appl. Phys. 40, 4128 (1969).
7. S.M. Marcus and S.R. Butler, Phys. Lett. 26A, 518 (1968).
8. R.W. Vest, Personal Communication.
9. J.M. Honig and A.J. Strauss, unpublished research.
10. J. Feinleib, W.J. Scouler, and A. Ferretti, Phys. Rev. 165, 765 (1968).
11. J.B. Goodenough, Bull. Soc. Chim. France 4, 1200 (1965).
12. J.B. Goodenough, Czech. J. Phys. B17, 304 (1967).
13. J.M. Honig, J.O. Dimmock and W.H. Kleiner, J. Chem. Phys. 50, 5232 (1969).
14. L.F. Mattheiss, Phys. Rev. 181, 987 (1969).
15. M.D. Banus and T.B. Reed, Proc. Conf. Chemistry of Extended Defects in Nonmetallic Solids, (L. Eyring and M. O'Keefe eds.) Scottsdale, Ariz. 1969. (North-Holland, Amsterdam 1970), p.488.
16. J.M. Honig and W.E. Wahnsiedler, unpublished research.
17. C.N.R. Rao, W.E. Wahnsiedler, and J.M. Honig, J. Solid State Chem., 1970, (in press).

18. D.S. Chapin, J.A. Kafalas and J.M. Honig, *J. Phys. Chem.* 69, 1402 (1965).
19. D.S. Rodbell, J.M. Lommel, and R.C. De Vries, *J. Phys. Soc. Japan* 21, 2430 (1966).
20. H.P.R. Frederikse and W.R. Hosler, *Phys. Rev.* 161, 822 (1967).
21. S.H. Wemple, M. Di Domenico, Jr., and A. Jayaraman, *Phys. Rev.* 180, 547 (1969).
22. L.C. Walters and R.E. Grace, *J. Phys. Chem. Solids* 28, 239 (1967).
23. J.M. Honig, *J. Chem. Ed.* 43, 76 (1966).
24. J.M. Honig, A.A. Cell and J.G. Cornwell, Proc. 3rd Rare Earth Conference (K. Vorres, Ed.) Clearwater, Fla., p. 557 (1963); also see, C.N.R. Rao and coworkers, *Trans. Farad. Soc.* 63, 1295 (1967) and *J. Solid State Chem.*, 1970 (in press).
25. R.W. Vest, personal communication.
26. F.J. Morin, *Phys. Rev. Lett.* 3, 34 (1959).
27. R.M. Moon, to be published.
28. D. Adler, J. Feinleib, H. Brooks, and W. Paul, *Phys. Rev.* 155, 851 (1967).
29. J. Feinleib and W. Paul, *Phys. Rev.* 155, 841 (1967).
30. N.F. Mott, *Phil. Mag.* 20, 1 (1969).
31. I.G. Austin and C.E. Turner, *Phil. Mag.* 19, 939 (1969).
32. V.P. Zhuze, A.A. Andreev, and A.I. Shelykh, *Soviet Physics - Solid State* 10, 2914 (1969).
33. D.B. McWhan and T.M. Rice, *Phys. Rev. Lett.* 22, 887 (1969).
34. D.B. McWhan and T.M. Rice, *Phys. Rev. Lett.* 23, 1384 (1969).
35. J.B. Goodenough, *J. Appl. Phys.* 39, 403 (1968).
36. R.M. Moon, T. Riste, W.C. Koehler, and S.C. Abrahams, *J. Appl. Phys.* 40, 1445 (1969).

37. L.L. Van Zandt, J.M. Honig, and J.B. Goodenough, *J. Appl. Phys.* 39, 594 (1968).
38. J.M. Honig and T.B. Reed, *Phys. Rev.* 174, 1020 (1968).
39. J.M. Honig, L.L. Van Zandt, T.B. Reed, and J. Sohn, *Phys. Rev.* 182, 863 (1969).
40. C.N.R. Rao, R.E. Loehman, and J.M. Honig, *Phys. Lett.* 27A, 271 (1968).

SPECTROSCOPY OF METALS AND SEMICONDUCTORS

R. Sharan

Department of Electrical Engineering

Indian Institute of Technology, Kanpur, India

I. INTRODUCTION

Spectroscopy is the science of study of the interaction of electromagnetic radiation (or matter waves) with solids, liquids or gases. In the present article, we would be interested in studying the interaction of a small range of electromagnetic waves (from the near infrared, through visible, to vacuum ultraviolet) with single-crystalline metals and semiconductors. The semiconducting materials, which are considered here, are important in the fabrication of many modern electronic devices.

Along with silicon and germanium, which are well-known for their uses in fabrication of diodes and transistors, some other Group III-V compound semiconductors are also becoming technologically important. These include the nine stoichiometric compounds formed by the combination of the Group III A elements, Al, In and Ga, with the Group V A elements, P, As and Sb. Particular mention can be made of the use of GaAs for the fabrication of light-emitting diodes and semiconductor lasers and of InSb for the detection of infrared radiation. It is obvious that a proper understanding of the optical properties of these semiconducting materials would facilitate the design of better opto-electronic devices.

However, even without keeping the device potentialities in mind, the optical properties of metals and semiconductors are important in their own right. These have been extensively investigated, both theoretically and experimentally, and several review papers on the subject have been published. A good review of the optical properties of solids has been given by Stern¹, and a very

readable elementary treatment on optical properties of metals has been published by Ehrenreich². With regard to the semiconducting materials there is a wealth of information available in the book on optical properties of III-V compounds³.

While studying the optical properties of solids one needs to consider the macroscopic as well as the microscopic picture. The former is adequately described by the classical electrodynamic theory of Maxwell. We have briefly presented the macroscopic theory in the following section along with the experimental methods. For the microscopic picture an idea of the band theory of solids is necessary and this has been treated in Section III, which leads directly to the discussion of the optical properties of metals and semiconductors in Sections IV and V, respectively.

II. MACROSCOPIC THEORY AND EXPERIMENTAL METHODS

From the macroscopic point of view, the interaction of electromagnetic radiation with matter is described by Maxwell's equations which are given below:

$$\nabla \cdot \vec{E} = \rho/\epsilon \quad (1)$$

$$\nabla \cdot \vec{H} = 0 \quad (2)$$

$$\nabla \times \vec{E} = -\frac{\mu}{c} \frac{\partial \vec{H}}{\partial t} \quad (3)$$

$$\nabla \times \vec{H} = \frac{\epsilon}{c} \frac{\partial \vec{E}}{\partial t} + \frac{4\pi\sigma'}{c} \vec{E} \quad (4)$$

Here, \vec{E} and \vec{H} are the electric and magnetic vectors characterising the electromagnetic wave. The properties of the matter are introduced into these equations through the dielectric constant ϵ , the magnetic permeability μ and the electrical conductivity σ' . Since only nonmagnetic materials will be considered here μ can be taken to be unity and one has to evaluate or measure only two quantities ϵ and σ' . The remaining quantities ρ and c , in the above equations are the charge density and the velocity of light respectively.

The magnetic field may be eliminated from equation (3) by taking curl of both sides and using equation (4):

$$\nabla \times \nabla \times \vec{E} = -\frac{\epsilon}{c^2} \frac{\partial^2 \vec{E}}{\partial t^2} - \frac{4\pi\sigma'}{c^2} \frac{\partial \vec{E}}{\partial t}$$

Now $\nabla \times \nabla \times \vec{E} = \nabla(\nabla \cdot \vec{E}) - \nabla^2 \vec{E}$. For a conducting medium $\rho = 0$ and from equation (1) we get $\nabla \cdot \vec{E} = 0$. Therefore,

$$\nabla^2 \vec{E} = \frac{\epsilon}{c^2} \frac{\partial^2 \vec{E}}{\partial t^2} + \frac{4\pi\sigma}{c^2} \frac{\partial \vec{E}}{\partial t} \quad (5)$$

This is the well known wave equation for the electric vector and its solution represents the propagation of an electric field with dissipation in the medium. If the direction of propagation is taken to be along x-axis and frequency to be ω , then the solution of the wave equation⁴ is given as:

$$\vec{E}(t, x) = \vec{E}_0 \exp \left\{ i\omega \left(\frac{nx}{c} - t \right) \right\} \exp \left(- \frac{kax}{c} \right) \quad (6)$$

where n and k are the index of refraction and the extinction coefficient, respectively, of the medium. By substituting equation (6) in equation (5), we obtain:

$$N = (n + ik) = (\epsilon + i \frac{4\pi\sigma}{\omega})^{\frac{1}{2}} = (\epsilon_1 + i \epsilon_2)^{\frac{1}{2}} \quad (7)$$

where N is called the complex index of refraction. By equating the real and imaginary parts of equation (7), we get:

$$n^2 - k^2 = \epsilon_1 \text{ and } 2nk = \frac{4\pi\sigma}{\omega} = \epsilon_2 \quad (8)$$

Thus we have two equivalent sets of parameters — n , k and ϵ_1 , ϵ_2 to describe the properties of the medium. Sometimes another parameter, the absorption coefficient (α) is also used, where:

$$\alpha = \frac{4\pi k}{\lambda} \quad (9)$$

and λ is the wavelength of the incident radiation. One important thing to note is that $(n^2 - k^2)$ and $2nk$ are not independent, but are related by Kramers-Kronig dispersion relations⁴. These relations are:

$$n^2(\omega) - k^2(\omega) = \frac{2}{\pi} \int_0^\infty \frac{\omega' \{ 2n(\omega')k(\omega') \}}{\omega'^2 - \omega^2} d\omega' + \text{Const} \quad (10)$$

$$2n(\omega)k(\omega) = - \frac{2\omega}{\pi} \int_0^\infty \frac{n^2(\omega') - k^2(\omega')}{\omega'^2 - \omega^2} d\omega' \quad (11)$$

where only the principal part of the integral is taken. The advantage of the dispersion relations is that if either $(n^2 - k^2)$ or $2nk$ is known over the whole frequency range, the other can be theoretically calculated. This cuts down the number of experiments to be done.

Experimental Methods. There is a large number of experimental methods to determine n and k . However, the most convenient to

measure are the transmission coefficient (T) and the reflectivity (R). Neglecting interference effects, the transmission coefficient is determined by,

$$T = \frac{I_t}{I_0} = \frac{(1-R)^2}{\exp(\alpha d) - R^2 \exp(-\alpha d)} \quad (12)$$

where I_0 and I_t are the intensities of incident and transmitted radiation and d is the thickness of the sample. The reflectivity is given as:

$$R = \frac{(n-1)^2 + k^2}{(n+1)^2 + k^2} \quad (13)$$

and is very sensitive function of the surface conditions of the sample. To determine n and k both T and R should be measured over the required frequency range.

Generally, it is not possible to measure both transmission and reflection coefficients over a given frequency range. Transmission measurement is done for samples where $\alpha(\omega)d \ll 1$, otherwise the transmitted beam will be too weak to be detected. When $\alpha(\omega)d \gg 1$, then reflectance measurements give useful information.

For the measurement of transmission or reflection coefficients, the basic apparatus needed is a spectrophotometer, the details of which are given in many text-books⁵. The main components of a spectrophotometer are (i) the source of radiation, (ii) the monochromator, and (iii) the detector and associated electronics. Tungsten lamps and globar are generally used as sources for visible and near infrared radiation, whereas discharge lamps are used in ultraviolet region. The monochromatic waves are obtained with the help of prism or grating. Photomultipliers are used as detectors in case of ultraviolet and visible radiation. In the near infrared region a thermocouple or PbS cell gives better sensitivity.

A complication arises in the case of ultraviolet radiation of higher frequencies because air becomes opaque in this region. As a result, the measurements must be performed in vacuum for this energy range of photons. Such an apparatus has been described by Johnson⁶ and is in wide use.

III. MICROSCOPIC THEORY OF OPTICAL PROPERTIES

Classical Theory. A very qualitative idea about the optical

effects in solids can be obtained by considering the behaviour of an oscillating electric field $E(t)$, where we consider only a single Fourier component at frequency ω . The equation of motion of the electron is given by:

$$\ddot{x} + \gamma' \dot{x} + \omega_s^2 x = \frac{eE(\omega)}{m} e^{-i\omega t} \quad (14)$$

where $\gamma' \dot{x}$ is a damping term characterized by strength γ' , and $m\omega_s^2 x$ is the restoring force of a spring whose natural frequency is ω_s . Since equation (14) is a linear equation, we have $x(t) = x(\omega)e^{-i\omega t}$. When this value is substituted in Eqn. (14), it becomes an algebraic equation and yields the value of $x(\omega)$. Then the induced polarization in the material is given as:

$$P(\omega) = en x(\omega) = \frac{n e^2 E(\omega)/m}{\omega^2 + i\omega\gamma - \omega_s^2} \quad (15)$$

Here n is the number of electrons. Induced polarization is also defined (in c.g.s. units) as:

$$P(\omega) = \frac{\epsilon'(\omega) - 1}{4\pi} E(\omega) \quad (16)$$

Equating equations (15) and (16) one obtains

$$\epsilon'(\omega) = 1 - \frac{\omega_p^2}{\omega^2 + i\omega\gamma - \omega_s^2} = \epsilon_1(\omega) + i\epsilon_2(\omega) \quad (17)$$

$$\text{where } \omega_p = \left(\frac{4\pi ne^2}{m} \right)^{\frac{1}{2}} \quad (18)$$

The dielectric constant $\epsilon'(\omega)$ in equation (17) is a complex quantity and its real and imaginary parts should be compared with the square of equation (7). The behaviour of the real and imaginary parts of $\epsilon'(\omega)$ are shown in Fig. 1 for $\omega_s \neq 0$ and $\omega_s = 0$. The former is the case of an insulator and is generally known as Lorentz model. When $\omega_s = 0$, there is no restoring force and it is the case of free electrons in a metal and leads to Drude model. The plasma frequency (ω_p) defined by equation (18) is associated with the collective movement of electrons and as we shall show later, is a very important parameter. The above picture is quite simplified but predicts the general nature of the optical properties of metals and insulators. However, there are some optical properties which cannot be explained unless quantum mechanical theory is taken into account. One of the effects of this theory in solids is to give rise to a band structure. Optical properties are very sensitive to actual band structure and hence in the following section we would consider the band structure of metals and

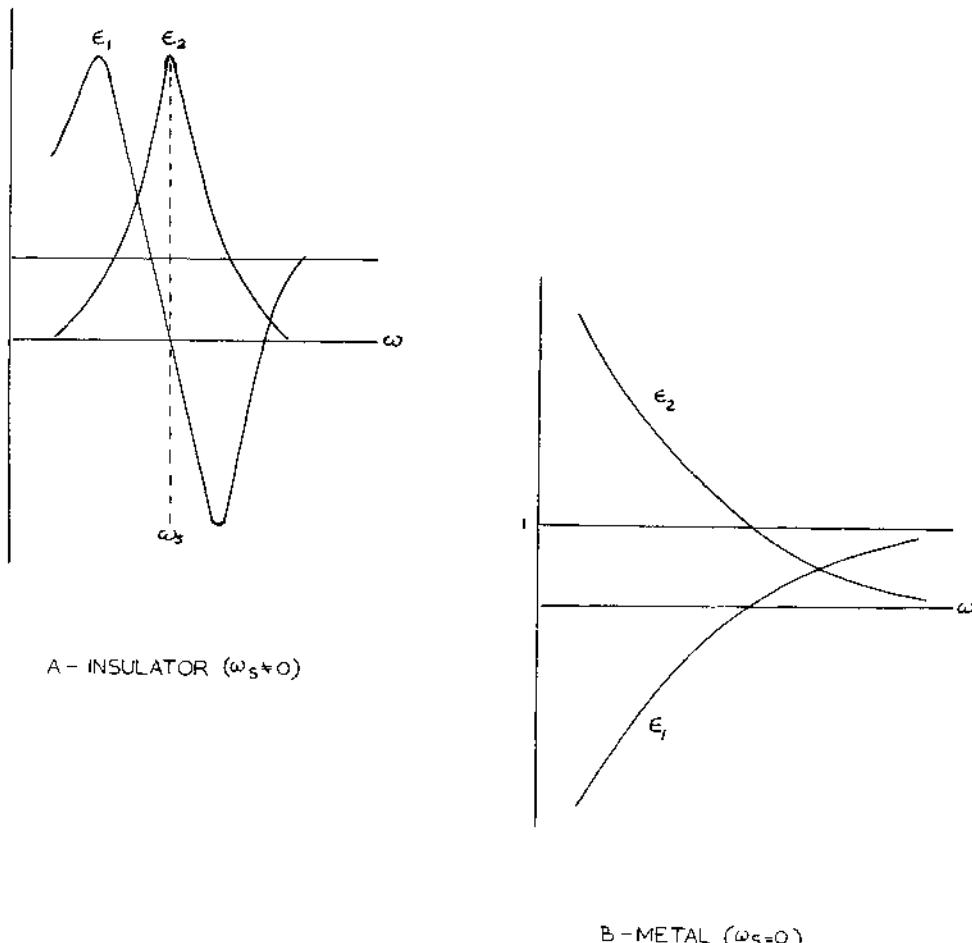


Fig. 1. Real and imaginary parts of dielectric constant from equation (17). From Ref. (2).

semiconductors.

Band Theory. Band structure is a quantum-mechanically calculated E-k diagram where E is the energy and \vec{k} is the momentum of the electron. A very simplified diagram of a portion of the band structure of a typical metal is shown in Fig. 2. The diagram is in two different k-directions and extends upto the boundary of the Brillouin zone. Two different bands labelled n and n' are shown and the Fermi level is indicated by dotted lines.

The band structure of a semiconductor is simplest when both the conduction and the valence bands are parabolic (or, equivalen-

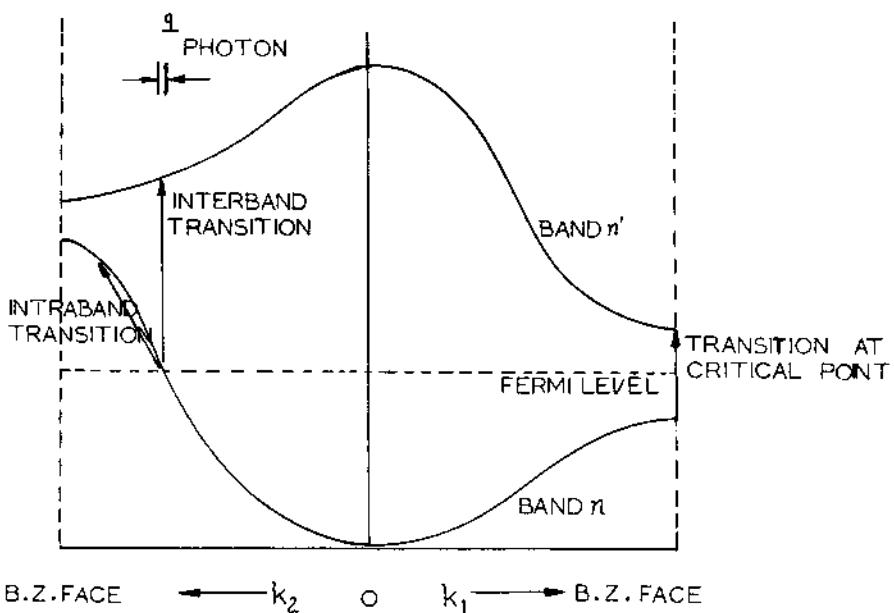


Fig. 2. Simple band structure of metals. From Ref. (2).

tly, spherical in k -space) and their extrema occur at $k = 0$. This is shown in Fig. 3a, and is mathematically represented by:

$$E_c = E_g + \frac{\pi^2}{2m_e} k^2 \quad (19)$$

$$E_v = - \frac{\pi^2}{2m_h} k^2 \quad (20)$$

where E_g is the energy band gap, m_e and m_h are the effective masses of the electrons and holes respectively and the reference level of the energy is at the top of the valence band. Since the extrema of both the conduction and valence bands occur at the same k -point, such materials are called "direct" gap semiconductors. Typical examples of this class are InSb, InAs, GaAs, etc.

In an actual case the valence band is six-fold degenerate but due to spin-orbit coupling splits into a four-fold degenerate band

at $k = 0$ and another two fold split-off band. This is shown in Fig. 3b. The degeneracy of the valence band is important when optical transitions to impurity levels are considered.

For many semiconductors, the minimum of the conduction band is not at $k = 0$, but along some symmetry direction in the Brillouin Zone (Fig. 3c). Such materials are known as "indirect" gap semiconductors and typical examples of this class are Si (minima in <100> direction), Ge (minima in <111> direction), GaP, etc. The mechanism of optical transitions in 'indirect' gap materials is more involved than the ones in direct gap materials.

Density of States. A quantity of interest in the calculation of the optical properties of metals and semiconductors is the optical density of states $D(E)$, which depends only on the band structure, i.e., the E-K diagram. For a general case $D(E)$ is given by⁷:

$$D(E) = \frac{2V}{(2\pi)^3} \iint \frac{d\vec{s}}{|\nabla_k(E_i - E_j)|} \quad (21)$$

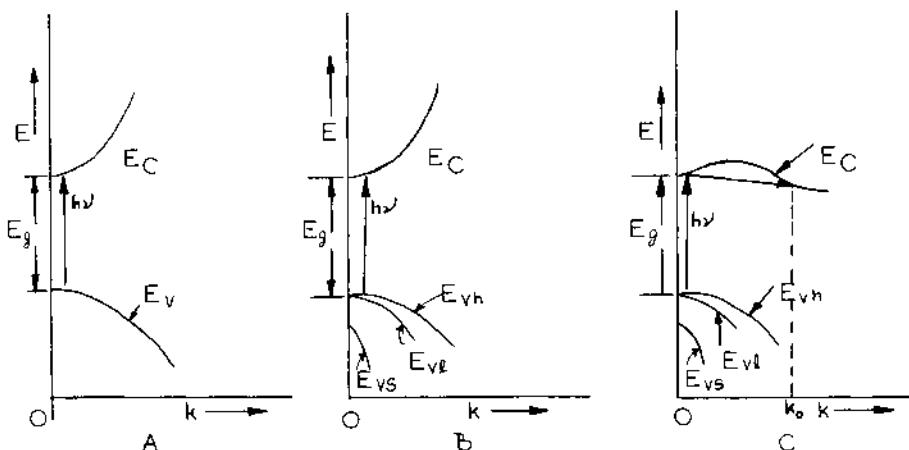


Fig. 3. Band structure of semiconductors. (A and B - direct energy gap; C - indirect energy gap).

where the factor 2 arises from spin orientations, V is volume of the specimen and $d\vec{s}$ is the element of area in K-space of the surface of constant energy. In most of the cases, it is difficult to analytically evaluate $D(E)$. Hence, for the present, we would note some of the qualitative features of $D(E)$ and evaluate it for the simple case of a parabolic, non-degenerate, direct-band-gap material.

$D(E)$ shows strong variations near some "critical points" (called van-Hove singular points) of the band diagram. The critical points occur when,

$$\nabla_k (E_i(k) - E_j(k)) = 0 \quad (22)$$

As evident from equation (21), $D(E)$ goes to infinity at these points. There are four types of van-Hove singularities⁷. We would only consider the simplest type, which occurs when:

$$\nabla_k E_i(k) = \nabla_k E_j(k) = 0 \quad (23)$$

i.e., both the gradients vanish separately. This is the case for the bands of Fig. 3a. From equations (19) and (20) we have

$$E = E_c - E_v = E_g + \frac{\hbar^2 k^2}{2m_r} \quad (24)$$

where

$$\frac{1}{m_r} = \frac{1}{m_e} + \frac{1}{m_h} \quad (25)$$

Substitution of equation (24) in equation (21) gives:

$$D(E) = \frac{8\sqrt{2}\pi}{h^3} m_r^{3/2} (E-E_g)^{1/2} \quad (26)$$

In case of complicated band structures, the expression for the density-of-states becomes quite involved and numerical calculations are required to evaluate it.

Mechanisms of Optical Transitions. There are three important mechanisms which give rise to the observed optical properties of solids. These are intraband and interband transitions and plasma oscillations. When the incident photons impart their energy to electrons, the electrons make a transition from an initial to a final energy state. Depending upon the position of the initial and final states one distinguishes between intraband and interband transitions.

Moreover, it should be noted that for a transition to take place both the conservation of energy and that of momentum have to be satisfied. According to the band theory if the system consists of only the interaction of photons and electrons, then the conservation laws require that:

$$E_i + h\nu = E_f \quad (27)$$

$$k_i + q_p = k_f \quad (28)$$

where i and f are the subscripts for initial and final states and $h\nu$ and q_p are the energy and momentum of the incident photon. Since q_p is much smaller than the dimensions of the Brillouin Zone, equation (28) reduces to $k_i \approx k_f$, i.e., the transitions are vertical.

For non-vertical transitions to be possible, the extra momentum must be provided by phonons or other crystal imperfections. Thus, the system would then consist of the interactions of the incident photon, the electron and the imperfection and the modified conservation laws would be:

$$E_i + h\nu \pm (E)_{\text{imperfection}} = E_f \quad (29)$$

$$k_i \pm (k)_{\text{imperfection}} = k_f \quad (30)$$

(i) Intraband Transitions. In this case both the initial and the final electronic states are located in the same band. Thus a vertical transition is not possible and an imperfection must interact for non-vertical transitions to take place. Such transitions are also known as free-carrier transitions and, for metals, are shown in Fig. 2. These are important in explaining the high reflectance of metals and can be adequately discussed within the framework of classical theory. It is experimentally observed that the reflectance of many metals in the infrared region can be simply predicted by the Drude model of Fig. 1B. In polyvalent metals like aluminium, which are nearly free-electron like, this qualitative behaviour extends far into ultraviolet region.

(ii) Interband Transitions. As the name implies, here, the initial and the final states are in different energy bands. Thus it is possible to get both vertical and non-vertical transitions depending upon the position of the extrema of the initial and final states. These are the most important transitions in a semiconductor since they give rise to the "fundamental" absorption. Recently it has been realized that the interband transitions are also important in explaining the behaviour near a dip in the reflectance spectra of noble and transition metals. These transitions are shown for metals and semiconductors in Fig. 2 and Fig. 3,

respectively.

(iii) Plasma Oscillations. At certain frequency ω_p , given by equation (18), the electrons in a solid act co-operatively due to Coulomb interaction among them and give rise to a resonance which is called plasma resonance. As evident from equation (18), ω_p depends only on the number of electrons per unit volume and their effective mass. In metals, ω_p corresponds to ultraviolet range of energy (10-20 eV), whereas in semiconductors it occurs in infrared region due to lower number of electrons per unit volume. However, since this number of electrons in semiconductors can be controlled by varying the doping, the plasma frequency can be varied within certain range.

IV. THE OPTICAL PROPERTIES OF METALS

Extensive work on the optical properties of metals has been done by Enrenreich and Philipp⁸ (EP). Similar experiments have been carried out on the semiconducting materials⁹ and the metal ReO_3 ¹⁰. We would briefly summarize the major features of these investigations.

The experimental results on the optical properties of metals fall into two categories: (a) those which can be explained by classical theory and (b) those which need quantum mechanical interpretation. In the following we would distinguish between these two cases.

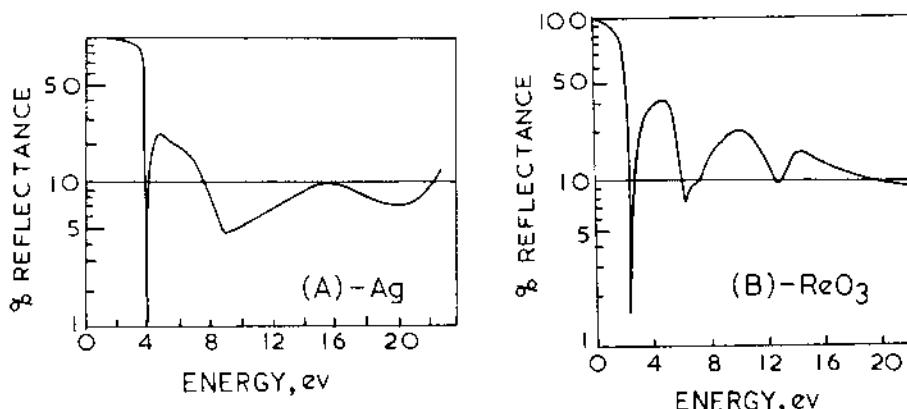


Fig. 4. Reflectance spectra of Ag and ReO_3 . Ref. (8, 10).

Let us consider the reflectance measurements in two metals - Ag and ReO_3 . These reflectance measurements have been done by EP⁸ and Feinleib *et al.*¹⁰ respectively between the photon energy range of 1 to 25 eV and are shown in Fig. 4. The experimental results show that at low frequencies there is nearly 100% reflectance which is characteristic of free-electron range and can be adequately explained by classical Drude model (Fig. 1B). According to this model the reflectance stays at 100% as long as ϵ_1 is negative. When ϵ_1 passes through zero at the plasma frequency, the reflectance falls off rapidly. Actually, in case of Ag and ReO_3 , there is a dramatically sharp dip near 4.0 and 2.30 eV, respectively, with the reflectance dropping from nearly 100% to less than 1% in a fraction of eV of photon energy. However, the reflectance rises again. This rise is due to the interband transitions (from d bands to Fermi level) and is a quantum mechanical effect. It is important to note that it is this transition that is responsible for the colour of noble metals. Similar behaviour, except for some minor deviations, is also observed in Cu and Au.

For analysis of the experimental data, which is the value of the reflectivity $R(\omega)$ measured, for normal incidence, over a wide frequency range, a quantity r is defined such that:

$$r = |r| e^{i\theta} = \frac{n - ik - 1}{n + ik - 1} \quad (31)$$

By separating the real and imaginary parts of r , we get,

$$R = |r|^2 \quad (32)$$

and $\tan \theta = \frac{2k}{n^2 + k^2 - 1} \quad (33)$

The phase angle $\theta(\omega)$ is related to $R(\omega)$ by Kramers-Kronig dispersion relation. Since $R(\omega)$ is measured over a very wide frequency range, $\theta(\omega)$ can be numerically computed by using this dispersion relation. Thus with $R(\omega)$ and $\theta(\omega)$ known, equations (32) and (33) can be simultaneously solved to obtain n and k at any required frequency ω_0 . With n and k known, ϵ_1 and ϵ_2 can also be determined from equation (8). The curves of $\epsilon_1(\omega)$ and $\epsilon_2(\omega)$, as deduced from the experimental results of Ag and ReO_3 , are shown in Fig. 5. These should be compared to the classical picture of a metal given in Fig. 1B. To do this one can calculate, say, $\epsilon_1(\omega)$ from the classical equation (17) and then subtract it from the experimentally determined curve of $\epsilon_1(\omega)$ to determine its deviation from free-electron theory. This has been done by EP and the case of silver is shown in Fig. 6A. The deviation from the free-electron calculation is shown as $\delta\epsilon_1^{(b)}$ and

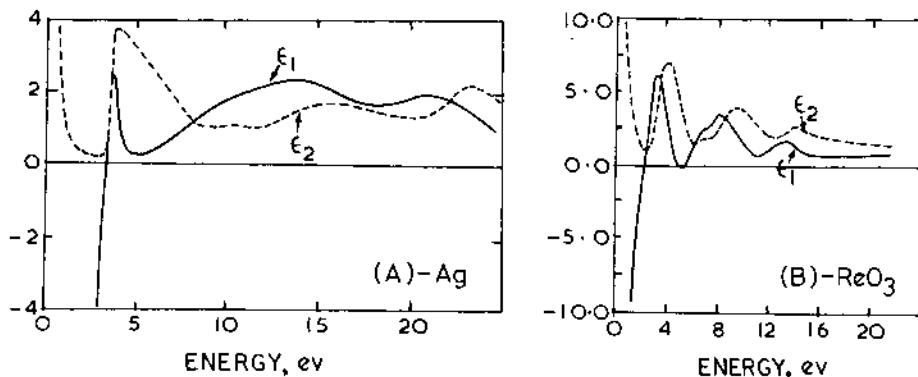


Fig. 5. Experimental spectral dependence of the real and imaginary parts of the dielectric constant. (A-Ag; B-ReO₃). From Ref. (8, 10).

this appears because of interband transitions. Let us consider the effect of these transitions on plasma oscillation. From Fig. 6A one sees that in the absence of interband transition the zero of $\epsilon_1(\omega)$ would be at 9.2 eV (for Ag). This energy should correspond to the plasma frequency, ω_p , given by equation (18). In the actual case, however, the interband transitions cause the frequency corresponding to $\epsilon_1 = 0$ to be reduced to a value slightly below ω_p (~ 3.9 eV), thus lowering the frequency of this hybrid plasma resonance. This effect is also observed in ReO₃.

In case of Cu also the curve of $\epsilon_1(\omega)$, as derived from experimental results, can be separated into the free-electron, $\epsilon_{1(f)}$ and interband transitions. This is shown in Fig. 6B. Here one finds that the interband transitions (denoted by $\delta\epsilon_1(b)$) peak up near 2.1 eV and the strength of these transitions is stronger than those of Ag at 3.9 eV. In spite of the strength of these transitions at 2.1 eV, the total ϵ_1 in Cu does not cross zero. This is because at the frequency (~ 2.1 eV) where interband transitions are strong, the free-electron component is still quite negative. Hence a plasma resonance is not observed in Cu.

Similar analysis can also be done for $\epsilon_2(\omega)$, i.e., one separates the contribution of free electron theory, calculated from equation (17), and that of interband transitions. Simultaneous consideration of $\epsilon_1(\omega)$ and $\epsilon_2(\omega)$ would give a better insight

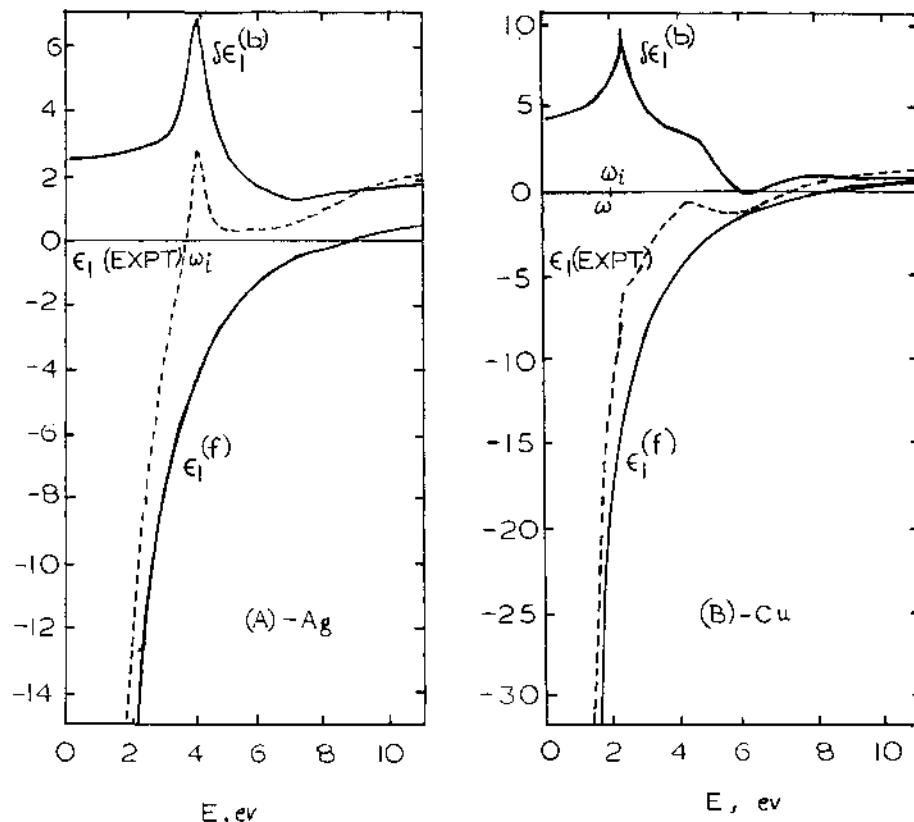


Fig. 6. Decomposition of the experimental values of ϵ_1 , into free ($\epsilon_1^{(f)}$) and interband ($\delta\epsilon_1^{(b)}(\omega)$) contributions. The threshold energy for interband transitions is indicated by ω_i . From Ref.(8).

in the physical mechanisms of transition.

Recently quite intensive efforts have been made to understand the roles of interband transitions and plasma resonance in transition metals, metal-oxides and semiconductors. Such work on Gr. IV and Gr. III-V semiconductors has been reviewed by Philipp and Ehrenreich⁹. Most recently, Rao et. al.¹¹ have investigated the plasma resonance in TiO, VO and NbO. They have found that all these oxides show reflectance minima due to plasma edges. The energy corresponding to this minimum in TiO varies with stoichiometry (in the homogeneity range) and this phenomenon has been related to the transport properties of these oxides.

V. THE OPTICAL PROPERTIES OF SEMICONDUCTORS

In case of semiconductors, it is observed that in the region where photon energies are approximately equal to the energy band gap, there is a rapid increase in the absorption coefficient. This is known as fundamental absorption.

Fundamental Absorption in Direct Gap Semiconductors. If the conduction band is completely empty and the valence band is completely full, then the expression for the absorption coefficient is¹²,

$$\alpha(h\nu) = \left(\frac{e\hbar^2}{2\pi n} \right) \frac{1}{h\nu} \frac{W_{ij}}{|A_0|^2} \quad (34)$$

where W_{ij} is the transition probability and A_0 is the amplitude of the vector potential of the incident radiation field. The transition probability W_{ij} denotes the number of transitions, per unit volume per unit time, from the initial state i to the final state j , and is given by the Fermi-golden-rule¹³:

$$W_{ij} = \frac{2\pi}{h} |H_{ij}|^2 D(E) \quad (35)$$

Here H_{ij} is the optical dipole matrix element. Generally, H_{ij} varies very slowly with K-vector. Hence, most of the structure of $\alpha(h\nu)$ is due to the structure of the density of states $D(E)$.

For a semiconductor which has a simple band structure, i.e., equations (19) and (20) represent the bands adequately, it is expected that the absorption coefficient $\alpha(h\nu)$ would vary as $(h\nu - E_g)^2/h\nu$. This type of variation is shown in Fig. 7, where the experimental points for InSb at 78°K are also shown. Fig. 7 shows that there is a good agreement between the experimental and theoretical results. The value of photon energy where the sharp absorption arises corresponds to the band gap (E_g) of the semiconductor. This is a very practical and direct means of measuring the temperature dependence of the band gap which is otherwise a difficult quantity to measure. Many other direct band gap semiconductors like GaAs, InAs, etc. show an absorption behaviour similar to that of InSb.

Indirect Gap Semiconductors. In this case, the minimum of the conduction band and the maximum of the valence band are at different K-points. Hence, vertical transitions cannot take place. The transitions, actually, take place with the emission or absorption of an acoustical phonon as given by equations (29) and (30). One example of such transitions is shown in Fig. 3C. Here a vertical transition takes place from the valence band to a virtual state in the conduction band and then the electron is scattered in the

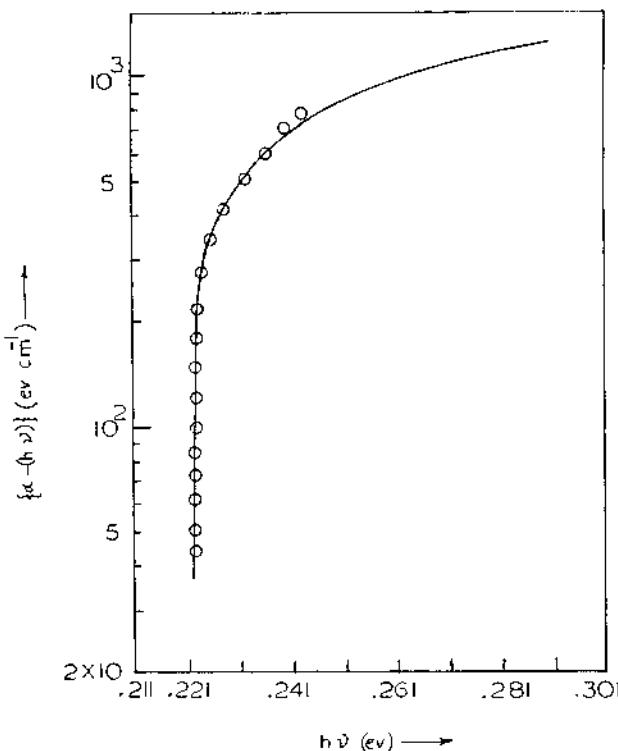


Fig. 7. Fundamental absorption in direct energy gap semiconductor. (InSb at 78°K).

conduction band with the emission of a phonon. Alternative transitions with the absorption of phonons are also possible. The net absorption is a combination of all such processes. For theoretical analysis of this photon-electron-phonon interaction one needs a second order quantum-mechanical perturbation theory and we would not go in details here.

The effect of indirect transitions on the experimental curve is that the fundamental absorption rises more slowly than was the case in Fig. 7. This is observed for Si, Ge, GaP, etc., which are thus taken to be indirect gap materials.

Impurity Absorption. It is well known that in Si and Ge the elements from Group IIIA and VA constitute the simple shallow acceptors and donors, respectively. Similarly III-V semiconductors can also be suitably doped to make them p- or n-type. An example is the doping of InSb with Cd to obtain p-type material.

Theoretical analysis of such impurities is done by considering

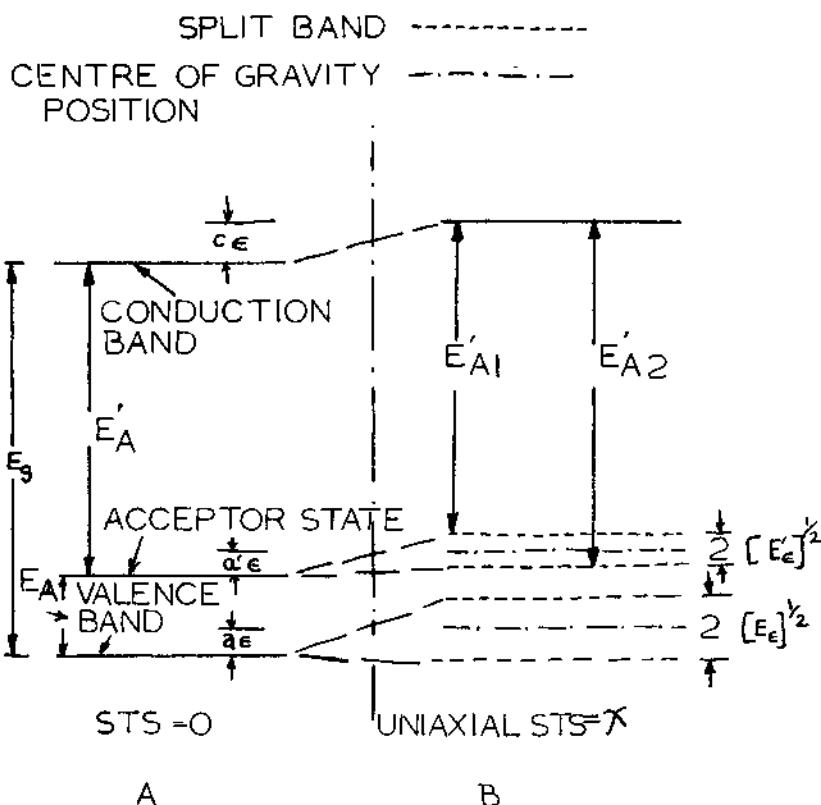


Fig. 8. Piezo-splitting in InSb at $K = 0$, including the splitting of shallow acceptor state.

them as a hydrogen-like atom with an electronic effective-mass different from the free-electron mass and with the dielectric constant of free space replaced by that of the parent crystal. This analogy to the hydrogen atom implies spectra which are similar to that of atomic hydrogen and, therefore, suggests a number of excited states between the ground state of the impurity and the nearest band edge. These states are given by:

$$E_n = -\frac{1}{2} \cdot \frac{e^2}{2\pi m^*} \cdot \frac{n}{m^* \epsilon} \quad n = 1, 2, \dots \quad (36)$$

where m^* is the effective-mass of the carrier and ϵ is the low-frequency dielectric constant of the medium. The ground state ($n=1$) of an acceptor for a direct band gap material is shown in Fig. 8A.

The optical transitions between the ground and the excited states of the hydrogen-like model take place in the far infrared

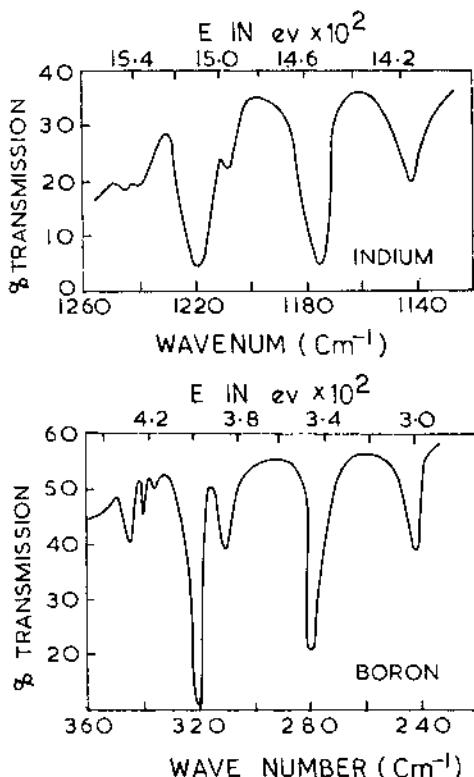


Fig. 9. Hydrogen-like spectra of shallow impurities in semiconductors at 4.2°K. (Indium and Boron in Si). From Ref. (14).

and can be observed only at very low temperatures ($\leq 4^{\circ}\text{K}$). Such spectra due to indium and boron impurities in silicon¹⁴ are shown in Fig. 9. We would not concern ourselves with these transitions here. Instead, we would consider the transitions from the acceptor state to the conduction band (or from the valence band to the donor state) which occur just before the fundamental absorption starts. These transitions have the same spectral dependence as the fundamental absorption but appear as very weak shoulders which are very difficult to detect. For Cd-doped InSb and some III-V compounds, these have been detected at 10°K with a high resolution spectrometer¹⁵. We would further study these transitions in the next section.

Perturbation Methods. In order to gain a better insight into the band structure of semiconductors, their optical properties are studied with various perturbations put on the crystal. Common examples of such perturbations are electric-field (Franz-Keldysh effect), magnetic field, variation of temperature, mechanical stress (hydro-

static or uniaxial) or a combination of these.

Also, while studying the optical spectra, with or without perturbation, one is confronted with the problem of detecting a very faint absorption line or a small shoulder. The detection of these requires a very high resolution and stability of the spectrophotometer which are difficult to achieve. To circumvent this problem, one suggestion has been to directly measure $(dI/d\lambda)$ versus λ instead of the conventional measurement of I versus λ , where λ is the wavelength of the radiation. The spectrophotometer can be easily modified to achieve this¹⁶ and this method is generally known as "differential-wavelength-modulation".

Using this differential-wavelength-modulation and by applying uniaxial stress to differently oriented crystals of Cd-doped p-InSb at 99°K, we have studied the transitions from the acceptor state to the conduction band¹⁷. The effect of the uniaxial component of stress is to split the degenerate bands and that of hydrostatic component is to shift the energy of different bands. Thus with the application of uniaxial stress, the valence band and the ground state of the acceptor are each split in two two-fold degenerate bands as shown in Fig. 8B. Here ϵ is the strain and a , c and a' are the deformation-potential constants of the valence and conduction bands and the acceptor state respectively. These constants define the shift of energy bands per unit deformation and are very important in calculating the transport properties of semiconductors. E_ϵ and $E_{\epsilon'}$ are related to the energy splitting of the valence band and acceptor state, respectively, with uniaxial stress.

A representative optical spectrum with uniaxial stress applied parallel to [111] crystal direction is shown in Fig. 10 for Cd-doped p-InSb at 99°K. The differential wavelength modulation has given $(dI/d\lambda)$ directly. Here, with no stress there is only one dip α_1 , which is associated with an electron-optical-phonon transition¹⁸. With a stress of 382.0 Kg/cm² there are three dips, where α_2 and α_3 correspond to energies E_{A_1}' and E_{A_2}' in Fig. 8B. At the applied stress of 1526 Kg/cm², α_2 and α_3 separate farther in energy. Fig. 11 shows a plot of the energy positions of the dips α_2 and α_3 as a function of stress. When the plot is extrapolated, one obtains the ionization energy of Cd impurity in p-InSb as 9.7 meV, which is in very good agreement with other measurements. A detailed analysis of spectra like Fig. 10, but with uniaxial stress applied in different crystal directions, yields the values of different uniaxial deformation-potential-constants. Such type of studies of the optical properties of semiconductors are becoming very useful.

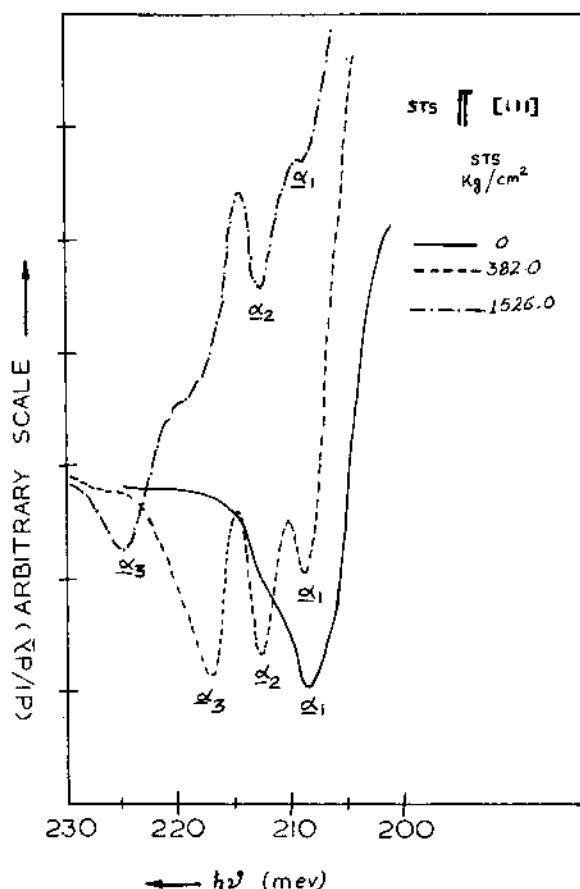


Fig. 10. Derivative transmission spectra of Cd-doped p-InSb at 99°K for stress parallel to [111]. Sample thickness = 0.92 mm. Applied stress values are indicated. From Ref. (17).

VI. CONCLUSIONS

The study of the optical properties of metals and semiconductors has helped in understanding their band structure. However, there is lots more that one does not know and vigorous research is going on in this area. The study of the optical properties of semiconductors has got an added motivation because of its direct applicability in the design of opto-electronic devices. Semiconductor lasers, light-emitting diodes and radiation detectors are some such devices. Many more are on the design table and will be the devices of tomorrow. For example, an intensive effort is being made to replace mechanical card readers in computers by

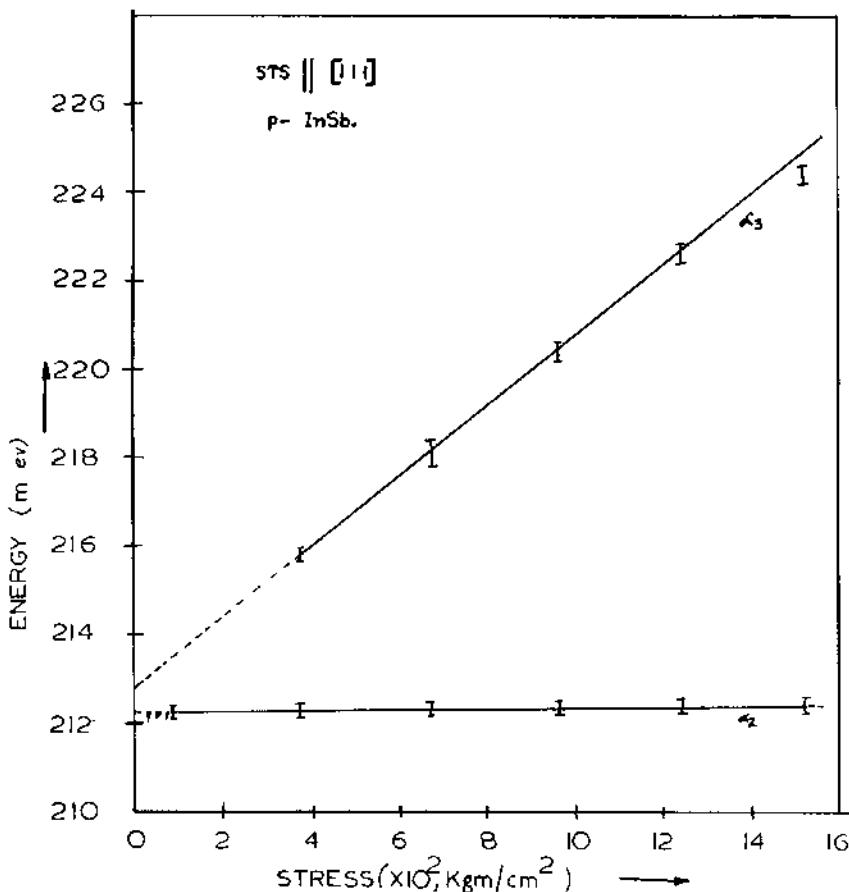


Fig. 11. Slopes of extrema α_2 and α_3 for stress parallel to $[111]$ for p-InSb.

opto-electronic card readers.

REFERENCES

1. F. Stern, Solid State Physics, p. 299, Vol. 15, F. Seitz and D. Turnbull, Eds., Academic Press, Inc., New York (1963).
2. H. Ehrenreich, IEEE Spectrum, 2, 162 (1965).
3. Semiconductors and semimetals, Vol. 3-optical properties of III-V compounds, R.K. Willardson and A.C. Beer, Eds., New York, Academic Press Inc. (1967).

4. J.M. Ziman, *Principles of the theory of solids*, p. 219, Cambridge University Press, (1964).
5. N.B. Colthup, L.H. Daly and S.E. Wiberley, *Introduction to infrared and Raman spectroscopy*, p. 37, Academic Press Inc., New York (1964).
6. P.D. Johnson, *J. Opt. Soc. Am.*, **42**, 278 (1952).
7. Ref. 4, pages 46, 114.
8. H. Ehrenreich and H.R. Philipp, *Phys. Rev.*, **128**, 1622 (1962), hereafter referred to as EP.
9. H.R. Philipp and H. Ehrenreich, Ref. 3, page 93.
10. J. Feinleib, W.J. Scouler, and A. Ferretti, *Phys. Rev.*, **165**, 765 (1968).
11. C.N.R. Rao, W.E. Wahnseidler and J.M. Honig, *J. Solid State Chem.*, (in print).
12. E.J. Johnson, page 125, Ref. 3.
13. L.I. Schiff, *Quantum mechanics*, p. 199, McGraw-Hill, New York (1949).
14. *Semiconductors*, ed. N.B. Hannay, Reinhold Publishing Corporation, New York (1960), page 465.
15. E.J. Johnson and H.Y. Fan, *Phys. Rev.*, **139**, A1991 (1965).
16. I. Balslev, *Phys. Rev.*, **143**, 636 (1966).
17. R. Sharan and E.L. Heasell, *Proc. Intern. Conf. Phys. Semiconductors*, Moscow, (1968). Publishing House "Nauka", Leningrad, Vol. II, p. 1194.
18. R. Sharan and E.L. Heasell, *J. Phys. Chem. Solids*, **31**, 541 (1970).

PHASE TRANSFORMATIONS IN SOLIDS

C.N.R.Rao

Department of Chemistry

Indian Institute of Technology, Kanpur, India

I. INTRODUCTION

The transformation of one polymorph to another constitutes an important aspect of phase transformations in solids. Phase transformations in many solids are accompanied by interesting changes in electrical, magnetic, dielectric and thermal properties in addition to changes in the crystal structure of the material. The subject of phase transformations in solids is of interest to workers in a number of disciplines dealing with solid materials and some aspects of interest to chemists have been reviewed by Ubbelohde¹, Rao and Rao² and Smoluchowski³. A brief account of various types of crystal structure transformations will be presented in this chapter.

II. THERMODYNAMIC CONSIDERATIONS

Classical thermodynamics gives a simple treatment of the equilibrium between two phases. The equilibrium properties of each phase are defined by the Gibbs free energy function, G, the pressure, P, and the temperature, T. The G-P-T surfaces of two phases are considered to be independent of each other and in a phase transformation the surfaces intersect at the transition point. If the transformation takes place from a low-temperature to a high-temperature form, there will be an increase in entropy and an associated volume change. The two thermodynamic relations which describe the differences between the slopes of tangents at the intersection are:

$$(\partial G / \partial T)_P = -S \quad (1)$$

$$(\partial G / \partial P)_T = V \quad . \quad (2)$$

These two relations are related by the Clausius-Clapeyron equation.

The classical thermodynamic approach in describing phase transformations is inadequate since a variety of systems are known to exhibit transformations occurring over a wide range of temperatures. These systems show anomalous changes in specific heats and specific volumes and show evidence of premonitory phenomena. The transition temperature, T_t , in these transformations is taken to be that in which the heat capacity or the coefficient of expansion shows a maximum variation. This class of transformations has often been referred to as gradual, smeared or diffuse transformations.

The earliest attempt to treat the thermodynamics of the gradual transitions was by Ehrenfest who considered the case where the free energies as well as their differential coefficients of two forms or structures are equal, but there is a discontinuity in their second differential coefficients. In such cases, the intersections of the G-P-T surfaces will not be sharp, but the surfaces may come into contact with each other at the transition point. According to Ehrenfest, when the G-P-T surfaces touch each other with the same slope but different curvatures at a temperature, there will be discontinuity in the second or higher derivatives of free energies:

$$(\partial^2 G / \partial T^2) = -C_p / T \quad (3a)$$

$$(\partial^2 G / \partial T \partial P) = (\partial V / \partial T) \quad (3b)$$

$$(\partial^2 G / \partial P^2) = (\partial V / \partial P) \quad (3c)$$

The free energy-temperature curves of three typical cases are shown in Fig. 1. The curves in Fig. 1(a) represent transformations predicted by classical thermodynamics where the G - curves intersect sharply at a finite temperature. The diffuse transitions, on the other hand, can be represented by curves in Fig. 1(b). Justi and Von Laue have pointed out that in gradual transitions, the third differential coefficients of G change and not the second at T_t , so that the free energy curves of forms I and II can cross each other. This argument is not entirely satisfactory since it is difficult to conceive of two forms of a substance separated by the anomaly as two phases existing on both sides of the transition point.

From the discussion above, it becomes apparent that one can distinguish, in principle, two types of transformations based on thermodynamics: The discontinuous or the first-order transitions which follow the predictions of classical thermodynamics and the

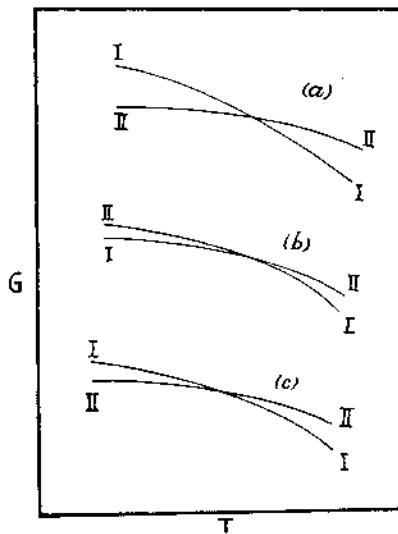


Fig. 1. Free energy-temperature relations in phase transformations:
(a) first-order; (b) and (c) higher order.

continuous (or the diffuse or the so-called second (or higher) order) transformations.

Classical phase rule ($N_f = 2 + C - P$) cannot be strictly applied to phase transformations where there are contributions from surface energy, γ , and the strain energy, ϵ , to the total degrees of freedom. In such cases

$$G = f(p, T, \gamma, \epsilon) \quad (4)$$

$$\text{and} \quad N_f = (C+2) - P + \sum \Pi \quad (5)$$

where Π stands for the additional degrees of freedom.

In the first-order or discontinuous transformations, the difference in the free energies of the two phases is given by:

$$\begin{aligned} 0 &= \Delta G = G_2 - G_1 = \Delta H - T \Delta S = \Delta E - T \Delta S + P \Delta V \\ &= \Delta E + P \Delta V - T (S_2 - S_1) \end{aligned} \quad (6)$$

where ΔE is the difference in internal energies of two phase; if $P \Delta V$ in a transformation is small, one can compare the stabilities of the two phases in terms of their Helmholtz free energies:

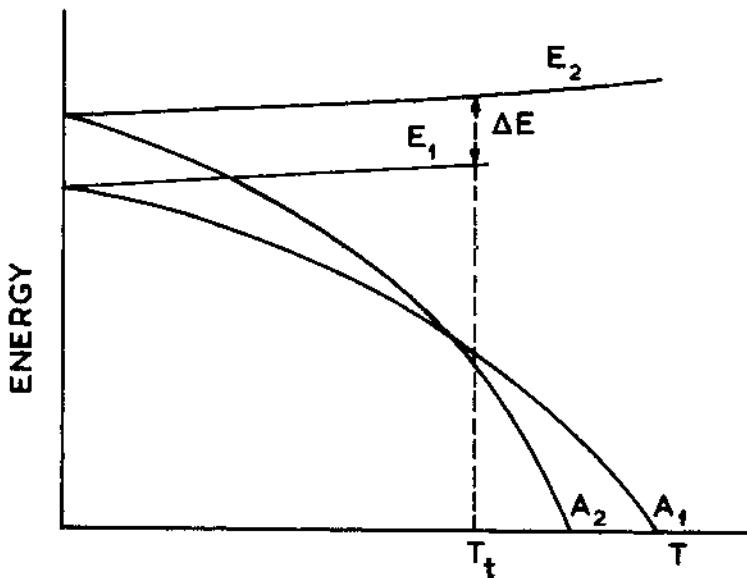


Fig. 2. Variation of the internal energies and free energies with temperature in first-order transformations.

$$0 = \Delta A = \Delta E - T \Delta S = (E_2 - E_1) - T(S_2 - S_1) . \quad (7)$$

According to this equation, the transformation results from the compensation of the lattice energy difference by the entropy difference at the transition temperature. The thermodynamic relations in first-order transitions are graphically represented in Fig. 2. It can be seen that at T_t , the free energy curves intersect each other and $\Delta A = 0$. These transformations involve step-wise increase in energy with increase in temperature and show discontinuity in energy and all other properties at T_t .

Many instances are known where phase transformations are induced by application of pressure. Whenever there is a decrease in volume during a transformation, pressure would favour the transformation and the thermodynamic variables are simply correlated by the Clausius-Clayperon equation. Since the phase produced by the application of pressure will have a lower volume, the decrease in volume is compensated by an increase in the E or in the coordination number. The increase in entropy may be seen as a decrease in disorder.

In higher-order or continuous transformations, the energy increases gradually with temperature until T_t where the rate of increase of energy falls sharply. As a result of this, discontinuity is found in specific heats or specific volumes. Since lambda-shaped specific heat curves are generally obtained in these

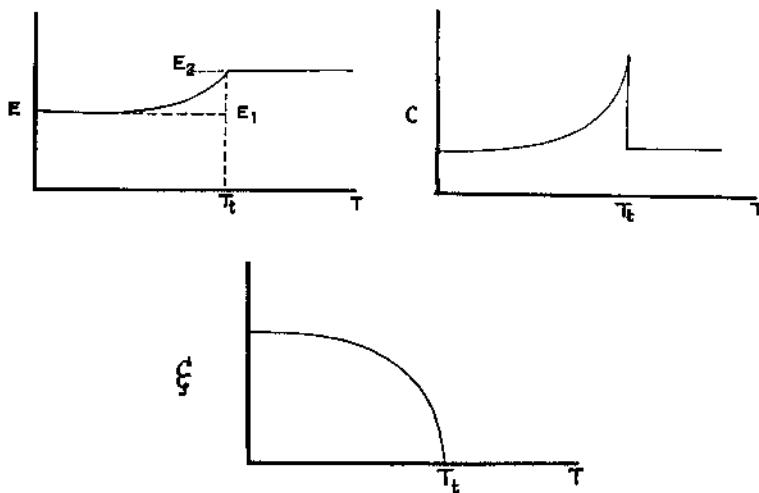


Fig. 3. Variation of internal energy, E , specific heat, C and order, ξ , in second-order transformations.

transformations, these are also called lambda-transformations (Fig. 3). Second or higher-order transformations occur over a temperature interval and are generally associated with increase in disorder with rising temperature. If a solid has perfect order at absolute zero of temperature, a rise in temperature disturbs the order of the structure and with progressive increase in temperature, the structure gets more and more disordered until the transformation temperature T_t is reached. One can introduce an order parameter ξ which is equal to unity at absolute zero and becomes zero at T_t (Fig. 3). The free energy function for such a system may be written as:

$$A(T, \xi) = E(\xi) - TS(\xi) . \quad (8)$$

Although for the sake of simplicity, we may classify transformations as first-order and second-order types, in reality it is difficult to find out exactly whether a transformation belongs to either of these two types. By definition, a first order transformation should occur sharply at one temperature and two phases of the same substance should not coexist at temperatures other than T_t . If a single crystal of a substance is taken through a transition point, the crystal should break up into a random assembly of one or more crystals of the other phase at T_t . If the temperature is then decreased below T_t , the first phase should reform randomly from the second phase and there should be no persistence of the crystal axis of the first phase.

There are a number of transformations which have been classified as first order: tin (grey-white, $\sim 19^\circ\text{C}$); cesium chloride ($\text{CsCl-NaCl}, \sim 480^\circ\text{C}$); NH_4Cl ($\text{CsCl-NaCl}, \sim 184^\circ\text{C}$); sulphur (rhombic-monoclinic, $\sim 96^\circ\text{C}$). While the transformations classified as first order are structurally discontinuous and may involve drastic variations in energy, there are indications of higher order effects (as well as of premonitory phenomena) and thermal hysteresis in many of these transformations. Many transformation have been found to proceed through the formation of hybrid crystals where the two structures coexist within a general pattern of orientation.

There are many phase transformations which strictly belong to neither first nor second order. For example, the ferroelectric transformation of KH_2PO_4 should theoretically be first order but conforms more closely to the second order. The transformations of BaTiO_3 and related materials show lambda-like changes in properties as well as small changes in latent heat. There is super-position of second-order behaviour on first-order transformations in the transformations of alkali sulphates. Many transformations are actually mixed transformations and can not be placed in either category. Even the order-disorder change in second-order transformations is often seen as an abrupt change towards the termination.

Many reversible transformations show hysteresis effects either in terms of the formation of metastable structures in one of the directions or in terms of differences in the transformation temperatures in the two directions⁴. Observation of thermal hysteresis clearly shows that the equilibrium is not complete at all the stages of the transformation as required by thermodynamics. Hysteresis is a necessary consequence of the coexistence mechanism of continuous transformations. An examination of the available data on reversible transformations shows that the thermal hysteresis is related to the volume change in a transformation; thermal hysteresis is generally small when the volume change in the transformation is small (e.g., α - β inversion in quartz). A detailed study of thermal hysteresis in transformations has been carried out by Rao and coworkers⁴.

Occurrence of metastable intermediate phases has been noted in some transformations. The metastable phases in such instances are probably related crystallographically to either of the two phases. The occurrence of a metastable phase in KNO_3 is shown in Fig. 4. Bi_2O_3 transforms from the monoclinic phase to the cubic phase at 730°C and on cooling the cubic phase, it reverts back to the monoclinic phase through a metastable tetragonal phase (Fig. 4).

Irreversible crystal structure transformations are found in many systems where one of the polymorphs is metastable in a particular temperature (or pressure) range. In such transformations, a polymorph transforms to the other form on heating above a particular

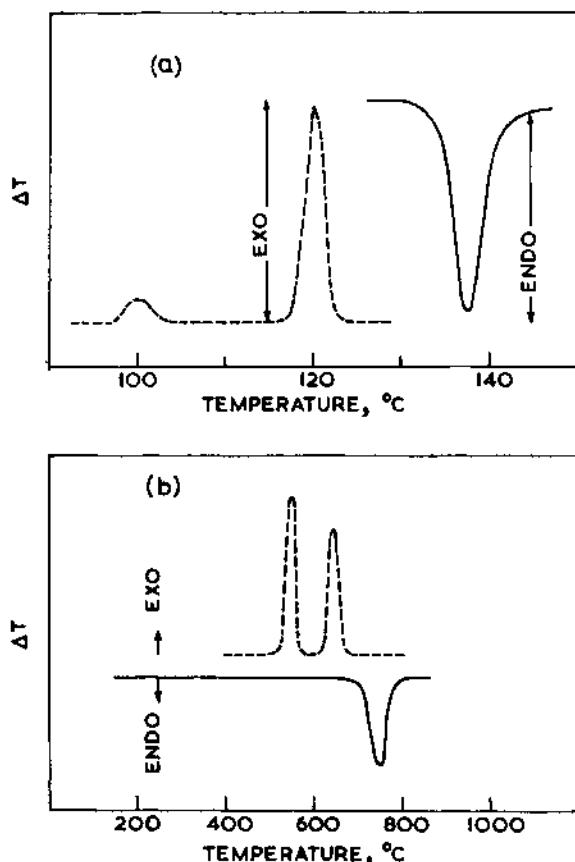


Fig. 4. Occurrence of metastable phases as shown by differential thermal analysis curves: (a) KNO_3^4 , (b) Bi_2O_3^5 . Broken lines are cooling curves.

temperature and remains in the new transformed structure even after cooling. It is not possible to assign any critical transformation temperature in an irreversible transformation and the transformation will be a function of time as well as temperature. Typical examples of the irreversible transformations are the following: anatase - rutile (TiO_2); aragonite-calcite (CaCO_3); cubic-hexagonal or cubic-monoclinic of rare-earth sesquioxides.

Simple thermodynamic considerations indicate that one polymorph of a solid can be more stable under a specified set of conditions than the others. It is however, difficult to strictly define the thermodynamic stability of a polymorph since a variety of physical factors as well as compositional differences drastically affect the stability of polymorphs. Thus, if there is any

difference in the stoichiometry between two polymorphs, one can not really classify them as belonging to the same polymorphic group. An interesting example of a transformation where physical factors affect the stability of a polymorph is provided by ZrO_2 ⁶. ZrO_2 undergoes a reversible monoclinic-tetragonal transformation at 1175°C (1035°C on cooling). The tetragonal ZrO_2 can be made stable at ordinary temperatures if the particle size is small; the stabilization is apparently due to the excess surface energy of the small particle size samples. The effects of particle size on phase transformations have been investigated by Rao and coworkers^{4,4a}, who find that the transformation temperature, the enthalpy and the thermal hysteresis are significantly affected by particle size.

If different samples of a solid contain different quantities of impurities, the regions of thermodynamic stability will vary. Thus, calcium carbonate with a small amount of strontium will tend to make the aragonite phase more stable than the calcite phase⁷. Presence of small quantities of sulphate ions or other impurities in TiO_2 (anatase) enhances the temperature of transformation to rutile considerably⁸. It is therefore difficult to define the thermodynamic stability unless the detailed history of the sample including its exact chemical composition is known.

III. STRUCTURAL CHANGES IN TRANSFORMATIONS

Buerger⁹ has classified various types of thermal transformations based on structural changes involving the primary or higher coordination. Transformations where there are changes in the primary coordination involve a more drastic change in energy and structure rather than those where there are changes in the higher coordination. Accordingly, one can classify thermal transformations of the first order into two categories: first coordination transformations and higher-coordination transformations.

Changes in primary coordination can take place by a reconstructive transformation, where the first-coordination bonds are broken and reformed. Such transformations will involve higher energies of activation and will be sluggish. Further, there may be no symmetry relation between the two phases (Fig. 5). Reconstructive transformations give rise to large changes in cell dimensions, symmetry, internal energy, specific heat etc. Typical of these transformations is the aragonite-calcite transformation of $CaCO_3$ (400°C) where the coordination number changes from 6 to 9. Changes in primary coordination may take place through a dilatational mechanism as suspected in the thermal transformation of cesium chloride (480°C) or ammonium chloride (184°C). Dilatational transformations are likely to be rapid compared to reconstructive transformations. Transformations of $BaTiO_3$ and $PbTiO_3$ may belong to this category.

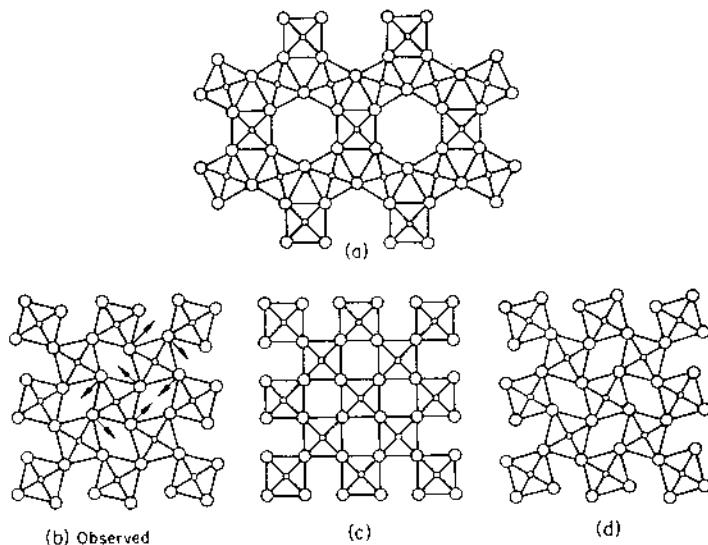


Fig. 5. Transformation of structure (a) to (b), (c) or (d) involves changes in first coordination while change from (b) to (c) is a displacive transformation (After Buerger).

since the coordination of Ti changes from 6 to 5 in the transformation.

Many transformations involving changes in higher coordination may also proceed through the breaking of the primary bonds and for this reason, the energy changes and other features of the reconstructive transformations involving higher coordination may resemble those of the reconstructive first coordination transformations. Typical examples of these are the sphalerite-wurtzite (ZnS), cristobalite-tridymite-quartz (SiO_2) transformations. Transformations between polytypes may also fall under this category. The transformation of TiO_2 (anatase or brookite to rutile) involves changes in secondary coordination, but show the characteristics of first coordination transformations.

In some transformations, changes in higher coordination can be effected by a distortion of the primary bond. Such transformations may be called distortional or displacive transformations (Fig. 5). These transformations may involve considerably smaller changes in energy and are usually fast. In the displacive transformations, the high-temperature form is usually more open and has higher specific volume, specific heat and symmetry. Examples of displacive transformations include the high-low transformations of

quartz (575°C), tridymite (160° and 105°C). Anatase (TiO_2) shows a displacive transformation prior to its reconstructive transformation to rutile.

IV. KINETICS OF TRANSFORMATIONS

The rate of a phase transformation is determined by the energy barrier opposing the process and an energy of activation will be required to surmount the energy barrier to enable the reaction to proceed. The basic equation relating rate constants with temperature in a thermal transformation is the classical Arrhenius equation:

$$k = A \exp(-E_a/RT) \quad (9)$$

where E_a is the energy of activation, k is the rate constant and A is a constant for the transformation.

The magnitude of E_a of a transformation depends on the mechanism of the transformation and the changes in coordination involved. Any transformation which involves changes in primary coordination (directly or indirectly) will need a much larger E_a than those involving changes only in higher coordination.

Phase transformations are generally initiated by the formation of a number of small particles or nuclei of the new phase which then grow until the transformation is complete. The presence of the phase boundary or surface between two phases increases the free energy and provides the energy barrier in the kinetics of nucleation. Nucleation involves the assembly of suitable kinds of atoms or units by diffusion or some other mechanism, change of the structure into the intermediate structure(s), followed by the formation of the nuclei of the new phase. An activation energy may be associated with each of these steps, but the experimental energy of activation will refer to the rate-determining step.

Nucleation may be homogeneous or heterogeneous. Heterogeneous nucleation takes place preferentially at grain boundaries, foreign impurities or dislocations. Homogeneous nucleation can occur in the absence of such defects; once a nucleus of the new phase is formed, however, it produces an imperfection in the host structure. The number of nuclei formed per unit volume is given by,

$$n = N \exp(-\Delta G_n/RT) \quad (10)$$

where ΔG_n is the free energy of formation of a nucleus. Also,

$$\Delta G_n = \Delta G_c + \Delta G_s + \Delta G_E \quad (11)$$

where the subscripts c, s and E refer to the free energy change due to the chemical change, the surface free energy and the elastic strain energy respectively. ΔG_s and ΔG_E are both positive, while ΔG_c is negative; ΔG_c is proportional to the volume of the nuclei and ΔG_s to the surface area of the nuclei. Ignoring ΔG_E , we can therefore define the energy change, ΔG_n , for the formation of one nucleus as,

$$\Delta G'_n = (4/3)\pi r^3 \Delta G'_c + 4\pi r^2 \Delta G'_s \quad (12)$$

where r is the radius of the nucleus and $\Delta G'_c$ and $\Delta G'_s$ stand for the free energy change per unit volume and surface free energy per unit area, respectively. The critical size of nuclei is found by setting the derivative of equation (12), with respect to r equal to zero.

$$r^* = -2\Delta G'_s / \Delta G'_c \quad (13)$$

In Fig. 6, the dependence of the free energy change on the radius

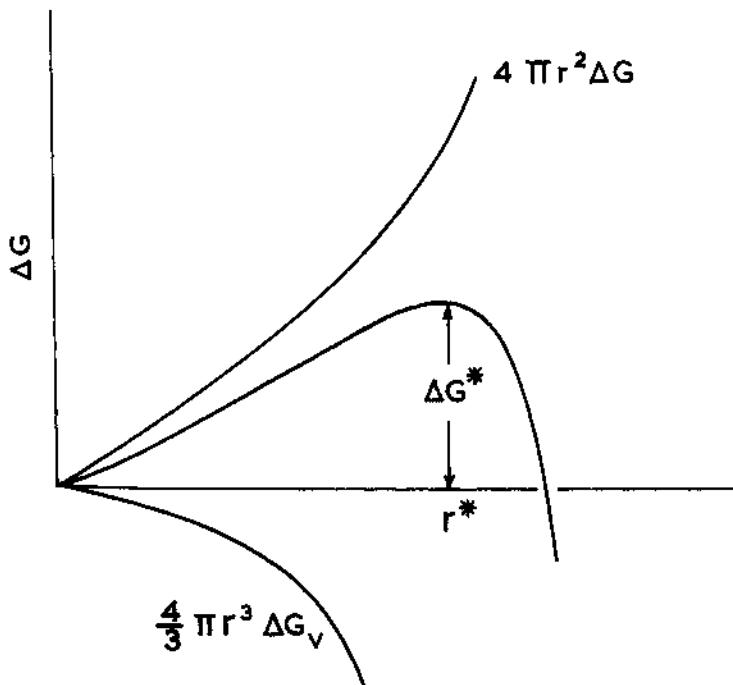


Fig. 6. Dependence of the free energy change upon the radius of the nucleus.

of the nucleus is shown schematically. Imperfections in solids generally decrease ΔG_n° and favour the nucleation process. In such heterogeneous nucleation processes, the interfacial energy between the new phase and the imperfection would be an important factor.

The over-all activation energy and the transformation rate depend not only on the kinetics of nucleation, but also the propagation (or growth) processes. The propagation process may involve material transfer across the phase boundary into the new phase and the transfer into the interior of the new phase by diffusion. Each of these steps is associated with an activation energy. In experimental studies, the kinetics of phase transformations are usually discussed in terms of the rates of nucleation and the propagation steps. Most of the activation energy of a transformation is generally utilized in the nucleation process.

In recent years, the kinetics of various phase transformations have been reported in the literature. Thus, the kinetics of the transformation of tin¹⁰ (white-grey) are found to comply with Avrami's equation¹¹:

$$(1 - f) = \exp(-At^k) \quad (14)$$

where f is the fraction transformed, t is the time and $k = 3$ (Fig. 7). The results seem to be in agreement with the three-dimensional growth of the nuclei formed during the beginning of the transformation. For the grey-white transformation of tin the value of k varied between 1.5 and 2. The results show that the rate of nucleation and not the rate of propagation, determines the value of k . For a value of $k = 1$, the above equation (14) becomes the first order rate equation. For $k > 1$ the reaction is autocatalytic, spontaneously giving rise to nuclei after initial induction. Order-disorder theory has also been applied to the study of phase transition kinetics^{11a} (see the chapter on order-disorder theory by J.M. Honig in this volume).

Several other forms of Avrami's equation^{8,12} have been employed to interpret kinetics of phase transformations, but we shall not discuss them presently. Many workers have reported that the first-order rate-law is obeyed in some crystal structure transformations (Fig. 8). Typical of these are the transformations of brookite (TiO_2) and anatase (TiO_2) to rutile. The first-order law has also been found in the transformations of germanium dioxide¹³. Transformations of defect cubic oxides of the rare earths to the hexagonal or monoclinic forms are also found to obey the first order law^{14,15}. Impurities affect transformation rates markedly; in some systems these effects have been documented^{2,7,8}.

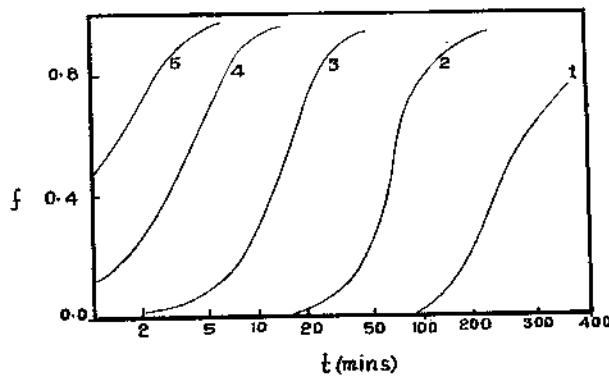


Fig. 7. Transformation of grey tin¹⁰: 1, 25°; 2, 27.5°; 3, 30°; 4, 32.5° and 5, 35°C.

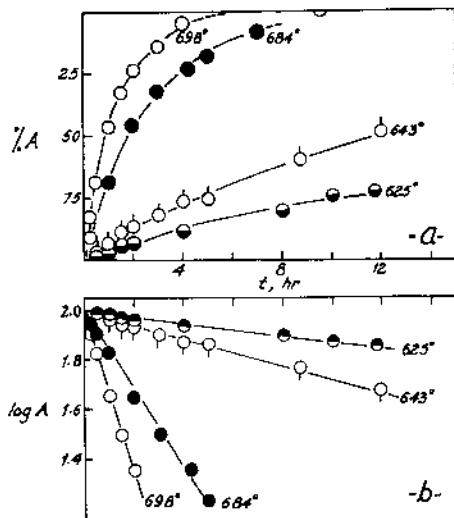


Fig. 8. First order rate data on the transformation of spectroscopically pure anatase to rutile^{8a}.

V. ORDER-DISORDER TRANSITIONS

Many substitutional solid solutions show ordered structure at

low temperatures, if they are properly annealed. The ordered structure transforms into the disordered structure over a wide temperature range (typical of second-order transformations) and the change becomes very rapid near the transition temperature. The transformation temperature is often referred to as the Curie point of order. The degree of order in the arrangement of atoms in the lattice site is measured by a long-range order parameter, s , and a local order parameter, σ . Both s and σ decrease (from a value of 1) with increase in temperature. The rate of decrease is large at T_t typical of co-operative phenomena. Good statistical theories for the variation of order and associated physical properties with temperature have been developed.¹⁶

We shall now briefly consider the theory of long-range order of Bragg and Williams (Also see J.M. Honig's chapter on Order-Disorder Theory in this volume) for the simple case of a binary alloy containing equal numbers of atoms A and B distributed over a total number of N sites. In the state of complete order the A and the B atoms occupy the sites α and β , respectively. When there is long-range disorder, we can define s in such a way that $s = 1$ when there is complete order and $s = 0$ when there is complete disorder. Thus, we obtain the following relations for the numbers of A and B atoms on different sites in terms of s and N:

$$\text{Number of A atoms on } \alpha \text{ sites} = (1+s)N/4 = N_{A\alpha}$$

$$\text{Number of B atoms on } \beta \text{ sites} = (1+s)N/4 = N_{B\beta}$$

$$\text{Number of A atoms on } \beta \text{ sites} = (1-s)N/4 = N_{A\beta}$$

$$\text{Number of B atoms on } \alpha \text{ sites} = (1-s)N/4 = N_{B\alpha} .$$

Since $\alpha = \beta = N/2$, the number of AA pairs (or BB pairs) formed is given by,

$$N_{AA} = N_{A\alpha} [ZN_{A\beta}/(N/2)] = \frac{(1+s)N}{4} \frac{8(1-s)N}{(N/2)4} = (1-s^2)N \quad (15)$$

where the coordination number Z is taken as 8 in the case of a body centered cubic lattice (For closest packing Z = 12). Similarly,

$$N_{BB} = (1-s^2)N \quad (16)$$

The number of AB pairs is given by,

$$\begin{aligned} N_{AB} &= [N_{A\alpha} N_{B\beta} + N_{A\beta} N_{B\alpha}] Z/(N/2) \\ &= 2(1+s^2)N \end{aligned} \quad (17)$$

By assuming that only the nearest neighbor interactions contribute to the internal energy, we find that at any given temperature,

$$E = N_{AA} E_{AA} + N_{BB} E_{BB} + N_{AB} E_{AB} + \int_0^T C_p dt \quad (18)$$

$$E = N(1-s^2) (E_{AA} + E_{BB}) + 2N(1+s^2) E_{AB} + \int_0^T C_p dt \quad (19)$$

The configurational entropy is calculated in terms of the number of possible ways of arranging A and B on α and β sites. Remembering that $N_{A\alpha} = N_{B\beta}$ and $N_{A\beta} = N_{B\alpha}$, we get,

$$S = k \ln \left[\frac{(N/2)!}{N_{A\alpha}! N_{A\beta}!} \right]^2 \quad (20)$$

Using Stirling's approximation

$$S \approx k N \ln 2 - \frac{kN}{2} \left[(1+s) \ln(1+s) + (1-s) \ln(1-s) \right] \quad (21)$$

From the above relation, it can be readily seen that when there is complete order, $s = 1$ and $S = 0$. Similarly when there is complete disorder, $s = 0$ and $S = kN \ln 2$.

The free energy is given by,

$$A = E - TS - T \int C_p d \ln T \quad (22)$$

where E and S are given by equations (19) and (21) respectively. At $T < T_c$, the free energy minimum will be near $s = 1$; the minimum is exactly at $s = 1$ only when $T = 0^\circ K$. Perfect order is therefore not found at ordinary temperatures. At $T = T_c$, there is some order present, but long-range order disappears. By differentiating the free energy with respect to s and setting the derivative equal to zero we obtain the temperature dependence of s shown in Fig. 9.

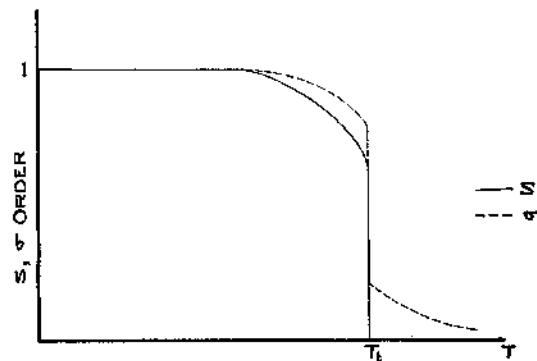


Fig. 9. Variation of order parameters with temperature.

Order-disorder transformations are accompanied by crystal structure changes. The existence of ordered structures is evidenced by x-ray diffraction patterns in terms of the appearance of the super lattice reflections in addition to the fundamental reflections. Order-disorder transformations are generally investigated by measurements of specific heat, electrical resistivity, mechanical (elastic) properties, and magnetic properties.

Specific heat measurements have shown that the curve does not fall to zero at $T \geq T_c$, but levels off at a small value. This is because of the assumptions made in the theory; since disorder is caused by the thermal motions of atoms long-range order is slowly replaced by short-range order. In short-range order, we have small domains of order whose sizes vary with temperature. Further, the energy of interaction in unlike pairs is generally less than in like pairs; $E_{AB} < 1/2 (E_{AA} + E_{BB})$. Bethe defined the short-range order parameter as,

$$\sigma = \frac{2}{AB} - 1 \quad (23)$$

where AB is the fraction of AB pairs at a given temperature. The fraction of unlike pairs is given by $1/2 (1-\sigma)$, while the fraction of like pairs is given by $1/2(1+\sigma)$.

$$\frac{\frac{1}{2}(1+\sigma)}{\frac{1}{2}(1-\sigma)} = \exp(-\Delta\varepsilon/kT) \quad (24)$$

where $\Delta\varepsilon = E_{AB} - 1/2 (E_{AA} + E_{BB})$. The dependence of σ on temperature is shown in Fig. 9; σ does not come to zero at the critical temperature, but approaches zero.

One of the typical examples of a system exhibiting order-disorder transformations is that of copper-gold. The composition 0.5 Cu-0.5 Au is a completely random solid solution at high temperatures and it approaches complete order at low temperatures. The 0.75 Cu - 0.25 Au solid solution is also random at high temperatures and it goes into an ordered Cu₃Au structure. In the ordered state of β -brass (ZnCu), one type of atom is at the body centers while the other type of atom is at the corners of the unit cell. Another interesting example of order-disorder transitions is provided by manganese with small amounts of copper in it. This Mn-Cu alloy has the same FCC structure of manganese. The alloy shows a cubic-tetragonal (disorder-order) transition around 250°C which does not involve any major atomic rearrangements but is seen in terms of the alignment of electron spins. Below 250°C, the exchange forces favor an ordered antiparallel spin orientation in nearest neighbors, while above 250°C the spins of nearest neighbors

are randomly oriented. Such spin orientation can be readily examined by neutron diffraction experiments.

VI. MARTENSITE TRANSFORMATIONS

One of the best documented phase transformations in solids is the Martensite transformation, commonly encountered in metal systems¹⁷. Unlike the nucleation and growth transformations, the Martensite transformations proceed by coordinated movements of a large number of atoms within a relatively short time. These transformations are essentially diffusionless since the atomic displacements are very small. It is not possible to stop or slow down a martensite transformation by thermal quenching because of the high rates associated with the transformation (even at very low temperatures). Martensite products have little in common with regard to the structure and there appears to be no simple way to predict the structure of the martensite product on the basis of the structure of the parent.

Martensite may denote the product phase resulting from the reaction or the mechanism of a specific type of phase transformation. Martensite transformations are mainly found in metal systems and originally referred to the transformation of austenite (FCC, Fe-C) to martensite (BCT). There are also some compounds which are known to show martensite transformations. NaCN which exists in the NaCl structure above 15°C (and in the orthorhombic structure below this temperature) shows the presence of an ordered rhombohedral structure when heated above 150°C; the rhombohedral structure persists even on cooling.¹⁸

VII. FERROELECTRIC TRANSFORMATIONS

Ferroelectrics form a subgroup under the pyroelectrics and are characterized by the reversibility of their permanent polarization by an applied electric field¹⁹. Such a reversal gives rise to non-linear dielectric behaviour and to a hysteresis loop. In ferroelectric materials the temperature at which the free energies of the polar state and non-polar state (or the less-polar state), approach each other, is called the Curie temperature. At the Curie temperature, where $\Delta G = 0$, a phase transformation is observed as evidenced by the spontaneous polarization. Phase transformations from a ferroelectric phase I to another ferroelectric phase II, from a ferroelectric phase to a paraelectric phase or from a ferroelectric phase to an antiferroelectric phase have been observed in a variety of substances. Accompanying these transformations one finds anomalies in heat capacity, breaks in thermal expansion curves and variations in optical properties (including

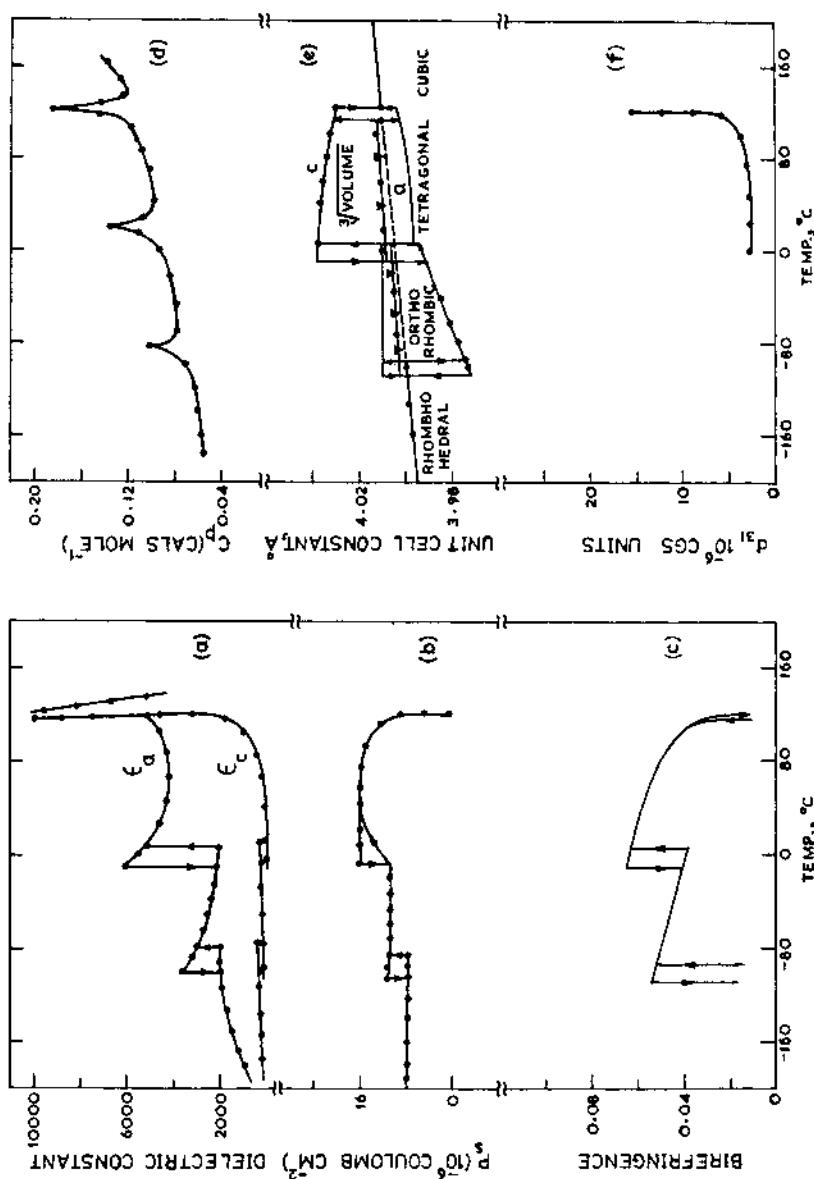


Fig. 10. Transformations in BaTiO_3 accompanied by changes in (a) dielectric constant (b) spontaneous polarization (c) birefringence (d) heat capacity (e) lattice dimensions and (f) piezoelectric coefficient d_{31} . (see ref. 19).

vibrational spectra), in addition to changes in crystal structure and dielectric properties Fig. 10). For detailed discussion of ferroelectric transitions, the reader is referred to the chapter by E.C. Subbarao in this volume.

VII. MAGNETIC TRANSITIONS

Many of the magnetic solids show interesting transitions at the Curie or the Neel temperature²⁰ (see N. Menyuk's chapter on Magnetism for details). For example, the high-temperature paramagnetic phases of ferrites (MFe_2O_4 or $MFeO_3$ where M is a metal ion) transform to the low-temperature ferromagnetic phases below the T_c . Similarly, many magnetic solids (e.g., MnO, FeO, NiO) transform to the antiferromagnetic state at T_N . These transformations involve the cooperative transformation of magnetic dipoles and show the features of the order \rightleftharpoons disorder transformations. At the transition temperature marked anomalies are noticed in the specific heat, thermal expansion coefficient, electrical resistivity and magnetic susceptibility. Some of the transitions exhibit thermal hysteresis similar to first order transitions.

Besides paramagnetic \rightleftharpoons ferromagnetic and paramagnetic \rightleftharpoons antiferromagnetic transitions, magnetic solids may also undergo paramagnetic \rightleftharpoons ferromagnetic, antiferromagnetic \rightleftharpoons ferrimagnetic or antiferromagnetic \rightleftharpoons ferromagnetic transitions.

IX. SEMICONDUCTOR-METAL TRANSITIONS

Solids have been classified as metals, semiconductors or insulators on the basis of their electrical resistivities (see the chapter by J. M. Honig on Transport Properties). Many of the common substances such as NiO, MnO, Fe_2O_3 , TiO_2 , Pr_6O_{11} , and MnS_2 are either insulators or semiconductors. There are many inorganic compounds which are metals, typical of these being TiO , NbO , ReO_3 and CrO_2 .

Some solid state materials are semiconducting upto a specific temperature at which the conductivity rises sharply by several orders of magnitude. Such semiconductor \rightleftharpoons semimetal transitions have been observed in VO_2 , V_2O_3 , Ti_2O_3 and a few other substances²¹. The semiconductor \rightleftharpoons semimetal transitions are generally accompanied by marked changes in lattice dimensions, magnetic susceptibility and heat capacity of the solid²¹ (Figs. 11 and 12). The metallic conductivity above the transition temperature probably results from the formation of a conduction band through direct cation-cation (or cation-oxygen-cation) bonding since the radii of ions like

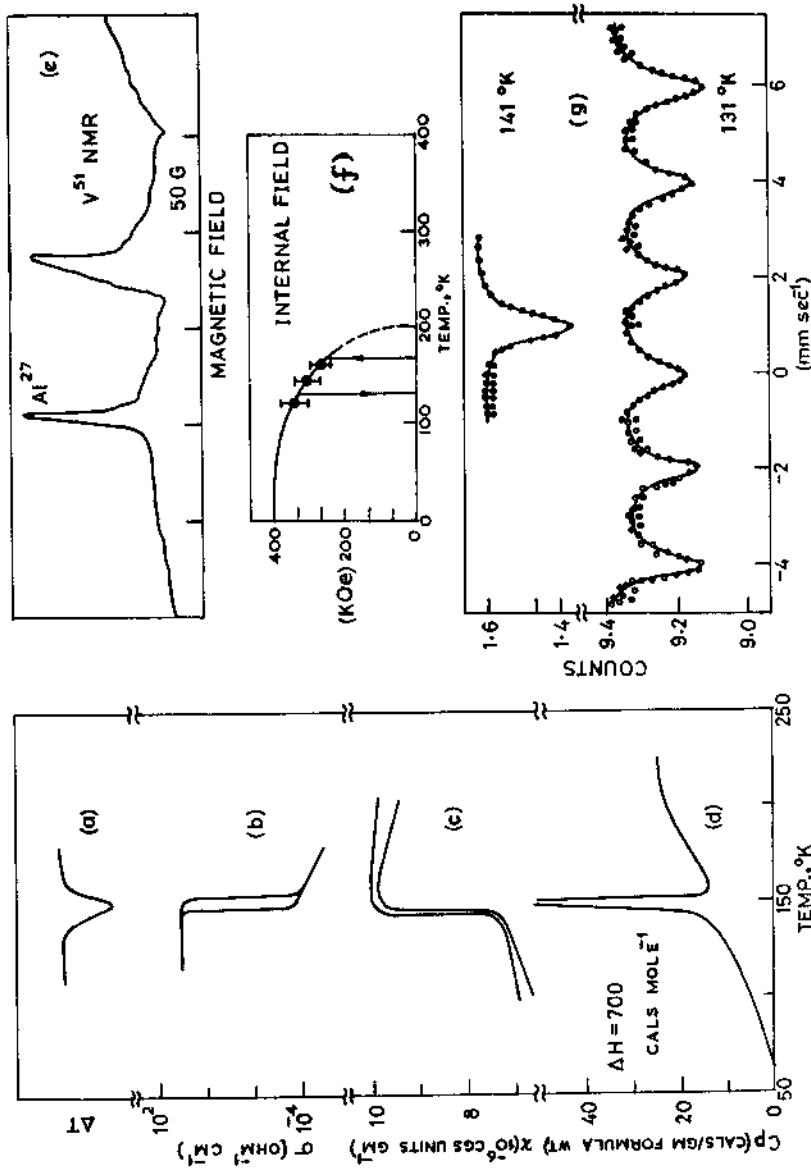


Fig. 11. Semiconductor-metal transition observed in V_2O_3 by (a) DTA anomaly^{27(a)}; measurements of (b) magnetic susceptibility²⁵ and (d) heat capacity as a function of temperature²⁹; (e) NMR spectrum of ^{51}V in V_2O_3 above T_{t}^{27} . The signal disappears below T_{t} ; (f) Variation of internal field of ^{57}Fe doped V_2O_3 ; (g) Mössbauer spectra of ^{57}Fe doped V_2O_3 ²⁸.

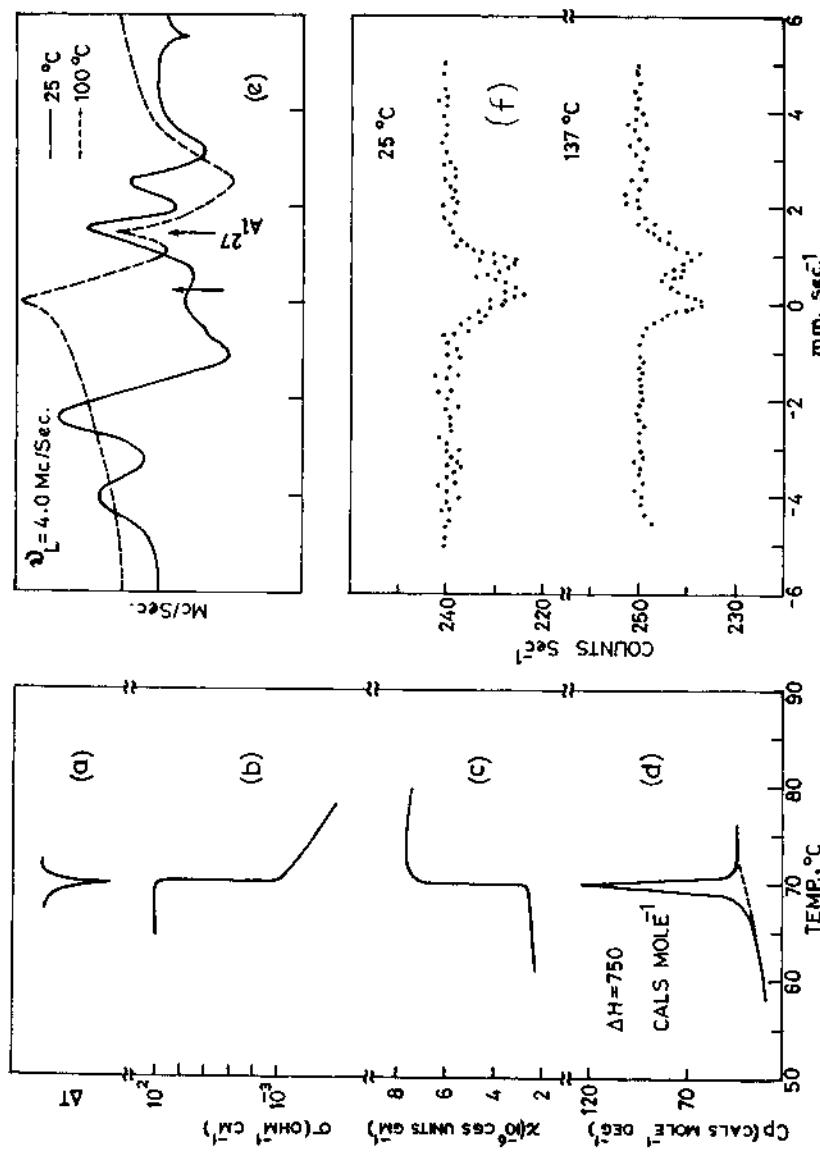


Fig. 12. Semiconductor-metal transition observed in VO_2 by (a) DTA anomaly⁵³; measurements of (b) electrical conductivity⁵³, (c) magnetic susceptibility⁵⁰ and (d) heat capacity as a function of temperature; (e) NMR spectra of ^{51}V in VO_2 below and above T_{t31} . (f) Mössbauer spectra of ^{57}Fe doped VO_2 below and above T_{t32} .

Ti^{+3} or V^{+3} are quite large. In order to account for the origin of the energy gap in the low temperature phase several models have been proposed. Morin²² suggested that the transition is associated with a change from a paramagnetic to an antiferromagnetic state. Mott proposes that the substances are metallic or non-metallic depending on the degree of overlap between the 3d wave functions; the non-metallic state according to him should be in a ferromagnetic or antiferromagnetic spin arrangement. Goodenough²⁰, on the other hand, suggests that the origin of the energy gap in the low-temperature phase is not magnetic ordering but a change in chemical bonding (due to the pairing of neighbouring cations). Adler and Brooks²³ have proposed that the transition originates from crystalline distortion and is not related to magnetic ordering. We shall now briefly examine the features of some of the semiconductor-semimetal transitions; these transitions have been discussed by J. M. Honig in his chapter on Transport Properties in this volume.

V_2O_3 has the corundum structure (rhombohedral symmetry) at room temperature and transforms to a monoclinic symmetry around 150°K with a volume change of 3.5% and thermal hysteresis of about 18° (first order transition). The electrical resistivity changes from 10^4 to 10^{-3} ohm cm at the transition temperature T_t ²⁴. The magnetic susceptibility also shows an anomaly at T_t ²⁵. Recent neutron diffraction studies have shown evidence of magnetic ordering in the low temperature phase²⁶, NMR (V^{51})²⁷ and Mössbauer studies²⁸ also establish antiferromagnetic ordering in this phase (see Fig. 11 for changes in various properties at T_t). A specific heat anomaly is seen to accompany the transformation²⁹. If the antiferromagnetism of the low-temperature phase is ignored the Goodenough model appears to be satisfactory in explaining the gross features of the transition.

VO_2 transforms from monoclinic phase (MoO_2 type) to tetragonal phase (rutile type) at 340°K; there is also an appreciable thermal hysteresis (~15°) (first-order transition). Magnetic susceptibility measurements³⁰ as well as NMR (V^{51})³¹ and Mössbauer studies³² show no evidence of magnetic ordering in VO_2 . The electrical resistivity shows a sharp transition at 340°K³³ (with a change by a factor of 10⁵). All the experimental observations (see Fig. 12 for changes in various properties at T_t), seem to fit the model of Goodenough. As one would expect, the magnitude of the change in susceptibility accompanying the magnetic transition of VO_2 decreases in solid solutions of VO_2-TiO_2 as the percentage of TiO_2 is increased³⁴. The magnitude of crystal distortion, enthalpy change as well as the conductivity jump decrease with % TiO_2 ; at 60% TiO_2 the solid solution exhibits no transition. Below 60% TiO_2 the transitions are semiconductor-semiconductor type^{34a}.

Unlike V_2O_3 and VO_2 , Ti_2O_3 (corundum structure) shows an

electrical transition over a wide range of temperatures (390-470°K) which is not accompanied by a change in crystal symmetry^{34b} (higher order transition). Lattice dimensions, however, show changes in the same temperature range^{34c}. The volume change in the transition is about 7%. Recent neutron diffraction studies show no evidence of antiferromagnetic ordering in Ti_2O_3 ^{34b}. The magnetic susceptibility shows a small change around the transition temperature. The transition is also evidenced by a specific heat anomaly. All the experimental observations on the Ti_2O_3 transition can be explained in terms of the distentions of the unit cell and the accompanying changes in the band gap^{34b}. Interestingly, it is found that a solid solution of 5% V_2O_3 in Ti_2O_3 does not show the electrical transition; instead it is metallic throughout the temperature range. The lattice parameters of this solid solution are nearly the same as the corresponding values for pure Ti_2O_3 after it undergoes the transition^{34c}.

The three systems discussed above, V_2O_3 , VO_2 and Ti_2O_3 , provide three different behaviours: (i) V_2O_3 with first-order transition and antiferromagnetism of the semiconducting phase, (ii) VO_2 with first-order transition, but no antiferromagnetism in the low-temperature phase and (iii) Ti_2O_3 with a higher-order transition and no antiferromagnetism in the low-temperature phase. Obviously, there is need for a unified theoretical interpretation of these observations; the theory should combine the features of the theories of Morin, Goodenough and Adler-Brooks.

X. HIGH PRESSURE TRANSFORMATIONS

Considerable progress has been made in recent years in attaining high pressures³⁵ and these new developments have made it possible to examine the P-V-T relations as well as various phenomena in solids under high pressures. High pressures may be produced by static methods or by the shock method. Crystal structure changes under high pressures have been studied in a variety of systems by employing optical, x-ray diffraction and electrical measurements. The high pressure phases are always more dense and are generally associated with higher coordination numbers. In addition to phase changes, other interesting changes in physical properties are often noticed under the influence of high pressures. Thus, many insulators are transformed into metals at high pressures; graphite is converted to diamond at high pressures and temperatures.

Pressure transformations of different kinds have also been studied in many metals and alloys³⁵. The types of transformations studied include diffusionless martensite transformations and order-disorder transformations. We shall not deal with these metal systems presently; the reader is referred to the book edited by Paul and Warchauner³⁵ for details.

The reconstructive transformation of graphite to diamond has been recently shown to take place under high pressures; the transformation is accompanied by a change in the coordination of carbon from 3 to 4. It has been shown that diamond on application of pressure transforms to a metallic state. It is interesting to note that while yellow phosphorus does not give any new modification under pressures upto 20,000 kg. cm⁻², the red phosphorous exhibits a reversible transformation even at about 4000 kg. cm⁻² and at 600°C.

The halides of potassium and rubidium have been shown to transform from the rock salt structure to the CsCl structure under pressure. The fluorides do not exhibit any change in structure under these conditions. The pressure transformation data of alkali halides have been successfully employed in the calculation of lattice energies and repulsive parameters in ionic halides (see Section XI). Calcium carbonate undergoes a transformation from calcite to aragonite on application of pressure; the transformation is reconstructive and is accompanied by a change in the primary coordination. The transformation of aragonite to calcite takes place on heating. The same change is also effected by grinding. TiO₂ and AgI are well-known examples where grinding produces the more stable form at the temperature of grinding. The transformation takes place under the influence of shearing-stresses superimposed on quasi-static pressures as verified in the transformations of SiO₂ (quartz - coesite), and CaCO₃ (calcite-aragonite). Transformations which do not normally take place below 300°C can be made to occur between 0° and 100°C by the application of displacive shearing stresses; rates of reactions are also considerably enhanced.

In the transformation of quartz to coesite at high pressures only changes in the secondary coordination take place, coesite having higher secondary coordination. The transformation temperature of high-low quartz is decreased by application of pressure. The coordination number increases with increasing pressures as evidenced in the case of aluminium silicates.

High pressure pulses of a short duration can be accomplished by the 'shock method'. Shock-induced phase transition studies have been investigated and compared with studies of transition under static pressures. Under shock-induced pressures, no transition is possible unless the phase change can occur within a microsecond, and instead, the system may remain in a metastable state. Application of a one-dimensional shock to a sodium chloride single crystal in different orientations indicated that the decrease in the interatomic distances is in part compensated by the redistribution of atoms. The static and shock data do not differ much. Potassium chloride transformed within the shock time and on removing the shock, the crystal assumed the original structure; the transformation is reversible well within one μ sec. since the

pressure release is very fast. Red phosphorous was known to transform to the black variety at 45 kbars at room temperature in the presence of shear. But with shock technique, the transformation takes place around 30 kbars. Diamond becomes metallic at about 600 kbars. The transformation of iodine to metallic iodine takes place at about 240 kbars.

XI. BORN TREATMENT OF THE PHASE TRANSFORMATIONS IN ALKALI HALIDES

The Born expression³⁶ for the potential energy of an ionic solid as modified by Huggins and Mayer³⁷ is given by:

$$U = (\alpha e^2/r) + (C/r^6 + D/r^8) - B(r) + \phi_0 \quad (25)$$

where U is the lattice energy of the ionic solid per pair of ions, r , the equilibrium interionic distance, ϕ_0 , the zero point energy, C and D are the van der Waals terms for dipole-dipole and dipole quadrupole interactions respectively. Different forms of the repulsive energy $B(r)$ have been employed in the literature (see the chapter on Ionic Solids by C.N.R. Rao). The most general form of the repulsive term is given by the four parameter expression:

$$B(r) = M_1 b_1 \exp(-r/\rho_1) + M_2 b_2 \exp(-ar/\rho_2) \quad (26)$$

in which M_1 and M_2 are the numbers of nearest and next to nearest neighbours, b_1 , b_2 , ρ_1 and ρ_2 are repulsive parameters and a is the ratio of the distance between next-to-nearest and nearest neighbours. Other forms of equation (26) have been employed by (i) setting $\rho_1 = \rho_2$ (three parameters) (ii) by setting $b_1 = b_2$ and $\rho_1 = \rho_2$ (two parameters) or (iii) by ignoring the second term due to the non-nearest neighbour interactions. Equation (25) with a two parameter repulsive term has not been very successful in explaining the relative stabilities of the $Fm\bar{3}m$ and $Pm\bar{3}m$ structures of the alkali halides^{36,38-42}.

In an attempt to explain the greater stability of the $Pm\bar{3}m$ structure of $CsCl$, May³⁸ assumed a higher van der Waals term with a three parameter repulsive term given by:

$$U = (\alpha e^2/r) + K(C/r^6 + D/r^8) - B(r) + \phi_0 \quad (27)$$

where $K = 3.6$ and $B(r)$ is given by equation (26) with $\rho_2 = \rho_1$. Tosi and Fumi³⁹ have objected to the use of higher values of $K > 1$ and have instead employed structure-dependent repulsive parameters to explain the pressure transitions of alkali halides of $Fm\bar{3}m$ structures. The use of three or four-parameter repulsive term in the phase transition studies necessitates employing $K > 1$

in most alkali halides; otherwise b_2 comes out to be negative.

Rao and coworkers⁴²⁻⁴⁴ have examined the pressure transitions of several alkali halides as well as the thermal transformation of CsCl and its solid solutions with RbCl, KCl and CsBr, in some detail, by employing the four parameter repulsive term. They have obtained satisfactory solutions by employing $K > 1$ (equation 27). They find, however, that at a critical value of $K (=K^*)$, b_2 changes sign from negative to positive (i.e., $b_2 = 0$). Rao and coworkers⁴²⁻⁴⁴ could therefore employ $K = K^*$ and the associated repulsive terms as well to obtain satisfactory solutions. The need for increased van der Waals term has been established by Rao and coworkers^{43,45} by employing the new van der Waals coefficients of Hajj⁴⁶ and Lynch^{46a}, to obtain solutions for phase transitions of alkali halides.

XII. A FEW COMMENTS ON THE EXPERIMENTAL TECHNIQUES EMPLOYED TO STUDY PHASE TRANSFORMATIONS

A variety of techniques have been employed for the study of phase transformations in addition to x-ray diffraction which is an essential part of the study of any crystal structure transformation. Once a phase transition is identified in a solid by any one of the techniques, one has to investigate the same, by other methods depending on the particular properties of interest. We recall the various methods listed in sections VII and IX in the study of ferroelectric as well as semiconductor-to-metal transitions. The transformation of KNO_2 provides another interesting example where a variety of methods (Fig. 13) of relevance have been employed to study the transformation⁴⁷. The reversible transition of KNO_2 was established by differential thermal analysis (DTA); x-ray diffraction was utilized to prove that the change was from monoclinic to cubic structure. Based on the crystal structures and also analogy with NaNO_2 ⁴⁸, it was suspected that the transformation may be associated with a change from a ferroelectric to a paraelectric state. Accordingly, the dielectric hysteresis loop disappeared around 40°C (Curie temperature); the dielectric anomaly was also observed at this temperature. A plot of electrical conductivity against temperature indicated a break at 40°C . A study of the temperature-dependent infrared spectrum of KNO_2 showed that the absorbance of the NO_2 deformation frequency (830 cm^{-1}) showed a marked increase in intensity at the Curie temperature⁴⁷; this is similar to the behaviour observed in the ferroelectric-paraelectric transformation of KNO_3 ⁴⁹. The literature has many such examples where several techniques have been employed to examine phase transformation in solids.

X-ray diffraction methods are by far the most important tools for the study of crystal structure transformations. Identification

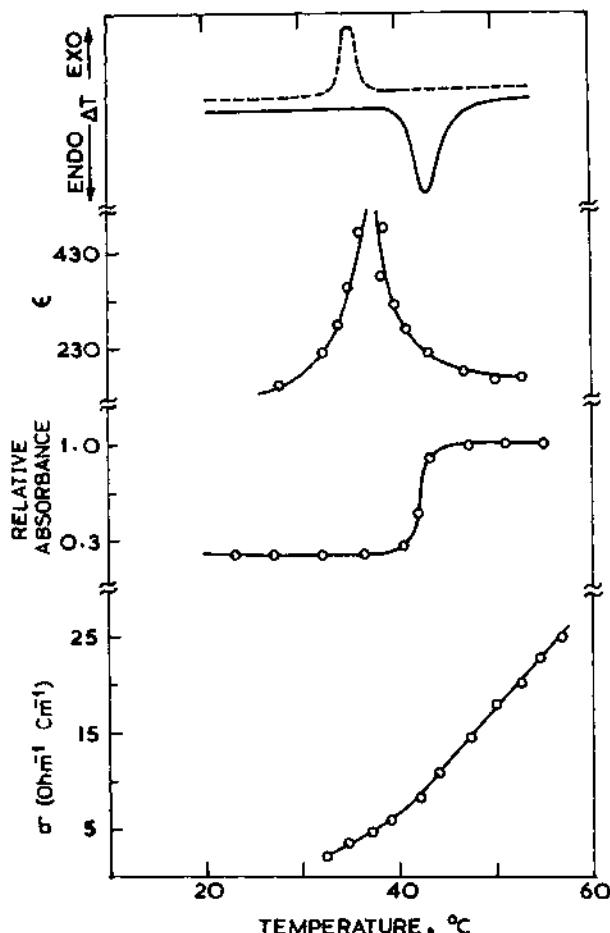


Fig. 13. Investigation of the ferroelectric transition in KNO_2 ⁴⁷ by (a) DTA anomaly; (b) dielectric constant (c) absorption spectra (IR) and (d) electrical conductivity measurements as a function of temperature.

of structures of various phases by x-ray diffraction is very essential irrespective of what other techniques one employs to identify or examine the changes in the system accompanying the transformation. Single crystal work with Laue and Weissenberg photographs are particularly useful for a detailed knowledge of the mechanism of a phase transformation. Thus, Ubbelohde and co-workers^{49a} have examined the co-existence of phases or formation of hybrid crystals during transformations. Change in the intensities of the x-ray lines also furnish valuable information. Single crystal studies enable one to find out the persistence or otherwise of the crystal axis throughout the transformation. Neutron

diffraction has been employed effectively as a tool to examine the positions of light atoms as well as to study magnetic structures. For example, the antiferromagnetism of the semiconducting phase of V_2O_3 or the absence of antiferromagnetism in Ti_2O_3 could be established conclusively by neutron diffraction studies.

The microscope has proved a very valuable tool for studying phase transformations, particularly with respect to movement of boundaries, growth of nuclei and changes in grain size. Many pressure transitions are directly identified by optical methods.

Both electronic and infrared spectroscopy are useful tools provided a characteristic absorption band of the substance shows variations during a transformation. Thus, the phase transitions of KNO_3 , $NANO_3$ ^{49,50} and KNO_2 ⁴⁷ have been examined by infrared spectroscopy⁵¹.

Wide-line NMR has been useful in the study of transformations containing suitable nuclei. Gutowski and coworkers^{52,53} examined the phase transformations in $NaCN$ and $NaHS$ employing NMR spectroscopy. ESR spectroscopy has been employed to study phase transformations by doping a crystal undergoing the transition with a paramagnetic ion.⁵⁴

Measurements of (a.c. or d.c.) conductivity and dielectric constant are important techniques for the study of phase transformations. The transitions are indicated by a break in the conductivity versus temperature curves. For instance, the transformations of $CsCl$ ⁵⁵ ($CsCl$ structure \rightleftharpoons $NaCl$ structure) and rare earth sesquioxides⁵⁶ (cubic \rightarrow hexagonal) have been examined by measuring the electrical conductivity as a function of temperature. Electrical conductivity measurements are particularly useful in the study of order-disorder transformations of alloys and semiconductor-metal transitions. Needless to point out the importance of dielectric constant measurements in the study of transformations associated with ferroelectricity. Many other transformations (e.g. kinetics of anatase-rutile transformation) have been examined by dielectric measurements.

Magnetic transitions (paramagnetic \rightleftharpoons ferromagnetic) or paramagnetic-antiferromagnetic) are generally studied by employing magnetic susceptibility measurements. Changes of magnetic anisotropy as a function of temperature has also been used to study polymorphic changes. Mössbauer spectroscopy containing (or doped with) the appropriate nuclei provides valuable information on such transitions if the spectra are recorded at different temperatures.

Measurements of heat capacities as a function of temperature are valuable in the study of phase transformations (particularly

in the case of second or higher order transitions). Differential thermal analysis is another powerful tool in the study of phase transformations. This technique has been employed to study the enthalpy change, activation energy and thermal hysteresis in transformations^{4,4a}.

In principle, any technique that can measure a property of the substance which undergoes marked change during the transformation can be employed for the study of phase transformations. Dilatometric measurements, change in Young's modulus, thermal expansion of solids are some of the important classical methods for the study of phase transformations.

REFERENCES

1. A.R. Ubbelohde, Quart. Rev., 11, 246 (1957).
2. C.N.R. Rao and K.J. Rao, in 'Progress in Solid State Chemistry', ed. by H. Reiss, Vol. 4 (Pergamon Press, Oxford, 1967).
3. "Phase Transformations in Solids" ed. by R. Smoluchowski, John Wiley & Sons, Inc., New York (1957).
4. K.J. Rao and C.N.R. Rao, J. Materials Sci., 1, 238 (1966).
- 4a. M. Natarajan, A.R. Das, C.N.R. Rao, Trans. Faraday Soc., 65, 3081 (1969).
5. C.N.R. Rao, G.V. Subba Rao and S. Ramdas, J. Phys. Chem., 73, 672 (1969).
6. R.C. Garvey, J. Phys. Chem., 69, 1991 (1965).
7. G.V. Subba Rao, M. Natarajan and C.N.R. Rao, J. Am. Ceram. Soc., 51, 179 (1968).
8. S.R. Yoganarasimhan and C.N.R. Rao, Trans. Faraday Soc., 58, 1579 (1962).
- 8a. C.N.R. Rao, Canadian J. Chem., 39, 498 (1961).
9. M.J. Buerger, in "Phase Transformations in Solids", ed. by R. Smoluchowski, John Wiley, New York, 1957.
10. W.G. Buergers and L.J. Gruen, Discussions of the Faraday Soc., 23, 183 (1957).

11. M. Avrami, J. Chem. Phys., 21, 988 (1951).
- 11a. C.N.R. Rao, S.R. Yoganarasimhan and P.A. Faeth, Trans. Faraday Soc., 57, 504 (1961) and the references cited therein.
12. J.H. de Boer, Discussions of the Faraday Soc., 23, 171 (1957).
13. Y. Kotera and M. Yonemura, Trans. Faraday Soc., 59, 147 (1963).
14. Unpublished results of C.N.R. Rao and coworkers from this laboratory.
15. S. Stecura, U.S. Bur. Mines Report. 6616 (1965).
16. F.C. Nix and W. Shockley, Rev. Modern Phys., 10, 1 (1938).
17. J.W. Christian, "The Theory of Transformations in Metals and Alloys", Pergamon Press, Oxford (1965).
18. L.A. Seigel, J. Chem. Phys., 17, 1146 (1949).
19. F. Jona and G. Shirane, "Ferroelectric Crystals", (Pergamon Press, New York, 1962).
20. J.B. Goodenough, "Magnetism and the Chemical Bond", (Interscience Publ., Inc. and John Wiley Inc., New York, 1963).
- 21a. "International Conference on the Metal-Non-Metal Transitions", San Francisco, California, USA, March, 1968, Publ. in Rev. Modern Physics, 40(4), 677-844 (1968).
 - b. C.N.R. Rao and G.V. Subba Rao, phys. stat. solidi (a), 1, 597 (1970).
22. F.J. Morin, Phys. Rev. Letters, 3, 34 (1959).
23. D. Adler and H. Brooks, Phys. Rev., 155, 826 (1967).
- 23a. K. Kosuge, J. Phys. Chem. Solids, 28, 1618 (1967).
24. J. Feinleib and W. Paul, Phys. Rev., 155, 841 (1967).
25. P.H. Carr and E. Foner, J. Appl. Phys., 31, 345S, (1960).
26. R.M. Moon and coworkers, as communicated by J.M. Honig, December, 1969.
27. E.B. Jones, Phys. Rev., 137, A 978 (1965).

28. T. Shinjo and K. Kosuge, J. Phys. Soc. (Japan), 21, 2622 (1966).
29. C.T. Anderson, J. Am. Chem. Soc., 58, 564 (1936).
30. T. Kawakubo and T. Nakagawa, J. Phys. Soc., Japan, 19, 517 (1964).
31. J. Umeda, H. Kusumoto, K. Narita and E. Yamada, J. Chem. Phys., 42, 1458 (1965).
32. C.N.R. Rao and coworkers (to be published); Also see, K. Kosuge, J. Phys. Soc. Japan, 22, 551 (1967).
33. L. Ladd and W. Paul, Solid State Comm., 1, 425 (1969).
34. S.M. Ariya and G. Grossmann, Soviet Physics, Solid State, 2, 1166 (1960).
- 34a. C.N.R. Rao, M. Natarajan, G.V. Subba Rao, R.E. Loehman, Unpublished results.
- 34b. J.M. Honig, Rev. Mod. Phys., 40, 748 (1968).
- 34c. C.N.R. Rao, R.E. Loehman and J.M. Honig, Phys. Letters, 27A, 271 (1968); J. Phys. Chem., 73, 1781 (1969).
35. "Solids Under Pressure", ed. by W. Paul and D.M. Warchauner, McGraw-Hill, New York (1963).
- 36a. M. Born and K. Huang, "Dynamical Theory of Crystal Lattices", Oxford University Press (1954).
- 36b. M. Born and J.E. Mayer, Z. Phys., 75, 1 (1932).
37. M.L. Huggins and J.E. Mayer, J. Chem. Phys., 1, 637 (1935).
38. A. May, Phys. Rev., 52, 339 (1937); 54, 629 (1938).
39. M.P. Tosi and F.J. Fumi, J. Phys. Chem. Solids, 23, 359 (1962).
40. R.B. Jacobs, Phys. Rev., 54, 468 (1938).
41. F. Hund, Z. Phys., 34, 833 (1925).
42. K.J. Rao, G.V. Subba Rao and C.N.R. Rao, Trans. Faraday Soc., 63, 1013 (1967).
43. M. Natarajan, K.J. Rao and C.N.R. Rao, Trans. Faraday Soc., 66, 0000, (1970).

44. K.J. Rao and C.N.R. Rao, Proc. Phys. Soc., (London) 91, 754 (1967).
45. K.J. Rao, G.V. Subba Rao and C.N.R. Rao, J. Phys. C (Proc. Phys. Soc.), Ser. 2, 1, 1134 (1968).
46. F. Hajj, J. Chem. Phys., 44, 4618 (1966).
- 46a. D.W. Lynch, J. Phys. Chem. Solids., 28, 1941 (1967).
47. K.J. Rao and C.N.R. Rao, Brit. J. Appl. Phys., 17, 1653 (1966).
48. S. Tanisaki, J. Phys. Soc. Japan, 18, 1181 (1963).
49. R.K. Khanna, J. Lingscheid and J.C. Decius, Spectrochim. Acta, 20, 1109 (1964).
- 49a. A.R. Ubbelohde in "Reactivity in Solids", ed. by J.H. de Boer, Elsevier Publishing Co., Amsterdam (1961).
50. R.M. Hexter, Spectrochim. Acta, 10, 281 (1958).
51. C.N.R. Rao, "Chemical Applications of Infrared Spectroscopy" (Academic Press, New York, 1963).
52. C.K. Coogan and H.S. Gutowski, J. Chem. Phys., 40, 3419 (1964).
53. C.K. Coogan, G.G. Belford and H.S. Gutowski, J. Chem. Phys., 39, 3061 (1963).
54. K.N. Shrivastava and P. Venkateswarlu, Proc. Indian Acad. Sci., 64, 275 (1966).
55. I.M. Hoodless and J.M. Morrison, J. Phys. Chem., 66, 557 (1962).
56. P.N. Mehrotra, G.V. Chandrashekhar, E.C. Subba Rao and C.N.R. Rao, Trans. Faraday Soc., 62, 3582 (1966).

SUBJECT INDEX

A

Acceptor levels, 32
Adsorption, 344, 352, 425
Alkali halides
 lattice energies, 21
 phase transitions, 613
Anisotropic conductivity, 385
Antiferroelectric, 155
Antiferromagnetism, 165, 170
Antiphase boundary, 83

B

Band structure, 505
Band theory, 497, 572
Barium azide, 468
 BaTiO_3 , 147, 323, 594, 596, 606
BCS theory, 262
Benzene solid (NMR), 246
Bethe-Peierls-Weiss method, 169
Bloch equations, 237
Bloch function, 499, 503
Block structures, 91, 92
Borate glasses, 405
Born-Haber cycle, 17
Born treatment, 18, 613
Bragg-Williams theory, 353, 602
Brillouin function, 168, 175, 268
Brillouin zone, 9, 508, 572
Burgers circuit, 450

C

CaCO_3 , 595, 596
 $\text{CaF}_2\text{-YF}_3$, 58, 59, 471

Catalysts

classification, 427
electronic properties, 425
metals, 429
semiconductors, 435

CeO_x , 56, 57, 59, 79

Chalcogenides
 of transition metals, 79

Chemical shift, 250

Chemisorption, 426, 430

Clausius-Clepeyron equation,
 590, 592

Clausius-Mosotti equation, 140

CoCr_2O_4 , 181

Cohesive energies, 17

Cole-Cole plot, 143

Collective electrons, 202

Conductivity

 electronic, 520

 four-probe, 524

 ionic, 110, 377

 of glasses, 412

 of hoppers, 541

 of semiconductors, 520

Configurational entropy, 35,
 65, 108

Controlled imperfections, 44

Controlled valence, 45

CoO , 101, 225

Cooper pair, 262

CrO_2 , 554

Crystal growth, 457

Crystal structure, 221

Curie law, 161

Curie point, 153, 165, 168, 175,
 222, 322, 605, 614

Curie-Weiss Law, 153, 169, 171, 205

D
d-character, 430
Debye model, 307
Debye-Hückel theory, 108
Debye T³ law, 13
Debye-Waller factor, 222, 306
Decomposition of
azides, 490
carbonates, 489
solids, 460, 484
Defect cluster, 60
Defect complexes, 54, 57, 70
Defect equilibria, 29, 69
in FeO_x, 61
in KBr, 41
Defect interactions, 121, 455
Defect ordering, 77
Degenerate semiconductors, 515
Deuterium resonance, 245
Diamagnetism, 159
Dielectric hysteresis, 150
Dielectric loss, 140
Dielectric properties, 129
of glasses, 416
Dielectric relaxation, 144
Differential thermal analysis, 595, 616
Diffusion coefficient, 359, 363, 366
Diffusion in glasses, 411
Diffusion in solids, 359
experimental procedures, 367
and reactions, 479
Diffusion profiles, 483
Diketene, 247
Dipolar interaction, 244, 256
Dislocations, 447, 467
and biochemical phenomena, 462
and crystal growth, 457
and dissolution, 558
and evaporation, 458
and precipitation, 459
and solid reactions, 450, 460, 465
elastic fields of, 452

hollow, 456
interaction with point defects, 455
motion of, 455
structure of, 447
Donors, 517
Donor levels, 32
Doppler broadening, 293
Doppler effect, 293
second order, 313, 319
Double exchange, 200
Double resonance, 257
Drift mobility, 521

E
Effective mass, 332, 509
Einstein law of diffusion, 379, 542
Electron paramagnetic resonance, 232, 255
Electronic polarization, 134
ENDOR, 257
Etch pits, 458
Ettinghausen effect, 533
Ettinghausen-Nernst effect, 533
Exhaustion range, 517, 518
Extended defect, 58
Extrinsic semiconduction, 515, 535

F
Fermi contact interaction, 256
Fermi-Dirac integral, 513
Fermi-Dirac statistics, 33, 343, 509, 518
Fermi level, 511, 535
Fermi surface, 267, 332
Fermi temperature, 202, 267
Ferrimagnetism, 173
Ferrites, 280, 607
Ferroelectrics, 147, 151, 245, 320, 594, 605
Ferromagnetism, 168, 169, 270
Ferrous oxide, 61, 101
Fick's law, 359, 371
Force balance method, 115
Fowler-Guggenheim isotherm, 352

- Frenkel defects, 107
 in ionic solids, 126
- Frenkel equilibrium, 36
- Frequency distribution function, 9
- G
- Galvanic cells (solid), 66
- Garnets, 281
- Germanate glasses, 407
- Germanium
 band structure, 517
 semiconductors, 547
- Glasses, 391
 constitution of, 391
 dielectric properties, 416
 transport properties, 411
- Glass formation, 393
 kinetic theory, 400
- Glass forming systems, 403
- Growth of
 crystals, 457
 nuclei, 487
- Gruneisen parameter, 21
- H
- Halide glasses, 407
- Hall effect, 531
- Heat capacity, 13
 of metals, 512
- Helicoidal spin structures, 282
- Hildebrand equation, 20
- Hole, 509, 515
- Hollow dislocations, 456
- Hoppers, 519, 541
- Hydrogen bonding (NMR), 245
- I
- Impurity absorption, 582
- Induced valence, 45
- Inorganic glasses, 391
- Insulator, 515, 560
- Interband transitions, 576
- Interstitials, 29, 125, 378
- Intraband transitions, 576
- Intrinsic semiconductors, 515, 535
- Inverse mixed catalysts, 443
- Ionic conductivity, 110, 377
 anisotropy in, 385
 jump mechanisms, 378
 of glasses, 412
- Ionic polarization, 134
- Ionic solids
- Born treatment, 18, 613
 cohesive energies, 17
 conductivity of, 110, 377
 point defects, 107
 positron annihilation, 337
- J
- Jahn-Teller distortion, 191
- Joule effect, 273
- K
- Kinetics
 of phase transformations, 598
 of solid reactions, 465
- Kirkendall effect, 364, 482
- Knight shift, 248, 261, 267
- KNO_2 , 614
- KNO_3 , 595, 614
- Koch cluster, 74
- Kondo effect, 267
- Kroger-Penney model, 499
- Kroger-Vink diagram, 40
- Kroger-Vink treatment, 36
- KTaO_3 , 556
- L
- Lamb-Mössbauer factor, 308, 316
- Lambda transition, 593
- Lande g-factor, 232
- Langevin function, 160
- Langmuir isotherm, 344
- Larmour frequency, 233
- Lattice dynamics, 1
- Lattice energies, 17, 21
- Lattice symmetry, 4
- Layer growth, 473
- Ligand fields, 185
- Line phases, 31, 41, 51
- Line shape function, 239

- Long range order, 50, 168, 229, 353, 602
 Lorentz force, 531
 Lorenz-Lorentz equation, 140
 Lorenz number, 526
 Luttinger-Tisza method, 179
- M
 Madelung constant, 19
 Magnetic anisotropy, 276
 Magnetic form factor, 224
 Magnetic resonance, 231
 experimental arrangement, 240
 spectrometer, 242
 Magnetic transitions, 607, 616
 Magnetism, 159
 Magnetocrystalline anisotropy, 194
 Magnetogyric ratio, 232
 Magnetostriction, 273
 in ferrites, 280
 in garnets, 281
 in metal systems, 274
 Martensite transformations, 605
 Matano surface, 372
 Maxwell's equations, 568
 Metals
 as catalysts, 429
 oxidation of, 472
 spectroscopy, 573
 transport properties, 497, 515
 Mie-Gruneisen equation, 21
 MnCr_2O_4 , 180
 MoO_x , 84, 86
 Mosotti catastrophe, 153
 Mössbauer effect, 287, 616
 applications of, 316
 in alloys, 250
 mechanisms of, 301
 Mössbauer's experiment, 299
 Mott-Littleton method, 115
- N
 NaCN , 605
 NaNO_2 , 614
 NbO , 554, 580
 NbO_x , 86, 95
 $\text{Nb}_2\text{O}_5\text{-TiO}_2$, 95
- $\text{Nb}_2\text{O}_5\text{-WO}_3$, 91, 94
 Neel temperature, 171, 222
 Nernst effect, 535
 Nernst-Einstein-Bardeen relation, 379
 Neumann's law, 385
 Neutron diffraction, 219
 Neutron scattering, 13, 221, 226
 NiO , 224, 476, 483
 Nitrate glasses, 407
 NMR spectroscopy, 231
 chemical shift, 250
 line shape, 239, 242
 of alloys, 266
 of metals, 248
 of non-metallic solids, 242
 of superconductors, 261
 Nonstoichiometry, 30, 46
 gross, 51
 Nuclear moments, 231
 Nuclear quadrupole moment, 232
 Nuclear quadrupole resonance, 232, 253
 Nucleation, 485, 598
- O
 Onsager reciprocity relations, 385, 527
 Orbital angular momentum
 quenching of, 190
 Order, 49
 Order-disorder theory, 343, 355
 applications, 346
 Order-disorder transformation, 601
 Order parameter, 593, 604
 Orientational polarization, 134
- P
 Paramagnetism, 160
 Particle-in-a-box, 497
 Pauli paramagnetism, 165, 205
 Perovskites, 145, 209
 Phase transformations, 589
 Born treatment, 613
 high pressure, 592, 611
 kinetics, 598

- thermal hysteresis, 594
thermodynamics, 589
shock induced, 612
Phonon annihilation, 226
Phonon creation, 226
Phosphate glasses, 405
Photolysis of
 barium azide, 468
 silver halides, 455
Physical adsorption, 426
Plasma edge, 554
Plasma oscillations, 577
Plastic flow, 448
 $p-n$ junction, 374
Point defects, 29, 107, 465
 equilibria, 29
 experimental studies, 109
 interaction energies, 121
 interaction with dislocations, 455
 migration energies, 124
 notation, 31
 random, 52
 role in reactions, 465
 theory of, 107
Polarization, 130, 133
Positive holes, 32, 509, 515
Positron annihilation, 329
 experimental methods, 330, 335
 in ionic solids, 331
 in metals, 337
 in molecular media, 338
 in semiconductors, 337
Positronium, 329
 chemistry, 338
Precipitation, 459
 PrO_x , 57, 59, 79, 557
Pyrochlores, 59, 62
- Q
- Quadrupole moment, 242, 252
Quadrupole resonance, 252
Quadrupole splittings, 247
Quartz (α - β), 594
- R
- Radius ratio criterion, 393
- Random point defects, 52
Reactions in solids, 65, 455
Reactivity of solids, 466
Reciprocal lattice, 9
Reflectance spectra, 478, 570, 577
Relaxation dielectrics, 145
 ReO_3 , 549, 578
 ReO_3 structure, 85, 86, 211
Resonance absorption
 of gamma rays, 304
Resonance fluorescence, 267
Righi-Leduc effect, 535
 RuO_2 , 548
- S
- Schottky defects, 107
 in ionic solids, 112, 119
Schottky equilibrium, 35, 40
Screw dislocation, 450
Seebeck coefficient, 528, 531
Self-diffusion coefficient, 366
Semiconductors, 515
 catalysts, 435
 compensated, 518
 extrinsic, 518
 degenerate, 515
 intrinsic, 515
 positron annihilation, 337
 semimetal transitions, 506, 607
Semimetal, 519
Shear plane, 83
Shear structures, 82
Shear vector, 83
Short range order, 50, 604
Silicate glasses, 403
Sintering, 469
 SiO_2 , 594, 597
Solid decompositions, 460, 484
Solid-gas reactions, 472
Solid-solid reactions, 477
Solid solutions, 30, 482
Solid state reactions, 465
Space charge polarization, 136
Specific heats of crystals, 13
Spectroscopy
 of metals, 573, 577
 of semiconductors, 581

- Spinels, 177, 179, 199, 327, 480
 Spin coupling, 250
 Spin flip, 220
 Spin-lattice relaxation, 234
 Spin-orbit coupling, 193
 Spin structure, 221
 Spin susceptibility, 262, 265, 266
 Spin waves, 170
 Spiral magnetism, 177
 Spontaneous polarization, 151
 SrTiO_3 , 556, 594
 Superconductors, 261
 Superexchange interactions, 195
 Superstructure, 50, 77
 Surface defects, 467
 Surface energies, 26, 456
- Tarnishing reactions, 474
 Tellurite glasses, 407
 Thermal conductivity, 525
 Thermal decompositions, 484
 Thermal neutrons, 220
 Thermodynamics
 of Crystals, 11
 of point defects, 29
 of FeO , 61
 Thermoelectric properties, 526, 545
 $\text{ThO}_2\text{-Y}_2\text{O}_3$, 65
 TiC_x , 52
 Tight-binding approximation, 505
 Tin transformation, 601
 TiO , 101, 126, 206, 580
 Ti_2O_3 , 563, 607, 616
 TiO_2 , 98, 471, 595, 597
 TiO_x , 86, 98, 554
 Transport theories, 497
 T-T-T diagram, 402
- Tungsten bronze, 52, 99
- U
 UO_x , 56, 57, 58
- V
 Vacancies, 29
 Vacancy mechanism, 378
 Vacancy pairs, 121
 Vanadate glasses, 407
 van der Waals forces, 22
 Van Hove singularities, 10
 Van Vleck paramagnetism, 165
 Vibrational thermodynamics, 11
 Vibrational states, 8
 VC_x , 52
 VO , 101, 126, 206, 561, 580
 VO_2 , 560, 607
 V_2O_3 , 560, 607, 616
- W
 Weiss molecular field, 165
 Wiedemann-Franz law, 526
 Willis cluster, 58
 WO_x , 85, 86
 Wüstite, 61, 75
- X
 X-ray scattering, 13
- Y
 Yafet-Kittel model, 177
- Z
 Zachariasen-Warren model, 394
 Zeeman energy levels, 232, 252
 ZrO_2 , 596
 $\text{ZrO}_2\text{-CaO}$, 65