EDITORIAL DEPARTMENT

Alfred N. Goldsmith, *Editor Emeritus*

John R. Pierce, *Editor*

E. K Gannett, *Managing Editor*

Marita D. Sands, *Production Manager*

•

EDITORIAL BOARD

John R. Pierce, *Chairman*
D. G. Fink
E. K. Gannett
T. A. Hunter
W. R. Hewlett
J. A. Stratton
W. N. Tuttle

•

George W. Bailey, *Executive Secretary*

•

John B. Buckley, *Chief Accountant*

Laurence G. Cumming, *Technical Secretary*

Evelyn Davis, *Assistant to the Executive Secretary*

Emily Sirjane, *Office Manager*

Responsibility for the contents of papers published in the PROCEEDINGS of the IRE rests upon the authors. Statements made in papers are not binding on the IRE or its members.

PROCEEDINGS OF THE IRE

Published Monthly by

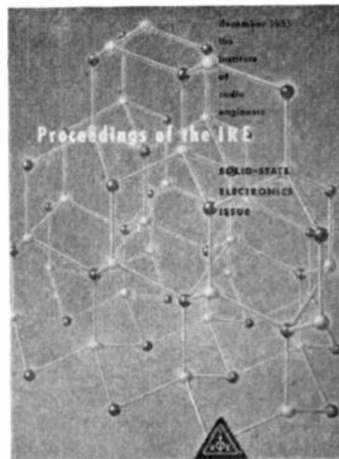
The Institute of Radio Engineers, Inc.

VOLUME 43 — *December, 1955* — NUMBER 12

CONTENTS

Solid-State Electronics . . . *The Editor* 1701
5573. The Electronic Energy Band Structure of Silicon and Germanium . . *Frank Herman* 1703
5574. Nonlinear Dielectric Materials . . . *E. T. Jaynes* 1733
5575. Some Aspects of Ferroelectricity . . . *G. Shirane, F. Jona, and R. Pepinsky* 1738
5576. History of Semiconductor Research . . . *G. L. Pearson and W. H. Brattain* 1794
5577. Junction Transistor Electronics . . . *J. L. Moll* 1807
5578. Photoconduction in Germanium and Silicon . . . *M. L. Schultz and G. A. Morton* 1819
5579. Photoeffects in Intermetallic Compounds . . . *H. P. R. Frederikse and R. F. Blunt* 1828
5580. Photoconductivity of the Sulfide, Selenide, and Telluride of Zinc or Cadmium . . . *R. H. Bube* 1836
5581. Performance of Photoconductors . . . *Albert Rose* 1850
5582. Lead Salt Photoconductors . . . *T. S. Moss* 1869
5583. Theory and Experiments on a Basic Element of a Storage Light Amplifier . . . *Jenny Rosenthal* 1882
5584. An Electroluminiscent Light-Amplifiying Picture Panel . . . *B. Kazan and F. H. Nicoll* 1888
5585. Opto-Electronic Devices and Networks . . . *E. E. Loebner* 1897
5586. Cathodoluminescence . . . *G. F. J. Garlick* 1907
5587. Electroluminescence and Related Topics . . . *G. Destriau and H. F. Ivey* 1911
5588. The Physical Chemistry of Crystal Phosphors . . . *F. A. Kröger* 1941
5589. Some Properties of Ferrites in Connection with Their Chemistry . . . *E. W. Gorter* 1945

Contents continued on following page



THE COVER—The complex arrangement of atoms on this month's cover forms the hexagonal-type crystal structure of cadmium sulfide. Because its conductivity increases when it is exposed to light, cadmium sulfide promises to become an increasingly important material to electronic engineers, offering multiple potential uses in detection and control equipment, electrophotography, light amplifiers, solar generators and other photoelectronic devices.

Cadmium sulfide is just one of a number of solid-state materials discussed in this special issue of PROCEEDINGS which are rapidly finding important new uses in the field of Radio Engineering. The cover portrayal of the atomic skeleton of this substance correctly suggests that this issue is primarily concerned with the internal mechanisms which produce useful electronic properties in solids.

Copyright, 1955, by the Institute of Radio Engineers, Inc.

EDITORIAL DEPARTMENT

Alfred N. Goldsmith, *Editor Emeritus*

John R. Pierce, *Editor*

E. K Gannett, *Managing Editor*

Marita D. Sands, *Production Manager*

EDITORIAL BOARD

John R. Pierce, *Chairman*
D. G. Fink
E. K. Gannett
T. A. Hunter
W. R. Hewlett
J. A. Stratton
W. N. Tuttle

George W. Bailey, *Executive Secretary*

John B. Buckley, *Chief Accountant*

Laurence G. Cumming, *Technical Secretary*

Evelyn Davis, *Assistant to the Executive Secretary*

Emily Sirjane, *Office Manager*

Responsibility for the contents of papers published in the PROCEEDINGS of the IRE rests upon the authors. Statements made in papers are not binding on the IRE or its members.

PROCEEDINGS OF THE IRE

Published Monthly by

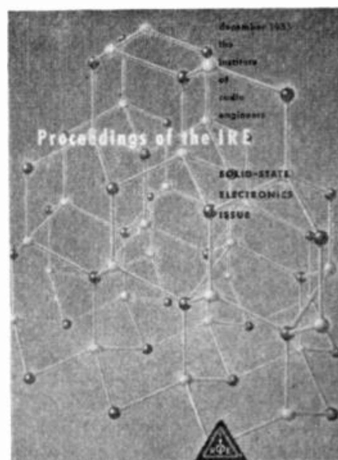
The Institute of Radio Engineers, Inc.

VOLUME 43 — *December, 1955* — NUMBER 12

CONTENTS

Solid-State Electronics ... *The Editor* 1701
5573. The Electronic Energy Band Structure of Silicon and Germanium ... *Frank Herman* 1703
5574. Nonlinear Dielectric Materials ... *E. T. Jaynes* 1733
5575. Some Aspects of Ferroelectricity ... *G. Shirane, F. Jona, and R. Pepinsky* 1738
5576. History of Semiconductor Research ... *G. L. Pearson and W. H. Brattain* 1794
5577. Junction Transistor Electronics ... *J. L. Moll* 1807
5578. Photoconduction in Germanium and Silicon ... *M. L. Schultz and G. A. Morton* 1819
5579. Photoeffects in Intermetallic Compounds ... *H. P. R. Frederikse and R. F. Blunt* 1828
5580. Photoconductivity of the Sulfide, Selenide, and Telluride of Zinc or Cadmium ... *R. H. Bube* 1836
5581. Performance of Photoconductors ... *Albert Rose* 1850
5582. Lead Salt Photoconductors ... *T. S. Moss* 1869
5583. Theory and Experiments on a Basic Element of a Storage Light Amplifier ... *Jenny Rosenthal* 1882
5584. An Electroluminiscent Light-Amplifiying Picture Panel ... *B. Kazan and F. H. Nicoll* 1888
5585. Opto-Electronic Devices and Networks ... *E. E. Loebner* 1897
5586. Cathodoluminescence ... *G. F. J. Garlick* 1907
5587. Electroluminescence and Related Topics ... *G. Destriau and H. F. Ivey* 1911
5588. The Physical Chemistry of Crystal Phosphors ... *F. A. Kröger* 1941
5589. Some Properties of Ferrites in Connection with Their Chemistry ... *E. W. Gorter* 1945

Contents continued on following page



THE COVER—The complex arrangement of atoms on this month's cover forms the hexagonal-type crystal structure of cadmium sulfide. Because its conductivity increases when it is exposed to light, cadmium sulfide promises to become an increasingly important material to electronic engineers, offering multiple potential uses in detection and control equipment, electrophotography, light amplifiers, solar generators and other photoelectronic devices.

Cadmium sulfide is just one of a number of solid-state materials discussed in this special issue of PROCEEDINGS which are rapidly finding important new uses in the field of Radio Engineering. The cover portrayal of the atomic skeleton of this substance correctly suggests that this issue is primarily concerned with the internal mechanisms which produce useful electronic properties in solids.

Copyright, 1955, by the Institute of Radio Engineers, Inc.

BOARD OF DIRECTORS, 1955

J. D. Ryder, *President*
Franz Tank, *Vice-President*
W. R. G. Baker, *Treasurer*
Haradan Pratt, *Secretary*
John R. Pierce, *Editor*
J. W. McRae, *Senior Past President*
W. R. Hewlett, *Junior Past President*

●

1955
S. L. Bailey
A. N. Goldsmith
A. V. Loughren
C. J. Marshall (R5)
L. E. Packard (R1)
J. M. Pettit (R7)
B. E. Shackelford
C. H. Vollum
H. W. Wells (R3)

●

1955-1956
E. M. Boone (R4)
J. N. Dyer (R2)
J. T. Henderson (R8)
A. G. Jensen
George Rappaport
D. J. Tucker (R6)

●

1955-1957
J. F. Byrne
Ernst Weber

●

ADVERTISING DEPARTMENT

William C. Copp,
Advertising Manager

Lillian Petranek
Assistant Advertising Manager

●

Change of address (with 15 days advance notice) and letters regarding subscriptions and payments should be mailed to the Secretary of the IRE, 1 East 79 Street, New York 21, N. Y.
All rights of publication, including foreign language translations are reserved by the IRE. Abstracts of papers with mention of their source may be printed. Requests for republication should be addressed to The Institute of Radio Engineers.

PROCEEDINGS OF THE IRE

Published Monthly by

The Institute of Radio Engineers, Inc.

(*Continued*)

Correspondence:

5590. Noise Factor Measurement . . . *A. C. Hudson* 1974
5591. Germanium Transistor Amplifiers Stable to 95°C. . . *W. Greatbatch and W. Hertreiter* 1974
5592. Power Flow . . . *W. M. Gottschalk* 1974
5593. Rebuttal . . . *J. R. Pierce* 1975
5594. Nonuniform Transmission Lines as Impedance-Matching Sections . . . *J. Willis and N. K. Sinha* 1975
Contributors . . . 1976

IRE News and Radio Notes:

Calendar of Coming Events . . . 1980
Operations Research Subject of December Symposium . . . 1980
IRE Southwestern Conference Scheduled for February . . . 1980
Vehicular Communications Group Holds Conference . . . 1981
PIB Sponsors Nonlinear Circuit Analysis Symposium in April . . . 1981
M. J. Kelley and Sir Gordon Radley Receive First Columbus Communication Prize . . . 1982
R. H. Ranger to Be President of Audio Engineering Society . . . 1983
National Service Foundation Announces Fellowship Program . . . 1983
5th Fall Symposium of PG on Broadcast Transmission Systems . . . 1984
Professional Group News . . . 1984
Technical Committee Notes . . . 1984

Books:

5595. Recent Books . . . 1985
5596. "Électronique Industrielle," by G. Goudet . . . *Reviewed by A. G. Clavier* 1985
5597. Abstracts and References . . . 1986
Annual Index to PROCEEDINGS OF THE IRE . . . Follows Page 2000

ADVERTISING SECTION

Meetings with Exhibits . . . 6A
News—New Products . . . 14A
IRE People . . . 20A
Industrial Engineering Notes . . . 74A
Section Meetings . . . 84A
Professional Group Meetings . . . 92A
Membership . . . 102A
Positions Wanted . . . 142A
Positions Open . . . 152A
Advertising Index . . . 189A

Solid-State Electronics

A new child in the radio engineering family has grown to astonishing stature during the last few years. Its personality is so radically new and its future is so promising that there can be little doubt that we are witnessing the early stages of a new era in radio-electronics, the solid-state electronics era.

"Solid-state electronics" may be described broadly as dealing with the control and utilization of the electric, magnetic, and photic properties of solids. The knowledge that special effects can be produced in some materials when energized by electric or magnetic fields or by light is not new. Early investigations of these properties by scientists date back to the last century. Nor is the application of solid-state materials in radio new. One can go back over the years and find several examples, notably in detectors, rectifiers, resonators, and transducers.

It is only recently, however, that we have begun to make real headway in understanding and utilizing solid-state materials on a large scale. We now see ferromagnetic and ferroelectric substances employed for memory cells, ferrites for microwave attenuators, photoconductors for automatic headlight dimmers, and electroluminescent materials for experimental light amplifiers, to name but a few.

The most startling progress to date has been in the field of semiconductors. The development of the transistor in 1948 was a major milestone in electronic progress and already this subject alone has become a major field of endeavor. And yet the transistor represents only one of several species of phenomena inherent to germanium (or silicon), germanium is only one genus of semiconductor material, and semiconductors are only one order in the family of solid-state materials. There remain many other species, genera, and orders to be further explored and utilized and many other devices to be developed.

Most of our fundamental knowledge about solid-state materials has been developed by physicists, chemists, and metallurgists, and much of this information has not yet filtered across to the engineer. In order for the radio engineer to use these materials in the creation of new devices he must first have some understanding of their fundamental properties. It is to this end that the IRE is devoting this special issue of its PROCEEDINGS to Solid-State Electronics—to bridge the gap between the scientist and the engineer.

CONTENTS OF THIS ISSUE

In this issue leading authorities in the field have been invited to discuss the principal classes of solid-state materials, their properties, and their applications. These discussions review prior work in the field, outline in tutorial fashion our present understanding of the subject, and indicate in what direction future progress lies. Emphasis is given to those topics about which little has heretofore appeared in the engineering literature, especially to those materials which can generate, store, or are actuated by light.

The issue starts with a tutorial introduction by Frank Herman to a subject which is basic to an understanding of solids; namely, crystals. The reader will find that this discussion of crystal geometry, energy band structure, and electrical and optical properties will provide him with an excellent vantage point from which to view with greater clarity all of the papers that follow.

Ferroelectric materials are of considerable interest today because of their applications as electromechanical transducers and computer memory elements and because they exhibit useful dielectric properties which are quite similar to the magnetic properties of ferromagnetic materials. The next paper (Jaynes) briefly explains the physics of nonlinear dielectrics and is followed by a detailed discussion of ferroelectric crystals and their dielectric behavior (Shirane, Jona, and Pepinsky).

The next group of papers deals with what is thus far the most important class of solid-state materials—semiconductors. The history of semiconductor research (Pearson and Brattain) is not only a very interesting story in itself, but provides some valuable object lessons on how slow we are to accept new ideas and give up old ones. The most useful semiconductor device to be developed so far is the junction transistor. Thanks to intensive research we now have an almost complete understanding of how it works and are making steady progress in improving its performance (Moll). The widespread use of germanium and silicon for transsistors and rectifiers has tended to obscure the fact that these materials are also excellent photoconductors (Schultz and Morton) and are already being put to good use as such.

The subject of photoconductors leads us into the general area of photoelectronic phenomena, an area which is very new and very promising. The recent knowledge gained from work on silicon and germanium has led to a much better understanding of photo-effects in intermetallic semiconductors (Frederikse and Blunt), has greatly stimulated research into the use of cadmium and zinc compounds for photoconductors (Bube), and in general has provided considerable insight into the basic mechanism by which the conductivity of a material is increased by exposure to light (Rose). Important strides have recently been made also in the development of lead salt photoconductors for use in infrared detectors (Moss).

As a result of the above advances in photoconductivity, together with the very recent and rapid progress in the development of electroluminescent materials, a great deal of attention is now being given to light amplifiers and related devices. Several successful experimental models have been announced just within the past year and predictions are now frequently heard of the many new wonders which are waiting around the corner, such as picture-on-the-wall television. Among the devices currently being developed is a storage light amplifier (Rosenthal). Once an image is projected onto its screen, the projector may be turned off and the screen will store and continue to reradiate the image, but at an intensity no greater than the projected image. This is not to be confused with a true light amplifier (Kazan and Nicoll) which reradiates a greatly intensified image, but only as long as the projector is on. Numerous other uses of photoconductive and electroluminescent cells are being investigated, including color converting screens, counting and switching devices, and logical networks for computers (Loebner).

It can be seen from the above that not only is there great practical interest in photoconductive or light-actuated materials, but also in luminescent or light-generating substances. Recent advances in our knowledge of cathodoluminescence, or electron-excited luminescence (Garlick), are having an important bearing on the study of cathode-ray-tube screen performance and efficiency. Another important type of luminescence can be obtained in some materials by application of an electric field. A thorough understanding of this phenomenon, called electroluminescence, will be of interest and importance to all engineers because of the host of applications it suggests (Destriau and Ivey). Our study of luminescence would not be complete without a survey of present views regarding the constitution and preparation of phosphors (Kroger).

The next paper (Gorter) deals with a class of material which in a relatively short time has found widespread application throughout the radio engineering field, both as perhaps the most popular form of storage element in computers and in providing for the first time practical nonreciprocal devices for the microwave art. This discussion of ferrites, their chemistry, how they are prepared, the history of their development, and their many applications closes the issue.

ACKNOWLEDGMENT

We wish to express our deep appreciation to Lloyd T. DeVore, general manager of the Stewart-Warner Electric Division and former manager of the General Electric Co. electronics laboratory, for accepting and discharging so well the heavy responsibility of planning the contents and procuring the material for this entire issue. He has succeeded in bringing together an outstanding group of papers that bring into engineering focus a many-faceted field of science which will profoundly affect the future of radio engineering.—*The Editor*

The Electronic Energy Band Structure of Silicon and Germanium*

FRANK HERMAN,† ASSOCIATE MEMBER, IRE

Summary—This article, which is of a tutorial character, is concerned with three broad subjects: (*a*) the theory of electronic energy bands in a perfect crystal; (*b*) the electronic energy band structure of silicon and germanium crystals; and (*c*) the relationship between some of the electrical and optical properties of these crystals and their energy band schemes. The article is essentially an introduction to the quantum theory of crystals, with silicon and germanium serving as illustrative examples. As such, the article should appeal particularly to electronic engineers and physicists working in the field of solid state electronics. A knowledge of quantum mechanics is probably not essential to the understanding of major portions of this paper.

The following topics are treated: crystal symmetry and crystal geometry; electronic quantum states in a perfect crystal; the energy band scheme; occupancy of the electronic quantum states; the hole concept; the effective mass tensor; velocity and acceleration of electrons and holes; the spin-orbit interaction and its consequences; the energy band structures of silicon and germanium; the band structure of germanium-silicon alloys; theory of lattice vibrations; the phonon concept; collisions between electrons or holes and phonons; the electrical conductivity; optical absorption and emission processes.

I. Introduction

SOLIDS CAN usually be classified as crystalline or noncrystalline. Both types of solids are formed from atoms which are held together by electrical forces. In a crystalline solid, or a crystal, the atoms are arranged in space in the form of a regular array. Thus, a crystal is a periodic structure. If an atomic aggregate does not exhibit a distinctive periodic structure, it is classed as a noncrystalline solid. In this paper, we will be interested in crystalline, rather than in noncrystalline solids.

Crystals may be classified as perfect crystals, nearly perfect crystals, and disordered crystals. In a perfect crystal, the atomic array is flawless. Each position in the spatial array is occupied by the proper type of atom, and there are no atoms present which do not belong to the array. A model of a perfect diamond-type crystal is shown in Fig. 1. [The diamond-type crystals include grey tin, germanium, silicon, and diamond itself.] Each ball in the model represents an atomic position, and each rod a chemical bond joining two atoms together. The normal atomic positions are called substitutional sites, while the positions not belonging to the array which an atom might occupy under special conditions are interstitial sites. In a perfect germanium crystal, for example, each substitutional site is occupied by a germanium atom, and each interstitial site is vacant.

In a nearly perfect crystal, the atomic array contains a small number of flaws or structural imperfections. If a small fraction of the substitutional sites in a germanium crystal are occupied by atoms other than germanium, or if some of the interstitial sites are occupied, the crystal would be called a nearly perfect crystal. Similarly, if a small fraction of the substitutional sites were vacant, or if the atomic array were not in proper registry in a few localized regions, the crystal would be a nearly perfect crystal.

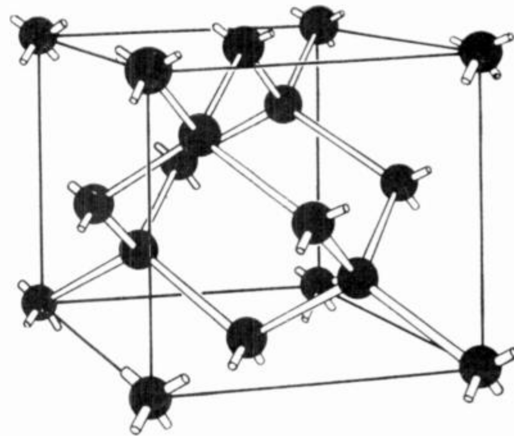


Fig. 1—Model of the atomic arrangement in diamond-type crystals. (After Shockley.)

A disordered crystal may be described in terms of an example. Consider a perfect crystal composed of two types of atoms. In such a crystal, the atoms of each type are arranged in the form of a regular array. Some of the substitutional sites are occupied by the atoms of one type, and the remaining sites by the atoms of the other type. If by some means we spoil the ordering, such that the atoms of both types are arranged at random among the substitutional sites appropriate to the perfect crystal, a disordered crystal results.

It is common knowledge that the crystals which occur in nature and those grown in the laboratory are never perfect. Such crystals usually contain a variety of structural imperfections. The very best natural and synthetic crystals are nearly perfect crystals. Some types of disordered crystals can be grown in the laboratory, while other types can be produced from nearly perfect crystals by suitable processing. Many of the crystals used in solid state devices such as transistors are nearly perfect crystals.

From a theoretical standpoint, perfect crystals are easier to treat than nearly perfect crystals; nearly perfect crystals are easier to treat than disordered crystals; and disordered crystals are easier to treat than noncrystalline solids. The theory of noncrystalline solids

* Original manuscript received by the IRE, September 19, 1955.
† RCA Labs. Princeton, N. J.

has not advanced very far due to mathematical and physical difficulties. Some progress has been made in dealing with disordered crystals theoretically. The theory of perfect and nearly perfect crystals is fairly well developed.

The usual practice is to treat a nearly perfect crystal as a perfect crystal in a first approximation, and then to take the structural imperfections into account in a second approximation. When a nearly perfect crystal is studied in this manner, it is found that some of the physical properties of the crystal are relatively insensitive to the structural imperfections, while others are highly sensitive. Many of the electrical and optical properties of nonmetallic crystals, *i.e.*, semiconductors and insulators, are extremely sensitive to the nature and the concentration of the structural imperfections. It is for this reason that imperfections are often introduced intentionally into synthetic nonmetallic crystals in order to give these crystals desirable electrical or optical properties. In contrast, the thermal and elastic properties of nonmetallic crystals are relatively insensitive to the presence of small numbers of imperfections.

In this paper, we will confine ourselves to the theory of perfect nonmetallic crystals. [The only exception to this rule occurs in Section XI, where we consider germanium-silicon alloys, which are essentially disordered crystals.] We have imposed this restriction in order to keep the length of the manuscript within reasonable bounds. We direct the reader's attention to the article by Shockley [1] and to the references given at the end of this paper for information concerning the effect of imperfections on the physical behavior of crystals, and the theory of crystals in general. Extensive bibliographies can be found in most of the terminal references.

Broadly speaking, the theory of crystals is concerned with the relationship between the physical and chemical properties of a crystal and the individual and collective behavior of the electrons and nuclei belonging to the crystal. While many problems in crystal physics can be handled by classical methods, some problems can be treated only within the framework of quantum mechanics. It is with the latter class of problems that we are concerned in this paper.

According to quantum theory, an electron in a crystal can exist in only a limited number of discrete states. The allowed states are called the electronic quantum states. When an electron occupies one of these quantum states, it has an energy which is characteristic of the state. Thus, the energy of an electron in a crystal is limited to a definite set of values, namely, the energy values corresponding to the allowed states. In a periodic array of atoms, the quantum states tend to cluster into nearly continuous groups of allowed energy levels called energy bands. The existence of energy bands greatly simplifies the description of the distribution of the allowed states in crystals.

The major subject of this paper is the energy band scheme, which determines in part the behavior of the electrons in a crystal. The energy band scheme has been widely treated in the literature [see the general bibliography]. Some of the accounts are rather elementary, while others are intended only for the specialist. Our version is meant to be intermediate between the more elementary and the more advanced treatments.

Why is there an interest in the energy band scheme of a crystal? Many of the physical properties of a crystal depend upon the nature of the available electronic quantum states, upon the manner in which the electrons in the crystal are distributed among these quantum states, and upon the readiness with which electrons make transitions between different quantum states under the action of an applied force. If enough is known about the energy band scheme of a crystal, one can interpret, analyze, and predict many of its physical characteristics, especially those depending upon the electronic behavior. In brief, one can gain a good deal of insight into the physical properties of a crystal from a knowledge of its energy band scheme.

Much current research is devoted to the elucidation of the energy band schemes of crystals having scientific or technological significance. The information gained from such research can prove very valuable in many applications, as the literature will testify.

In the early portions of this paper, we will consider the theory of electronic quantum states in perfect crystals from a general point of view, though we favor nonmetallic crystals over metallic crystals in many discussions. The emphasis was placed on nonmetallic crystals because such crystals are commonly employed in solid state devices such as rectifiers, transistors, photocells, etc., and thus are worthy of special attention. In the later portions, we are concerned almost exclusively with the energy band schemes for the silicon and germanium crystals. Throughout the paper, geometrical ideas are illustrated by the case of diamond-type crystals. The necessary background material is provided in those discussions where it is likely that the general reader will benefit from the inclusion of such material.

Since we are attempting to cover a rather broad area, even with the many restrictions we have imposed on the content of this paper, our treatment is rather superficial in many places. We have not given the proofs of most of the statements which are made in the course of the presentation. It may be taken for granted that the proofs are to be found in the references cited at the end of the paper.

We have chosen the cases of silicon and germanium to illustrate the general theory for the following reasons: First, these crystals are the basic constituents of many solid state devices. An understanding of their properties is of great practical importance. Second, silicon and germanium have been carefully investigated, and more is known about the energy band schemes of these crystals than is known about the band schemes of other crystals. Third, the energy band schemes of silicon and germanium contain many of the features which are likely to be

found in the band schemes of other crystals. The better we understand the band schemes of these two crystals, and the relationship between the band schemes and the physical properties, the easier it will be to understand crystals in general.

The organization of this paper is as follows: In Section II, we discuss the symmetry properties of crystals, and certain geometrical concepts related to these symmetry properties. The theory of electronic quantum states in crystals, and the theory of energy bands are developed in Section III. The manner in which the electrons are distributed among the available quantum states under equilibrium conditions is described in Section IV. We also introduce the hole concept in Section IV. The structure of an energy band near its extrema is considered in some detail in Section V. We show in Section VI how the velocity and the acceleration of an electron or a hole can be determined from a knowledge of the energy band structure. The effect of the spin-orbit interaction on the energy band structure of crystals is considered in Section VII.

Having outlined the foundations of the subject, we turn to the band structure of silicon and germanium. Recent theoretical and experimental developments bearing on the energy band structure of these two crystals are reviewed in Section VIII. We summarize the current state of our knowledge concerning the valence and conduction band structures of silicon and germanium in Sections IX and X. The band structure of germanium-silicon alloys is considered in Section XI.

The remaining sections are concerned with the physical implications of what has come before, and with additional background material. The theory of lattice vibrations and the phonon concept are treated briefly in Section XII. In Section XIII we discuss the collisions between electrons or holes and phonons. The relationship between some of the electrical and optical properties of silicon and germanium and the band structures of these crystals is considered in Sections XIV and XV. The significance of the recent developments in the theory of silicon and germanium is summarized in the final section.

II. Geometrical Preliminaries

The present section is devoted to a brief discussion of crystal geometry and crystal symmetry. Crystal geometry provides a means for classifying crystals and the symmetry properties determine the coordinate systems in which mathematical descriptions of the properties of a crystal will be tractable. For the purposes of this section, we will assume that the crystal under consideration is a perfect crystal of infinite extent.

Let us begin by defining the term "symmetry" according to its technical, rather than its common usage. An object is said to possess (spatial) symmetry if certain operations, such as rotations or translations, can be performed on the object, with the result that each of these operations carries the object to a position in space which is indistinguishable, by any known test, from its original position. An object is said to be invariant to each of its symmetry operations. The symmetry of an object is uniquely defined as soon as we enumerate all the operations which have the property of transforming the object into itself.

The translational symmetry of a crystal can be specified by any three independent identity periods. By definition, an identity period is the smallest possible translation in a given direction which carries the crystal into itself. Let us choose a convenient set of identity periods, and let us denote these vector quantities by the symbols $\boldsymbol{a}_1$, $\boldsymbol{a}_2$, and $\boldsymbol{a}_3$. Since these three vectors are required to be independent, they must not all lie in the same plane.

Consider the set of vectors

$$\boldsymbol{d} = d_1\boldsymbol{a}_1 + d_2\boldsymbol{a}_2 + d_3\boldsymbol{a}_3 \tag{1}$$

where d_1, d_2, d_3 are any three integers. Note that each vector in the set $\boldsymbol{d}$ is the sum of integral multiples of the three identity periods. It is obvious that the crystal will be transformed into itself by any translation $\boldsymbol{d}$.

The endpoints of the vectors belonging to the set $\boldsymbol{d}$ form a three-dimensional lattice of points known as the direct lattice. The vectors $\boldsymbol{d}$ are called the vectors of the direct lattice, or, more simply, the direct lattice vectors. The primitive (smallest) direct lattice vectors are the identity periods $\boldsymbol{a}_1$, $\boldsymbol{a}_2$, and $\boldsymbol{a}_3$.

Since a crystal is a periodic structure, it must possess a unit cell. It is convenient to choose as the unit cell a parallelopiped each of whose edges is a primitive translation vector. In Fig. 2 we show a network of points.

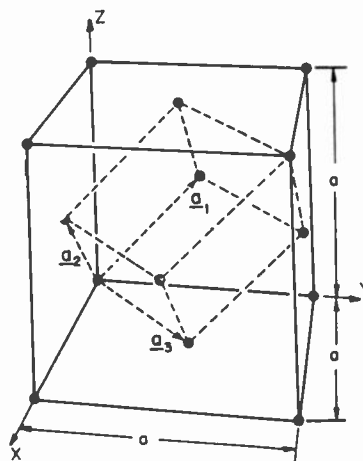


Fig. 2—Direct lattice for diamond-type crystals. The primitive translation vectors and the parallelopiped unit cell are shown.

These points form a face-centered cubic lattice. If we assume that the network of points corresponds to the direct lattice of some crystal, the parallelopiped defined by the dotted lines may be regarded as the unit cell of the crystal.

The atomic arrangement in diamond-type crystals is suggested by the model shown in Fig. 1. In this model,

each ball represents an atomic position, and each rod a chemical bond joining two atoms together. A comparison of Figs. 1 and 2 reveals that a diamond-type lattice is composed of two inter-penetrating face-centered cubic lattices. For convenience, let us call these two sub-lattices A and B. The four atoms neighboring an atom in A belong to B, and the four atoms neighboring an atom in B belong to A. It is readily established that a diamond-type lattice has the same translational periodicity as A or B. Therefore, A or B can be chosen as the direct lattice for the diamond-type crystal. Since the sub-lattices A and B are physically equivalent, it doesn't really matter which we choose. For the sake of argument, let us take the sub-lattice A as the direct lattice. If we assume that the sub-lattice A is pictured in Fig. 2, then it follows that the primitive translation vectors and the unit cell shown in Fig. 2 apply to the model in Fig. 1 as well. We note in passing that the density of atoms in a diamond-type crystal is two per unit cell.

In addition to its translational symmetry, a crystal may have other symmetry properties. For example, a diamond-type crystal is invariant to an inversion taken about the point midway between any two adjacent atomic positions. A diamond-type crystal is also invariant to certain rotations and reflections. These are indicated in Fig. 3, where we have shown an atom belonging to a diamond-type crystal together with its four nearest neighbors. It should be noted particularly that the four nearest neighbors lie at the vertices of a regular tetrahedron, and that each of the symmetry operations indicated in Fig. 3 has the property of transforming a regular tetrahedron into itself.

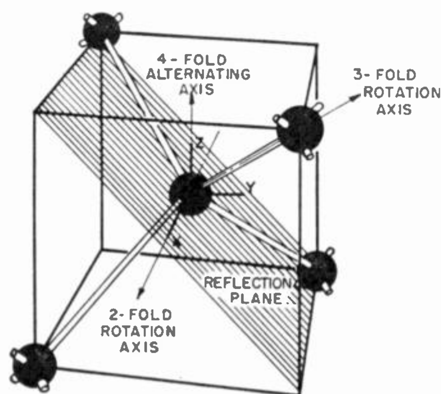


Fig. 3—Some point symmetry operations for diamond-type crystals.

The terminology used in Fig. 3 may be explained as follows: A rotation axis is said to be an n fold axis if the allowed angles of rotation are given by the expression $2\pi p/n$ radians, where n is an integer, and where $p=0$, $1, 2 \cdots n-1$. A rotation about an axis is known as a proper rotation. If a crystal can be transformed into itself by a rotation about an n fold axis, followed by a reflection in the plane perpendicular to this axis, the crystal is said to possess an n fold alternating axis. An alternating axis is sometimes called a rotation-reflection axis, and the rotation followed by the reflection a rotary-reflection or an improper rotation. An ordinary reflection is an improper rotation by the angle zero, while an inversion is an improper rotation by the angle π radians. Since proper and improper rotations leave at least one point of the crystal fixed in space, such symmetry operations are known as point symmetry operations.

We now return to the general theory. Let us introduce three new vectors $\boldsymbol{b}_1$, $\boldsymbol{b}_2$, and $\boldsymbol{b}_3$, and let us require that these vectors satisfy the following conditions:

$$\boldsymbol{a}_i \cdot \boldsymbol{b}_j = 2\pi\delta_{ij} = \begin{cases} 2\pi & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}; \qquad i, j = 1, 2, 3. \quad (2)$$

Note that the vector $\boldsymbol{b}_1$ is perpendicular to the plane passing through $\boldsymbol{a}_2$ and $\boldsymbol{a}_3$, etc. Consider the set of vectors

$$\boldsymbol{h} = h_1\boldsymbol{b}_1 + h_2\boldsymbol{b}_2 + h_3\boldsymbol{b}_3, \quad (3)$$

where h_1, h_2, and h_3 are any three integers. The endpoints of the set of vectors $\boldsymbol{h}$ form a lattice known as the reciprocal lattice. [Each vector $\boldsymbol{h}$ has the dimensions of a reciprocal length.] The vectors $\boldsymbol{h}$ are known as reciprocal lattice vectors. The vectors $\boldsymbol{b}_1$, $\boldsymbol{b}_2$, and $\boldsymbol{b}_3$ are the primitive translation vectors of the reciprocal lattice.

The concept of a lattice in reciprocal space finds an immediate application in the representation of periodic functions in the crystal (direct) space by three-dimensional Fourier series. Let the function $F(\boldsymbol{r})$ have the same symmetry properties as the crystal. For example, $F(\boldsymbol{r})$ might be the crystal potential or the crystal charge density. Since $F(\boldsymbol{r})$ must be invariant to any translation by a direct lattice vector $\boldsymbol{d}$, we must have:

$$F(\boldsymbol{r}) = F(\boldsymbol{r} + \boldsymbol{d}). \quad (4)$$

It is easily demonstrated that $F(\boldsymbol{r})$ can be represented by a three-dimensional Fourier series of the form:

$$F(\boldsymbol{r}) = \sum_{\boldsymbol{h}} f(\boldsymbol{h}) \exp(i\boldsymbol{h}\cdot\boldsymbol{r}), \quad (5)$$

where the $f(\boldsymbol{h})$ are Fourier coefficients, and where the summation is carried over all reciprocal lattice vectors $\boldsymbol{h}$. The proof is elementary. Suppose we apply the transformation $\boldsymbol{r}\rightarrow\boldsymbol{r}+\boldsymbol{d}$ to (5). This gives:

$$F(\boldsymbol{r} + \boldsymbol{d}) = \sum_{\boldsymbol{h}} f(\boldsymbol{h}) \exp[i\boldsymbol{h}\cdot(\boldsymbol{r} + \boldsymbol{d})]. \quad (6)$$

Consider the factor $\exp(i\boldsymbol{h}\cdot\boldsymbol{d})$. Making use of the definitions (1) and (3), and the relations (2), we readily establish that $\boldsymbol{h}\cdot\boldsymbol{d}=2\pi(h_1d_1+h_2d_2+h_3d_3)=2\pi$ times an integer. Hence, $\exp(i\boldsymbol{h}\cdot\boldsymbol{d})=\exp[2\pi i(\text{integer})]=1$, and it follows that (6) = (5). But this is essentially the periodicity condition (4). We may conclude that the periodic function $F(\boldsymbol{r})$ can be represented by the Fourier series given by (5).

It can be shown that the lattice reciprocal to a face-centered cubic lattice is body-centered cubic. Since the direct lattice for a diamond-type crystal is face-centered cubic, the reciprocal lattice for such a crystal must be

body-centered cubic. We show a body-centered cubic lattice in Fig. 4. It is convenient to choose for the unit cell of reciprocal space the truncated octahedron shown in Fig. 4, rather than a parallelopiped. To demonstrate how the truncated octahedra fill all of reciprocal space, we have displayed in Fig. 5 four truncated octahedra centered at four adjacent lattice sites of the body-centered cubic (reciprocal) lattice.

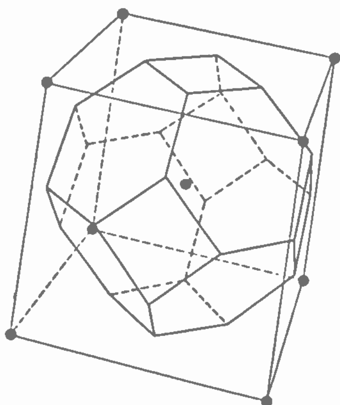


Fig. 4—Reciprocal lattice for diamond-type crystals. The truncated octahedron is the unit cell, or the reduced zone.

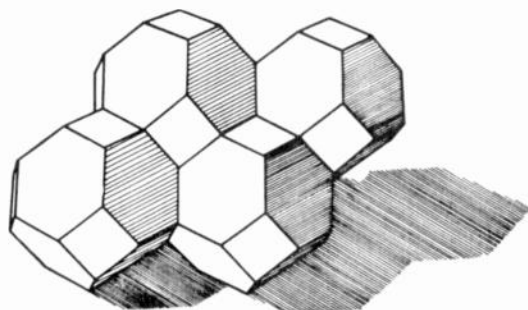


Fig. 5—Cluster of four truncated octahedra.

III. Electronic Quantum States in a Perfect Crystal

In this section, we will discuss the general nature of the quantum states which are available to electrons in a perfect crystal. We will also show how the available quantum states can be cataloged with the aid of the energy band scheme.

When an electron occupies a given quantum state, it can be represented by a wave function which is characteristic of this state, and it will have an energy which is also characteristic of this state. The wave functions and the energy levels corresponding to the various quantum states can be determined by solving a certain set of wave equations.

The wave equation (Schrödinger equation) for an electron may be written in the following form:

$$\left[-\frac{h^2}{8\pi^2 m}\nabla^2 + V(\boldsymbol{r})\right]\psi(\boldsymbol{r}) = E\psi(\boldsymbol{r}), \qquad (7)$$

where h is Planck's constant, m the mass of an electron in free space, and ∇^2 the Laplacian differential operator. $V(\boldsymbol{r})$ is the potential acting on the electron represented by the wave function $\psi(\boldsymbol{r})$. E is an energy parameter.

In the course of its motion through a crystal, an electron will be attracted by the nuclei, and repelled by the other electrons. The crystal potential $V(\boldsymbol{r})$ is so constructed as to take all of these attractions and repulsions into account in some suitable manner. For the purposes of the present discussion, we will assume that all the electrons in the crystal are acted upon by the same crystal potential, $V(\boldsymbol{r})$. We make this assumption in the interest of simplicity.

The most characteristic feature of the crystal potential $V(\boldsymbol{r})$ is its symmetry. The crystal potential must have the same symmetry properties as the crystal itself. In particular, $V(\boldsymbol{r})$ must be invariant to any translation which carries the crystal into itself. Thus, we may write:

$$V(\boldsymbol{r}) = V(\boldsymbol{r} + \boldsymbol{d}), \qquad (8)$$

where $\boldsymbol{d}$ is any direct lattice vector. It is possible to solve the wave equation (7) in a relatively simple and straightforward manner because the potential $V(\boldsymbol{r})$ is a periodic function.

The crystal wave equation (7) will have nontrivial solutions which satisfy the imposed boundary conditions (see below) only if the energy parameter E is set equal to any one of a number of distinctive values. These special values of E are known as the energy eigenvalues of the wave equation. The solutions $\psi(\boldsymbol{r})$ corresponding to these choices of E are called the crystal eigenfunctions. Each quantum state is characterized by a particular energy eigenvalue and by the corresponding crystal eigenfunction. Thus, when an electron is in some quantum state, its energy will be given by the energy eigenvalue appropriate to this quantum state.

In order to avoid the complications which arise when we attempt to impose realistic boundary conditions on the crystal under consideration, it is desirable to work with a mathematical construct known as a cyclic crystal. Briefly, a cyclic crystal is a representaitve sample of an infinite crystal. The cyclic crystal may be defined as the portion of an infinite crystal enclosed by a parallelopiped with concurrent edges $N\boldsymbol{a}_1$, $N\boldsymbol{a}_2$, $N\boldsymbol{a}_3$, whose center coincides with the origin of coordinates. The quantity N is an arbitrary, large integer, and $\boldsymbol{a}_1$, $\boldsymbol{a}_2$, and $\boldsymbol{a}_3$ are the primitive translation vectors of the direct lattice. It is readily established that the cyclic crystal contains exactly N^3 unit cells, each of volume $\boldsymbol{a}_1 \cdot \boldsymbol{a}_2 \times \boldsymbol{a}_3$.

The cyclic crystal concept is fully defined only after we introduce the so-called cyclic boundary conditions. The cyclic boundary conditions require that any two points in space are to be regarded as physically equivalent if they are separated by a vector of the type:

$$\boldsymbol{N} = N[n_1\boldsymbol{a}_1 + n_2\boldsymbol{a}_2 + n_3\boldsymbol{a}_3], \qquad (9)$$

where n_1, n_2, and n_3 are any three integers. In the following discussion, we will assume that all the functions with which we deal are defined only within the confines of the cyclic crystal, and that the cyclic boundary conditions apply to all these functions.

As first shown by Bloch, each solution of the crystal wave equation (7) satisfying the cyclic boundary conditions

$$\psi(\boldsymbol{r}) = \psi(\boldsymbol{r} + \boldsymbol{N}) \tag{10}$$

must have the form

$$\psi(\boldsymbol{r}) = \exp(i\boldsymbol{k}\cdot\boldsymbol{r})u(\boldsymbol{r}), \tag{11}$$

where $u(\boldsymbol{r})$ is a modulating function having the same translational periodicity as the crystal, and where $\boldsymbol{k}$ is a quantity which we will tentatively call a wave vector. A wave function of the form (11) is known as a Bloch function.

The wave vector $\boldsymbol{k}$ appearing in (11) is restricted to the following set of values:

$$\boldsymbol{k} = N^{-1}(k_1\boldsymbol{b}_1 + k_2\boldsymbol{b}_2 + k_3\boldsymbol{b}_3], \tag{12}$$

where N is the large integer defining the size of the cyclic crystal; $\boldsymbol{b}_1$, $\boldsymbol{b}_2$, and $\boldsymbol{b}_3$ are the primitive translation vectors of the reciprocal lattice, and k_1, k_2, and k_3 are any three integers. If $\boldsymbol{k}$ were chosen to be different from any of the values given by (12), the Bloch function (11) would not satisfy the cyclic boundary conditions. It is easy to demonstrate that the allowed values of $\boldsymbol{k}$ are uniformly distributed in reciprocal space, and that their density is N^3 per unit volume.

One can show that any two Bloch functions whose wave vectors differ by a reciprocal lattice vector describe the same type of wave behavior. Therefore, such wave functions are physically equivalent. The inclusion of physically equivalent solutions in our catalogue of solutions can be avoided by restricting the range of $\boldsymbol{k}$ in reciprocal space. This is most easily accomplished by requiring that $\boldsymbol{k}$ lie in the central unit cell of the reciprocal space. ($\boldsymbol{k}$ may also lie on the boundaries of this unit cell.) Wave vectors which lie in the central unit cell of reciprocal space, or on the boundaries of this unit cell, are known as reduced wave vectors. Henceforth, the symbol $\boldsymbol{k}$ will be used to designate a reduced wave vector, rather than any wave vector, and the central unit cell will be called the reduced zone. [The reduced zone for silicon and germanium is the truncated octahedron shown in Fig. 4.]

To summarize: Each crystal eigenfunction of the wave equation (7) satisfying the cyclic boundary conditions (10) is a Bloch function. Each Bloch function is characterized by a reduced wave vector $\boldsymbol{k}$. There are N^3 allowed values of $\boldsymbol{k}$, and these are uniformly distributed in the reduced zone.

The crystal wave equation actually has a number of eigenfunctions and energy eigenvalues for each of the allowed values of $\boldsymbol{k}$. These may be distinguished from each other by introducing an integral index β. Each eigensolution of the wave equation is uniquely specified by a reduced wave vector $\boldsymbol{k}$ and by an index β. The assignment of the index β to the solutions for each choice of $\boldsymbol{k}$ will be discussed shortly. For the present, we merely note that each energy eigenvalue can be written as $E_\beta(\boldsymbol{k})$, and that each Bloch function, or crystal eigenfunction, can be expressed as follows:

$$\psi_\beta(\boldsymbol{k}, \boldsymbol{r}) = \exp(i\boldsymbol{k}\cdot\boldsymbol{r})u_\beta(\boldsymbol{k}, \boldsymbol{r}), \tag{13}$$

where $u_\beta(\boldsymbol{k}, \boldsymbol{r})$, for each choice of β and $\boldsymbol{k}$, is a modulating function having the same translational periodicity as the crystal; that is,

$$u_\beta(\boldsymbol{k}, \boldsymbol{r}) = u_\beta(\boldsymbol{k}, \boldsymbol{r} + \boldsymbol{d}), \tag{14}$$

for all $\boldsymbol{k}$ and β.

We may now write down a set of crystal wave equations, one for each quantum state β, $\boldsymbol{k}$:

$$\left[-\frac{h^2}{8\pi^2 m}\nabla^2 + V(\boldsymbol{r})\right]\psi_\beta(\boldsymbol{k}, \boldsymbol{r}) = E_\beta(\boldsymbol{k})\psi_\beta(\boldsymbol{k}, \boldsymbol{r}). \tag{15}$$

The available electronic quantum states are given by the solutions to this set of wave equations. Since each electron in the crystal must occupy one of the available states, the only allowed energy levels are those corresponding to the energy eigenvalues $E_\beta(\boldsymbol{k})$. Any energy level not included among the energy eigenvalues $E_\beta(\boldsymbol{k})$ may be termed a forbidden energy level.

We will not attempt to explain here how the crystal potential $V(\boldsymbol{r})$ is calculated in practice, nor will we attempt to describe the methods whereby the crystal eigenfunctions $\psi_\beta(\boldsymbol{k}, \boldsymbol{r})$ and the energy eigenvalues $E_\beta(\boldsymbol{k})$ are actually evaluated. Rather, we will proceed on the following three assumptions: (*a*) the set of wave equations (15) provides an adequate means for determining the available quantum states; (*b*) the crystal potential $V(\boldsymbol{r})$ is known; and (*c*) the solutions to the set of wave equations have been determined.

The allowed energy levels $E_\beta(\boldsymbol{k})$ form a spectrum on an energy scale, which may be called the electronic energy spectrum. The spectrum for a typical crystal is composed of a number of deep-lying, widely spaced lines, and a number of higher-lying, closely spaced bands. The deep-lying lines correspond to the energy levels for the core electrons. While the core electrons are involved in many physical processes, such as X-ray emission and absorption, they play no important role in those processes with which we will be principally concerned, namely, electrical conduction and optical absorption and emission processes. Therefore, we will not consider the core energy levels further.

The broad bands, which lie above the core energy levels, may be divided into two classes, the valence bands, and the conduction bands. In all crystals, the conduction bands lie above the valence bands. The distinction between valence and conduction bands in nonmetallic crystals will be considered in the next section. The electrons whose energies lie in the energy intervals spanned by the valence bands are appropriately called valence electrons, and the electrons whose energy levels belong to the conduction bands are called conduction electrons or free electrons.

We wish to emphasize that the energy levels for the valence and the conduction electrons are confined to

very definite energy ranges, and that these energy ranges are often separated from each other by finite energy intervals which are called forbidden bands. In the case of nonmetallic crystals, with which we will be primarily concerned, the only forbidden band of interest is the one between the highest valence band and the lowest conduction band. Whenever we speak of a forbidden band in the remainder of this paper, we will always have this particular forbidden band in mind.

So far, we have only treated the energy spectrum in terms of allowed and forbidden levels on an energy scale. We will now consider the energy spectrum in a more general sense. Consider the set of energy levels $E_\beta(\boldsymbol{k})$ corresponding to a particular value of $\boldsymbol{k}$, namely, $E_1(\boldsymbol{k})$, $E_2(\boldsymbol{k})$, $E_3(\boldsymbol{k}) \cdots$. If $\boldsymbol{k}$ lies at a general point of the reduced zone, each energy level will usually occur only once in the catalogue of energy levels. In such a case, we can assign the index β to the various quantum states in the order of ascending energy; that is, we label the states with β such that:

$$E_1(\boldsymbol{k}) < E_2(\boldsymbol{k}) < E_3(\boldsymbol{k}) \cdots. \tag{16}$$

Once this assignment is made, we can define, in an unambiguous manner, a set of energy functions $E_\beta(\boldsymbol{k})$ satisfying the following relation:

$$E_1(\boldsymbol{k}) \leqq E_2(\boldsymbol{k}) \leqq E_3(\boldsymbol{k}) \cdots \tag{17}$$

for all choices of $\boldsymbol{k}$. In (16), the symbol $E_\beta(\boldsymbol{k})$ denotes the energy level corresponding to the quantum state β, $\boldsymbol{k}$. In (17), the symbol $E_\beta(\boldsymbol{k})$ denotes an energy function which describes the $\boldsymbol{k}$-dependence of the energy eigenvalues specified by the index β. In this article, we will sometimes use the symbol $E_\beta(\boldsymbol{k})$ in the first sense, and sometimes in the second sense. The context will usually suggest which sense is intended.

The quantum state β, $\boldsymbol{k}$ is said to be nondegenerate if the energy level $E_\beta(\boldsymbol{k})$ occurs only once in the catalog of energy levels specified by the reduced wave vector $\boldsymbol{k}$. If a given energy level is repeated n times, for some choice of $\boldsymbol{k}$, the n quantum states having this energy level in common are said to be n fold degenerate. Thus, if $E_\beta(\boldsymbol{k}) = E_{\beta'}(\boldsymbol{k})$, the states β, $\boldsymbol{k}$ and β', $\boldsymbol{k}$ are two-fold degenerate. Energy degeneracies occur most often for values of $\boldsymbol{k}$ lying at symmetry points or along symmetry axes of the reduced zone. It is usually possible to predict the occurrence of energy degeneracies by symmetry arguments. The equality signs in (17) allow for energy degeneracies. The states corresponding to a value of $\boldsymbol{k}$ lying at a general point of the reduced zone are almost always nondegenerate.

We now come to the heart of the subject, the concept of an energy band. An energy band is defined as the set of N^3 quantum states associated with the N^3 allowed values of $\boldsymbol{k}$ and with a particular value of β. An energy band may also be defined as the set of N^3 energy levels $E_\beta(\boldsymbol{k})$ associated with the N^3 allowed values of $\boldsymbol{k}$ and with some value of β. The energy function $E_\beta(\boldsymbol{k})$ describes the $\boldsymbol{k}$-dependence of the energy band β.

The energy band functions $E_\beta(\boldsymbol{k})$ have several interesting properties. These will be described briefly:

The energy band function $E_\beta(\boldsymbol{k})$ is a continuous function of $\boldsymbol{k}$ throughout the reduced zone. The function $E_\beta(\boldsymbol{k})$ is also a differentiable function of $\boldsymbol{k}$ throughout the reduced zone, with the exception of those values of $\boldsymbol{k}$ for which the band β comes into contact with some other band or bands; that is, $E_\beta(\boldsymbol{k})$ will be differentiable where it is nondegenerate, and nondifferentiable where it is degenerate. Where $E_\beta(\boldsymbol{k})$ is differentiable, all the derivatives of $E_\beta(\boldsymbol{k})$ with respect to the components of $\boldsymbol{k}$, in all orders, can be uniquely defined. The differentiability of $E_\beta(\boldsymbol{k})$ will prove important in Section V.

The function $E_\beta(\boldsymbol{k})$ has the same rotational symmetry in the reduced zone as does the reduced zone in reciprocal space. Let R be any proper or improper rotation which transforms the reduced zone—or the reciprocal lattice, for that matter—into itself. [The center of rotation is placed at the central point of the reduced zone.] When R acts on $\boldsymbol{k}$, it carries $\boldsymbol{k}$ into a new position in the reduced zone. We will denote the new position by the symbol $R(\boldsymbol{k})$. Since $E_\beta(\boldsymbol{k})$ has the same rotational symmetry as the reduced zone, $E_\beta(\boldsymbol{k})$ must be invariant to each of the rotations R. Therefore, we may write:

$$E_\beta(\boldsymbol{k}) = E_\beta[R(\boldsymbol{k})]. \tag{18}$$

If $E_\beta(\boldsymbol{k})$ attains its maximum or its minimum value at $\boldsymbol{k} = \boldsymbol{k}_0$, it must have symmetrically equivalent maxima or minima at each of the points $R(\boldsymbol{k}_0)$.

It can be shown that the inversion operation is always included among the symmetry operations R. Thus, whether or not the crystal has a center of inversion, the function $E_\beta(\boldsymbol{k})$ will be invariant to inversion about the central point of the reduced zone; that is, we will have, for all crystals, the following relation:

$$E_\beta(\boldsymbol{k}) = E_\beta(-\boldsymbol{k}). \tag{19}$$

Let us denote by the symbol $\boldsymbol{K}$ any point in the reciprocal space. [When $\boldsymbol{K}$ lies in the reduced zone, it becomes a reduced wave vector $\boldsymbol{k}$.] It is not difficult to prove that $E_\beta(\boldsymbol{K})$, regarded as a function of $\boldsymbol{K}$, has the same symmetry properties as the reciprocal lattice. In particular, $E_\beta(\boldsymbol{K})$ has the same translational periodicity in reciprocal space as the reciprocal lattice. Hence, we may write:

$$E_\beta(\boldsymbol{K}) = E_\beta(\boldsymbol{K} + \boldsymbol{h}), \tag{20}$$

where $\boldsymbol{h}$ is any reciprocal lattice vector. It follows from (20) that the energy function $E_\beta(\boldsymbol{k})$ must assume equal values at any two points on the surfaces of the reduced zone which can be joined by a reciprocal lattice vector.

What is the relationship between the nature of the crystal potential $V(\boldsymbol{r})$ and the nature of the energy band functions $E_\beta(\boldsymbol{k})$? The symmetry properties (18) and (20) are associated with the translational symmetry of the crystal potential. The occurrence of energy degeneracies, *i.e.*, contacts between adjacent energy bands, is related to all the symmetry properties of the crystal

WRH

potential, particularly the rotational symmetry properties. It is not necessary to solve the set of crystal wave equations (15) in order to determine the nature of the contacts between energy bands and the symmetry of the various energy bands. Such information can be obtained by symmetry arguments alone. On the other hand, the exact form of the various energy bands—by which we mean the width and the location of the maxima and the minima and similar items—can be determined only by solving the wave equations (15); that is, the form of the various energy bands depends upon the exact nature of the crystal potential, rather than upon its symmetry properties alone. Thus, there is a limit to what symmetry arguments can tell us about the energy band structure of a crystal.

It is often necessary to have a firm picture of the $\boldsymbol{k}$-dependence of an energy function $E_\beta(\boldsymbol{k})$. The $\boldsymbol{k}$-dependence of an energy function can be indicated in a variety of ways. For example, $E_\beta(\boldsymbol{k})$ can be represented by a family of constant energy surfaces in the reduced zone. This type of representation requires a three-dimensional drawing. It is sometimes more convenient to show the intersection of the family of constant energy surfaces with some plane in the reduced zone, usually a plane passing through the central zone point. The intersection, which may be drawn in two dimensions, consists of a family of equi-energy contours, one contour for each constant energy surface. The intersection may also be shown as a conventional contour map in three dimensions, with energy plotted vertically, and the plane laid out horizontally.

The representation used most frequently in practice shows $E_\beta(\boldsymbol{k})$ plotted against $\boldsymbol{k}$ for a particular direction of $\boldsymbol{k}$. This gives a line—an energy profile—for the energy band β. If $E_\beta(\boldsymbol{k})$ is plotted against $\boldsymbol{k}$ for a few well-chosen directions of $\boldsymbol{k}$, a reasonably complete picture of the $\boldsymbol{k}$-dependence of $E_\beta(\boldsymbol{k})$ can be developed. We will have occasion to use various representations for $E_\beta(\boldsymbol{k})$ in the course of this article.

It is desirable to define in precise terms the expression "width of the forbidden band" for the case of a nonmetallic crystal. Let us designate the highest valence band and the lowest conduction band by the symbols "val" and "cond." The width of the forbidden band, or the energy gap E_{gap}, is given by:

$$E_{\text{gap}} = E_{\text{cond}}(\boldsymbol{k}_c) - E_{\text{val}}(\boldsymbol{k}_v), \tag{21}$$

where $\boldsymbol{k}_c$ is a point in the reduced zone where the lowest conduction band reaches its lowest energy value, and where $\boldsymbol{k}_v$ is a point in the reduced zone where the highest valence band attains its highest energy value.

Let the energy separation between the highest valence band and the lowest conduction band at the point $\boldsymbol{k}$ be denoted by the symbol $E_{\text{vert}}(\boldsymbol{k})$ ["vert" for vertical]. By definition, we have:

$$E_{\text{vert}}(\boldsymbol{k}) \equiv E_{\text{cond}}(\boldsymbol{k}) - E_{\text{val}}(\boldsymbol{k}). \tag{22}$$

If $\boldsymbol{k}_v = \boldsymbol{k}_c$, as may be the case for some nonmetallic crystals, then:

$$E_{\text{gap}} = E_{\text{vert}}(\boldsymbol{k}_v) = E_{\text{vert}}(\boldsymbol{k}_c). \tag{23}$$

On the other hand, if $\boldsymbol{k}_v \neq \boldsymbol{k}_c$, as is the case for silicon and germanium, we may expect to have:

$$E_{\text{gap}} < E_{\text{vert}}(\boldsymbol{k}), \tag{24}$$

for all values of $\boldsymbol{k}$. The distinction between the minimum value of $E_{\text{vert}}(\boldsymbol{k})$ and the minimum energy separation E_{gap} is an important one for crystals in which $\boldsymbol{k}_v \neq \boldsymbol{k}_c$. We will return to this when we discuss the optical properties of nonmetallic crystals in Section XV.

IV. Occupancy of the Electronic Quantum States

In this section we will explain how the electrons are distributed among the available quantum states under equilibrium conditions.

The occupancy of each quantum state β, $\boldsymbol{k}$ is strictly limited by the Pauli Exclusion Principle. According to this principle, no more than two electrons can occupy a given quantum state β, $\boldsymbol{k}$. If two electrons occupy the same quantum state, they must have oppositely-oriented spins. The state of an electron is really not fully specified if all we know about the electron is which quantum state β, $\boldsymbol{k}$ it occupies. In addition to this information, we must be given its spin orientation, which can assume one of two possible values. For simplicity, we will sometimes assume that the specification of the electron spin orientation is included in the index β. Whenever we refer to an electronic quantum state by the indices β and $\boldsymbol{k}$, we will usually have in mind the state associated with either spin orientation, the particular orientation being given by β. In certain contexts, we will use the symbols β and $\boldsymbol{k}$ to refer to the pair of quantum states having β and $\boldsymbol{k}$ in common. On such occasions, we will generally speak of the quantum state β, $\boldsymbol{k}$ as a double quantum state; *i.e.*, a quantum state capable of double occupancy by two electrons having oppositely-oriented spins. It should be noted that the two electrons occupying the (double) quantum state β, $\boldsymbol{k}$ have the same energy, $E_\beta(\boldsymbol{k})$, and the same Bloch function $\psi_\beta(\boldsymbol{k}, \boldsymbol{r})$.

The distribution of the electrons among the various quantum states under conditions of thermal equilibrium is governed by the Fermi-Dirac statistics. These statistics take into account the restriction on the occupancy of states imposed by the Pauli Exclusion Principle. According to the Fermi-Dirac statistics, the probability that a quantum state with energy E is occupied by an electron of either spin orientation is:

$$f_0(E, T) = \frac{1}{\exp\left[(E - \xi)/kT\right] + 1}, \tag{25}$$

where ξ is the so-called Fermi energy, T the absolute temperature, and k Boltzmann's constant. [Although we use the symbol k both for the Boltzmann constant and the magnitude of the reduced wave vector $\boldsymbol{k}$, there

should be no confusion since the Boltzmann constant always appears together with T in the combination kT.] $f_0(E, T)$ is called the Fermi distribution function.

At the absolute zero of temperature, the Fermi distribution function has the following characteristics:

$$f_0(E, 0) = \begin{cases} 1 & E < \xi \\ \frac{1}{2} & E = \xi. \\ 0 & E > \xi \end{cases} \qquad (26)$$

There is a sharp cutoff in the distribution of the electrons among the available states. The states for which $E < \xi$ are fully occupied, and the states for which $E > \xi$ are totally unoccupied.

At any finite temperature, the cutoff in the state occupancy becomes somewhat diffuse. It is easily shown that the Fermi distribution function is appreciably different from either unity or zero only in a range of energy of the order of kT centered at the Fermi energy level ξ. The Fermi function is equal to $\frac{1}{2}$ at the Fermi energy level for all values of the temperature.

The following approximations are sometimes useful:

$$f_0(E, T) = \exp[-(E - \xi)/kT]; \quad E - \xi > 0; \quad E - \xi \gg kT. \qquad (27)$$

$$1 - f_0(E, T) = \exp[-(\xi - E)/kT]; \quad \xi - E > 0; \quad \xi - E \gg kT. \qquad (28)$$

The fraction of the available states which will be occupied at temperature T in the range $E, E+dE$, where E lies considerably above ξ, is given by (27). The fraction of the available states which will be unoccupied at temperature T in the range $E, E+dE$, where E lies considerably below ξ, is given by (28). [The distribution of the occupied and unoccupied states is worked out for a special case in the appendix.]

In a nonmetallic crystal, *i.e.*, in a semiconductor or an insulator, the Fermi energy lies in the forbidden energy band separating the highest valence band from the lowest conduction band. At the absolute zero of temperature, all the core and valence band states will be fully occupied (see below), and all the conduction band states will be totally empty. In a fully occupied band, there are no higher energy states within the band into which an electron can jump. Since the electrons in a fully occupied band cannot gain energy from an applied electric field, such electrons cannot contribute to the electrical conductivity. The nonmetallic crystal will behave like a perfect insulator at absolute zero temperature because all of its electrons belong to fully occupied bands.

At finite temperatures, some electrons will be thermally excited from states lying near the top of the highest valence band to states lying near the bottom of the lowest conduction band. The electrons in a partially filled band can gain energy from an applied electric field because they can jump into the unfilled states in the band. Consequently, the electrons in the partially occupied conduction band and the electrons in the partially unoccupied valence band can contribute to the electrical conductivity, and the nonmetallic crystal will lose its insulating properties at finite temperatures.

The Fermi energy will normally lie near the center of the forbidden band in a (perfect) nonmetallic crystal. In the temperature range for which $E_{gap} \gg kT$, the distribution of the occupied electron states in the conduction band is given by (27), and the distribution of the unoccupied electron states in the valence band is given by (28). When the electronic distribution can be described by (27) and (28), the electrons are said to obey the Maxwell-Boltzmann or the classical statistics. The Fermi-Dirac statistics reduce to the classical statistics whenever the exponential term in the denominator of (25) is vastly different from unity. This condition has been anticipated in (27) and (28).

In a metallic crystal, the Fermi energy lies in an allowed energy band, rather than in a forbidden energy band. This allowed band is called the (lowest) conduction band. Since the (lowest) conduction band is partially filled, and since the electrons in a partially filled band can conduct electricity, the metallic crystal will conduct electricity at all values of the temperature, including absolute zero. In practice, nonmetallic and metallic crystals can be distinguished by the difference in the temperature dependence of their electrical conductivities over a wide temperature range.

Let us now explore the nature of fully occupied bands in some detail. It will be recalled that an energy band β is formed by the N^3 (double) quantum states associated with the N^3 allowed values of k and with a particular band index β. If each of the N^3 quantum states is doubly occupied (by two electrons having oppositely oriented spins), the band is said to be fully occupied. According to this usage, a fully occupied band accommodates exactly $2N^3$ electrons.

A necessary condition for a crystal to be nonmetallic is that the number of electrons per unit cell be even. Consider a cyclic crystal containing N^3 unit cells, and containing Q electrons per unit cell. Such a crystal will have QN^3 electrons in all. Since each energy band can accommodate no more than $2N^3$ electrons, it follows that the QN^3 electrons can occupy an integral number of bands only if Q is even. At the absolute zero of temperature the $Q/2$ lowest bands will be fully occupied, and the remaining bands totally unoccupied, provided Q is even, and provided there is a finite energy separation between the band $\beta = Q/2$ and the band $\beta = Q/2+1$. But this is just the condition for the crystal to be nonmetallic.

The condition that Q be even is not a sufficient condition for the crystal to be nonmetallic. If Q is even, and if the energy band $\beta = Q/2+1$ touches the energy band $\beta = Q/2$, there will be no finite energy separation between the highest occupied state and the lowest unoccupied state at absolute zero temperature. In this particular case, the crystal will be metallic, rather than nonmetallic.

If Q is odd, the crystal will necessarily be metallic since an integral number of energy bands will not be fully occupied at absolute zero temperature. In fact, if Q is odd, one band—the (lowest) conduction band—will be exactly half filled at the absolute zero of temperature.

We now introduce the hole concept, one of the most important concepts in the physics of nonmetallic crystals. Consider a nonmetallic crystal at the absolute zero of temperature. At this temperature, the core and the valence band states will be fully occupied, and the conduction band states will be totally empty. Let us remove a single electron from the crystal. The state left vacant by the removal of the electron is known as a defect electron state or a hole state.

The vacant state is an available state into which an electron in the band containing the vacant state can jump. The existence of an available state into which an electron can jump makes electrical conduction possible. This type of conduction is known as hole conduction since it involves a hole or a defect in an otherwise filled band.

It can be shown that the behavior of all the remaining electrons in the crystal is equivalent to the behavior of a single positively charged particle known as the hole. By suitable manipulation, the properties of a hole can be deduced from a knowledge of the properties of the missing electron to which it corresponds. This theme is developed further in Section VI, where we show how the velocity and the acceleration of a hole can be calculated. It is interesting to note that a hole can be represented by the wave function of the missing electron. In writing down a wave equation for a hole, care must be taken to give the hole a positive, rather than a negative charge. When the crystal is in its state of lowest energy, the hole will occupy a state at the top of the highest valence band; that is, there will be a vacant electronic state at the top of the highest valence band.

Let us turn to another nonmetallic crystal at absolute zero temperature. This time, we will inject an electron into the crystal. Since all the valence and core states are fully occupied, the electron must assume a conduction band state. When the crystal is in its state of minimum energy, the injected electron will occupy a state at the bottom of the lowest conduction band. Such a state is called an excess electron state. More generally, the states occupied by the electrons in a nearly empty band are called excess electron states. Similarly, the unoccupied states in a nearly filled band are called defect electron states or holes states.

A hole owes its existence, at least its conceptual existence, to the fact that the electrons in a nonmetallic crystal can be arranged into open and closed groups, *i.e.*, into partially and totally filled bands. If an open group contains relatively few electrons, the behavior of these electrons can be treated in a straightforward manner. The electrons are regarded as occupying excess electron states, and are handled accordingly. On the other hand, if a group is a few electrons short of being a closed group, the electrons in this group can be treated more conveniently in terms of the defect electrons or holes.

The hole concept, then, is a convenient means for representing the electrons in a nearly closed group in terms of the electrons which are needed to close the group; that is, the behavior of the electrons in a nearly filled band can be treated in terms of the holes corresponding to the electrons, which, if present, would complete the filling of the band.

As we have already mentioned, the electrons in a filled band, *i.e.*, a closed group, cannot contribute to the electrical conductivity. On the other hand, the electrons in a partially filled band, *i.e.*, an open group, can contribute to the conductivity. In a nonmetallic crystal, the electrons in the lowest conduction bands and those in the highest valence bands are the only ones forming open groups. The electrons belonging to the lower valence bands and to the core bands generally form closed groups. Thus, in many problems, the only states of interest are those lying near the top of the highest valence band or bands—where the holes reside—and those lying near the bottom of the lowest conduction band or bands—where the conduction or "free" electrons reside.

Since many of the dynamical attributes of electrons and holes can be deduced directly from the energy band structure, and since electrons and holes occupy states near the extrema of energy bands, it is desirable to investigate the extrema of energy bands in some detail. This is done in the next section.

V. Structure of an Energy Band Near Its Extrema

In this section we will examine the structure of an energy band having a number of extrema located at a set of symmetrically equivalent points in the reduced zone. We will be concerned chiefly with the structure close to the extrema, which may be minima or maxima. For the present, we will assume that the energy band under consideration is nondegenerate at its extrema; that is, the energy band does not touch any other band or bands at its extrema. We will remove this restriction toward the end of this section.

Since we will be dealing with a single energy band in the discussion to follow, we will suppress the band index β and write the band function as $E(\mathbf{k})$. As we have already mentioned, the function $E(\mathbf{k})$ is differentiable at those points in the reduced zone where it is nondegenerate. Since $E(\mathbf{k})$ is assumed to have its extrema at a set of points where it is nondegenerate, $E(\mathbf{k})$ can be expanded as a Taylor series at each extremum. Because the various extrema are symmetrically equivalent, it is sufficient to treat a single extremum, say the one occurring at $\mathbf{k}=\mathbf{k}_0$. The Taylor series will not contain first order terms because $E(\mathbf{k})$ is an extremum at $\mathbf{k}=\mathbf{k}_0$. We will retain only the constant term and the second order terms, and drop all higher order terms, because we are

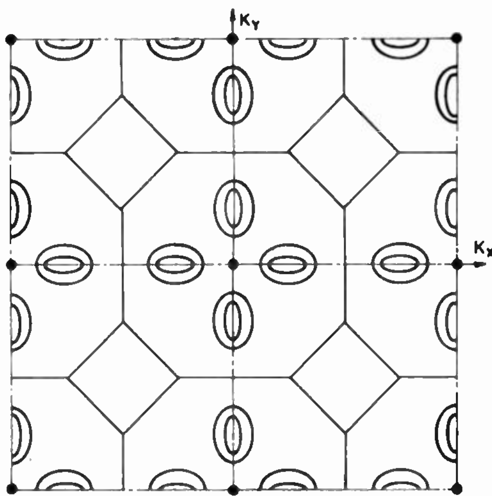


Fig. 6—Two equi-energy contours in the plane $K_z=0$, for a band having six extrema located at the midpoints of the 100 axes in the reduced zone. The contours are shown in a portion of the extended reciprocal space as well as in the reduced zone

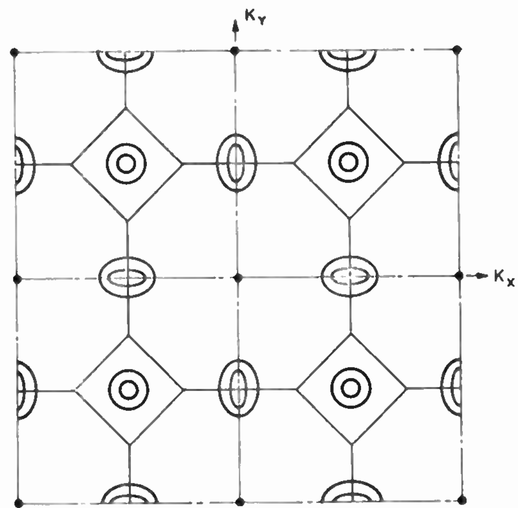


Fig. 7—Two equi-energy contours in the plane $K_z=0$, for a band having three extrema located at the midpoints of the square face centers of the reduced zone. The contours are shown in a portion of the extended reciprocal space as well as in the reduced zone.

interested only in the portion of the band lying close to k_0. For reasons which will become apparent, we write the Taylor series as follows:

$$E(k') \equiv E(k - k_0) = E(k_0) + (h^2/8\pi^2) \sum_i \sum_j (k_i' k_j' / m_{ij}^*), \quad (29)$$

where $i, j = x, y, z$, and where:

$$1/m_{ij}^* \equiv (2\pi/h)^2 [\partial^2 E(k')/\partial k_i' \partial k_j']_{k'=0} \quad (30)$$

is the ij component of a certain tensor which is known as the effective mass tensor because its components have the dimensions of a reciprocal mass. The effective mass tensor measures the curvature of an energy band at its point of definition. The physical significance of the effective mass tensor will be considered in the next section.

It is always possible to choose a coordinate system in which the nondiagonal ($i \neq j$) components of the effective mass tensor vanish. The (nonvanishing) diagonal components in this coordinate system are known as the principal components of the effective mass tensor Denoting the coordinates of $k' = k - k_0$ in the new frame of reference by the symbols k_a', k_b', and k_c', and writing the principal components as m_a^*, m_b^*, and m_c^*, we may replace (29) by the diagonal expression:

$$E(k') = E(k_0) + \frac{h^2}{8\pi^2} \left[\frac{(k_a')^2}{m_a^*} + \frac{(k_b')^2}{m_b^*} + \frac{(k_c')^2}{m_c^*} \right]. \quad (31)$$

Since the effective mass tensor measures the curvature of the energy surface at its point of definition, the three principal components are positive at an energy band minimum, and negative at an energy band maximum.

Let us now consider the nature of the constant energy surfaces

$$E(k) = \text{constant} \quad (32)$$

in the neighborhood of the extremum at $k = k_0$. If the three principal components are all equal, the constant energy surfaces will be spherical. When two components are equal, and the third different, the constant energy surfaces will be ellipsoids of revolution. If the three components are all different, the constant energy surfaces will be general ellipsoids.

The constant energy surfaces near the top or bottom of an energy band having several equivalent extrema will consist of a number of symmetrically equivalent ellipsoids. For the purpose of illustration, consider an energy band for a crystal having the translational periodicity of a face-centered cubic lattice. For such a crystal, the reduced zone is the truncated octahedron shown in Fig. 4. If the extrema occur along the 100 axes or along the 111 axes with the zone, the constant energy surfaces will consist of six or eight ellipsoids of revolution, respectively. If the extrema occur at the square face centers or at the hexagonal face centers, the constant energy surfaces will consist of three or four ellipsoids of revolution, respectively. For other locations of the multiple extrema, the constant energy surfaces will consist of a certain number of (general) ellipsoids. If the band has a single extremum, the extremum must lie at the central zone point, and the constant energy surfaces near the extremum must be spheres.

In Figs. 6 and 7 we have shown the intersection of two constant energy surfaces with the plane $K_z=0$ in a portion of the extended reciprocal space, as well as in the reduced zone. The heavy dots in these figures represent the loci of points belonging to the reciprocal lattice [cf. Figs. 4 and 5]. We have arbitrarily drawn the ellipsoids of revolution as prolate, rather than oblate. A portion of the extended reciprocal space has been included in these figures in order to dramatize the fact that $E(K)$ has the same symmetry properties as the reciprocal lattice. [cf. (20) above.] Symmetry relations (18) and (19) are also illustrated by these drawings.

Up to this point, we have assumed that the energy band under consideration is nondegenerate at its extrema. By making this assumption, we were able to expand the band function as a Taylor series in the neighborhood of each extremum. We now turn to a more complicated case, namely, an energy band whose extrema occur at points of degeneracy. In this case, the band function cannot be differentiated at its extrema. It can be shown that the constant energy surfaces near each extremum must be warped surfaces, rather than ellipsoidal surfaces. We will give specific examples of such warped surfaces when we come to the valence band structure of silicon and germanium in a subsequent section.

By way of summary, we have shown in Fig. 8 several types of band structures. In Fig. 8(a), we show an energy band having a single extremum at the central point of the reduced zone. The band is assumed to be nondegenerate at the central point. As we have already indicated, the constant energy surfaces are spherical.

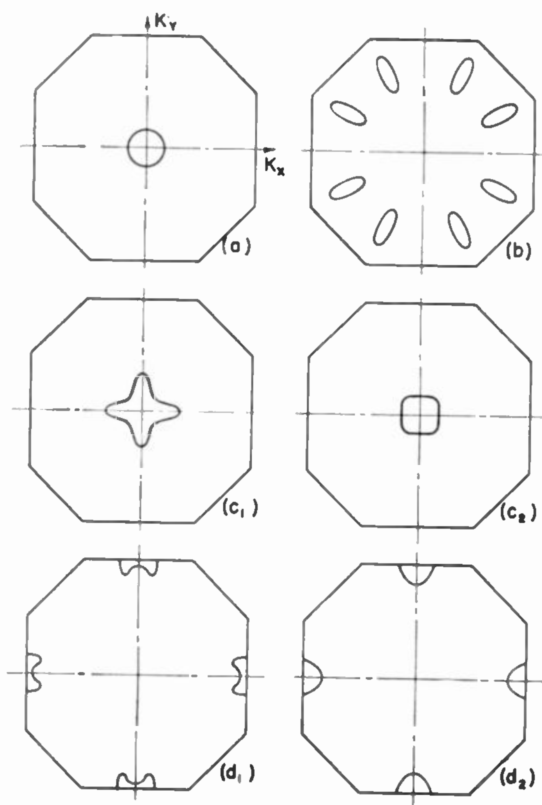


Fig. 8—Typical equi-energy contours in the plane $k_z=0$ for several cases. In (*a*), the contour is a circle, in (*b*), the contours are ellipses; in the remaining figures, the contours are warped.

We have drawn a circular equi-energy contour to represent this situation. In Fig. 8(b), we have an energy band with extrema occurring along the 24 [1, ½, 0) axes. Eight of these lie in the plane $k_z=0$. The band is again assumed to be nondegenerate at the extrema. The constant energy surfaces near the extrema consist of 24 general ellipsoids. We have drawn the elliptical equi-energy contours lying in the plane $k_z=0$ to represent this case.

In Figs. 8(c_1) and 8(c_2), we show the warped contours corresponding to two energy bands which come into contact with each other at the central zone point. The equi-energy contours for the two highest valence bands in silicon and germanium have the forms given by these figures. In Figs. 8(d_1), and 8(d_2) we have drawn the warped contours corresponding to two energy bands which are degenerate at the square face centers of the reduced zone. These two figures should be compared with the elliptical contours shown in Fig. 7, which correspond to a nondegenerate band.

VI. Velocity and Acceleration of Electrons and Holes

In this section, we show how the instantaneous velocity and the instantaneous acceleration of an electron or a hole subjected to an applied force can be determined directly from the band structure.

A few introductory remarks are in order. Under equilibrium conditions, the electrons in a crystal occupy quantum states whose properties do not change with time. Such quantum states art called stationary quantum states. The eigensolutions of the crystal wave equation (15) describe stationary quantum states. Under nonequilibrium conditions, which exist whenever the crystal is subjected to an external force, the electrons in the crystal occupy quantum states whose properties change with time. Such quantum states are called nonstationary quantum states. An electron in a nonstationary state must be represented by a wave function appropriate to such a state.

It is possible to construct a nonstationary state wave function from the wave functions of the equilibrium stationary states by superimposing a number of Bloch functions with adjacent values of $\boldsymbol{k}$, such that the superposition forms a wave packet. Let us construct a wave packet from the Bloch functions belonging to the energy band β, and let the mean reduced wave vector of the wave packet be denoted by $\boldsymbol{k}(t)$. We use the symbol $\boldsymbol{k}(t)$ to indicate that the composition of the wave packet, and hence the value of the mean reduced wave vector of the packet, changes with time. It can be shown that the time rate of change of $\boldsymbol{k}(t)$ is directly proportional to the applied force. The actual relationship is:

$$\dot{\boldsymbol{k}}(t) = (2\pi/h)\boldsymbol{F}_\beta[\boldsymbol{k}(t)], \tag{33}$$

where $\boldsymbol{F}_\beta[\boldsymbol{k}(t)]$ is the instantaneous value of the applied force acting on the electron in the nonstationary state β, $\boldsymbol{k}(t)$, *i.e.*, on the electron represented by the wave packet described above.

The instantaneous average velocity of an electron in the nonstationary state β, $\boldsymbol{k}(t)$ is given by:

$$\boldsymbol{v}_\beta[\boldsymbol{k}(t)] = (2\pi/h)\ \text{grad}_k\ E_\beta[\boldsymbol{k}(t)]. \tag{34}$$

The electron velocity, which corresponds to the group velocity of its wave packet, is given, then, by the gradient of the energy function $E_\beta(\boldsymbol{k})$, evaluated at the position of the mean reduced wave vector of the wave packet

The instantaneous average electric current produced by the motion of the electron is:

$$i_\beta[k(t)] = -ev_\beta[k(t)], \quad (35)$$

where e is the magnitude of the electron charge.

The instantaneous value of the force exerted on an electron in the state β, $k(t)$ by an applied electric field $\mathcal{E}(t)$ and an applied magnetic field $\mathcal{H}(t)$ is:

$$F_\beta[k(t)] = -e[\mathcal{E}(t) + (1/c)v_\beta\{k(t)\} \times \mathcal{H}(t)], \quad (36)$$

where c is the velocity of light. This expression is valid for steady applied fields as well as for applied fields which vary slowly in time. [If the periods of the applied fields are long compared with the mean time between successive collisions of the electron with the lattice, the fields may be regarded as slowly varying; otherwise, the fields are rapidly varying, and (36) does not apply.]

The instantaneous acceleration of an electron in the state β, $k(t)$ is related to the applied force in the following manner:

$$v_\beta[k(t)] = \{(2\pi/h)^2 \mathrm{grad}_k \mathrm{grad}_k E_\beta[k(t)]\} \cdot F_\beta[k(t)], \quad (37)$$

where

$$(2\pi/h)^2 \mathrm{grad}_k \mathrm{grad}_k E_\beta[k(t)] \quad (38)$$

is the effective mass tensor appropriate to the instantaneous position of $k(t)$ on the energy surface $E_\beta(k)$.

An electron (or a hole) moving through a crystal will collide with the lattice from time to time. [We will discuss such collisions in Section XIII.] Whenever a collision occurs, the electron (or the hole) will either gain or lose a certain amount of energy. As a result, the electron (or the hole) will make a transition from one quantum state to another. The various expressions given above apply to an electron during the time interval between successive collisions. The analogous expressions for a hole will be described shortly.

In the course of its wanderings through the crystal, a conduction band electron will occupy a certain number of quantum states in succession. These quantum states will all lie near the bottom of the conduction band. Therefore, a conduction electron will always have a positive effective mass tensor; that is, the electron will always occupy states for which the principal components of effective mass tensor are positive. [At the bottom of a band, the curvature of the energy band is upward; hence the effective mass tensor is positive in such a region.]

We now turn to the case of a hole. A hole can be represented by the wave packet for the missing electron to which it corresponds. For the purposes of the following discussion, we will assume that the missing electron state is specified by the wave packet β, $k(t)$. Hence, the hole may be denoted by β, $k(t)$. The instantaneous average velocity of the hole β, $k(t)$ is given directly by (34) above. The current due to the motion of the hole is given by (35), with $-e$ replaced by $+e$. The change in sign is required because the hole carries a positive, rather than a negative charge. The force exerted by applied electric and magnetic fields on the hole is given by (36), with $-e$ replaced by $+e$.

The effective mass tensor for the hole β, $k(t)$ is given by the negative of (38). Since the state β, $k(t)$ lies near the top of the valence band, the principal components of the effective mass tensor (38) are negative. [The curvature of the energy surface is downward at the top of a band.] Hence, the effective mass tensor for a hole is a positive quantity. The instantaneous average acceleration of the hole is given by (37), provided two changes of sign are made. The first change of sign occurs in the force term [$-e$ replaced by $+e$], and the second change occurs in the effective mass term [the negative of (38) is taken]. The double sign change is of course equivalent to no sign change at all. We may summarize the case of a hole by saying that a hole behaves like a positively charged particle having a positive effective mass tensor.

The essential idea we wish to convey here is that the instantaneous velocity and the instantaneous acceleration of an electron or a hole subjected to an applied field can be determined directly from the energy band structure. These attributes are among the few attributes of electrons and holes which can be obtained in the absence of any knowledge concerning the wave functions for these particles. Most of the properties of electrons and holes cannot be determined unless the wave functions for these particles are known.

VII. Spin-Orbit Interaction

In this section we describe spin-orbit interaction and its effect on energy band structure of a crystal [2].

The interaction between the magnetic dipole field associated with the spin of an electron and the magnetic dipole field associated with the orbital motion of the same electron is known as the spin-orbit interaction. The energy associated with the spin-orbit interaction, or the spin-orbit energy, is usually a small fraction of the total energy of an electron. The gross features of the band structure of a crystal can be obtained even if the spin-orbit interaction is neglected since the spin-orbit energy is so small. However, the fine details in an energy band scheme cannot be derived without taking the spin-orbit energy into account.

In Section III, where we developed the theory of energy bands, the spin-orbit interaction was omitted in the interest of mathematical simplicity. Since the effect of the spin-orbit interaction is actually quite small, it is permissible to solve for the energy band structure with the spin-orbit interaction neglected, and then to take account of the spin-orbit interaction by perturbation methods. When such a program is carried out, the energy level $E_\beta^{\mathrm{modified}}(k)$ in the modified (improved) energy band scheme can be related to the corresponding energy level $E_\beta^{\mathrm{original}}(k)$ in the original band scheme in the following manner:

$$E_\beta^{\mathrm{modified}}(k) = E_\beta^{\mathrm{original}}(k) + \Delta E_\beta(k), \quad (39)$$

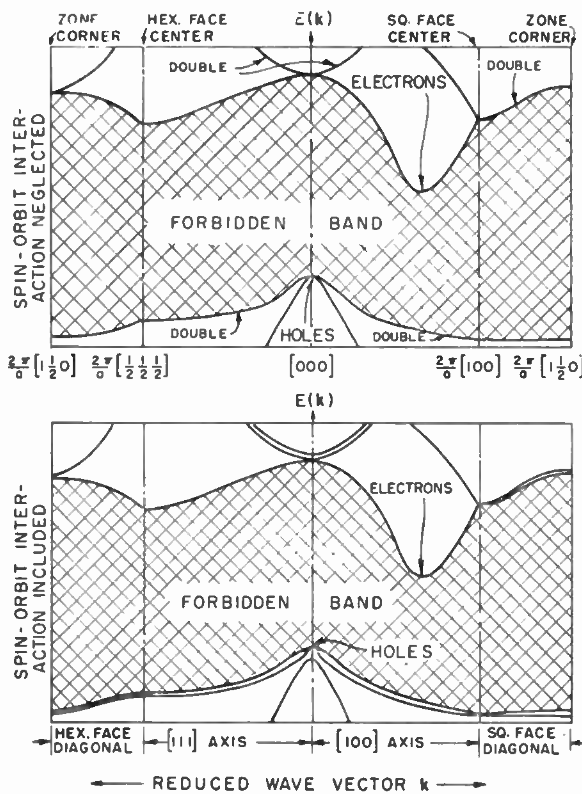


Fig. 9—Schematic diagrams of the energy band structure of silicon. The band scheme with the spin-orbit interaction neglected is shown at the top, and the band scheme with the spin-orbit interaction taken into account is shown at the bottom. The conduction band edge occurs along the 100 axes, while the valence band edge occurs at the central zone point. The width of the forbidden band is 1.08 ev at room temperature. The spin-orbit splitting (in the lower drawing) at the valence band edge is probably about 0.035 ev. In the interest of clarity, the spin-orbit splitting is not shown to scale.

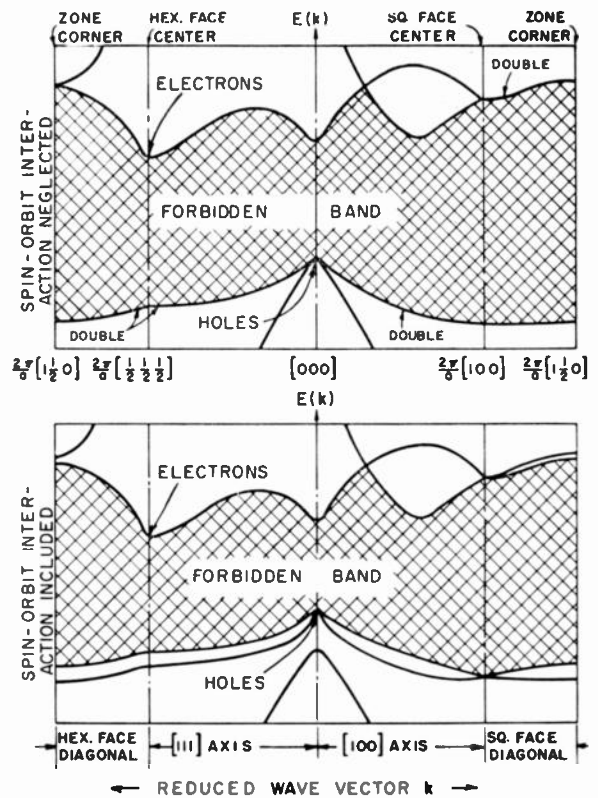


Fig. 10—Schematic diagrams of the energy band structure of germanium. The spin-orbit splitting is omitted from the upper figure, but included in the lower figure. The conduction band edge occurs at the hexagonal face centers, while the valence band edge occurs at the central zone point. The width of the forbidden band is 0.65 ev at room temperature. The spin-orbit splitting at the valence band edge is approximately 0.28 ev. The lowest conduction band has three types of minima: four 111 minima (at the band edge); one 000 minimum; and six 100 minima. The 000 minimum lies 0.1 ev (or slightly more) above the 111 minima; the 100 minima lie 0.18 ev above the 111 minima.

where $\Delta E_\beta(k)$ is the change in the energy level for the quantum state β, k due to the spin-orbit interaction. The various expressions given previously for the velocity and acceleration of an electron, for the effective mass tensor, etc., remain valid for the modified energy band scheme.

In crystals having a center of inversion, such as the diamond-type crystals silicon and germanium, the number of energy bands does not change when the spin-orbit interaction is taken into consideration. In such crystals, each quantum state β, k can be occupied by no more than two electrons, both in the original and in the modified energy band scheme. This follows from the Pauli Exclusion Principle. In the original band scheme, two electrons must have oppositely oriented spins if they occupy the same (double) quantum state β, k. In the modified band scheme, the situation is more complicated, and we will not attempt to describe it.

If the state β, k in the original band scheme is a nondegenerate state, and if this state is well separated in energy from the neighboring states having the same value of k, the spin-orbit shift $\Delta E_\beta(k)$ [cf. (39)] will not produce any marked effect apart from the shift itself. On the other hand, consider two states β, k and β', k which are degenerate in the original band scheme; *i.e.*, $E_\beta^{\text{original}}(k) = E_{\beta'}^{\text{original}}(k)$. In some cases, the shifts $\Delta E_\beta(k)$, $\Delta E_{\beta'}(k)$ will be equal, and again nothing remarkable happens. However, if the shifts in the two states due to the spin-orbit interaction are different, the energy degeneracy will be resolved, and the two states will be separated in energy in the modified band scheme. The separation in energy produced by the spin-orbit interaction is called the spin-orbit splitting.

It can be shown that the detailed nature of the energy band structure is strongly modified by the spin-orbit interaction in a region where the spin-orbit interaction is responsible for the resolution of an energy degeneracy. The detailed nature of the energy band structure is usually modified only slightly in a region where the spin-orbit interaction does not resolve an energy degeneracy, and in a region where there is no energy degeneracy to resolve.

To set the stage for the remaining sections, and to illustrate the general nature of spin-orbit splitting, we show in Figs. 9 and 10, above, our best estimates of the energy band structures of silicon and germanium. These

estimates are based partly on the available experimental evidence [3], and partly on the theoretical calculations performed by the author [4] and by others [5]. The drawings are schematic, rather than exact.

What we have actually shown in Figs. 9 and 10 are the energy profiles of the higher-lying valence bands and the lower-lying conduction bands along the 100 and 111 axes of the reduced zone. The energy profiles for these bands along the diagonals of the square and the hexagonal faces of the reduced zone are also shown. The profiles marked "double" represent the profiles of two energy bands which are degenerate in the particular directions shown.

For the convenience of the reader, we have again drawn the reduced zone in Fig. 11, with certain points conspicuously marked.

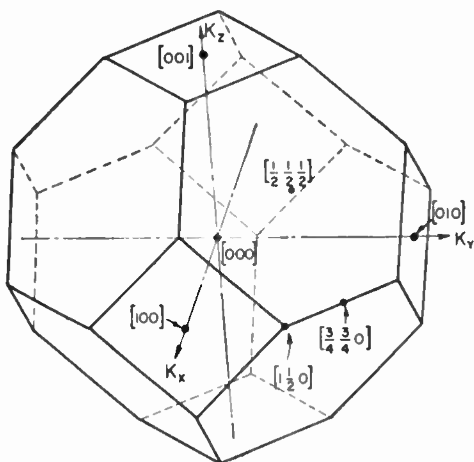


Fig. 11—Reduced zone for silicon and germanium. The factor $(2\pi/a)$ has been omitted in labeling the various points in the reduced zone.

The band schemes for silicon and germanium are displayed according to the original scheme (spin-orbit interaction neglected), and also according to the modified scheme (spin-orbit interaction included), for the purposes of comparison. The modified band schemes correspond to reality; the original band schemes are merely approximations to the true state of affairs.

The more interesting features of the band structures of silicon and germanium, with the spin-orbit interaction taken into account, will be considered in the next three sections.

VIII. Band Structure of Silicon and Germanium: Recent History

In this section, we will indicate some of the landmarks in the recent research which has led to our present understanding of the energy band structure of silicon and germanium, particularly the energy band structure in the neighborhood of the valence and the conduction band edges.

A few definitions are in order: By a band edge we will mean the position or positions in the reduced zone at which an energy band function $E_B(\mathbf{k})$ reaches its minimum or its maximum value or values. Suppose a given energy band has an extremum at $\mathbf{k}=\mathbf{k}_0$. The region of the energy band function associated with values of $\mathbf{k}$ lying close to $\mathbf{k}_0$ will be called the neighborhood of the band edge at $\mathbf{k}_0$. If an energy band is nondegenerate at its extremum or extrema, the corresponding band edge is defined by a single energy band, *i.e*, by the energy band in question. A band edge is defined by two energy bands if the two energy bands are degenerate at their common extremum or extrema. The terms valence and conduction band edges will always refer to the edges of the highest valence band and the lowest conduction band.

The early work [6] on the electrical and the optical properties of silicon and germanium was based on the following model of the valence and the conduction bands: (*a*) The valence and the conduction band edges were both assumed to occur at the central point of the reduced zone; (*b*) The valence and the conduction band edges were each defined by a single energy band; (*c*) The constant energy surfaces near the valence and the conduction band edges were spherical in form.

This model—which we will call the simple model—was adopted for two very good reasons. First, the model offered decided mathematical advantages. Most of the theoretical calculations associated with the electrical and the optical properties could be carried out in a relatively straightforward manner in terms of this simple model. Second, the intrinsic properties of silicon and germanium were not known with sufficient accuracy to warrant the use of a more refined model. Up until about five years ago, the simple model proved adequate for most purposes.

The situation began to change as soon as careful measurements were performed on highly purified single crystals of silicon and germanium. It was found that the results of such measurements did not agree in certain respects with the theoretical predictions stemming from the simple model. The large and anisotropic magnetoresistance effect observed in germanium by Pearson and Suhl [7], and the mobility anomalies obtained by Pearson, Haynes, and Shockley [8] for the same crystal suggested that the simple model was not the correct one. Shockley [9] believed that the experimental data might be interpreted as evidence for the following model: The valence and the conduction band edges were each defined, not by one spherical energy surface, but by three warped energy surfaces which came into contact with each other at the band edge. According to this picture, the valence and the conduction band edges both occurred at the central point of the reduced zone. Since this three band model had many attractive features, it was adopted in many circles.

An alternate possibility for the conduction band structure was suggested by the work of Herman [10] and Herman and Callaway [11]. Their theoretical calculations indicated that the conduction band edge in diamond-type crystals might not occur at the central

zone point at all, but, rather, at a number of equivalent points away from the zone center. The most likely possibility was a conduction band having six equivalent minima located at six symmetrically equivalent points lying along the 100 axes of the reduced zone. Since the calculation for germanium [11] was based on rather crude assumptions, the results were not very reliable. However, this work pointed the way to a new type of model, which, for convenience, we will call the one-band, many-valley model.

[The qualifying term "one-band" is used to indicate that each of the equivalent minima (or valleys) is defined by a single energy band. In the case of a one-band many-valley model—which is also known as a non-degenerate, many-valley model—the constant energy surfaces near each of the equivalent minima are ellipsoidal in form. On the other hand, if two bands were to come into contact with each other at the various positions in the reduced zone of the conduction band edge, the structure would be called a two-band, many-valley model, or a degenerate, many-valley model. Here, the constant energy surfaces near each minimum would consist of two families of warped surfaces.]

The work of Herman [10] and Herman and Callaway [11] indicated that the three-band model was the correct one for the valence band structure of diamond, silicon, and germanium.

Shortly after this work was reported, it occurred to Shockley [12] that the method of cyclotron resonance absorption might be used to determine the nature and form of the valence and conduction band edges in a semiconductor such as germanium. In a cyclotron resonance experiment, a microwave electric field is applied to a sample in a direction perpendicular to an external magnetic field $\mathcal{H}$. Under the influence of the magnetic field, the charge carriers (electrons or holes) rotate in closed orbits. An absorption occurs when the angular frequency of the microwave field, ω, corresponds to the rotation frequency or cyclotron frequency of the charge carriers. Under these conditions, the charge carriers pick up energy from the electric field. The rotation frequency at resonance is given by:

$$\omega = \frac{e\mathcal{H}}{m^*c} \cdot \qquad (40)$$

Three cases are to be distinguished. [For simplicity, we will consider the resonance associated with the charge carriers belonging to a single band.] (*a*) Spherical energy surfaces (the simple model): In this case, there will be only one absorption peak. The frequency at which this peak occurs is independent of the orientation of the crystal with respect to the applied field directions. Here, the quantity m^* is a direct measure of the (scalar) effective mass of the carriers. (*b*) Ellipsoidal energy surfaces (one-band, many-valley model): Let a, b, and c be the direction cosines between the direction of the applied magnetic field and the principal axes of a particular elliposid, and let m_a^*, m_b^*, and m_c^* be the principal components of the effective mass tensor for this ellipsoid. Resonance due to carriers rotation associated with this ellipsoid occurs at frequency $\omega = e\mathcal{H}/m^*c$, where:

$$m^* = \left[\frac{m_a^* m_b^* m_c^*}{a^2 m_a^* + b^2 m_b^* + c^2 m_c^*}\right]^{1/2}. \qquad (41)$$

For a general orientation of the crystal relative to the applied field directions, there will be a different resonance peak for the carriers in each of the several ellipsoids. The principal components of the effective mass tensor, and the orientation of the various ellipsoids relative to the axes of the reduced zone, can be determined by examining the behavior of the absorption spectrum as the orientation of the crystal relative to the applied field directions is changed. (*c*) Warped energy surfaces (three-band model; two-band, one- or many-valley model): In this case, no general relation between the effective mass m^* appearing in (40) and the components of the mass tensor can be stated. However, if an analytical form involving certain arbitrary constants can be assumed for the energy surfaces on the basis of theoretical considerations, the values of these constants can be determined by examining the absorption spectrum, particularly its dependence on the orientation of the crystal in the resonant cavity.

A great deal of pertinent information bearing on the structure of a band near a band edge can be deduced by cyclotron resonance absorption measurements because the absorption frequencies are directly related to the curvature of the energy band surface near the band edge. [The effective mass tensor measures the curvature of an energy band at its point of definition.]

The first successful cyclotron resonance measurements on germanium and silicon were performed by the Berkeley and the Lincoln Laboratory groups [13]. The results of these experiments indicated that the one-band, many-valley model was the correct one for the conduction bands of both germanium and silicon. In the case of germanium, it was found that the equivalent minima were located along the 111 axes of the reduced zone, while in the case of silicon, the equivalent minima were located along the 100 axes of the reduced zone. Further details will be given in a subsequent section.

At about the same time that the cyclotron resonance work on germanium was being done, Meiboom and Abeles [14] succeeded in establishing—by a theoretical analysis of the magnetoresistance data of Pearson and Suhl [7]—that the conduction band edge in germanium occurred at a number of equivalent points lying along the 111 axes. The Meiboom and Abeles result was in substantial agreement with that obtained by the cyclotron resonance absorption measurements, both as regards the orientation of the ellipsoids relative to the axes of the reduced zone, and the ratio of the longitudinal and transverse effective mass components. Subsequent magnetoresistance measurements on silicon by Pearson and Herring [15] confirmed cyclotron resonance result for conduction band structure of silicon.

WRH

The magnetoresistance effect is the change in the electrical resistance due to the presence of a magnetic field. The magnitude of the change is influenced by the relative orientation of the crystal axes, the magnetic field direction, and the direction of the electrical field which is responsible for the current flow. The anisotropy of the magnetoresistance effect is closely related to the anisotropy of the energy band containing the charge carriers. Under favorable conditions, it is possible to infer the nature of the anisotropy of the energy band from the nature of the anisotropy of the magnetoresistance effect. However, the interpretation of magnetoresistance data is not as clear-cut as is that of cyclotron resonance data.

Another experiment which is capable of shedding some light on the structure of energy bands is the piezoresistance effect. This is the change in the electrical resistance due to a shearing stress, *i.e.*, a nonhydrostatic stress. The piezoresistance effect in silicon and germanium was studied by Smith [16]. Smith interpreted his data for silicon as evidence for a conduction band structure involving 100 ellipsoids. (He was not able to make unique predictions for the remaining cases. However, in the light of what we now know about the band structures of silicon and germanium, all of Smith's results appear quite reasonable.)

We will now leave the conduction band structure and turn to the valence band structure.

The validity of the three-band model for the valence structure of germanium was first brought into serious question by the infrared absorption measurements performed by Briggs and Fletcher [17], and by Kaiser, Collins, and Fan [18]. These investigators found three temperature dependent peaks in the absorption spectrum which they attributed to inter-valence band hole transitions. However, this fine structure in the absorption spectrum could not be reconciled with the three-band model. The nature of the valence structure of germanium remained somewhat of a mystery until the cyclotron resonance absorption measurements by the Berkeley and the Lincoln Laboratory groups [13] provided the essential clue.

The cyclotron resonance absorption work [13] indicated that the valence band edge in germanium occurred at the central point of the reduced zone, and that it was defined by two bands which came into contact with each other at the central zone point. This result implied that the three-band model was a correct first approximation, but that an essential ingredient had been left out of the theory of energy bands in its then current state, namely, the effect of the spin-orbit interaction.

It was soon established by Dresselhaus, Kip, and Kittel [19], and by Elliott [20], that when the spin-orbit interaction is properly taken into account, the three top-most valence bands in germanium break up into two higher-lying bands, and one lower-lying band, in the neighborhood of the central zone point. Thus, the valence band edge is actually defined by two bands, rather than three. The two-and-one band model provided a natural explanation for the anomalous results obtained by the infrared absorption measurements [17, 18]. Kahn [21] was able to fit the experimental data of Briggs and Fletcher very nicely to a theory he developed based on the two-and-one band model.

Cyclotron resonance absorption measurements on silicon [13] indicated that the two-and-one band model also applied to the valence structure of this crystal. Although this result has not been definitely confirmed by other types of measurements, the validity of this result does not seem open to question.

To summarize: The conduction bands of silicon and germanium are described by the one-band, many-valley model, while the valence structures of these crystals are described by the two-and-one band model.

IX. Band Structure Near the Valence Band Edge in Silicon and Germanium

As we have already noted in the previous section, the band structure near the valence band edge in silicon and germanium was determined by cyclotron resonance absorption measurements. The experimental evidence indicates that the valence band edge in both crystals occurs at the central point of the reduced zone, and further, that the valence band edge is defined by two bands which come into contact with each other at the central zone point. In the neighborhood of the band edge, the $\boldsymbol{k}$-dependence of the energy surfaces for these two bands is given by the following expression:

$$E(k) = -\frac{h^2}{8\pi^2 m}\left[Ak^2 \pm \sqrt{B^2k^4 + C^2(k_x^2k_y^2 + k_y^2k_z^2 + k_z^2k_x^2)}\right], \quad (42)$$

where A, B, and C are constants, and where k_x, k_y, and k_z are the cartesian coordinates of the reduced wave vector $\boldsymbol{k}$. The magnitude of $\boldsymbol{k}$ is denoted by k. In writing down this expression, we have placed the zero of energy at the valence band edge.

The negative and positive roots correspond to the highest and to the second highest valence bands, respectively. We will refer to the former as $V1$, and to the latter as $V2$. The experimental values for A, B, and C are as follows:

Silicon [22]:	$A = 4.1$	$B = 1.4$	$C = 3.7$
Germanium [22]:	$A = 13.0$	$B = 8.7$	$C = 11.4$.

The form of (42) was suggested by the work of Dresselhaus, Kip, and Kittel [19], and of Elliott [20], on the theory of spin-orbit splitting in diamond-type crystals.

In the region of the valence band edge, the constant energy surfaces for $V1$ and $V2$ are warped surfaces. The warped surfaces for $V1$ are anisotropic, *i.e.*, nonspherical, to an appreciable extent. The warped surfaces for $V2$ are nearly spherical; that is, their anisotropy is quite small. These remarks apply both to silicon and to germanium.

In those applications where mathematical simplicity is of the utmost importance, it is desirable to neglect the warping and treat the surfaces as though they were spherical. The warped surfaces can be approximated by spherical surfaces having the same average curvature by writing:

$$E(k) \approx -\frac{h^2}{8\pi^2 m}[A \pm \sqrt{B^2 + C^2/6}]k^2. \quad (43)$$

The plus and minus signs refer to $V2$ and $V1$, respectively. Use of this spherical approximation introduces a negligible error in the case of $V2$, but a significant error in the case of $V1$.

Since the constant energy surfaces for $V1$ and $V2$ are warped, there will be a spread in the effective masses for the holes occupying these two valence bands. The spread in the effective mass values is appreciable for the holes in $V1$, but quite small for the holes in $V2$. The spherical approximation given above can be used to determine average values for the effective masses of the holes in $V1$ and $V2$. The results are as follows:

Silicon:
$\langle m^*_{V1}\rangle = 0.49\ m$ $\langle m^*_{V2}\rangle = 0.16\ m$ Ratio $= 3.1$

Germanium:
$\langle m^*_{V1}\rangle = 0.28\ m$ $\langle m^*_{V2}\rangle = 0.044\ m$ Ratio $= 6.4$

A third valence band, $V3$, is separated from $V1$ and $V2$ by the effect of the spin-orbit interaction. The maximum of $V3$ lies slightly below the common maximum of $V1$ and $V2$ at the central point of the reduced zone. Denoting the magnitude of the spin-orbit splitting at this point by the symbol E_{so}, we may describe the k-dependence of the energy surfaces for $V3$ near the band edge of $V3$ by the following expression [19, 20]:

$$E(k) = -E_{s0} - (h^2/8\pi^2 m)A k^2. \quad (44)$$

The constant energy surfaces for $V3$ are strictly spherical near the band edge. The effective masses for the holes in $V3$ in silicon and germanium are readily determined; they are:

Silicon: $m_{V3}^* = 0.24\ m$. Germanium: $m_{V3}^* = 0.077\ m$.

The best current estimates for E_{s0} are [21]:

Silicon: $E_{s0} = 0.035$ ev; Germanium: $E_{s0} = 0.28$ ev.

The most remarkable feature of the valence band structure of silicon and germanium is the presence of three valence bands near the valence band edge. This means that there are three classes of holes, corresponding to unoccupied states in the three valence bands $V1$, $V2$, and $V3$. The holes in each class have different dynamical characteristics. For example, the effective masses for the three types of holes are different. Therefore, the holes in $V1$, $V2$, and $V3$ will contribute to the electrical conductivity and to other electronic transport processes in different ways (for a further discussion, see references to Section XIV).

What are the relative hole populations in the three valence bands? At temperatures for which $kT \ll E_{s0}$ [E_{s0} is the spin-orbit splitting at the central zone point], the number of holes in $V3$ will be negligible compared with the number of holes in either $V1$ or $V2$. There will be an appreciable hole population in $V3$ only when kT is of the order of or greater than E_{s0}. At room temperature, where $kT = 0.025$ ev, there will be an appreciable number of $V3$ holes in silicon, but virtually no $V3$ holes in germanium. In the case of silicon, all three types of holes will be involved in electrical conduction at normal temperatures. In germanium, only two types of holes, those in $V1$ and $V2$, will contribute to the electrical conductivity at normal temperatures.

X. Band Structure Near the Conduction Band Edge in Silicon and Germanium

The conduction band edge occurs along the 100 axes of the reduced zone in silicon, and along the 111 axes in germanium. The constant energy surfaces are prolate ellipsoids of revolution near each conduction band minimum in both crystals. The longitudinal and transverse effective mass components, m_l^* and m_t^*, for the electrons in silicon and germanium have the following values at 4°K, the temperature at which cyclotron resonance experiments are performed:

Silicon [22]:
$m_l^* = 0.98\ m$ $m_t^* = 0.19\ m$ $m_l^*/m_t^* = 5.2$;

Germanium [22]:
$m_l^* = 1.57\ m$ $m_t^* = 0.082\ m$ $m_l^*/m_t^* = 19.0$.

Here, as elsewhere in this paper, the mass of an electron in free space is represented by the symbol m. Since the energy band structure varies slightly with temperature, the effective mass components are actually temperature dependent quantities. The room temperature values for m_l^* and m_t^* will be somewhat different from the values quoted above. It is possible to obtain values for the effective mass ratio m_l^*/m_t^* over a wide temperature range from magnetoresistance measurements [14, 15]. However, the values obtained in this manner are not as accurate as those obtained by cyclotron resonance experiments.

Cyclotron resonance and magnetoresistance experiments do not provide information bearing on the exact location of the conduction band minima in the reduced zone; these experiments merely indicate along which axes of the reduced zone the minima must lie. In order to determine the exact location of the minima along the appropriate axes of the reduced zone, it is necessary to turn to other types of measurements.

In the case of silicon, there are two lines of experimental evidence [23, 24] which suggest that the minima occur along the 100 axes within the zone, rather than at the square face centers, *i.e.*, at the intersection of the 100 axes with the zone boundaries. [Theoretical considerations support this view.] Kohn's analysis [23] indicates that the minima lie between $\frac{1}{2}$ and $\frac{3}{4}$ of the way

from the zone center to zone boundaries; that of Macfarlane and Roberts [24] suggests that the conduction band minima lie roughly 7/9 of the way from the central zone point to the square face centers. Therefore, it appears that there are six equivalent minima, rather than three, as would be the case if the minima were to occur at the square face centers.

In the case of germanium, the experimental evidence is conflicting. The work of Enz [25] and Conwell [26] suggests that the minima are more likely to occur at the hexagonal face centers than within the reduced zone along the 111 axes. [Our own work [4] supports this view.] On the other hand, the analysis of Macfarlane and Roberts [27] indicates that the conduction band minima lie roughly $\frac{2}{3}$ of the way from the central zone point to the hexagonal face centers. Since we favor the former alternative, we have placed the conduction band edge in germanium at the hexagonal face centers in Fig. 10. If this is correct, as we believe it is, the conduction band edge occurs at four equivalent points, rather than eight. In other words, there are four minima, rather than eight, at the conduction band edge.

The width of the forbidden band, E_{gap}, can be determined by several methods [28]. One of the most common involves a measurement of the change in the electrical conductivity with temperature. The conductivity is proportional to the number of electrons in the conduction band and holes in the valence bands, and the number of carriers (electrons and holes) in turn depends on the ratio of the width of the forbidden band to the temperature [cf. Appendix]. Hence, the width of the forbidden band can be obtained by studying the temperature dependence of the electrical conductivity.

Another common method for determining E_{gap} is to measure the spectral distribution of the optical absorption. The photon energy at the onset of absorption is a measure of the band gap. It is generally found that the values for E_{gap} determined by electrical conductivity and optical absorption measurements are in substantial agreement. Since the band structure changes with temperature, the width of the forbidden band is a temperature-dependent quantity. [The change in the band structure with temperature is due to the thermal expansion of the lattice.] Typical results for E_{gap} at room temperature are as follows:

Silicon [24]: $E_{gap}=1.08$ ev. Germanium [27]: $E_{gap}=0.65$ ev.

According to our best estimates, and to the available experimental evidence [29, 30], the conduction band in silicon approaches the valence band closely only in the neighborhood of the 100 minima. In the case of germanium, the conduction band appears to have three different types of minima; all three types of minima are not far removed from the valence band maximum on an energy scale. Optical measurements [29–31] suggest that the conduction band has a minimum at the central point of the reduced zone, and that this minimum lies 0.1 ev (or slightly more) above the 111 minima—which define the true conduction band edge. Our interpretation [32] of the observed variation [33] of the optical absorption edge with composition in germanium-silicon alloys [cf. Section XI] leads us to postulate the existence of six minima lying along the 100 axes of the reduced zone. We estimate the energy separation between the six 100 minima and the (four) lower-lying 111 minima to be about 0.18 ev. The 111 minima, the 000 minimum, and the 100 minima are all shown in Fig. 10. Thus, the conduction band in germanium approaches the valence band closely at points in the reduced zone other than the points at which the conduction band edge occurs.

The free, *i.e.*, the conduction band electrons in silicon will occupy states near the six 100 conduction band minima over the entire temperature range from absolute zero to the melting point. At low and moderate temperatures, the vast majority of the free electrons in germanium will occupy states near the four 111 conduction band minima. At elevated temperatures, a small fraction of the electrons will occupy states near the 000 minimum, and a still smaller fraction will occupy states near the six 100 minima. The electrons in different types of minima will have different effective masses and hence different dynamical characteristics. However, the electrons in the 111 minima will exert the dominating influence over such processes as the electrical conductivity over the entire temperature range, by reason of their superior numbers.

XI. Band Structure of Germanium-Silicon Alloys

Since silicon and germanium are both diamond-type crystals, and since these two crystals are formed from elements occupying adjacent positions in column IV of the periodic table, one should expect that the energy band structures of silicon and germanium are related to each other in some manner. Some insight into this relationship can be gained by examining the solid substitutional alloys of germanium and silicon over a range of composition starting with pure germanium and extending to pure silicon. In a germanium-silicon solid substitutional alloy, the germanium and silicon atoms are arranged at random at the lattice sites of a diamond-type lattice. Christian [34] has succeeded in growing single crystals of germanium-silicon alloys of various compositions in the germanium-rich and silicon-rich portions of the composition range.

Because an alloy is not a perfect crystal, but rather a disordered crystal [cf. Section I], it does not have an energy band structure in the strict sense. However, one can speak of the energy band structure for an alloy in a loose sense. The electronic energy spectrum of an alloy [35] contains a number of energy intervals in which there is a high density of allowed states; these intervals may be called allowed bands. The energy spectrum for an alloy also contains a number of energy intervals in which there is a low density of allowed states; these are essen-

tially forbidden bands. While the transition from an allowed band to an adjacent forbidden band is sharp in a perfect crystal, the transition is rather gradual in an alloy. One may visualize the electronic energy spectrum of an alloy as consisting of a number of allowed bands [regions of high state densities] having small tails [regions of low state densities] which extend into the forbidden bands [regions of negligibly low state densities].

Since the lattice constants of perfect silicon and germanium crystals differ by only a few per cent, and since germanium and silicon atoms have the same valency, it is likely that the tails in the energy spectrum of a germanium-silicon alloy represent a minor rather than a major feature of the energy spectrum. In the following discussion, we will treat the germanium-silicon alloys as though these alloys have energy spectra (energy band structures) which are very similar to the energy spectra of perfect crystals. We will assume that the optical gap—the photon energy at the onset of optical absorption due to the excitation of electrons across the forbidden band—is a measure of the width of the forbidden band.

The author has proposed a theory of the energy band structure of germanium-silicon alloys in order to explain the observed variation of the optical gap of such alloys with composition. The experimental results are as follows [33]: In the region 0 to 15 mol per cent silicon, the optical gap is found to increase linearly with composition. Between 15 and 100 mol per cent silicon, the variation of the optical gap with composition is nearly linear, but the slope of the optical gap vs composition curve in this region is roughly one half as large as the slope in the former region. Further, there is a sharp break in the optical gap vs composition curve at 15 mol per cent silicon.

The explanation put forward by the author [32] runs as follows: As silicon is added to germanium in the range 0 to 15 mol per cent silicon, the lowest conduction band moves away from the highest valence band. However, the 111 minima in the conduction band move away from the valence band at a more rapid rate than do the 100 minima. In the range just mentioned, the 111 minima lie below the 100 minima, and the variation of the optical gap with composition is determined by the rate at which the 111 minima move away from the valence band edge.

At 15 mol per cent silicon, the 111 and the 100 minima coincide on an energy scale. Between 15 and 100 mol per cent silicon, the 111 minima lie above the 100 minima, and the variation of the optical gap with composition is determined by the rate at which the 100 minima move away from the valence band edge. Since the 111 minima move more rapidly than the 100 minima, the slope of the optical gap vs composition curve is larger in the former range than in the latter.

The 000 conduction band minimum moves away from the valence band edge at a more rapid rate than either the 111 or the 100 minima. Consequently, the 000 minimum does not play a role in determining the optical gap over the entire range of composition from pure germanium to pure silicon. The valence band structure changes continuously over the composition range; the valence band edge remains at the central point of the reduced zone for all compositions.

Magnetoresistance measurements on germanium-silicon alloys have been performed by Glicksman [36] in order to test the validity of the theory described above. Glicksman's results tend to support the theory. Cyclotron resonance absorption measurements on a few samples of germanium-rich and silicon-rich alloys have been reported by the Berkeley group [37]. The cyclotron resonance results are consistent with our theory.

The fact that a theory of germanium-silicon alloys could be proposed solely on the basis of one piece of experimental evidence—the variation of the optical gap with composition—suggests that similar theories can be developed to explain the generic relationships which must exist between the energy band structures of similar or related crystals. The author [38] has taken a first step in the direction of linking the energy band structures of diamond-type crystals on the one hand with those of certain zinc-blende-type crystals on the other.

XII. Lattice Vibrations and Phonons

In this section we will make a brief departure from the theory of electronic energy bands, and discuss the theory of lattice vibrations. We will describe the general nature of lattice vibrations, and the manner in which the normal modes of vibration can be cataloged. It will be seen that the lattice vibrational spectrum has many features in common with the electronic energy band spectrum. The phonon concept will be considered briefly at the end of this section. The information to be presented below will find an immediate application in the next section, where we treat the collisions between electrons or holes and phonons.

The nuclei in a crystal can be set into oscillation about their equilibrium positions by the application of heat, by optical excitation, and by other means. Since the motions of the various nuclei are coupled together, a nuclear vibration will generally involve all the nuclei in the crystal. Any vibration of the nuclear assembly is called a lattice vibration because the nuclei in a crystal are arranged in the form of a lattice.

Any lattice vibration can be resolved into a certain number of normal modes of vibration, just as any periodic function can be analyzed into its harmonic components. Each normal mode describes a possible type of vibration. The total number of normal modes of vibration is equal to the total number of vibrational degrees of freedom of the nuclei in the crystal. Consider a cyclic crystal containing N^3 unit cells, with A nuclei per unit cell. Since each nucleus has 3 vibrational degrees of freedom, and there are AN^3 nuclei in the cyclic crystal, there must be $3AN^3$ vibrational degrees of freedom in all; hence $3AN^3$ normal modes of vibration.

Each normal mode of vibration is essentially an elastic wave. As such, it can be classified according to its frequency, its wavelength, and its direction of propagation. In special cases, it can be classified according to its direction of polarization as well. [If all the nuclei in the crystal vibrate in the same direction, this direction is the direction of polarization; if different nuclei oscillate in different directions, a polarization direction cannot be assigned.]

Each normal mode can be characterized by a reduced wave vector $\boldsymbol{q}$. The wavelength, λ, of a normal mode is given by the expression: $\lambda = 2\pi/q$, where q is the magnitude of $\boldsymbol{q}$. The direction of $\boldsymbol{q}$ defines the direction of propagation of the mode. The reduced wave vector $\boldsymbol{q}$ has the same geometrical properties as the reduced wave vector $\boldsymbol{k}$ employed previously in describing the electronic quantum states. Thus, the reduced wave vector $\boldsymbol{q}$ is restricted to a set of N^3 allowed values. The allowed values of $\boldsymbol{q}$ are uniformly distributed in the reduced zone.

We will denote the angular frequency of a normal mode by the symbol $\omega_\mu(\boldsymbol{q})$, where μ is an integer ranging from 1 to $3A$. [A is the number of nuclei in the unit cell.] For each choice of $\boldsymbol{q}$, the subscript μ labels the angular frequencies according to the following scheme:

$$0 \leq \omega_1(\boldsymbol{q}) \leq \omega_2(\boldsymbol{q}) \cdots \leq \omega_{3A}(\boldsymbol{q}). \tag{45}$$

Since there are N^3 choices of $\boldsymbol{q}$, and $3A$ choices for μ, the $3AN^3$ normal modes are fully accounted for.

It can be shown that $\omega_\mu(\boldsymbol{q})$, regarded as a function of $\boldsymbol{q}$, is a continuous function of $\boldsymbol{q}$ throughout the reduced zone, for each choice of μ. The N^3 normal modes associated with each index μ are said to form a branch of the vibrational spectrum.

The long wavelength modes belonging to the three lowest branches of the vibrational spectrum of any crystal correspond to sound waves. For this reason, these three branches are called the acoustical branches. It can be shown that:

$$\omega_\mu(\boldsymbol{q}) \rightarrow 0 \text{ as } q \rightarrow 0 \text{ for } \mu = 1, 2, 3. \tag{46}$$

In the region of very small q (very long wavelength), $\omega_\mu(\boldsymbol{q})$ is directly proportional to q for each direction of propagation. Hence the phase velocity is a constant for the long wavelength modes of each acoustical branch, for a given direction of propagation. This is a well known characteristic of sound waves.

Since sound waves can generally be classified as either longitudinal or transverse, we can further classify the acoustical branches as follows: The branch containing the longitudinal sound waves is called the longitudinal acoustical (LA) branch; the other two are known as the transverse acoustical (TA) branches.

If a crystal has only one nucleus per unit cell, there will be only three branches in its vibrational spectrum, namely, the three acoustical branches. However, if there is more than one nucleus per unit cell, there will be additional branches. These additional branches are called the optical branches. [In polar crystals, which contain two nuclei per unit cell, the modes in the additional branches can be excited optically; hence the designation "optical branches."] In a crystal having A nuclei per unit cell, there will be $3A-3$ optical branches.

If we examine the function $\omega_\mu(\boldsymbol{q})$ for any optical branch, we find that as q approaches zero, $\omega_\mu(\boldsymbol{q})$ approaches a finite value rather than zero, as was the case for the acoustical branches.

For each mode belonging to the long wavelength region of an acoustical branch, all the nuclei in each unit cell vibrate in the same direction. For each mode belonging to the long wavelength region of an optical branch, the nuclei in each unit cell generally vibrate in different directions. This is an important distinguishing feature of acoustical and optical branches.

Since diamond-type crystals have two nuclei per unit cell, the vibrational spectrum for such crystals contains three acoustical branches and three optical branches. The two nuclei in each unit cell vibrate in opposite directions for any mode belonging to the long wavelength region of any optical branch. If the common direction of vibration is parallel (perpendicular) to the direction of propagation, the mode is termed a longitudinal (transverse) mode. The optical branch containing the longitudinal modes is called the longitudinal optical (LO) branch; the other two are called the transverse optical (TO) branches.

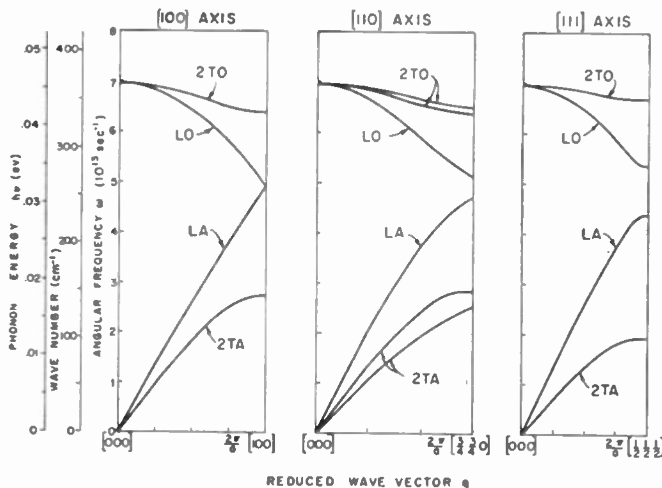


Fig. 12—Lattice vibrational spectrum for germanium, based on the calculations by Hsieh. The transverse branches are degenerate along the 100 and the 111 axes, and are non-degenerate along the 110 axis. The symbols TA, LA, LO, and TO refer to the transverse acoustical, longitudinal acoustical, longitudinal optical, and transverse optical branches, respectively.

The lattice vibrational spectra for silicon and germanium have been calculated by Hsieh [39] on the basis of an approximational method developed by Born [40] and later elaborated by Smith [41]. The lattice vibrational spectrum for germanium, as given by Hsien's work (cf. Table I of reference [39]), is shown in Fig. 12 above. We have plotted $\omega_\mu(q)$ vs $\boldsymbol{q}$ for $\boldsymbol{q}$ lying along the 100, the 110, and the 111 axes of the reduced zone. The six

WRH

branches are labeled in accordance with the notation described above: LA stands for the longitudinal acoustical branch, etc. The transverse branches are degenerate along the 100 and the 111 axes, but not along the 110 axis. The vibrational spectra for silicon and germanium are very similar, at least so far as form is concerned. A good picture of the vibrational spectrum for silicon can be obtained from Fig. 12 by doubling the energy scale. Since the work of Hsieh is based on a number of approximations, the results should be regarded as approximate, rather than exact.

We now return to the general theory. It can be shown that the vibrational energy of each normal mode of vibration is quantized. That is, the vibrational energy of a mode can assume any one of a certain set of discrete values, and no others. The allowed values of the vibrational energy of a normal mode are:

$$(n + 1/2)h\nu \tag{47}$$

where n is a positive integer or zero, and where ν is the frequency of the mode. The state of excitation of the normal mode is specified by the quantum number n.

When a crystal is in thermal equilibrium, the total vibrational energy of the lattice is distributed in a certain manner among the various normal modes. It can be shown that, under conditions of thermal equilibrium, the average value of n, $\langle n\rangle$, of a normal mode having frequency ν is given by Planck's law:

$$\langle n\rangle = \frac{1}{\exp(h\nu/kT) - 1}\cdot \tag{48}$$

The average vibrational energy $\langle E\rangle$ of the mode is given by:

$$\langle E\rangle = [\langle n\rangle + \tfrac{1}{2}]h\nu. \tag{49}$$

The total vibrational energy of the lattice can be obtained by summing $\langle E\rangle$ over all the normal modes.

At the absolute zero of temperature, $\langle n\rangle = 0$ for all the normal modes. At very low temperatures, the only normal modes which are excited to any appreciable extent are those having very low frequencies. These modes are of course the long wavelength acoustical modes. At moderate temperatures, all the acoustical modes become excited. The optical modes do not become excited to any significant extent until the temperature reaches relatively high values.

The vibrational amplitude associated with a normal mode is directly proportional to the square root of the vibrational energy of the mode. As the temperature is raised, the vibrational amplitude of each normal mode increases. It follows that the nuclei will oscillate more and more violently about their equilibrium positions as the temperature is raised. However, the nuclear displacements will generally be quite small compared with an inter-nuclear spacing, over the entire temperature range from absolute zero to the melting point.

It is convenient to describe the state of excitation of the various normal modes in terms of the phonon concept. A phonon represents a unit quantum excitation of a normal mode of vibration. The state of excitation of a given normal mode is given by the quantum number n, which is a measure of the number of phonons contained in the mode. If a normal mode is excited by a unit amount, a phonon is created; if a normal mode is de-excited by a unit amount, a phonon is destroyed. It should be noted that a phonon is not a lattice vibration as such, but rather a unit quantum excitation of some normal mode of vibration.

To sum up: The possible vibrations of the nuclei in a crystal are given by the normal modes of vibration. The normal mode frequencies form a spectrum, the lattice vibrational spectrum. The energy content of each normal mode is quantized, and the state of excitation of a given normal mode is given by the number of phonons (n) contained in the mode. The distribution of phonons, and the distribution of vibrational energy, among the various normal modes, under conditions of thermal equilibrium, are given by the functions (42) and (43), respectively. The vibrational amplitudes of the nuclei in a crystal increase with increasing temperature.

XIII. Collisions Between Electrons or Holes and Phonons

In the earlier sections, we were concerned with the nature of the available electronic quantum states, and with the manner in which the electrons are distributed among these quantum states under equilibrium conditions. In Section VI, we considered the behavior of electrons and holes under nonequilibrium conditions. We showed how the instantaneous velocities and accelerations of electrons and holes, between successive collisions with the lattice, can be determined from a knowledge of the energy band structure. In the previous section, we treated the vibrational quantum states of the lattice vibrations. In the present section, we will tie some loose ends together, and consider the collisions between electrons or holes and the lattice. We will explain how an electron or a hole can be forced from one quantum state to another by suffering a collision with the lattice. In order to avoid duplication, we will only treat electron-lattice collisions. The remarks we are about to make concerning such collisions apply to hole-lattice collisions as well.

Whenever a collision between an electron and the lattice occurs, energy will be transferred from the electron to the lattice, or vice versa. If the lattice gains (loses) energy, one of its normal modes may be excited (de-excited) by a unit amount. While other mechanisms of energy transfer exist, the one just mentioned is by far the most important. Therefore, we will ignore the others, and confine ourselves to the collision process involving the unit excitation or de-excitation of some normal mode.

Consider an electron initially in the state β, $\mathbf{k}$. Let us assume that the electron interacts with the normal mode ν, $\mathbf{q}$, and that as a result of the collision, the elec-

tron makes a transition to some state β', $\boldsymbol{k}'$ which was initially unoccupied. In any collision, energy and momentum must be conserved. In quantum mechanics, these conservation requirements limit the possible transitions between states. The rules governing possible transitions are called selection rules. The momentum conservation rule takes the form:

$$\boldsymbol{k}' = \boldsymbol{k} + \boldsymbol{q} + \boldsymbol{h}, \quad (50)$$

where $\boldsymbol{h}$ is a reciprocal lattice vector. This is a momentum conservation rule because the product of any quantity in (50) and $(h/2\pi)$ has the dimensions of a momentum. The reciprocal lattice vector $\boldsymbol{h}$ appears in (50) to guarantee that $\boldsymbol{k}'$, $\boldsymbol{k}$, and $\boldsymbol{q}$ all lie in the reduced zone or on the boundaries of the reduced zone. Since the propagation (or wave) vector of the electron is changed as a result of the collision, the electron is deflected, or scattered, by its interaction with the normal mode.

The second selection rule is an energy conservation condition: The energy gained or lost by the electron must be equal to the energy lost or gained by the normal mode. That is:

$$E_{\beta'}(\boldsymbol{k}') = E_{\beta'}(\boldsymbol{k} + \boldsymbol{q} + \boldsymbol{h}) = E_{\beta}(\boldsymbol{k}) \pm h\nu, \quad (51)$$

where $E_\beta(\boldsymbol{k})$ and $E_{\beta'}(\boldsymbol{k}')$ are the initial and the final electron energies, respectively. If the normal mode is excited by a unit amount [the minus sign in (51)], a phonon of energy $h\nu$ is created; if the normal mode is de-excited by a unit amount [the plus sign in (51)], a phonon of energy $h\nu$ is destroyed. The interaction between a normal mode and an electron leading to the creation or to the destruction of a phonon is commonly called an electron-phonon collision.

We will confine our remarks in the following discussion to the collisions between electrons and phonons in non-polar crystals such as silicon and germanium. First, we will consider a nonpolar crystal in which the conduction band has a single nondegenerate minimum located at the central point of the reduced zone. [This situation might obtain in the case of grey tin.] In such a crystal, the phonons most likely to be involved in electron-phonon collisions are the long wavelength (longitudinal) acoustical phonons. When an electron collides with a long wavelength acoustical phonon, it will gain or lose a small amount of energy, and it will be deflected only slightly; *i.e.*, its reduced wave vector will change only slightly. [Explanation: Long wavelength acoustical phonons have small reduced wave vectors $\boldsymbol{q}$ and small energies $h\nu$.] Hence, electron will make a transition from its initial state to an adjacent final state.

Second, let us consider a nonpolar crystal in which the conduction band has a number of nondegenerate minima located at well-separated, symmetrically equivalent points in the reduced zone. [This is the case for both silicon and germanium.] In such a crystal, the electrons can collide with both long and short wavelength acoustical phonons under suitable conditions. If an electron collides with a long wavelength acoustical phonon, it will make a transition between two states both of which lie near the bottom of the same conduction band minimum. This type of collision is quite similar to the one described above, in that the electron gains or loses a small amount of energy, and changes its reduced wave vector only slightly. If an electron collides with a short wavelength acoustical phonon, it can make a transition between two states which lie near the bottoms of different conduction band minima. In this case, the electron will be strongly deflected: the short wavelength phonon has a large reduced wave vector $\boldsymbol{q}$. Further, the electron will gain or lose an appreciable amount of energy: the short wavelength phonon has an appreciable energy $h\nu$.

At very low temperatures, the collisions with long wavelength acoustical phonons will account for nearly all the collisions. At moderate and high temperatures, both types of collisions will occur. At high temperatures, collisions with the long and short wavelength optical phonons will also occur. [In polar crystals such as sodium chloride, collisions with (longitudinal) optical phonons are the most important ones at all temperatures.]

The probability per unit time of an electron making a collision is $1/\tau$, where τ is called the collision time. The collision time for an electron will depend upon its dynamical characteristics, such as its energy, its effective mass, and its direction of propagation. These in turn depend upon the state it occupies prior to the collision. The collision time will usually exhibit a marked temperature dependence. In general, the collision time will decrease as the temperature is increased.

In the next section we will discuss briefly the theory of electrical conductivity and show how the electron-phonon collisions provide a dissipative mechanism whereby the energy gained by electrons from an applied electric field is transferred, in an irreversible manner, to the lattice. The energy thus gained by the lattice appears in the form of heat energy. We will see that the collision time plays a central role in the theory of electrical conductivity.

XIV. Electrical Conductivity

In this section we will show how the electrical conductivity of a crystal is related to the quantum states which the charge carriers occupy, and to the collision times of these carriers. We will also give references to recent work dealing with the electrical properties of silicon and germanium.

When a crystal is in a state of thermal equilibrium, its electrons are distributed among the available quantum states in the manner described in Section IV. In a nonmetallic crystal, there will be a certain number of electrons in the conduction bands, and a like number of holes in the valence bands. In the interest of simplicity, we will confine our attention to the conduction band electrons, and assume all the electrons occupy states in the lowest conduction band. Since we will deal only with a single band, we may suppress the band index β.

Under equilibrium conditions, the current density due to the motion of the electrons (in the lowest conduction band) is given by the following expression:

$$J_0 = -(2)(1/2\pi)^3(e)\int v(k)f_0(k)dk, \quad (52)$$

where $dk = dk_1dk_2dk_3$ [cf. (12) above]. The integral is carried over the entire reduced zone. The factor 2 in front of the integral sign takes the double occupancy of each quantum state k into account. $v(k)$ is the velocity of an electron in the state k. We will assume that the electronic distribution among the available conduction band states is governed by the Maxwell-Boltzmann or the classical statistics. Thus, we have:

$$f_0(k) = \exp\left[-\{E(k) - \xi\}/kT\right], \quad (53)$$

where $E(k)$ is the energy of an electron in the state k, and where ξ is the Fermi energy.

It is a simple matter to prove that J_0 is identically zero. The proof involves the following considerations: (*a*) The velocity $v(k)$ is equal to $(2\pi/h)$ $\text{grad}_k E(k)$ [cf. (34) above]. (*b*) The energy function $E(k)$ satisfies the symmetry property $E(k) = E(-k)$ [cf. (19) above]. (*c*) $v(k) = -v(-k)$ since $\text{grad}_k E(k) = -\text{grad}_k E(-k)$. (*d*) To each electron in a state k, there corresponds an electron in the state $-k$; *i.e.*, $f_0(k) = f_0(-k)$. It follows that the currents associated with the motions of the various electrons cancel in pairs, leaving a zero net current. What we have shown is that there is no current flow when the crystal is in thermal equilibrium.

Suppose a steady electric field $\mathcal{E}$ is applied to the crystal. In this case, the crystal will be in a state of non-equilibrium, and the electronic distribution will depart from $f_0(k)$, the equilibrium distribution function. We will label the non-equilibrium states by $k[=k(t)]$, and we will denote the non-equilibrium distribution function by $f(k)$. For convenience, we will suppress the time dependence of k, and write $k(t)$ simply as k.

In the presence of the field $\mathcal{E}$, there is a nonvanishing electric current density given by

$$J = -(e/4\pi^3)\int v(k)f(k)dk \quad (54)$$

It should be noted that while $f_0(k) = f_0(-k_n)$, $f(k) \neq f(-k)$. As we will see below, it is possible to express J as follows:

$$J = \sigma\mathcal{E}. \quad (55)$$

This relation defines σ, the conductivity tensor. In cubic-type crystals such as silicon and germanium, the nondiagonal components of σ vanish identically, and the electrical conductivity tensor reduces to a scalar quantity. In other words, the current flow will be parallel to the direction of the applied field in silicon and germanium.

The electrical conductivity associated with the conduction band electrons can be determined if the energy band structure is known, *i.e.*, if $v(k)$ is a known function, and if the nonequilibrium distribution function $f(k)$ is known. For the sake of argument, we will assume that $v(k)$ is given. We will now consider the manner in which $f(k)$ is evaluated.

The distribution of the electrons among the available quantum states may vary because the electrons are accelerated by the applied field during their motion through the crystal. The distribution may also vary as a result of the collisions between the electrons and the phonons. The rate of change of $f(k)$ due to the action of the applied field is given by the expression: $-e(2\pi/h)\mathcal{E}\cdot\text{grad}_k f(k)$. We will denote the rate of change of $f(k)$ due to collisions by the symbol $[\partial f(k)/\partial t]_{\text{coll}}$. Under steady state conditions, the following equality must be satisfied:

$$-e(2\pi/h)\mathcal{E}\cdot\text{grad}_k f(k) = \left[\frac{\partial f(k)}{\partial t}\right]_{\text{coll}}. \quad (56)$$

The collision term is generally a complicated integral which can be explicitly evaluated only if the collision mechanism is known in sufficient detail.

In many cases of physical interest, it is possible to express the right-hand-side of (56) in terms of the departure from equilibrium in the following manner:

$$[\partial f(k)/\partial t]_{\text{coll}} = -[f(k) - f_0(k)]/\tau(k). \quad (57)$$

This relation defines the quantity $\tau(k)$, which is called the relaxation time for an electron in the state k. In order to establish the physical significance of the relaxation time $\tau(k)$, we will assume that the distribution $f(k)$ is set up by an applied force which is suddenly removed at time $t=0$. The rate of approach to equilibrium is then given by:

$$\partial[f(k) - f_0(k)]/\partial t = -[f(k) - f_0(k)]/\tau(k). \quad (58)$$

The solution of this differential equation has the form:

$$[f(k) - f_0(k)]_t = [f(k) - f_0(k)]_{t=0}\exp[-t/\tau(k)]. \quad (59)$$

Thus, $\tau(k)$ is the time constant of the decay of the system from its initial, nonequilibrium state to its final, equilibrium state. The relaxation time $\tau(k)$ is essentially the collision time defined at the end of the previous section. The terms relaxation time and collision time can both be used to refer to the quantity defined by (57) above.

We will assume that a relaxation time can be defined in the sense of (57); that is, we will assume that the collision term appearing in (56) can be expressed in the form (57). Thus, we may write:

$$-(2\pi e/h)\,\mathcal{E}\cdot\text{grad}_k f(k) = -[f(k) - f_0(k)]/\tau(k). \quad (60)$$

This equation, which is known as the Boltzmann transport equation, may be rewritten as:

$$f(k) = f_0(k) + (2\pi e/h)\tau(k)\mathcal{E}\cdot\text{grad}_k f(k). \quad (61)$$

Since $f(k)$ departs only slightly from $f_0(k)$ for the fields $\mathcal{E}$ normally of interest, we may replace $f(k)$ by $f_0(k)$ on the right-hand-side of (61). This gives:

$$\begin{aligned} f(k) &= f_0(k) + (2\pi e/h)\tau(k)\mathcal{E}\cdot \text{grad}_k f_0(k) \\ &= f_0(k) + e\tau(k)v(k)\cdot \mathcal{E}\partial f_0(k)/\partial E(k). \end{aligned} \tag{62}$$

Writing:

$$J_i = \sum_j \sigma_{ij}\mathcal{E}_j; \qquad i, j = x, y, z, \tag{63}$$

we have for the ij component of the electrical conductivity tensor:

$$\sigma_{ij} = -(e^2/4\pi^3)\int \tau(k)v_i(k)v_j(k)\partial f_0(k)/\partial E(k)dk. \tag{64}$$

The electrical conductivity component σ_{ij} depends upon the weighted average of the product $\tau(k)v_i(k)v_j(k)$, the energy gradient of the function $f_0(k)$ serving as the weighting function. It is not difficult to show that the electrical conductivity is directly proportional to the number of electrons in the conduction band. The temperature dependence of the electrical conductivity is due to the change in the number of electrons with temperature, and to a lesser degree to the change in the relaxation time with temperature. The electrical conductivity and its temperature dependence can be determined if $\tau(k)$ and $E(k)$ are known.

We have gone through the preceding analysis not so much to obtain specific results, as to bring out the intimate relationship between the energy band structure and the electrical conductivity tensor. By studying the electrical conductivity, its temperature dependence, and certain related electrical properties, such as the magnetoconductivity and the piezoconductivity, it is often possible to obtain useful information bearing on the energy band structure. Reversing the argument, we may say that a knowledge of the energy band structure —and a knowledge of the relaxation time—is essential to the proper understanding of the transport of electricity (or heat) by the electrons, and of the change in this transport produced by a magnetic field, a mechanical force, etc.

Since the primary emphasis of this paper is on the energy band structure, rather than on the electronic transport properties, we will not attempt to describe the electrical properties of silicon and germanium here. Instead, we will conclude this section by listing the various types of transport processes which have been investigated in order to elucidate the energy band structure of silicon and germanium. The interested reader might consult the original papers for further information. The theory of electronic transport processes in silicon and germanium is treated most comprehensively in the papers by Herring [42], Herring and Vogt [43], and Brooks [44].

The transport processes of interest are the following:

(*a*) Electrical conductivity and its temperature dependence [42–45]
(*b*) Hall effect and magnetoresistance [14, 26, 42–44, 46]
(*c*) Electrical conductivity and its pressure dependence [44, 47]
(*d*) Piezoresistance [16, 42–44, 48]
(*e*) Thermoelectric power [42–44, 49]
(*f*) Dielectric constant at microwave frequencies [42–44, 50]
(*g*) Magnetic susceptibility [25]

XV. Optical Absorption and Emission

In this section, we will describe some of the mechanisms which are responsible for the absorption and emission of optical energy in perfect (nonmetallic) crystals. We will be primarily interested in absorption or emission processes involving the destruction or the creation of a photon having a frequency lying in the infrared or visible region of the electromagnetic spectrum.

As is well known, a photon represents a unit quantum excitation of a monochromatic electromagnetic wave. We will denote the frequency of the photon by the symbol ν_{photon}, and the propagation vector of the photon by the symbol $\boldsymbol{q}_{\text{photon}}$: The energy of the photon, or light quantum, is simply $h\nu_{\text{photon}}$. The photon wavelength is given by: $\lambda_{\text{photon}} = 2\pi/q_{\text{photon}}$, where q_{photon} is the magnitude of $\boldsymbol{q}_{\text{photon}}$.

One of the simplest absorption acts involves the destruction of a photon and the excitation of an electron from some initial state β, k to some final state β', k'. The final state must be unoccupied before the excitation process occurs [Pauli Principle]. Energy and momentum must both be conserved during the absorption act. The momentum conservation selection rule takes the form: $k' = k + \boldsymbol{q}_{\text{photon}}$. Since the wavelength of an infrared or a visible photon is very large compared with a lattice constant, the photon momentum is a negligible quantity. Therefore, we may drop the term $\boldsymbol{q}_{\text{photon}}$, and rewrite the momentum conservation rule as follows:

$$k' = k. \tag{65}$$

Since the reduced wave vectors of the initial and final electronic states are equal, the electronic transition is called a vertical transition. [Visualize energy plotted vertically and k plotted horizontally. The line connecting the initial and the final states on an energy band diagram is a vertical line.] Energy conservation rule is:

$$E_{\beta'}(k') = E_{\beta'}(k) = E_\beta(k) + h\nu_{\text{photon}}, \tag{66}$$

where $E_\beta(k)$ and $E_{\beta'}(k')$ are the initial and final electronic energies, respectively. The corresponding emission act is governed by the same rules as govern the absorption act: (*a*) The final state must be unoccupied before the emission takes place. (*b*) The electronic transition is vertical, *i.e.*, $k' = k$. (*c*) Energy is conserved:

$$E_{\beta'}(k') = E_{\beta'}(k) = E_\beta(k) - h\nu_{\text{photon}}, \tag{67}$$

where the various symbols have the same meaning as before.

Consider a crystal for which the valence and the conduction band edges occur at the same point in the reduced zone. The minimum photon energy required to induce an electron transition across the forbidden band at the absolute zero of temperature is a direct measure of the width of the forbidden band at this temperature. Suppose we excite an electron from a state at the top of the valence band to a state at the bottom of the conduction band. If the electron returns to the initial valence state by emitting a photon, the energy of the photon will be equal to the width of the forbidden band. The emission act may be described as the radiative recombination of an electron and a hole.

A second type of absorption or emission involves (*a*) the destruction or the creation of a photon, (*b*) the transition of an electron from some initial state to some final state, and (*c*) the creation or the destruction of a phonon. As before, the electron must make a transition into an unoccupied state. For these processes, the momentum conservation condition takes the form:

$$\mathbf{k}' = \mathbf{k} + \mathbf{q}_{\text{phonon}} + \mathbf{h}. \tag{68}$$

where $\mathbf{q}_{\text{phonon}}$ is the reduced wave vector of the phonon, and where $\mathbf{h}$ is a reciprocal lattice vector. The term $\mathbf{q}_{\text{photon}}$ has again been omitted. $\mathbf{h}$ appears in (68) to insure that $\mathbf{k}$, $\mathbf{k}'$, and $\mathbf{q}_{\text{phonon}}$ all lie in the reduced zone or on the boundaries of the reduced zone. An electronic transition involving the creation or the destruction of a phonon is called a nonvertical transition because the reduced wave vector of the electron changes as a result of such a transition.

The energy conservation conditions for the four possible types of nonvertical electronic transitions may be written as follows:

$$\begin{aligned} E_{\beta'}(\mathbf{k}') &= E_{\beta'}(\mathbf{k} + \mathbf{q}_{\text{phonon}} + \mathbf{h}) \\ &= E_{\beta}(\mathbf{k}) \pm h\nu_{\text{phonon}} \pm h\nu_{\text{photon}}, \end{aligned} \tag{69}$$

where the minus signs refer to creation processes, and the plus signs to destruction processes.

Consider a crystal in which the valence and the conduction band edges occur at different points in the reduced zone, namely, $\mathbf{k}_v$ and $\mathbf{k}_c$, where $\mathbf{k}_v \neq \mathbf{k}_c$. This situation obtains in both silicon and germanium. At absolute zero temperature, there are no phonons present, and the absorption or emission of a photon cannot proceed with the destruction of a phonon. However, the absorption or the emission of a photon can proceed if a phonon is created. The minimum photon energy required to induce an electronic transition from the top of the valence band to the bottom of the conduction band at absolute zero temperature is:

$$\begin{aligned} h\nu_{\text{photon}} &= E_{\text{cond}}(\mathbf{k}_c) - E_{\text{val}}(\mathbf{k}_v) + h\nu_{\text{phonon}} \\ &= E_{\text{gap}} + h\nu_{\text{phonon}}. \end{aligned} \tag{70}$$

Thus, the threshold for optical absorption due to nonvertical electronic transitions occurs at a frequency higher than that corresponding to the width of the forbidden band, E_{gap}. As the temperature is raised, the phonon population increases. At moderately high temperatures, there will be enough phonons present for optical absorption at the threshold to proceed by the destruction of a phonon. In this case, we have:

$$\begin{aligned} h\nu_{\text{photon}} &= E_{\text{cond}}(\mathbf{k}_c) - E_{\text{val}}(\mathbf{k}_v) - h\nu_{\text{phonon}} \\ &= E_{\text{gap}} - h\nu_{\text{phonon}}. \end{aligned} \tag{71}$$

In this temperature range, the optical absorption edge occurs at a frequency lower than that corresponding to the width of the forbidden band.

It follows that the optical absorption edge will exhibit a temperature dependence which is related to the temperature dependence of the phonon population. As the temperature is raised, the width of the forbidden band will change slightly due to the thermal expansion of the lattice. The change in E_{gap} with temperature also influences the temperature dependence of the optical absorption edge.

A third mechanism for absorption of radiant energy involves the excitation of the normal modes of vibration. In this type of absorption, the electrons do not make transitions between states; only the phonons are involved. Since this mechanism is not directly related to the electronic energy spectrum, we will not discuss it any further. The interested reader might consult the paper by Lax and Burstein [51] for information concerning the lattice absorption in nonmetallic crystals.

It is instructive to examine the relationship between the optical properties of silicon and germanium and their energy band structures [30, 44, 52].

The optical absorption in crystals such as silicon and germanium is due to four effects: (*a*) lattice absorption; (*b*) absorption due to the excitation of electrons or holes within their respective bands; (*c*) absorption due to the excitation of free holes from one valence band to another; and (*d*) absorption due to the excitation of electrons across the forbidden band.

The lattice contribution to the absorption spectrum consists of a number of bands. These bands have not yet been unambiguously correlated with the lattice vibrational spectrum, though attempts in this direction have been made [51].

The absorption due to the excitation of free electrons from states in the conduction band to other states in the conduction band is directly proportional to the number of free electrons present. This type of absorption is characterized by an inverse square dependence on the wavelength of the incident light, and can be observed at wavelengths above the wavelength corresponding to the threshold for fundamental absorption, *i.e.*, absorption due to the excitation of electrons across the forbidden band. At wavelengths below the threshold for fundamental absorption, the free electron absorption is masked by the much stronger fundamental absorption.

We have already discussed the absorption due to free holes in germanium [21] in Section XIII. The most in-

teresting feature of this absorption is the fine structure in the absorption spectrum due to interband hole transitions. The absorption due to free holes in silicon exhibits an inverse square dependence on wavelength to the highest wavelengths at which measurements have thus far been performed. A fine structure due to interband hole transitions is expected in the far infrared, a region of the spectrum which has not yet been fully explored. The fine structure in germanium appears in the near infrared because the spin-orbit splitting in germanium is considerably larger than the spin-orbit splitting in silicon.

The width of the forbidden band, E_{gap}, and the temperature dependence of E_{gap} can be determined by examining the behavior of the threshold for fundamental absorption at low levels of absorption as a function of temperature. A theory of the absorption near the absorption edge in silicon and germanium has been developed by Macfarlane and Roberts [24, 27]. This theory is essentially a phenomenological extension of the work of Hall, Bardeen, and Blatt [53]. Macfarlane and Roberts assume that the only phonons which participate in the absorption are the phonons in the longitudinal acoustical branch of the vibrational spectrum. According to the selection rule (68), the reduced wave vector of the phonons must be equal to $\boldsymbol{k}_c$, the reduced wave vector for electrons at the bottom of the conduction band. [In silicon and germanium, $\boldsymbol{k}_v=(0, 0, 0)$, and $\boldsymbol{h}=(0, 0, 0)$.] The theory of Macfarlane and Roberts [24, 27], leads directly to the values for E_{gap} and $h\nu_{phonon}$, and indirectly to the value of $\boldsymbol{k}_c$. Typical results for E_{gap} are as follows:

Silicon:

$$E_{gap}=1.14 \text{ ev at } T=0°\text{K};$$
$$E_{gap}=1.08 \text{ ev at } T=300°\text{K}.$$

Germanium:

$$E_{gap}=0.73 \text{ ev at } T=0°\text{K};$$
$$E_{gap}=0.65 \text{ ev at } T=300°\text{K}.$$

The values obtained for $h\nu_{phonon}$, the energy of the longitudinal acoustical phonon, are:

Silicon: $h\nu_{phonon}=0.052$ ev;
Germanium: $h\nu_{phonon}=0.022$ ev.

In order to determine the values of $\boldsymbol{q}_{phonon}$ [$=\boldsymbol{k}_c$], it is necessary to make some assumptions concerning the nature of the lattice vibrational spectrum. Macfarlane and Roberts calculated the spectra for silicon and germanium by the same procedure that Hsieh [39] employed, namely, the method developed by Smith [41]. In the case of silicon, Macfarlane and Roberts found that $\boldsymbol{k}_c$ lies 7/9 of the way from the central zone point to the square face centers. A similar value for $\boldsymbol{k}_c$ was obtained by Kohn [23] on the basis of his interpretation of hyperfine splitting data. We think that the Macfarlane and Roberts result for $\boldsymbol{k}_c$ is a reasonable one.

In the case of germanium, they found that $\boldsymbol{k}_c$ lies $\frac{2}{3}$ of the way from the central zone point to the hexagonal face centers. On the other hand, Enz [25] and Conwell [26], by density-of-states arguments, conclude that $\boldsymbol{k}_c$ is more likely to lie at the hexagonal face centers —a conclusion with which we are in agreement. It is possible that the Macfarlane and Roberts analysis of the optical data is oversimplified, or that the lattice vibrational spectrum which they use is unreliable.

The absorption spectrum for germanium exhibits a second rise beyond the threshold for fundamental absorption [29, 30]. This second rise may be attributed to electronic transitions from the top of the valence band to the 000 minimum of the lowest conduction band. While the electronic transitions from the valence band edge to the conduction band edge (the 111 minima) require the assistance of phonons, the electronic transitions from the valence band edge (which occurs at 000) to the 000 minimum do not. The second rise is pronounced because the intensity for direct (vertical) transitions is considerably greater than that for indirect (nonvertical) transitions.

Further experimental evidence bearing on the energy separation between the 111 and the 000 minima of the conduction band and the valence band edge comes from radiative recombination measurements [31]. The intensity of the light emitted when 000 electrons recombine with holes is much greater than the intensity of the light emitted when 111 electrons recombine with holes. The relationship between corresponding absorption and emission processes can be studied fruitfully with the aid of the principle of detailed balancing.

XVI. Significance

The theoretical and experimental work which has been done on silicon and germanium during the past few years has taught us a great deal about the energy band structures of these crystals. Many of the important features of the energy band structures are now fairly well understood. The realization that the spin-orbit interaction can often play a profound role in determining the detailed nature of the energy band structure is of major significance.

The band structures of silicon and germanium have been found to be remarkably complex. Instead of a single minimum at the conduction band edge, there are six minima in the conduction band of silicon, and four minima in the conduction band of germanium, at the band edge. The conduction band of germanium has auxiliary minima lying above the four minima which define the conduction band edge. The structure near the valence band edge in silicon and germanium is complicated by the warping of the constant energy surfaces in this region. The structure is further complicated by the existence of three valence bands, two of which define the valence band edge, the third being separated from the first two by the spin-orbit interaction.

The results for silicon and germanium suggest that the band structures of other nonmetallic crystals are likely to be complicated; the band structures of some

nonmetallic crystals may even be more complex than those of silicon and germanium. The experience we have acquired in investigating these two crystals should prove extremely valuable in connection with the exploration of the band structures of other crystals.

The theory of the electrical and optical properties of silicon and germanium is greatly complicated by the complex nature of their energy band structures. While some of these properties can be correlated in a satisfying and convincing manner with the energy band structures, others cannot, at least at present. Much work remains to be done before the theory of silicon and germanium can be regarded as a completed theory. While many old problems have been solved during the past few years, we have become aware of many new problems, and some of these still defy solution.

Appendix

Calculation of the Number of Electrons and Holes in a Nonmetallic Crystal Having a Particular Energy Band Structure

In this appendix we will show how the position of the Fermi energy level in the forbidden band of a nonmetallic crystal having a particular type of band structure can be determined. We will also indicate how the electrons and holes are distributed in the conduction and valence bands, and what their total numbers are.

We will work with a crystal having the following band structure: The valence band edge is defined by two valence bands which come into contact with each other at the central point of the reduced zone. The upper valence band will be denoted by the symbol $V1$, and the lower, by $V2$. We will assume that the constant energy surfaces for $V1$ and $V2$ can be approximated by spherical surfaces in the neighborhood of the valence band edge. Let us place the zero of energy at the valence band edge, and let us denote the scalar effective masses for the holes in $V1$ and $V2$ by the symbols m_{V1}^* and m_{V2}^*. The energy band functions for the two valence bands are:

$$E_\beta(k) = -(h^2/8\pi^2)(k^2/m_\beta^*); \quad \beta = V1, V2. \quad (72)$$

Since $V1$ lies above $V2$, the curvature of $E_{V1}(k)$ must be less than that of E_{V2} at the valence band edge, and we have: $m_{V1}^* > m_{V2}^*$.

We will assume that the lowest conduction band has M minima which are located within the reduced zone at a set of M symmetrically equivalent points. We will assume further that this conduction band is nondegenerate at each of its minima. The conduction band edge is defined, then, by the M symmetrically equivalent and nondegenerate minima of the lowest conduction band. The constant energy surfaces close to each of the M minima are ellipsoids. Let one of these minima occur at $k = k_0$, and let us introduce the definition: $k' = (k_a', k_b', k_c') \equiv k - k_0$, where the components of k' are measured along the principal axes of the ellipsoids associated with the minimum at k_0. The energy of an electron occupying a state in the neighborhood of k_0 is given by the following expression:

$$E_{cond}(k') = E_{gap} + (h^2/8\pi^2)[(k_a')^2/m_a^* + (k_b')^2/m_b^* + (k_c')^2/m_c^*], \quad (73)$$

where E_{gap} is the width of the forbidden band, and where m_a^*, m_b^*, and m_c^* are the principal components of the effective mass tensor for the minimum at k_0. The zero of energy is again at the valence band edge.

Let the quantity $D_{elec}(E)dE$ denote the number of electrons per unit volume whose energies lie in the range E, $E+dE$. The electrons in all M minima, of both spin types, are counted by the density of states function $D_{elec}(E)$. It can be shown by simple algebraic arguments that $D_{elec}(E)$ is given by:

$$D_{elec}(E) = 8(2)^{1/2}(\pi M/h^3)(m_a^* m_b^* m_c^*)^{1/2}(E - E_{gap})^{1/2}; \quad E \geqq E_{gap}. \quad (74)$$

The density of states for the holes in the two valence bands is:

$$D_{holes}(E) = D_{V1}(E) + D_{V2}(E), \quad (75)$$

where:

$$D_\beta(E) = 8(2)^{1/2}(\pi/h^3)(m_\beta^*)^{3/2}(-E)^{1/2}; \quad \beta = V1, V2; \quad E \leqq 0. \quad (76)$$

We will assume that the distribution of the occupied (electron) states in the conduction band and the distribution of the unoccupied (hole) states in the valence bands are governed by the classical statistics. Therefore the electron distribution function is

$$f_{elec}(E) = \exp[-(E - \xi)/kT]; \quad E \geqq E_{gap}, \quad (77)$$

where ξ is the Fermi energy, measured with respect to the valence band edge. The classical distribution function for the holes is:

$$f_{holes}(E) = \exp[-(\xi - E)/kT]; \quad E \leqq 0. \quad (78)$$

The total number of electrons in the conduction band, N_{elec}, is obtained as follows:

$$N_{elec} = \int_{E_{gap}}^{\infty} D_{elec}(E) f_{elec}(E) dE = 4(2)^{1/2}\pi^{3/2}(M/h^3)(m_a^* m_b^* m_c^*)^{1/2}(kT)^{3/2} \cdot \exp[-(E_{gap} - \xi)/kT]. \quad (79)$$

Similarly, the total number of holes in the two valence bands, N_{holes}, is:

$$N_{holes} = \int_{-\infty}^{0} D_{holes}(E) f_{holes}(E) dE \quad (80)$$
$$= 4(2)^{1/2}\pi^{3/2}(1/h^3)[(m_{V1}^*)^{3/2} + (m_{V2}^*)^{3/2}](kT)^{3/2} \cdot \exp[-\xi/kT].$$

Since we are dealing with a perfect crystal, the number of occupied states in the conduction band must be equal to the number of unoccupied states in the valence bands. In other words, $N_{elec} = N_{holes}$. By equating (79) and (80), we can readily determine the value for the Fermi energy:

$$\xi = (1/2)E_{gap} - (1/2)(kT) \log \phi, \quad (81)$$

where:

$$\phi = \frac{M(m_a{}^* m_b{}^* m_c{}^*)^{1/2}}{(m_{V1}{}^*)^{3/2} + (m_{V2}{}^*)^{3/2}}. \quad (82)$$

Substituting the value for ξ given by (81) back into (79) or (80), we obtain the desired result:

$$N_{elec} = N_{holes} = 4(2)^{1/2}\pi^{3/2}\phi^{1/2}h^{-3}(kT)^{3/2} \cdot \exp\left[-E_{gap}/2kT\right]. \quad (83)$$

The temperature dependence of N_{elec} or N_{holes} should be particularly noted. It should be observed that the energy band structure enters in two places in (83), namely, in ϕ [cf. (82) above], and in E_{gap}.

Acknowledgment

The author wishes to thank various members of the RCA Laboratories staff, particularly Drs. M. Glicksman, J. O. Kessler, and L. S. Nergaard, for helpful comments on the manuscript.

Bibliography

[1] Shockley, W., "Transistor Electronics: Imperfections, Unipolar and Analog Transistors." Proceedings of the IRE, Vol. 40 (November, 1952), pp. 1289–1314, see also: Schultz, M. L., and Morton, G. A., "Photoconduction in Germanium and Silicon." Proceedings of the IRE, Vol. 43, pp. 1819–1827 (this issue).

[2] Kittel and Elliott were among the first to appreciate the importance of the spin-orbit interaction in energy band theory [cf. references 19 and 20].

[3] Kittel, C., "Experimental Evidence on the Band Structure of Germanium and Silicon." *Physica*, Vol. 20 (November, 1954), pp. 829–833; see also references [67] and [68].

[4] Herman, F., "Some Recent Developments in the Calculation of Crystal Energy Bands—New Results for the Germanium Crystal." *Physica*, Vol. 20 (November, 1954), pp. 801–812; Herman, F., unpublished.

[5] Woodruff, T. O., "Solution of the Hartree-Fock-Slater Equations for Silicon Crystal by the Method of Orthogonalized Plane Waves," *Physical Review*, Vol. 98 (June, 1955), pp. 1741–1742; see also references [19] and [20].

[6] See, for example, Shockley, W., *Electrons and Holes in Semiconductors*. New York, D. Van Nostrand and Company, Inc., 1950.

[7] Pearson, G. L., and Suhl, H., "The Magneto-Resistance Effect in Germanium." *Physical Review*, Vol. 83 (August, 1951), pp. 768–776.

[8] Pearson, G. L., Haynes, J. R., and Shockley, W., "Comment on Mobility Anomalies in Germanium." *Physical Review*, Vol. 78 (May, 1950), pp. 295–296.

[9] Shockley, W., "Energy Band Structures in Semiconductors." *Physical Review*, Vol. 78 (April, 1950), pp. 173–174; see also reference [6], pp. 336–341, and footnote on page 174.

[10] Herman, F., "Electronic Structure of the Diamond Crystal." *Physical Review*, Vol. 88 (December, 1952), pp. 1210–1211; for a more complete account, see Herman, F., "Calculation of the Energy Band Structures of the Diamond and Germanium Crystals by the Method of Orthogonalized Plane Waves." *Physical Review*, Vol. 93 (March, 1954), pp. 1214–1225.

[11] Herman, F., and Callaway, J., "Electronic Structure of the Germanium Crystal." *Physical Review*, Vol. 89 (January, 1953), pp. 518–519.

[12] Shockley, W., "Cyclotron Resonances, Magnetoresistance, and Brillouin Zones in Semiconductors." *Physical Review*, Vol. 90 (May, 1953), p. 491.

[13] Dresselhaus, G., Kip, A. F., and Kittel, C., "Observation of Cyclotron Resonance in Germanium Crystals." *Physical Review*, Vol. 92 (November, 1953), p. 827; Lax, B., Zeiger, H. J., Dexter, R. N., and Rosenblum, E. S., "Directional Properties of the Cyclotron Resonance in Germanium." *Physical Review*, Vol. 93 (March, 1954), pp. 1418–1420; Dexter, R. N., Zeiger, H. J., and Lax, B., "Anisotropy of Cyclotron Resonance of Holes in Germanium." *Physical Review*, Vol. 95 (July, 1954), pp. 557–558; Dexter, R. N., Lax, B., Kip, A. F., and Dresselhaus, G., "Effective Masses of Electrons in Silicon." *Physical Review*, Vol. 96 (October, 1954), pp. 222–223; Dexter, R. N., and Lax, B., "Effective Masses of Holes in Silicon." *Physical Review*, Vol. 96 (October, 1954), pp. 223–224; Dresselhaus, G., Kip, A. F., and Kittel, C., "Cyclotron Resonance of Electrons and Holes in Silicon and Germanium Crystals." *Physical Review*, Vol. 98 (April, 1955), pp. 368–384; see also the papers by Kip, Lax *et al.* and Kittel in reference [64].

[14] Meiboom, S., and Abeles, B., "Theory of the galvanomagnetic Effects in N-Germanium." *Physical Review*, Vol. 93 (March, 1954), p. 1121; Abeles, B., and Meiboom, S., "Theory of the Galvanomagnetic Effects in Germanium." *Physical Review*, Vol. 95 (July, 1954), pp. 31–37; see also: Shibuya, M., "Magnetoresistance Effect in Cubic Semiconductors with Spheroidal Energy Surfaces." *Physical Review*, Vol. 95 (September, 1954), pp. 1385–1393; see also the papers by Meiboom and Abeles, and Shibuya in reference [64].

[15] Pearson, G. L., and Herring, C., "Magneto-resistance Effect and the Band Structure of Single Crystal Silicon." *Physica*, Vol. 20 (November, 1954), pp. 975–978.

[16] Smith, C. S., "Piezoresistance Effect in Germanium and Silicon," *Physical Review*, Vol. 94 (April, 1954), pp. 42–49.

[17] Briggs, H. B., and Fletcher, R. C., "Absorption of Infrared Light by Free Carriers in Germanium." *Physical Review*, Vol. 91 (September, 1953), pp. 1342–1346.

[18] Kaiser, W., Collins, R. J., and Fan, H. Y., "Infrared Absorption in P-type Germanium." *Physical Review*, Vol. 91 (September, 1953), pp. 1380–1381.

[19] Dresselhaus, G., Kip, A. F., and Kittel, C., "Spin-orbit Interaction and the Effective Masses of Holes in Germanium." *Physical Review*, Vol. 95 (July, 1954), pp. 568–569; for a more complete account, see: Dresselhaus, G., Kip, A. F., and Kittel, C., "Cyclotron Resonance of Electrons and Holes in Silicon and Germanium Crystals." *Physical Review*, Vol. 98 (April, 1955), pp. 368–384.

[20] Elliott, R. J., "Theory of the Effect of Spin-orbit Coupling on Magnetic Resonance in Some Semiconductors." *Physical Review*, Vol. 96 (October, 1954), pp. 266–279; Elliott, R. J., "Spin-orbit Coupling in Band Theory—Character Tables for Some 'Double' Space Groups." *Physical Review*, Vol. 96 (October, 1954), pp. 280–287.

[21] Kahn, A. H., "Theory of the Infrared Absorption of Carriers in Germanium and Silicon," *Physical Review*, Vol. 97 (March, 1955), pp. 1647–1652. See also Dingle, R. B., "Theory of the Infrared Absorption by Carriers in Semiconductors," *Physical Review*, Vol. 99 (September, 1955), pp. 1330–1331.

[22] Lax, B., reported orally at the Symposium on Physics of Semiconductor Materials, University of Michigan, June 7, 1955. These are the most recent results obtained by the Lincoln Laboratory group. The results of the Berkeley group are substantially the same as those reported by Lax.

[23] Kohn, W., "Spin Resonance Widths of Electrons in Donor States in Silicon." *Physical Review*, Vol. 98 (June, 1955), p. 1561.

[24] Macfarlane, G. G., and Roberts, V. "Infrared Absorption of Silicon Near the Lattice Edge." *Physical Review*, Vol. 98 (June, 1955), pp. 1865–1866.

[25] Enz, C., "Ladungsträger-suszeptibilität und Cyclotron-resonanz von Germanium." *Helvetica Physica Acta*, Vol. 28 (1955), pp. 158–166.

[26] Conwell, E. M., "Hall Effect and Density of States in Germanium," *Physical Review*, Vol. 99 (August, 1955), pp. 1195–1198.

[27] Macfarlane, G. G., and Roberts, V., "Infrared Absorption of Germanium Near the Lattice Edge," *Physical Review*, Vol. 97 (March, 1955), pp. 1714–1716.

[28] Smith, R. A., "The Electronic and Optical Properties of the Lead Sulphide Group of Semi-Conductors." *Physica*, Vol. 20 (November, 1954), pp. 910–929.

[29] Dash, W. C., Newman, R., and Taft, E. A., "Optical Absorption in Single Crystals of Ge and Si." *Physical Review*, Vol. 98 (May, 1955), p. 1192. For a more complete account, see Dash, W. C., and Newman, R., "Intrinsic Optical Absorption in Single-Crystal Germanium and Silicon at 77°K and 300°K," *Physical Review*, Vol. 99 (August, 1955), pp. 1151–1155.

[30] Fan, H. Y., Shepherd, M. L., and Spitzer, W., "Infrared Absorption and Energy Band Structure of Germanium and Silicon." in *Photoconductivity Conference*. New York, John Wiley and Sons, Inc., in press.
[31] Haynes, J. R., "New Radiation Resulting from Recombination of Electrons and Holes in Germanium." *Physical Review*. Vol. 98 (June, 1955), pp. 1866–1868. See also van Roosbroeck, W., and Shockley, W., "Photon-Radiative Recombination of Electrons and Holes in Germanium," *Physical Review*, Vol. 94 (June, 1954), pp. 1558–1560.
[32] Herman, F., "Speculations on the Energy Band Structure of Ge-Si Alloys." *Physical Review*, Vol. 95 (August, 1954), pp. 847–848; see also reference [4].
[33] Johnson, E. R., and Christian, S. M., "Some Properties of Germanium-Silicon Alloys." *Physical Review*, Vol. 95 (July, 1954), pp. 560–561.
[34] Christian, S. M., private communication.
[35] Parmenter, R. H., "Energy Levels of a Disordered Alloy." *Physical Review*, Vol. 97 (February, 1955), pp. 587–598.
[36] Glicksman, M., "Magnetoresistance of Germanium-Silicon Alloys," *Physical Review*, (November 15, 1955 issue).
[37] Dresselhaus, G., Kip, A. F., Ku, Han-Ying, Wagoner, G., and Christian, S. M., "Cyclotron Resonance in Ge-Si Alloys." *Physical Review*, in press.
[38] Herman, F., "Speculations on the Energy Band Structure of Zinc-Blende-Type Crystals." *Journal of Electronics* (London), Vol. 1 (September, 1955), pp. 103–114.
[39] Hsieh, Yü-Chang, "The vibrational Spectrum and the Specific Heat of Germanium and Silicon." *Journal of Chemical Physics*, Vol. 22 (February, 1954), pp. 306–311.
[40] See, for example, Brillouin, L., "Wave Propagation in Periodic Structures." New York, McGraw-Hill Book Company, Inc., 1946.
[41] Smith, H. M. J., "The Theory of the Vibrations and the Raman Spectrum of the Diamond Lattice." *Philosophical Transactions of the Royal Society*, London, Vol. A241 (July, 1948), pp. 105–145.
[42] Herring, C., "Transport Properties of a Many-Valley Semiconductor." *Bell System Technical Journal*, Vol. 34 (March, 1955), pp. 237–290.
[43] Herring, C., and Vogt, E., "Transport and Deformation Potential Theory for Many-Valley Semiconductors with Anisotropic Scattering." *Physical Review*, in press.
[44] Brooks, H., "Theory of the Electrical Properties of Germanium and Silicon," in L. Marton, ed. *Advances in Electronics and Electron Physics*, (New York), Academic Presss, (Vol. VII, 1955).
[45] Prince, M. B., "Drift Mobilities in Semiconductors I; Germanium." *Physical Review*, Vol. 92 (November, 1953), pp. 681–687. Morin, F. J., "Lattice-Scattering Mobility in Germanium." *Physical Review*. Vol. 93 (January, 1954), pp. 62–63; Prince, M. B., "Drift Mobilities in Semiconductors II: Silicon." *Physical Review*, Vol. 93 (March, 1954), pp. 1204–1206; Morin, F. J., and Maita, J. P., "Conductivity and Hall Effect in the Intrinsic Range in Germanium." *Physical Review*, Vol. 94 (June, 1954), pp. 1525–1529; Morin, F. J., and Maita, J. P., "Electrical Properties of Silicon Containing Arsenic and Boron." *Physical Review*, Vol. 96 (October, 1954), pp. 28–35.
[46] Willardson, R. W., Harman, T. C., and Beer, A. C., "Transverse Hall and Magneto-Resistance Effects in P-type Germanium." *Physical Review*, Vol. 96 (December, 1954), pp. 1512–1518.
[47] Paul, W., and Brooks, H., "Pressure Dependence of the Resistivity of Germanium." *Physical Review*, Vol. 94 (June 1954), pp. 1128–1133; Paul, W., and Pearson, G. L., "Pressure Dependence of the Resistivity of Silicon." *Physical Review*, Vol. 98 (June, 1955), pp. 1755–1757.
[48] Adams, E. N., "Elastoresistance in P-type Ge and Si." *Physical Review*, Vol. 96 (November, 1954), pp. 803–804.
[49] Frederikse, H. P. R., "Thermoelectric Power of Germanium Below Room Temperature." *Physical Review*, Vol. 92 (October, 1953), pp. 248–252; Geballe, T. H., and Hull, G. W., "Seeback Effect in Germanium." *Physical Review*, Vol. 94 (June, 1954), pp. 1134–1140; Herring, C., "Theory of the Thermoelectric Power of Semiconductors." *Physical Review*, Vol. 96 (December, 1954), pp. 1163–1187; Geballe, T. H., and Hull, G. W., "Seebeck Effect in Silicon." *Physical Review*, Vol. 98 (May, 1955), pp. 940–947.
[50] Benedict, T. S., and Shockley, W., "Microwave Observation of the Collision Frequency of Electrons in Germanium." *Physical Review*, Vol. 89 (March, 1953), pp. 1152–1153; Benedict, T. S., "Microwave Observation of the Collision Frequency of Holes in Germanium." *Physical Review*, Vol. 91 (September, 1953), pp. 1565–1566; Goldey, J. M., and Brown, S. C., "Microwave determination of the Average Masses of Electrons and Holes in Germanium." *Physical Review*, Vol. 98 (June, 1955), pp. 1761–1763,
[51] Lax, M., and Burstein, E., "Infrared Lattice Absorption in Ionic and Homopolar Crystals." *Physical Review*, Vol. 97 (January, 1955), pp. 39–52; see also: Collins, R. J., and Fan, H. Y., "Infrared Lattice Absorption Bands in Germanium, Silicon, and Diamond." *Physical Review*, Vol. 93 (February, 1954), pp. 674–678.
[52] Burstein, E., Picus, G. S., and Sclar, N., "Optical and Photoconductive Properties of Silicon and Germanium," in *Photoconductivity Conference*, New York, John Wiley and Sons, Inc., in press.
[53] Hall, L. H., Bardeen, J., and Blatt, F. J., "Infrared Absorption Spectrum of Germanium." *Physical Review*, Vol. 95 (July, 1954), pp. 559–560; Bardeen, J., Blatt, F. J., and Hall, L. H., "Indirect Transitions from the Valence to the Conduction Bands," in *Photoconductivity Conference*, New York, John Wiley and Sons, Inc., in press.

General Bibliography

Textbooks

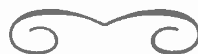
[54] Born, M., and Huang, K., *Dynamical Theory of Crystal Lattices*, London, Oxford University Press, 1954.
[55] Brillouin, L., *Wave Propagation in Periodic Structures*, New York, McGraw-Hill Book Company, Inc., 1946.
[56] Kittel, C., *Introduction to Solid State Physics*, New York, John Wiley and Sons, Inc., 1953.
[57] Peierls, R. E., *Quantum Theory of Solids*, London, Oxford University Press, 1955.
[58] Seitz, F., *The Modern Theory of Solids*, New York, McGraw-Hill Book Company, Inc., 1940.
[59] Shockley, W., *Electrons and Holes in Semiconductors*, New York, D. Van Nostrand and Company, Inc., 1950.
[60] Slater, J. C., *Quantum Theory of Matter*, New York, McGraw-Hill Book Company, Inc., 1951.
[61] Spenke, E., *Elektronische Halbleiter*, Berlin, Springer-Verlag, 1955.
[62] Wilson, A. H., *The Theory of Metals*, 2nd ed., Cambridge, Cambridge University Press, 1953.

Proceedings of Recent Conferences

[63] *Imperfections in Nearly Perfect Crystals*, New York, John Wiley and Sons, Inc., 1952. Symposium Held at Pocono Manor, October, 1950. See particularly: Seitz, F., "Imperfections in Nearly Perfect Crystals, A Synthesis," pp. 3–76
[64] "Proceedings of the International Conference on Semiconductors, Amsterdam, June 29–July 3, 1954," *Physica*, Vol. 20 (November, 1954).
[65] *Report on the Conference on Defects in Crystalline Solids, University of Bristol, July, 1954*, London, The Physical Society, 1955
[66] *Photoconductivity Conference*, New York, John Wiley and Sons, Inc., in press. Conference held at Atlantic City, November, 1954.

Recent and Forthcoming Review Articles

[67] Burstein, E., and Egli, P., "Physics of Semiconducting Materials," in L. Marton, ed., *Advances in Electronics and Electron Physics* (New York), Academic Press, (Vol. VII, 1955).
[68] Brooks, H., "Theory of the Electrical Properties of Germanium and Silicon," in L. Marton, ed., *Advances in Electronics and Electron Physics* (New York), Academic Press, (Vol. VII, 1955).
[69] Bube, R. H., Herman, F., and Leverenz, H. W., "The Solid State," in G. K. Rollefson, ed., *Annual Review of Physical Chemistry* (Stanford), Annual Reviews, Inc., Vol. 5 (1954), pp. 199–218.
[70] Ewald, P. P., "Properties of Solids and Solid Solutions," in F. D. Rossini, ed., *Thermodynamics and Physics of Matter* (Princeton, 1955), Princeton University Press, pp. 554–640.
[71] Lark-Horovitz, K., "The New Electronics," in *The Present State of Physics* (Washington, D. C., 1954), American Association for the Advancement of Science, pp. 57–127.



Nonlinear Dielectric Materials*

E. T. JAYNES,† SENIOR MEMBER, IRE

Summary—**The following is a brief description of nonlinear dielectrics from the standpoint of the fundamental physics involved. Specific available materials and technical applications are considered only by way of illustration of general properties. It is shown that, although dielectric nonlinearity and ferroelectricity are quite different phenomena, the fact that ferroelectrics have high dielectric constants makes them the most likely materials to exhibit a high degree of nonlinearity at electric field strengths safely below the breakdown values.**

Introduction

THE MACROSCOPIC electromagnetic properties of matter are commonly described by specifying the current density J, electric displacement D, and magnetic induction B as functions of the electric and magnetic "intensities" E and H. In the majority of substances these relations are found to be linear, leading to the definitions of the conductivity, dielectric constant, and permeability tensors. The meaning of the word "linear" must be made more precise as soon as we consider time-varying fields; in particular, we must be careful to distinguish between nonlinearity and dispersion. The simple statement that $D(t)$ is proportional to $E(t)$ will not do; by common usage it is understood that linearity is concerned with such a proportionality in the frequency domain rather than in the time domain. Thus, consider a Fourier integral representation of the fields:

$$E(t) = \int_{-\infty}^{\infty} E(\omega)e^{i\omega t}d\omega$$
$$D(t) = \int_{-\infty}^{\infty} D(\omega)e^{i\omega t}d\omega \qquad (1)$$

By linearity we mean that the material is characterized by some unique function $\epsilon(\omega)$ such that

$$D(\omega) = \epsilon(\omega)E(\omega). \qquad (2)$$

This corresponds in general to no simple relation between $D(t)$ and $E(t)$. Thus, a linear dielectric material is characterized intuitively by the following conditions:

1. No frequencies are present in $D(t)$ which not are present in $E(t)$.
2. If $E_1(t)$ produces $D_1(t)$, which we write compactly as $E_1 \rightarrow D_1$ and $E_2 \rightarrow D_2$, then $(a_1E_1 + a_2E_2) \rightarrow (a_1D_1 + a_2D_2)$, with similar definitions for linear conductors and magnetic media.

It is possible to have one but not both of these conditions satisfied: for example, consider the corresponding magnetic properties of water placed in a strong but inhomogenous magnetic field. For applied frequencies within the range of the proton Larmor frequencies, condition 1 is satisfied but not the superposition condition 2.[1]

Strictly speaking, all such linear laws may be regarded as approximations for several different reasons. In the first place, they would presumably fail in any material substance at sufficiently high field strengths; *i.e.*, breakdown or saturation effects would occur. Secondly, there really are no unique relations between the above vectors since in the work of the highest accuracy one would always expect to find that, for example, the electric displacement does not depend only on the electric field, but also on every other physical condition of the material such as temperature, state of stress and strain, degree of illumination, and even the entire past history of the specimen. Thus, even if it should be found that D is exactly proportional to E at constant temperature, the fact that the dielectric constant so defined is a function of temperature gives rise to electrocaloric effects, in which a sudden change in electric field produces a change in temperature. Thus, the linearity or nonlinearity of a substance could depend on its degree of thermal contact with its surroundings; *i.e.*, on whether it is operated under isothermal or adiabatic conditions. Similarly, if a dielectric is linear under conditions of constant stress, it might not be so under conditions of constant strain. In practice, however, these are usually extremely small effects.

Finally there are more esoteric examples provided by modern physical theories, according to which even a perfect vacuum should have nonlinear properties. In quantum electrodynamics one finds that the phenomenon of scattering of light by light (a violation of condition 2) should occur due to the formation and subsequent annihilation of electron-positron pairs;[2] the cross-section for this process is, however, so small that experimental confirmation is not to be thought of. Another effect is predicted by General Relativity; an electromagnetic field contains energy and therefore mass. This produces a gravitational field which can in turn deflect a light beam, again in violation of condition 2. Once again, we do not expect any experimental confirmation in the laboratory! In order to be extremely cautious about the experimental situation, however, we note that electrical measurements of the highest accuracy are never performed with intense fields; if appreciable deviations from Maxwell's equations did occur in free space at field strengths in excess

* Original manuscript received by the IRE, October 14, 1955.
† Microwave Laboratory, Stanford University, Stanford, Calif.

[1] F. Bloch, *Phys. Rev.*, vol. 70, p. 460; 1946.
A. Bloom, *Phys. Rev.*, vol. 98, p. 1105; 1955.
[2] O. Halpern, *Phys. Rev.*, vol. 44, p. 855; 1934.
H. Euler and B. Kockel, *Naturwiss.*, vol. 23, p. 246; 1935.

of 10^5 v/cm or 10^6 oersted, they would almost certainly have escaped experimental detection thus far.[3]

In spite of the above considerations, the vast majority of substances is found to be linear to a high degree of accuracy at all field strengths commonly attained, hence we denote as "nonlinear" only those substances in which there are substantial and easily demonstrable effects arising from violation of conditions 1 or 2 above, occurring at easily produced field strengths. Each such substance is potentially capable of important technical applications, since by violation of condition 1 we generate harmonics and beat frequencies, while violation of condition 2 enables one to modulate one signal by the presence of another.

In the case of conductivity and permeability, substances which are nonlinear in the above sense have long been known; thyrite and, to a certain extent, electrolytic cells and various solid-state devices, such as rectifiers, may be regarded as media with nonlinear conductivity (more correctly as circuit elements with nonlinear conductance), while the hysteresis and saturation effects in ferromagnetic materials respresent great nonlinearity. By contrast, nonlinear dielectric materials, although not entirely unknown in the past, have not until recently been available in forms of interest to electrical engineers. Probably the earliest known example of dielectric nonlinearity was the discovery, over 80 years ago, of the Kerr electro-optical effect.[4] Glass and many liquids, in particular nitrobenzene, develop optical birefringence in fairly strong electric fields; thus the dielectric constant at optical frequencies varies with a low-frequency electric field in violation of the superposition condition. An example of a nonlinear capacitance at microwave frequencies is provided by crystal rectifiers, particularly the germanium welded-contact variety,[5] in which the barrier capacitance varies strongly with bias voltage. By far the most important nonlinear dielectrics, however, are the ferroelectric crystals or ceramics.

Ferroelectrics

From a phenomenological point of view, ferroelectricity may be defined as the electric analog of ferromagnetism, and the fundamental criterion of ferroelectricity is the existence, at certain temperatures, of hysteresis between D and E. The static electric displacement then depends not only on the applied electric field but also on the past history, in such a way that with sufficiently slow periodic variation of E we obtain a D-E hysteresis loop exactly like the familiar B-H curves of ferromagnetic materials. Above a certain temperature T_c, called the Curie point, this hysteresis disappears, but the relation between D and E may remain appreciably nonlinear up to temperatures far above T_c. Since oscillograms illustrating these effects have been published recently in this journal,[6] they will not be repeated here. In the neighborhood of the Curie point more complicated phenomena are sometimes found.[7]

Several distinct classes of ferroelectrics, with widely different chemical composition and crystal structure, are now known. The first to be discovered was Rochelle salt (sodium potassium tartrate tetrahydrate), widely used for its piezoelectric properties.[8] This substance appears to be unique in that it possesses two Curie temperatures (-18°C and $+23$°C) and is ferroelectric between them. Other ferroelectric tartrates[9,10] show only a single Curie point. Mueller[11] has shown that the electrical, mechanical, and thermal properties of Rochelle salt can be correlated very satisfactorily by a single thermodynamic free-energy function valid on both sides of the Curie points. This is important not only from the standpoint of economy of description, but it indicates that the ferroelectric-"paralectric" phase transition at the Curie points is probably not a very drastic rearrangement from a molecular point of view as is the case in many phase transitions, for example, that between diamond and graphite.[12] This conclusion seems well established also for the other classes[13] of ferroelectrics.

Another class of ferroelectrics is represented by the salt potassium dihydrogen phosphate, KH_2PO_4 and other substances of similar chemical composition and crystal structure (*i.e.*, the alkali and ammonium phosphates and arsenates).[14] Although they have found applications based on their electro-optical properties, the fact that their Curie points are at liquid-air temperatures limits their usefulness as nonlinear dielectrics in the purely electrical sense. KH_2PO_4 is at present unique in that the molecular mechanism of its properties (different arrangements of hydrogen bonds) is probably

[3] There is, of course, indirect evidence associated with atomic theory suggesting that the laws of electrostatics hold at field strengths far beyond this limit; for example, the fact that the wavelengths of the spectral lines of hydrogen can be calculated to great accuracy on the assumption that the electric field of the nucleus is a coulomb field. However, this could hardly be called an electrical measurement.

[4] M. Born, "Optik," p. 365, J. Springer, Berlin; 1933.

[5] H. C. Torrey and C. A. Whitmer, "Crystal Rectifiers," Chap. 13, M.I.T. Radiation Laboratory series No. 15; McGraw-Hill Book Co., Inc., New York; 1948.

[6] W. P. Mason and R. F. Wick, "Ferroelectrics and the dielectric amplifier," Proc. IRE, vol. 42, pp. 1606–1620; November, 1954.

[7] W. J. Merz, *Phys. Rev.*, vol. 91, p. 513; 1953.

[8] W. G. Cady, "Piezoelectricity," McGraw-Hill Book Co., Inc., New York, 1946.

[9] W. J. Merz, *Phys. Rev.*, vol. 82, p. 562; 1950, vol. 83, pp. 226, 656; 1951.

[10] B. T. Matthias and J. K. Hulm, *Phys. Rev.*, vol. 82, pp. 108; 1951.

[11] H. Mueller, *Phys. Rev.*, vol. 47, p. 175; 1935, vol. 57, pp. 829–842; 1940, vol. 58, pp. 565–805; 1941, *Zeit. Krist.*, vol. 99, p. 122, 1938, *Ann. N. Y. Acad. Sci.*, vol. 40, p. 321; 1940.

[12] R. Smoluchowski, et al., "Phase Transformations in Solids," John Wiley & Sons, New York; 1951.

[13] H. R. Danner and R. Pepinsky, *Phys. Rev.*, vol. 99, p. 1215; 1955.

[14] G. Busch and P. Scherrer, *Naturwiss.*, vol. 23, p. 737; 1935.
G. Busch, *Helv. Phys. Acta.*, vol. 11, p. 269, 1938.
C. C. Stephenson and J. G. Hooley, *Phys. Rev.*, vol. 56, p. 121; 1939.
W. Bantle and P. Scherrer, *Nature*, vol. 143, p. 980; 1939.
J. and K. Mendelssohn, *Nature*, vol. 144, p. 595, 1939.

understood with greater certainty than for any other ferroelectric.[15,16]

The third class of ferroelectrics,[17] represented by the cubic form of barium titanate, $BaTiO_3$ and several similar substances, is at the present time the most interesting from a theoretical standpoint and the most useful in terms of applications. The Curie point of $BaTiO_3$ is at 120°C, and at room temperature the most nearly perfect crystals grown to date[7] exhibit hysteresis loops with coercive force as low as 600 v/cm and a spontaneous polarization of about 26 microcoulombs /cm², very much higher than in the first two classes of ferroelectrics. The dielectric properties of $BaTiO_3$ single crystals above the Curie point have been measured by Drougard, Landauer, and Young,[18] with results that may be summarized as follows. For crystals that are free to distort in accordance with their electrostrictive properties (*i.e.*, under conditions of zero stress), the behavior is given within experimental error by the free-energy function

$$F(P, T) = F_0(T) + A(T)P^2 + B(T)P^4, \qquad (3)$$

where P is the dielectric polarization, F_0 the free energy at zero polarization (which is irrelevant for dielectric properties, although it largely determines the specific heat of the material), and A, B are linear functions of temperature, given by the empirical equations

$$\begin{aligned} A &= 3.8 \times 10^{-5}\,(T - 105) \\ B &= 4.5 \times 10^{-15}\,(T - 175), \end{aligned} \qquad (4)$$

which are in cgs units, with the temperature in degrees centigrade. The electric field is, from thermodynamics,

$$E = \partial F/\partial P = 2AP + 4BP^3. \qquad (5)$$

Therefore, the incremental (small signal) dielectric constant is

$$\begin{aligned} \epsilon &= 1 + 4\pi(\partial P/\partial E) = 1 + 4\pi/(2A + 12BP)^2 \\ &\simeq 4\pi A^2/(2A^3 + 3BE^2), \end{aligned} \qquad (6)$$

the approximation being valid at field strengths for which the cubic term in (5) is small compared to the linear one. Since B is negative in the temperature range where these experiments were performed (119°C to 150°C), we have the rather surprising result that application of a biasing field increases the dielectric constant of the crystal. This phenomenon is seen clearly in the oscillograms of Merz,[7] and may be shown by thermodynamic arguments[19] to be connected with the fact that the crystals exhibit a first-order transition at the Curie point; *i.e.*, as the temperature is lowered, the spontaneous polarization jumps discontinuously from zero to a finite value (about 18 microcoulombs /cm²). Some of the first crystals grown, which were less perfect, exhibited a second-order transition and, as required by thermodynamics, a positive sign of B.[20] It must be remembered, however, that this increase due to bias occurs only under conditions of zero stress, and therefore can be seen only at sufficiently low frequencies, below all of the mechanical vibration modes of the crystal. Measurements made at Stanford University by Mr. V. Varenhorst at frequencies of 20, 40, and 120 mcs showed in all cases a decrease in dielectric constant with bias voltage. The results were complicated by temperature hysteresis effects which persist above the Curie point and are not understood, but in a typical case at 40 mcs and 130°C, application of a biasing field of 1700 v/cm lowered the dielectric constant from 6,700 to 6,000.

Growth of good single crystals of $BaTiO_3$ is still a difficult and costly art, and most of its applications to date have involved the ceramic material, often with various added impurities. The ceramic also exhibits dielectric nonlinearity, a typical result[6] being a decrease of dielectric constant with biasing field such that a field of 10 kv/cm lowers ϵ from 1,400 to 1,100, while a field of 30 kv/cm lowers it to 700. Similar results have been found at Stanford University, with an interesting additional qualitative observation that the loss tangent of a ceramic at radio frequencies may be lowered substantially by application of a biasing field of a few kv/cm. Further data on properties of ceramics have been given by von Hippel.[21]

Many details concerning the physical properties of $BaTiO_3$ have been omitted here; in particular the phenomena of domain formation and motion which are essential to an understanding of the properties of single crystals below the Curie point. These have been described by Forsbergh,[22] Merz,[23] and Little.[24] A recent discussion of the theory of ferroelectrics has been given by Devonshire.[25]

Theory of Dielectrics

It might be supposed that with modern knowledge of the properties of atoms and molecules, it would be a straightforward matter to calculate the dielectric constant of any material of known composition from first principles. Unfortunately, this turns out to be an extremely complicated problem on which little progress has been made; only in the case of gases where the dielectric constant is very close to unity and the similar case of dilute solutions of polar molecules in nonpolar liquids can one claim quantitative success. Although it

[15] J. C. Slater, *Jour. Chem. Phys.*, vol. 9, p. 16; 1941.
[16] C. C. Stephenson and J. G. Hooley, *Jour. Am. Chem. Soc.*, vol. 66, p. 1937; 1944.
[17] von Hippel, Breckenridge, Chesley, and Tisza, *Jour. Ind. Eng. Chem.*, vol. 38, p. 1097; 1946.
[18] M. E. Drougard, R. Landauer, and D. R. Young, *Phys. Rev.*, vol. 98, p. 1010; 1955.
[19] E. T. Jaynes, "Ferroelectricity," Princeton University Press, Princeton, N. J., Chap. 3; 1953.
[20] W. J. Merz, *Phys. Rev.*, vol. 76, p. 1221; 1949.
[21] A. von Hippel, "Dielectric Materials and Applications," John Wiley & Sons, New York; 1954.
[22] P. W. Forsbergh, Jr., *Phys. Rev.*, vol. 76, p. 1187; 1949.
[23] W. J. Merz, *Phys. Rev.*, vol. 88, p. 421, 1952; vol. 95, p. 690; 1954.
[24] E. A. Little, *Phys. Rev.*, vol. 98, p. 978; 1955.
[25] A. F. Devonshire, *Phil. Mag. Suppl.*, vol. 3, p. 85; 1954.

is easy to describe in words the mathematical procedure that would give a rigorous treatment, in practice one must make from the start drastic simplifications which make it difficult to interpret whatever agreement or disagreement with experiment is found. The basic reason for this is that the properties of a crystal are not merely the properties of a single molecule, multiplied by the number of molecules present, but the interactions between them are an essential part of the picture, and these are never taken into account in a correct way. From the point of view of quantum mechanics, we would have to say that it is not meaningful in any precise way to speak of the states and behavior of individual molecules or atoms, but the only really correct attitude is the global one in which we enumerate the possible quantum states (wave functions) of the crystal as a whole. Almost without exception, however, theoretical treatments of dielectric properties of solids have been based on the concepts developed by Clausius and Mosotti about 100 years ago. Here one regards a solid as composed of a large number of polarizable objects (in various cases the mechanism of polarization may be thought of as distortion of electronic distributions of atoms or ions, motion of ions, or rotation of molecular aggregates having a permanent dipole moment) each with polarizability α, so that each object, in an electric field F, develops a dipole moment

$$M = \alpha F. \tag{7}$$

The field F is not, however, the same as the macroscopic applied field E; because of the interaction of the polarizable objects with each other, there is an additional term commonly taken as proportional to the net polarization, with a proportionality constant β, known as the Lorentz factor.

$$F = E + \beta P. \tag{8}$$

Lorentz showed that if the polarizable objects are arranged in a cubic or random array, and each maintains the same constant dipole moment (no thermal agitation effects), β would have the value $4\pi/3$. If there are N of these polarizable objects per unit volume, the polarization is

$$P = N\alpha F = N\alpha(E + \beta P),$$

so that the dielectric susceptiblity becomes

$$\chi = P/E = N\alpha/(1 - N\alpha\beta). \tag{9}$$

Introducing the dielectric constant

$$\epsilon = 1 + 4\pi\chi, \tag{10}$$

and assuming the Lorentz value $\beta = 4\pi/3$, we arrive at the well-known Clausius-Mosotti formula

$$\frac{\epsilon - 1}{\epsilon + 2} = \frac{4\pi N\alpha}{3}, \tag{11}$$

which is presented in some textbooks as if it were a rigorous relation.

Many refinements of this treatment have been made, and a very complete account of them may be found in the recent book of Böttcher.[26] They have led to some improvement in agreement with experiment but not to any appreciably deeper understanding, because the basic concepts remain the local polarizability and local field F, which in modern theory no longer have a precise meaning. Nevertheless, this classical treatment contains enough of the truth to be very useful in giving a qualitative understanding of dielectrics, provided certain precautions are observed. In the first place, (11) cannot be used when the polarizability is due to freely rotating permanent dipoles (*i.e.*, polar molecules) except in the case of high dilution when it reduces to

$$\epsilon - 1 = 4\pi N\alpha \ll 1.$$

From statistical mechanics, one can calculate the polarizability of a rotating dipole of moment M, with the result $\alpha = M^2/3kT$, with k Boltzmann's constant and T the temperature in degrees Kelvin. Eq. (11) predicts an infinite dielectric constant, *i.e.*, ferroelectricity, when $N\alpha\beta > 1$, so that this should occur at sufficiently low temperatures for any substance with rotating dipoles. However, if we insert the numerical values, we find that many polar substances should be ferroelectric at temperatures far above their boiling points! This is the famous "$4\pi/3$ catastrophe," which was resolved by Onsager[27] with the observation that strong correlations between the motions of nearby dipoles reduce the effective Lorentz factor; the results of his approximate treatment are obtained if we formally replace β in the above equations by $4\pi/(2\epsilon+1)$; the opposite conclusion is then obtained that ferroelectricity does not occur unless the polarizability becomes infinite.

The fact that ferroelectricity is so easy to "explain" when one uses poor mathematical approximations has long plagued theoreticians and has delayed any reliable understanding of the true cause of ferroeelctricity. For example, Rochelle salt was for many years treated as an assembly of rotating dipoles exhibiting the catastrophe of (11). Another example of a model which predicts ferroelectricity as a result of poor approximation is an array of harmonic oscillators interacting with each other. Such an oscillator with a particle of charge e, mass m, and resonant frequency ω has a temperature-independent polarizability of $e^2/m\omega^2$, and therefore according to the above equations one can always produce a ferroelectric array by making the resonant frequency sufficiently small. However, this model is so simple that it can be treated rigorously; an orthogonal transformation of coordinates enables one to find the states of the array as a whole, and it is found when the problem is treated correctly that ferroelectricity cannot occur unless the polarizability of a single oscillator becomes infinite. We

[26] C. J. F. Böttcher, "Theory of Electric Polarisation," Elsevier Press, Amsterdam; 1952.
[27] L. Onsager, *Jour. Am. Chem. Soc.*, vol. 58, p. 1486; 1936.

also note that Slater's theory of KH_2PO_4, already mentioned as probably the best available theory of a ferroelectric, does not make use of electrostatic interactions but rather ones of a more direct mechanical nature. In spite of these and other considerations,[28] many people believe that in barium titanate we have the realization of the classical polarization catastrophe; in the present writer's opinion it is doubtful whether any existing theory is free of the effects of poor approximations of the above type, and interactions other than electrostatic may well be the essential ones.

If, disregarding these precautions, we assume that a classical type of theory should have at least a qualitative usefulness, what can be said about the expected occurrence of dielectric nonlinearity? According to the above equations, this would require that either the polarizability or the Lorentz factor be field-dependent. We consider separately the three cases that the polarization is due to: (a) rotating permanent dipoles; (b) translational motion of ions; or (c) electronic distortion of atoms or ions.

(a) *Rotating Dipoles*. The polarizability $\alpha = M^2/3kT$ given above is an approximation valid only at field strengths F such that $MF \ll kT$. The exact expression, first calculated by Langevin in 1905, is

$$\alpha = (M/F)L(a) = M^2/3kT - M^4F^2/45k^3T^3 + \cdots,$$

where $L(a) = \coth a - a^{-1}$ is the Langevin function and $a = MF/kT$. From this we find that at $a = 1$ the polarizability is lowered by about 6 per cent, and for $a > 5$, $L(a)$ is essentially equal to unity, so that α varies inversely as the internal field F. Since molecular dipole moments are of the order of 10^{-18} esu, we find that at room temperature an appreciable nonlinearity could be expected only for F greater than about 4×10^4 esu, or 1.2×10^7 v/cm. Since we must remember to use the Onsager field for F, the applied field E would have to be of the same order of magnitude; thus a measurable nonlinearity due to saturation of rotating dipoles could be expected only at very low temperatures and intense field strengths.

(b) *Translational Motion of Ions*. Here the prospects are considerably brighter. In many types of crystals the size of the lattice is determined by the larger ions that have to fit into it, and if small ions are also present, they may be free to move in the interstices, through distances of an appreciable fraction of an Angstrom, but cannot move further due to contact with the larger ions. It is seen without any calculation that this results in a contribution to the polarization of the crystal which saturates rather abruptly at a certain value. KH_2PO_4 is undoubtedly of this type, with movable hydrogens; other crystals of similar structure, even though not ferroelectric, might be expected to show nonlinearity. However, materials with high dielectric constants should provide the most favorable possibilities, since according to the above equations we then obtain an internal field F which is considerably "amplified" above the applied field E. If appreciable motion occurs, the Lorentz factors might also vary.

(c) *Electronic Distortion*. As a simple example, consider an atom which has a ground state ψ_0 with energy E_0, and an excited state ψ_1 with energy E_1, such that the matrix element of the dipole moment operator between them,

$$M_{01} = e\int \psi_0^* z\psi_1 dV$$

does not vanish.[29] Using quantum mechanics and statistical mechanics,[30] the following formula for polarizability may be obtained:

$$\alpha = \frac{|M_{01}|^2}{kT}\cdot\frac{\tanh a}{a}$$

where

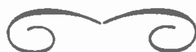
$$a = \frac{[\frac{1}{4}(E_1 - E_0)^2 + |M_{01}|^2F^2]^{1/2}}{kT}$$

It is seen that appreciable nonlinearity requires that a be of order unity or greater and that the term in F^2 must contribute substantially to a. Therefore, since M_{01} will typically be of the order of magnitude 10^{-18} esu, if the two energy levels are sufficiently close together the situation is about the same as in the case of rotating dipoles. If a high dielectric constant leads to great internal field strengths, these conditions might be met, although it appears that the case of movable ions remains the most favorable to development of strong nonlinearity.

[28] J. M. Luttinger and L. Tisza, *Phys. Rev.*, vol. 70, p. 954, 1946; vol. 72, p. 257; 1947.

[29] L. I. Schiff, "Quantum Mechanics," McGraw-Hill Book Co., Inc., New York, sec 25; 1949.

[30] See ref. 19 (pp. 58–60) for a similar calculation.



Some Aspects of Ferroelectricity*

G. SHIRANE†, F. JONA‡, AND R. PEPINSKY§

Summary—Some recent developments in the study of ferroelectricity are described. Special consideration is given to crystal and domain structures, and their contributions to the dielectric behavior of ferroelectrics and antiferroelectrics.

I. Introduction

CONSIDERABLE attention has been devoted to ferroelectric crystals in recent years, both because of their practical value for electronic and acoustical applications, and because they pose very intriguing and fundamental problems in solid state physics.

Rochelle Salt, KH_2PO_4 and $BaTiO_3$ are well-known ferroelectrics. They exhibit certain remarkable dielectric properties which are in many ways analogous to the magnetic properties of ferromagnetics. These analogies are the sole basis for terminology *ferroelectric* (used both as noun and adjective) and *ferroelectricity*. Just as ferromagnetic materials show a hysteresis effect in relationship of magnetic induction and field, ferroelectrics show hysteresis in dielectric displacement vs applied electric field relation. This behavior appears only in certain temperature ranges, depending on material concerned.

In its ferroelectric phase a crystal is spontaneously electrically polarized. The most important property of a ferroelectric is that the *direction* of polarization can be altered by an applied electric field. In typical ferroelectrics the spontaneous polarization diminishes as the crystal is heated, and it disappears at a temperature which is called the *ferroelectric Curie point*. The dielectric constant in the direction of spontaneous polarization is generally high and shows a very high peak at the Curie point; above this point, further heating results in a rapid decrease of the dielectric constant, according to the *Curie-Weiss law*. This law was first evolved to describe the temperature dependence of magnetic permeability in ferromagnetics.

The spontaneous polarization is accompanied by a spontaneous strain, and piezoelectric and elastic anomalies accompany the dielectric anomaly. Due to the polarization and strain, the symmetry of a crystal in its ferroelectric phase is lower than that of the nonpolar phase. This departure from the higher symmetry is slight, and a macroscopic ferroelectric crystal generally consists of multiple twins. A twin individual is known as a *domain*. Within a domain the polarization is uniformly directed; adjacent domains are twin individuals oriented and, hence, polarized in different directions. Domain orientations can be altered by application of a field. This domain reorientation is responsible for the D vs E hysteresis loop.

These are basic properties of ferroelectric crystals. They are treated in detail, and for various examples in the following sections. The field of ferroelectricity is already rich in experimental, theoretical, and practical developments, and we cannot hope to review all of these. Our special aim will be to discuss crystal and domain structures, and to relate dielectric properties to these as far as seems presently feasible.

II. Some General Considerations

A. Crystallographic Properties

It is well-known that crystals can be divided into 32 classes, according to their rotational symmetries (cf. Cady [1], p. 17). Of these classes, 11 have a center of symmetry, and are called centrosymmetric; the remaining 21 are noncentrosymmetric. Crystals in 20 out of the 21 noncentrosymmetric classes show the phenomenon of piezoelectricity; *i.e.*, upon application of an external stress, an electrical polarization is created within the crystal. 10 out of the 20 piezoelectric classes are called *pyroelectric*. Crystals in these latter classes are already polarized, since they possess at least one axis which shows properties at one end different from those at the other. The electrical polarization is usually masked by surface charge, and occasionally by twinning; but it can usually be observed in a nonconducting crystal if the temperature of the crystal is altered, and a *change* in the polarization is therewith induced. Hence the name pyroelectricity: electricity released by heat.

Very occasionally the direction of polarity of a pyroelectric crystal can be reversed by application of a sufficiently high electric field. *Such reversible pyroelectrics are ferroelectrics*. The reversal, of course, must be possible at field strengths less than the breakdown strength of the crystal.

Whereas one can deduce the presence of piezo- or pyro-electricity as soon as the crystal class is established (by optical or X-ray means, *e.g.*), dielectric measurements alone can establish the presence of ferroelectricity. The latter word is thus a dielectric, not a crystallographic, term.

* Original manuscript received by the IRE, August 22, 1955.

† Dept. of Physics, Tokyo Inst. of Technology, Oh-okayama, Tokyo. Permanent address: The Pennsylvania State University, University Park, Pa.
‡ Dept. of Physics, The Pennsylvania State University, University Park, Pa.
§ Physics Dept., Brookhaven National Lab., Upton, L. I., N. Y. Permanent address: The Pennsylvania State University, University Park, Pa.

Certain materials, generally closely related to known ferroelectrics, show a dielectric behavior which merits the term *antiferroelectric* (see Section III, *H*). It is not possible to define the term *antiferroelectric* crystallographically; the definition again involves dielectric behavior as well as structural properties. This point is deserving of present emphasis, since it has been confused in certain previous publications.

B. Analogies between Ferroelectricity and Ferromagnetism

The words *ferroelectric* and *antiferroelectric* are perhaps unhappy choices, and are based solely upon analogies between the dielectric behavior of certain crystals and the magnetic behavior of ferromagnetic materials. Firstly, it is found that ferroelectric crystals are generally comprised of multiple twins. In each twin individual the spontaneous polarization is directed along a specific crystallographic direction. But the twin individuals are disposed at various angles to one another crystallographically; hence the polarization is in a different direction from one individual to the next. These individual regions are called *ferroelectric domains*, in analogy with magnetic domains in ferromagnetic crystals. The properties and macroscopic physical effects of ferroelectric domains will be considered in detail in Section VII.

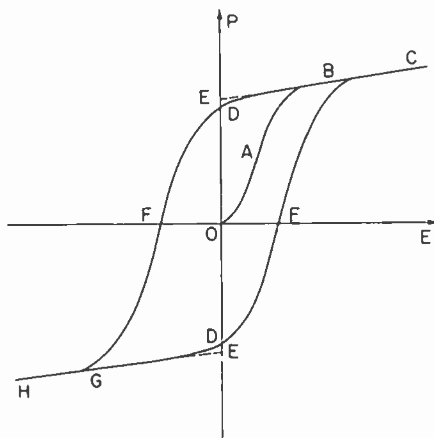


Fig. 1—Ferroelectric hysteresis loop (schematic). OE = Spontaneous polarization P_s, OF = Coercive field strength E_c.

Secondly, the very characteristic hysteresis loop, which results when one plots magnetization vs alternating magnetic field in ferromagnetic materials, has a precise analogy for the electric case. A ferroelectric hysteresis loop is shown schematically in Fig. 1. Its cause is immediately understandable on the basis of the domain concept. Consider first a crystal consisting of an equal number of positive and negative domains; *i.e.* the domains are anti-parallel with respect to some given crystallographic direction. Upon increasing the field in the positive direction, the positive domains grow at the expense of the negative domains. The polarization increases very rapidly (Fig. 1, OA), and reaches a saturation value (BC), at which point all domains are aligned in the direction of the field. This means that the crystal now has a "single domain" structure. When the field is reduced to zero again, a few domains remain aligned. At zero applied field a finite value of the polarization can be measured, called the ***remanent polarization*** P_r (OD). Extrapolation of the linear portion BC of the hysteresis loop back to the polarization axis yields the value of the ***spontaneous polarization*** P_s (OE). In order to annihilate the remanent polarization P_r, we must apply an electric field in the opposite (negative) direction. The field needed for this purpose is called the *coercive field* E_c (OF). Upon further increase of the field in the negative direction, uniform alignment of the dipoles can again be achieved, this time in the direction opposite to the previous one (GH).

The ferroelectric hysteresis loop can be directly observed on a cathode ray oscilloscope by means of a simple circuit first described by Sawyer and Tower [2]. The value of the spontaneous polarization P_s can be determined by measuring the distance OE on the observed loop on a calibrated screen of the cathode ray tube. It is also possible, of course, to determine the temperature dependence of the spontaneous polarization P_s of the ferroelectric crystal, by observing the change of the distance OE as a function of the temperature of the crystal.

Most ferroelectric crystals have a transition temperature above which they are no longer polar, even though they may still be piezoelectric. This transition temperature, which marks a phase change, is called the *Curie temperature* or *Curie point*. The dielectric constant, generally quite high, shows a very high peak at the Curie temperature; above it the dielectric constant decreases very rapidly according to a relation which is known, again in analogy to ferromagnetism, as the Curie-Weiss law:

$$\epsilon = \epsilon_0 + \frac{C}{T - \theta}. \quad (1)$$

Here ϵ represents the dielectric constant, ϵ_0 the electronic contribution to the dielectric constant, C the *Curie constant*, T the absolute temperature and θ the *characteristic temperature* (also called the *extrapolated Curie temperature*). θ is generally lower than the transition temperature.

C. The Dielectric Constant

Both the anomaly of the dielectric constant at the transition point and the Curie-Weiss law are characteristic features of a ferroelectric. In fact, they are predicted by the thermodynamic theory of the transition. The definition of the dielectric constant of a ferroelectric crystal deserves particular attention. The dielectric

constant is linearly related to the slope of the P vs E curve. Well above the Curie temperature, in the non-polar state, the relation between polarization P and field E is linear, and, therefore, the dielectric constant is field-independent. In the ferroelectric region, however, the existence of a hysteresis loop clearly demonstrates that the value of the dielectric constant depends on the field strength with which it is measured. If the applied field is very small, no new domains are created and no movement of the domain boundaries takes place. In this case one measures the dielectric constant of the crystal with no interference from the domain structure. This quantity, which is called the *initial* dielectric constant, is directly proportional to the slope of the *virgin curve* OA (Fig. 1) at the zero point; and it is this constant to which we refer when we speak of *the dielectric constant* of a ferroelectric crystal.

D. Other Properties of Ferroelectrics, and their Interrelations and Interpretation

It can be shown, on the basis of thermodynamic calculations, that a number of other changes are to be expected at the transition temperature. The spontaneous polarization is accompanied by the spontaneous strains, consequent upon the atomic shifts taking place within the lattice. Not only does the dielectric constant show an anomaly at the Curie temperature, but so also do the piezoelectric moduli and the elastic constants which are connected with the spontaneous strains. Moreover, there will generally be an anomaly of the specific heat at the transition temperature, and the shape of this anomaly will depend upon whether the transition is of the first or of the second kind.

Several phenomenological theories have been developed to relate the results of macroscopic measurements. Among these are treatments by Mueller [3] for Rochelle Salt and by Devonshire [4, 5] for $BaTiO_3$. The general procedure is to expand the free energy as a function of independent variables such as polarization and stress, and to determine the coefficients of expansion in such a manner as to match the experimental results.

A main purpose of both theoretical and experimental studies is to account for physical behavior in terms of specific crystal-structure models. Phenomenological theories cannot hope to provide specific model information, but they can disclose what thermodynamic properties are important, and what features are to be sought, in a model. An excellent recent article by Devonshire [6] discusses these matters.

The most fruitful approach to development of specific models is via X-ray and neutron crystal diffraction analyses. Large advances have been made with such studies in the past two or three years, and some of these are reported below. Another method which appears promising as a source of information at the atomic level is that of nuclear quadrupole resonance, as illustrated by the study of Cotts and Knight on $KNbO_3$ [7].

When crystal structural principles are established, some useful results may be achievable—in the matter of predicting new ferroelectrics, *e.g.*—by seeking common structural features within a given family of ferroelectrics. The best illustration of the procedure is Matthias' consideration of conditions for ferroelectric activity in the oxygen-octahedra class [8].

The above problems are concerned with single-domain properties. A series of very significant problems is concerned with domain structures and interactions: how and why domains are formed, and how they respond to external fields and stresses. These matters are reviewed to some extent in Section VII.

In the next section we introduce some specific ferroelectric crystals and some aspects of their behavior.

III. Some Representative Ferroelectrics

A. Four Families of Ferroelectrics

We may group known ferroelectrics into four families, on crystal-chemical bases.

The first to be discovered was Rochelle Salt (abbreviated RS), $NaKC_4H_4O_6 \cdot 4H_2O$, a double tartrate of sodium and potassium with four molecules of water. Rochelle Salt is one of several ferroelectric tartrates, which we shall consider to form the "tartrate family." This family is comprised of two groups: the first contains RS, and isomorphous mixed crystals with the same water content; the second contains isomorphous double tartrates with lithium as one of the cations, and one water molecule per tartrate group. Although the groups differ one from the other structurally, the disposition of the tartrate ions is remarkably similar in both.

The second group to be discovered was that of the alkali metal dihydrogen phosphates and arsenates, represented by KH_2PO_4 (abbreviated KDP). Also in this group are $NH_4H_2PO_4$ and the corresponding arsenate, which are antiferroelectric.

The third class of ferroelectrics will be entitled the oxygen-octahedra family, on the basis of their crystal chemistry, to be discussed later. This includes the very important sub-group whose structures are slight distortions of that of the mineral *perovskite*, $CaTiO_3$; the sub-group is thus called the perovskite-type ferroelectrics. The best known member is barium titanate, $BaTiO_3$; but the number of ferroelectrics and antiferroelectrics to be found among simple and mixed perovskite-type crystals (solid solutions) is tremendously great. Other sub-groups include: the so-called rhenium trioxide type, of which WO_3 is a representative; and the pyrochlore type, represented by cadmium niobate, $Cd_2Nb_2O_7$, and isomorphic mixed crystals. The entire family will be considered in detail in the present article.

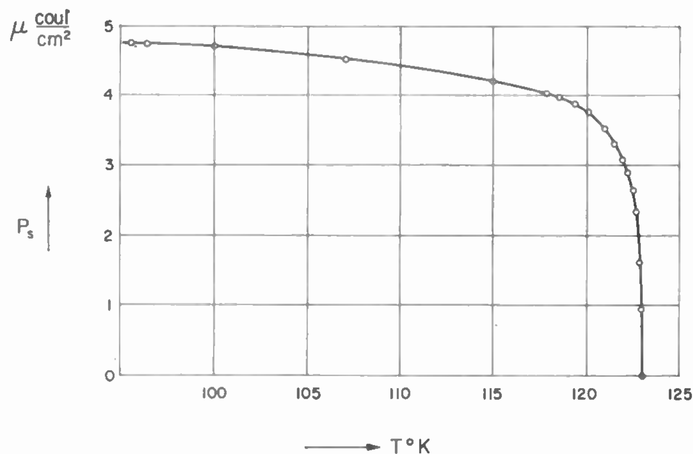


Fig. 2—Spontaneous polarization of KH_2PO_4 as a function of temperature (after von Arx and Bantle [13]).

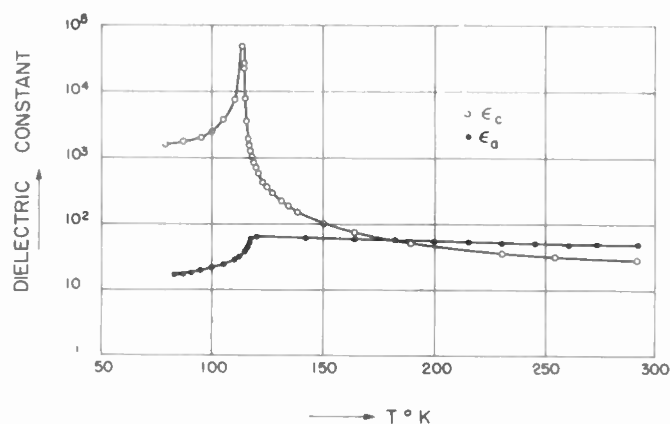


Fig. 3—Dielectric constant of KH_2PO_4 as a function of temperature (after Busch [11]). Temperature scale should be shifted so that peak appears at 123°K.

The final family is the latest to be found, and is a quite new and unrelated type. It is represented by guanidine aluminum sulfate hexahydrate (abbreviated GASH), $NHC(NH_2)_2AlH(SO_4)_2 \cdot 6H_2O$. The structure of this group is unknown at the present writing, but is probably readily obtainable by X-ray methods. The crystals decompose before a Curie temperature can be reached.

Large crystals of the ferroelectric tartrates and dihydrogen phosphates can be grown easily from solution. The properties of these crystals are described in detail in the treatises of Cady [1] and Mason [9], and in review articles such as that by Baumgartner, Jona, and Kaenzig [10]. Earlier difficulties in the growth of single crystals of $BaTiO_3$ have been overcome (see below), and studies of these are in progress in many laboratories. GASH and its isomorphs grow readily from solution in good crystals.

B. Potassium Dihydrogen Phosphate

KDP satisfies perfectly the general description of a ferroelectric crystal given above, and we discuss it first for this reason. Its Curie temperature lies at about 123°K [11]. The crystal has nonpolar tetragonal symmetry above the Curie point, and is piezoelectric. The symmetry changes to orthorhombic below the transition point [12], where the crystal shows ferroelectric properties (and is, of course, still piezoelectric). The direction of spontaneous polarization lies along the orthorhombic c axis, which coincides with the original tetragonal c axis (Fig. 2) [13]. The spontaneous strain is a shear of about 27′ in the xy plane. The behavior of the dielectric constant along the c axis above the Curie point is well represented by the Curie-Weiss law (cf. Fig. 3) with $\epsilon_0 = 4.5$, $C = 3{,}100$°K, and $\theta = 121°$. The dielectric behavior in the neighborhood of the Curie temperature was extensively investigated by Baumgartner [14]. KDP is representative of a group of ferroelectrics, all members of which are isomorphous with it; among these are KH_2AsO_4, RbH_2PO_4, the deuterated salts (crystals in which the hydrogen is replaced by deuterium), and various mixed crystals involving the above ions.

The structure of KDP, first determined by West [15], is comparatively simple, and is discussed in some detail in Section VI. Slater [16] developed a theory which explained the transition to the ferroelectric phase on the assumption of an ordering of the protons involved in hydrogen bonding. Experimental support for this theory was first provided by the X-ray investigation of Frazer and Pepinsky [17], who determined the crystal structures of the two forms directly on either side of the transition temperature. This study showed a lengthening of O—H · · · O bonds as the Curie temperature was passed from above, indicative of a tendency toward ordering of the H atoms in the ferroelectric state. This ordering was beautifully confirmed by neutron diffraction analyses of Peterson *et al.*, [18], Levy *et al.* [19], Pease and Bacon [20], Bacon and Pease [21].

C. Rochelle Salt

Rochelle Salt has a Curie point at about 23°C; but this crystal is very unusual in that a lower transition occurs at a temperature of −18°C. The latter temperature is usually called the "lower Curie point." *The crystal is ferroelectric only in the region between the two Curie points.* Above the upper Curie point, and below the lower, the symmetry is orthorhombic and the crystal is piezoelectric. In the ferroelectric region the symmetry is monoclinic, and spontaneous polarization occurs along crystallographic direction which corresponds to original orthorhombic a axis (see Fig. 4, next page). The spontaneous strain is a shear of approximately 3′ in yz plane. According to crystallographic convention the orthorhombic a axis of the ferroelectric crystal should now be denoted as the monoclinic b axis; for the sole symmetry element of the monoclinic phase is a twofold screw axis, and the polarization can be only along this direction. However, it has become customary to maintain the notation "a" for the polar axis in the

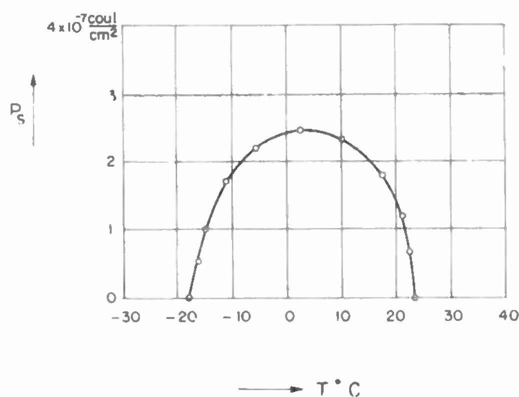


Fig. 4—Spontaneous polarization of Rochelle Salt as a function of temperature (after Hablnetzel [22]).

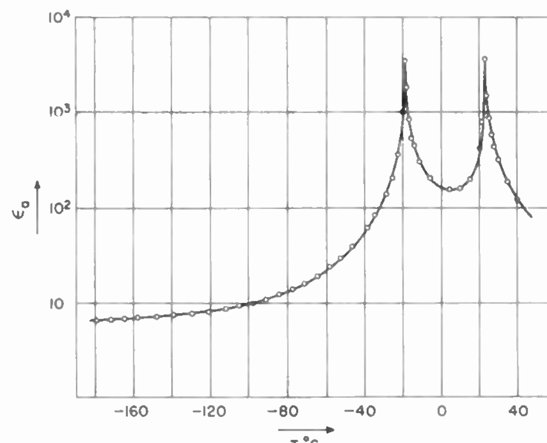


Fig. 5—Dielectric constant ϵ_a of Rochelle Salt as a function of temperature (after Hablnetzel [22]).

monoclinic phase; and we follow this practice to avoid confusion. The dielectric constant measured along a follows the Curie-Weiss law above the upper Curie point, with $C=2{,}240°K$ and $\theta=23°C$. The behavior of ϵ_a at the upper and lower Curie points is shown in Fig. 5 [22].

The phenomenological theory of the RS behavior has been thoroughly developed by Mueller [3]. The domain structure has been studied directly by optical means only recently by Mitsui and Furuichi [23, 24] (see Section VII). The crystal structure is much more complicated than that of KDP, and was reported originally by Beevers and Hughes in 1941 [25]. Several theories have been developed which make certain hydrogen bonds responsible for the occurrence of the phase transition [26]. These are based upon the Beevers-Hughes analysis. A careful X-ray re-examination of the structure by Sundara-Rao, Mazzi, and Pepinsky [27], and a similarly careful single-crystal neutron study by Frazer, Danner, and Pepinsky [28], reveal significant departures from the 1941 structure results. Among these discrepancies is one involving the particular hydrogen bond to which the polarization was chiefly attributed. This bond is now revealed as the least likely to be involved in the entire structure, and an entirely new mechanism is suggested on the basis of the more precise structure.

The nature of the lower transition is incompletely understood, and a satisfactory theory of the dielectric behavior in the neighborhood of the lower Curie point is not available.

Substitution of very small amounts of isosteric ions in place of potassium reduces the temperature range of ferroelectric activity in Rochelle Salt rapidly, substitution of about two per cent eliminating the spontaneous polarization entirely. Substitution of deuterium for those hydrogens which are easily replaceable in solution widens the ferroelectric temperature range somewhat. A surprising effect, first discovered by Kurtschatov [29], appears upon substitution of *twelve per cent or more* of potassium ions by NH_4^+. These partially "ammoniated" Rochelle Salt mixed crystals show single Curie points at low temperature, the exact position of the transition depending upon the composition. The dielectric properties of NH_4-substituted RS crystals have been re-examined recently [30], and Kurtschatov's results were confirmed and extended.

D. Other Tartrate Ferroelectrics

Ferroelectricity at low temperature has been recently discovered in two other double tartrates: $LiNH_4C_4H_4O_6\cdot H_2O$ (LAT) and $LiTlC_4H_4O_6\cdot H_2O$ (LTT), by Matthias and Hulm [31] and by Merz [32] (LAT only). LAT is orthorhombic and nonpolar[1] above 100°K approximately; below that temperature it is ferroelectric and probably monoclinic. The crystal structure of the nonpolar phase was first determined by Vernon and Pepinsky [33]. The structure is remarkable in that the tartrate ions are disposed in the lattice in a manner closely related to the arrangement of the tartrate groups in Rochelle Salt, although the spontaneous polarization in LAT is directed along the orthorhombic b axis, while in RS it is directed along the a axis. The dielectric behavior is notable in that the maximum of the dielectric constant, at the Curie point, is low (in the neighborhood of 100) compared to previously-known ferroelectrics.

LTT is still more remarkable in that its Curie point is at 10°K, and the direction of spontaneous polarization is at right angles to that in LAT—although in their nonpolar phases the crystals are isomorphous. LAT and LTT are discussed in a little more detail in Section VI.

E. Oxygen-Octahedra Ferroelectrics

A third family of ferroelectrics, after the tartrates and dihydrogen phosphates, is that of the oxygen-octahedra types. The best known and to date most

[1] The term paraelectric is often used to describe the phase in which a crystal is neither ferroelectric nor antiferroelectric. We may introduce it occasionally.

important sub-family is that of the perovskite-type ferroelectrics, of which the best known representative is barium titanate, $BaTiO_3$. The common chemical formula for the perovskite group is ABO_3, where A is mono- or di-valent. The structure of the nonpolar phase is shown in Fig. 6.

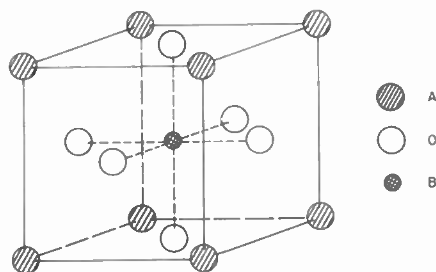


Fig. 6—Cubic perovskite-type structure ABO_3.

The basic feature of the oxygen-octahedra family is evidenced in this perovskite structure. Here a small, highly-polarizable ion is surrounded by an octahedron of oxygens[2]; and adjacent octahedra are closely bonded one to another through sharing of corners. Some or all spaces between the linked octahedra structure may or may not be filled with other ions. Tungsten trioxide, WO_3, is comprised solely of linked WO_6 octahedra, and is representative of the so-called rhenium-trioxide structure. In barium titanate, spaces remaining between TiO_6 octahedra are filled with Ba ions. In the nonpolar phases these structures are cubic. A large number of perovskite-type crystals lose the high cubic symmetry as their temperature is reduced below a specific temperature. They may show ferroelectric or other anomalous dielectric behavior as accompanying phenomena.

Another oxygen-octahedra-type ferroelectric is exemplified by the pyrochlore structures, represented by cadmium niobate, $Cd_2Nb_2O_7$ [34, 35]. In this structure the linking of NbO_6 octahedra is more complicated than in the perovskites. The structure will be discussed in Section VI.

Ferroelectric properties also have been discovered, recently, in $PbNb_2O_6$ [157].

Finally, certain structures which were assumed to be ilmenites (after the mineral, $FeTiO_3$, of this name) are thought to be ferroelectric [36, 37]. Unfortunately, more work needs to be done both on the dielectric and structural properties of these.

F. Some Properties of $BaTiO_3$

In contrast to KDP and RS, in which only one ferroelectric phase appears, $BaTiO_3$ exhibits *three different* ferroelectric phases. The temperature dependence of the dielectric constant [38] is shown in Fig. 7, and of the spontaneous polarization in Fig. 8. The Curie temperature is about 120°C; above it the dielectric constant follows the Curie-Weiss law, with $C = 170{,}000°K$ and $\theta = 100°C$. The Curie constant C has a much larger value than in the case of KDP or RS. Above 120°C the structure is cubic, as in Fig. 6. Below this temperature, and down to 5°C, the symmetry is tetragonal, and a spontaneous polarization exists along a former cube edge, which is now the tetragonal axis. Between 5°C and −80°C the symmetry is orthorhombic, and the direction of spontaneous polarization transfers to that of a face diagonal of the former cubic cell. Below −80°C the direction of polarization is along a former body diagonal of the cube, and the symmetry is rhombohedral. Spontaneous strains accompany the various polarizations. The strain in the tetragonal phase, for example, is an extension in the direction of polarization and a contraction in directions perpendicular to this.

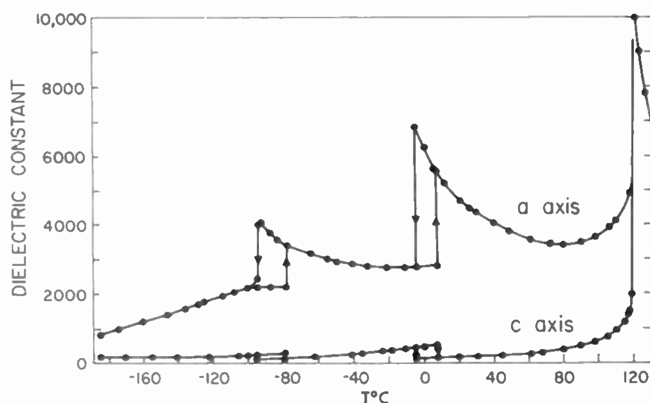


Fig. 7—Dielectric constant of $BaTiO_3$ single crystals as a function of temperature (after Merz [38]).

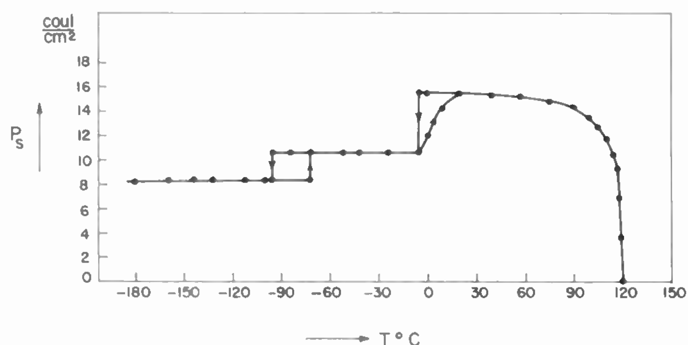


Fig. 8—Spontaneous polarization of $BaTiO_3$ single crystals as a function of temperature (after Merz [38]).

The discovery of ferroelectricity in $BaTiO_3$ in 1943 was a great stimulus to development of this field of solid state physics. The simple structure of the nonpolar phase, compared to the previously-known ferroelectrics, encouraged investigators to presume that the

[2] The six oxygens constitute the corners of an eight-faced octahedron (see Fig. 22).

mechanism of ferroelectric activity in the perovskite class could be more readily elucidated than in other materials. The stable chemical properties and high mechanical strength, coupled with advantageous dielectric properties, offered large promise for electromechanical transducer and other applications.

The realization of these advantages was retarded by difficulties of preparation of single crystals, both for investigative and practical purposes; early studies were restricted to ceramic specimens. Successful preparation of single crystals by Blattner, Matthias, Merz, and Scherrer [39] in 1947 opened a new investigative phase. Recently, Remeika [40] has developed a method for growing large single crystals out of a potassium fluoride flux. The availability of large crystals, from Remeika's process, has permitted several basic and practical advantages. The basic studies include dielectric observations near the Curie point by Merz [41], domain wall studies by Merz [42] and by Little [43], and a successful single-crystal structure analysis of the tetragonal phase by Frazer, Danner, and Pepinsky [44]. On the practical side, the feasibility of use of $BaTiO_3$ crystals for computer memory elements is under study in several laboratories.

Because of its theoretical and practical importance, $BaTiO_3$ properties will be discussed *in extenso* in Section IV.

G. Other Perovskite-Type Ferroelectrics

A number of perovskite-type crystals have been found to be ferroelectric. These include $PbTiO_3$, $KNbO_3$, and various solid solutions involving these and other cations. $KNbO_3$ shows a behavior closest to that of $BaTiO_3$, evidencing three successive transitions with decreasing temperature [45–47]. The phase diagrams and dielectric behavior of perovskite-type solid solutions is one of the major problems in ferroelectricity and one of extreme practical importance. The matter will be discussed to some extent in Section V.

The discovery of a number of oxygen-octahedra ferroelectrics was facilitated by consideration first advanced by Matthias [8]. He pointed out that the existence of small, highly polarizable ions with inert gas configurations, enclosed within interlinked oxygen octahedra, is highly favorable for the occurrence of ferroelectricity. In addition to new perovskite-type compounds, this principle led to the examination and discovery of ferroelectric behavior in $LiTaO_3$, reported to have an ilmenite-type structure [36].

H. Some Antiferroelectrics

A number of crystals have transitions above which the dielectric constant behaves as it does above a ferroelectric Curie point. Yet, below the transition temperature no spontaneous polarization occurs and hence no hysteresis loop is observed. Typical examples are $PbZrO_3$ [48–50], $NaNbO_3$ [51–52], $PbHfO_3$ [53], and various mixed crystals in the perovskite class [54–56]; $NH_4H_2PO_4$ [57–59], which has a room-temperature structure isomorphous with that of KH_2PO_4; $(NH_4)_2H_3IO_6$ [60] and $Ag_2H_3IO_6$ [61]. The structural nature of antiferroelectricity will be described in Section VI.

I. Guanidine Aluminum Sulfate Hexahydrate

Guanidine aluminum sulfate hexahydrate (GASH) has been found to be ferroelectric over a wide temperature range including room temperature [161]. The symmetry of this crystal is trigonal, which makes the trigonal axis the direction of spontaneous polarization. Dielectric measurements were done by Merz *et al.* [162], with the surprising result that, at room temperature, the dielectric constant along the trigonal (polar) axis is extremely small (about 6), while it is even smaller in the direction perpendicular to the polar axis (about 5). No Curie point can be detected because above 100°C the crystal starts to lose water of crystallization and reproducible results cannot be obtained. The spontaneous polarization is about 0.35 microcoulomb/cm² at room temperature and decreases linearly with increasing temperature, indicating the existence of a Curie point somewhere between 200 and 300°C. The coercive field strength is about 1,500 volt/cm, at room temperature and 60 cps, showing a strong frequency dependence.

It appears that a number of crystals isomorphous with GASH are also ferroelectric [161]. These crystals are obtained upon replacing the aluminum cation by Ga and/or Cr, and the sulfate anion by SeO_4. The ferroelectric properties of these isomorphs are supposedly very similar to those of GASH.

IV. Barium Titanate

A. Crystal Growth

The first single crystals of $BaTiO_3$ obtained by the Swiss group in 1947 were grown by using $BaCl_2$ as a flux [39]. Although single crystals obtained by this method were very small (a few millimeters on an edge), some important features of the three phase transitions were made clear by the domain structure study of Forsbergh [62], X-ray and optical studies by Kay [63], and Kay and Vousden [64], and dielectric measurements by Merz [38], all utilizing this material.

Several other methods, such as that of flame fusion and the Stockbarger process from a pure melt, were tried in attempts to obtain larger single crystals. The methods did give larger single crystals; but these were badly twinned, and it was almost impossible to cut thin plates, required for the dielectric measurements, which were free of large stresses. As we will see below, the dielectric properties of $BaTiO_3$ single crystals are very sensitive to strain.

It was a key experimental advance when Remeika [40] developed his method for growing large platelike single crystals, using KF as a flux. In this method, a mixture of $BaTiO_3$ and KF, with 0.02 per cent of Fe_2O_3 added, is placed in a platinum crucible, heated to about 1,200°C,

and then cooled slowly from 1,200°C to 900°C. At this point still-liquid flux is poured off and the crystals are annealed by slow cooling to room temperature. This method produces thin clear plates, always paired in a manner which leads them to be called "butterfly wings" (cf. Fig. 9) [65]. The small amount of Fe_2O_3 supplies acceptor ions to compensate for donor ions produced by loss of oxygen during crystal growth. This compensation avoids high conductivity in cooled crystals.

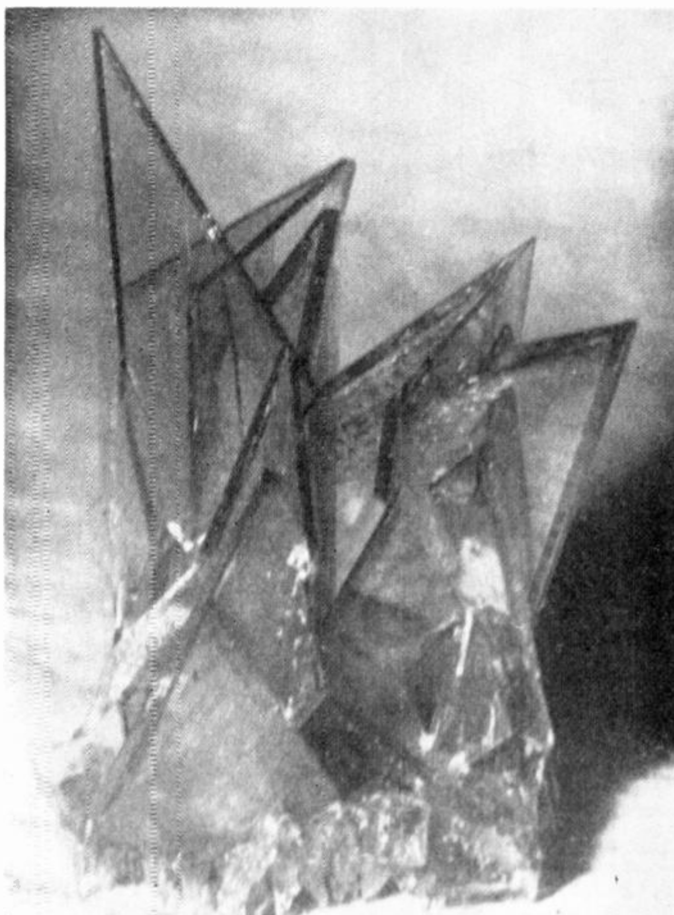


Fig. 9—Single crystals of $BaTiO_3$ grown by Remeika's method (after Smakula and Sils [65]).

Thin plates produced in this way can be as large as a few centimeters on edge, and they are more nearly perfect and "single domain" than crystals previously obtained. They show rectangular hysteresis loops and a low coercive force, both of which are important in applications such as for memory storage elements. They have, furthermore, permitted detailed studies of dielectric properties, such as that of Merz [41] and other studies previously noted [42, 44].

It must be mentioned in this discussion of crystal growth that an extensive study of the phase diagram of the BaO-TiO_2 system was carried out by Rase and Roy [66] (Fig. 10); and their results have revised, in several important ways, the phase diagram previously given by Statton [67]. Among the important results of Rase and Roy's study was the observation that $BaTiO_3$ transforms from the cubic to a hexagonal phase at 1,460°C, before it melts at 1,612°C. This hexagonal phase can be quenched to room temperature by rapid cooling through 1,460°C. This explains the occurrence of hexagonal crystals under some conditions in previous crystal-growth studies. The structure of this nonferroelectric, hexagonal $BaTiO_3$ was studied by Burbank and Evans [68], and it was shown that TiO_6 octahedra exist as an essential framework of the lattice, but the octahedra are linked differently than in cubic $BaTiO_3$. The phase diagrams of the $BaTiO_3$-$BaCl_2$ system, determined by Roy *et al.* [69], and of the $BaTiO_3$-KF system as elucidated by Karan *et al.* [70] were of large value for the successful method of crystal growth.

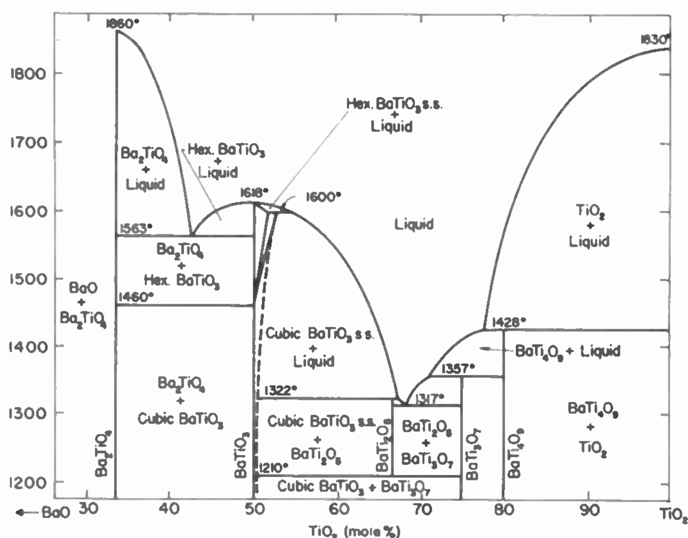


Fig. 10—Phase equilibrium diagram of the BaO-TiO_2 system; s.s. = solid solution. Some fields have been left unlabeled owing to lack of space (after Rase and Roy [66]).

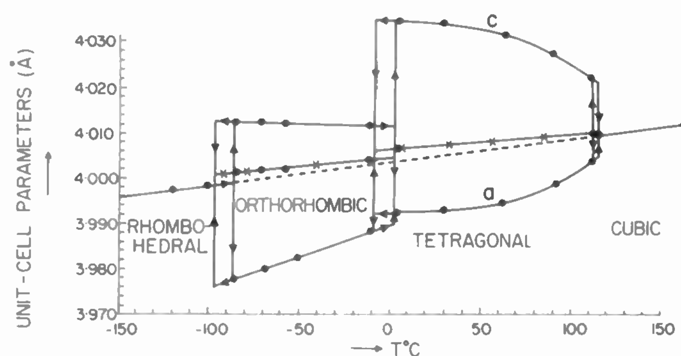


Fig. 11—Lattice parameters of $BaTiO_3$ as a function of temperature (after Kay and Vousden [64]).

B. Dielectric Properties

We now consider in some detail the temperature dependence of the dielectric constant, and its relation to the symmetry of the crystal (Fig. 7). Fig. 11 shows the temperature dependence of the spontaneous strain, as measured by Kay and Vousden [64]. Above the Curie point the crystal has cubic symmetry, and the value

of the dielectric constant is independent of the direction along which a small ac measuring field is applied. The validity of the Curie-Weiss law in the cubic region was already mentioned. At the Curie point, about 120°C, the dielectric constant reaches a very high value of about 10,000; then it shows a sudden drop when the crystal becomes ferroelectric, on further cooling.

The crystal assumes tetragonal symmetry below the Curie temperature, the tetragonal c axis being the direction of the spontaneous polarization, and with c only slightly longer than a. At room temperature, the distortion is given by the axial ratio $c/a = 1.01$ [71]. The crystal being tetragonal now, two dielectric constants are needed in order to describe the dielectric behavior of the crystal. One, ϵ_c, is measured upon applying the ac field along the c axis; the other, ϵ_a, is measured upon applying the ac field perpendicular to the c axis. However, a crystal plate which is cooled below the Curie point normally shows a rather complicated domain structure (see Section VII); thus measurement with a small ac field applied perpendicular to the major surfaces would furnish an *average* dielectric constant, whose value would depend on the relative ratio between numbers of c domains and a domains. In order to avoid this confusion, Merz selected, at room temperature, a single domain plate with the c axis perpendicular to the major surfaces for the measurement of ϵ_c; and for the measurement of ϵ_a he selected a single domain plate with the c axis *in* the major surfaces.

One important result revealed by this measurement was that ϵ_c is a few hundreds at room temperature, and is much smaller than ϵ_a, which is several thousands. The dielectric constant of ceramic $BaTiO_3$ is approximately $\epsilon = \frac{1}{3} \cdot (\epsilon_c + 2\epsilon_a)$, and it was known to be about 2,000. The physical meaning of this smaller ϵ_c is that the ions and electrons are easier to move in the direction perpendicular to the polarization under an electric field. A similar situation holds for the refractive index, which shows $n_c < n_a$. In this case only electrons are concerned.

Upon decreasing the temperature further, a second transition takes place in the vicinity of 0°C. The crystal changes symmetry to orthorhombic, while still remaining ferroelectric. The polar axis jumps discontinuously to a direction which corresponds to the original cubic [110]; this direction is usually identified with the orthorhombic a axis. The orthorhombic b axis is taken as lying in another cubic [1$\bar{1}$0] direction, perpendicular to a, while the orthorhombic c axis lies perpendicular to both a and b, and parallel to the direction of an original cube edge [100]. A thermal hysteresis of about 10°C is associated with the transition, in the sense that, upon cooling, the transition takes place at about -5°C while upon heating, it takes place at about $+5$°C. Since the crystal is now orthorhombic, its dielectric behavior will be described by three dielectric constants, ϵ_a, ϵ_b, and ϵ_c, along the orthorhombic axis a, b and c, respectively.

It is sometimes convenient to use monoclinic parameters $a' = b'$, c' and γ to describe this orthorhombic unit cell. They are related to orthorhombic parameters by:

$$a = 2a' \sin \frac{\gamma}{2}, \qquad b = 2a' \cos \frac{\gamma}{2}, \qquad c = c'.$$

(See Fig. 12.) (According to crystallographic convention the c' axis should be the monoclinic b.) These monoclinic parameters are directly related to the cubic unit cell and they can easily show a distortion from the cubic cell. The a' and c' axes are shown in Fig. 12; and $\gamma = 90°8'$.

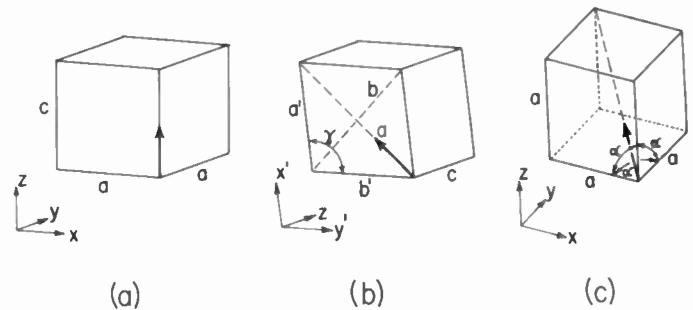


Fig. 12—Unit cells of the three different phases of $BaTiO_3$: (a) tetragonal, (b) orthorhombic, referred to monoclinic axes, (c) rhombohedral. Arrows show direction of spontaneous polarization. [In (b) the monoclinic axes have been labeled according to the convention used in most papers, which is in contradiction to the crystallographic convention.]

The portion of Fig. 7 between 0° and -80°C is now understandable on the basis of the structural change, and with the help of Fig. 13. Plate (B) used for the measurement of ϵ_c in the tetragonal phase now assumes the configuration (D), since this implies a jump of the polar axis of 45°, more likely than (E), which implies a jump of the polar axis of 90°. Since the measuring field is still applied along the original cube edge, one measures $\epsilon_{a'}$ in this case. The plate (C), which was

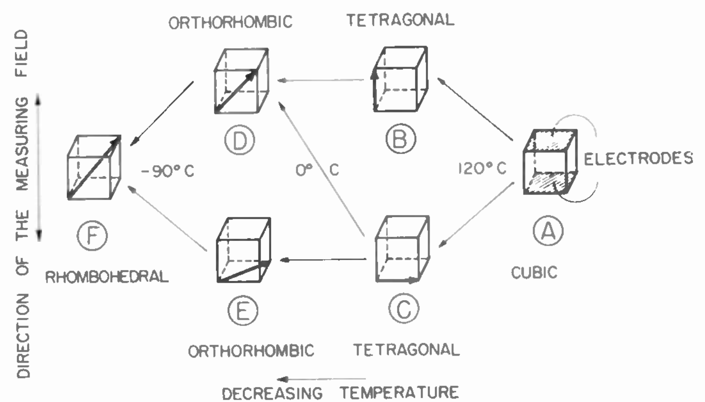


Fig. 13—Relative positions of the polar axis in the three phases of $BaTiO_3$ single domain crystals with respect to the cubic axes (schematic). The polar axis is indicated by the heavy arrow.

used for the measurement of ϵ_a in the tetragonal phase, can take the orientations (D) or (E). In the first case one would measure again $\epsilon_{a'}$, and in the latter one measures ϵ_c. The continuation of the experimental curve for ϵ_a should therefore represent some average of $\epsilon_{a'}$ and ϵ_c. The difference between this curve and $\epsilon_{a'}$ must be attributed to the ϵ_c component.

The preferred orientation at the transition can also be seen from the fact that c' is smaller than $a'=b'$, as seen in Fig. 11. Therefore the tetragonal c axis preferably changes to the monoclinic b' rather than to the monoclinic c', while the tetragonal a can take either a' or c'. From this experiment alone we cannot deduce the individual values ϵ_a, ϵ_b, and ϵ_c. It seems probable, from the experimental curve, that ϵ_c is the largest of these, however. The optical observation shows that $n_c > n_b > n_a$, which suggests a similar relation for the dielectric constants. If this is the case, the polarization direction again is the direction of strongest bondings for both electrons and ions.

Upon further decrease in temperature, a third transition takes place around $-80°C$, with a thermal hysteresis of approximately 15°C. The symmetry of the crystal changes to rhombohedral. The rhombohedral unit cell is equivalent to a slightly elongated cubic cell stretched along the [111] direction; this latter is the polar direction, and the structural parameters are given by $a=b=c$ and α which is about 89°52′. The only possible configuration now is (F), as shown in Fig. 13. In this case, an electric field applied along the original [100] direction would furnish only one value of the dielectric constant, which can be written in terms of ϵ_1 and ϵ_2, where ϵ_1 is the dielectric constant measured along the polar axis and ϵ_2 is the one which is measured perpendicular to it.

Even in the probable case in which the plate does not remain a single domain, but rather assumes a multidomain pattern, the measured dielectric constant should be the same whatever the original configuration at room temperature. The fact that the experimental curve still shows two branches below $-90°C$, therefore, is not understandable. The only possible explanation for this discrepancy is the existence of microscopic cracks, due to internal strains, in greater amount in some specimens than in others. In any case the dielectric measurements of the rhombohedral phase should be repeated in order to clarify this point.

The temperature dependence of the spontaneous polarization, P_s, was shown in Fig. 8, for crystals grown using $BaCl_2$ as a flux. The behavior of the polarization at the two lower transitions is characterized by sudden jumps of P_s. It must be noticed that the measuring electric field is always along the cubic [100] direction; therefore the polarization shown in the figure is the component of the polarization on the cube edge. To obtain the absolute value of the polarization, it must be multiplied by $\sqrt{2}$ in the orthorhombic phase and by $\sqrt{3}$ in the rhombohedral phase. After this correction the polarization becomes more or less constant in the three different phases, although there is no reason why it should be unchanged through the transitions.

Recently, Merz [41] carried out extensive dielectric measurements on "good" single crystals grown by Remeika. It was found that the dielectric constant as a function of temperature agrees fairly well with the old measurements; the Curie constant is about 1.7×10^5 degrees, and the dielectric constant ϵ_c at room temperature is about 160. It appeared, however, that the value of the polarization is much higher, being 26 microcoulomb/cm² at room temperature, compared with the previous value of 16 microcoulomb/cm². Furthermore, the polarization shows a very steep jump at the Curie temperature to 18 microcoulomb/cm², as shown in Fig. 14. The Curie temperature of this crystal is lower than for usual crystals probably as a result of the inclusion of the small amount of oxide added for better insulation. Very recently Drougard *et al.* [72] reported almost the same value for the spontaneous polarization.

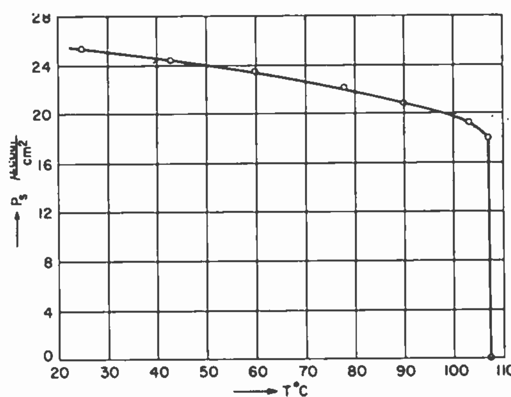


Fig. 14—Spontaneous polarization of a good $BaTiO_3$ single crystal as a function of temperature (after Merz [41]).

The higher value of P_s was attributed by Merz to the better quality of the crystal. This follows because an applied field can switch all or almost all parallel and anti-parallel domains, while the older, less perfect crystals have some anti-parallel domains which do not switch even at high field strength (see Section VII). It is remarkable how the reported measured value of the spontaneous polarization has changed, during the progress of these studies, starting with 5 to 6 microcoulombs/cm² for the ceramics.

The hysteresis loops obtained with new crystals are almost rectangular, as shown in Fig. 15 (next page). Coercive force E_c is much smaller than before: about 500 volts/cm at 60 cycles/sec. Jump of polarization at E_c is very sharp, and it shows almost perfect saturation. The flat part of the loop corresponds to a dielectric constant of between 160 and 200, while the steep part corresponds to a value of about 10^6.

It was shown by Kay and Vousden [64], and by Merz, [38], that in the tetragonal phase the spontaneous strain is proportional to the square of the spontaneous polarization:

$$z_z \sim \frac{\Delta c}{c} \sim P_s^2,$$

where Δc is estimated by extrapolating the thermal expansion curve in the cubic phase back to the ferroelectric

phase and comparing it with the actual lattice parameter at the temperature concerned. The birefringence $\Delta n = n_a - n_c$ was measured as a function of temperature, and it was established that

$$\Delta n \sim \frac{\Delta c}{c} \sim P_s^2.$$

These are results to be anticipated when the deviations from a cubic structure are small. Similar relations hold also in the orthorhombic and rhombohedral phases.

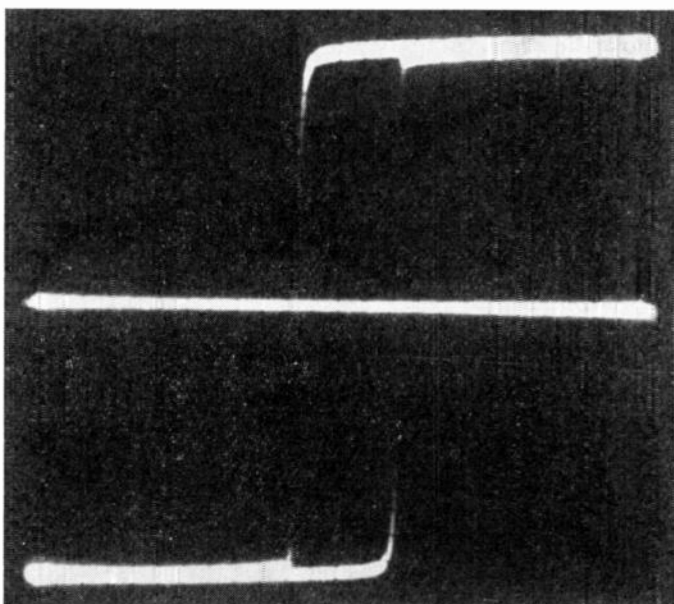


Fig. 15—Hysteresis loop of a good $BaTiO_3$ single crystal at 60 cps (after Merz [41]).

C. Thermodynamics of the Transition

Several efforts have been directed at theoretical treatments of the three $BaTiO_3$ phase transitions. Among these, the most successful is the phenomenological theory of Devonshire [4–6]. As already mentioned, the standard procedure is to expand the free energy function in terms of certain independent variables, such as polarization and stress, to use certain measured properties of the crystal to determine coefficients, and then to predict other properties. The advantage of such a procedure is that it is independent of any atomic model. It is always possible to describe the experimental results by adding as many terms to the expansion as necessary. The difficult point, however, is to explain the experimental facts with the smallest number of expansion coefficients possible, and on the basis of the most reasonable assumptions.

The appropriate expansion of the free energy function[3] depends on external conditions. Let us consider the crystal to be at a fixed temperature, and under zero stress, and let us expand the free energy in terms of increasing powers of polarization:

$$A = \tfrac{1}{2}\chi^X(P_x^2 + P_y^2 + P_z^2) + \tfrac{1}{4}\xi_{11}^X(P_x^4 + P_y^4 + P_z^4) + \tfrac{1}{2}\xi_{12}^X(P_y^2P_z^2 + P_z^2P_x^2 + P_x^2P_y^2) + \tfrac{1}{6}\zeta^X(P_x^6 + P_y^6 + P_z^6) + \cdots . \quad (2)$$

This equation and the notation used therein follow Devonshire [4]. P_x, P_y and P_z represent the components of the polarization in the x, y, and z directions. χ^X, ξ_{11}^X, ξ_{12}^X and ζ^X are expansion coefficients, all of which, strictly speaking, depend upon the temperature. The superscript X indicates that the quantity is to be taken at constant stress X; that is, the crystal is free to deform (which corresponds to the normal experimental condition). The superscript X is used in order to differentiate quantities measured at constant stress from those measured at constant strain, which latter would be designated by a superscript x.

In (2) the zero point of free energy is taken to be that of the unpolarized, unstressed crystal. As an approximation, we shall assume that only the coefficient χ^X depends upon temperature, and that all the other coefficients are temperature independent. It will be noticed that the expansion includes all terms of the second and fourth order, but only some of the sixth order. Terms in odd powers of the polarization components are omitted because we want the free energy function to be the same for reversal of the signs of any of the polarization components. Upon using the experimental values of the Curie temperature T_c, the extrapolated Curie temperature θ, the Curie constant C, and the spontaneous polarization P_s at one temperature in the tetragonal phase, Devonshire was able to account for the characteristics of the three phase transitions—including the temperature dependence of the dielectric constant, the spontaneous polarization, and the spontaneous strain, as shown in Figs. 16 and 17. He subsequently extended the theory to account for the piezoelectric and elastic constants [5].

We shall limit ourselves here to the behavior of the crystal in the cubic phase and the tetragonal phase. In the tetragonal phase, the crystal is ferroelectric, and the polarization vector lies along the c axis; thus $P_x = P_y = 0$ and if we write $P_z = P$, (2) takes the simple form:

$$A = \tfrac{1}{2}\chi^X P^2 + \tfrac{1}{4}\xi^X P^4 + \tfrac{1}{6}\zeta^X P^6, \quad (3)$$

where P represents the polarization along the c axis. From thermodynamic considerations, it can be shown readily that the applied electric field E is given by $\partial A/\partial P = E$. Thus:

$$\frac{\partial A}{\partial P} = E = \chi^X P + \xi^X P^3 + \zeta^X P^5, \quad (4)$$

where it can be seen that χ^X is the reciprocal suscepti-

[3] The free energy function used here is the *Helmholtz free energy function*, which is defined thermodynamically as $A = U - TS$, where U is the internal energy, T is the absolute temperature, and S is the entropy.

bility of the free crystal. We know from experiment that this reciprocal susceptibility is a linear function of the temperature in the nonferroelectric region (Curie-Weiss law), so that we can reasonably assume

$$\chi^X = \frac{\chi_0{}^X}{T_c - \theta}(T - \theta), \tag{5}$$

where T_c is the Curie temperature. If we set $T = T_c$, we see that $\chi_0{}^X$ is the reciprocal susceptibility at the Curie temperature. For simplicity, the coefficients ξ^X and ζ^X are still assumed to be temperature independent. Furthermore, ζ^X is assumed to be positive; otherwise infinite polarization would give the lowest value of the free energy.

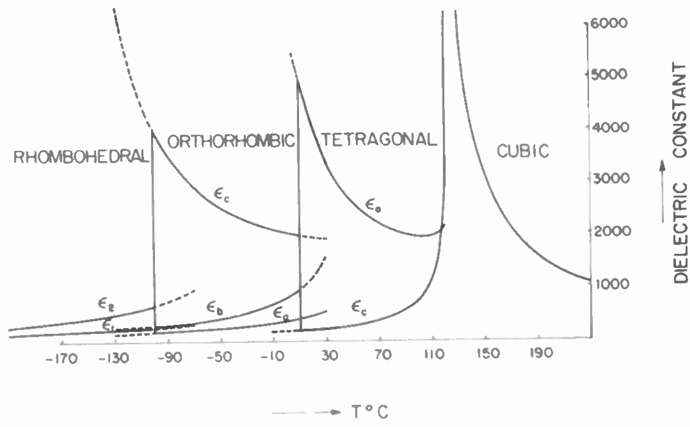


Fig. 16—Theoretical curve for the principal dielectric constant of $BaTiO_3$ (after Devonshire [4]).

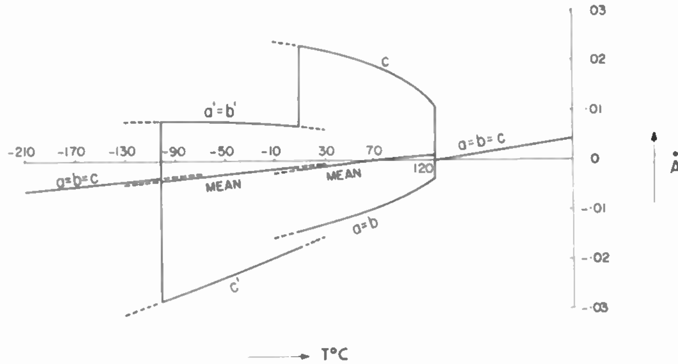


Fig. 17—Spontaneous strain of $BaTiO_3$ calculated theoretically by Devonshire [4], relative to the spacing at 120°C.

It is apparent from (4) that A has a stationary value for P equal to zero, and that this is a minimum when χ^X is positive. If χ^X and ξ^X are both positive, A will have only one minimum at $P=0$. If, as a result of its temperature dependence, χ^X becomes negative, then A will have a maximum at $P=0$ and will have a minimum at a nonzero value of P; that is, the crystal will be in a polarized state. The ferroelectric transition takes place at the point where the polarization changes from zero to a nonzero value. The order of the transition depends on the sign of the coefficient ξ^X. When ξ^X is positive, the phase change is of the second order; the polarization has no discontinuity at the Curie point, but its temperature dependence does. Putting $E=0$ in (4), we obtain, for χ^X small and negative and neglecting the small fifth power term in P:

$$P^2 = -\frac{\chi^X}{\xi^X}. \tag{6}$$

P has the value zero when χ^X is zero: that is, at the Curie temperature. This is the type of phase transition observed in Rochelle Salt and KDP. No latent heat is expected at the transition, but the specific heat shows a discontinuity.

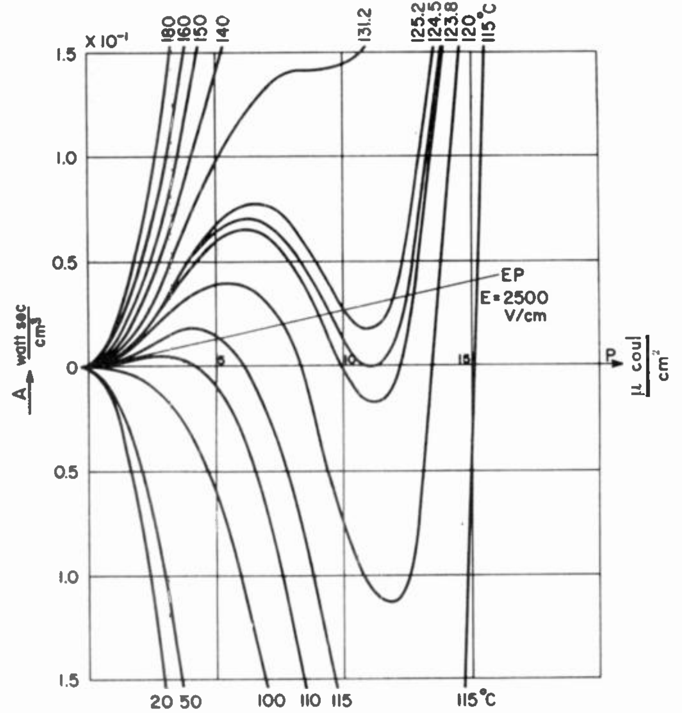


Fig. 18—Free energy curve of $BaTiO_3$ as a function of polarization according to Devonshire's theory (after Kaenzig and Maikoff [73]).

When ξ^X is negative, the phase transition is of the first order. This situation is well illustrated in Fig. 18, which represents the free energy A as a function of polarization, with temperature as the parameter [73]. It is possible for A to show two minima, one for zero polarization, the other for some nonzero value of polarization. When the two minima are equal the transition takes place and the polarization jumps discontinuously from zero to some finite value. The latter can easily be computed from the following two equations:

$$\chi^X + \xi^X P^2 + \zeta^X P^4 = 0; \tag{7}$$

$$\tfrac{1}{2}\chi^X P^2 + \tfrac{1}{4}\xi^X P^4 + \tfrac{1}{6}\zeta^X P^6 = 0. \tag{8}$$

The first equation is obtained from (4) by putting $E=0$ (for the polarization to be spontaneous, it must have a finite value even when the applied field is zero); the

second implies that the free energy A has the same value for $P=0$ and for $P=P_0$ at the transition. From these two equations we obtain the values at the transition:

$$P_0^2 = \frac{3}{4}\left(-\frac{\xi^X}{\zeta^X}\right); \quad (9)$$

$$\chi_0^X = \frac{3}{16}\frac{(\xi^X)^2}{\zeta^X}. \quad (10)$$

These relations show that at the transition temperature P has a discontinuity and χ^X is positive.

Eqs. (9) and (10) now permit determination of the coefficients ξ^X and ζ^X, since we can write

$$\xi^X = -4\frac{\chi_0^X}{P_0^2}, \quad (11)$$

and

$$\zeta^X = 3\frac{\chi_0^X}{P_0^4}. \quad (12)$$

On the right-hand side we find quantities that can be measured experimentally: χ_0^X, and P_0 (the value of the spontaneous polarization at the Curie point). The susceptibility $k^X = 1/\chi^X$ is related to the dielectric constant by

$$k^X = \frac{\epsilon^X - 1}{4\pi} \sim \frac{\epsilon^X}{4\pi}, \quad (13)$$

since ϵ^X has a high value at the Curie point. ϵ^X obeys the Curie-Weiss law (1); and hence it follows from (5) that

$$\chi_0^X \approx \frac{4\pi}{C}(T_c - \theta), \quad (14)$$

where C is the Curie constant and $T_c - \theta = 11°C$. Upon introducing experimental values on the right side of (11) and (12), Merz [41] obtained:

$$\xi^X = -10.8 \times 10^{-13} \text{ cgs units;}$$

$$\zeta^X = 28.2 \times 10^{-23} \text{ cgs units.}$$

Direct experimental determination of the coefficients, made via a double loop experiment described below, yields [41]

$$\xi^X = -6.8 \times 10^{-13} \text{ cgs units;}$$

$$\zeta^X = 22.8 \times 10^{-23} \text{ cgs units.}$$

The agreement is fairly good, but it has even been improved recently by the investigation of Drougard, Landauer, and Young [74]. Taking the derivative of (4), we can write for the dielectric constant

$$\frac{4\pi}{\epsilon} \approx \frac{\partial E}{\partial P} = \chi^X + 3\xi^X P^2 + 5\zeta^X P^4. \quad (15)$$

The latter authors measured the coefficient ξ^X by studying the polarizaton dependence of the capacitance of a sample. The polarization is developed by a low-frequency (60 cps) field of variable amplitude, and a very small amplitude radio-frequency field is superposed on the 60 cps field to permit measurement of the capacitance. Measurements are taken at different temperatures to show the temperature dependence of the coefficients under study. The experimental result, at 120°C, is:

$$\xi^X = -10.0 \times 10^{-13} \text{ cgs units.}$$

This is in excellent agreement with the value of -10.8×10^{-13} cgs units, achieved from Devonshire's theory. The same authors find, contrary to our original assumption, that the coefficient ξ^X is strongly temperature-dependent. The dependence is linear from the Curie point to 150°C; and these investigators point out that if the linearity holds at higher temperatures, the coefficient ξ^X should change sign in the vicinity of 175°C.

A knowledge of the coefficients of the free energy expansion permits the description of various properties of the crystal below the transition point. The spontaneous polarization P_s is obtained by solving (7); the dielectric constant can be computed from (15).

Another important relation can be derived for the spontaneous polarization P_s in terms of the Curie constant C. The entropy S, related to the polarization, is given by the relation

$$S = -\left(\frac{\partial A}{\partial T}\right)_{P,X}. \quad (16)$$

The change in entropy at the transition is then given by

$$\Delta S = S - S_0 \cong -\frac{1}{2}P^2 a\frac{\chi^X}{\partial T} = -\frac{1}{2}P_s^2\left(\frac{4\pi}{C}\right). \quad (17)$$

Here S_0 is the entropy for zero polarization. This relation holds for first- and second-order transitions. In the case of a first-order transition, $P_s = P_0$, and $\Delta S = L/T_c$, where L is the latent heat. In the case of a second-order transition, ΔS is obtained by integration of the anomalous specific heat vs temperature curve between the temperatures corresponding to $P=P_s$ and $P=0$. The importance of relation (17) was first emphasized by Jaynes [75]. Since all the terms in (17) are experimentally known, it is possible to test its validity. Observed and calculated entropy changes are in fact in satisfactory agreement, as can be seen from Table I (see also Devonshire [6], p. 108).

In yielding a negative value for ξ^X, the theory of Devonshire indicates that the 120°C transition of $BaTiO_3$ is of the first order. The experimental proof of this fact is not easy. A first-order phase change implies, indeed, a discontinuity in the spontaneous polarization and a discontinuous change in the lattice parameters, together with the existence of a latent heat at the transi-

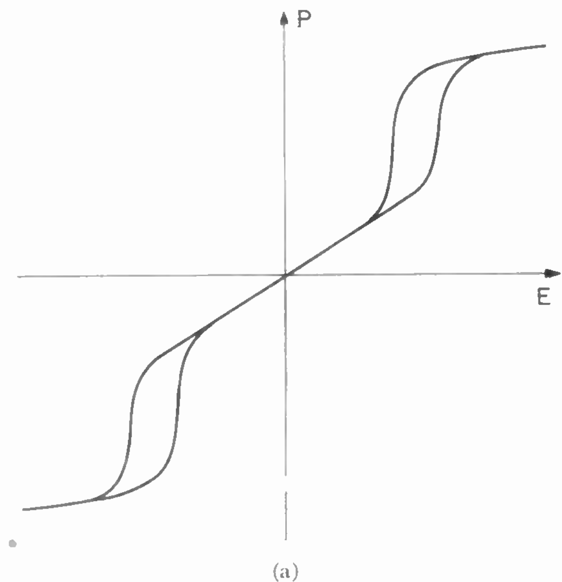
TABLE I

ENTROPIES OF POLARIZATION

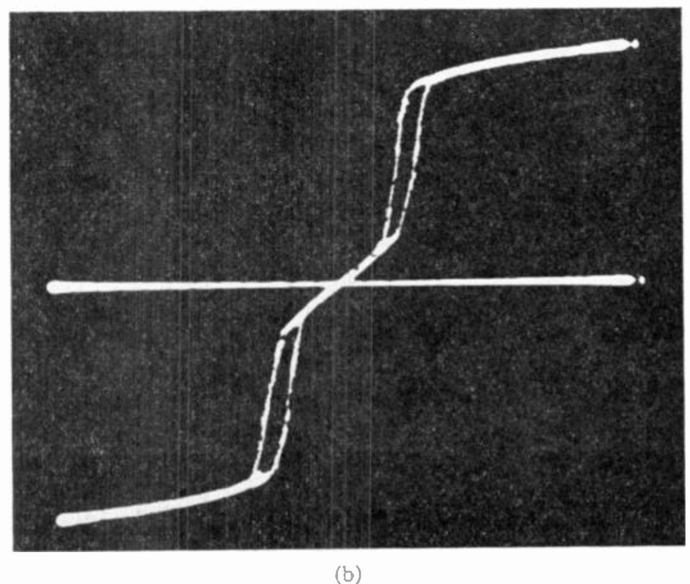
Substance	KH_2PO_4	$BaTiO_3$	$KNbO_3$	$Cd_2Nb_2O_7$
Spontaneous polarization in the vicinity of the Curie temperature (microcoulomb/cm²)	5[13]	20[41]	26[92]	5[35]
Curie Constant C (°C)	3,260[149]	170,000[41]	241,000[93]	100,000[35]
ΔS calculated from (17) (cal/mole degree)	0.60	0.12	0.15	0.02
ΔS measured (cal/mole degree)	0.47[149] 0.72[150]	0.12[151]	0.19[92] 0.28[47]	0.01[152]

tion temperature; but it does not place any limitation on the smallness of the discontinuities. In practice it is therefore very difficult to distinguish between really discontinuous changes and very sharp but continuous changes. This is especially true when internal strains, generally present in crystals, render the changes sluggish. In early stages of the research on $BaTiO_3$, it was generally believed that the transition at 120°C was of the second order. It was the dielectric investigation of Kaenzig and Maikoff [73] which first supported Devonshire's assertion of a first-order transition. Recently, with the availability of good single crystals, other dielectric measurements have supported the discontinuity of the spontaneous polarization [41] and that of the dielectric constant [72] at the transition point.

Decisive evidence for the first-order phase change is given by the so-called "induced transition": that is, upon inducing a transition by means of an electric field. This was first accomplished by Roberts [76] who applied a strong dc field to a ceramic specimen just above the Curie temperature. A sudden temperature increase of about 0.5°C was observed in the specimen. This is attributed to the latent heat released in a sudden transition from a weakly polarized state to a strongly polarized state. We can think of this as a transition from a nonferroelectric phase to a ferroelectric phase induced by the field. From the relation $L=\Delta T\cdot C_p$, taking $C_p=30$ cal/mol deg, the latent heat of the transition was estimated to be about 15 cal/mol. More recently the induced transition was demonstrated directly on the screen of a cathode ray tube by Merz [41] on very good single crystals. The circuit used for this experiment is the Sawyer-Tower arrangement commonly used to detect hysteresis loops [2]. A few degrees above the Curie point a strong ac field is applied to the crystal. At zero field the crystal is not ferroelectric, since its temperature is above the Curie point. Upon increasing the field strength [see Fig. 19(a)], the polarization increases linearly, as one would expect in an ordinary dielectric material. At a critical field strength a sudden jump in polarization occurs. Upon reducing the field, the polarization does not retrace its original path in the neighborhood of the jump, but traces out a hysteresis loop suggestive of the usual type of ferroelectric hysteresis loop. Upon applying a 60-cycle field, the phenomenon occurs 120 times a second. On the screen of the cathode



(a)



(b)

Fig. 19—(a) Double hysteresis loop of $BaTiO_3$ observed just above the Curie point (schematic). The induced transition is shown by the sudden jump in polarization. (b) Oscilloscope picture of a double hysteresis loop of $BaTiO_3$ at 111.4°C (after Merz [41]).

ray tube one observes a so-called "double hysteresis loop" [cf. Fig. 19(b)]. The sudden jump of the polarization at the critical field is clearly visible in the figure. Such double hysteresis loops can only be observed if the phase change is of the first order. This is understandable on the basis of the diagram represented in Fig. 18. In the presence of an applied field E, the induced phase transition occurs when $A-EP$ (rather than A) has the same value for two different values of P. This can only happen if ξ^x is negative, for only then does one get the type of curves shown in Fig. 18. Clearly, we can estimate the coefficients of the Devonshire expansion of the free energy by comparing these experimental P vs E curves with those obtainable from (4), which are represented in Fig. 20. It was in this way that Merz determined the values of ξ^x and ζ^x reported above, and confirmed Devonshire's claim for a first-order phase change.

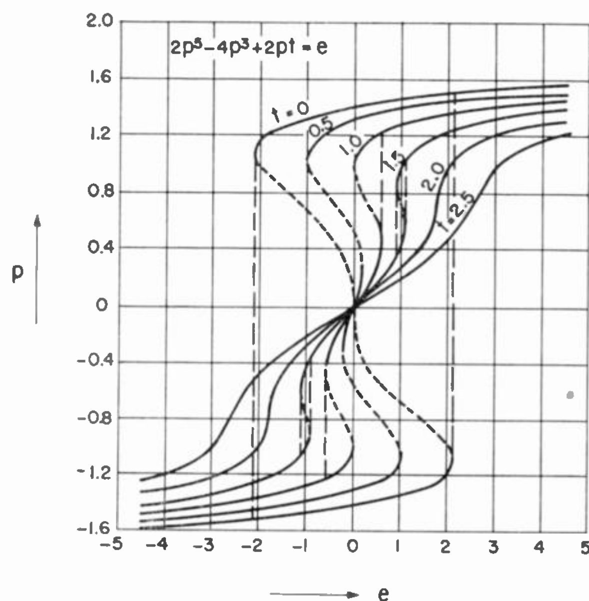


Fig. 20—Polarization as a function of electric field according to Devonshire's theory. p, e and t are linearly related with polarization, field and temperature, respectively (after Merz [41]).

Let us now study the phenomenon of the induced phase transition in more detail. It is generally considered (cf. Merz [41]) that the central part of the double loop corresponds to a paraelectric phase, while both ends of the loop correspond to the ferroelectric phase. It must be stressed, however, that the central part of the loop does not correspond to a cubic phase, but rather to a tetragonal one; because whenever we apply a field to the cubic phase along a [100] direction, it becomes tetragonal. Consequently the induced phase change represents a transition between two tetragonal phases; and in this way it is different from the normal phase change from cubic to tetragonal, observed at the Curie point upon descreasing the temperature with no field applied. The interesting question arising then is the following: can we induce a ferroelectric phase at any temperature above the Curie point simply by applying a strong enough field? The answer to this question is provided by a study of the P vs E plot of Fig. 20. We see from this figure that the transition can be field-induced only in a narrow temperature range (up to about 10°C above the Curie temperature). The parameter t of the curves plotted in Fig. 20 is a measure of the temperature according to the relation

$$t = 4\chi^x \frac{\zeta^x}{(\xi^x)^2}; \tag{18}$$

and it can be seen that when the temperature T becomes higher than a value T_0 which corresponds to $t=9/5$, then no induced transition is possible. The polarization changes continuously with increasing field, no matter how large the field strength applied.

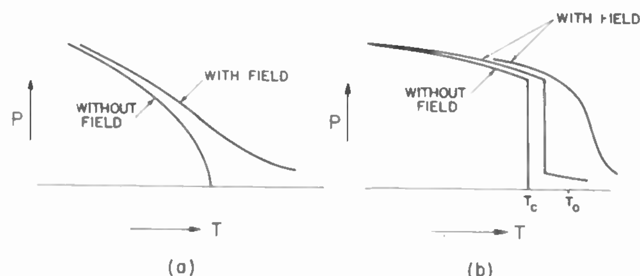


Fig. 21—Polarization as a function of temperature with and without a biasing field: (a) near a second-order transition, (b) near a first order transition (after Devonshire [6]).

We can also consider the problem from another viewpoint. Suppose that we apply a dc field to the crystal, and then vary the temperature in order to see when the transition takes place. Fig. 21 shows how the polarization changes as a function of temperature, with and without a biasing field. In the case of a first-order transition we see that for fields which are not too large a jump in the polarization can still be observed. However, we notice that here again the polarization is not zero at temperatures immediately higher than that corresponding to the jump. When the applied field is increased beyond some critical value, we no longer observe any discontinuity in the polarization or of its temperature derivative. In this case, the effect of the field is to blur the transition: if we measure the dielectric constant with a small ac field superposed upon a large dc biasing field, as a function of temperature, we notice a maximum of ϵ at some temperature. This maximum, however, does not correspond to a phase transition in the sense that we speak of the Curie temperature at 120°C. When the phase change is of the second order we see from Fig. 21(a) that no definite transition point occurs when a dc biasing field is applied.

It is seen from Fig. 21(b) that, as the biasing field is increased, the transition temperature increases also. The shift of the transition temperature with the field was measured by Kaenzig and Maikoff [73], upon super-

posing a small ac field on the biasing dc field. Their result is expressed in the equation

$$\frac{\Delta T_c}{\Delta E} = 1.2 \times 10^{-3} \text{ degree/volt cm}^{-1}. \quad (19)$$

The same quantity was estimated by Merz [41] by means of the double-loop experiments. Merz plotted the value of the critical field at which the induced transition occurs as a function of the temperature. This curve also represents a plot of the transition temperature vs the biasing field. The slope of this curve yields

$$\frac{\Delta T_c}{\Delta E} = 1.4 \times 10^{-3} \text{ degree/volt cm}^{-1}, \quad (20)$$

which is in good agreement with Kaenzig and Maikoff's direct measurement.

Let us now see what the theory has to say about this field dependence of the transition point. First of all, from thermodynamic considerations we can establish the following:

$$\frac{\partial T_c}{\partial E} = -\frac{\Delta P}{\Delta S}, \quad (21)$$

where ΔP represents the discontinuous jump of the polarization at the Curie point upon heating, which is equal to $-P_0$. Eq. (21) is similar to the well-known relation of Clausius-Clapeyron. Upon using $\Delta S = 0.12$ cal/mole degree [151] and $P_0 = 18$ μcoul/cm^2 [41], we obtain

$$\frac{\partial T_c}{\partial E} = 1.4 \times 10^{-3} \text{ degree/volt cm}^{-1},$$

in excellent agreement with experimental results. On the other hand, we can estimate the same quantity on the basis of the Devonshire theory. Starting from (21), and using the relations (9), (10), and (17), we obtain:

$$\frac{\partial T_c}{\partial E} = \frac{C}{4\pi}\left(-\frac{4}{3}\frac{\zeta^X}{\xi^X}\right)^{1/2}. \quad (22)$$

Using the value of ξ^X measured by Merz, and the value of ζ^X measured by Drougard *et al.* [74], we obtain

$$\frac{\partial T_c}{\partial E} = 1.6 \times 10^{-3} \text{ degree/volt cm}^{-1},$$

which is again in good agreement with the experimental results, and thus provides more support for the Devonshire phenomenological theory.

The behavior of the dielectric constant (measured with a small ac field above the Curie temperature) as a function of the dc biasing field, can be predicted on the basis of Fig. 19 and Fig. 20, since ϵ is given by the slope of the curve in Fig. 20. Upon increasing the biasing field the dielectric constant ϵ should increase, slightly at first, then strongly as the polarization jumps when the critical field is reached; and ϵ should finally drop rapidly to some constant value. In contrast with this prediction the study of Kaenzig and Maikoff revealed that the dielectric constant decreased under these conditions. The discrepancy is attributed by the authors to the minute cracks produced in the crystals by heating them through the transition. A re-examination of the phenomenon with good single crystals would be desirable.

There is another dependence of dielectric properties which the theory can predict and experiment can check: effects of various mechanical pressures. Merz [77] studied the influence of hydrostatic pressure on the Curie temperature, and found a linear decrease of the latter at the rate of -5.7×10^{-9} degree/dyne cm^{-2}. Theoretically the shift of Curie temperature with hydrostatic pressure can be computed from the Clausius-Clapeyron equation:

$$\frac{\partial T_c}{\partial p} = \frac{\Delta V}{\Delta S}, \quad (23)$$

where p is the pressure and ΔV represents the anomaly of the volume at the transition upon heating. Using $\Delta V = -0.06$ A^3, we obtain

$$\frac{\partial T_c}{\partial p} = -6.7 \times 10^{-9} \text{ degree/dyne cm}^{-2},$$

which, within the accuracy to be expected, is in agreement with Merz' result.

We can derive an expression for the same quantity from the theory of Devonshire [6]. This expression can be evaluated only if one knows the values of the electrostrictive constants, and also the dependence of the higher order terms on temperature and pressure. Neglecting this dependence, and using the available data for the electrostrictive constants, we obtain

$$\frac{\partial T_c}{\partial p} = -9 \times 10^{-9} \text{ degree/dyne cm}^{-2},$$

a result which is satisfactory if one considers the relatively low accuracy of the electrostrictive measurements.

The only discrepancy between theory and experiment which we find is in the effect of a two-dimensional pressure perpendicular to the c axis. This effect was studied by Forsbergh [78], who found the Curie temperature increased with the square of the applied pressure. On the other hand, both the Clausius-Clapeyron relation and Devonshire's theory predict a linear increase of the Curie temperature with pressure. The reason for this discrepancy has not yet been found.

Before we conclude this section, let us emphasize once more that the treatment reported above considers only the phase change of a free crystal, according to the normal experimental conditions. However, when a theoretical model is set up, it is much more convenient to assume a situation in which the strains are kept constant within the crystal. The coefficients of the free energy expansion in this latter case represent parame-

ters at constant strain x, *e.g.* χ^x, ξ^x, ζ^x. The thermodynamic treatment of this case has been developed by Devonshire [5]. The most interesting result of this theory applied to $BaTiO_3$ is that the coefficient ξ^X was found to be positive, although ξ^X was found to be negative. This means that the transition of the clamped crystal will be of the second order, although that of the free crystal is of the first order, an evidence that the first-order transition of $BaTiO_3$ must be due to the interaction between the polarization and strain. This fact is very important when an atomic model is considered for $BaTiO_3$.

Concluding, we can say that the phenomenological treatment of Devonshire is able to explain satisfactorily the different properties of $BaTiO_3$ with a comparatively small number of expansion coefficients. It must be kept in mind, however, that this type of theory by its nature is an approximation, not only in the sense that we are using a finite number of terms of the free energy, but also in that we assume the validity of the same free energy expansion for both the cubic and the tetragonal phase. This second assumption can be justified *a priori* by the fact that the ferroelectric phase is generally obtained by a small change in the parameters of the nonpolar phase, and *a posteriori* by the fact that the theory based on this assumption is in satisfactory agreement with the experimental evidence.

Various theories have been developed on the basis of different atomic models. We will discuss these after reporting on the structure of $BaTiO_3$.

V. Ferroelectricity and Antiferroelectricity in Perovskite-Type Compounds

A. Solid Solutions Containing $BaTiO_3$

One approach to an understanding of ferroelectricity in $BaTiO_3$-type crystals is to study the influence of individual ions on the dielectric properties.

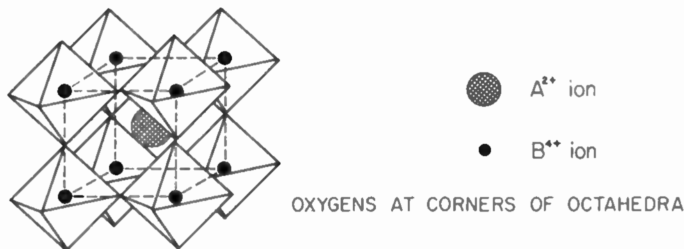


Fig. 22—Perovskite structure ABO_3 considered as a three dimensional framework of BO_6 octahedra (after Megaw [156]).

A first step in this direction is to consider the packing of the ions in the structure. The structure consists of a three-dimensional framework of TiO_6 octahedra sharing corners, as shown in Fig. 22, with Ba ions occupying spaces between the octahedra. The packing can be characterized by a *tolerance factor*, t, defined as

$$R_A + R_O = t \cdot \sqrt{2} \cdot (R_B + R_O),$$

where R_A, R_B, and R_O are the ionic radii of the A, B, and O ions respectively. When $t=1$, the packing is perfect: that is, the ions are just in contact with one another. Using Goldschmidt ionic radii, t assumes the value of 1.02 for $BaTiO_3$; and it lies between 1.0 and 0.8 for other perovskite compounds (see Table II). A tolerance factor greater than one means that there is too large a space available for Ti; therefore this ion can "rattle" inside of its octahedron. It was early considered that this too-large space for Ti, aided by Ti-O-Ti-O interaction along the c direction, was the main cause of ferroelectric polarization. The Ba ions were considered as spacers which aided in providing the extra room for Ti.

TABLE II

Values of Tolerance Factor t for Various Perovskite-Type Compounds ABO_3 (after Megaw [156])

		B Ion→	Ti^{4+}	Sn^{4+}	Zr^{4+}	Th^{4+}
		$R_B(A)$ $CN=6$	0.64	0.74	0.77	1.10
A Ion↓	$R_A(A)$ $CN=6$	$R_A(A)$ $CN=12$	Tolerance Factors t			
Ca^{2+}	1.06	1.16	0.89	0.85	0.84	(0.72)
Sr^{2+}	1.27	1.37	0.97	0.92	0.91	(0.78)
Ba^{2+}	1.43	1.52	1.02	0.97	0.96	0.83
Pb^{2+}	1.32	1.40	0.98	(0.92)	0.93	(0.79)
Cd^{2+}	1.03	1.11	0.88	(0.83)	(0.82)	(0.71)

R_A = Ionic radius of A ion, in Ångstrom Units (A).
CN = Coordination number.
Parentheses indicate compound not known as a perovskite.

One way to test this idea is to replace Ba or Ti by other ions. The perovskite-type structure can be realized by the combination of various ions with a wide range of ionic radii. Ba can be replaced by Sr, Ca or Pb; and Ti by Zr or Sn. The first solid solution system tried was $BaTiO_3$-$SrTiO_3$ [163] and it was found that the Curie temperature decreased linearly with increasing concentration of Sr. The lower two transition temperatures also decreased. These results seemed to support the idea of the "rattling Ti," since Sr is smaller than Ba, and consequently it provided smaller space for Ti. Effects of replacement of Ti in $BaTiO_3$ by Zr and Sn were also examined. Again the Curie temperature decreases with increasing concentration of Zr or Sn. Here the ionic radius of Zr or Sn is larger than that of Ti; therefore, the replacement decreases the effective volume for B ions.

The simple picture suggested by these results was upset by a study of the $BaTiO_3$-$PbTiO_3$ system [158]. The phase diagram in Fig. 23 shows that the Curie temperature *increases* with increasing concentration of Pb—even though the Pb ion is also smaller than Ba, as is Sr. It is obvious then that the role of the A ions in the structure is not expressed only through their ionic radii. Other factors, such as polarizability and the character of the bonding, must be taken into account. In this respect, it must be noticed that Sr belongs to the alkali earth metals, as does Ba, but Pb is situated in a dif-

ferent group in the periodic table, and has a different type of electronic configuration. Another important result in the (Ba, Pb)TiO_3 system is that the lower two transition temperatures decrease, while the upper one increases, with increasing Pb concentration.

The phase transition of pure $PbTiO_3$ at 490°C was studied by Shirane and Hoshino [79]. It was shown that the transition is quite similar to that of $BaTiO_3$ at 120°C, but all the anomalies at the transition are much more drastic in $PbTiO_3$. The axial ratio c/a is 1.02 just below the Curie temperature, and it reaches 1.06 at room temperature. Single crystals of $PbTiO_3$ were grown by Rogers [80], by using excess PbO as a flux.

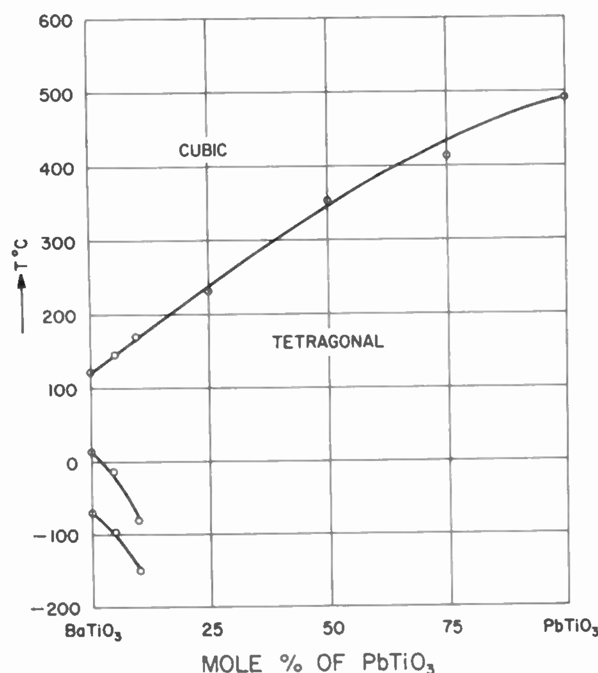


Fig. 23—Phase diagram of the $BaTiO_3$-$PbTiO_3$ system (after Shirane and Suzuki [158]).

Crystals more than a few mm in size always crack when they pass through the Curie temperature on cooling, because of a large spontaneous strain associated with the transition.

Why the ferroelectric properties of $PbTiO_3$ are much more exaggerated than those of $BaTiO_3$, and how this is related to the role of Pb ions in the lattice, is a very interesting problem. We will come back to this point after we discuss the crystal structure of $PbTiO_3$ in Section VI.

B. $PbZrO_3$ and Related Compounds

The dielectric properties and phase transition in $PbZrO_3$ were first studied by Roberts [81]. This compound was reported to have a tetragonal lattice of the perovskite type at room temperature, as does $BaTiO_3$. The dielectric constant of ceramic $PbZrO_3$ is about 100 at room temperature and it shows a sharp peak at the transition temperature of 230°C (Fig. 24). Above 230°C the Curie-Weiss law is followed, with $C = 1.6 \cdot 10^{5}$°C, and $\theta = 193$°C.

However, extensive dielectric studies by Sawaguchi, Shirane, and Takagi [48, 49] indicated that $PbZrO_3$ might not be ferroelectric, notwithstanding the close resemblance in its dielectric constant *vs* temperature curve with that of $BaTiO_3$. No ferroelectric hysteresis-loops were observed, even with a large electric field and even at a temperature just below the Curie point—where the coercive field, if the material is ferroelectric, would be expected to be small. Moreover, an application of a dc field decreases the Curie temperature, in contrast to the case of $BaTiO_3$. It was suggested that $PbZrO_3$ may be an antiferroelectric, with dipoles arranged in an antiparallel array so as to give no net polarization. This was proved by a structural study by Sawaguchi, Maniwa, and Hoshino [82], using single crystals.

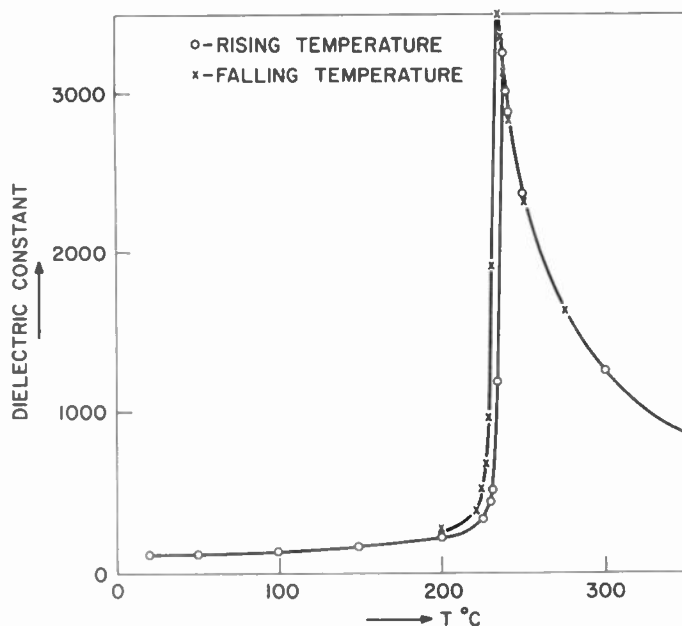


Fig. 24—Dielectric constant of $PbZrO_3$ ceramics as a function of temperature (after Roberts [81]).

The room temperature structure of $PbZrO_3$ was first reported from powder photographs to be tetragonal, with lattice parameters [156].

$$a_t = 4.16A, \qquad c_t/a_t = 0.99.$$

Attention is to be paid to this axial ratio c/a, which is smaller than unity, in contrast with the 1.01 value for $BaTiO_3$. The powder pattern contains some extra lines, which can only be explained by assuming a multiple unit cell. Later, an X-ray and optical study [82] of small $PbZrO_3$ single crystals revealed that the true symmetry at room temperature is orthorhombic, with

$$a = 5.88A = \sqrt{2}\, a_t, \qquad b = 2a, \qquad c = 8.20A = 2c_t,$$

where a_t and c_t are the pseudo-tetragonal parameters. One of the original cubic axes becomes an orthorhombic axis, and the other two axes lie at 45° to the cubic axis, as in the case of orthorhombic $BaTiO_3$ (see Fig. 12).

Then the true unit cell of $PbZrO_3$ contains 8 "molecules." It is a notable coincidence that we have exactly $b=2a$, which could be accounted for by pseudo-tetragonal parameters.

The X-ray study revealed anti-parallel shifts of Pb ions along the a axis, as shown in Fig. 25. This antiferroelectric configuration of Pb shifts seems to be a reasonable one from the viewpoint of dipolar interaction. Moreover it is now understandable why c_t/a_t is smaller than unity in $PbZrO_3$, if we recall that the spontaneous strain perpendicular to the polarization is always a contraction in $BaTiO_3$.

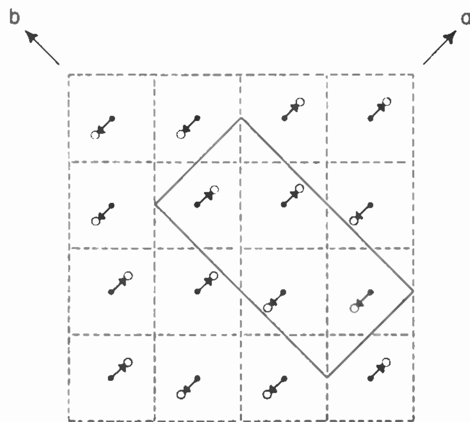


Fig. 25—Antiferroelectric structure of $PbZrO_3$. Arrows represent the directions of shift of the Pb ions. The solid line shows the orthorhombic unit cell (after Sawaguchi *et al.* [82]).

Single crystals of $PbZrO_3$ were grown by Jona *et al.* [50], using PbF_2 as a flux. Although the crystals obtained were too small for single-crystal dielectric measurements, they were adequate for an optical study. It was shown that the differences between the three principal refractive indexes are

$$n_b - n_a = 0.039, \quad \text{and} \quad n_c - n_a = 0.024.$$

Thus the refractive index is smallest along the a axis, along which the anti-parallel shifts of Pb ions were reported. (We should recall that the refractive index in the tetragonal and orthorhombic $BaTiO_3$ is smallest along the direction of the spontaneous polarization.) It must be noticed that the optical anisotropy is largest in the ab plane. Although dielectric measurements of an untwinned crystal are not available as yet, we may anticipate similar relations between the three principal dielectric constants. The structure of $PbZrO_3$ shown in Fig. 25 could also explain the large optical anisotropy in the ab plane. This structure, however, does not suggest why the lattice parameters are exactly $b=2a$. It might be that some other ions (such as oxygens) shift simultaneously in the b direction so as to give a large optical and dielectric anisotropy within the ab plane, retaining $b=2a$. Necessary more detailed structural study is in progress, with X rays and neutrons, by Jona *et al.* [83].

The properties of the $PbZrO_3$-$PbTiO_3$ system were studied, to determine the results of mixing these two different types of dielectrics. The phase diagram ob-

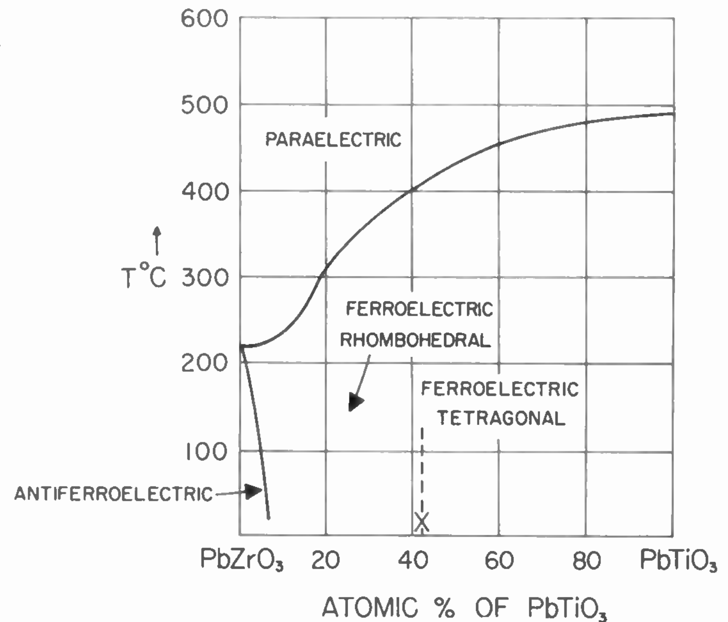


Fig. 26—Phase diagram of the $PbZrO_3$-$PbTiO_3$ system (after Shirane *et al.* [159], and Sawaguchi [56]).

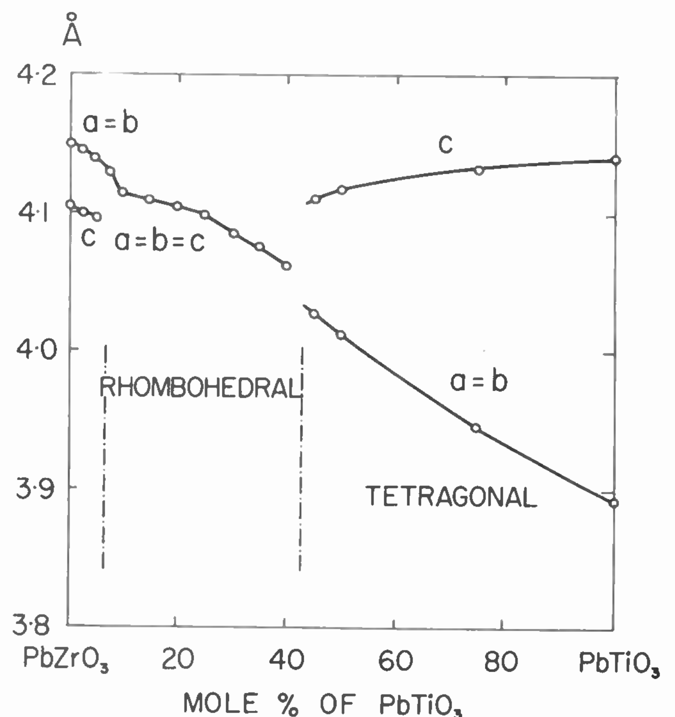


Fig. 27—Lattice spacing of $PbZrO_3$-$PbTiO_3$ at room temperature as a function of composition (after Shirane and Suzuki [84]).

tained [56, 84, 159] is shown in Fig. 26. The lattice spacings of solid solutions between $PbZrO_3$ and $PbTiO_3$ [84] are shown as a function of composition, at room temperature, in Fig. 27. $PbZrO_3$-$BaZrO_3$ system was also investigated [54] with results shown in Fig. 28 (facing page). One important revelation is that antiferroelectric $PbZrO_3$ phase is limited to a narrow composition range near pure $PbZrO_3$ side. When small amounts of $PbTiO_3$ or $BaZrO_3$ are added, a ferroelectric phase can be observed between the antiferroelectric and paraelectric phases. An X-ray study [55] revealed that this ferroelectric phase has a rhombohedral structure, which

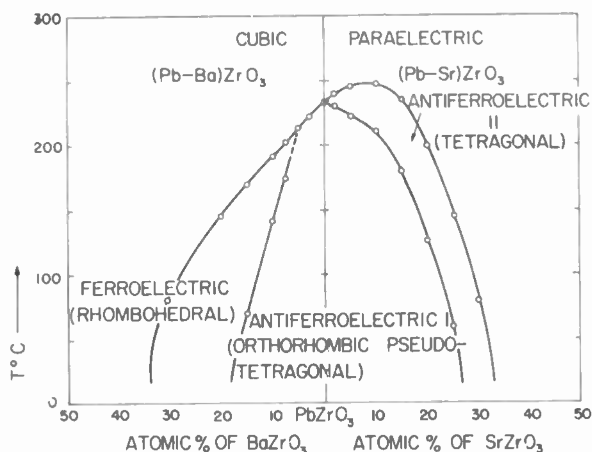


Fig. 28—Phase diagrams of (Pb, Ba)ZrO_3 and (Pb, Sr)ZrO_3 (after Shirane [54]).

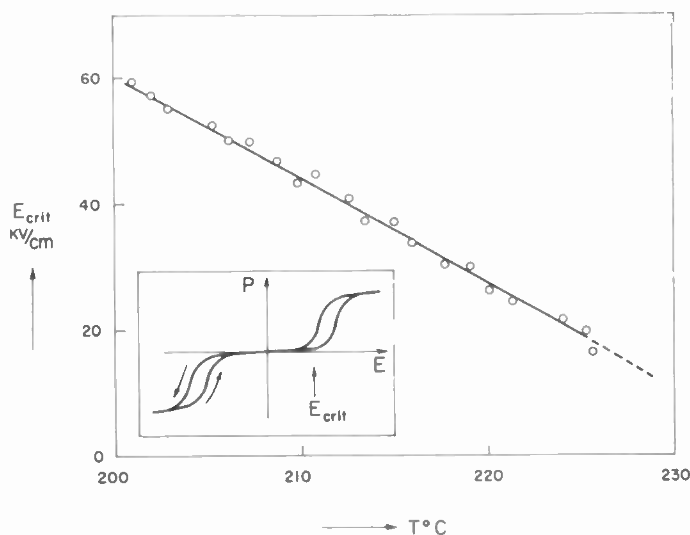


Fig. 29—Threshold field strength of $PbZrO_3$ as a function of temperature (after Sawaguchi and Kittaka [85]).

is similar to that of the lowest phase of $BaTiO_3$ (below $-80°C$). Thus the antiferroelectric phase in $PbZrO_3$ is very critical, in the sense that its free energy is very close to that of the ferroelectric phase, especially near the Curie temperature.

This point was clearly demonstrated by the "induced transition" in $PbZrO_3$ [49]. When a strong field was applied to a $PbZrO_3$ ceramic at temperatures just below the Curie point, a double hysteresis loop was observed [Fig. 19(a)]. The inner part of the loop is antiferroelectric and the outer part ferroelectric, this state having been induced by the electric field. It is quite natural that the free energy of the ferroelectric phase is lowered compared with that of the antiferroelectric phase. The crystal structure of this induced phase was studied by X rays, while a strong dc field was applied to keep it in the ferroelectric state. It was shown that this induced ferroelectric phase is also a rhombohedral lattice with $\alpha < 90°$. This double loop is quite similar in shape to those subsequently observed by Merz [41] in $BaTiO_3$ just above the Curie temperature [Fig. 19(b)].

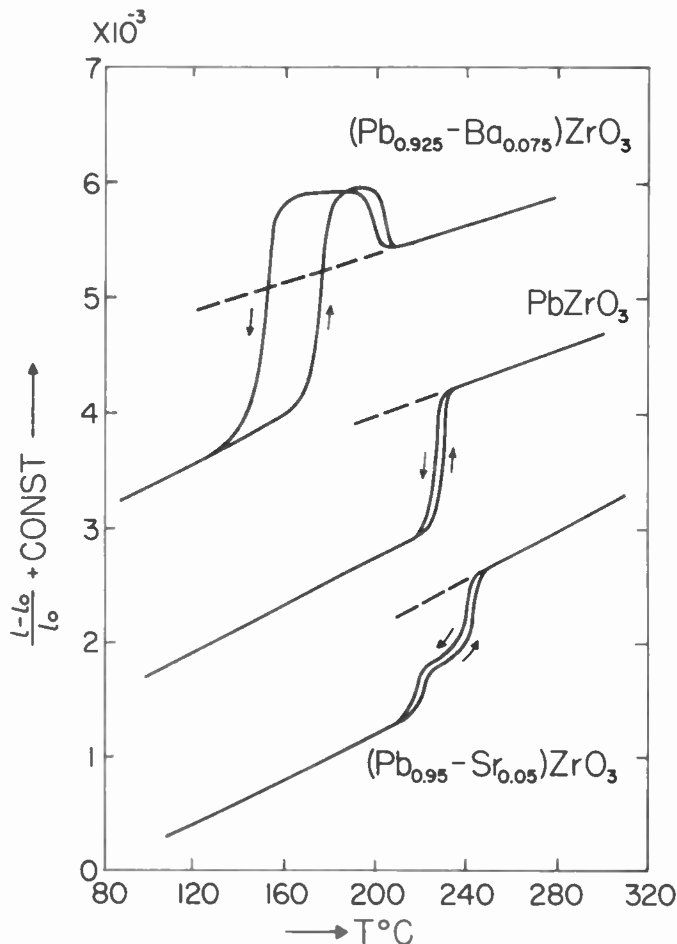


Fig. 30—Linear thermal expansion of $PbZrO_3$, $(Pb_{0.925}, Ba_{0.075})ZrO_3$ and $(Pb_{0.95}, Sr_{0.05})ZrO_3$ (after Shirane [54]).

While the induced transition is from antiferroelectric to ferroelectric in $PbZrO_3$, it is from paraelectric to ferroelectric in $BaTiO_3$. Fig. 29 shows the temperature dependence of the threshold field E_c at which the ferroelectric phase is induced [85]. It was estimated that the antiferroelectric phase of pure $PbZrO_3$ has a free energy which is lower by about 4 cal/mole at 230°C compared with that of the ferroelectric phase [56].

The solid solution system (Pb, Sr)ZrO_3 was also studied [54], with the result shown in Fig. 28. An intermediate phase was again observed; however, this phase is not ferroelectric, but instead is another antiferroelectric phase with a tetragonal structure. The same antiferroelectric phase was observed by Sawaguchi [56] in Pb(Zr, Ti)O_3 system, when the Ti concentration is very small. The relative stability of these two antiferroelectric phases, the ferroelectric phase and the paraelectric phase was discussed by Sawaguchi [56] in detail, as a function of temperature and composition.

The phase transitions of $PbHfO_3$ were studied by Shirane and Pepinsky [53], and similar antiferroelectric phases to those of (Pb, Sr)ZrO_3 composition were observed with transition temperatures of 160° and 215°C.

These various phases show quite different properties, not only dielectrically and structurally but also thermally. Fig. 30 shows the linear thermal expansion curves

for three typical compounds. If we compare the volume of each phase with the normal volume of the cubic phase, it is clear that the two antiferroelectric phases show a volume decrease, while the rhombohedral ferroelectric phase shows a volume increase. A similar volume increase was observed at the Curie points of $BaTiO_3$, $PbTiO_3$, and $KNbO_3$, although the phase changes in these crystals are from cubic to tetragonal.

A theory of antiferroelectricity was first developed by Kittel [86], before any crystal was identified as an antiferroelectric. A model treated by Kittel is a simple dipole array as shown in Fig. 31; and it was shown that antiferroelectricity may be expected to occur. His theory gives some criteria for the identification of antiferroelectric crystals. Takagi [87] then developed a theory involving a model which has rotatable dipoles at the cube corners and polarizable ions at the body centers. The theory was then extended to a general case in which the rotatable dipoles were replaced by induced dipoles [87a].

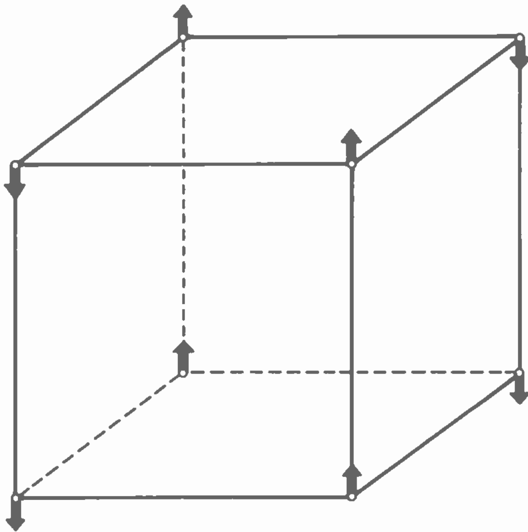


Fig. 31—An anti-ferroelectric array.

Kittel has shown that the peak value of the dielectric constant at the antiferroelectric Curie temperature is generally not pronounced for both first- and second-order phase changes. In addition to this general case, however, Takagi's theory emphasized that the dielectric constant can in some special case show a very sharp maximum at the Curie point. In such special cases a ferroelectric state instead of an antiferroelectric one can be realized by a small change of the polarizabilities of the constituent ions. This theoretical prediction is in good accordance with the observation of the sharp change of the dielectric constant in pure $PbZrO_3$, and the appearance of the ferroelectric phase in the $Pb(Zr, Ti)O_3$ or in $(Pb, Ba)ZrO_3$ systems. It was shown experimentally that the peak value of the dielectric constant at the transition in the $(Pb, Sr)ZrO_3$ system decreases rapidly with increasing concentration of Sr; this is in contrast to the case of the $Pb(Zr, Ti)O_3$ and $(Pb, Ba)ZrO_3$ systems, in which pronounced peaks were always observed at the ferroelectric Curie point. We can consider that $(Pb, Sr)ZrO_3$ compositions show the general characteristics of a small change in the dielectric constant at the antiferroelectric Curie temperature. Thus these theories for simple models could account for some important properties of antiferroelectrics. The structure of $PbZrO_3$ is much more complicated than are the models, unfortunately.

C. Niobates and Tantalates

So far we have discussed perovskite-type compounds with bivalent A ions and the tetravalent B ions. However, perovskite-type compounds can also be obtained with

$$A^{+1}B^{+5}O_3 \quad \text{and} \quad A^{+3}B^{+3}O_3.$$

Several ferroelectric compounds were reported by Matthias [88] in the first group: $KNbO_3$, $NaNbO_3$, $KTaO_3$ and $NaTaO_3$. The lattice perameters of $NaTaO_3$ and $KTaO_3$ were reported by Vousden [89] at room temperature, and the Curie temperatures were reported at 13°K for $KTaO_3$ [90] and 475°C for $NaTaO_3$ [88]. However, no detailed investigations have been published on the ferroelectric transitions of the tantalates, except for the dielectric properties of $KTaO_3$ near the Curie point at 13°K [90].

On the other hand, the dielectric properties and phase transitions of niobates have been studied in some detail by Matthias and Remeika [45], Wood [46], Pepinsky *et al.* [91], and Shirane *et al.* [47, 52]. It was shown that $KNbO_3$ undergoes three phase transitions in quite similar ways to those of $BaTiO_3$, achieving three ferroelectric phases of tetragonal, orthorhombic, and rhombohedral symmetry at the transition temperatures of 435°, 225°, and −10°C respectively.[4] If we reduce the three transition temperatures by dividing by the Curie temperature, they become respectively:

$BaTiO_3$:	1	0.69	0.49
$KNbO_3$:	1	0.71	0.38

It must be stressed that $KNbO_3$ is the only perovskite-type ferroelectric which has been found to show three transitions similar to those of $BaTiO_3$.

Small single crystals of $KNbO_3$ can be grown by using K_2CO_3 as a flux. A technique is not known, at present, for growing good $KNbO_3$ single crystals with plate-like habit. Dielectric constant vs temperature curve, in Fig. 32 (opposite), was obtained with a presumably multidomain crystal. It is obvious that this curve is essentially the same as the upper branch of Merz's curve for $BaTiO_3$, indicating that the domain configuration of the

[4] These transition temperatures are obtained from measurements of the dielectric constant for *increasing* temperature. On cooling, the three transitions occur at 410°, 200°, and −55°C. The temperature hysteresis at the lowest phase change is therefore about 45°C [47].

TABLE III

TRANSITION ENERGY ΔE (CAL/MOLE) AND ENTROPY CHANGE ΔS(CAL/MOLE DEGREE) AT THE THREE TRANSITIONS IN $BaTiO_3$ AND $KNbO_3$ (AFTER SHIRANE ET AL. [47])

		Cubic to Tetragonal	Tetragonal to Orthorhombic	Orthorhombic to Rhombohedral
$BaTiO_3$	ΔE ΔS	47–50 [151, 153, 154] 0.12–0.13	16–26 [151, 153–155] 0.06–0.09	8–14 [153–155] 0.04–0.07
$KNbO_3$	ΔE ΔS	190±15 0.28	85±10 0.17	32±5 0.12

tetragonal phase probably consists mainly of a domains. For further comparison with $BaTiO_3$, the dielectric properties near the Curie point should be known. A reliable value of the spontaneous polarization near the Curie temperature is difficult to obtain because of relatively high conductivity at these elevated temperatures. Recently, however, Triebwasser and Halpern [92, 93] overcame this difficulty by the use of a high frequency (5,000 cps) ac field for the hysteresis loop measurement, with the result that $P_s=26$ microcoulomb/cm² in the vicinity of the Curie temperature. The Curie constant is given as about 2.4×10^{5}°C [93]. It should be mentioned that a very extensive study of the phase diagram of the $K_2O-Nb_2O_5$ system was carried out by Reisman and Holtzberg [94]. The results are not only essential for crystal growth, but also explain the difficulties encountered in preparing ceramic specimens of $KNbO_3$.

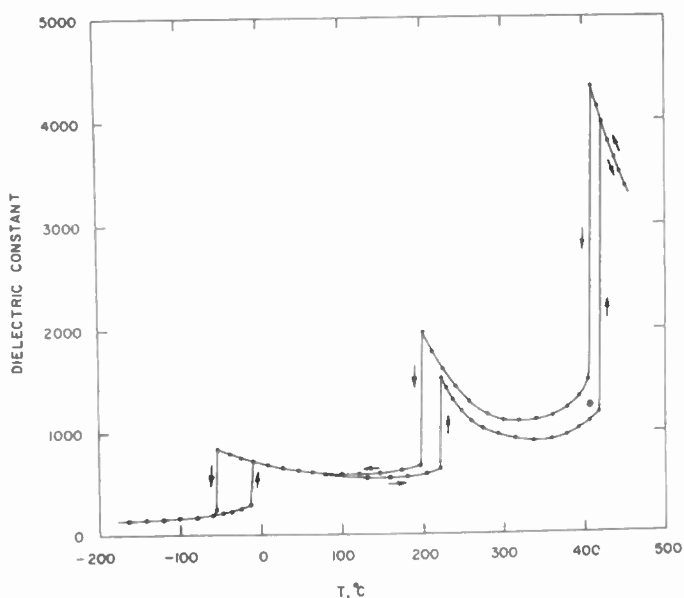


Fig. 32—Dielectric constant of a $KNbO_3$ multi-domain crystal as a function of temperature (after Shirane *et al.* [47]).

The temperature dependence of the lattice parameters of $KNbO_3$ was first studied by Wood [46] and then in detail by Shirane, Newnham, and Pepinsky [52]. The essential feature of the curve is quite similar to that of $BaTiO_3$, but the spontaneous strain is larger in $KNbO_3$ than in $BaTiO_3$. For example, the maximum tetragonality in $KNbO_3$ is $c/a=1.017$, compared with 1.009 in $BaTiO_3$. The volume decrease at the Curie temperature is 0.15 A³. A study was also made of the specific heat anomaly at the transitions. The results are shown in Table III above, together with data on $BaTiO_3$. The larger transition energies in $KNbO_3$ could be explained in terms of the larger lattice distortion. It may be interesting to point out that the relative entropy change of the three transitions is nearly the same in these two crystals; and moreover, the entropy changes at the Curie points are approximately proportional to their $(c/a)-1$ values in the tetragonal phase.

Thus the dielectric properties and the phase transitions of $KNbO_3$ are quite similar to those of $BaTiO_3$ from various viewpoints. This implies that the coefficients of the free energy expansion of the Devonshire type have similar relative magnitudes. This fact is quite interesting if we consider that these two substances have different A and B ions. It may be of considerable theoretical interest to explain this similarity from an atomic viewpoint. Some more information on $KNbO_3$, such as the atomic positions in the tetragonal phase, may be helpful for this purpose.

A recent study of nuclear quadrupole resonance of Nb^{93} in $KNbO_3$, carried out by Cotts and Knight [7], has special importance because of the unique and valuable information obtained concerning the crystalline field. What can be achieved from such experiments is the gradient of the electric field at the position of the Nb nucleus, and the asymmetric parameter η, which describes the dependence of the field gradient on the crystallographic direction. An adequate model for the ferroelectric properties of this crystal should be able to explain the observed values of the field gradient eq and η. Experiments gave $eq=3.10^{15}$ esu/cm³, and $\eta=0.806$. The field gradient at the Nb position was calculated for two different models: one assumed a point charge at the six nearest neighbor oxygens, and the second assumed a Megaw's model of a covalent bond [110, 114]. It was shown that the second model gives better agreement; but it is far from satisfactory. Unfortunately, the details of the models for the calculations are not shown; but it appears that the result may be very sensitive to the actual atomic positions in the orthorhombic phase which are not yet known. It may be mentioned that this

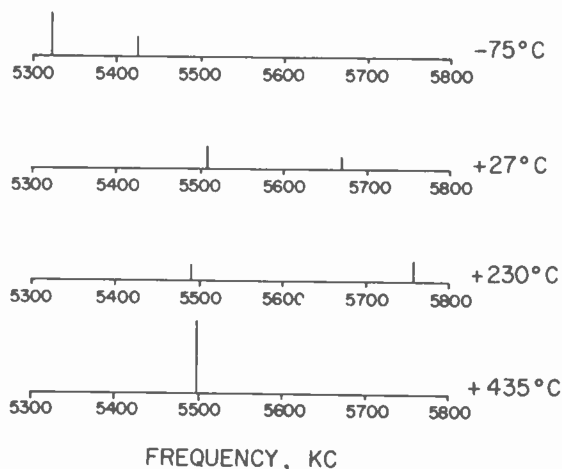


Fig. 33—Partial magnetic absorption spectrum of $KNbO_3$ at four different temperatures, showing the effects of crystal structure on the quadrupole splitting (after Cotts and Knight [7]).

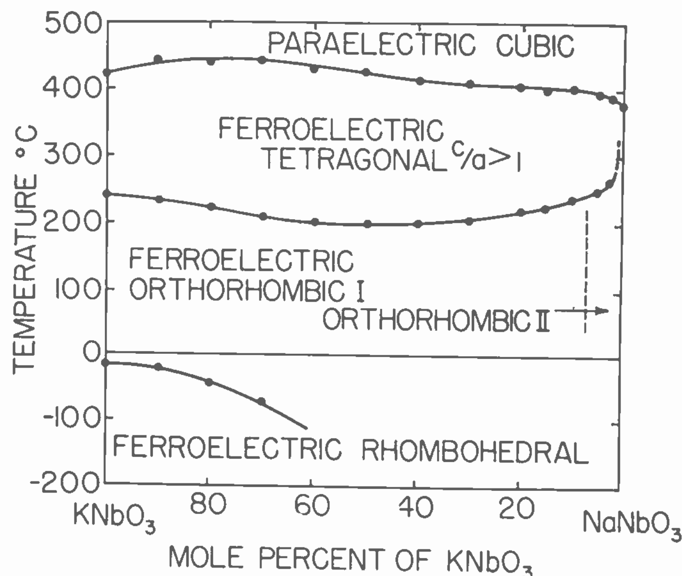


Fig. 34—Phase diagram of the $KNbO_3$-$NaNbO_3$ system. The two upper transition points of pure $NaNbO_3$ are not shown (after Shirane *et al.* [52]).

resonance experiment serves as a quite sensitive detector of the phase transition, as illustrated in Fig. 33.

The situation with respect to $NaNbO_3$ is quite complex, in contrast to the clear results for $KNbO_3$. $NaNbO_3$ was reported as ferroelectric by Matthias [45]. Then Vousden reported the structure to be nonpolar [51], which necessarily rejects ferroelectricity [95]. A further dielectric study indicated that the crystal may be antiferroelectric [52]. However, the phase diagram of the $KNbO_3$-$NaNbO_3$ system has revealed that the ferroelectric phase does persist for compositions close to pure $NaNbO_3$ (Fig. 34). It is therefore possible that a strong electric field could induce a ferroelectric phase, thus giving an observed anomalous P *vs* E relation on the oscilloscope. Three phase transitions were observed in $NaNbO_3$ at 350°, 480° and 640°C. [52, 45], but they are certainly of different nature than those observed in $BaTiO_3$ and $KNbO_3$. A very recent investigation by Cross and Nicholson [96] on single crystals of $NaNbO_3$ confirmed that the room temperature structure is nonpolar. Double hysteresis loops can be induced at room temperature, by very large fields (about 90 KV/cm) in the direction *perpendicular* to the orthorhombic c axis. At low temperatures, ferroelectric loops can be obtained for fields *parallel* to the c axis. In the most perfect crystals the ferroelectric Curie point seems to be below −209°C; but the ferroelectric phase may be induced by a high electric field above this temperature, and, once induced, it persists up to −55°C. The structure above 350°C is pseudotetragonal, truly orthorhombic, also antiferroelectric. The nature of the upper phase is not as yet well established.

D. Some Questions and Conclusions Concerning Perovskite Types

Among the numerous perovskite-type solid solutions so far studied, the $PbTiO_3$-$PnSnO_3$ system is of special interest in the light of Matthias's assumption that the B ion must have a noble gas configuration. The ferroelectric properties of the system were confirmed by Nomura [97] up to 80 per cent Sn concentration. It appears, however, that the compounds do not crystallize in the perovskite-type structure when the Sn content exceeds 80 per cent. Extensive studies were made by Graenicher and Jakits [98] of the $CaTiO_3$-$SrTiO_3$, (K, La)TiO_3-$BaTiO_3$ and (K, Nd)TiO_3-$BaTiO_3$ systems. Several different phases were observed in the first system. An unsuccessful attempt was made to interpret the results by the use of Roberts' theory [99, 100].

Now let us return to the question with which we opened this chapter. What is the determining factor for the Curie temperature of $BaTiO_3$? Detailed studies of several solid solution systems have not given an answer to this question; on the contrary, they have so far made the entire problem much more complicated. The question now is not only what will decide the Curie temperature of a perovskite-type compound, but also what type of phases will appear, and in what sequence. We have already seen that cubic perovskite compounds can transform into three different types of ferroelectric and several antiferroelectric phases. At present, there seems to be no simple explanation for these complicated phenomena. The situation is particularly difficult when one tries to deal with the compounds in which ferroelectric and antiferroelectric phases exist with small differences of energy. Although this delicate balance between two phases was nicely demonstrated by Takagi for a simple model, an explanation based on the real structure is an important matter which remains to be provided.

One other point needs emphasis. The phase diagrams we have discussed have shown that there are wide ranges of composition in which the cubic phase transforms directly into a rhombohedral phase, instead of taking

the tetragonal-orthorhombic-rhombohedral pathway realized in $BaTiO_3$ and $KNbO_3$. The symmetry of the rhombohedral phase is as high as a tetragonal one, and there is no obvious reason why the tetragonal phase should be preferred. It is also an interesting fact that there is no composition in which the cubic phase transforms directly into a ferroelectric orthorhombic phase.

VI. Some Crystal Structural Studies of Ferroelectrics and Antiferroelectrics

A. On the Need for Structural Information

The establishment of a phenomenological theory is the first step in the development of an understanding of the behavior of a particular ferroelectric crystal. As discussed earlier, this has been excellently accomplished for Rochelle Salt by Mueller [3], and for $BaTiO_3$ by Devonshire [4–6]. The next step is the examine the crystal structure of the material, preferably both in its nonpolar and polarized states, and to deduce as much as possible about the atomic mechanism of the transition.

Structural mechanisms have been proposed for the ferroelectric transitions in KDP, Rochelle Salt and $BaTiO_3$, and for the antiferroelectric transition in $NH_4H_2PO_4$, on the basis of the paraelectric structures only. Of these, Slater's explanation of the polarization mechanism in KDP was the first extensive effort at a model theory [16]. It was based upon West's excellent analysis of the structure of the nonpolar phase [15]. Subsequent X-ray [17] and neutron analyses [111, 18–21] of the crystal above and below its Curie point partly supported Slater's theory, and greatly extended knowledge of the mechanism. Nagamiya [57] was the first of a number of investigators [58, 59] who have proposed mechanisms for the antiferroelectric transition in $NH_4H_2PO_4$; and his scheme has been substantiated by an X-ray analysis of the crystal in both phases [58].

Several earlier attempts at structural explanations of the Rochelle Salt transition at the upper Curie point (cf.; *e.g.*, [26], and [9], pp. 235 *et seq.*) foundered on inadequate structural knowledge of the material in any of its phases. X-ray and neutron diffraction analyses in progress at Penn State [27] and Brookhaven [28] have clarified some aspects of the structure of the ferroelectric phase. In particular, they demonstrate that the mechanism formerly proposed for the polarization is quite incorrect; and the neutron study indicates one formerly unsuspected hydrogen shift of particular importance.

The structure of lithium ammonium tartrate (LAT) has been elucidated by Vernon and Pepinsky, in its nonpolar state [33]; and neutron observations confirm the change to a lower symmetry below the Curie point [101]. The mechanism of the transformation is quite uncertain, however, as is that of lithium thallium tartrate, since the actual structural differences between high and low temperature phases are still unknown. The mechanism of the ferroelectric transition in ammoniated Rochelle Salt, which was first discovered by Kurtschatov [29], is also unknown. Some X-ray work has been accomplished on the nonpolar phase of $[(NH_4)_{0.2}K_{0.8}]NaC_4H_4O_6\cdot 4H_2O$ [102]. This suggested that NH_4^+ ions go preferentially into one of two nonequivalent potassium positions.

The structures of the nonpolar phases of the oxygen-octahedra ferroelectrics, and in particular the perovskite and rhenium trioxide types, are certainly the simplest of all these anomalous crystals. The paraelectric materials are all cubic. This very high symmetry is in a large sense the source of the complexities of the ferroelectric and antiferroelectric structures, because the symmetry can be reduced in so very many ways. This is illustrated by the large number of phases which appear in pure and mixed perovskites, as briefly discussed in the preceding section. Until very recently no dependable positive results had been achieved in structure analyses of perovskite-type ferroelectrics; and model theories were consequently on unfirm ground. Now structures of tetragonal ferroelectric $BaTiO_3$ and $PbTiO_3$ have been deduced. the former by a single-crystal neutron study by Frazer *et al.* [44], and the latter by a combined X-ray and neutron analysis by Shirane *et al.* [103]. A single-crystal X-ray study of orthorhombic $KNbO_3$ is in progress, as is a neutron study of rhombohedral $Pb(Zr, Ti)O_3$. A refinement of Sawaguchi's X-ray analysis of antiferroelectric $PbZrO_3$ is also in progress, by combined X-ray and neutron methods [83].

A single-crystal analysis by Jona *et al.* [35] has refined Bystroem's [105] structure for cubic $Cd_2Nb_2O_7$; but the structure of the ferroelectric phase has not as yet yielded to an X-ray study.

Ueda and Kobayashi [106] have deduced a structure for the room-temperature, antiferroelectric phase of WO_3, following earlier studies by Bräkken [107] and Ueda and Ichenokawa [108]. The latter workers also examined the tetragonal high temperature modification, and their conclusions concerning the symmetry agree with those of Kehl *et al.* [109]. More structural work is certainly needed on the phases of this interesting material.

The condition of the model theory for the so-called ilmenite ferroelectrics, such as $LiNbO_3$ and $LiTaO_3$, is curious, in that Schweinler [37] has proposed a theory based on this structure, but there seems to be some doubt as to whether these lithium salts *are* ilmenite-type structures [110].

Structures are not available for the antiferroelectrics $(NH_4)_2H_3IO_6$ [60] and $Ag_2H_3IO_6$ [61]. The analyses of these will be difficult, and will probably require neutron methods. The structure of the newest ferroelectric, guanidine aluminum sulfate hexahydrate (GASH), is not yet available, but its analysis should be straightforward.

Successful analysis of the tetragonal ferroelectrics $BaTiO_3$, $PbTiO_3$ and KH_2PO_4, and of antiferroelectric $NH_4H_2PO_4$, are described in the following paragraphs. Some results on the Rochelle Salt and LAT structures

are also discussed, and the nonpolar structure of $Cd_2Nb_2O_7$ is reviewed. The completed analyses of ferroelectrics provide a sounder basis for model theories; and the incompleted work at least emphasizes the dangers of theorizing *before* reliable structural information is available.

One of the intriguing aspects of the structural studies is the contribution of neutron diffraction methods to the analyses. The techniques have been reviewed in several recent publications [111–113].

B. The Structure of Tetragonal $BaTiO_3$

We begin our discussion of ferroelectric and antiferroelectric structures with a study of room-temperature tetragonal ferroelectric $BaTiO_3$. The coordinates of the atoms in this structure can be expressed in terms of three parameters δz_{OI}, δz_{OII}, and δz_{Ti}, as follows:

Ba at (0, 0, 0); Ti at $(1/2, 1/2, 1/2 + \delta z_{Ti})$;

OI at $(1/2, 1/2, \delta z_{OI})$; OII at $(1/2, 0, 1/2 + \delta z_{OII})$ and $(0, 1/2, 1/2 + \delta z_{OII})$.

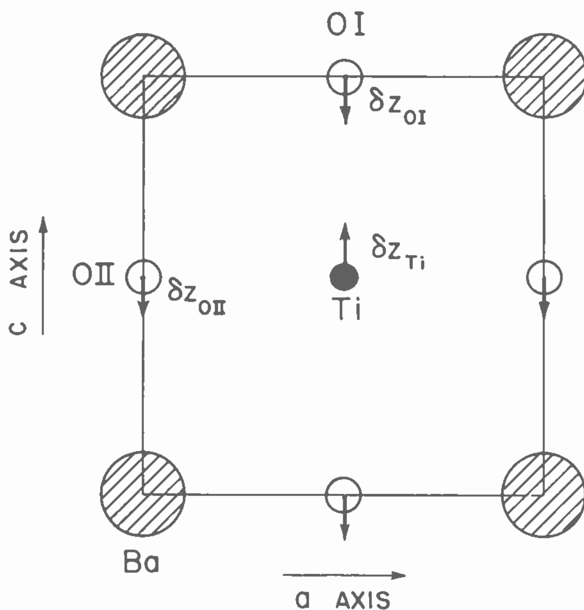


Fig. 35—The (010) projection of $BaTiO_3$ (schematic). The oxygen OII superposed upon Ti is omitted.

These coordinates are presented as fractions of the tetragonal cell edges: $a = 3.98$ A, $c = 4.026$ A. For the cubic structure, above the Curie point at 120°C, all of the δz values are zero. The parameters are indicated in the (0, 1, 0) projection, shown in Fig. 35, which is to be compared to the ideal perovskite lattice of Fig. 6. In addition to the three coordinate parameters, nine anisotropic temperature parameters are needed to describe the thermal vibrations of the atoms (two each for Ba, Ti and OI, and three for OII).

The first single-crystal structural study of $BaTiO_3$ was carried out by Kaenzig [114]. Kaenzig assumed that $\delta z_{OII} = 0$; *i.e.*, the oxygens on the faces parallel to the tetragonal axes have not shifted from their centered position, and hence the shortest distances in the cubic phase between Ba and OII ions remain equivalent in the tetragonal phase (the distances increase slightly and equally over those in the cubic phase, due only to the slight elongation along the c axis as the tetragonal symmetry develops below the Curie point). The subsequent analysis led to values $\delta z_{Ti} = +0.014$ and $\delta z_{OI} = -0.032$. The assumption concerning δz_{OII} arose from the model accepted at the time of the analysis, in which Ti and OI play an essential role in the ferroelectric behavior.

The structural problem was then taken up by Evans [115]. Evans avoided the assumption that $\delta z_{OII} = 0$. After an extensive and very careful study he was forced to conclude that the structure could not be solved uniquely by X-ray methods because of interactions between coordinate and temperature parameters along the tetragonal axis. This is equivalent to saying that agreement between observed and calculated X-ray intensities can be obtained with a rather wide range of different sets of parameters. Among the various possibilities, Evans found that reasonable assumptions concerning temperature oscillations led to the following coordinate parameters, which gave good agreement between observed and calculated intensities:

$$\delta z_{Ti} = +0.012, \quad \delta z_{OI} = -0.026, \quad \delta z_{OII} = 0.$$

The following parameters gave even a slightly better agreement in intensities, but required quite unreasonable temperature parameters:

$$\delta z_{Ti} = +0.015, \quad \delta z_{OI} = -0.024, \quad \delta z_{OII} = -0.020.$$

These two structures will be referred to as Evans' first and second structures, respectively. Because of the unreasonable temperature parameters, Evans disregarded the second model. His results thus tended to substantiate Kaenzig's assumption that $\delta z_{OII} = 0$; however, his over-all conclusion—as already stated—was that X-ray diffraction could not provide a definite answer to the problem.

It is worthwhile to examine the conditions which prevent a unique solution of the $BaTiO_3$ structure by X-ray diffraction. The chief sources of difficulty are: the massive scattering power of barium atoms for X rays; and the fact that we must deal with a noncentrosymmetric atomic arrangement, in which it is particularly difficult to separate the effects of small coordinate shifts and temperature oscillations of the lighter atoms because of the presence of a general phase angle in the crystal structure factors.

What we *observe*, in an X-ray diffraction study, are the *intensities* of reflected X rays, $I_{obs}(h, k, l)$, where h, k, l are indices which identify the set of reflecting crystallographic planes. The *magnitudes of observed structure factors* are given by the relation

$$|F_{obs}(h, k, l)| = K \cdot \sqrt{I_{obs}(h, k, l) \cdot L_p}. \qquad (24)$$

Here L_p is the *Lorentz-polarization factor*, which depends

only on the scattering angle, and the values of which are known; and K is a scale factor.

Structure factor can be *calculated* from the relation

$$F_{\text{calc}}(h, k, l) = \sum_{j}^{N} f_j e^{2\pi i(hx_j + ky_j + lz_j)}, \quad (25)$$

and

$$f_j = f_j^0 \cdot e^{-B_j(\sin\theta/\lambda^2)}. \quad (26)$$

Here f_j^0 is the tabulated value of the *atomic structure factor*, j referring to a specific atom in the unit cell; x_j, y_j and z_j are the coordinates of this jth atom in the cell, and N is the total number of atoms in the cell; B_j is that atom's *temperature parameter*, if the temperature oscillation is isotropic, and f_j is the scattering power for a temperature-oscillated atom; θ is the Bragg angle of scattering, and λ is the X-ray wavelength. If the temperature oscillations of the atoms are *not* isotropic, the expression for F_{calc} (h, k, l) is somewhat more complex.

In an X-ray analysis one must establish the coordinates x_j, y_j, z_j in some way, such that observed and calculated structure factors agree in amplitudes. We do not discuss this generally most difficult step in the analysis. In the special case of perovskite-type crystals, the approximate coordinates are very easily found.

F_{calc} (h, k, l) can be rewritten as

$$F_{\text{calc}}(h, k, l) = \sum_{j}^{N} f_j \cos 2\pi(hx_j + ky_j + lz_j)$$

$$+ i \sum_{j}^{N} f_j \sin 2\pi(hx_j + ky_j + lz_j)$$

$$= A(h, k, l) + iB(h, k, l). \quad (27)$$

In the case of $BaTiO_3$ the structure factor can be written (for any given set of indices h, k, l):

$$F = (A_{\text{Ba}} + A_{\text{Ti}} + A_{\text{O}}) + i(B_{\text{Ba}} + B_{\text{Ti}} + B_{\text{O}}). \quad (28)$$

If we substitute the coordinates for the various atoms, as given at the beginning of this subsection, we find that the real parts A_j contain linear combinations of $\cos 2\pi \cdot l \cdot \delta z_j$, and the imaginary parts B_j contain linear combinations involving $\sin 2\pi \cdot l \cdot \delta z_j$. Since the Ba atom is at the origin, $B_{\text{Ba}} \equiv 0$. When all δz_j's $= 0$, all B_j terms disappear.

Let us examine the contributions of these terms to the X-ray intensities. Recall, from (24), that

$$F^2 = K^2 \cdot L_p \cdot I. \quad (29)$$

We write

$$F^2 = (A + iB)(A - iB)$$

$$= A^2 + B^2$$

$$= A_{\text{Ba}}^2 + 2A_{\text{Ba}}(A_{\text{Ti}} + A_{\text{O}}) + (A_{\text{Ti}} + A_{\text{O}})^2$$

$$+ (B_{\text{Ti}} + B_{\text{O}})^2. \quad (30)$$

Table IV in column two shows the X-ray atomic scattering factor for Ba, f_{Ba}, is much larger than f_{Ti} and f_{O}. A_{Ba} is always much larger than the other A's and B's.

TABLE IV

COMPARISONS OF ATOMIC SCATTERING FACTORS OF SEVERAL ATOMS FOR X RAYS AND NEUTRONS

Element	Atomic Number	X-Ray Scattering Amplitudes f_x, 10^{-12} cm		Neutron Coherent Scattering Amplitudes
		$\theta = 0°$	$(\sin\theta)/\lambda = 0.5\text{Å}^{-1}$	f_{coh}, 10^{-12} cm
Ba	56	15.8	8.3	0.53
Pb	82	23.1	12.9	0.96
Ti	22	6.2	2.7	−0.38
O	8	2.25	0.62	0.58
H	1	0.28	0.02	−0.38
D	1	0.28	0.02	0.65
K	19	5.3	2.2	0.35
P	15	4.23	1.83	0.50

Now let us consider some simple model structures, such as those in Fig. 36. We arbitrarily set $\delta z_{\text{OII}} = 0$, but move Ti and OI atoms in the z directions. In (a), δz_{Ti} is +, and $\delta z_{\text{OI}} = 0$. In (b), δz_{Ti} is +, and δ_{OI} is −. In (c), both δz_{Ti} and δz_{OI} are +. All of these three models have identical A_{Ti} and B_{Ti}, and differ only in A_{O} and B_{O}. Models (b) and (c) have identical A_{O} values; but the signs of B_{O} for these cases are opposite.

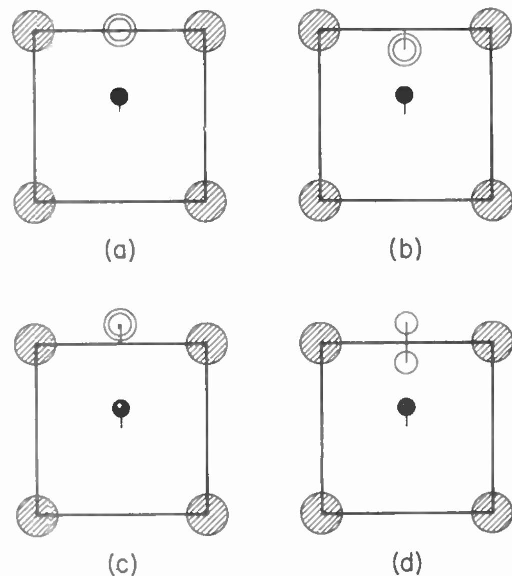


Fig. 36—Schematic representation of possible configurations resulting from the shift of the OI oxygen atom in $BaTiO_3$.

Changes in A_{O} produce significant changes in F^2, because A_{O} appears in a term $2 \cdot A_{\text{Ba}} \cdot A_{\text{O}}$. Thus we can distinguish model (a) from models (b) or (c). But it is difficult to distinguish between (b) and (c), because the difference is only in B_{O} in (30), and B_{O} is small compared with the A terms.

This difficulty is expressible in terms of the fundamental difficulty of diffraction: we can observe the absolute magnitude of F from an experiment, but not its real and imaginary components, A and B, separately. When B is much smaller than A, the absolute magnitude of F is not sensitive to the B term.

Let us now consider a fictitious model, such as that of (d) in Fig. 36, in which half of OI is shifted to $+\delta z$, and the other half to $-\delta z$. In this case B_{O} is zero, but A_{O} is

the same as for models (b) and (c). F^2 thus has nearly the same value for models (b), (c) and (d). When δz_{OI} is small in (d), this model cannot easily be distinguished from a model in which OI has a large anisotropic thermal vibration along the z direction. This is a qualitative argument only; but it does suggest the physical effects of the coordinate and temperature parameters on the calculated X-ray intensities, and thus illustrates the difficulties encountered by Evans.

The results of the X-ray analysis of $BaTiO_3$ were thus doubly negative. Firstly, an uncertainty exists in the coordinate and temperature parameters; and, secondly, the sign of the oxygen shift relative to Ti cannot be determined. Evans' assumption of motions of both OI and OII in directions opposite to the Ti shift, in his second model, was proved to be correct by a later neutron analysis. We will observe, below, that the same assumption would fail in the case of $PbTiO_3$.

When neutron diffraction is employed instead of X-ray scattering, the ratios of barium, titanium and oxygen scattering factors are entirely favorable for a structure determination. A comparison between X-ray and neutron scattering factors for several atoms is shown in Table IV. The neutron method does require a single crystal of considerable size, however. The analysis cannot be accomplished with a powder specimen, because the c/a ratio for the tetragonal cell is so close to 1 that various sets of reflections with different indices cannot be resolved in the diffraction pattern.

The crystal growth method of Remeika [40] has rendered single crystals of adequate size available for a neutron study; and a successful analysis has recently been accomplished by Frazer, Danner, and Pepinsky [44] at the Brookhaven reactor. Crystals of dimensions 0.3 mm by 2.5 mm by 15 mm were cut from larger c plates by Pepinsky's method [116], with the long dimension parallel to [1, 0, 0] or [1, 1, 0] directions. A strong dc field was applied in the c direction during the neutron exposures, to insure single-domain crystals. Some fifty ($h0l$) reflections were measured by standard neutron proportional-counter methods, and with high accuracy. No assumptions were made concerning δz_{OI} or δz_{OII} values. The following parameters were found, at 18°C:

$$\delta z_{Ti} = +0.0139; \quad \delta z_{OI} = -0.0230; \quad \delta z_{OII} = -0.0143.$$

Temperature factors found are (in 10^{16} cm^{-2}):

$$B_{Ti} = 0.153; \quad B_{OI} = 0.334; \quad B_{OII} = 0.266; \quad B_{Ba} = 0.274.$$

These are all isotropic. The displacements, in Ångstrom units, are:

$$\delta z_{Ti} = 0.06\text{A}; \quad \delta z_{OI} = -0.09\text{A}; \quad \delta z_{OII} = -0.06\text{A}$$

The disagreement factor,

$$R = \frac{\sum |F_{obs} - F_{calc}|}{\sum |F_{obs}|},$$

is less than 0.03, *for neutrons.*

This structure, which may be subject to slight alteration, is close to Evans' second model, but shows no anomalies in temperature oscillations.

Before discussing the physical significance of this structure, we will discuss the results of another recent analysis: that of tetragonal $PbTiO_3$ by Shirane, Pepinsky, and Frazer [103, 104].

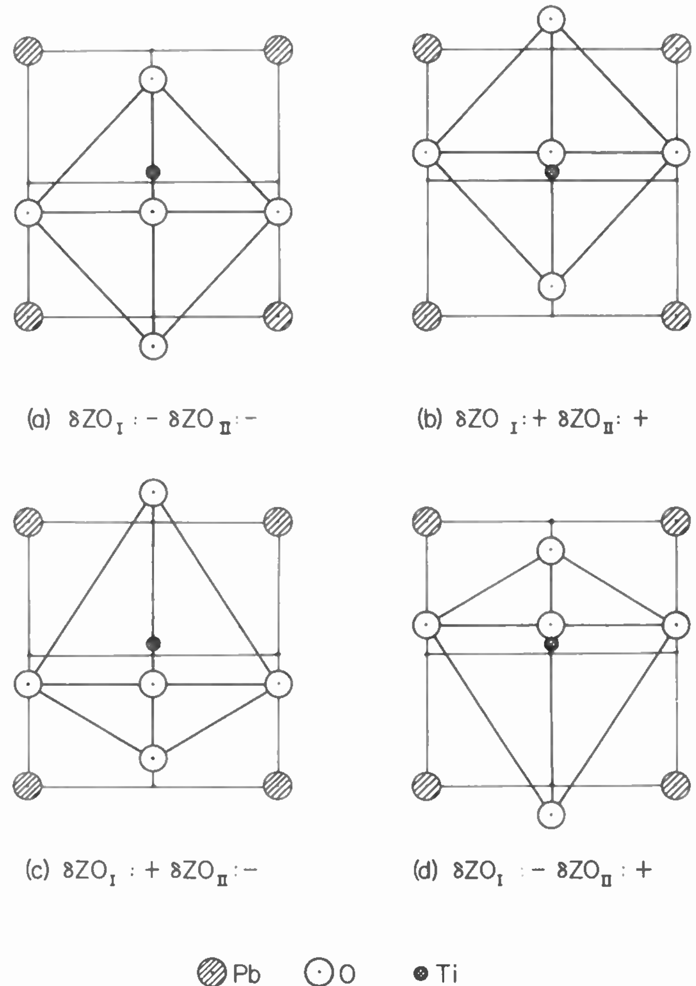


Fig. 37—Four possible models for $PbTiO_3$ structure.

C. The Structure of Tetragonal $PbTiO_3$

The conditions which rendered difficult the X-ray analysis of tetragonal $BaTiO_3$ also exist for tetragonal $PbTiO_3$. The X-ray scattering power of Pb is considerably larger than that of Ba (cf. Table IV). However, the ion shifts in the crystal may be much larger than those of $BaTiO_3$, as can be anticipated from the larger lattice distortion of $PbTiO_3$. Thus the ambiguity due to the interaction between coordinate and temperature parameters may be greatly reduced, but the ambiguity between models (b) and (c) of Fig. 36 may be more troublesome. In this case, as in $BaTiO_3$, the structure should be solved without ambiguity by a neutron diffraction study of single crystals.

At present a sufficiently large crystal of $PbTiO_3$ for a single-crystal neutron study is not available. Consequently, the analysis was carried out by the combination of an X-ray single crystal study and a powder diffraction study by neutrons. The ($h0l$) projection, based on

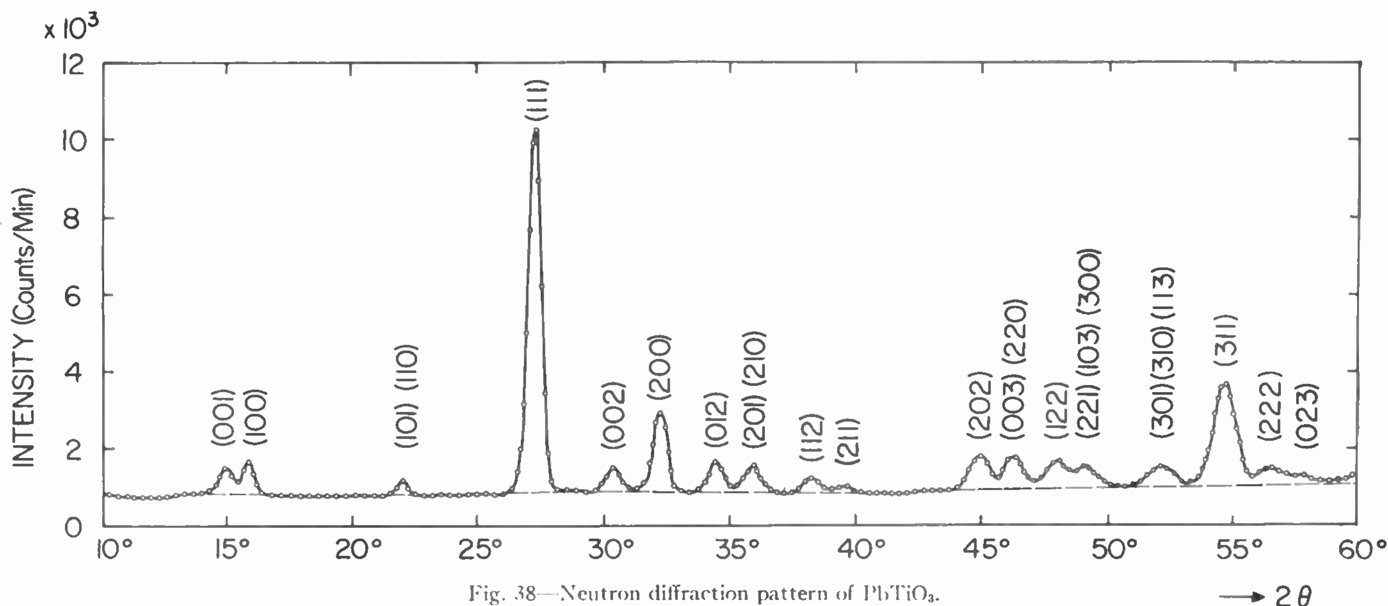


Fig. 38—Neutron diffraction pattern of $PbTiO_3$.

the intensity data collected by MoKα radiation, gave:

$$\delta z_{Ti} = +0.04, \quad \delta z_{OI} = \pm 0.11, \quad \delta z_{OII} = \pm 0.11.$$

Here we can take the sign of Ti shift as positive. There are four possible models corresponding to the combinations of signs of the oxygen shifts, as shown in Fig. 37. The R factors for all of the four models lie between 0.057 and 0.053, and the difference is certainly too small to be significant. It is thus not possible to decide between the four models on the basis of the X-ray intensities. An inspection of Fig. 37 reveals that models (c) and (d) are more improbable because they require drastic changes in O—O distances. Nothing could be concluded from geometrical considerations, however, as to a choice between models (a) and (b), even though model (a) has been assumed for $BaTiO_3$.

The neutron powder diffraction data permits an unambiguous decision between the structures, and proves that model (b) is correct. As shown in Fig. 38, the large tetragonality ($c/a = 1.06$) even allows resolution of the (100) and (001) reflections of the powder pattern. The agreement with model (b) is excellent, and the other three models are quite unsatisfactory. It is definitely not possible to account for the observed data with the assumption of $\delta z_{OII} = 0$. The final coordinate parameters are:

$$\delta z_{Ti} = +0.040, \quad \delta z_{OI} = +0.112, \quad \text{and} \quad \delta z_{OII} = +0.112.$$

This is a rather unexpected result, due to its departure from the case of $BaTiO_3$. Moreover it seems strange, at first sight, that the oxygens are shifted in the same direction as is Ti. It is misleading, however, to describe the structure in terms of shifts referring to the A atom at the origin. The z value of the origin is taken quite arbitrarily at the A atom; it could actually be at any point along the z axis. By fixing the origin at some point, for example at the A atom, we are implicitly fixing our viewpoint of the structure. The model with the origin at the A atom may be convenient for representing Kaenzig's and Evans' first structures for $BaTiO_3$, because only Ti and OI are shifted in these . But it is not at all suitable for describing the $PbTiO_3$ structure.

In speaking of atomic "shifts," one recalls the structure of the cubic phase. It is certainly preferable to describe the new structure in terms of alterations in bond lengths and angles. In the case of $PbTiO_3$, however, the equivalence of δz_{OI} and δz_{OII} leaves the oxygen octahedra undistorted except for an extension along the z axis. Since the close-packing of oxygens is an anticipated feature of both these titanates, and the above analyses reveal that this packing is not much disturbed as the transitions occur, it is both convenient and reasonable to place the origin at the z level of OII. Fig. 39 shows the

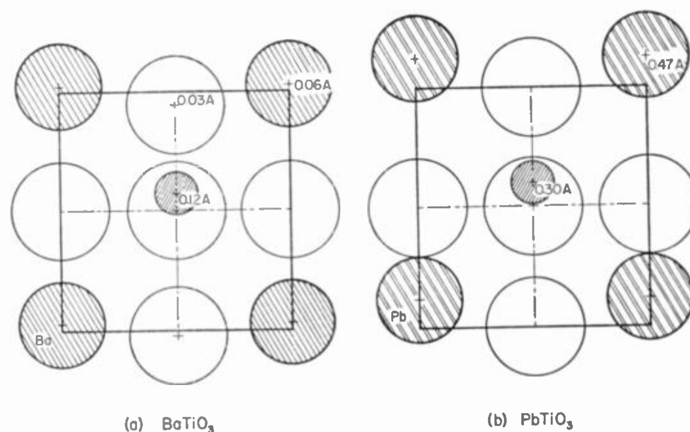


Fig. 39—Models for tetragonal $BaTiO_3$ and $PbTiO_3$ (schematic: the size of the ions shown was chosen arbitrarily).

tetragonal structures of both titanates considered in this way. (The reader should note that the figure for $PbTiO_3$ is inverted compared to the model with δz_{OI} and δz_{OII} both + in Fig. 37.)

From this point of view, the Ti in $PbTiO_3$ is shifted by 0.30 A with respect to the oxygen octahedron, and the Pb moves in the same direction by the larger

TABLE V

COMPARISONS OF δz VALUES AND BOND LENGTHS, IN A, FOR $BaTiO_3$ AND $PbTiO_3$, AT ROOM TEMPERATURE. OXYGEN OII IS PLACED AT $c/2$ (AFTER SHIRANE *et al.* [103, 104]).

		$BaTiO_3$ (in A)	$PbTiO_3$ (in A)
δ_z Values	δz_{OI}	−0.03	0.00
	δz_{Ti}	+0.12	+0.30
	δz_{Ba} or δz_{Pb}	+0.06	+0.47
Bond Distances, Ti to:	OI (upper)	1.86	1.78
	OII	2.00	1.98
	OI (lower)	2.17	2.38
Bond Distances, Ba or Pb to:	OII (upper)	2.80	2.53
	OI	2.82	2.80
	OII (lower)	2.88	3.20
Axial Lengths	a	3.99	3.90
	c	4.03	4.15

distance of 0.47 A. The structure of $BaTiO_3$ revealed by neutron diffraction indicates a shift of Ti by 0.12 A and a shift of Ba in the same direction but by a smaller distance of 0.06 A (see Table V, above). The oxygen octahedron is only slightly distorted in $BaTiO_3$. These models show features common to both crystals, together with important differences. Moreover they clearly reveal the important role which Pb ion plays in $PbTiO_3$.

D. A Discussion of the $BaTiO_3$ and $PbTiO_3$ Structures

Let us examine the changes of bond lengths in $BaTiO_3$ and $PbTiO_3$ due to the ion shifts associated with the ferroelectric phase transition (Table V). We will distinguish between the upper and lower OI atoms by the symbols OI(u) and OI(l). A remarkable change common to both crystals is the Ti-OI(u) distance, which is 1.85 A in $BaTiO_3$ and 1.78 A in $PbTiO_3$. These should be compared with the sum of the Goldschmidt radii, 1.96 A. The situation for $BaTiO_3$ is shown in Fig. 40,

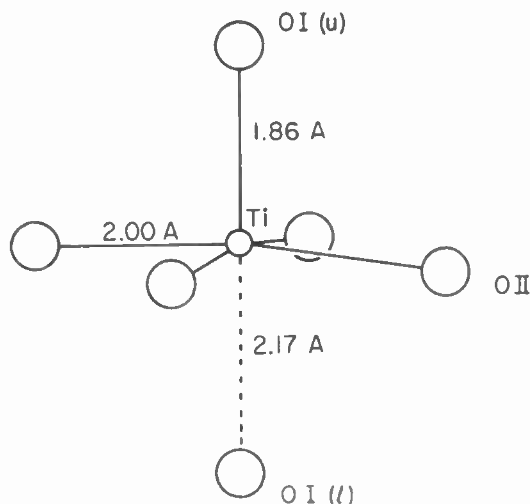


Fig. 40—Environment of Ti in tetragonal $BaTiO_3$.

and for $PbTiO_3$ in Fig. 41. The Ba and Pb atoms have the large coordination number 12, and consequently a large bond distance to surrounding oxygens. Therefore, the small shift of Ba does not give any noticeable change of Ba—O bond lengths in $BaTiO_3$. The situation is, however, quite different in $PbTiO_3$; the very large shift of Pb along the z direction does appreciably influence the Pb—O bond system. The Pb-OII(u) length is 2.53 A, which is again much shorter than the sum of Goldschmidt radii, 2.78 A. (See Fig. 42.) It appears that this large change of Pb—O bond length is the most pronounced difference between $PbTiO_3$ and $BaTiO_3$.

From these observed structures, we can estimate the polarizations due to ion shifts. Generally speaking, the spontaneous polarization consists of two parts: one is ionic, due to the ion shifts, and the second is electronic, due to the deformation of electron clouds. Of course this separation is rather formal, but it gives some idea of the nature of the polarization in ferroelectric crystals. The ionic part can be calculated from

$$\frac{1}{v}\sum_i n_i \delta z_i,$$

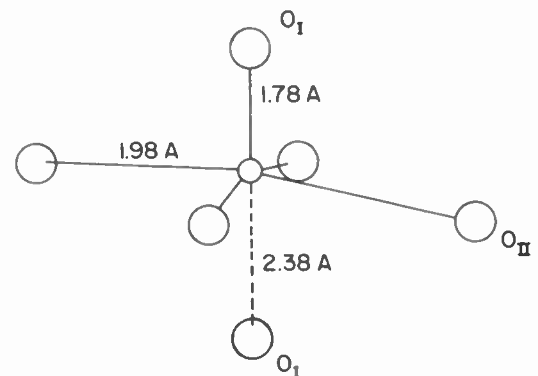


Fig. 41—Environment of Ti in $PbTiO_3$

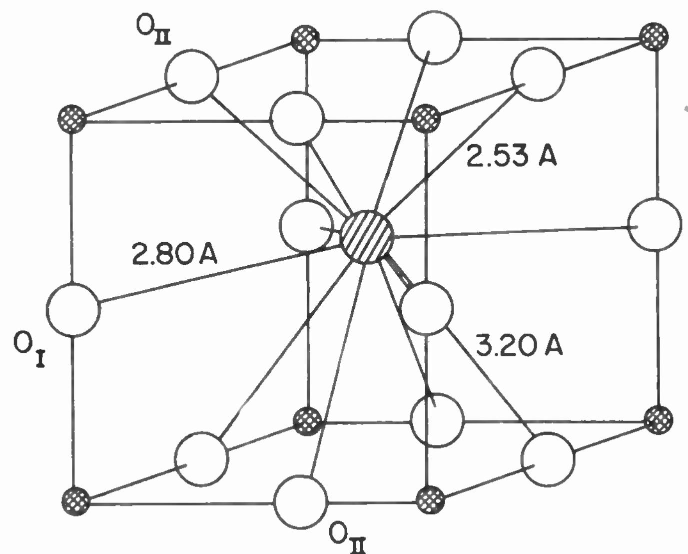


Fig. 42—Environment of Pb in $PbTiO_3$.

where n_i is the effective charge of the ith ion, and v is the unit cell volume. Since the individual δz's depend upon the choice of origin, no physical meaning can be given to the individual term. The sum is constant, independent of the choice of origin, however.

The effective charges of the ions are not known, unfortunately; but charges based on assumed completely ionic character for the crystals give a maximum possible contribution of the ionic part. Thus the assumption of a completely ionic crystal gives ionic polarizations of 17 and 54 microcoulomb/cm² for $BaTiO_3$ and $PbTiO_3$, respectively. The differences between these figures and the observed spontaneous polarizations are the minimum contributions to be attributed to the electronic part. The observed polarization value for $BaTiO_3$ is 26 microcoulomb/cm². No reliable value of P_s is available for $PbTiO_3$ at present.

Now let us examine the existing theories of $BaTiO_3$ type ferroelectrics in the light of the structures described above. Model theories of ferroelectrics may be classified into two general types. The first is based on dipoles with two positions of equilibrium. This is equivalent to an ion moving in the type of field shown in Fig. 43(b). The Mason-Matthias [117] theory of $BaTiO_3$ is typical of this type. It is related to the earlier Eucken-Büchner [118] theory of dielectric behavior in certain high dielectrics. Six equivalent positions are assumed for Ti in the cubic phase, and one of these six positions is preferentially occupied in the tetragonal phase due to a cooperative phenomenon of long range forces. Several fundamental objections to this theory have already been discussed [75]. It was shown by Devonshire that the Curie constant of this type of model cannot be as large as that observed in $BaTiO_3$. Moreover, Kaenzig's observation of the structure of $BaTiO_3$ above the Curie temperature did not show any evidence for such a model.

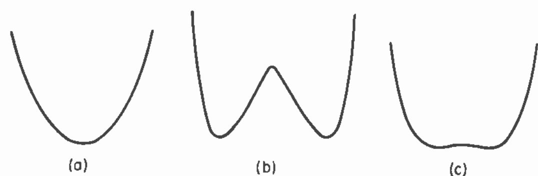


Fig. 43—Different types of potentials for ions in ferroelectrics (after Devonshire [6]).

The second model may be called the oscillating ion type. In this, ions are assumed to move in a potential well as shown in Fig. 43(a). The Curie constant of this type of model is determined by the reciprocal of the anharmonic coefficient of oscillation, b. The large Curie constant of $BaTiO_3$ can be accounted for by this model without difficulty, since b may be quite small. Slater [160] developed a theory along this line, assuming a Ti displacement only and using the Lorentz field based on the actual atomic arrangement. It was shown that the Ti and OI ions along the chain in the polarization direction exert very strong fields on each other. In this theory the ionic polarization of the Ti ion is an essential factor.

According to Devonshire's estimate, the oxygen ions are much more loosely bound to their symmetrical positions than are the Ti ions. Jaynes [119] has proposed the "oxygen displacement theory," which depends upon this concept. Each cubic unit cell has three oxygen ions, O_x, O_y, O_z, which are free to move in the x, y and z directions respectively. As the crystal is cooled, the mismatch of ionic sizes successively causes the O_z, then the O_y, then the O_x ions to be squeezed out of the planes of Ba ions, thus causing the spontaneous polarization first along a cube edge, then along a face diagonal, and finally along a body diagonal. Thus this theory could quite naturally account for the three phase transitions and the associated strains.

This oxygen displacement theory was supported by Kaenzig's and Evans' first structures, in which only OI and Ti are shifted with respect to Ba and OII. Here the Ti shift can be considered as being induced by the oxygen move. However, the $BaTiO_3$ and $PbTiO_3$ structures revealed by neutron diffraction suggest that the oxygen octahedra suffer little distortion in passing through the transition, in contradiction with the oxygen displacement theory. The observed structures are rather in accordance with Slater's model. in which the Ti move was considered with respect to the oxygen octahedron. The important role of Pb in $PbTiO_3$ is not considered in this theory.

The theoretical treatments we have just discussed are based on the long range forces of dipole interaction. Other investigators, and recently among them Megaw [110, 144], have approached the problem in rather another way. This is to consider the transition as associated with a change of the type of chemical bonding rather than a change in the dipole structure. The importance of the homopolar character of the bond is emphasized, and the bond angles at OII with Ti were considered as an essential factor for the tetragonal structure. This put the emphasis on the Ti-OII-Ti line, rather than on the usually emphasized Ti-OI-Ti chain. In $PbTiO_3$ Megaw [110] stressed the hypothesis that the homopolar character of the Pb—O bond may play an important role, and she pointed to the homopolar bond system in PbO crystals. In the PbO structure Pb—O bonds form a flat tetragonal pyramid with Pb at the apex, and the Pb to O distance is 2.33 A. The environment of Pb in $PbTiO_3$ may be considered as similar to that in PbO. This treatment is entirely based on the short range force due to bonding, and it is considered that the long range force is unnecessary to explain the transition.

The crystal structures of tetragonal $BaTiO_3$ and $PbTiO_3$ certainly indicate that the compounds are not purely ionic. This does not require that the homopolar bond character is a driving force which induces the ferroelectric transition. Whether the long-range dipolar interaction forces, or the short-range homopolar bond properties, are most important for the transition, can be answered in part at least by calculations based on the tetragonal structure. Computations by Cohen [143] had shown that the antiferroelectric state is slightly more stable than the ferroelectric state for $BaTiO_3$; and this result had been considered a weak point in the dipole theory. Takagi [87a] has recently shown that a CsCl-type crystal can become ferroelectric without invoking

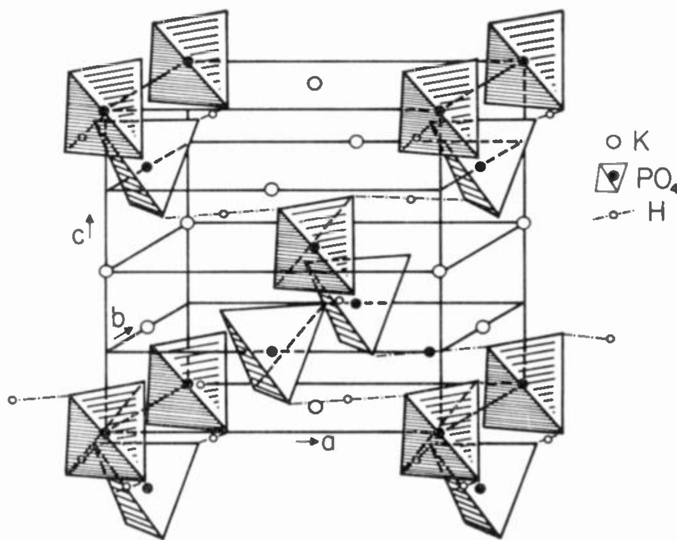


Fig. 44—Structure of KH_2PO_4 (after West [15]).

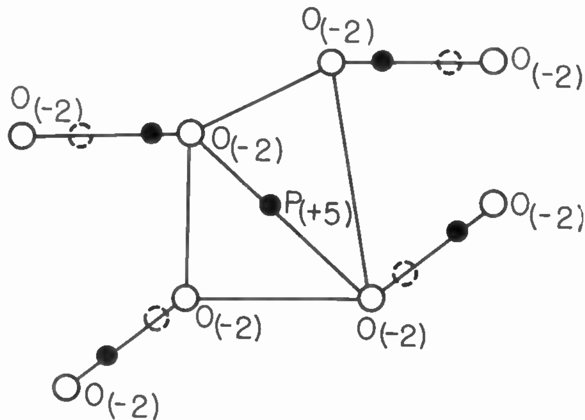


Fig. 45—One PO_4 tetrahedron and H bonding in KH_2PO_4 (after Mason [9]).

short-range forces. Moreover, the dipole-dipole interaction theory does indicate that for Evans' second model, the ferroelectric state is actually slightly more stable than the antiferroelectric one. The evidence to date is that dipole interactions *are* the main source of perovskite-type ferroelectricity, but that bond character must be considered as well.

E. Structural Changes in KH_2PO_4 at the Curie Point

The structure of KH_2PO_4, as deduced by West [15], is shown in Figs. 44, 45 and 46. It consists of a three-dimensional framework of PO_4 groups, linked together by hydrogen bonds: the two *upper* oxygens of one PO_4 tetrahedron are linked by H's to *lower* oxygens in two adjacent tetrahedra; the two *lower* oxygens of the first tetrahedron are linked to *upper* oxygens in two other neighboring tetraehedra.

Slater [16] assumed that proton ordering was responsible for the ferroelectric transition in this crystal. The spontaneous polarization is along c. He visualized the

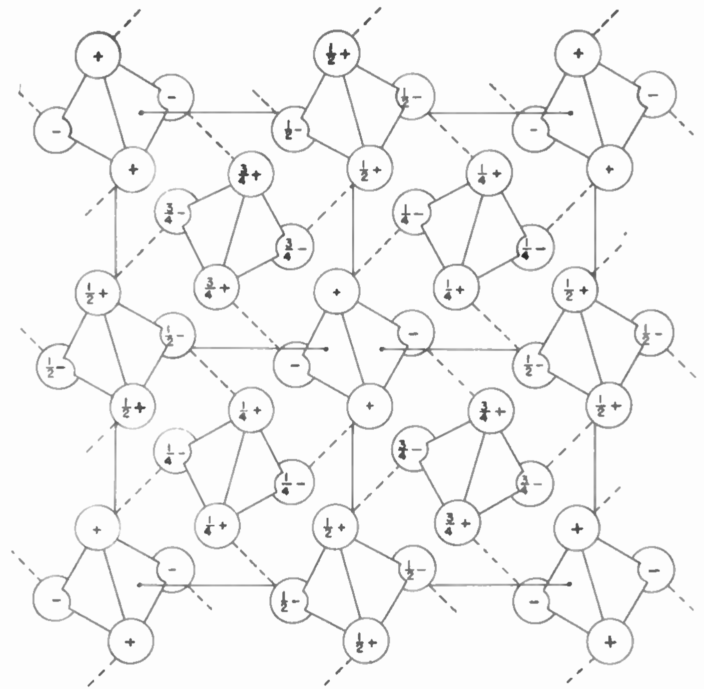


Fig. 46—The H_2PO_4 system looking down the tetragonal axis in $F\bar{4}d2$ (after Frazer and Pepinsky [17]).

formation of $H_2PO_4^-$ groups, by hydrogens associating more closely with the *upper* oxygens of each PO_4 group, *e.g.*, as shown in Fig. 45. Such an association is permitted, if the crystal symmetry is reduced by the transition. Slater further assumed that the H_2PO_4 with dipole orientation along $\pm c$ has a lower energy, by an amount Δ, than a group with its dipole axis perpendicular to c.

The results of Slater's statistical-mechanical treatment accounted for several important properties, including the form of the temperature-dependence of the dielectric constant ϵ_c, and the entropy change at the transition. Takagi [120] later showed that the theory could also explain the temperature dependence of the dielectric constant ϵ_a, perpendicular to the c-axis. Yomosa and Nagamiya [121] modified the theory to include deformation effects, and they were able to account for the anomalous piezoelectric effect and spontaneous strain in the ferroelectric region.

A difficulty with Slater's theory is that the dipole moment of each phosphate group, and therewith the moment of the entire crystal, is attributed directly to the positions of the positive protons relative to the negative phosphate groups. Yet upon reversal of polarization the displacements of the protons must be very nearly perpendicular to the c axis, as shown by crystallographic studies. Thus the displacements of the protons cannot explain the change in polarization *parallel* to the c axis. This discrepancy does not invalidate Slater's statistical mechanical analysis, but implies the presence of additional displacements which must have appreciable components parallel to the c axis.

The present picture, first afforded by the X-ray

analysis of the transition by Frazer and Pepinsky [17], is that hydrogen bond changes perpendicular to *c* do occur, and are accompanied by alterations in dispositions *of other atoms* along the *c* direction. The spontaneous polarization as well as the dielectric behavior in the *c* direction about the Curie point depend upon these latter alterations.

Frazer and Pepinsky carried out an extensive X-ray study of KDP at 126°K and 116°K. These temperatures are 3° *above* the transition point, and 7° *below* it. (De Quervain [12] had noted anomalies in scattering at temperatures closer to the Curie point.) This analysis was followed by a preliminary neutron diffraction study of the structure at room temperature by Pepinsky and Frazer [111], and by extensive neutron studies above and below the Curie point by Peterson *et al.* [18], Levy *et al.* [19], and Pease and Bacon [20, 21].

One of the concerns of the structural study is the behavior of the hydrogens through the transition. As West first showed, the hydrogen bond is almost perpendicular to the *c* axis, the angle with the *ab* plane being 2° or less. The H bonds are shown schematically as dotted lines in Figs. 44–46. Later measurements show the O—H. . . O directions to be approximately 0.5° out of parallelism with the *ab* plane.

Table VI lists O—H. . . O bond lengths as reported by the various investigators and at various temperatures. It is to be noted that the temperatures at which the low-temperature measurements were made are different for the different investigators. It must be emphasized here, and certainly later when motions of other atoms are considered, that analytical results accomplished at *different* temperatures in the neighborhood of the Curie point are not really comparable. This is to be anticipated not merely from de Quervain's results, but from the tremendous dielectric anomaly in the temperature range a few degrees on either side of the transition point. The measurements most likely to shed light on the mechanism itself are those closest to Curie point.

TABLE VI
HYDROGEN BOND LENGTHS IN KDP

	Room Temperature	Above Curie Point	Below Curie Point
West [15]	2.53 A	—	—
Frazer and Pepinsky [17]	—	2.44 A (at 126°K)	2.51 A (at 116°K)
Peterson, Levy, etc. [18, 19]	2.48 A	2.48 A (at 140°K)	2.50 A (at 113°K)
Bacon and Pease, etc. [20, 21]	2.49 A	2.48 A (at 132°K)	2.49 A (at 77°K)

Frazer and Pepinsky, observing at temperatures 3° above the Curie point and 7° below it, found O—H. . . O distances of 2.44 A and 2.51 A, respectively. This suggested a very strong H bond above the transition, and a weakening of this bond below. Pepinsky and Frazer's neutron study of KDP at room temperature [111] suggested a double minimum in the H position, with a half-hydrogen, on the average, in each minimum position. This tentative result was first disputed by both Peterson and Levy [18] and Bacon and Pease [21]; but in later communications both of the latter groups of workers tend to view the double minimum as quite possible on the basis of the neutron data, and undoubtedly the case on the grounds of thermal and infrared data. [164].

Levy *et al.* [19] found the O—H. . .O distance to increase only slightly between 140°K and 113°K, and Bacon and Pease [20, 21] found only a slight increase at 77°K over the separation at 132°K. These results are in contradiction with the X-ray results of Frazer and Pepinsky. It is probable that the measurements of the latter workers, at 126°K and 116°K, actually reflect structural conditions in the immediate neighborhood of the transition which do not obtain at larger temperature differences from it.

A second piece of evidence suggesting a mechanism for the transition was furnished by the X-ray study [17]. Projections perpendicular to the *c* axis, on the (1, 1, 0) plane, are shown in Figs. 47 and 48 (next page), at 126° K and 116°K respectively. These figures show a pair of PO_4 tetrahedra, the atoms of one being marked with primes. Two oxygens of each tetrahedron overlap in the projection (on the central horizontal line). Potassium atoms are shown above and below the respective tetrahedra. The reader should refer to Fig. 44, and think of the structure as viewed along the *b* direction.

An analysis by Frazer and Pepinsky of West's X-ray data, using modern analytical methods, indicated hydrogen peaks between oxygens in adjacent tetrahedra and also indicated a slight anisotropy in the thermal oscillations of K and P atoms along the *c* axis. At 126°K, Frazer and Pepinsky found the PO_4 tetrahedra to be regular, but each K atom, which is surrounded by eight O's, shows larger distances to the O atoms above and below it than to the side O's. The K atoms oscillate thermally with larger excursions along the *c* direction. This is in keeping with the nature of the O-packing around them. A similar thermal anisotropy is found for the P atoms.

This thermal anisotropy is in keeping with the dielectric behavior at 126°K. At this small temperature above the Curie point the dielectric constant along *c* is very large (nearly 10^4), but the constant along *a* is normal. The dielectric constant drops rapidly as the temperature is increased. At 132°K, where Bacon and Pease [21] observed the structure, the dielectric constant is of the order of 10^2; and at 140°K, where Peterson *et al.* [18] studied it, ϵ_c has dropped still more. It is reasonable to associate a high anisotropic polarizability with a high thermal oscillation anisotropy. The fact that neither Bacon's nor Peterson's group observed evidence of thermal anisotropy from neutron studies at 132° and 140°K is perhaps attributable to the fact that neither worked close enough to the transition temperature.

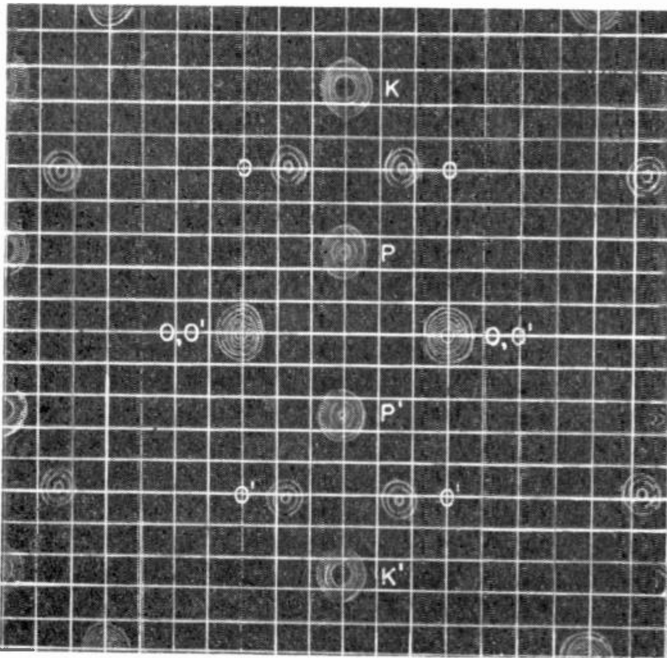


Fig. 47—Electron density projection of KH_2PO_4 on (110) in F$\bar{4}$d2 (equal to (100) in I$\bar{4}$d2) at 126°K (after Frazer and Pepinsky [17]). A better reproduction of the above figure may be seen in a previously published paper by Frazer and Pepinsky [17].

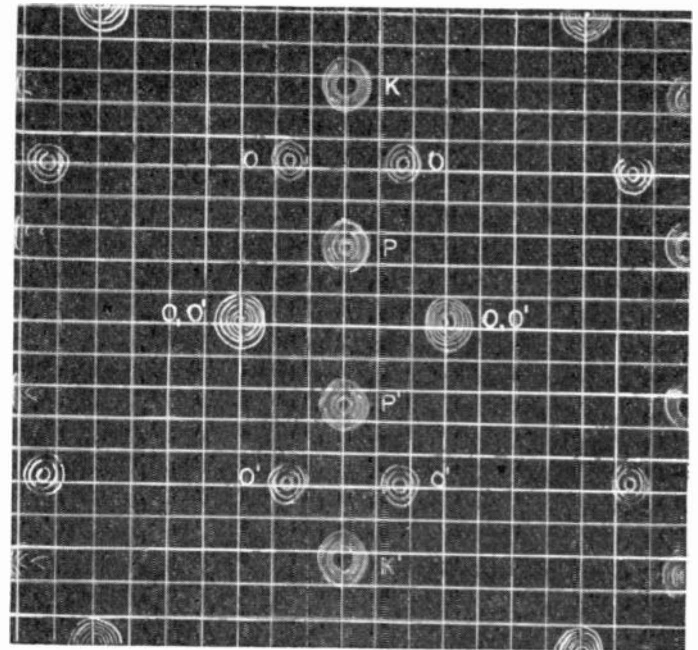


Fig. 48—Electron density projection of KH_2PO_4 on (110) in Fdd at 116°K (after Frazer and Pepinsky [17]). A better reproduction of the above figure may be seen in a previously published paper by Frazer and Pepinsky [17].

However, it is also to be mentioned that difference maps based on West's original excellent X-ray data also clearly indicate strong thermal oscillations for K and P atoms —or at least elongation of the K and P electron clouds— in the *c* direction. If the absence of thermal anisotropy very close to the Curie point can be confirmed by a further neutron study, then consideration must be given to the possibility that the electron clouds are elongated along *c*.

The X-ray study at 116°K showed not only the increased O—H. . .O bond distances reported above—thus suggesting a tendency toward *ordering* of the hydrogens —but also showed the P and K atoms to be displaced somewhat in the *c* direction. The O tetrahedra remain undistorted. The projection of this ferroelectric structure is shown in Fig. 48, viewed perpendicularly to the *c* axis. Comparison of Figs 47 and 48 shows the displacements of P and K atoms in the *c* direction, as the crystal becomes polarized. The P atom in the upper (unprimed) tetrahedron moves *down* slightly (0.03 A) from the central position; the K atom directly above this tetrahedron moves *upward* by 0.05 A. Thus the P and K move *away* from the upper two O's in this tetrahedron. It is to be assumed that these are then the O's to which the H's have become more closely associated. In short, what apparently happens is that the K and P atoms build up a preferred vibration along the *c* axis, as the crystal is cooled to just above 123°K. This anisotropic oscillation is facilitated by a change in the packing of O's around the K. The polarizability along *c* increases with the anisotropic oscillations of K and P. As the Curie point is attained, the K and P atoms *lock in* to off-center positions, and the H atoms order correspondingly: H's remain close to O's, away from which the P's and K's have moved.

A computation of the spontaneous polarization, from the positions of P, K and O's as found by Frazer and Pepinsky at 116°C, gives 4.1 microcoulomb/cm². The observed value of the spontaneous polarization is about 4.5 microcoulomb/cm².

Although neither Peterson *et al.* nor Bacon *et al.* find preferred oscillations of P and K atoms, at their temperatures of observation, both groups beautifully confirm the concept of ordering of the hydrogens below the Curie point, and also the displacements of K and P atoms. Projected views of KDP along the *c* axis are shown in Figs. 49, 50 and 51, taken from Bacon and Pease's papers. Fig. 49 shows a schematic of the projected structure on the *ab* plane, above the Curie point. Fig. 50 shows a contour map of density of neutron-scattering material, and represents the portion inside the dotted lines of Fig. 49, for the structure at room temperature. Fig. 51 represents hydrogen atoms, within the region outlined by dotted lines in Fig. 49, at 77°K.

Fig. 50 shows hydrogen nuclei as kidney-shaped dotted contours, between oxygens of adjacent tetrahedra. In extended analyses none of the neutron investigators were able to determine with complete assurance whether these maps represent a hydrogen in a single broad minimum, or two half-hydrogens in a double minimum, the two parts of which were statistically occupied. This is a difficult point to clear up by a structure analysis. Other evidence [164] clearly points to the latter type of disorder. *Below* the Curie point the hydro-

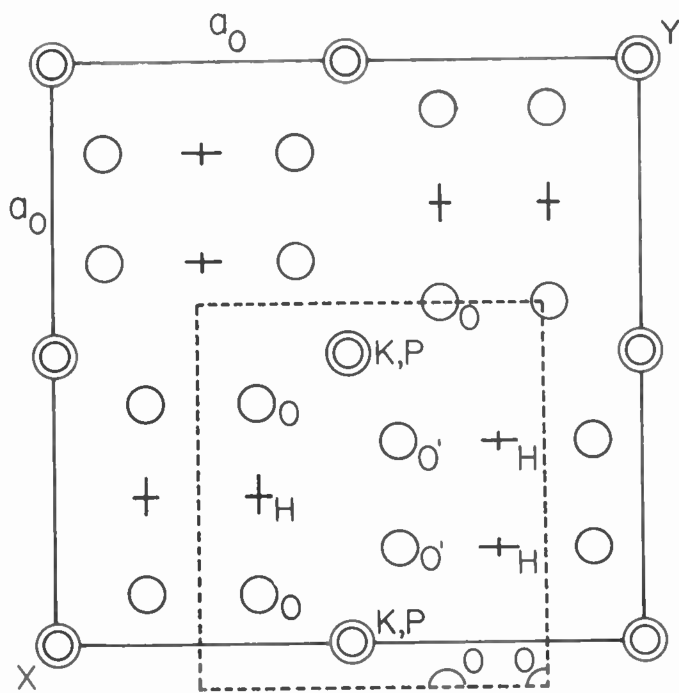


Fig. 49—Projection of KH_2PO_4 on (001) (after Bacon and Pease [21]).

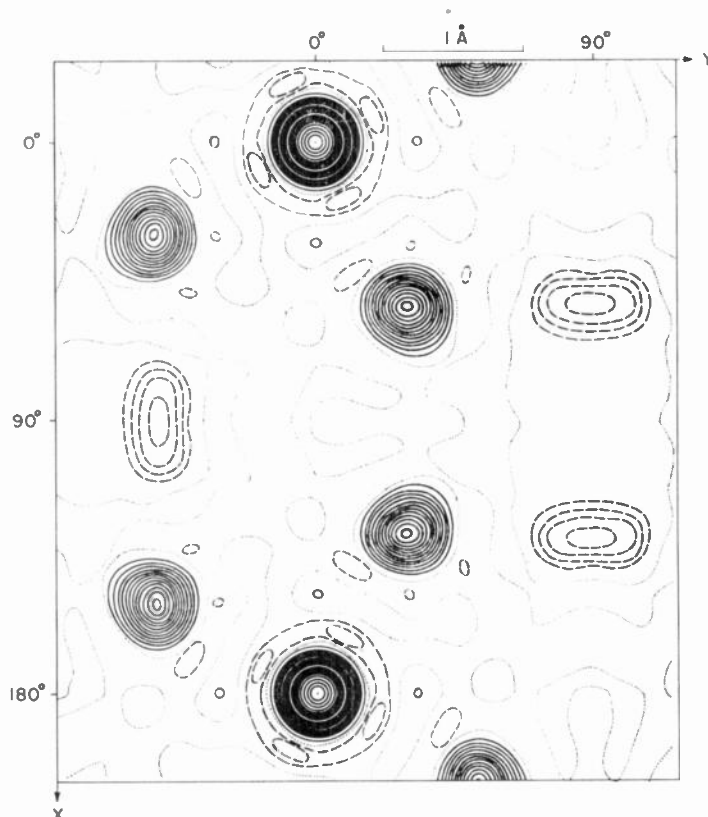


Fig. 50—Projection on the (001) plane of KH_2PO_4 at room temperature. The broken lines represent negative contours, thus showing hydrogen. This figure corresponds to the region outlined by the dotted lines in Fig. 49 (after Bacon and Pease [21]).

gens are clearly ordered, as shown in Fig. 51. In this last figure the effects of K, P and O atoms have been subtracted out of the pattern. The hydrogen is observed at 0.21 A from the *center* of the O—H. . .O bond of 2.49 A: 1.04 A from one O atom, and 1.46 A from the other O. This pattern was of course for a single-domain crystal, maintained in this state by an applied electric field When the field direction was reversed, the hydrogen changed its equilibrium position so that it was closer to the other oxygen.

Summarizing these results on KDP: Slater's brilliant quasi-model theory is shown to be well founded, and X-ray and neutron analyses have succeeded in clarifying most aspects of the transition mechanism. The transition occurs due to remarkable cooperation between *hydrogen bond effects* and *changes in thermal oscillations and mean positions of K and P atoms.* The tremendously delicate balance of forces within a crystal, and the effect of thermal energy upon this balance, is well illustrated by the ferroelectric transition in KDP.

F. The Structural Changes in $NH_4H_2PO_4$ at the Antiferroelectric Transition

Although ammonium dihydrogen phosphate, ADP, is isomorphous with KH_2PO_4, the dielectric behavior of ADP at the transition point at 148°K is quite different from that of KDP. In ADP it is the dielectric constant perpendicular to the c axis, ϵ_a, which shows the more pronounced anomaly at the transition. It was first suggested by Nagamiya [57] that this crystal may be antiferroelectric below the transition point, in a direction

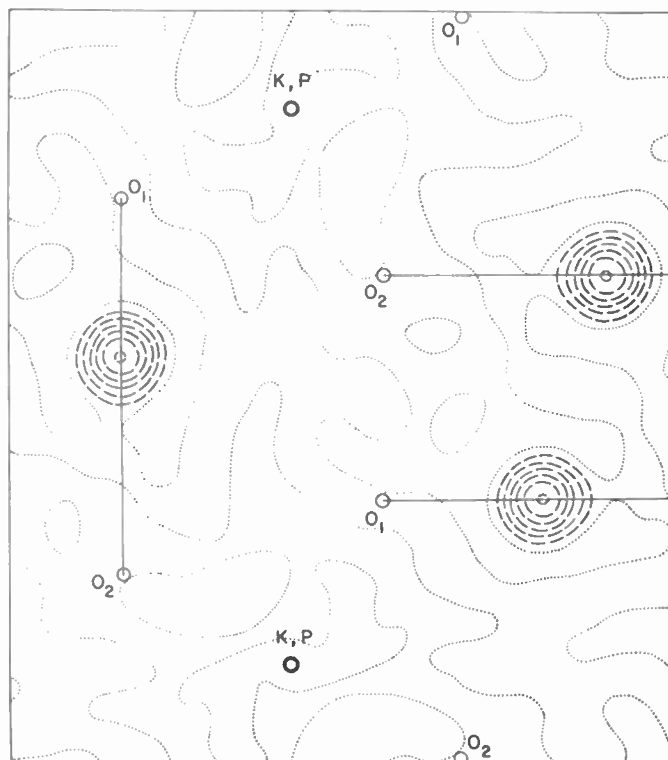


Fig. 51—Projection on (001) of the density of neutron scattering due to hydrogen atoms in KH_2PO_4 at −180°C. Broken lines are negative contours and the zero contour is shown dotted; no positive contours appear. The small circles mark positions of potassium, phosphorous and oxygen atoms (after Pease and Bacon [20]).

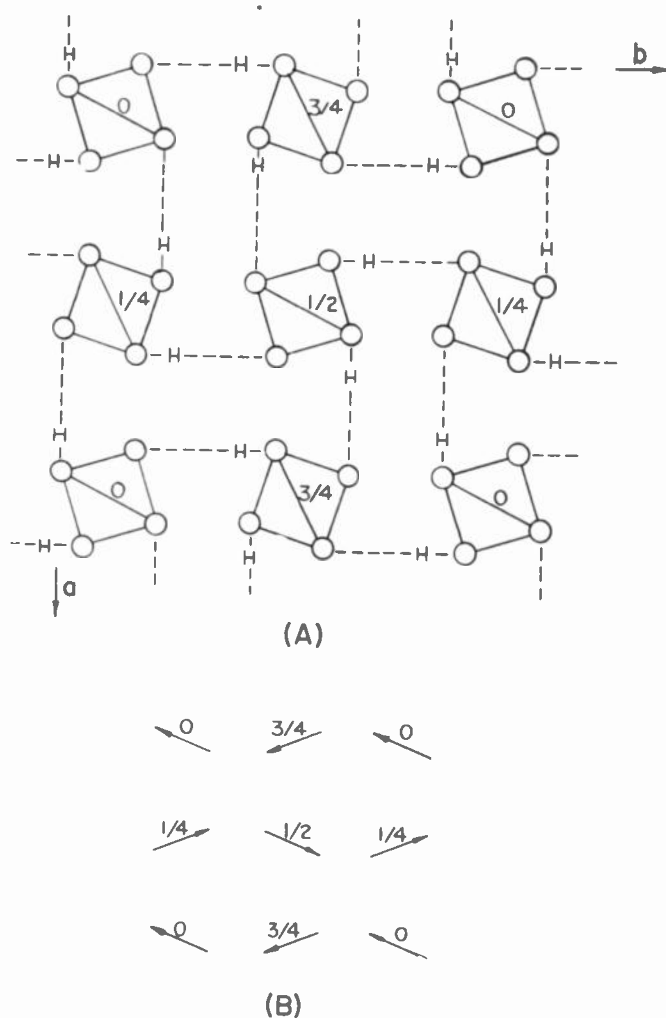


Fig. 52—(a) Proposed ordering of hydrogen in $NH_4H_2PO_4$. (b) Dipoles resulting from hydrogen ordering in $NH_4H_2PO_4$ (after Nagamiya [57]).

perpendicular to the *c* axis. An ordering of hydrogens was proposed, differing from that proposed by Slater for KDP in that the two hydrogens of H_2PO_4 groups become associated with one "upper" and one "lower" oxygen of the PO_4 tetrahedra rather than with two "upper" (or two "lower") oxygens. This ordering is shown in Fig. 52(a). Only the H_2PO_4 groups have been included in the figure, which is to be compared with Fig. 46 for KDP [the *a* axes in Fig. 46 are 45° from the *a* axis in Fig. 52(a); but the *c* axes are in the same direction for both]. The dipole arrangement expected from this ordering of hydrogens is represented schematically in Fig. 52(b) by a system of arrows, representing dipole moments perpendicular to the *c* axis. Note dipole moments in Fig. 52(a) are along the *b* axis and not the *a* axis.

Deuterated ADP, $ND_4D_2PO_4$, was reported by Matthias [122] to have a transition similar to that in ADP, but the transition temperature was shifted from 148°K to 242°K. An X-ray and dielectric study by Frazer, Keeling, and Pepinsky [58] supported Nagamiya's hypothesis of the antiferroelectric transition. Further support was provided in a subsequent X-ray and optical study by Wood *et al.* [123], and through a dielectric

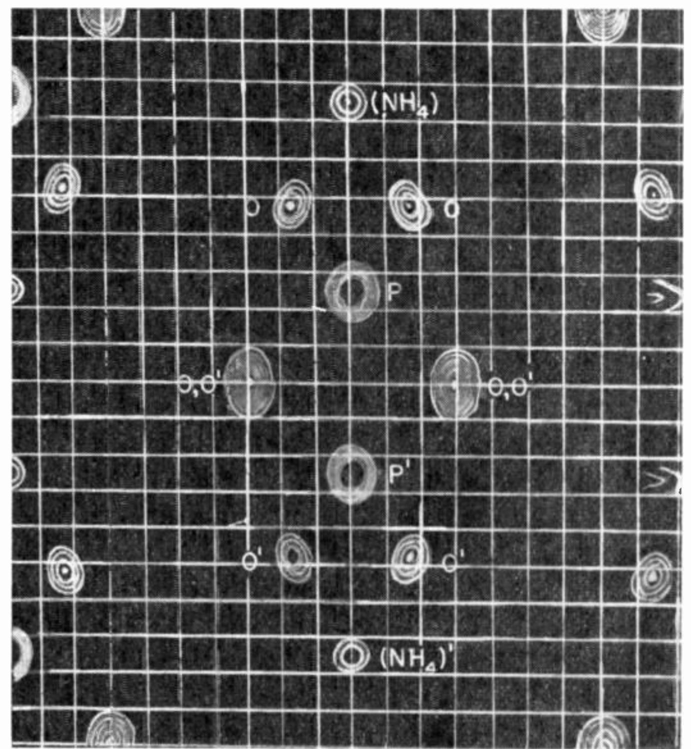


Fig. 53—Electron density projection of $NH_4H_2PO_4$ on (100) in $I\bar{4}d2$ at 153°K (after Keeling and Pepinsky [58]). A better reproduction of the above figure may be seen in a previously published paper by Keeling and Pepinsky [58].

study by Mason and Matthias [59]. The symmetry is orthorhombic below transition and tetragonal above.

Keeling and Pepinsky [58] have carried out an extensive X-ray study of the ADP transition, at 153°K and 143°K (5° on either side of the transition point). Results of this analysis are shown in Fig. 53 above and Figs. 54 and 55 opposite. Figs. 53 and 54 are projections of electron densities which correspond to Figs. 47 and 48 for KDP. Whereas the *c* axis is the polar axis in KDP below its Curie point, a twofold screw axis exists perpendicular to the *c* axis in ADP below its transition point; and this twofold (screw) axis causes the projection of Fig. 54 to be centrosymmetric. No net polarization can appear in any direction, because all three crystallographic axes are of the same twofold screw character.

A schematic of the dipole arrangement in this crystal, as viewed along the *a* axis, is shown in Fig. 55.

The mechanism of this transition is not well understood, in contrast to the case of KDP. Surely the H atoms of the NH_4 group are involved in production of the transition, which intervenes before a ferroelectric Curie point can be reached. Whereas N—H. . .O bond distances are all 2.88 A at 153°K, at 143°K four unequal distances appear from an N atom of an NH_4 group to neighboring oxygens:

$$N—H\ldots O_1 = 3.04 \text{ A},$$
$$N—H\ldots O_2 = 2.76 \text{ A},$$
$$N—H\ldots O_3 = 2.99 \text{ A},$$
$$N—H\ldots O_4 = 2.88 \text{ A}.$$

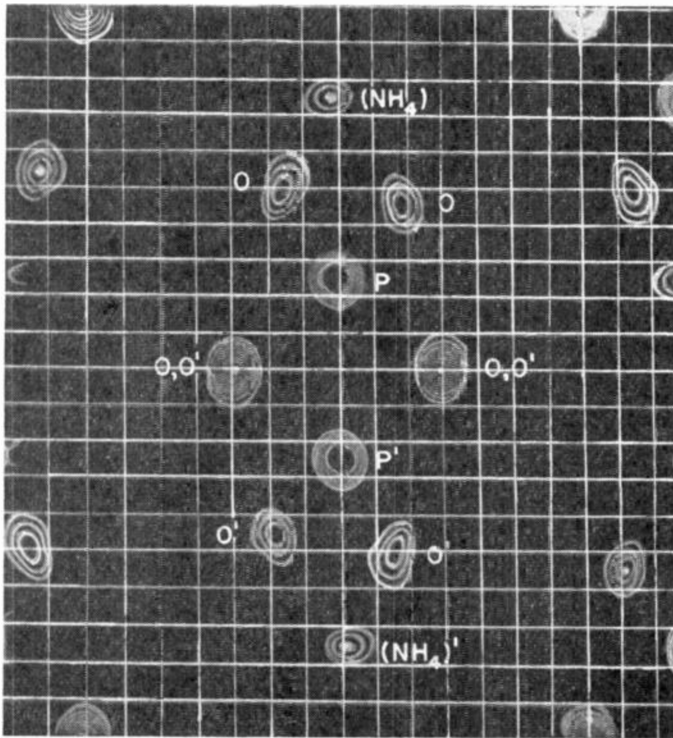


Fig. 54—Electron density projection of $NH_4H_2PO_4$ on (100) in $P2_12_12_1$ at 143°K (after Keeling and Pepinsky [58]). A better reproduction of the above figure may be seen in a previously published paper by Keeling and Pepinsky [58].

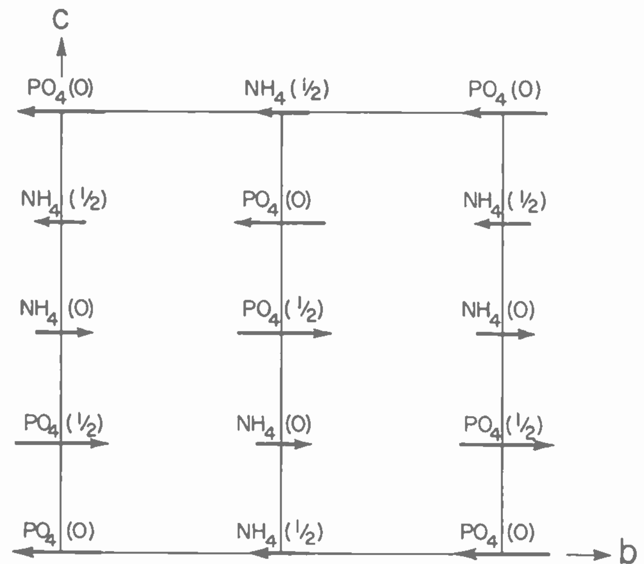


Fig. 55—Antiferroelectric dipole arrangement in $NH_4H_2PO_4$. Numbers in parentheses indicate positions along a axis (after Keeling and Pepinsky [58]).

This suggests different degrees of ordering of the H atoms in these four bonds. A neutron diffraction analysis of the crystal is in progress by Frazer, Danner, and Pepinsky at Brookhaven, in order to examine the H-bond system.

Macroscopic crystals of $NH_4H_2PO_4$ shatter at the transition point, 148°K. This would render a neutron diffraction analysis of the antiferroelectric crystal difficult if not impossible, since large crystals (1 to 2 mm in diameter, 1 to 2 cm in length) are required, with present neutron sources. An ingenious procedure which will permit examination of the anti-ferroelectric phase has been suggested by Frazer [124]. Substitution of deuterium for hydrogens in $NH_4H_2PO_4$ raises the transition point by 94° [122]. Substitution of As for P raises the transition point still further, since the transition temperature for $NH_4H_2AsO_4$ is 216°K. The transition temperature for $ND_4D_2AsO_4$ was found by Frazer to be 31°C. (See also Stephenson *et al.* [165]) Thus it is possible to grow $ND_4D_2AsO_4$ from solution at room temperature *in the antiferroelectric phase.* Such crystals are now in preparation for neutron examination.

G. X-Ray and Neutron Diffraction Studies of Rochelle Salt

The difficulties of structural analyses of the perovskite-type and alkali dihydrogen phosphate transitions pale in comparison with the analyses of the tartrates. The number of atoms whose coordinates and temperature parameters must be determined in the tartrate cells is very much greater than in the structures already described above. Whereas in $BaTiO_3$ or $PbTiO_3$ *three* positional parameters and at most nine anisotropic temperature parameters must be determined in three dimensions, in Rochelle Salt 112 positional parameters must be determined, and at least 12 atoms must have special temperature parameters determined for them (more atoms may need such special treatment, later), for only one two-dimensional projection! The Rochelle Salt structure is complex indeed in its nonpolar states, above the upper Curie point and below the lower one, but in the ferroelectric state the structural problems are very much greater. Furthermore, the ferroelectric transitions are more subtle, in the sense that thermal anomalies at the transitions are much less than in the case of perovskite-type or dihydrogen phosphate transitions. In lithium ammonium tartrate (LAT) the dielectric anomaly is also extraordinarily low.

It is actually only within the last several years that the theory of structure analysis refinement and computational methods have advanced to the point that a sufficiently accurate X-ray analysis of Rochelle Salt is really possible. The development of neutron diffraction methods, to the point where they can cope with such complex crystals, is also of very recent date.

As already stated, the structure of Rochelle Salt at 20°C was first reported by Beevers and Hughes [25] in 1941. The analysis was a tremendously difficult one at that time, and the achievements of these earlier analysts is not in any way to be deprecated. The difficulties and limitations of the early analysis were not entirely recognized by theorists desirous of a model to account for the dielectric behavior of this crystal, however; and thus great weight has been placed on features of the Beevers-Hughes structure which are not confirmed in a re-examination of the structure.

The structure of the tartrate ion (actually as determined from the analysis of LAT, to be described later) is shown in Fig. 56. This figure shows no hydrogen

atoms, although these are essential features of the molecule. The ion consists of a chain of four carbon atoms C_1 to C_4. C_1 and C_4 are in carboxyl (COOH) groups; C_2 and C_3 each have attached one H and one OH group. The hydroxyl groups are marked by O_5 and O_6 in the figure; and the carboxyl oxygens are O_1 to O_4. Rochelle Salt has the stoichiometric formula NaK(tartrate)$\cdot 4H_2O$; so there are four water molecules, one Na and one K for each tartrate ion.

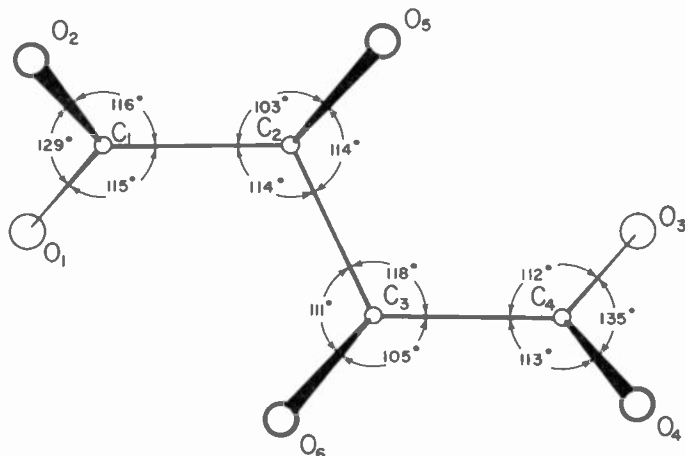


Fig. 56—Tartrate molecule.

Beevers and Hughes showed that there are four RS molecules per unit cell, the cell having dimensions $a=11.93$ A, $b=14.30$ A, $c=6.17$ A. The space group symmetry reported by them was $P2_12_12$, which is orthorhombic and nonpolar. Actually, such a symmetry excludes the possibility of ferroelectric activity. It was Jaffe [125] who first emphasized that in the ferroelectric phase Rochelle Salt must actually be monoclinic, space group $P2_1$. The monoclinic angle was found by Mueller [3b] to be 90°3′ at 11°C.

The reason for the orthorhombic assignment of Beevers and Hughes lay in the fact that X-ray evidence for the orthorhombic symmetry is overwhelming. The crystal is surely orthorhombic above the upper Curie point; this has been confirmed by a neutron study to be discussed below. In the ferroelectric phase two of the three 2_1 or 2 axes must disappear, if the crystal is to be polar; the remaining 2_1 axis must be along the monoclinic b direction.

A confusion in notation arises here. What should actually be monoclinic b has been denoted as an a direction by Beevers and Hughes. We will retain the Beevers-Hughes notation, to avoid confusion with established procedure for Rochelle Salt, and denote the true monoclinic b as the orthohombic a direction. For discussion of the confusion in axial designations, see Section VII, *C*.

In the monoclinic cell, the only missing reflections should be $(h, 0, 0)$ reflections with h odd. If $(0, k, 0)$ reflections with k odd are also missing, it would ordinarily be an indication of the existence of a 2_1 axis in the y direction. This would immediately require orthorhombic, not monoclinic, symmetry. Beevers and Hughes found the odd $(0, k, 0)$ reflections were indeed missing. This was the basis for their space-group assignment. An X-ray check of the $(0, k, 0)$ reflections at Penn State failed to reveal the odd-k terms, thus confirming the earlier observations.

The true symmetry was only revealed in a neutron study by Frazer and Pepinsky at Brookhaven [28]. Using either normal Rochelle Salt, or material in which all of the hydrogens *except those directly attached to* C_2 *or* C_3 were replaced by deuteriums (called deuterated RS, henceforth), the odd $(0, k, 0)$ neutron reflections *were* missing for the phases above the upper Curie point and below the lower one. This, plus the same conditions on $(h, 0, 0)$ reflections, confirmed the orthorhombic character of these nonferroelectric phases. The odd $(0, k, 0)$ were clearly *present*, however, in the ferroelectric region, although the odd $(h, 0, 0)$ were still missing—showing that the crystal had really become monoclinic there.

This observation was further tested by a technique introduced by Pepinsky [126]. Hydrogen nuclei have a scattering factor of -0.38×10^{-12} cm, and deuterium nuclei a scattering factor $+0.65\times10^{-12}$ cm. The replacement of H by D nuclei permits application of a very important technique in crystal analysis known as the *isomorphous replacement method* (cf. Robertson [127], pp. 147 *et seq.*; Pepinsky [111]). Pepinsky suggested the proper partial and statistical replacement of H by D atoms, so that the negative and positive neutron scattering powers of H and D nuclei cancel, and no net coherent scattering power remains for the replaceable H's. He applied this specific technique to Rochelle Salt, grown out of a proper heavy-light water solution. Crystals with the proper H to D ratio at these replaceable positions showed *no* odd $(0, k, 0)$ neutron reflections in the ferroelectric phase. This clearly established the fact that it was motion of H (or D) atoms which was largely responsible for the spontaneous polarization in Rochelle Salt. Since the H electrons are very difficult to find by means of X-rays, these results explained the difficulties encountered by Beevers and Hughes.

The structure which Beevers and Hughes evolved is shown in Fig. 57. This is a projection on the ab plane. Water molecules are marked with numbers $(H_2O)_7$ through $(H_2O)_{10}$. Other numbering is as in Fig. 56. The Fourier map from which this schematic is derived is not shown, but appears in Beevers and Hughes' paper. The direction of spontaneous polarization is along a. The Beevers-Hughes figures show no effects of polarization; they are based upon the assumption of nonpolar symmetry. Furthermore, the R factor for this projection is about 0.40, indicating a rather rough agreement between observed and calculated structure factors. With such agreement little reliance can be placed on the interatomic distances which the analysts derived.

WRH

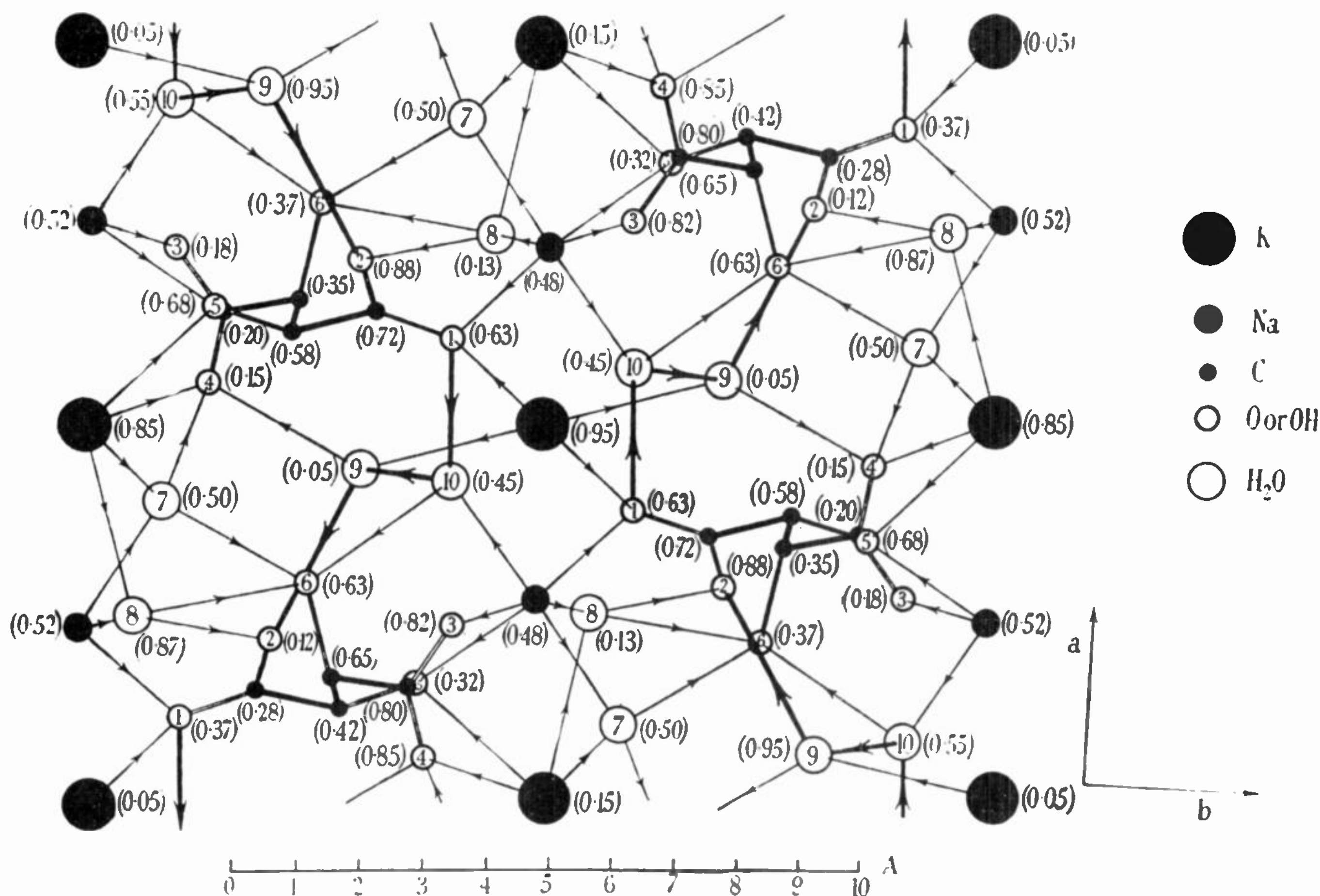


Fig. 57—Projection on (001) plane of the structure of Rochelle Salt (after Beevers and Hughes [25]).

The distance on which chief attention has been directed, following the Beevers-Hughes results, is that between O_1 and $(H_2O)_{10}$. This was stated to be 2.56 A—indicating a fairly strong hydrogen bond. Several investigators have based theories of ferroelectric activity in RS on this bond.

A refinement of the Rochelle Salt structure has been carried out in the Penn State X-ray laboratory [27, 128]. The R factor has been reduced to 0.14, which is very good for so complex a structure. This immediately revealed a number of interesting features. In the first place, the O_1—$(H_2O)_{10}$ bond is not short at all; it is one of the longer H bonds in the structure, and does not change measurably at the transitions. In the second place, when the orthorhombic symmetry is assumed, a good number of the electron density maxima appear to be *split*. This splitting is a consequence of extra, false symmetry. Restriction to monoclinic symmetry, and an extensive analysis of the resulting noncentrosymmetric projection, eliminates the splitting and shows the deviation in all but H atoms from the nonpolar structure.

Of particular interest, from X-ray results, is disposition of K atoms. The K's appear not in special positions found previously, but slightly off of these positions. Coordination of oxygens around these appears to change as crystal passes through two Curie points.

The X-ray study of the ferroelectric and the two nonpolar phases is still in progress. Meanwhile, a neutron single-crystal analysis of deuterated Rochelle Salt has been carried on at Brookhaven [28, 128]. This is the most ambitious neutron structural study yet attempted, and indeed compares in complexity with the largest structures yet successfully examined with X rays.

An intermediate stage of this neutron analysis is in Fig. 58, figure having been furnished through kindness of Dr. B. C. Frazer. Atoms can be identified by comparison with Fig. 57. Neutron projection shows all H or D atoms, and nature of H-bond system. Deviations from nonpolar structure are not in this figure.

The most significant polar deviations are found to be: displacements of K atoms from the special positions in Figs. 57 or 58; displacements of certain of the H_2O groups; and a large and very significant shift in the H atom on $(OH)_5$ (under C_4 in Fig. 57). The H atom on O_5 has two possible orientations. One of these orientations is adopted in one part of the structure, on one of the O_5 atoms on a pseudo-symmetrically-related mate to O_5 the H atom has another orientation. The result is a net hydrogen displacement of almost one Ångstrom unit, with the principal component parallel to the ferroelectric a axis. The "short" H bond between O_1 and $(H_2O)_{10}$ plays no part in the polarization.

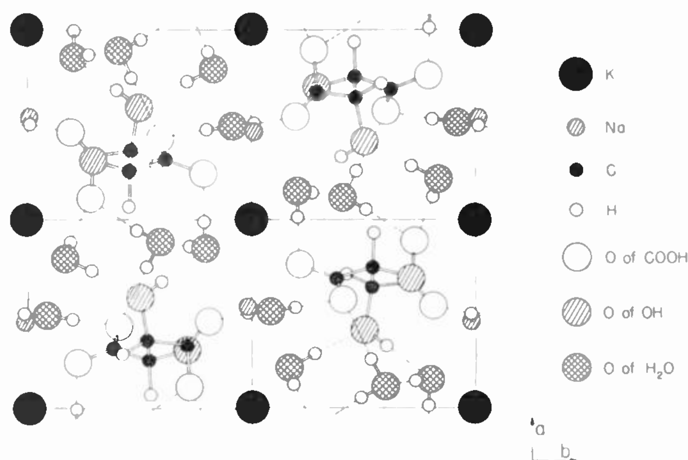


Fig. 58—Projection on (001) plane of the structure of Rochelle Salt obtained from neutron diffraction data. To facilitate comparison with Fig. 59, the oxygens belonging to OH groups, to H_2O, and to carboxyls, are here distinguished from each other (after Frazer *et al.* [28]).

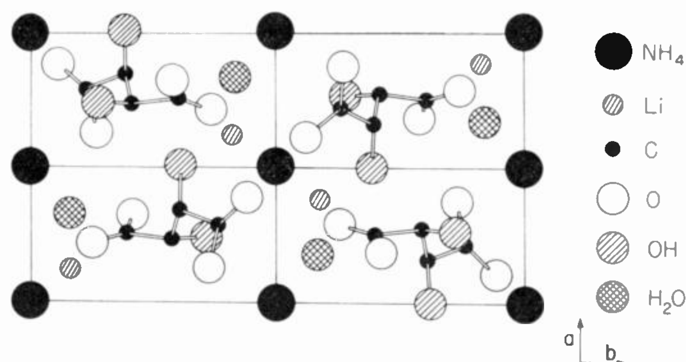


Fig. 59—Projection on (001) plane of the structure of lithium ammonium tartrate monohydrate obtained from X-ray diffraction data (after Vernon and Pepinsky [33]).

II. X-Ray Analysis of LAT at Room Temperature

The ferroelectric transition in lithium ammonium tartrate monohydrate, LAT, was discovered by Matthias and Hulm [31] and by Merz [32], following data and suggestions of Scholz [129], Jaffe [130], and Mason [131]. The dielectric anomaly at the Curie point is remarkable in that the peak value of the dielectric constant, ϵ_b, along the polar axis, is ~35—in contrast to the tremendous values of ϵ achieved along the ferroelectric axes of the materials already discussed in this section. Furthermore, lithium thallium tartrate, LTT, is isomorphous with LAT in the nonpolar phase, but has a ferroelectric transition at 10°K, *with the a axis as the ferroelectric direction;* and lithium rubidium tartrate, LRT, is also isomorphous with paraelectric LAT and LTT, but does not become ferroelectric at all.

LAT is of further interest in that the temperature-dependence of ϵ_b does not appear to follow a Curie-Weiss law.

The structures of LAT and LRT at room temperature have been determined by Vernon and Pepinsky [33], using X-ray diffraction. The space group for these is the same as for Rochelle Salt, $P2_12_12$, and the cell dimensions for all three crystals are, in A:

	LAT	LRT	RS
a	7.86	7.87	11.93
b	14.60	14.68	14.30
c	6.47	6.35	6.17

The similarities in *b* and *c* axes are striking.

The structures were solved by standard methods, and projections were obtained on the *ab* and *bc* planes. The projection of LAT on the *ab* plane is shown in Fig. 59. This is drawn to the same scale as the *ab* projection of Rochelle Salt shown in Fig. 58. The purpose of this is to facilitate comparison of the two structures—which, it is then clear, are remarkably similar. The steric arrangement of the tartrate molecules in the RS and LAT cells are essentially identical. The chief difference between the structures is the absence of the water molecules $(H_2O)_7$, $(H_2O)_9$ and $(H_2O)_{10}$ in LAT, and the consequent reduction in the separation of neighboring tartrate groups along the *a* direction. The similarity in *b* and *c* axial lengths to those in RS is understandable. The lithium atom appears to be in the neighborhood of O_1. It is to be noted that the X-ray analysis has not indicated H positions. It has shown, however, the probable existence of H bonding.

In discussing the structure, the numbering of Fig. 57 for Rochelle Salt is followed; and NH_4 ions at the projection corners and center are denoted as $(NH_4)_1$, while those on the centers of the edges of the cell projection are called $(NH_4)_2$. The coordination number for oxygens and hydroxyls around $(NH_4)_2$ ions is six: there are two O_1 and two O_4 carboxyl oxygens at 2.99 A, and two O_5 hydroxyl oxygens at 2.90 A. These distances suggest hydrogen bonding between the $(NH_4)_2$ nitrogen and the indicated oxygens. The coordination of oxygens around the $(NH_4)_1$ ions is four at H-bonded distances, and four at Van der Waals distances: two O_5 hydroxyls at 2.84 A (this is indicative of a strong H-bond); two O_1 carboxyl oxygens at 3.02 A, and two O_3 at 3.34 A; and two H_2O oxygens at 3.35A. The remarkable feature of the distribution of O's around $(NH_4)_1$ ions is the nearness of approach to one another of the O_3 carboxyl oxygens, which are related by the two-fold axis through an $(NH_4)_1$ ion. The O_3-O_3 separation is 2.62 A. Is a hydrogen present *between* the two O_3 oxygens? If so, it must be on the 2-fold axis, or statistically distributed about that axis; and which H is involved? Is this bond, and the oxygen coordination around $(NH_4)_1$, involved directly in the polarization? Or is the hydroxyl hydrogen at O_5 involved, as it certainly is in the case of Rochelle Salt?

A series of neutron diffraction observations by Pepinsky and Rock [101] at Brookhaven has clearly shown that the 2-fold screw axis along *a*, and hence also the 2-fold axis along *c*, are destroyed as the crystal passes from the paraelectric to ferroelectric states; and these axes reappear as the Curie point is passed from below. Neutron observations are being extended by the Penn State-Brookhaven group, using both normal and deuterated

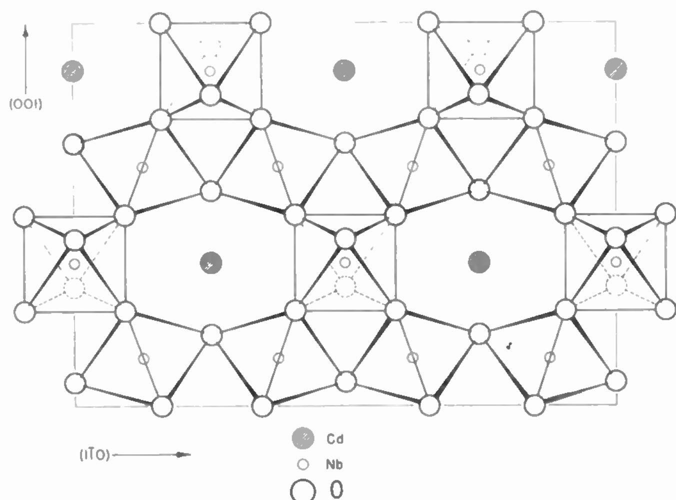


Fig. 60—NbO_6 octahedron configuration in $Cd_2Nb_2O_7$ projected on (110) plane (after Jona *et al.* [35]).

crystals, as well as crystals containing isotopically pure Li^6 or pure Li^7. Further speculations concerning the nature of the transitions in both LAT and LTT will best await completion of these studies.

I. Some Miscellaneous X-ray Studies

One other ferroelectric which is under examination is $Cd_2Nb_2O_7$. The crystal is cubic, with a Curie point at 185°K; and a further transition occurs just below 90°K. The paraelectric structure was first examined by Bystroem [105], using X-ray powder diffraction. A single-crystal room-temperature X-ray analysis by Jona, Shirane and Pepinsky [35] refined the Bystroem structure somewhat. The structure is shown projected on (110) in Fig. 60, reproduced from the paper by Jona *et al.*, and in the photograph of Fig. 61.

The nonpolar phase is cubic, with eight $Cd_2Nb_2O_7$ groups per cell, $a=10.372$ A, and space group Fd3m. The structure is a three-dimensional framework of corner-sharing NbO_6 octahedra. Cd ions and the extra O occupy otherwise open spaces in the framework. If the NbO_6 tetrahedra are *regular*, an oxygen parameter x should be 0.312. The value of this parameter was found to be 0.305 ± 0.003.

The symmetry change at the Curie point is not yet certain. Thus no proper basis for a model theory of the dielectric behavior is available.

A new structure type, represented by lead metaniobate, $PbNb_2O_6$, has been reported as ferroelectric by Goodman [157]. The Curie temperature is given as 570°C, with the Curie constant $C=2.95\times10^5$°C. The room-temperature symmetry is orthorhombic, with cell dimensions $a\simeq25$ A, $b\simeq25$ A, $c=7.0$ A, and 40 molecular weight units of $PbNb_2O_6$ per cell. A structure analysis is likely to prove quite difficult.

A preliminary X-ray examination has been conducted on $(NH_{4\,0.2}, K_{0.8})NaC_4H_4O_6\cdot4H_2O$, in the nonpolar phase [102]. The results suggest *preferential* replacement by NH_4 ions of the K ions at the corners and center of the

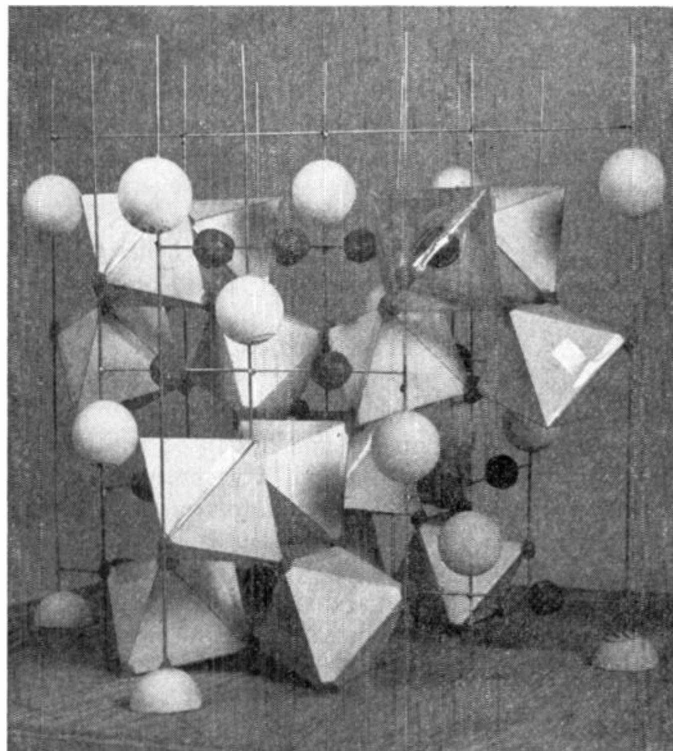


Fig. 61—Model of $Cd_2Nb_2O_7$. The paper octahedra represent NbO_6 groups: Nb is at the center of each octahedron and oxygens are at the six corners. The large white balls represent the seventh oxygen atoms, the small dark balls represent cadmium atoms.

cell of Fig. 57. This problem is much more amenable to neutron diffraction than to X-ray study. The structural nature of the ferroelectric transition is not at all understood.

VII. Domain Structure of Ferroelectric Crystals

A. Origins of Domain Structure

It is well known that Weiss introduced the concept of *domains* in ferromagnetic materials to account for the fact that such materials may show no macroscopic magnetization in the absence of an applied external field, and yet may be magnetized to a very high degree by application of a relatively very small external field. Within each domain there is a large spontaneous magnetization, but the direction of this magnetization may vary from domain to domain in such a way that the specimen as a whole may show no net macroscopic magnetization.

The fact that domains should be expected in ferromagnetic crystals, even in single crystals, was shown by Landau and Lifshitz [132] on energetic grounds. When spontaneously magnetized regions are all aligned, a demagnetizing field is produced. The interaction between the demagnetizing field and the magnetization of the crystal involves a large magnetostatic energy. The formation of domains causes the surface of the crystal to be checkered with north pole regions and south pole regions, and, in some instances, with regions of no polarity. As a result, demagnetizing field will be reduced, and with it magnetostatic interaction energy.

The strong analogies between ferromagnetic and ferroelectric behaviors extend somewhat to domain behavior; but there are qualitative as well as quantitative differences in the effects. The lattice distortions accompanying ferroelectric transitions are very much greater than those consequent upon the development of spontaneous magnetization. Therefore strain effects may play a larger part in ferroelectric domain behavior than in the magnetic case. Furthermore, the existence of free charges in the electric case, in contrast to the nonexistence of free poles in the magnetic case, plays an important role in ferroelectricity in neutralizing surface polarity and thereby decreasing the depolarizing field.

The energetic arguments first advanced by Landau and Lifshitz for ferromagnetism have been applied to ferroelectrics. We will follow here a treatment by Kaenzig and co-workers [133]. Consider an isolated, nonconducting ferroelectric crystal, polarized uniformly. The crystal is subjected to a depolarizing electric field, arising from charges produced on the surfaces perpendicular to the polarization direction. The energy of this depolarizing field can be written:

$$U_e = \tfrac{1}{2}\Gamma P^2 V, \qquad (31)$$

where P is the polarization, V the crystal volume, and Γ a depolarization factor. (In the simple case of a spherical single-domain crystal in a vacuum, $\Gamma = 4\pi/3$.) The depolarizing field renders the assumed uniformly-polarized state unstable, unless a large external field is applied. Introduction of domains, with adjacent domains having anti-parallel directions of spontaneous polarization, greatly reduces the depolarizing field. The number of domains per unit volume is limited by the fact that energy is stored in the boundary layers or "*walls*" between the domains. The increase of this wall energy must eventually balance the decrease of energy of the depolarizing field, as the number of domains (and hence the total wall surface) is increased. For a given temperature, then, there is an equilibrium distribution of domains within a macroscopically unstrained, isolated, nonconducting ferroelectric crystal.

The domain structures of ferroelectric crystals can readily be studied in polarized light under a microscope. This is possible because the materials are transparent and the refractive indices for polarized light vary with crystallographic direction. Thus, in a ferroelectric crystal, it is possible to relate the direction of spontaneous polarization, which always has a specific orientation with respect to crystallographic axes, to the appearance of the crystal in polarized light.

The optical properties of a crystal are generally specified by describing an ellipsoidal surface, called the *indicatrix*. This surface has the following significance. The radius vector from the origin of the indicatrix to the surface of the indicatrix is directly proportional to the refractive index, in the crystal, of a light wave whose electric vector is vibrating parallel to the radius vector. Except for triclinic crystals, one or more of the axes of the indicatrix lie along crystallographic axes.

The orientation of the indicatrix can be found for any crystal under the polarizing microscope. Since the orientation of the indicatrix with respect to the crystallographic axes can be experimentally determined for any given material, the orientation of ferroelectric domains can be established in polarized light. This is a simple and convenient type of measurement which can be performed under widely-varied conditions of temperature, electric field, etc.

B. Domain Structure in KDP

We consider the domain structure in dihydrogen phosphate family of ferroelectric crystals first, because in this group the relation between crystallographic directions and optical properties is easily visualized. The prototype of the group is the well-studied KH_2PO_4, which is ferroelectric at temperatures lower than 123°K = −150°C. The existence of a domain structure in this region had been conclusively demonstrated by earlier X-ray studies (De Quervain [12], Ubbelohde and Woodward [134]), and by the study of the change in double refraction in the presence of an electric field (Zwicker and Scherrer [135]). The domain structure was directly observed by De Quervain and Zwicker [136] and more recently by Mitsui and Furuichi [23]. Some experimental difficulties arise, in the case of KH_2PO_4, due to the fact that the ferroelectric phase is limited to low temperatures.

Above 123°K, KH_2PO_4 belongs to the tetragonal symmetry class $D_{2d}^{12} - I\bar{4}2d$. Because of the tetragonality, it is optically uniaxial, and the c axis is the optic axis. This means that a plate cut perpendicular to the c axis (a so-called c plate), if observed on the polarizing microscope between crossed nicols, reveals no extinction position. The indicatrix is an ellipsoid with rotational symmetry around the crystallographic c axis (which coincides with the Z axis of the indicatrix). Below the Curie point, the crystal polarizes spontaneously along the c direction. The phase change is associated with a spontaneous shear x_y^0, and the crystal acquires orthorhombic symmetry C_{2v}^{19}-Fdd. The relative orientation of the tetragonal axes and that of the orthorhombic axes is shown schematically in Fig. 62. The crystal becomes optically biaxial: the orthorhombic indicatrix no longer has rotational symmetry. Its axes lie parallel to the crystallographic (orthorhombic) axes. This implies that a c plate, which showed no birefringence above the Curie point, will show birefringence below.

The expected domain structure is shown schematically (and exaggerated) in Fig. 63, which presents the relative orientation of the orthorhombic axes of two adjacent domains I and II. Since the spontaneous shear x_y^0 is about 27′ at 103°K (De Quervain [12]), the extinction positions of domain I will differ from the extinction positions of domain II by 27′ at this temperature. This difference, although small, is of the order of magnitude

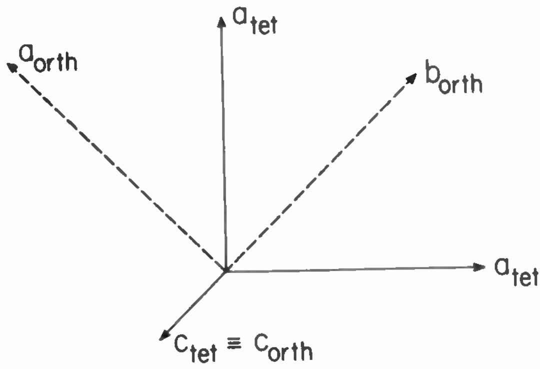


Fig. 62—Orientation of the crystallographic axes of tetragonal KH_2PO_4 (T>123°K) relative to the crystallographic axes of orthorhombic KH_2PO_4 (T<123°K) (schematic).

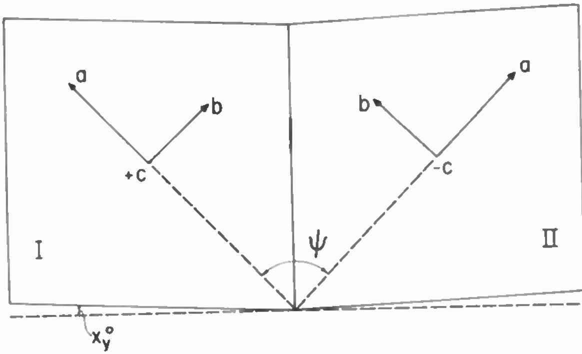


Fig. 63—Schematic domain structure of KH_2PO_4. The spontaneous shear x_y is about 27′ at 100°K. The angle ψ is therefore equal to 90°−27′.

of detectable angles within the experimental technique of the polarizing microscope. Mitsui and Furuichi were not able to measure this difference. This is very probably due to the unavoidable complications of the measurements at the low temperatures. These authors were able, however, to confirm that a *c* plate is birefringent in the Curie region (see Fig. 64), and that the domain boundaries are parallel to the original tetragonal *a* axis. The width of the domains appears to be a few hundredths of a millimeter.

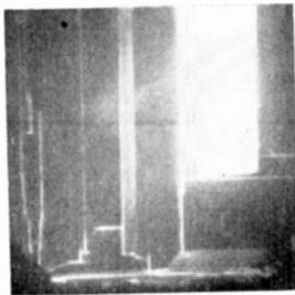


Fig. 64—*c* plate of KH_2PO_4 observed between crossed Nicols below the Curie point (after Mitsui and Furuichi [23]).

A plate cut perpendicularly to the direction of the tetragonal *a* axis (*a* plate) must actually change in optical properties at the Curie point; but this change is very small, since it involves only a very slight change in the axial lengths of the indicatrix.

C. Domains in Rochelle Salt

In the case of Rochelle Salt, too, the existence of domains has been conclusively proved by means of X-ray investigations (Miyake [137], Staub [138]). The direct observation of the domain structure was accomplished by Mitsui and Furuichi [23], by means of the polarizing microscope. Above the upper Curie point at +24°C, Rochelle Salt belongs to the orthorhombic symmetry class D_2^3-$P2_12_12$, and is, therefore optically biaxial. The acute bisectrix is the *a* axis, the plane of the optic axis is the (010) plane, and the crystal is optically positive. This fixes the position of the axes of the indicatrix with respect to the crystallographic axis, as shown in Fig. 65: $a_{orth} = Z_{orth}$, $b_{orth} = Y_{orth}$, $c_{orth} = X_{orth}$, where X_{orth}, Y_{orth}, Z_{orth} are the indicatrix axes. An *a* plate therefore extinguishes at 90°, and the extinction positions correspond to the directions of the axes X_{orth} and Y_{orth} of the indicatrix.

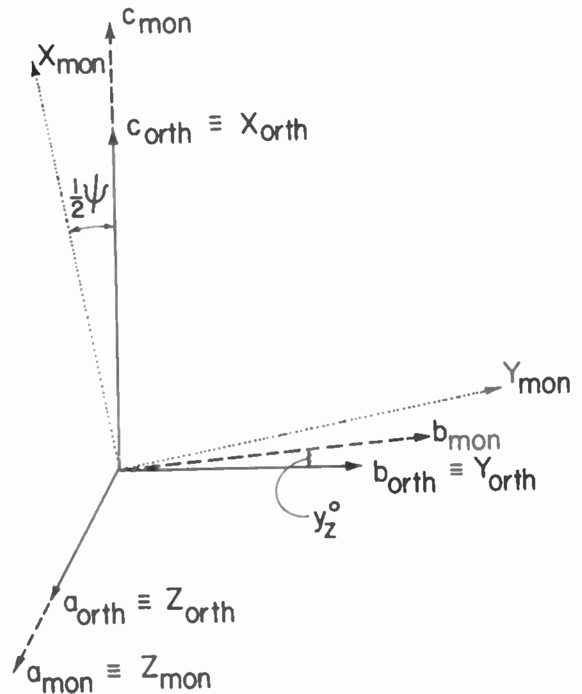


Fig. 65—Rochelle Salt (schematic): orientation of the crystallographic axes (a_{orth}, b_{orth}, c_{orth}) and the axes of the orthorhombic indicatrix (X_{orth}, Y_{orth}, Z_{orth}) (for temperatures T>24°C), relative to monoclinic crystallographic axes (a_{mon}, b_{mon}, c_{mon}) and the axes of the monoclinic indicatrix (X_{mon}, Y_{mon}, Z_{mon}) (for temperatures 24°C>T>−18°C). Spontaneous shear y_z^0 is about 3′ at 11°C. (For contradiction with crystallographic convention, see text).

Between +24°C and −18°C a Rochelle Salt crystal polarizes spontaneously in the direction of the orthorhombic *a* axis. The symmetry becomes monoclinic, C_2^2-$P2_1$, owing to the spontaneous shear y_z^0. This means that in a single domain crystal one of the monoclinic axes deviates from the direction of the original orthorhombic axis by an angle equal to y_z^0. This deviation is very small. Mueller [3(b)] made a direct measurement of variation with temperature of the angle between the original *b* and *a* faces of a Rochelle Salt block, and found that the deviation from 90° is about 3′ at 11°C.

It is customary in crystallography to designate the *unique* axis of a monoclinic crystal as the b axis; thus in Rochelle Salt what *was* the orthorhombic a axis becomes the monoclinic b, while the monoclinic a lies perpendicular to the latter and deviates from the orthorhombic b axis by the small angle associated with y_z^0. Because of previous usage, and in order to avoid confusion in the reading of earlier papers, the crystallographic axes of monoclinic Rochelle Salt, parallel or almost parallel to the orthorhombic axes a_{orth}, b_{orth}, c_{orth}, will simply be called a_{mon}, b_{mon}, c_{mon} respectively. Now, in monoclinic Rochelle Salt one axis of the indicatrix coincides with the crystallographic a axis; but the other two, perpendicular to each other, do not necessarily coincide with the crystallographic b and c axes (see Fig. 65). It is their direction which determines the extinction positions of a monoclinic a plate; their deviation from the original orthorhombic axes will determine the difference in the extinction positions of the orthorhombic and monoclinic phases. This is obtainable from the difference in extinction positions of two adjacent domains, which is twice deviation angle from original orthorhombic axes.

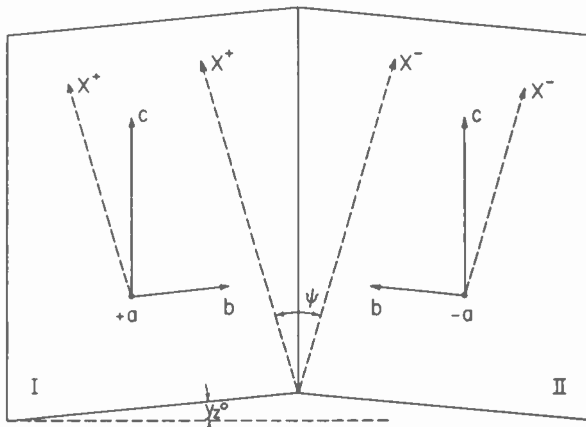


Fig. 66—Schematic domain structure of Rochelle Salt. Only the x axis of the monoclinic indicatrix is shown. The angle ψ, which is the angle between the extinction positions of the two domains I and II, is about 2.5° at 17°C.

Fig. 66 shows the situation for two adjacent domains I and II. It appears that the angle between the X axes of the indicatrices of two anti-parallel domains is large enough to permit the observation of a difference in the extinction positions of the two domains. This angle is about 2.5° at 17°C [23]. Fig. 67 shows a photograph of the observed domain structure. The patterns disappear above 24°C and below −18°C.

In b and c plates no domain structure is observable.

The domains are mostly slabs perpendicular to the orthorhombic b axis, with widths ranging from 5×10^{-1} to 1×10^{-4} cm. They are called c domains, since they are parallel to the orthorhombic c direction. Less often domains are observed which are parallel to the original orthorhombic b axis. These are called b domains.

Since the crystals are strongly piezoelectric, it is to be expected that the domain structure can be changed by means of an externally applied stress. In fact, when a shearing stress Y_z is applied, it is found that one set of domains narrows while the other widens. Upon applying a sufficiently strong stress, the crystal can be changed into a single domain. Moreover, since both the spontaneous deformation and polarizations are functions of temperature, a change in the domain structure can be brought about by temperature variation.

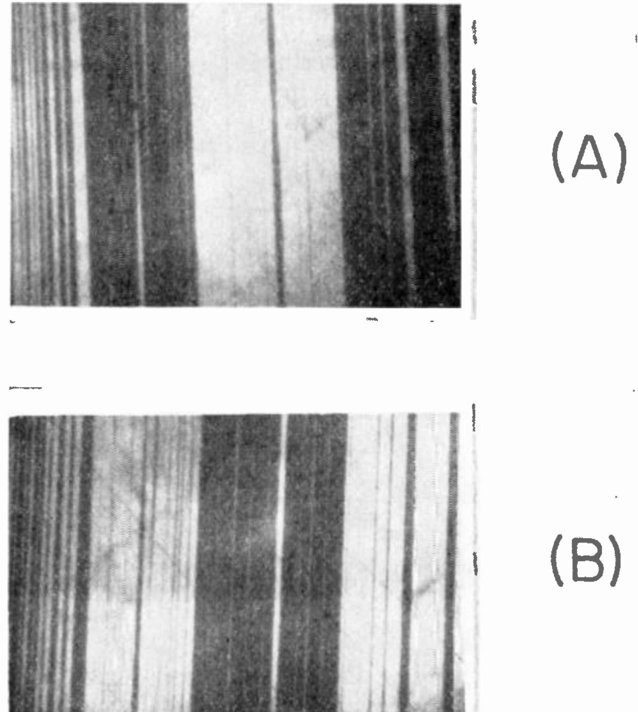


Fig. 67—a plate of Rochelle Salt observed between crossed Nicols (×50). The c axis is rotated from the plane of a Nicol: (A) to the left, (B) to the right (after Mitsui and Furuichi [23]).

An interesting correlation has been found between the thickness D and the width W of the domains: the domain width W is almost perfectly proportional to the square root of the thickness D:

$$W = k\sqrt{D}, \tag{32}$$

where the constant k is of the order of 10^{-3} cm$^{1/2}$. This functional dependence can also be derived theoretically, as shown by Mitsui and Furuichi, by computing the domain width which makes the internal energy of the crystal a minimum, and taking into account the contribution of the electrostatic energy. The latter plays an important role, together with the internal stress and piezoelectric effect, in determining the domain structure —especially when the crystal polarizes more or less abruptly and the surface charge is not immediately neutralized. If the domain structure formed in this way persists even after the charge is neutralized, a relationship between the domain width and the thickness of the crystal might be expected. Moreover, if the domain wall moves after neutralization, a charge may develop on the surfaces across which the wall has moved, which is pro-

portional to the spontaneous polarization. The charge is the source of an electric restoring force, which acts on the wall and can be removed only upon neutralization of the charge itself. In fact, it has been observed that displacements of the domain walls caused, for example, by thermal motions, occur much more freely when the crystal is immersed in an electrolytic solution, where the surface charge and thus the restoring force cannot develop, than when the crystal surface is dry.

All these experimental results are a confirmation of the fact, to be expected theoretically, that the domain configuration of a ferroelectric crystal cannot be determined from considerations of static equilibrium only, as is the case for ferromagnetic domains, but that other factors, such as conductivity of the crystal, and time, must be taken into account. In the case of ferromagnetics, there is no possibility of neutralizing surface poles, because of the nonexistence of free magnetic poles, and in fact the relationship (32) for the width of domains was first obtained for *ferromagnetic* domains.

D. Domains in Barium Titinate

Above 120°C barium titanate belongs to the cubic crystal class O_h^1-Pm3m; between 120°C and 0°C it has the tetragonal symmetry C_{4v}^1-P4mm, the c axis being slightly longer than the a axis ($c/a = 1.01$ at 20°C). The crystal polarizes spontaneously in the directions of any one of the former cube edges, and this behavior often produces complicated domain structures. The latter have been studied by a number of authors: originally by Blattner, Kaenzig, Merz, and Sutter [139]; later, in greater detail, by Forsbergh [62] and by Kay and Vousden [64]; and most recently by Merz [42] and by Little [43]. Possible domain configurations of a $BaTiO_3$ plate at room temperature can be visualized, schematically, with the help of Fig. 68. The figure represents a cube of $BaTiO_3$ consisting of domains I and II. The direction of the spontaneous polarization, which coincides with the c axis, is indicated on each one of the three visible cube faces.

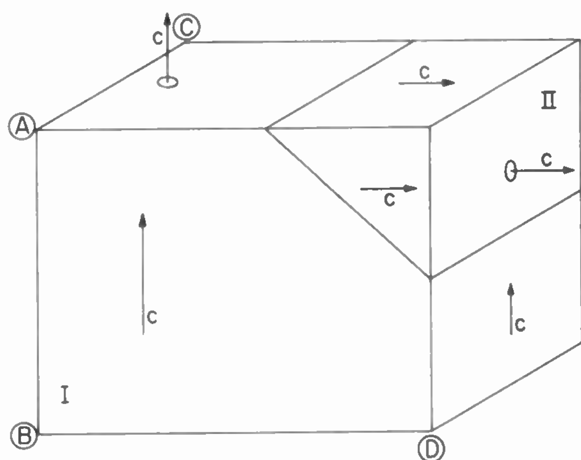


Fig. 68—Schematic domain structure of a tetragonal $BaTiO_3$ crystal with 90° domain walls.

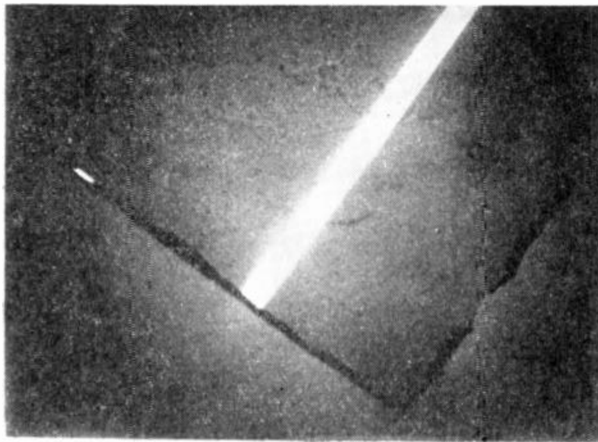
Suppose that we cut a thin plate out of this cube, perpendicular to the direction AB and close to the edge AC. We then observe this plate under the polarizing microscope, between crossed nicols, for light traveling in the direction AB. Domain I is oriented in such a way as to have its optic (c) axis parallel to the light, and is called a c domain; it will always look dark upon rotation of the plate on the microscope stage. The optic axis of domain II, on the other hand, lies *in* the plane of the plate: this domain will therefore have two extinction positions, at 90° from each other. This domain is called an a domain, since one of the a axes is perpendicular to the plate surface. It is very easy, then, to distinguish between the two domains on the polarizing microscope. We obtain the same situation by slicing the cube perpendicularly to the direction BD and close to D, and observing it between crossed nicols for light traveling along BD. The roles of the domains are interchanged (domain II remains dark, domain I is bright except at the two extinction positions), but the over-all picture remains the same. The boundary between an a and a c domain is called a 90° wall, because the two domains are polarized at 90° to each other. The 90° wall is a crystallographic (110) plane, as required by the condition that no surface charges are to be found on the wall (the components of polarization perpendicular to the wall are of equal magnitude in the two domains and have a head-to-tail arrangement). Fig. 69 (next page) is a photo of a $BaTiO_3$ plate containing a 90° wall.

Suppose now that we cut a plate from the cube in Fig. 68, perpendicular to the direction AC and close to the face ABD, and observe it between crossed nicols for light traveling in the direction AC. In this case the wall between domain I and domain II is still a 90° wall, since the domains are polarized at 90° to each other. However, now we are looking at two a domains. Upon rotation of the plate on the microscope stage, both domains will appear bright except at the two extinction positions. The extinction positions of the two domains deviate slightly from each other according to the tetragonal distortion. This angular deviation is

$$2 \text{ arc tangent } c/a \approx 36',$$

since $c/a = 1.01$ at room temperature. The wall between these domains can be seen directly because of the total reflection of the light on the boundary; and, moreover, the direction of polarization of the two domains can be determined by using a quarter-wave plate. Fig. 70 (next page) shows a photograph of a $BaTiO_3$ crystal containing many 90° walls between a domains.

In addition to the walls just described, there is another important type of domain wall in $BaTiO_3$: 180° walls, which separate domains with anti-parallel polarization. Fig. 71 (next page) shows this schematically. These walls are not visible under the polarizing microscope unless special conditions are applied. To see them it is necessary to apply a dc field in the direction perpendicular to the polarization of the domains. The effect of the field



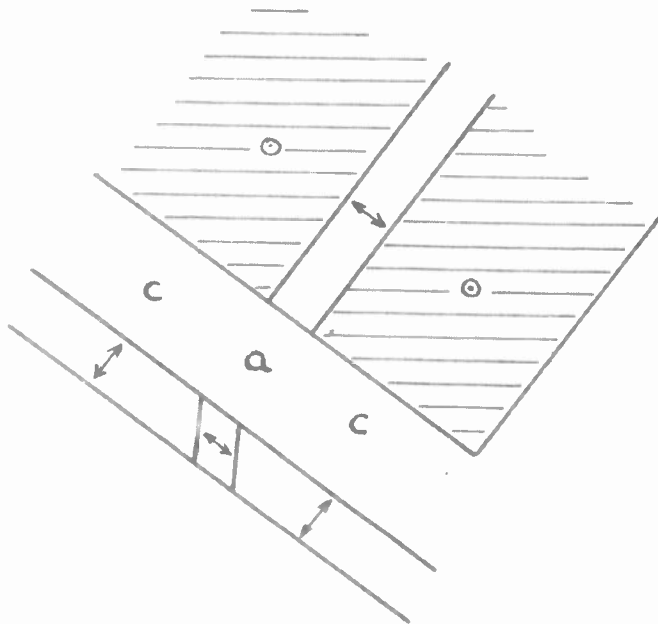
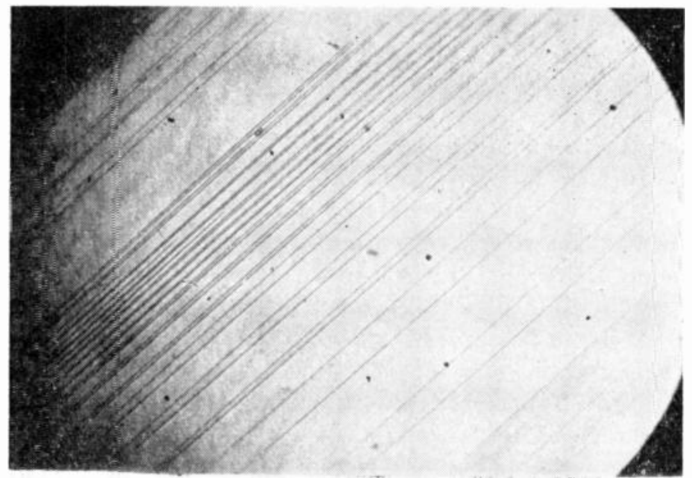


Fig. 69—Photograph and schematic sketch of 90° walls between an *a* domain and two *c* domains in $BaTiO_3$ (×50) (after Merz [42]).



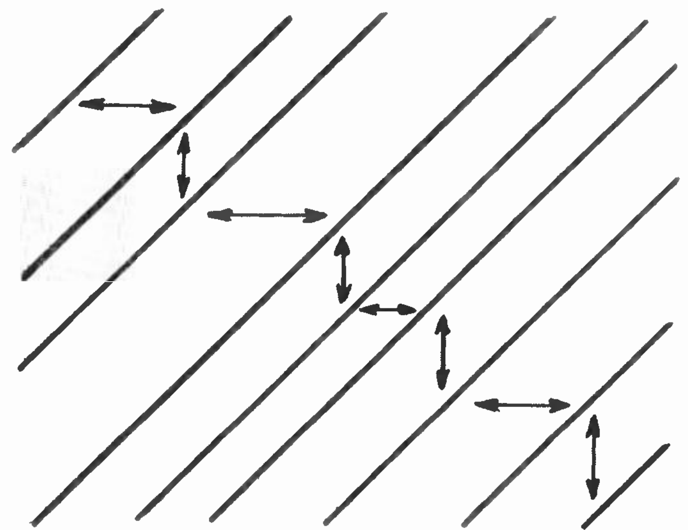


Fig. 70—90° walls between *a* domains in $BaTiO_3$ (×50) (after Merz [42]).

is to rotate the polar axes of the two domains slightly in opposite direction, as indicated in Fig. 71 by the dotted lines. In this way the internal strains are increased, and the extinction positions of the two domains deviate from each other by a small angle. With this technique Merz [42] was able to show that there is a fine-structure in the domains, in the sense that every supposedly single domain consists in reality of many domains with anti-parallel polarization. The width of these domains is of the order of 10^{-4} cm. Fig. 72 shows a photograph of the edge of a $BaTiO_3$ plate strained by a field. Fig. 73 shows both 90° and 180° walls: the oblique lines are 90° walls, while the zigzag lines represent 180° walls.

It is now clear that the arrangement of the domains and the orientation of the electric polarization in each domain can be easily determined by means of a polarizing microscope. What the usual optical methods cannot do is to provide information as to the positive or negative directions of the polarization, without the aid of externally applied fields or stresses. Hooton and Merz

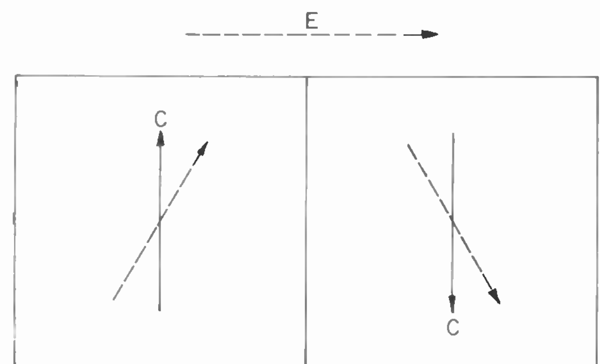


Fig. 71—180° domain wall, between anti-parallel domains (schematic). An electric field *E* applied perpendicular to the *c* axis rotates the latter in opposite directions (dotted lines, exaggerated).

[140] have developed a beautiful technique, by which it is possible to distinguish easily and rapidly between the positive and negative ends of the electric dipoles. They etch the $BaTiO_3$ crystals for a few minutes in concen-

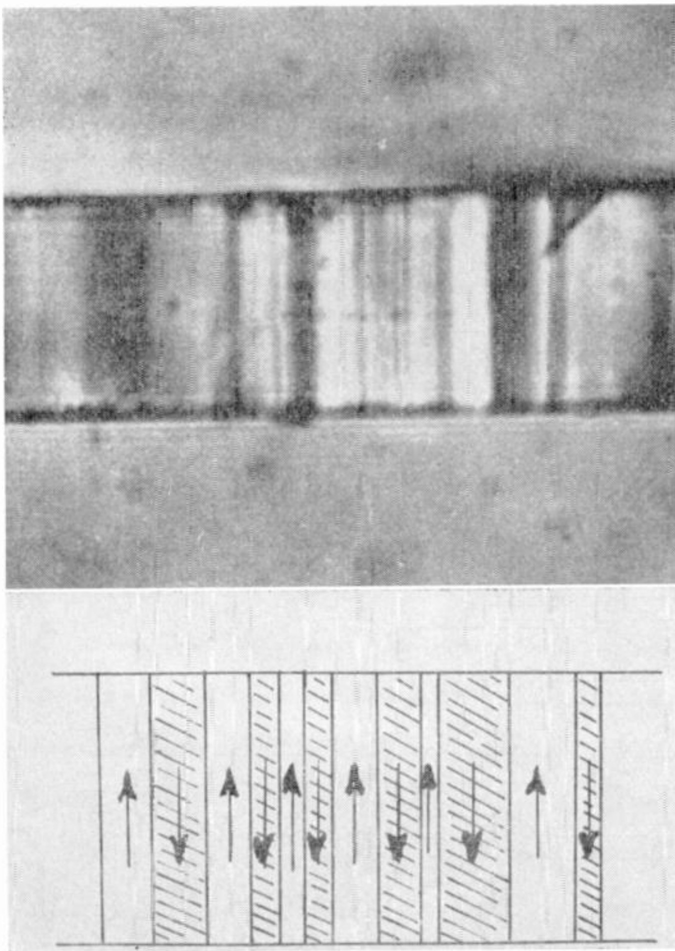
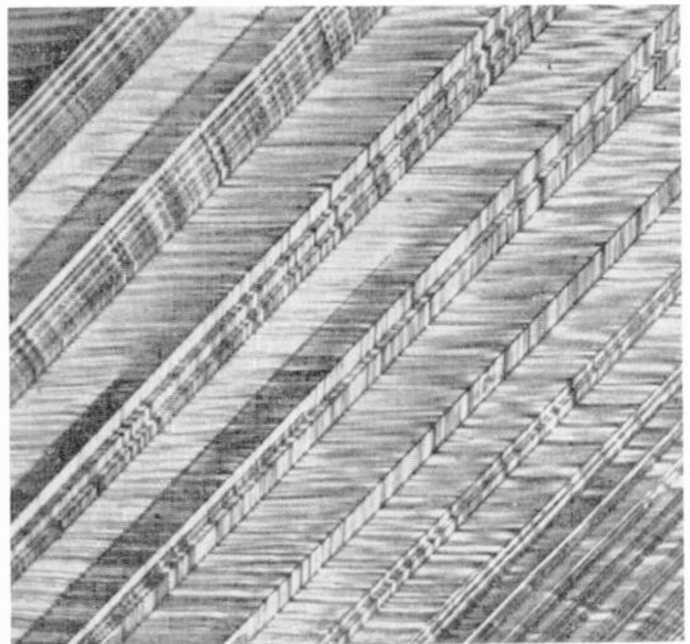


Fig. 72—180° walls in $BaTiO_3$: edge of a c plate strained by an applied field (after Merz [42]).



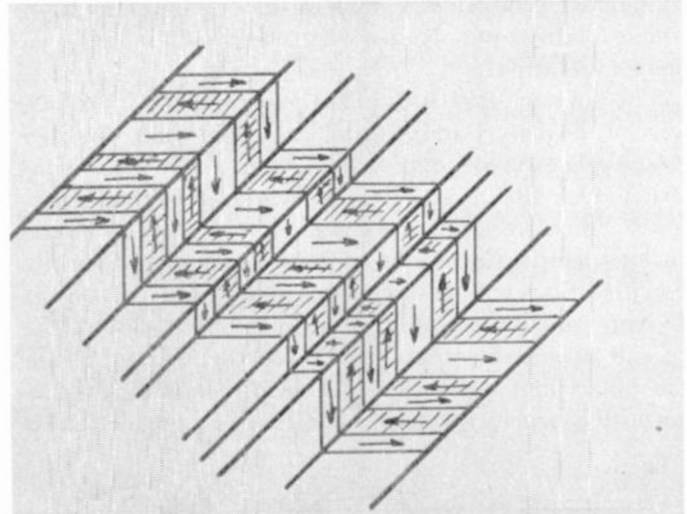


Fig 73—180° and 90° walls in a $BaTiO_3$ crystal surface (after Merz [42]).

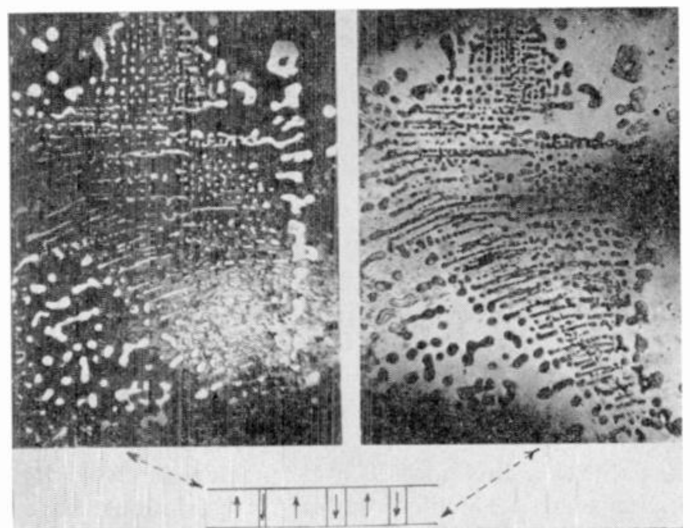


Fig. 74—Photograph of the etched top and bottom surfaces of a poorly poled $BaTiO_3$ crystal showing many anti-parallel polarized domains. (One print is reversed) (after Hooton and Merz [140]).

trated hydrochloric acid at room temperature. When the washed crystals are observed by reflected light under the microscope, it is seen that the etch rate is different for the positive (head) ends and the negative (tail) ends of the dipoles. A dipole head etches rapidly, while the tail etches very slowly if at all. Fig. 74 shows a photograph of the two surfaces of a poorly poled $BaTiO_3$ crystal plate, as it appears after the etching treatment. The dark areas correspond to regions where the positive ends of the dipoles are at the surface, the bright areas to negative ends of the dipoles. The islands, which are anti-parallel regions, go right through the crystal, as shown by the fact that the photograph of the lower surface is a perfect mirror image of the upper surface. Moreover a domains can easily be distinguished from anti-parallel c domains, since the etch rate perpendicular to the dipole direction is intermediate between the rapid rate for the dipole head and the slow rate for the dipole tail; the a domains appear greyish.

By means of these etch patterns, Hooton and Merz were able to prove that strong coupling of the dipoles takes place in the forward direction. The head to tail arrangement at a 90° wall required by the absence of space charges on such a wall, is the cause for the "reflection" of an anti-parallel surface region through an a

domain, so that a mirror image of the same region can occasionally be observed on the surface, on the other side of an a domain. In addition to this, there is also a coupling through c domains, which the authors attribute to the strong piezoelectric effect of tetragonal $BaTiO_3$ and electrostatic interactions. When wedge-shaped a domains are introduced into a c domain plate the crystal is strained in a complicated way; in particular, a shear stress acts along the 90° wall, which induces a polarization through the piezoelectric effect. This induced polarization propagates through the crystal by electrostatic coupling and can, in this way, influence the orientation of the polar axis in regions located further away. For the details of this mechanism the reader is referred to the original paper. Whatever the explanation of the coupling through the domains might be, it should be emphasized that the strains introduced into the crystal either by a wedges, by mechanical or thermal treatment, or by local imperfections, play an essential part, through the piezoelectric coupling, in determining the domain configuration of other regions of the crystal. This is, of course, of great importance when the crystal is used as a memory element.

The domain structures in the lower phases of $BaTiO_3$ were studied by Forsbergh [62] and Kay and Vousden [64] by means of microscope stages especially constructed for the observation of crystals at low temperatures. Since we know the symmetry of $BaTiO_3$ in the various temperature ranges, we also know what extinction directions will be shown by a single domain block at a given temperature, and for a given orientation of the crystal axes with respect to the direction of the light in the polarizing microscope. These directions are schematically represented in Fig. 75. The extinction is

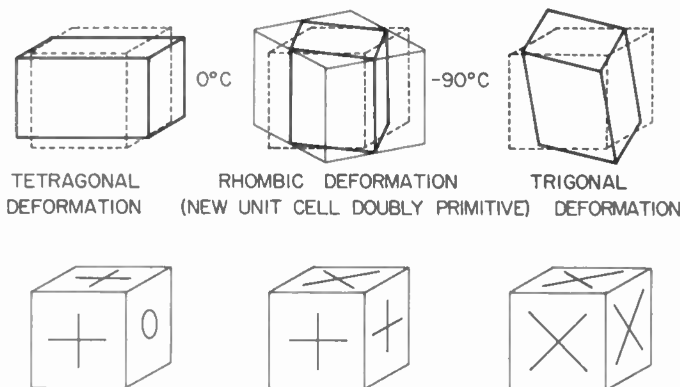


Fig. 75—The three deformations of $BaTiO_3$ and the corresponding extinction directions in a single domain cube (after Forsbergh [62]).

called "parallel" when the directions lie parallel to the original cube axes; and it is called "symmetrical" when the extinction directions lie at 45° (symmetric) with the original axes. Detailed optical investigations have shown that in the orthorhombic phase the boundary between two domains with symmetrical extinction is an orthorhombic (101) plane; and the wall between domains with parallel extinction, as well as that between one domain with parallel and one with symmetrical extinction, is an orthorhombic (111) plane. The situation is shown schematically in Fig. 76. This figure also shows the observed twinning of a $BaTiO_3$ plate in rhombohedral phase. It is evident that a plate with cubic habit consists of domains, all of which show symmetrical extinction.

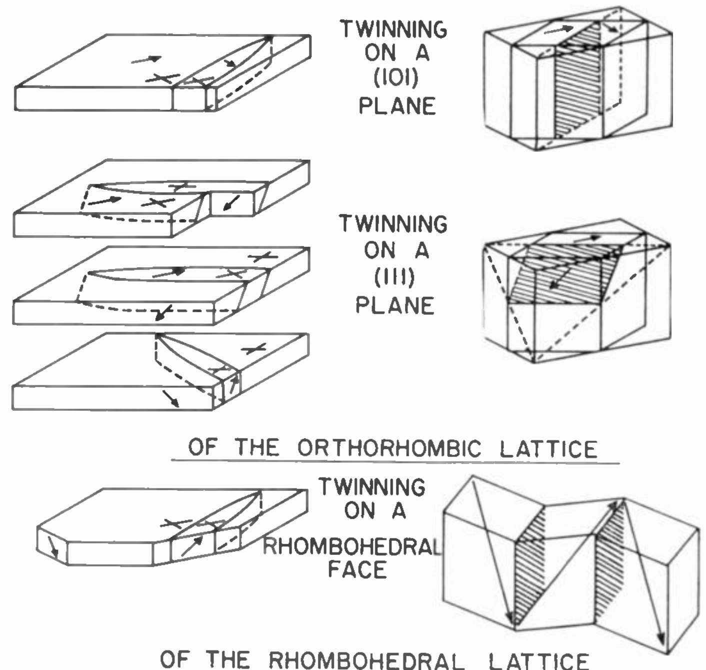


Fig. 76—The predominant twinning in the orthorhombic and rhombohedral phases of $BaTiO_3$ (after Forsbergh [62]).

E. Domain Walls

From the above considerations of the *geometry* of domain walls, which can be related to the symmetry of the crystals concerned, we proceed to a discussion of wall energy.

A first approach to the problem can be made by treating the ferroelectric material as a continuum, and applying thermodynamic methods. A detailed treatment of this sort has been developed for the first time by Mitsui and Furuichi [23], who used the results of their theory in a calculation of the wall width and energy of Rochelle Salt domains. These authors deal, theoretically, with a semi-infinite crystal which is to be polarized in the direction of the x axis, and assume that the polarization $P(y)$ is a continuous function of the coordinate y. The wall energy F_{wall} is considered as comprised of interaction and elastic energy; and the calculation yields the relation:

$$F_{wall} = \text{const.} \cdot P_0^3. \tag{33}$$

The constant appearing in (33) is a function of the elastic and the electrostrictive constants, as well as of the coefficients involved in the free energy expansion for the crystal under consideration. The introduction of the

numerical values of the constants for Rochelle Salt yields

$$F_{wall} = 0.057 \text{ erg/cm}^2 \text{ at } 0°\text{C},$$

and

$$F_{wall} = 0.010 \text{ erg/cm}^2 \text{ at } 20°\text{C}.$$

The thickness of the wall is given by an expression of the type

$$t = \frac{\text{const.}}{P_0}; \tag{34}$$

and numerical evaluation yields

$$t = 24 \text{ A at } 0°\text{C}$$

and

$$t = 43 \text{ A at } 20°\text{C}.$$

This result is remarkable, since it tells us that, within the approximation of the continuum theory, the thickness of a domain wall in Rochelle Salt is of the order of a few lattice constants—in contrast with ferromagnetic materials, where domain boundaries are much thicker (*e.g.*, about 280 lattice constants, or approximately 1,000 A, in iron).

This difference, however, can be anticipated on the basis of qualitative differences between the magnetic and electric cases. In the ferromagnetic case the domain wall consists of a transition region in which the magnetization vector turns over gradually from a given direction to, let us say, the opposite direction. The wall thickness is the result of a compromise between the tendency of the exchange energy to make the transition region between anti-parallel domains as wide as possible and the tendency of the anisotropy energy to make the same region thin in order to avoid the directions of "hard magnetization." The contribution of magnetostriction is negligible. The exchange term is roughly 10^5 times larger than the anisotropy term; and as a consequence a relatively thick boundary results. In ferroelectrics, the exchange energy is replaced by the energy of interaction between the dipoles. The nature of the electric dipole interaction is such that it is energetically disadvantageous for the dipoles to undergo a *gradual* transition from their maximum value in one direction to their maximum value in the opposite direction. In addition, the anisotropy energy in the sense given above is very large indeed. The elastic energy plays an important role in the ferroelectric case, being of the same order of magnitude as the interaction energy, and it of course also favors a thin transition region. We can conclude from these qualitative considerations that a ferroelectric domain wall will cover a region of a few lattice constants only. The results of the continuum theory then are in accord with expectation.

It is readily possible to evaluate the results of the continuum theory for the case of KH_2PO_4, as done by Kaenzig and Sommerhalder [141], and for that of $BaTiO_3$. The wall thickness turns out to be a few lattice constants, in both cases.

The next step appears to be that of the development of a *molecular theory* of the domain wall. This has been done by Kaenzig and Sommerhalder for the case of KH_2PO_4 [141]. The task is not easy. Firstly, a molecular theory can be developed only if one has a *model*; and a model for ferroelectric theory is bound to be a simplification of reality. In the second place, the calculations can be performed precisely only for absolute zero of temperature, since consideration of thermal motion would complicate the problem prohibitively.

Kaenzig's model is that of point dipoles arranged in a cubic body-centered lattice, with the dipoles aligned in the (001) direction. A wall is assumed to be parallel to the (100) plane, and its energy calculated from the difference between the interaction energies of the state *with* the wall and the state *without* a wall, respectively. The coupling between the dipoles is taken care of by the introduction of an *internal field*. The result is, energy of wall per cm^2 is given by an expression of the type

$$F_{wall} = \text{const.} \cdot P^2, \tag{35}$$

where the constant can be evaluated numerically on the basis of experimental results (see this Section, paragraph *K*) on KH_2PO_4 crystals, yielding

$$F_{wall} = 47 \text{ erg/cm}^2,$$

and

$$t \simeq 2 \text{ or } 3 \text{ lattice constants}$$

at about 100°K, that is, about 20 degrees below the Curie point. The first conclusion to be reached from this result is that the wall energy of KH_2PO_4 domains appears to be about 1,000 times larger than that of Rochelle Salt domains. The quantitative value of the ratio between these wall energies, of course, does not mean much, since they are both derived on the basis of assumptions which are oversimplified. But it is reasonable that the wall energy is higher in KH_2PO_4 than in Rochelle Salt, because of the differences in temperature and hardness of the two crystals. The wall thicknesses, on the other hand, appear to be of the same order of magnitude. One must not overestimate the reliability of these results; but they do suggest that, whatever the model assumed, these ferroelectric domains appear to have walls only a few lattice constants thick.

A similar result can be obtained for the case of 180° domain boundaries in $BaTiO_3$, as Merz [42] has shown by translating the theory for the wall energy of ferromagnetic domains in a purely formal way. Here again the magnetic exchange energy is replaced by the interaction energy, and the anisotropy energy by the elastic energy. The results for $BaTiO_3$ at room temperature (i.e., about 100° below the Curie point) are:

$$F_{wall} = 7 \text{ erg/cm}^2$$

and

$$l \simeq 1 \text{ lattice constant.}$$

As for the 90° walls in $BaTiO_3$, Little [43] assumes two contributions to the wall energy. The first is given by a linear and continuous variation of the angle between the dipoles within the wall—an assumption which is questionable; and the second is given by a term whose microscopic origins are not detailed, but which is postulated as being proportional to $1/l$. In this way, Little obtains the following values for the wall energy:

$$F_{\text{wall}} = 65 \text{ erg/cm}^2,$$

and for the wall thickness

$$l = 4{,}000 \text{ A or about } 1{,}000 \text{ lattice constants.}$$

These values are very probably too high.

F. Dynamic Properties of the 180° Walls in $BaTiO_3$

In the study of the behavior of 180° walls in $BaTiO_3$ under an applied field, Merz finds that there is practically no sidewise motion of the boundaries. As a consequence, the polarization produced by a set of charged electrodes at one point of the crystal surface does not interfere with the polarization caused by another set at another point, even if the two sets are very close to each other (10^{-2} cm or less). This is a very valuable property for a memory device. The absence of sidewise motion of the 180° walls implies that reversal of polarization takes place by way of nucleation of many new domains with opposite polarization. This has been directly confirmed by Merz, who applied an ac voltage to a c domain crystal, and observed the formation and the growth of new domains by illuminating the crystal with a stroboscopic light source whose frequency was equal to that of the ac field. This technique had been utilized by several previous investigators.

Under the action of the electric field it is easier for the crystal to create new domains, which then grow rapidly in the forward direction, than to move the walls of the already existing domains sidewise. This fact is understandable from the results obtained in the last section. The wall is only a very few lattice constants thick, but its energy is comparatively large (~ 7 erg/cm^2). If the wall is set in motion its energy increases further, because of the inertia of the ions which must shift to new equilibrium positions as the dipole moments change direction (cf. Kittel [142]). In order to move the wall one unit cell to the side, an amount of work will have to be provided which is of the order of magnitude of the surface energy of the wall itself. The energy gained by such a motion is of the order of $E \cdot P$. With $P \sim 25 \cdot 10^{-6}$ coulombs/cm^2, and, for example, $E = 10^4$ volts/cm, then $E \cdot P = 0.1$ erg/cm^2. The nucleation of new domains, on the other hand, seems to be easier. In every crystal there are always quite a few local imperfections, and at the surface of the crystal, especially, local strains are to be found. These, with the help of an applied electric field, favor the flipping over of a few dipoles. Theoretically, the choice between the two alternatives, domain wall motion or nucleation of new domains, can be made only upon comparison of the activation energies involved in either process. The energy barrier for wall motion can be computed exactly only with the knowledge of the detailed manner in which the wall moves, while the activation energy for nucleation is strictly related to the shape and size of the domain nuclei. This process of domain nucleation takes place for fields as low as about 2 kv/cm (cf. Little [43]). The nuclei, in the form of spikes, initially grow very fast in the forward direction probably by means of a sequence of dipole flips which propagate along the polar axis. When the electrostatic energy of the depolarizing field increases, the propagation velocity decreases, and the growth continues as neutralizing charge accumulates on the domain walls. The minimum field needed for the further growth of nuclei is about 2.4 kv/cm (cf. Little [43]). The fact that the limiting fields for nucleation and for growth of anti-parallel domains are nearly the same is considered an experimental support for the hypothesis that both processes involve the reversal of individual dipole moments.

Merz [42] has explained the fact that the nucleated domains grow in the forward direction only, by considering that the forward coupling of the electric dipoles is large while the sidewise coupling is very small. Whether the neighboring dipoles on the side are parallel or anti-parallel is not too important to a given dipole. In $BaTiO_3$ the parallel alignment is usually preferred because the crystal is ferroelectric. However, Cohen [143] has shown that in the perovskite structure an anti-ferroelectric arrangement is energetically almost as likely as a ferroelectric one. The strong forward coupling of the dipoles, on the other hand, puts an important emphasis on the O_I—Ti—O_I chains, rather than the O_{II}—Ti—O_{II} chains—in contrast to the hypothesis advanced by Megaw [110, 144].

The mechanism of reversal of the polarization involves two processes: nucleation of domain spikes (with *no* charge compensation), and growth of domain wedges (*with* charge compensation). This picture is in line with the results of electrical pulsing experiments of Merz [42]. Merz applied square pulses to a crystal in series with a resistance. The electrode area was of the order of 10^{-4} cm^2, and the crystal thickness was $d = 5 \times 10^{-3}$ cm. The characteristics of the pulses were: rise time (0.02 microsec) was small compared to the time required by the field to reverse the polarization within the crystal; length of the pulses (1 to 10 microsec) was longer than this time. If repeated pulses of one sign only are applied to the crystal, the change in polarization is small, because the conditions correspond to the flat part of the hysteresis loop (see Fig. 1); and the current measured through the series resistance is small. This corresponds to a binary "zero" when the specimen is used as a memory device. If alternate positive and negative pulses

are applied, the change in polarization is large, because the conditions are then those for the steep part of the hysteresis loop; therefore, the measured current is large, and this corresponds to a binary "one." The plot of the measured peak current i_{max} vs the applied field E shows that the curve can be divided into two portions: a curved low-field-strength portion (from $E=2$ kv/cm to $E=6$ kv/cm), and a linear high-field-strength portion (for $E>6$ kv/cm).

The low-field-strength portion corresponds to the process of nucleation and growth of spikes without charge compensation. The dependence is exponential:

$$i_{max} \sim e^{-\alpha(T)/E},$$

where the temperature-dependent quantity $\alpha(T)$ represents some kind of activation field, and the $1/E$ dependence in the exponent can be understood upon assuming dagger-shaped nuclei (cf. Merz [42]). In the high field region the current i_{max} is linearly dependent on the applied field strength E:

$$i_{max} \sim \mu(T)(E - E''),$$

where the mobility μ is about 2.5 cm²/volt sec at room temperature, and E'' is of the order of 2.5 kv/cm. In this region of high field strength the nucleation is extremely fast, and the velocity of propagation of the domains along the polar axis is of the order of 10^4 cm/sec for fields of about 10^4 volts/cm. The crystal behaves like an ohmic resistance R, with R directly proportional to the square of the thickness of the specimen, and inversely proportional to the area A of the electrodes. The switching time t_s is proportional to the thickness of the sample, and, above a certain field, inversely proportional to the field strength E. For example: for a crystal 0.005 cm thick, at room temperature and for a field of 11,500 v/cm, it takes 0.2 microsecond to reverse the polarization. At constant field the switching time is a function of the temperature, being shorter for higher temperatures—as is to be expected from the fact that the rate of nucleation increases while the anisotropy of the crystal decreases with increasing temperatures. These limiting switching times, as pointed out by Mason and Wick [145], provide a natural limit to the speed of operation of storage devices, and also limit the maximum frequencies for dielectric amplifiers. For thin crystals the maximum frequencies are of the order of megacycles.

G. Ferroelectrics as Storage Elements

Although this field of research is not yet completely explored, the application of $BaTiO_3$ crystals as devices for storing digital information in computers and switching systems appears to hold great promise. Some possible types of circuits for the operation of these devices have been described by Anderson [146]. The advantages of the ferroelectric elements are that they require low power consumption while storing or reading out information, they have small size per memory unit, they have memory access times of one microsecond or less, and their construction is simple and inexpensive. According to Anderson, up to 2,500 bits of information per square inch can be stored in a surface only a few thousandths of an inch thick, using pulses less than a microsecond long. Operation can be provided from low voltage circuits (10 volts or less) such as transistors. No power consumption occurs during storage period, and storage can be provided for long periods of time without regeneration.

What properties are important in a ferroelectric, for it to be useful as a storage element? Among the many points which have to be considered, an essential one concerns the mechanical stresses introduced upon reversing the polarization at a point. For example, let us consider Rochelle Salt. Clearly, this crystal is not suitable because it is mechanically weak, chemically unstable, its ferroelectric properties disappear at temperatures slightly above room temperature, and its hysteresis loops are not sufficiently "square." These, however, are not the most decisive factors. Generally, when we apply a field to a small area of the surface of a single domain crystal in order to reverse the polarization of that small region, what we require is that the new domain remains where it was created even after the removal of the field, without expanding (crosstalk) or disappearing (forgetting). Such a situation is schematically represented in Fig. 77(a) and it seems that this is the case for a $BaTiO_3$ plate. In Rochelle Salt, however, the situation is rather like that represented in Fig. 77(b). The reversal of the polarization at the point indicated by the + domain implies a change in sign of the spontaneous shear y_z. Now, if the field is removed the new domain disappears as a consequence of the mechanical stress caused by the spontaneous shear; if the field is maintained the domain grows by means of wall motion, thus causing crosstalk. The same situation is found for KDP. All three phases of $BaTiO_3$ appear satisfactory, however, and so does GASH.

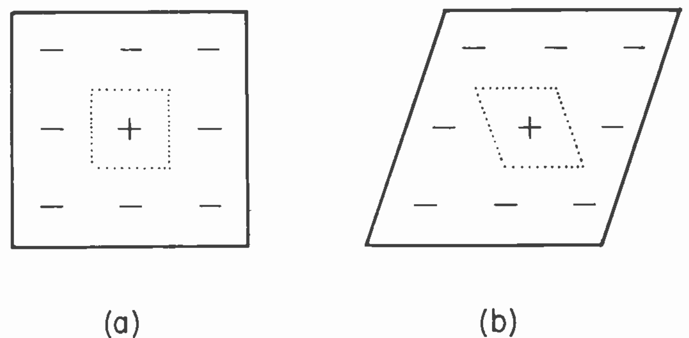


Fig. 77—Creation of an anti-parallel domain in a single domain crystal (schematic): (a) $BaTiO_3$ type, (b) Rochelle Salt and KH_2PO_4 type.

In connection with the dynamic properties of the 180° wall in $BaTiO_3$ it must be mentioned that the recent investigation by Little [43] revealed that even though the activation energy for motion prependicular to the polar axis is very high, such motion was actually ob-

served, in contrast with Merz's electrical measurements. The reasons for this discrepancy have not been clarified. We might speculate that a source of difference could be that the geometry of the experiment is different for the two authors. While Merz observed the edge of a crystal plate strained by means of a field directed perpendicularly to the polar axis (Fig. 72), Little studied the behavior of the 180° walls upon applying the electric field at 45° to the direction of the polar axis in an *a* domain plate (Fig. 78). The angle between the domain wall and the field is different for the two methods. Also there might be a different effect of the external surface of the crystal on the 180° wall. A second explanation could be that the crystals used by the two investigators show different properties, and a difference in conductivity might be particularly important. Little's crystals were grown using $BaCl_2$ as the flux; Merz's crystals were grown by Remeika's method. The smaller conductivity of the latter might have impeded the sidewise motion of the 180° walls.[5]

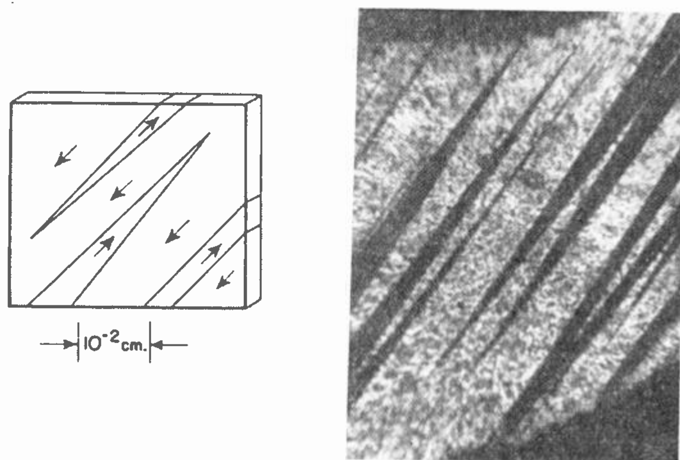


Fig. 78—Anti-parallel wedges in an *a* plate of $BaTiO_3$ (after Little [43]).

H. Dynamic Properties of the 90° Domain Walls in $BaTiO_3$

Both nucleation and growth of 90° domains have been studied by Little [43], on crystals which initially were single-domain. By means of electrodes on the opposite edges of an *a* domain plate, a dc field was applied in the direction perpendicular to the polarization axis. Upon increasing the field slowly, it is observed that 90° wedges are nucleated at the crystal boundaries at a critical field strength, E_{crit}. These wedges, from 2 to 10 microns wide, extend across a crystal of 0.05-cm width in less than 0.01 sec. E_{crit} is a function of temperature, its behavior being very similar to that of the spontaneous polarization and strain in a single domain crystal (Fig. 8). At room temperature E_{crit} is about 2.4 kv/cm. When the field is removed the wedges shrink slowly at first, then disappear suddenly. They are forced out of the crystal by stresses produced by the domains. Upon applying a field of opposite polarity, it is seen that nucleation starts from the other side of the crystal, indicating that the wedges normally start at the cathode side.

The initial growth of 90° wedges was studied upon applying a square pulse to the *a* domain plate, and recording the number of wedges N in the crystal immediately afterwards. N is, of course, a function of the pulse height E and the pulse time t. For a given field strength there exists a threshold pulse length t_0, below which no wedges are formed. At low field strength, t_0 depends on the field according to the equation

$$t_0 \cdot (E - E_{crit})^2 = \text{constant}, \tag{36}$$

while at high fields

$$t_0 = t_s e^{B/E}, \tag{37}$$

where $B = 25$ kv/cm, and $t_s = 0.05$ microsec. From this formula Little was able to estimate the limiting velocity of a wedge at high fields, upon interpreting t_0 as the minimum time required for wedges to grow large enough to be observed. This velocity is about 10^6 cm/sec; *i.e.*, of the order of the velocity of sound.

After the wedges have been introduced into and grown across a crystal, domain growth can be described in terms of wall motion: the two now parallel walls separate and move sideways with decreasing velocity and some irregularity. The main difference between these 90° walls and the 180° walls described in the preceding sections is that the 90° walls seem to be considerably thicker, so that their motion is not a process of discrete molecular steps, but is rather to be thought of as a slip of one domain with respect to its neighbor. Clearly this process is strongly influenced by the mechanical stresses involved with the small deviation of the crystal axes from 90° across a domain wall—so that the crystal changes its shape during the motion of the wall, and also by the piezoelectric effect connected with it. In fact, the experiments show that it is possible to cause the domain wall to oscillate, upon applying a small alternating field; but the wall motion is damped out as the acoustic resonance frequency is approached, so that the wall is probably no longer induced to move by finite field strengths at frequencies much above 3 mc/sec.

I. Interactions Between 180° and 90° Walls in $BaTiO_3$

The complicated interactions occurring in a $BaTiO_3$ crystal when 90° and 180° walls are simultaneously active have been studied by Little [43]. The remarkable result of this investigation is that head-to-head, or tail-to-tail, configurations of the 90° wall can occur as a consequence of this interaction. The interfacial polarization is neutralized by free charges, and the resulting domain configuration is called "pseudo-saturated" because, even though there exist 90° domains, they cannot be removed by a field as long as the energy barrier for rotation is too high.

[5] It should be mentioned, however, that in another place Merz also reports movement of the domain wall (see Merz [42], p. 691). He makes no attempt to correlate the differing observations.

The interactions between 180° and 90° walls may play an important role in determining the state of strain in a crystal, thus contributing to the aging phenomena observed both in ceramics and single crystals. A more detailed investigation of the problem is to be awaited.

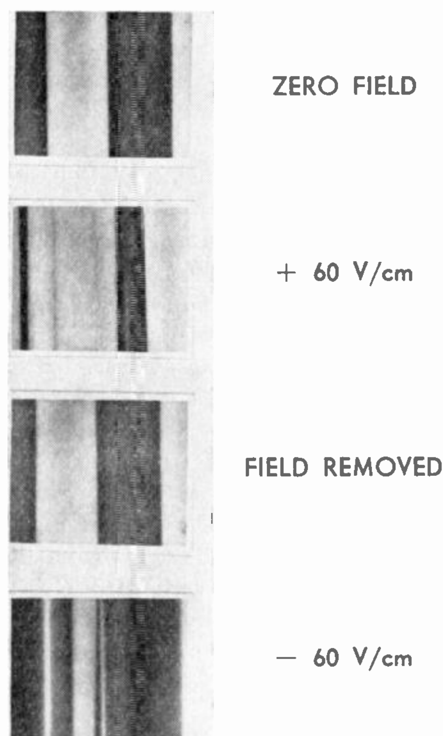


Fig. 79—Effect of an electric dc field on the domain structure of Rochelle salt (after Mitsui and Furuichi, kindly furnished by T. Mitsui).

J. Dynamic Properties of Rochelle Salt Domains

The dynamic properties of Rochelle Salt domains have been studied by Marutake [147], and by Mitsui and Furuichi [24]. Qualitatively, the behavior is similar to that of 90° walls in $BaTiO_3$. An electric field applied to a Rochelle Salt crystal along the a axis causes many new domains to appear. When application and withdrawal of the field is repeated, it is observed that the new domains always appear at the same place in the crystal. This fact again suggests that nucleation of the domains takes place at crystal imperfections. The relaxation time for appearance of new domains depends exponentially on the applied field strength, according to the equation

$$t_s = t_{s0} \cdot e^{C/E}, \tag{38}$$

where both t_{s0} and C are constants, in analogy to Merz's results for $BaTiO_3$. The relaxation time t_s is of the order of 2 microseconds for a field E of about 200 v/cm at about 19°C; *i.e.*, five degrees below the upper Curie point. Once the domains have grown across the crystal, the electric field causes a motion of the domain wall (see Fig. 79). The propagation velocity v is directly proportional to the electric field strength E. The magnitudes are smaller than in the case of $BaTiO_3$: v is about 0.2 cm/sec for a field E of about 200 v/cm at 18.4°C, and increases with temperature at constant field.

K. The Effect of Particle Size

We have seen that an isolated nonconducting ferroelectric crystal which is forced to form a single domain is subjected to a very strong depolarizing field. The energy of this field was denoted by $U_e = \frac{1}{2}\Gamma P^2 V$, where, in the case of a spherical crystal, $\Gamma = 4\pi/3$. If the same crystal is provided with short-circuited electrodes on the faces perpendicular to the ferroelectric directions, and then undergoes a spontaneous polarization, the energy of the crystal will decrease. The change in energy of the short-circuited ferroelectric crystal is referred to as *polarization energy*:

$$U_P = -\tfrac{1}{2} f P^2 V, \tag{39}$$

where P and V represent polarization and volume, respectively, while f is a constant whose order of magnitude lies, for known ferroelectric crystals, between 2 and 0.03. In the particular case of a spherical crystal, the balance between U_e and U_P is clearly in favor of the former, thus impeding the uniformly polarized state. What actually happens is that the energy of the depolarizing field is decreased by the formation of domain walls, and consequently an homogeneous spontaneous polarization can take place within every domain.

Suppose now that the dimensions of the crystal are diminished. In this case, the relative contributions of the different energy terms to the total domain energy change, and surface energies become more important than volume energies. The energy of the domain walls is a surface energy; upon decreasing the crystal size, its relative contribution to the total energy of the crystal becomes larger and larger. When very small dimensions are reached there will be a point at which it will be energetically more favorable to do without the domain walls, so that the whole crystal particle becomes one single domain. Then the spontaneous polarization becomes impossible because of the depolarizing field. The crystal particle is no longer ferroelectric

This idea was developed by Kaenzig and co-workers [133, 141] and confirmed by them in a series of experiments on KH_2PO_4. These experiments involved X-ray and electron diffraction, dielectric measurements, and electron microscopy on colloidal particles of KH_2PO_4 imbedded in an insulating medium with an effective dielectric constant of about 5. KH_2PO_4 is particularly suitable because the Curie temperature is so low that both crystal and imbedding medium can be considered as perfect insulators; moreover, the spontaneous polarization can only lie along the c axis, so that no closure of the electrical flux is possible. The experiments show that no spontaneous polarization occurs in particles with diameter smaller than 1,500 A, while particles larger than 4,000 A show the normal ferroelectric behavior. Clearly the critical particle size is dependent upon the dielectric constant of the imbedding medium, and in

fact it is found to decrease with the increase of the latter. If the imbedding medium is conducting, even the smallest particles produced (about 500 A) show normal ferroelectric behavior. From the critical particle size it is possible to compute the surface energy of the domain boundaries; the results of these computations were given in Section VII, *E*. In spite of the fact that, below the critical size, the ferroelectric transition is forbidden, an anomaly of the dielectric constant can still be detected about one degree below the Curie temperature of the macroscopic KH_2PO_4 crystal. The nature of this transition is not yet known.

Similar experiments were performed, also by Kaenzig and co-workers [148], on colloidal particles of $BaTiO_3$. It is clear that in this case different results will have to be expected, since the two conditions that made KH_2PO_4 most advantageous are no longer satisfied; the Curie temperature of $BaTiO_3$ is much too high for the crystal and the imbedding medium to be considered as insulators; and in $BaTiO_3$ spontaneous polarization can occur along three directions perpendicular to each other, which admits the possibility of domain structures with closed electrical flux. Consequently, it is not to be expected that a critical particle size of $BaTiO_3$ will exist, below which no spontaneous polarization can occur, but rather that domain configurations with closed flux will take place. Actually, an effect of particle size was found, in the sense that the ferroelectric properties change gradually with decreasing crystal dimensions. The 120°C transition between the polar tetragonal and the non-polar cubic phase, a very sharp transition in macroscopic $BaTiO_3$ crystals, becomes smeared out over a temperature range which is the larger, the smaller the particle size. It is possible to detect the tetragonal distortion even at temperatures a few hundred degrees above the Curie temperature of the macroscopic crystal! At room temperature the tetragonal distortion decreases with decreasing particle size.

It was possible, by means of electron diffraction experiments, to prove that there is a discrepancy between the interior of the crystal and a surface layer of about 100 A thickness. This is the cause for the deviations from the normal ferroelectric behavior. The surface layer shows a large spontaneous tetragonal strain and a very high Curie temperature, while the interior of the crystal particle prefers the cubic symmetry at temperatures above the Curie point of the macroscopic crystal. Between the surface layer and the interior there is a transition region with large atomic misfit. Kaenzig [166] suggested lattice vacancies and imperfections might be responsible for the spontaneous deformation of the surface. The existence of the latter seems to be confirmed, even in macroscopic crystals, by the fact that new domains are nucleated, preferentially at the surface (see Section VII, *F*).

VIII. Some Conclusions and Acknowledgments

In the preceding discussion we have concerned ourselves with some thermodynamic, crystal structural, and domain properties of most of the known ferroelectrics. We have not considered such practical matters as the use of these materials in such applications as electromechanical transducers or dielectric amplifiers. We have not considered such pressing problems as aging effects in ceramic and single-crystal ferroelectrics. Knowledge of subjects which we have discussed we have found incomplete; and we have left the problems scarcely improved in this respect. After our survey, it should not be necessary to emphasize the vast amounts of work yet to be accomplished.

We are very grateful for helpful discussions with Prof. Y. Takagi of the Tokyo Institute of Technology, Prof. Jack Tessman of The Pennsylvania State University, and Dr. B. C. Frazer of the Brookhaven National Laboratory. We express gratitude as well to the many investigators and journals whose figures and tables have been used in this article and whose names we have tried diligently to associate with their respective figures and tables.

Our own research on ferroelectricity at Pennsylvania State University has been supported by contracts with the Wright Air Development Center, the Signal Corps Engineering Laboratory, the Office of Naval Research, the Atomic Energy Commission, the Gulf Research and Development Company, and by the Brookhaven National Laboratory. We are pleased for another opportunity to express our appreciation for this support.

Bibliography

[1] Cady, W. G., *Piezoelectricity.* New York, McGraw-Hill Book Co., Inc., 1946.
[2] Sawyer, C. B., and Tower, C. H., "Rochelle Salt as a Dielectric." *Physical Review*, Vol. 35 (February 1, 1930), pp. 269–273.
[3] (a) Mueller, H., "Properties of Rochelle Salt." *Physical Review*, Vol. 47 (January 15, 1935), pp. 175–191.
(b) Mueller, H., "Properties of Rochelle Salt, II." *Physical Review*, Vol. 57 (May 1, 1940), pp. 829–839.
(c) Mueller, H., "Properties of Rochelle Salt, III." *Physical Review*, Vol. 58 (September 15, 1940), pp. 565–573.
(d) Mueller, H., "Properties of Rochelle Salt, IV." *Physical Review*, Vol. 58 (November 1, 1940), pp. 805–811.
(e) Mueller, H., "The Dielectric Anomalies of Rochelle Salt." *Annals of the New York Academy of Science*, Vol. 40 (December 31, 1940), pp. 321–356.
[4] Devonshire, A. F., "Theory of Barium Titanate, Part I." *Philosophical Magazine*, Vol. 40 (October, 1949), pp. 1040–1063.
[5] Devonshire, A. F., "Theory of Barium Titanate, Part II." *Philosophical Magazine*, Vol. 42 (October, 1951), pp. 1065–1079.
[6] Devonshire, A. F., "Theory of Ferroelectricity." *Philosophical Magazine, Supplement*, Vol. 3 (April, 1954), pp. 85–130.
[7] Cotts, R. M., and Knight, W. D., "Nuclear Resonance of Nb^{93} in $KNbO_3$." *Physical Review*, Vol. 96 (December 1, 1954), pp. 1285–1293.
[8] Matthias, B. T., "Ferroelectricity." *Science*, Vol. 113 (May, 1951), pp. 591–596.
[9] Mason, W. P., *Piezoelectric Crystals and Their Application to Ultrasonics.* New York, D. Van Nostrand Company, 1950.
[10] Baumgartner, H., Jona, F., and Kaenzig, W., "Seignette-Elektrizität." *Ergebnisse der Exakten Naturwissenshaften*, Vol. 23 (1950), pp. 235–282.
[11] Busch, G., "Neue Seignette-Elektrika." *Helvetica Physica Acta*, Vol. 11 (June 18, 1938), pp. 269–298.
[12] de Quervain, M., "Röntgenometrische Untersuchungen an Kaliumphosphat bei tiefen Temperaturen." *Helvetica Physica Acta*, Vol. 17 (December 23, 1944), pp. 509–552.
[13] von Arx, A., and Bantle, W., "Polarisation und spezifische Wärme von KH_2PO_4." *Helvetica Physica Acta*, Vol. 16 (June 30, 1943), pp. 211–214.
[14] Baumgartner, H., "Elektrische Sättigungserscheinungen und elektrokalorischer Effekt von Kaliumphosphat, KH_2PO_4." *Helvetica Physica Acta*, Vol. 23 (Dec. 10, 1950), pp. 651–696.
[15] West, J., "A Quantitative X-Ray Analysis of the Structure of

Potassium Dihydrogen Phosphate (KH_2PO_4)." *Zeitschrift für Kristallographie*, Vol. 74 (1930), pp. 306–332.

[16] Slater, J. C., "Theory of the Transition in KH_2PO_4." *Journal of Chemical Physics*, Vol. 9 (January, 1941), pp. 16–33.

[17] Frazer, B. C., and Pepinsky, R., "X-Ray Analysis of the Ferroelectric Transition in KH_2PO_4." *Acta Crystallographica*, Vol. 6 (March, 1953), pp. 273–285.

[18] Peterson, S. W., Levy, H. A., and Simonsen, S. H., "Neutron Diffraction Study of Tetragonal Potassium Dihydrogen Phosphate." *Journal of Chemical Physics*, Vol. 21 (November, 1953), pp. 2084–2085.

[19] Levy, H. A., Peterson, S. W., and Simonsen, S. H., "Neutron Diffraction Study of the Ferroelectric Modification of Potassium Dihydrogen Phosphate." *Physical Review*, Vol. 93 (March 1, 1954), p. 1120.

[20] Pease, R. S., and Bacon, G. E., "Ferroelectric Structure of Potassium Dihydrogen Phosphate." *Nature*, Vol. 173 (March, 1954), pp. 443–444.

[21] Bacon, G. E., and Pease, R. S., "Neutron Diffraction Study of Potassium Dihydrogen Phosphate by Fourier Synthesis." *Proceedings Royal Society A*, Vol. 220 (1953), pp. 397–421.
"A Neutron-Diffraction Study of the Ferroelectric Transition of Potassium Dihydrogen Phosphate." *Proceedings of the Royal Society (A)*, Vol. 230 (June 21, 1955), pp. 359–381.

[22] Habluetzel, J., "Schweres Seignettesalz, Dielektrische Untersuchungen an $KNaC_4H_2D_2O_6 \cdot 4D_2O$-Kristallen." *Helvetica Physica Acta*, Vol. 12 (November 4, 1939), pp. 489–510.

[23] Mitsui, T., and Furuichi, J., "Domain Structure of Rochelle Salt and KH_2PO_4." *Physical Review*, Vol. 90 (April 15, 1953), pp. 193–202.

[24] Mitsui, T., and Furuichi, J., "Kinetic Properties of the Domains in Rochelle Salt." *Physical Review*, Vol. 95 (July 15, 1954), p. 558.

[25] Beevers, C. A., and Hughes, W., "The Crystal Structure of Rochelle Salt, Sodium Potassium Tartrate Tetrahydrate, $NaKC_4H_4O_6 \cdot 4H_2O$." *Proceedings of the Royal Society*, Vol. 177 (January 10, 1941), pp. 251–259.

[26] Mason, W. P., "Theory of the Ferroelectric Effect and Clamped Dielectric Constant of Rochelle Salt." *Physical Review*, Vol. 72 (November 1, 1947), pp. 854–865.

[27] Sundara Rao, R. V. G., Mazzi, F., and Pepinsky, R., "Structural Study of Rochelle Salt." To be published.

[28] Frazer, B. C., McKeown, M., and Pepinsky, R., "Neutron Diffraction Study of Rochelle Salt Single Crystals." *Physical Review*, Vol. 94 (June 1, 1954), p. 1435.
Frazer, B. C., Danner, H. R., and Pepinsky, R., "Neutron Diffraction Study of Rochelle Salt." To be published.

[29] Kurtschatov, I. V., *Le Champ Moléculaire dans les Diélectriques (le sel de Seignette)*. Paris, Herman et Cie, 1936.

[30] Melmed, A., Jona, F., and Pepinsky, R., "Ferroelectric and High Dielectric Crystals: Dielectric Properties of Na(K, NH_4)-$C_4H_4O_6 \cdot 4H_2O$ Mixed Crystals." Technical Report No. 20, Contract AF33(616)-2133, Wright Air Development Center; January, 1954.

[31] Matthias, B. T., and Hulm, J. K., "New Ferroelectric Tartrates." *Physical Review*, Vol. 82 (April 1, 1951), pp. 108–109.

[32] Merz, W. J., "Lithium Ammonium Tartrate Monohydrate, $LiNH_4C_4H_4O_6 \cdot H_2O$, A New Ferroelectric Crystal." *Physical Review*, Vol. 82 (May 15, 1951), pp. 562–563.

[33] Vernon, R. C., and Pepinsky, R., "Structure Determination of $LiNH_4C_4H_4O_6 \cdot H_2O$." Program and Abstracts, American Crystallographic Association, Camp Tamiment, Pa.; June 18, 1952.
"X-Ray Analysis of $LiNH_4$tartrate·H_2O and LiRbtartrate·H_2O." Final Report, Contract DA-36-039-SC-21, Signal Corps Engineering Laboratories; January, 1953.

[34] Cook, W. R., Jr., and Jaffe, H., "Ferroelectricity in Oxides of Face-Centered Cubic Structure." *Physical Review*, Vol. 89 (March 15, 1953), pp. 1297–1298.
"Ferroelectricity in Oxides of Fluorite Structure." *Physical Review*, Vol. 88 (December 15, 1952), p. 1426.

[35] Jona, F., Shirane, G., and Pepinsky, R., "Dielectric, X-Ray and Optical Study of Ferroelectric $Cd_2Nb_2O_7$ and Related Compounds," *Physical Review*, Vol. 98 (May 15, 1955), pp. 903–909.

[36] Matthias, B. T., and Remeika, J. P., "Ferroelectricity in the Ilmenite Structure." *Physical Review*, Vol. 76 (December 15, 1949), pp. 1886–1887.

[37] Schweinler, H. C., "Ferroelectricity in the Ilmenite Structure." *Physical Review*, Vol. 87 (July 1, 1952), pp. 5–11.

[38] Merz, W. J., "The Electric and Optical Behavior of $BaTiO_3$ Single Domain Crystals." *Physical Review*, Vol. 76 (October 15, 1949), pp. 1221–1225.

[39] Blattner, H., Matthias, B. T., Merz, W., and Scherrer, P., "Investigations of Barium Titanate Single Crystals." *Experientia*, Vol. 3 (1947), pp. 148–149.

[40] Remeika, J. P., "A Method for Growing Barium Titanate Single Crystals." *Journal of the American Chemical Society*, Vol. 76 (February, 1954), pp. 940–941.

[41] Merz, W. J., "Double Hysteresis Loop of $BaTiO_3$ at the Curie Point." *Physical Review*, Vol. 91 (August 1, 1953), pp. 513–517.

[42] Merz, W. J., "Domain Formation and Domain Wall Motion in Ferroelectric $BaTiO_3$ Single Crystals." *Physical Review*, Vol. 95 (August 1, 1954), pp. 690–698.

[43] Little, E. A., "Dynamic Behavior of Domain Walls in Barium Titanate." Technical Report 87, Laboratory for Insulation Research, Massachusetts Institute of Technology, October, 1954, *Physical Review*, Vol. 98 (May 15, 1955), pp. 978–984.

[44] Frazer, B. C., Danner, H., and Pepinsky, R., "The Crystal Structure of Tetragonal $BaTiO_3$ by Neutron Diffraction." *Physical Review*, Vol. 100 (October 15, 1955).

[45] Matthias, B. T., and Remeika, J. P., "Dielectric Properties of Sodium and Potassium Niobate." *Physical Review*, Vol. 82 (June 1, 1951), pp. 727–729.

[46] Wood, E. A., "Polymorphism in Potassium Niobate, Sodium Niobate and Other ABO_3 Compounds." *Acta Crystallographica*, Vol. 4 (July, 1951), pp. 353–362.

[47] Shirane, G., Danner, H., Pavlovic, A., and Pepinsky, R., "Phase Transitions in Ferroelectric $KNbO_3$." *Physical Review*, Vol. 93 (February 15, 1954), pp. 672–673.

[48] Sawaguchi, E., Shirane, G., and Takagi, Y., "Phase Transitions in Lead Zirconate." *Journal of the Physical Society of Japan*, Vol. 6 (September–October, 1951), pp. 333–339.

[49] Shirane, G., Sawaguchi, E., and Takagi, Y., "Dielectric Properties of Lead Zirconate." *Physical Review*, Vol. 84 (November 1, 1951), pp. 476–481.

[50] Jona, F., Shirane, G., and Pepinsky, R., "Optical Study of $PbZrO_3$ and $NaNbO_3$ Single Crystals." *Physical Review*, Vol. 97 (March 15, 1955), pp. 1584–1590.

[51] Vousden, P., "The Structure of Ferroelectric Sodium Niobate at Room Temperature." *Acta Crystallographica*, Vol. 4 (November, 1951), pp. 545–551.

[52] Shirane, G., Newnham, R., and Pepinsky, R., "Dielectric Properties and Phase Transitions of $NaNbO_3$ and (Na, K)NbO_3." *Physical Review*, Vol. 96 (November 1, 1954), pp. 581–588.

[53] Shirane, G., and Pepinsky, R., "Phase Transitions in Antiferroelectric $PbHfO_3$." *Physical Review*, Vol. 91 (August 15, 1953), pp. 812–815.

[54] Shirane, G., "Ferroelectricity and Antiferroelectricity in Ceramic $PbZrO_3$ containing Ba or Sr." *Physical Review*, Vol. 86 (April 15, 1952), pp. 219–227.

[55] Shirane, G., and Hoshino, S., "Crystal Structure of the Ferroelectric Phase in $PbZrO_3$ containing Ba or Ti." *Physical Review*, Vol. 86 (April 15, 1952), pp. 248–249.
"X-Ray Study of Phase Transitions in $PbZrO_3$ containing Ba or Sr," *Acta Crystallograph*, Vol. 7 (Feb., 1954), pp. 203–210.

[56] Sawaguchi, E., "Ferroelectricity versus Antiferroelectricity in the Solid Solution of $PbZrO_3$ and $PbTiO_3$." *Journal of the Physical Society of Japan*, Vol. 8 (September–October, 1953), pp. 615–629.

[57] Nagamiya, T., "On the Theory of the Dielectric, Piezoelectric and Elastic Properties of $NH_4H_2PO_4$." *Progress Theoretical Physics* (Japan), Vol. 7 (March, 1952), pp. 275–284.

[58] Frazer, B. C., Keeling, R. O., Jr., and Pepinsky, R., "Dielectric and X-Ray Investigations of $NH_4H_2PO_4$ and (NH_4, Rb) H_2PO_4." Program and Abstracts, American Crystallographic Association, Camp Tamiment, Pa., June 18, 1952.
Keeling, R. O., Jr., and Pepinsky, R., "An X-Ray Diffraction Study of the Transition in $NH_4H_2PO_4$ at 148°K." *Zeitschrift für Kristallographie*, Vol. 106 (April, 1955), pp. 236–265.

[59] Mason, W. P., and Matthias, B. T., "The Piezoelectric, Dielectric and Elastic Properties of $ND_4D_2PO_4$ (Deuterated ADP)." *Physical Review*, Vol. 88 (November 1, 1952), pp. 477–479.
Mason, W. P., "Properties of a Tetragonal Antiferroelectric Crystal." *Physical Review*, Vol. 88 (November 1, 1952), pp. 480–484.

[60] Busch, G., Kaenzig, W., and Meier, W. M., "Antiseignetteelektrizität von $(NH_4)_2H_3IO_6$." *Helvetica Physica Acta*, Vol. 26 (June 15, 1953), p. 385.

[61] Gränicher, H., Meier, W. M., and Petter, W., "Über die Antiferroelektrizität und die Struktur von $Ag_2H_3IO_6$." *Helvetica Physica Acta*, Vol. 27 (June 30, 1954), pp. 216–217.

[62] Forsbergh, P. W., "Domain Structure and Phase Transition in Barium Titanate." *Physical Review*, Vol. 76 (Oct. 15, 1949), p. 1187.

[63] Kay, H. F., "Preparation and Properties of Crystals of Barium Titanate, $BaTiO_3$." *Acta Crystallographica*, Vol. 1 (November, 1948), pp. 229–230.

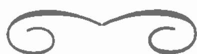
[64] Kay, H. F., and Vousden, P., "Symmetry Changes in Barium Titanate at Low Temperatures and Their Relation to Its Ferroelectric Properties." *Philosophical Magazine*, Vol. 40 (October, 1949), pp. 1019–1040.

[65] Smakula, A., and Sils, V., "Growing of $BaTiO_3$ from Ternary Melts." Progress Report No. XV, Laboratory for Insulation Research, Massachusetts Institute of Technology, pp. 39–40; June, 1954.

[66] Rase, D. E., and Roy, R., "Phase Equilibria in the System

$BaO-TiO_2$." *Journal of the American Ceramic Society*, Vol. 38 (March, 1955), pp. 102–113.
[67] Statton, W. O., "The Phase Diagram of the $BaO-TiO_2$ System." *Journal of Chemical Physics*, Vol. 19 (June, 1951), pp. 33–40.
[68] Burbank, R. D., and Evans, H. T., Jr., "The Crystal Structure of Hexagonal Barium Titanate." *Acta Crystallographica*, Vol. 1 (December, 1948), pp. 330–336.
[69] Rase, D. E., and Roy, R., "The System $BaCl_2-BaTiO_3$." Report for Contract No. DA 36-039-SC-5594, The Pennsylvania State University; July 15, 1952.
[70] Karan, C., and Skinner, B. J., "$BaTiO_3$-KF Phase Diagram." *Journal of Chemical Physics*, Vol. 21 (December, 1953), p. 2225.
[71] Megaw, H. D., "Temperature Changes in the Crystal Structure of Barium Titanium Oxide." *Proceedings of the Royal Society (A)*, Vol. 189 (April 17, 1947), pp. 261–283.
[72] Drougard, M. E., and Young, D. R., "Dielectric Constant Behavior of Single-Domain, Single Crystals of Barium Titanate in the Vicinity of the Curie Point." *Physical Review*, Vol. 95 (September 1, 1954), pp. 1152–1153.
Drougard, M. E., Funk, H. L., and Young, D. R., "Dielectric Constant and Loss Measurements on Barium Titanate Single Crystals While Traversing the Hysteresis Loop." *Journal of Applied Physics*, Vol. 25 (September, 1954), pp. 1166–1169.
[73] Kaenzig, W., and Maikoff, N., "Ist die seignette-elektrische 120°C-Unwandlung von Bariumtitanat erster oder zweiter Art?" *Helvetica Physica Acta*, Vol. 24 (Sept. 20, 1951), p. 343.
[74] Drougard, M. E., Landauer, R., and Young, D. R., "Dielectric Behavior of Barium Titanate in the Paraelectric State." *Physical Review*, Vol. 98 (May 15, 1955), pp. 1010–1014.
[75] Jaynes, E. T., *Ferroelectricity*. Princeton University Press, 1953.
[76] Roberts, S., "Adiabatic Study of the 120°C Transition in Barium Titanate." *Physical Review*, Vol. 85 (March 1, 1952), pp. 925–926.
[77] Merz, W. J., "The Effect of Hydrostatic Pressure on the Curie Point of Barium Titanate Single Crystals." *Physical Review*, Vol. 78 (April 1, 1950), pp. 52–54.
[78] Forsbergh, P. W., Jr., "Effect of a Two-Dimensional Pressure on the Curie Point of Barium Titanate." *Physical Review*, Vol. 93 (February 15, 1954), pp. 686–692.
[79] Shirane, G., and Hoshino, S., "On the Phase Transitions in Lead Titanate." *Journal of the Physical Society of Japan*, Vol. 6 (July–August, 1951), pp. 265–270.
[80] Rogers, H. H., "The Growing of Lead Titanate Crystals and Some of Their Properties." Technical Report No. 56, Laboratory for Insulation Research, Massachusetts Institute of Technology, pp. 1–17; 1952.
[81] Roberts, S., "Dielectric Properties of Lead Zirconate and Barium-Lead Zirconate." *Journal of the American Ceramic Society*, Vol. 32 (February, 1950), pp. 63–66.
[82] Sawaguchi, E., Maniwa, H., and Hoshino, S., "Antiferroelectric Structure of Lead Zirconate." *Physical Review*, Vol. 83 (September 1, 1951), p. 1078.
[83] Jona, F., Shirane, G., Mazzi, F., and Pepinsky, R., "X-Ray and Neutron Diffraction Study of Lead Zirconate." *Abstracts of Thirteenth Annual Pittsburgh Conference*, November 3, 1955.
[84] Shirane, G., and Suzuki, K., "Crystal Structure of $Pb(Zr, Ti)O_3$." *Journal of the Physical Society of Japan*, Vol. 7 (1952), p. 333.
[85] Sawaguchi, E., and Kittaka, T., "Antiferroelectricity and Ferroelectricity in Lead Zirconate." *Journal of the Physical Society of Japan*, Vol. 7 (1952), pp. 336–337.
[86] Kittel, C., "Theory of Antiferroelectric Crystals." *Physical Review*, Vol. 82 (June 1, 1951), pp. 729–732.
[87] Takagi, Y., "Ferroelectricity and Antiferroelectricity of a Crystal Containing Rotatable Polar Molecules." *Physical Review*, Vol. 85 (January 15, 1952), pp. 315–324.
[87(a)] Takagi, Y., "Ferroelectricity and Antiferroelectricity in some Ionic Crystals." *Proceedings of the International Conference of Theoretical Physics, 1953, Kyoto and Tokyo*, pp. 824–838; Science Council of Japan, Ueno Park, Tokyo (1954).
[88] Matthias, B. T., "New Ferroelectric Crystals." *Physical Review*, Vol. 75 (June 1, 1949), p. 1771.
[89] Vousden, P., "A Study of the Unit-Cell Dimensions and Symmetry of Certain Ferroelectric Compounds of Niobium and Tantalum at Room Temperature." *Acta Crystallographica*, Vol. 4 (July, 1951), pp. 373–376.
[90] Hulm, J. K., Matthias, B. T., and Long, E. A., "A Ferroelectric Curie Point in $KTaO_3$ at Very Low Temperatures." *Physical Review*, Vol. 79 (September 1, 1950), pp. 885–886.
[91] Pepinsky, R., Thakur, R., and McCarty, C., "Low Temperature Dielectric Behavior of Potassium Niobate." *Physical Review*, Vol. 86 (May 15, 1952), p. 650.
[92] Triebwasser, S., and Halpern, J., "Curie Constant, Spontaneous Polarization, and Latent Heat in the Ferroelectric Transition in $KNbO_3$." *Physical Review*, Vol. 98 (June 1, 1955), p. 1562.
[93] Triebwasser, S., Private Communication. (June, 1955).
[94] Reisman, A., and Holtzberg, F., "Phase Equilibria in the System $K_2CO_3-Nb_2O_5$ by the Method of Differential Thermal Analysis." *Journal of the American Chemical Society*, Vol. 77 (April 20, 1955), pp. 2115–2119.
[95] Pepinsky, R., "Remarks on Vousden's Structure of Ferroelectric Sodium Niobate." *Acta Crystallographica*, Vol. 5 (March 1952), p. 288.
[96] Cross, L. E., and Nicholson, B. J., "The Optical and Electrical Properties of Single Crystals of Sodium Niobate." *Philosophical Magazine*, Vol. 46 (May, 1955), pp. 453–466.
[97] Nomura, S., "Dielectric Properties of Titanates Containing Sn^{4+} Ions. I," *Journal of the Physical Society of Japan*, Vol. 10 (February, 1955), pp. 112–119.
[98] Graenicher, H., and Jakits, O., "Ueber die dielektrischen Eigenschaften und Phasenumwandlungen bei Mischkristallsystemen vom Perowskittyp." *Supplemento Nuovo Cimento*, Vol 11 (August 10, 1954), pp. 480–520.
[99] Roberts, S., "Dielectric Constants and Polarizabilities of Ions in Simple Crystals and Barium Titanate." *Physical Review*, Vol. 76 (October 15, 1949), pp. 1215–1220.
[100] Roberts, S., "Polarizabilities of Ions in Perovskite-Type Crystals." *Physical Review*, Vol. 81 (March 1, 1951), pp. 865–868.
[101] Pepinsky, R., and Rock, E. J., unpublished.
[102] Megaw, H., and Pepinsky, R., "Ferroelectric and High Dielectric Crystals." Technical Report No. 28, Contract AF33(616)-2133 between the Wright Air Development Center and The Pennsylvania State University; (September, 1953).
[103] Shirane, G., Pepinsky, R., and Frazer, B. C., "X-Ray and Neutron Diffraction Study of Ferroelectric $PbTiO_3$." *Physical Review*, Vol. 97 (February 15, 1955), pp. 1179–1180.
[104] Shirane, G., Pepinsky, R., and Frazer, B. C., "X-Ray and Neutron Diffraction Study of Ferroelectric $PbTiO_3$." *Acta Crystallographica* (1955), in press.
[105] Bystroem, A., "X-Ray Analysis of $Ca_2Sb_2O_7$ and Compounds of Similar Composition." *Arkiv för Kemi Mineralogi och Geologi* Vol. 18A (June, 1944), No. 21.
[106] Ueda, R., and Kobayashi, J., "Antiparallel Dipole Arrangement in Tungsten Trioxide." *Physical Review*, Vol. 91 (September 15, 1953), p. 1565.
[107] Bräkken, H., "Die Kristallstrukturen der Trioxide von Chrom, Molybdän und Wolfram." *Zeitschrift für Kristallographie*, Vol. 78 (July, 1931), pp. 484–488.
[108] Ueda, R., and Ichinokawa, T., "On the Phase Transition of Tungsten Trioxide." *Physical Review*, Vol. 82 (May 15, 1951), pp. 563–564.
[109] Kehl, W. L., Hay, R. G., and Wahl, D., "High Temperature Phases of WO_3." *Physical Review*, Vol. 82 (June 1, 1951), p. 774.
[110] Megaw, H., "Ferroelectricity and Crystal Structure II." *Acta Crystallographica*, Vol. 7 (February, 1954), pp. 187–194.
[111] Pepinsky, R., "Some Aspects of Neutron Single Crystal Analysis." *Science*, Vol. 117 (January 2, 1953), pp. 1–3.
[112] Hastings, J. M., and Corliss, L. M., "Neutron Diffraction," in A. Weissberger, ed., *Physical Methods of Organic Chemistry* (Interscience, New York, 1954), 2nd ed., Chap. 37.
[113] Bacon, G. E., *Neutron Diffraction*. Oxford, Clarendon Press, 1955.
[114] Kaenzig, W., "Röntgenuntersuchungen über die Seignetteelektrisität von Bariumtitanat." *Helvetica Physica Acta*, Vol. 24 (April 10, 1951), pp. 175–216.
[115] Evans, H. T., Jr., "An X-Ray Diffraction Study of Barium Titanate." Technical Report No. 58, Laboratory for Insulation Research, Massachusetts Institute of Technology, January, 1953.
[116] Pepinsky, R., "Method for Cutting and Shaping Fragile Crystals." *Review of Scientific Instruments*, Vol. 24 (May, 1953), p. 403.
[117] Mason, W. P., and Matthias, B. T., "Theoretical Model for Explaining the Ferroelectric Effect in Barium Titanate." *Physical Review*, Vol. 74 (December 1, 1948), pp. 1622–1636.
[118] Eucken, E., and Büchner, A., "Die Dielektrisitätskonstante schwach polarer Kristalle und ihre Temperaturabhängigkeit." *Zeitschrift für physikalische Chemie B*, Vol. 27 (December, 1934), pp. 321–349.
[119] Jaynes, E. T., "Displacement of Oxygen in $BaTiO_3$." *Physical Review*, Vol. 79 (September 15, 1950), pp. 1008–1009.
[120] Takagi, Y., "Theory of the Transition in KH_2PO_4 (I)." *Journal of the Physical Society of Japan*, Vol. 3 (July-August, 1948), pp. 271–272.
"Theory of the Transition in KH_2PO_4 (II)." *Journal of the Physical Society of Japan*, Vol. 3 (July–August, 1948), pp. 273–274.
[121] Yomosa, S., and Nagamiya, T., "The Phase Transition and the Piezoelectric Effect of KH_2PO_4." *Progress of Theoretical Physics* (Japan), Vol. 4 (July–September, 1949), pp. 263–274.
[122] Matthias, B. T., "Phase Transition of $ND_4D_2PO_4$." *Physical Review*, Vol. 85 (January 15, 1952), p. 141.
[123] Wood, E. A., Merz, W. J., and Matthias, B. T., "Polymorphism

in $ND_4D_2PO_4$." *Physical Review*, Vol. 87 (Aug. 1, 1952), p. 544.
[124] Frazer, B. C., "A Phase Transition in $ND_4D_2AsO_4$." *Physical Review*, Vol. 91 (July 1, 1953), p. 246.
[125] Jaffe, H., "Polymorphism of Rochelle Salt." *Physical Review*, Vol. 51 (January 1, 1937), pp. 43–47.
[126] Pepinsky, R., "Hydrogen-Deuterium Replacement in Neutron Diffraction Analysis." Program and Abstracts, Eleventh Annual Pittsburgh Conference on X-Ray and Electron Diffraction, pp. 32–33; November 6, 1953.
[127] Robertson, J. M., *Organic Crystals and Molecules*. Ithaca, Cornell University Press, 1953.
[128] Sundara Rao, R. V. G., Frazer, B. C., and Pepinsky, R., "Some Observations on the Rochelle Salt Problem." Program and Abstracts, Eleventh Annual Pittsburgh Conference on X-Ray and Electron Diffraction, p. 31, November 6, 1953, p. 31.
[129] Scholz, H., Dissertation, Göttingen, (1940). Reported by G. Joos in *FIAT Review of German Science*, Vol. 9. *Physik der festen Körper, Part II* (July, 1948), pp. 54–55.
[130] Jaffe, H., Brush Development Company, Final Report to U. S. Army Signal Corps, Contract No. W28-003-SC-1583.
[131] Cf. reference 9, p. 233.
[132] Landau, L., and Lifshitz, E., "On the Theory of the Dispersion of Magnetic Permeability in Ferromagnetic Bodies." *Physikalische Zeitschrift der Sowjetunion*, Vol. 8 (1935), pp. 153–169.
[133] Jaccard, C., Kaenzig, W., and Peter, M., "Das Verhalten von kolloidalen Seignetteelektrika I, Kaliumphosphat KH_2PO_4." *Helvetica Physica Acta*, Vol. 26 (Sept. 15, 1953), pp. 521–544.
[134] Ubbelohde, A. R., and Woodward, I., "Laue Photographs of Sub-crystalline Regions in 'Hybrid' Crystals of Potassium Dihydrogen Phosphate." *Nature*, Vol. 156 (July 7, 1945), p. 20.
[135] Zwicker, B., and Scherrer, P., "Elektrooptische Eigenschaften der seignetteelektrischen Kristalle KH_2PO_4 und KD_2PO_4." *Helvetica Physica Acta*, Vol. 17 (Sept. 6, 1944), pp. 346–373.
[136] de Quervain, M., and Zwicker, B., "Beobachtungen über elektrische Elementarbezirke an Seignetteelektrika." *Helvetica Physica Acta*, Vol. 16 (June 30, 1943), pp. 216–218.
[137] Miyake, S., "Effect of Temperature Variation and Electric Field on X-Ray Intensity Reflected from Rochelle Salt Crystal. Part I: Temperature Effect." *Proceedings of the Physico-Mathematical Society of Japan*, Vol. 23 (May, 1941), pp. 377–395.
"Effect of Temperature Variation and Electric Field on X-Ray Intensity Reflected from Rochelle Salt Crystal. Part II: The Effect of Electric Field." *Proceedings of the Physico-Mathematical Society of Japan*, Vol. 23 (October, 1941), pp. 810–819.
[138] Staub, H., "Untersuchung der dielektrischen Eigenschaften des Seignettesalzes mittels Röntgenstrahlen." *Physikalische Zeitschrift*, Vol. 34 (1933), No. 7, pp. 292–296.
"Ueber den Nachweis des inneren elektrischen Feldes des Seignettesalzes mittels Röntgenstrahlen." *Physikalische Zeitschrift*, Vol. 35 (1934), No. 18, pp. 720–725.
[139] Blattner, H., Kaenzig, W., Merz, W., and Sutter, H., "Die Domänenstruktur von $BaTiO_3$-Kristallen." *Helvetica Physica Acta*, Vol. 21 (August 10, 1948), pp. 207–209.
[140] Hooton, J. A., and Merz, W. J., "Etch Patterns and Ferroelectric Domains in $BaTiO_3$ Single Crystals." *Physical Review*, Vol. 98 (April 15, 1955), pp. 409–413.
[141] Kaenzig, W., and Sommerhalder, R., "Das Verhalten von kolloidalen Seignetteelektrika II, Theoretische Betrachtungen." *Helvetica Physica Acta*, Vol. 26 (November 16, 1953), pp. 603–610.
[142] Kittel, C., "Domain Boundary Motion in Ferroelectric Crystals and the Dielectric Constant at High Frequency." *Physical Review*, Vol. 83 (July 15, 1951), p. 458.
[143] Cohen, M. H., "Ferroelectricity *versus* Antiferroelectricity in Barium Titanate." *Physical Review*, Vol. 84 (Oct. 15, 1951), p. 369.
[144] Megaw, H. D., "Origin of Ferroelectricity in Barium Titanate and Other Perovskite-Type Crystals." *Acta Crystallographica*, Vol. 5 (November, 1952), pp. 739–749.
[145] Mason, W. P., and Wick, R. F., "Ferroelectrics and the Dielectric Amplifier." PROCEEDINGS OF THE IRE, Vol. 42 (November, 1954), pp. 1606–1620.
[146] Anderson, J. R., "Ferroelectric Storage Elements for Digital Computers and Switching Systems." *Electrical Engineering*, Vol. 71 (October, 1952), pp. 916–922.
[147] Marutake, M., "The Effect of Electric Field on the Domain Structures in Rochelle Salt." *Journal of the Physical Society of Japan*, Vol. 7 (January–February, 1952), pp. 25–29.
[148] Anliker, M., Brugger, H. R., and Kaenzig, W., "Das Verhalten von kolloidalen Seignetteelektrika III, Bariumtitanat $BaTiO_3$." *Helvetica Physica Acta*, Vol. 27 (May 1, 1954), pp. 99–124.
[149] Bantle, W., "Die spezifische Wärme seignette-elektrischer Substanzen. Dielektrische Messungen an KD_2PO_4-Kristallen." *Helvetica Physica Acta*, Vol. 15 (July 8, 1942), pp. 373–404.
[150] Stephenson, C. C., and Hooley, J. G., "The Heat Capacity of Potassium Dihydrogen Phosphate from 15 to 300°K. The Anomaly at the Curie Temperature." *Journal of the American Chemical Society*, Vol. 66 (August, 1944), pp. 1397–1401.
[151] Blattner, H., Kaenzig, W., and Merz, W., "Herstellung und Untersuchung von $BaTiO_3$ Einkristallen." *Helvetica Physica Acta*, Vol. 22 (February 15, 1949), pp. 35–65.
[152] Danner, H., and Pepinsky, R., "Specific Heat Anomaly of Ferroelectric cadmium Niobate at the Curie Temperature." *Physical Review*, Vol. 99 (August 15, 1955), pp. 1215–1217.
[153] Shirane, G., and Takeda, A., "Transition Energy and Volume Change at Three Transitions in Barium Titanate." *Journal of the Physical Society of Japan*, Vol. 7 (Jan. Feb., 1952), pp. 1–4.
[154] Volger, J., "The Specific Heat of Barium Titanate between 100°K and 410°K." *Philips Research Reports*, Vol. 7 (1952), pp. 21–27.
[155] Todd, S. S., and Lorenson, R. E., "Heat Capacities at Low Temperatures and Entropies at 298.16°K of Metatitanates of Barium and Strontium." *Journal of the American Chemical Society*, Vol. 74 (April, 1952), pp. 2043–2045.
[156] Megaw, H. D., "Crystal Structure of Double Oxides of the Perovskite-Type." *Proceedings of the Physical Society*, Vol. 58 (March, 1946), pp. 133–152.
[157] Goodman, G., "Ferroelectric Properties of Lead Metaniobate." *Journal of the American Ceramic Society*, Vol. 36 (November, 1953), pp. 368–372.
[158] Shirane, G., and Suzuki, K., "On the Phase Transition in Barium-Lead Titanate, I." *Journal of the Physical Society of Japan*, Vol. 6 (July–August, 1951), pp. 274–278.
Shirane, G. and Takeda, A., "On the Phase Transition in Barium-Lead Titanate, II." *Journal of the Physical Society of Japan*, Vol. 6 (September–October, 1951), pp. 329–332.
[159] Shirane, G., and Takeda, A., "Phase Transitions in Solid Solutions of $PbZrO_3$, I. Small Concentrations of $PbTiO_3$." *Journal of the Physical Society of Japan*, Vol. 7, (January–February, 1952), pp. 5–11.
Shirane, G., Suzuki, K., and Takeda, A., "Phase Transitions in Solid Solutions of $PbZrO_3$ and $PbTiO_3$, II. X-Ray Study." *Journal of the Physical Society of Japan*, Vol. 7 (January–February, 1952), pp. 12–18.
[160] Slater, J. C., "The Lorentz Correction in Barium Titanate." *Physical Review*, Vol. 78 (June 15, 1950), pp. 748–761.
[161] Holden, A. N., Matthias, B. T., Merz, W. J., and Remeika, J. P., "New Class of Ferroelectrics." *Physical Review*, Vol. 98 (April 15, 1955), p. 546.
"Guanidine Aluminum Sulfate Hexahydrate—A New Ferroelectric Material." *Physical Review*, Vol. 99 (July 15, 1955), p. 626.
[162] Merz, W. J., Remeika, J. P., Holden, A. N., and Matthias, B. T., "Electrical Properties of Guanidine Aluminum Sulfate Hexahydrate (G.A.S.H.) and Some Isomorphs." *Physical Review*, Vol. 99 (July 15, 1955), p. 626.
[163] Rushman, D. F., and Strivens, M. A., "The Permittivity of Polycrystals of the Perovskite Type." *Transactions of the Faraday Society*, Vol. 42A (1946), pp. 231–238.
[164] Rundle, R. E., and Parasol, M., "O-H Stretching Frequencies in Very Short and Possibly Symmetrical Hydrogen Bonds." *Journal of Chemical Physics*, Vol. 20 (September, 1952), pp. 1487–1488.
[165] Stephenson, C. C., Corbella, J. M., and Russell, L. A., "Transition Temperatures In Some Dihydrogen and Dideutero Phosphates and Arsenates and Their Solid Solutions." *Journal of Chemical Physics*, Vol. 21 (June, 1953), p. 1110.
[166] Kaenzig, W., "Space Charge Layer near the Surface of a Ferroelectric." *Physical Review*, Vol. 98 (April 15, 1955), pp. 549–550.



History of Semiconductor Research*

G. L. PEARSON† AND W. H. BRATTAIN†

Summary—**This paper presents a running story of semiconductor research from its earliest beginnings up to the present day, with special emphasis on the inception of new ideas and the resolution of older discrepancies. At several points in the story, short interludes are taken to fill in the status as of that time. Semiconductor research began quite inconspicuously about 120 years ago with some observations on the electrical properties of silver suphide. Progress was very slow for the next 50 years and then, about 1885, a mild interest developed with the discovery of point contact rectifiers. These devices were used as detectors until displaced by the vacuum tube around 1915. Development of selenium and cuprous oxide rectifiers about 1930 revived interest and the publication of a good theory of semiconductors in 1931 added still more momentum. The next period of active interest came around World War II when the catwhisker diode was revived and developed into an excellent radar detector. The announcement of the transistor in 1948 gave this field of research such a boost that it has become a real giant in the last few years and semiconductor electronics now rates as a major field of endeavor.**

INTRODUCTION

THE HISTORY of semiconductor research is a very interesting story, in fact it is so interesting that the present authors doubt that they are really doing it justice. One quickly realizes that a thorough job requires a prohibitive amount of time and energy. Two good review articles, one by K. Lark-Horowitz [1] and one by E. W. Herold [2], have recently been published. Each author takes a somewhat different viewpoint and the first mentioned article includes a comprehensive list of references up to 1951. There is, of course, no point in reproducing such a bibliography or in simply rewriting the material in these publications.

We have chosen to emphasize in the present article the interplay of ideas in the development of semiconductor research which has impressed us during our long connection with semiconductors starting in the early 1930's, and have chosen to write from this viewpoint. We are likewise impressed with the length of time required for certain key ideas to become accepted as well as how long some misconceptions lasted. There should be an object lesson in this for future workers. We realize, however, that our perspective becomes clouded as the story approaches the present time. This not only makes the writing more difficult, but likewise makes it almost impossible to distinguish between the current key ideas and misconceptions of today. The lesson is, therefore, of no particular value except possibly to the rare individual who can see the present in its true perspective.

If any of our friends feel slighted because no mention is made of their work, let them remember that the material to be covered is actually too large for one short article and that merely listing all the published papers during the last five years would fill a good sized book [3]. For the sake of brevity, our bibliography often lists one comprehensive article rather than the numerous original contributions. Finally, we freely admit that the latter part of the story draws heavily on the work carried out by our colleagues at the Bell Telephone Laboratories.

* Original manuscript received by the IRE, August 25, 1955.
† Bell Telephone Labs., Inc., Murray Hill, N. J.

EARLY HISTORY OF SEMICONDUCTORS

Probably the first significant observation in the semiconductor field was made by Michael Faraday [4] in 1833 when he found that silver sulfide had a negative temperature coefficient of resistance. This characteristic set it apart from other conductors (metals) whose resistance increased with increase in temperature. The observation that a photovoltage could be produced by shining light on the surface of one electrode in an electrolyte by Becquerel [5] in 1839 was another early contribution. The next significant step was W. Smith's discovery [6] in 1873 that resistance of selenium could be reduced by shining light on it (photo-conductivity).

In 1874 came the discovery that contacts between certain materials would rectify, or more precisely, that the resistance did not obey Ohm's law but depended on the magnitude and sign of the applied voltage. F. Braun [7] observed this for contacts between metals and various sulfides such as galena and pyrites, and A. Schuster [8] for contacts between untarnished and tarnished (probably copper oxide) copper wires. Some workers thought that this phenomenon was a thermal effect, and this idea persisted up to about 1906. The first photoelement of the modern barrier layer type was made of selenium by W. G. Adams and R. E. Day [9] in 1876. While continuing this work on selenium, C. E. Fritts [10] in 1883 produced the first large area dry rectifier. Thus by 1885 four of the fundamental properties of semiconductors—(1) negative temperature coefficient of resistance, (2) rectification, (3) photoconductivity, and (4) photoelectromotive force—had been observed, although not all on the same material.

The demonstration of the existence of radio waves by H. Hertz in 1888 created a potential demand for a suitable detector, but it was not realized until 1904 that semiconductor rectifiers were well suited for this purpose. J. C. Bose [11], H. H. C. Dunwoody [12], L. W. Austin [13], and G. W. Pierce [14] found that point contacts (cat whiskers) on galena, silicon carbide, tellurium, silicon, etc. were good detectors of radio waves. Silicon detectors were found by experience to be most stable, while galena detectors had the best sensitivity. It was G. W. Pierce, by the way, who went to a great deal of

effort at this time to show that these devices did not operate on a thermal basis. With the advent of the vacuum tube at about this time, interest in the point contact detector lagged and little of scientific interest was contributed on such detectors for a number of years.

The next period, beginning in the 1920's, saw the development of barrier layer rectifiers and photocells as commercial devices. Some good science and a large amount of art was involved. As we shall see later, the real understanding had to await the help of quantum mechanics and its application to the understanding of solids. The major contributions in this period were the development of the copper oxide rectifier and photocell by L. O. Grondahl and P. H. Geiger [15] and the almost parallel developments using selenium, described by B. Lange [16]. Commercial use of these devices as rectifiers, battery chargers, photographic exposure meters, and in the case of copper oxide as modulators and nonlinear circuit elements, created demand for better scientific understanding of underlying phenomena.

Though not generally accepted until the 1930's, it now appeared that rectification and photovoltage were surface properties occurring at the interface between the semiconductor and metal contacts or between two semiconductors. On the other hand, the mechanism of electrical conductivity which gave rise to the negative temperature coefficient of resistivity and the change of resistance under the influence of light (photoconductivity) appeared to be a body property within the homogeneous semiconductor. Various workers now proposed that electrical conduction within the body obeyed Ohm's law and that photoconductivity was due to an increase in the number of carriers by light excitation. The contributions of B. Gudden and R. W. Pohl [17] (beginning in 1921) on photoconductivity in the alkali halides should be mentioned in this connection. While these materials were more nearly insulators than semiconductors, they were nevertheless much less complicated than copper oxide or selenium. Also they could be fabricated into large *single crystals*, thus giving a more ideal solid for study. Most important of all, however, was the division of the problem into two parts, one of surface phenomena and the other of body properties. In what follows we will consider developments first in one category and then the other, switching back and forth as the story unfolds.

Developments in the Understanding of Body Properties up to 1942

It was natural that semiconductor research should borrow from the experimental progress and understanding of electrical conduction in metals. The discovery by E. H. Hall [18] in 1879 that electrons flowing in a metal could be deflected by a magnetic field perpendicular to the direction of current flow was, of course, a landmark.[1] The interpretation of this effect to give the density of charge carriers was a very important tool for investigating conduction processes. Likewise the interpretation of the ratio of Hall field to electric field to give the mobility of the charge carriers was of great importance.

Workers in this field were pleased that the Hall effect in metals had the right sign (i.e., negative) which indicated that the carriers were negative charges and undoubtedly electrons. They were greatly disturbed, however, when it was found that some borderline metals gave positive Hall coefficients thus indicating positive charge carriers. Furthermore, the measured mobilities at room temperature were such as to rule out positive ions. This indicated, if you will, that there were apparently positive electrons long before anyone suspected the existence of positrons or holes.

When Hall effect measurements were performed on semiconductors several conclusions were at once apparent [19, 20, 21]. The number of charge carriers were much smaller than in most metals and the mobilities were somewhat larger. Some semiconductors, for example copper oxide, have a positive Hall coefficient, while others like zinc oxide had a negative coefficient. Disregarding for the moment the primary difficulty of positive carriers with mobilities of the same order of magnitude as electrons, these measurements were a great step forward. They enabled the researchers of this time to separate, conceptually at least, the conductivity σ into its two component parts: (1) the density n of carriers and (2) their mobility μ. The familiar relation $\sigma = en\mu$ had been shown to hold for electronic semiconductors.

It was found that the mobility was a property of the solid or crystal lattice and that for a given lattice it decreased with increase in temperature. On the other hand, the density of carriers increased *very rapidly* with temperature for any given semiconducting specimen, such as copper oxide. In addition it was found that n varied over *wide limits* at a given temperature from specimen to specimen of the same semiconducting material. In other words, n was both temperature and structure sensitive.

The difference between metals and semiconductors had now been pinned down. In metals the mobility was only slightly structure sensitive and it decreased with increase in temperature, whereas the number of carriers was almost constant and of the order of one per atom of solid. The resistivity of metals had a positive temperature coefficient because of the change of mobility with temperature. In semiconductors the mobility varied with temperature as in metals, but the carrier density was orders of magnitude less and increased rapidly with temperature. The exact value of n at any given temperature was very structure sensitive; at room temperature it varied in the range of one charge carrier per thousand atoms of the solid to one per hundred million. The electrical resistivity of semiconductors was therefore a structure sensitive or *extrinsic* property. The negative

[1] The electron itself was not clearly identified until eighteen years later when J. J. Thomson measured the ratio of charge to mass of the cathode rays.

temperature coefficient was primarily due to changes in n, the variation in number usually swamping out the dependence of mobility on temperature.

At high temperatures it was observed that the density of charge carriers was no longer structure sensitive but equal for all samples of a given semiconductor. In this region the density of carriers increased even more rapidly with temperature and in general obeyed a simple law. The logarithm of the density was a linear function of the reciprocal of the absolute temperature. It was of course natural to interpret this behavior in terms of an activation energy proportional to the slope of this straight line. This activation energy turned out to be the same for any given semiconductor, such as copper oxide, and therefore was an *intrinsic* property of the solid. Attempts were also made to interpret the low temperature characteristics in terms of a second and smaller activation energy. With material available at the time, however, the results were questionable since this activation energy was found to be structure sensitive and, in general, a function of temperature.

Other important contributions were made in this period by chemically-minded workers such as Carl Wagner [22, 23]. It was found that in compound semiconductors the sign of the Hall coefficient could be correlated with small deviations from the stoichiometric composition. If, for example, in copper oxide there is a deficiency of copper in the solid, say a few parts per million below that necessary to satisfy exactly the chemical formula Cu_2O, then the sign of the Hall coefficient at low temperatures is positive, and this was called a *defect* semiconductor. If the compound is *over-oxidized*, it has a *positive* Hall coefficient. If on the other hand there is usually excess metal, as in zinc oxide (ZnO), then the Hall effect is negative, and this was called an *excess* semiconductor. The compound is slightly *reduced*. Finally at high temperatures in the intrinsic region all semiconductors were found to have negative Hall coefficients.

In the meantime, quantum mechanics was making great strides in interpreting the behavior of electrons in atoms and molecules. The experiments of C. J. Davisson and L. H. Germer and those of G. P. Thomson demonstrated the wave nature of electrons. The concepts of energy levels with one electron per level, electron spin, Pauli exclusion principle, and Fermi-Dirac statistics all led to a clarification of metallic conduction by A. Sommerfeld [24] and his co-workers. The concept of electrons as Bloch waves in a crystal contributed greatly to understanding the behavior of electrons in all types of solids.

In 1931, A. H. Wilson [25] presented a quantum mechanical model of a solid semiconductor which has since become fundamental for understanding the behavior of semiconductors. In retrospect it turns out that the Wilson model can be described from a number of different viewpoints. One of these is the picture of the electrons as waves throughout the solid or *crystal lattice*. At certain frequencies there is interference between these waves and the regularity of the crystal lattice, just as had been found for X-rays. Electron waves of such frequency cannot exist in the lattice. From the relation between frequency and energy, certain energies were thus excluded. This led to the concept of *energy bands* in the solid.

At low temperatures the electrons in the lattice fill the lowest energy bands, one electron of each spin per level, until there is just enough charge to make the entire lattice neutral. When the band structure is such that the last or *valence* electrons completely fill their band, there is a gap in energy E_g from the last filled level to the next possible energy site. Now the process of conduction involves accelerating the electrons by means of an electric field, thus adding to their energy. In order for this to happen there must be, in the quantum mechanical sense, empty energy levels available to receive the electrons. If there are none available (energy gap) then the electron waves cannot be made to drift in the direction of the field. Under these conditions, for every wave traveling in one direction through the lattice there is another in the opposite direction, and this situation cannot be altered except by a discrete jump in electron energy equal to the energy gap.

At any finite temperature there is of course a certain amount of thermal excitation, maintaining some electrons above the gap and thus leaving an equal number of vacant energy levels below the gap. This picture was ideal for explaining the *intrinsic* conductivity in semiconductors at the higher temperatures. The thermally excited electrons would be free to conduct and their density would depend on temperature as found experimentally, the energy of excitation being equal to the gap energy.

The above description is of course over-simplified. Quantum mechanics predicted that vacant sites, or *holes*, should also be free to conduct with possibly as great a mobility as the excited electrons. For each electron excited there would therefore be two carriers available and the product of the carrier densities would increase with temperature according to an activation energy equal to the energy gap. The apparent activation energy derived from the experimental data as described above is thus actually equal to one-half the energy gap. With this refinement it is seen that the Hall effect should be zero for equal mobilities of holes and electrons. The fact that a negative Hall sign was always obtained in the intrinsic range meant that the electrons are more mobile than the holes. The Hall effect measured, in a sense, the difference in mobilities whereas the conductivity measured, in a like sense, the average mobility.

The low-temperature, structure-sensitive part of the conductivity data was explained by the model as follows. In cases where the semiconducting lattice is not exactly perfect, such as a small Cu atom deficiency in copper oxide, the vacant copper site is an irregularity in the lattice resulting in one too few electrons to fill the

energy band. This missing electron (hole) is usually bound, at very low temperatures, to the vacant copper position but can be thermally excited so that it becomes a charge carrier. The energy necessary to excite the holes is much less than the energy gap. The hole, when freed, is equivalent to an electron at the defect and so the defect is called the *acceptor state*. Such a state is, of course, localized at the defect and is in this sense a higher energy level to which an electron in the filled band can be excited, leaving a free-to-move hole in the valence band. When thus activated the acceptor site has a localized negative charge.

The explanation of excess conductivity is quite similar. The excess Zn atom in zinc oxide has an extra electron. The energy band is full, but the extra electron can be excited to the higher band where it is free to conduct, leaving a localized positive charge in a *donor state*.

The same general picture of conduction in the body of a semiconducting material can be obtained from another viewpoint (tight binding approximation). Chemical binding in solids is electronic in nature and is now quite well understood quantum mechanically. Such binding can be more or less arbitrarily divided into the following three classes: (1) *ionic* as in NaCl, (2) *covalent* as in H_2 where energy is gained by sharing electrons and pairing spins, and (3) *metallic* where, in a sense, positive ions float in a cloud of electrons. Semiconductors generally range from almost complete covalent binding to various admixtures of covalent and ionic. In any event, the valency is complete and there are just sufficient electrons in the perfect lattice to satisfy all the bonds. It takes considerable energy E_g to free an electron from a given bond. Such excitation leaves a hole as well as producing a free electron. Defects consist of imperfections in the structure where there is either one too many or one too few electrons to satisfy the bonding. Small energies are necessary to *free* the electron or hole from the site of the defect. The filled band mentioned in our first picture is the *valence* band, the next band of possible energy levels in the lattice is the *conduction* band and is more or less empty.

Both pictures represent different quantum mechanical approximations to the actual state of affairs. If the calculations could be carried out with sufficient accuracy in each case, the final results would agree. Since this cannot be done in practice, one gets a better mental picture by thinking in terms of both pictures instead of using one to the exclusion of the other.

In the above description of Wilson's theory of semiconductors we have considered only the thermal mechanism for activating free electrons and free holes. Light quanta are equally effective for this purpose, and the phenomena of photoconductivity are easily described from either the energy band or the chemical valence viewpoint. The energy of a light quantum increases with decrease in wavelength. Light of sufficiently short wavelength has large enough energy quanta to produce hole-electron pairs in intrinsic semiconductors. Longer wavelengths can excite holes or electrons only in the impurity range, and of course there will be some free carrier absorption all the way out to infinite wave length (dc) but this can be very small when the number of free carriers is small.

It should be pointed out that all that has been said here about the above model was not immediately apparent to the workers in the field after reading Wilson's paper. In fact it took about fifteen years for the full light to dawn. One of the blind spots[2] arose from the fact that it was much simpler to consider a semiconductor with one type of defect or the other, rather than both donors and acceptors at the same time. This was a case of plain lazy thinking on the part of many investigators. It is now perfectly obvious that, when one is concerned with defects of the order of one part per million, it would be surprising to find, in nature, defects of only one kind. The statistics of this more general case were discussed by J. H. deBoer and Van Geel [26]. Also it was thought at first that the electron wave viewpoint and the valence bond picture were not consistent with each other.

H. Dember [27] in 1931 found a potential difference between illuminated and unilluminated portions of a cuprous oxide wafer. J. Frenkel [28] explained this effect on the basis that the light created hole-electron pairs, pointing out that this potential difference arises from the unequal diffusion coefficients of the holes and the electrons. Frenkel also explained the photomagnetoelectric effect first discovered by I. K. Kikoin and M. M. Naskov [29] in 1934 by a similar analysis. The explanation of these effects involved both the majority and the minority carriers. The fact that minority carriers might play an important role in the understanding of semiconductor phenomena was more or less overlooked by other investigators. As we shall see later, this was another blind spot.

Developments in the Understanding of Surface Properties Up to 1942

As late as 1935 some investigators thought they had evidence that rectification was a body property. The concept of rectification at a surface may seem obvious now, but it was a controversial subject at that time.

Due to their commercial uses, the best known rectifiers at that time were Cu_2O grown on copper and selenium melted and spread on a metal washer. In each case a second contact was added to complete the electrical circuit. For the cuprous oxide unit this was graphite or aquadag plus a lead washer held under pressure. In the case of selenium the second contact was a low melting-point alloy sprayed on the selenium surface with a great amount of "art" involved in the process.

Another commercial product was the silicon carbide

[2] There were obviously many blind spots in the working concepts about semiconductors in the nineteen thirties. Since the authors were part of the group that was so blind, they offer no apology for being so blunt in this matter.

voltage regulator or lightning protector [30]. This was a symmetrical device in that it did not rectify, it was nonohmic in that it had a very high resistance at low voltages but would pass very large currents at higher voltages. It was made by compressing SiC granules in a ceramic flux and heat treating at elevated temperatures. Sprayed or evaporated metal contacts were applied on either side of the fired discs.

In the case of the large area rectifiers there was no geometrical asymmetry between the two electrodes as in the case of the point contact rectifier or catwhisker radio detector. Assuming rectification at contacts between different materials, why did these configurations not leave one with two opposing rectifiers having a symmetrical current voltage characteristic? Gradually it was realized that any contact between a metal and a semiconductor would rectify; some more and some less, depending on the method of fabrication. Potentiometric probe measurements indicated that the material comprising the body of the semiconductor obeyed Ohm's law. Silicon carbide was an especially difficult material to study since it was nearly impossible to make a low resistance potential probe contact. Single pieces of SiC, while generally found to be ohmic, had internal boundaries across which large nonohmic potential drops were observed.

While the above mentioned puzzles were being untangled, it became clear that rectification and nonohmic properties must be surface effects. Although rectifier production was an art and not a science, most contacts to semiconductors were found to rectify to some degree. The art consisted in making one contact as good a rectifier as possible and the other as poor (*i.e.*, as ohmic) as possible. Two such combinations in parallel opposed was an obvious model for the SiC device, the current flowing either way in the easy direction through one unit or the other.

The big question in the minds of all active workers in the field at this time was: What is the scientific explanation of rectification at the surface? Very few were concerned with the fact that explaining the ohmic contact might be equally difficult. In copper oxide the rectifying surface was obviously at the interface between the copper base and the oxide surface layer. In selenium, after much travail, it was shown to be at the sprayed interface. The direction of easy flow in each case was that of positive charges moving from the semiconductor across the rectifying contact. Both copper oxide and selenium had positive Hall coefficients, and it was found that other semiconductors with negative Hall coefficients tended to rectify in the opposite direction.

It was noted that rectifying contacts, when illuminated by a bright light, produced a photovoltage and that in general good photo emf cells were good rectifiers, although some experts disagreed on this correlation as late as 1932. Where it could be shown that the photovoltage was definitely at one surface and not at the other (this was often very difficult to determine), it was found that defect semiconductors became positively charged, while excess semiconductors became negatively charged. Another effect under active investigation at this time was that of thermoelectric emf. The thermal voltages in semiconductors are some orders of magnitude larger than those in metals, so that to a good approximation almost any metal may be taken as the zero of reference. It was found that semiconductors having a given sign of the Hall effect all had a like sign of the thermal voltage with reference to metals and that those of opposite Hall sign had the opposite sign of the thermal voltage. Although it was not understood why there should be mobile carriers with a positive sign (holes) in addition to the well-established negative carriers (electrons), nevertheless the signs of the Hall coefficient, rectification, photovoltage, and thermal voltage were all consistent with the concept that some semiconductors had negative carriers while others had positive ones.

One quantum effect considered theoretically in this period was the ability of electron waves to penetrate potential barriers. This phenomenon was seized upon in an attempt to explain rectification. A number of calculations were carried out for various shaped barriers but the predicted sign of rectification did not agree with experiment. It was well known at this time that surfaces of metals must possess charge double layers. Because of the large density of free electrons these double layers could not be more than about one Angstrom (10^{-8} cm) in thickness. When different metals were put in contact these double layers interacted to give just the proper potential difference across the contact so that there was no net flow of electrons in either direction across the boundary, providing that the metals were at the same temperature (*thermal equilibrium*). These double layers of the order 10^{-8} cm in thickness offered no barrier since the electron waves could easily penetrate them.

W. Schottky [31], N. F. Mott [32] and B. Davydov [33] suggested in 1939 that, due to the low density of carriers in a semiconductor, any such double layer must penetrate to depths of the order of 10^{-4} cm. Simple electrostatic reasoning (Poisson's equation) showed that in such an arrangement, most of the potential difference would be spread out on the semiconductor side of the barrier. It is now easy to see that this situation can explain rectification since penetration of such a broad potential barrier by electrons must be insignificant. Consider a metal contact to an excess (electron) semiconductor and assume that the contact potential difference is such that the charge on the semiconductor is positive and that on the metal negative. The result is a potential hill for electrons between metal and semiconductor.

This situation is almost exactly analogous to two hot cathodes in a vacuum providing both are at the same temperature. In equilibrium (no applied voltage), the contact potential is such as to prevent any net flow in either direction. If a voltage is now applied with such a polarity as to make it more difficult for electrons to go

from the lower work function cathode (1) to the higher work function cathode (2), that is (1) is made more positive than (2), the net current will be electrons from (2) to (1) since the counter flow from (1) to (2) has been reduced. As the voltage is increased the current saturates at the total emission value of cathode (2). When the polarity is reversed, the current will at first increase exponentially with voltage since the total emission from cathode (1), with its lower work function, can be orders of magnitude larger than for (2). This is the direction of *easy flow.* This analog has the proper sign to explain the experimental results for excess semiconductors.

For the case of a defect semiconductor where the charge carriers are holes, the picture is analogous with a reversal of the sign of rectification. The transfer of charge in this case is due to the flow of electrons into or out of the valence band and is entirely equivalent to turning the whole picture upside down and considering the vacant places in the electron distribution of the metal as holes. On basis of above considerations, degree of rectification depends on contact potential difference and if contact potential difference is of wrong sign, no rectification will occur. That is, low work function metals in contact with excess semiconductors should not rectify and vice versa for defect semiconductors.

One might be inclined to think that the theory was now in good order. However, while in some cases on cuprous oxide and selenium the low work function metals made better rectifiers than the higher work function metals with the various metals falling in the more or less proper order, the amount of variation and degree of rectification obtained experimentally was far less than predicted by theory [45]. As we shall see later there was a naive assumption hidden here, another blind spot. It should be emphasized that this was a tremendous advance in the theory and was an important building block for future progress.

B. Davydov [34] suggested that rectification might occur at the boundary between an excess and a defect semiconductor. He thought this might be the explanation in copper oxide and suggested that in a thin layer of the oxide near the copper the conductivity was of excess type instead of defect as in the remainder of the oxide layer. In this picture the contacts to both sides of the oxide layer were ohmic and the rectification occurred at the interface between excess and defect layers of the semiconductor. In working out the theory of this effect he recognized the importance of the minority carriers, *i.e.*, holes in the excess layer and electrons in the defect layer. In a companion paper [35] on the photoelectromotive force in semiconductors Davydov points out that it is also necessary in this case to consider the role of the minority carrier. With the emphasis on the success of the space charge theory of rectification these two theoretical papers attracted little attention from other investigators and this blind spot regarding the role of the minority carrier continued to persist.

After having exploited the longer radio wavelengths, research workers began in the nineteen thirties to restudy the short wavelengths which Hertz used in his original experiments and to proceed to even shorter wavelengths. They quickly found that ordinary vacuum tubes were of little use in this region. Some new type of detector was necessary for this purpose. It was only natural for those who had started with crystal detectors to return for a second look. This thought apparently occurred to a number of workers at the time. G. C. Southworth, for one, remembers visiting the Cortlandt Street (New York City) radio market where he ferreted out some old silicon detectors, then almost obsolete, for this purpose.

The next step, initiated by R. S. Ohl about 1935, was a very significant one. In attempting to improve these old detectors he turned to the chemists and metallurgists to obtain purer silicon. First R. O. Grisdale prepared some melts from the purest silicon obtainable and later J. H. Scaff, H. C. Theuerer and E. E. Schumacher [36] conducted a thorough investigation into the whole silicon materials problem. Improvements came rapidly. Not only did silicon detectors become practical devices (far removed from the old catwhisker type), thus making radar feasible in World War II, but important scientific discoveries were made. To start with, the metallurgists learned how to make silicon of either excess or defect type.

Since the easy flow direction for a point contact on excess silicon occurred when the silicon was made negative, it was termed *n*-type and the oppositely poled defect material was termed *p*-type. One dreads to think what would have happened if these investigators had taken the polarity of the metal point instead. These designations have now become universal terms, *n* for electrons and *p* for holes [37]. One of the melts made by Scaff and Theuerer was *n*-type at one end of the melt and *p*-type at the other, with a rather sharp boundary where the two met. A section cut from this melt perpendicular to and including the interface boundary was found by Ohl [38] to be an excellent rectifier and to exhibit a phenomenal photoelectromotive force. This was the *p-n* junction which has developed into a most important circuit element in our present-day semiconductor electronics.

Another development of great significance arose from the conviction of Scaff, Theuerer and Schumacher [36] that segregation of impurities was taking place in the silicon ingots and that both the *n*-type and *p*-type impurities tended to remain in the molten silicon as the ingot solidified from one end. They further deduced that this tendency was greater for one type than for the other. Thus if both were present in about equal amounts in the original molten silicon and the melt were allowed to freeze progressively from one end to the other, the resulting solid would be *p*-type on one end and *n*-type on the other. These workers went on to isolate these impurities and showed that those from the third column of the periodic table are acceptors and that those from

the fifth column are donors. They recognized that impurities from the third column could compensate those from the fifth column, in other words, that electron or hole concentrations were determined by the net difference in impurity concentration as suggested by Wilson's theory.

E. Merritt [39] in 1925 found rectification between a metal point contact and germanium, the next element below silicon in the fourth column of the periodic table, and the Siemens and Halske Works were developing germanium diodes as early as 1941.[3]

It is appropriate at this point to discuss the high-frequency limitations of rectifiers and in particular to indicate why the point contact rectifiers should have much higher cutoff frequencies than the larger area barrier devices such as copper oxide or selenium. It was generally recognized that these differences were a matter of geometry rather than some fundamental property of the devices themselves. The equivalent circuit of such a rectifying device consists of the rectifier in parallel with a capacity and this combination in series with an ohmic resistance. When that frequency is reached at which the capacitive impedance is equal to or less than the series ohmic resistance, the rectifier is shunted out. Very little is gained by making a copper oxide barrier rectifier smaller in area since the series resistance and capacitive impedance are both inversely proportional to the area. In a point contact device, however, the series resistance in the forward direction is reduced by its spreading character and is inversely proportional to the radius of the point contact, while the capacity is proportional to the square of this radius. It is thus seen that making the point contact smaller improves the frequency response. This advantage was pushed to the limit in the design of radar detectors.

Active Semiconductor Devices Up to 1942

The analogy between the vacuum tube diode and the semiconductor diode was fairly obvious to any worker in this field in the nineteen thirties. A vacuum tube diode can be operated as a negative resistance and as such is an active circuit element. Some semiconductor diodes also exhibited negative resistance at frequencies as high as several megacycles although they were usually rather unstable and difficult to reproduce. The fundamental physics of these phenomena were not understood, but the most widely accepted explanation was based on thermal effects arising from the large negative temperature coefficient of resistance. Modern counterparts of these early negative resistance diodes are now known as negative or positive gap diodes.

The analogy with the vacuum tube diode also suggested to many workers of those days that what one really should do was put a grid in the semiconductor diode and, eureka, the result would be an active triode with amplifying possibilities. The insurmountable block to this experiment was, of course, the technique of placing this grid at the proper spacing in a region only 10^{-4} cm thick. R. Hilsch and R. W. Pohl [40], working with alkali halide crystals in which the space-charge layer could be made of the order of one centimeter in thickness, did put in a grid and made in principle an active solid-state triode circuit element. The frequency cutoff of this experimental device was of the order of one cycle per second or less.

The Situation After the War

With the advent of World War II many things happened to upset research work in general and semiconductor research work in particular. Scientists in one field were recruited into another and vice versa. It was human nature for those of us who went into new fields to rediscover what experts in that field had already known. This situation was further aggravated when secrecy closed down. Things done in the late 1930's which would normally have been published were placed under secrecy. For example, there were parallel developments on silicon point contact diodes in England [41] and probably in France and Germany, although the literature on this subject is quite sparse. The war work was of course device-minded, however much was done to fill in the scientific picture.

In the United States most of this war work was done at Massachusetts Institute of Technology, Purdue University, University of Pennsylvania, General Electric Company, and Bell Telephone Laboratories. H. C. Torrey and C. A. Whitmer [111] prepared a book entitled *Crystal Rectifiers*, published in 1948 as a part of the Radiation Laboratory Series,[4] which gives an excellent account of this war effort.

After the war these well-equipped semiconductor laboratories continued along their previous lines, excepting that there was generally a swing from device development toward more fundamental semiconductor research. As a starting point it was natural to ask the question why there had not been more success in understanding semiconductor phenomena considering all the work that had been done. This question may seem foolish in context because this story is being written with the perspective of 1955, but it was not foolish then. The answer was, of course, that semiconductor phenomena were very complicated and structure-sensitive. There was a ray of hope, however, in that two relatively simple semiconductors, silicon and germanium, had been extensively investigated during the war.

Silicon and germanium are elements, they have high melting points, and the binding forces which hold the solid together are almost purely of the simple covalent type. This type of bond, thanks to quantum mechanics, was well understood. Substitutional solid solutions of fifth column impurities acted as donors and likewise the third column impurities acted as acceptors. Com-

[3] Private communication from Karl Siebertz.

[4] McGraw-Hill Book Co., Inc., New York, N. Y.

pared with silicon and germanium, cuprous oxide was a mess and selenium, though an element, had a very low melting point and solidified into complicated interlocking chains. It thus appeared that the best course of action was to concentrate on these two simple semiconductor elements and try to understand them first. This was where the semiconductor group, organized at Bell Telephone Laboratories after the war, took off. This group was fortunate in two respects. While predominantly made up of physicists, it had the close cooperation of both chemists and metallurgists. R. B. Gibney,[6] a physical chemist, was a member of the group and contributed essentially to its progress. J. H. Scaff and H. C. Theuerer made available their metallurgical experience with silicon and germanium. In many cases it was possible to ask for, and receive quickly, samples of either silicon or germanium having specified concentrations of given impurity elements. It can be said in retrospect that progress from this point on was everywhere largely dependent on close cooperation between physicists, chemists, and metallurgists.

Surface Properties After 1942

As the dust began to settle after the war, it became more and more obvious that the theory of rectification did not fit all of the pertinent experimental facts. In the first place, theory predicted that the rectification between a metal and a semiconductor should depend on the contact potential difference between the metal and the semiconductor. While a qualitative agreement had been found for large area metal contacts on cuprous oxide and on selenium, the quantitative variation of rectification was too small by several orders of magnitude. For metal point contacts on silicon or germanium, the work function of the metal made little or no difference.

Theory also predicted a contact potential difference between *n*- and *p*-type silicon, but the experiments of W. E. Meyerhof [42] failed to verify this point. Although theory indicated that contacts between similar *n*-type or between similar *p*-type semiconductors should be ohmic, the experiments of S. Benzer [43] with medium doped *n*-type germanium wedges showed that such contacts acted like two opposing rectifiers. He found that a heavily doped germanium wedge produced good rectification with either of the above wedges. He also showed that the two reverse-current voltage characteristics obtained in the former experiment were in the one direction determined by one wedge, and in the other direction determined by the other wedge.

W. Shockley reasoned that if the contact potential field at the surface could produce a space charge layer in the semiconductor surface, then one should be able to produce such a space charge layer at will with an externally applied electric field. Furthermore, if the semiconductor was very thin, so that this space charge was an appreciable part of the thickness, then the applied field could be used to modulate the conductivity of the semiconductor sample and amplification would be possible (field effect). This experiment [44], although tried in many ways, always produced an effect less than predicted. It appeared to those working in this field that all of the above-described phenomena were related and that an explanation of one would clarify this entire branch of semiconductor research.

The break came when J. Bardeen [45] saw that if he assumed special energy states at the surface (*surface states*) then he could explain this group of puzzling experiments. I. Tamm [46] and W. Shockley [47] had previously suggested the possibility of such states at the free surface of a solid, but the importance of such states at a semiconductor surface had not been realized. In other words, it was naive to assume that the free surface of a semiconductor, not in contact with another conductor, had no charge double layer. The more reasonable assumption was that the space charge layer at the surface was a property of the semiconductor and its surface, so that no matter how complicated the surface (adsorbed atoms, etc.), the surface must be in equilibrium with its interior. As a result, a surface with no space charge layer would be very unlikely indeed. It was also deduced that a very low density of surface states, many times less than one per surface atom, would be sufficient to shield the interior from any reasonable surface field. The rectifying barrier was actually there at the surface before the contact was made. Only in special cases where the surface state density was not too large could the contact potential field modify the space charge layer a little one way or the other. This new understanding of rectifying contacts made the ohmic contact more difficult to understand, and we will discuss this point in more detail a little later.

Having postulated a space charge layer at the free surface of a semiconductor, the question now arose as to the best experiment for verifying its existence. W. Shockley pointed out that, according to this picture, the contact potential between *n*- and *p*-type samples should increase with doping. This experiment was successfully performed [48] and out of it came an estimate of the surface state density under fixed experimental conditions. It was likewise realized that if a space charge layer resided at the free surface, then one should be able to change the surface potential by shining a light on it, and this effect should be adaptable to experimental verification. Suitable experimental arrangements were set up; the effect was not only measurable, but the change in contact potential was of the right sign [49]. Further experiments along these lines led to the use of an electrolyte to bias the surface, and it was during the course of this work by J. Bardeen and W. H. Brattain [50] that the point contact transistor was born.

In the course of these experiments it became evident that the minority carrier, even in small concentrations, played a very important role. A large part of the re-

[6] Now at the Los Alamos Scientific Laboratory.

verse current in point contacts to n-type germanium consisted of holes flowing to the metal and in the forward direction the resistance was modulated by injection of holes [51]. It is of course not surprising that this blind spot persisted for so long. The minority carriers were, after all, present in too small concentrations in most semiconductors to matter very much. In cuprous oxide (hole conductor), for example, the electrons have a density of about one charge carrier per cubic centimeter! In silicon and germanium the density was greater but still very small until really high purity ingots were obtained.

The point contact transistor was a three-dimensional device. No good physicist likes to work with such a complicated case if it can be reduced to one dimension. This simplification led to the fundamental studies of W. Shockley, *et al.* [52] on carrier injection in germanium rods. This, along with the better understanding of point contacts [53], laid the groundwork for semiconductor electronics.

Another contribution of major importance was that of W. Shockley [54] in working out the theory of p-n junctions and junction transistors. This led to the construction and evaluation of p-n-p and n-p-n junction transistors [55, 56] using the grown crystal method and, shortly thereafter, using metal alloy-diffusion techniques [57, 58]. This latter method of preparing p-n junctions was likewise adapted to both germanium [59] and silicon rectifiers [60] for both low and high power uses. Another method for preparing p-n junctions is that of gaseous diffusion [61, 62] of the desired impurity atoms into the surfaces of semiconductors. This technique has produced high power silicon rectifiers [63] and highly efficient photovoltaic solar energy converters [64].

It is really too early to assess the impact of these developments on the physics of the solid state. Nevertheless, several generalities should be pointed out at this time. The p-n junction in silicon and germanium is probably the simplest surface or phase boundary known, and our understanding of its properties is complete in great detail [65]. Devices such as rectifiers, solar batteries, and transistors made from one of the most abundant elements, namely silicon, will undoubtedly have an important impact on technology. One now sees that a general definition might be proposed for an active circuit element as follows: Two phase boundaries close enough together (within a diffusion length) so that deviations from equilibrium occurring at either will influence the other and with electrical connections to each of the *three* regions bounded by the two phase boundaries. Note that this definition covers the vacuum tube triode as well as the transistor, and also the active diode where the connection to the middle region is missing.

Finally we should explain ohmic contacts, at least in principle. Either holes or electrons can flow easily across the surface of a semiconductor. Most of the non-ohmic phenomena occur as a result of deviations from equilibrium densities. If holes and electrons could recombine fast enough or be generated fast enough at the surface to maintain equilibrium for all current densities, then regardless of the dipole layer, the contact would be ohmic. This explains why mechanical damage of the surface is useful in making ohmic contact. Another possibility is to start with a given type of semiconductor and dope it more and more. Then as one approaches the surface the contact becomes essentially that between two metals since the dipole layer is so thin that it is penetrable. Doping in general lowers the lifetime so that this technique works both ways. It should be pointed out that an *abrupt* n-$n+$ or p-$p+$ boundary has interesting properties of its own and is not ohmic.

Body Properties After 1942

In the case of silicon and gemanium it soon became apparent that there was a real hope of obtaining an exact relationship between impurity concentration and conductivity. Data were accumulating from measurements of conductivity and Hall effect on more and more samples of both silicon and germanium and under a wide variety of conditions. These samples were polycrystalline and, as it turned out later, far from perfect. Nevertheless, the analysis of these data to give mobilities and concentrations of electrons and holes began to show definite consistencies. This started with work done at the University of Pennsylvania and at Purdue University. K. Lark-Horovitz and V. A. Johnson [66, 67] did an analysis of the germanium data and G. L. Pearson and J. Bardeen [68] did an even more complete job on silicon. Some of the highlights of these studies are given below.

X-ray studies by E. S. Greiner [68] on the lattice constant as a function of impurity concentration definitely established that phosphorus and boron impurity atoms in silicon occupy normal silicon sites in the lattice: they go in substitutionally rather than interstitially. The inference was very strong that this is also the case for all group III and group V impurities in both silicon and germanium.

It was found that the statistics of electron and hole concentrations obeyed the Wilson model quite well. In particular the various interactions between electrons, holes, donors, acceptors, and energy levels of the solid could be represented by the appropriate mass action laws, and the energy differences between the various levels could be determined. The simplest relation to come out of all this was that, regardless of impurity content, the product of the electron and hole densities was always given by a universal function of temperature, namely

$$np = K \exp[-E_g/kT],$$

where the chief difference between different semiconductors is in the value of E_g, since K changes very little.

For an intrinsic semiconductor, n is always equal to p. Equal concentrations of donors and acceptors leave this equality unchanged but unequal concentrations, at low enough temperatures, make $n>p$ or $p>n$ without changing the product. An apparently simple relation covering this latter situation is

$$n/p = \exp [2(E_f - E_i)/kT],$$

where E_f measures the Fermi level, or electrochemical potential, and E_i is its value for the intrinsic case where $n=p$. This is actually a complicated relationship since E_f is an involved function of the difference in donor and acceptor concentrations as well as of the absolute temperature T.

Another result of these analyses was a better understanding of the factors that control the mobility. The scattering of holes and of electrons could be separated into two parts. In the case of such covalent nonionic lattices as those of silicon and germanium, one part consists of the interaction or scattering of the charge carriers by the elastic thermal vibrations of the lattice (*phonons*). This effect increased with temperature; *i.e.*, the lattice part of the mobility decreases with increase in temperature. It was found in general that electron mobility is greater than hole mobility at any given temperature. For relatively pure samples at high enough temperatures, this was the whole story. The mobilities are an intrinsic property of the lattice alone and are not structure sensitive. At low temperatures and high impurity concentrations, on the other hand, the second part, impurity scattering, is important. That this part is due chiefly to scattering of holes and electrons by ionized impurity atoms was shown by E. Conwell and V. F. Weisskopf [69]. The temperature coefficient of this effect is opposite in sign to lattice scattering and is therefore limiting only at low temperatures and high concentrations of donors or acceptors.

Finally it was found, at first only qualitatively, that the added concentrations of either electrons or holes (whichever the case) is equal to the difference between donor and acceptor impurity concentrations. It was later shown, using radioactive antimony, that this relation is quantitative [70].

The following simple picture of an elementary semiconductor thus emerged as a result of the work described above. A group V impurity atom will enter into the crystal lattice and replace a group IV atom. Four of the impurity atom's valence electrons are used in covalent electron bonds with each of four tetrahedrally placed nearest neighbors. The fifth electron is loosely bound to the impurity atom by a force that is just about that between an electron and a singly charged ion in a dielectric medium. When excited, thermally or otherwise, this electron is free to conduct. The same reasoning applies to group III impurities which leave an extra hole in the valence band. Simultaneous addition of equal amounts of group V and group III impurity atoms exactly cancel their charge carriers. The extra electron fills the hole in the impurity band, thus leaving all covalent bonds satisfied. However, the donor atom is left positively charged and the acceptor atom negatively charged.

During this time various improvements were being made in semiconductor materials. A big step forward was taken when G. K. Teal and J. B. Little [71] succeeded in growing single crystals of germanium. Sometime later single crystals of silicon were also obtained by G. K. Teal and E. Buehler [72]. In all of this work steps were continually being taken to produce even more perfect crystals both as to lattice perfection and degree of chemical purity. A major breakthrough in this regard came with the ingenious development of zone refining by W. G. Pfann [73]. The impurities in germanium, and in many other chemical elements and compounds, tend to segregate into the molten portion of a solidifying ingot. By passing successive molten zones through a long ingot of germanium, Pfann was able to sweep these inherent impurities into one end of the ingot where they could be removed by cropping. Following a thorough study of the segregation coefficients associated with the various impurities [74], it was possible to prepare germanium ingots in which the harmful impurity content was less than 10^{12} atoms per cc. This amounts to less than one impurity atom for each ten billion germanium atoms and is probably the highest purity attained in any commercially available chemical element or compound. Although silicon is more difficult to zone refine than germanium, it has recently been purified to about the same level by a combination of zone refining and chemical purification. The progress achieved in semiconductor research during the past few years is closely associated with these advances in the preparation of materials.

The ability to produce large silicon and germanium single crystals having a high degree of lattice perfection did not essentially change the over-all picture although it did result in many refinements. There was a period of several years when the mobilities of both electrons and holes in germanium showed a linear increase with the date at which the experiment was performed, finally saturating at the presently accepted values about 1953.

Another very significant advance came with the discovery of transistor action. It then became possible to attain nonequilibrium concentrations of holes and electrons and the importance of the minority carrier concentration was finally realized. The classical experiments of J. R. Haynes and W. Shockley [75] made it possible to measure the drift mobility of holes and electrons directly. Small but significant differences were found between Hall and drift mobilities. Still another parameter was added by the determination of lifetimes of nonequilibrium concentrations of electrons and holes.

The work of M. Becker and H. Y. Fan [76] and the contemporary but independent work of H. B. Briggs [77] on the optical properties in germanium and silicon contributed essentially to the over-all picture as did the

quantum efficiency work of F. S. Goucher [78] which showed that one hole-electron pair was produced for each light quantum absorbed.

The work of P. P. Debye [79] and F. J. Morin [80, 81] in carrying out very careful measurements on really good single crystal samples of germanium and silicon over wide ranges of tempertaure and impurity concentrations was another significant contribution. Table I lists the presently accepted physical constants of these two well understood semiconductors.

TABLE I
PROPERTIES OF GERMANIUM AND SILICON

	Ge	Si
Melting point	936°C	1420°C
Density 25°C	5.323 gm/cm³	2.330 gm/cm³
Thermal expansion coeff. (25°C)	6.1×10^{-6}/°C	4.2×10^{-6}/°C
Thermal conductivity (25°C)	0.14 cal/sec cm°C	0.20 cal/sec cm°C
Specific heat (0–100°C)	0.074 cal/gm°C	0.181 cal/gm°C
Atomic weight	72.60	28.08
Lattice constant 25°C	5.657×10^{-8} cm	5.429×10^{-8} cm
Atoms/cc	4.42×10^{22}	4.96×10^{22}
Volume compressibility	1.3×10^{-12} cm²/dyne	0.98 cm²/dyne
Dielectric constant	16	12
Covalent bond ionization energy 0°K	1.2 electron volts	1.2 electron volts
Impurity atom ionization energy	~0.01 electron volts	~0.04 electron volts
Intrinsic resistivity at 300°K	47 ohm cm	230,000 ohm cm
Electron mobility at 300°K	3800 cm²/volt sec	1300 cm²/volt sec
Hole mobility at 300°K	1800 cm²/volt sec	500 cm²/volt sec

One could accordingly characterize a given sample of germanium at a given temperature by stating its electron and hole concentrations, the respective mobilities and the lifetime. It now became possible to formulate the mathematical problem of semiconductor electronics. This was accomplished in complete generality, at least for germanium at room temperature, by W. van Roosbroeck [82]. Significant contributions in this field were also made by J. Bardeen [83], W. Shockley [113], C. Herring [84], R. C. Prim [85] and H. Brooks [86]. Theory of recombination by traps was developed by R. N. Hall [87] and by W. Shockley and W. T. Read [88].

We are now getting so close to the forefront that the perspective is insufficient to give a good general picture. It can certainly be said, however, that the availability of such pure and perfect single crystals as we have in present-day silicon and germanium amounts to a major revolution in the physics of solids. New phenomena are turning up all around us. On the one hand the results from experiments on magneto-resistance [89, 90], piezoresistance [91] and cyclotron resonance [92, 93] have shown that the energy band structure in silicon and germanium cannot be characterized by a single isotropic effective mass for the charge carriers. It turns out that the situation is quite complicated [94, 95]. In some cases there are groups of carriers each with anisotropic effective masses. In other cases the effective mass is velocity dependent. The electron spin resonance results [96] in silicon and their analysis [97] lead to a better understanding of impurity energy levels. The new thermal emf results [98] contribute to the knowledge of interaction between phonons and charge carriers. The experiments on radiation damage [99, 100] and dislocations [101–104] lead to an improved understanding of solid structure in general.

The thorough understanding of germanium and silicon is contributing essentially to the understanding of other semiconductors such as lead sulfide [105] and cadmium sulfide [106]. An entirely new series of semiconductors, the group III–group V compounds, first investigated by H. Welker [107], is currently attracting a great deal of attention and some of these may very well rival silicon and germanium in the not too distant future. Table II lists the currently available constants of this series of semiconducting materials. Space as well as perspective just does not permit a complete story of the latest developments. We suggest that the reader consult the current literature.

TABLE II
PROPERTIES OF GROUP III–GROUP V INTERMETALLIC COMPOUNDS

	Melting Point °C	E_g (Optical) eV 300°K	n or p cm⁻³	μ_n (Hall) cm²/v sec 300°K	μ_p (Hall) cm²/v sec 300°K
InSb	523 (*a*)	.18 (*b*)	10^{15} (*b*)	77,000 (*b*)	~1,250 (*b*)
InAs	936 (*a*)	.35 (*b*)	5×10^{16} (*b*)	>15,000 (*b*)	
GaSb	720 (*a*)	.77 (*b*)	10^{17} (*b*)	2,500 (*b*)	700 (*b*)
InP	1070 (*a*)	1.25 (*a*)		3,400 (*a*)	650 (*a*)
GaAs	1240 (*a*)	1.35 (*c*)	2.7×10^{17} (*c*)	4,040 (*c*)	
AlSb	1080 (*a*)	1.7 (*b*)	10^{17} (*b*)	~100 (*b*)	

(*a*) H. Welker, "Semiconducting Intermetallic Compounds," *Physica*, vol. 20, p. 893; (November) 1954.
(*b*) H. J. Hrostowski and M. Tannenbaum, "Recent Work on Group III Antimonides and Arsenides," *Physica*, vol. 20, p. 1065; (November) 1954.
(*c*) R. Barrie, F. A. Cunnell, J. T. Edmond, and I. M. Ross, "Some Properties of Gallium Arsenide," *Physica*, vol. 20, p. 1087; (November) 1954.

BIBLIOGRAPHY

[1] Lark-Horovitz, K., "The New Electronics." See [117], p. 57.
[2] Herold, E. W., "Semiconductors and Transistors." *Journal of the Franklin Institute*, Vol. 259 (February, 1955), pp. 87–106.
[3] For a recent bibliography on transistors see: Krull, A. R., "Transistors and Their Applications. A Bibliography 1948–1953." *Transactions of the Institute of Radio Engineers*, Vol. ED-1 (September, 1954), p. 40.
[4] Faraday, M., *Experimental Researches in Electricity*, Bernard Quaritch, London (1839). Vol. I, pp. 122–124.
[5] Becquerel, A. E., "On Electric Effects under the Influence of Solar Radiation." *Comptes Rendus de l'Academie des Sciences*, Vol. 9 (November 21, 1839), pp. 711–714.
[6] Smith, W., "The Action of Light on Selenium." *Jour. of the Soc. Telegraph Engineers*, Vol. 2, No. 1 (1873), pp. 31–33.
[7] Braun, F., "Ueber die Stromleitung durch Schwefelmetalle." *Annalen der Physik und Chemie*, Vol. 153, No. 4 (1874), pp. 556–563.
[8] Schuster, A., "On Unilateral Conductivity." *Philosophical Magazine*, Vol. 48 (October, 1874), pp. 251–257.
[9] Adams, W. G., and Day, R. E., "The Action of Light on Selenium." *Proceedings of the Royal Society of London*, Vol. 25 (June, 1876), pp. 113–117.

[10] Fritts, C. E., "A New Form of Selenium Cell." *American Journal of Science*, Vol. 26 (December, 1883), pp. 465–472.
[11] Bose, J. C., U. S. Patent 755840, 1904.
[12] Dunwoody, H. H. C., U. S. Patent 837616, 1906.
[13] Austin, L. W., "A High Resistance Thermo-Electric Detector for Electrical Waves," *Physical Review*, Vol. 24 (June, 1907), pp. 508–510.
[14] Pierce, G. W., "Crystal Rectifiers for Electric Currents and Electrical Oscillators." *Physical Review*, Vol. 25 (July, 1907), pp. 31–60, Vol. 28 (March, 1909), pp. 153–187, Vol. 29, (November, 1909), pp. 478–484.
[15] Grondahl, L. O., and Geiger, P. H., "A New Electronic Rectifier." *Transactions of the American Institute of Electrical Engineers*, Vol. 46 (February, 1927), pp. 357–366.
[16] Lange, B., *Die Photoelemente und Ihre Anwendung*. Leipzig, Johann Ambrosius Barth, 1936.
[17] Pohl, R. W., "Electron Conductivity and Photochemical Processes in Alkali Halide Crystals." *Proceedings of the Physical Society*, Vol. 49 (extra part (1937)), pp. 3–31.
[18] Hall, E. H., "On a New Action of the Magnet on Electric Currents." *American Journal of Mathematics* Vol. 2 (November, 1879), pp. 287–291.
[19] Vogt, W., "Elektrisches und Optisches Verhalten von Halbleitern III." *Annalen der Physik*, Vol. 7 (October 31, 1930), pp. 183–204.
[20] Englehard, E., "Elektrisches und Optisches Verhalten von Halbleitern IX." *Annalen der Physik*, Vol. 17 (July 8, 1933), pp. 501–542.
[21] Schottky, W., and Waibel, F., "Die Elektronenleitung des Kupperoxyduls." *Physikalische Zeitschrift*, Vol. 34 (December 1, 1933), pp. 858–864.
[22] Wagner, C., and Schottky, W., "Theorie der Geordneten Mischphases." *Zeitschrift für Physikalische Chemie*, Vol. B11 (December, 1930), pp. 163–210.
[23] Wagner, C., "Theorie der Geordneten Mischphasen." *Zeitschrift für Physikalische Chemie*, Vol. B22 (July, 1933), pp. 181–194.
[24] Sommerfeld, A., "Zur Elektronentheorie der Metalle auf Grund der Fermischen Statistik." *Zeitschrift für Physik*, Vol. 47 (February, 1928), pp. 1–32, Vol. 47 (February, 1928), pp. 43–60.
[25] Wilson, A. H., "The Theory of Electronic Semiconductors." *Proceedings of the Royal Society of London*, Vol. A133 (October, 1931), pp. 458–491, Vol. A134 (November, 1931), pp. 277–287.
[26] DeBoer, J. H., and Van Geel, W. C., "Rotgrenze des Inneren Photoeffektes und Ablosungsarbeit bei Halbleitern." *Physica*, Vol. 2 (January, 1935), pp. 286–298.
[27] Dember, H., "Über eine Photoelektromotorische Kraft in Kuperoxydul-Kristallen." *Physikalische Zeitschrift*, Vol. 32 (July, 1931), pp. 554–556, Vol. 32 (November, 1931), pp. 856–858.
[28] Frenkel, J., "Theory of Some Photoelectric and Photomagnetoelectric Phenomena in Semiconductors." *Physikalische Zeitschrift der Sowjetunion*, Vol. 8, No. 2 (1935), pp. 185–203.
[29] Kikoin, I. K., and Naskov, M. M., "A New Photoelectric Effect in Cuprous Oxide." *Physikalische Zeitschrift der Sowjetunion*, Vol. 5, No. 4, (1934), pp. 586–596.
[30] McEachron, K. B., "Thyrite, A New Material for Lightning Arresters." *General Electric Review*, Vol. 33 (February, 1930), pp. 92–99.
[31] Schottky, W., "Zur Halbleitertheorie der Sperrschichtrichter und Spitzengleichrichter," *Zeitschrift für Physik*, Vol. 113 (July 21, 1939), pp. 367–414.
[32] Mott, N. F., "The Theory of Crystal Rectifiers." *Proceedings of the Royal Society of London*, Vol. 171 (May 1, 1939), pp. 27–38.
[33] Davydov, B., "On the Contact Resistance of Semiconductors." *Journal of Physics* (USSR), Vol. 1, No. 2 (1939), pp. 167–174.
[34] Davydov, B., "The Rectifying Action of Semiconductors." *Technical Physics of the USSR*, Vol. 5, No. 2 (1938), pp. 87–95.
[35] Davydov, B., "On the Photo-electromotive Force in Semiconductors." *Technical Physics of the USSR*, Vol. 5, No. 2 (1938), pp. 79–86.
[36] Scaff, J. H., Theurer, H. C., and Schumacher, E. E., "*p*-type and *n*-type Silicon and the Formation of the Photovoltaic Barrier in Silicon Ingots." *Transactions of the American Institute of Mining and Metallurgical Engineers*, Vol. 185 (June, 1949), pp. 383–388. Although a major portion of the work reported in this reference was done before the war, publication was held up for security reasons.
[37] Scaff, J. H., U. S. Patent 2,402,582. Filed April 4, 1941.
[38] Ohl, R. S., U. S. Patent 2,402,662. Filed May 27, 1941.
[39] Merritt, E., "On Contact Rectification by Metallic Germanium." *Proceedings of the National Academy of Sciences*, Vol. 11 (December, 1925), pp. 743–748.
[40] Hilsch, R., and Pohl, R. W., "Steuerung von Elektronenströmen mit einem Dreielektrodenkristall und ein Modell einen Sperrschicht." *Zeitschrift für Physik*, Vol. 11 (December 8, 1938), pp. 399–408.
[41] Bleaney, B., Ryde, J. W., and Kinman, T. H., "Crystal Valves." *Journal of the Institution of Electrical Engineers*, Part IIIA, Vol. 93 (May, 1946), pp. 847–854.
[42] Meyerhof, W. E., "Contact Potential Difference in Silicon Crystal Rectifiers." *Physical Review*, Vol. 71 (May 15, 1947), pp. 727–735.
[43] Benzer, S., "Ge-Ge Contacts," *Physical Review*, Vol. 71 (January 15, 1947), p. 141.
[44] Shockley, W., and Pearson, G. L., "Modulation of Conductance of Thin Films of Semiconductors by Surface Charges." *Physical Review*, Vol. 74 (July 15, 1948), pp. 232–233.
[45] Bardeen, J., "Surface States and Rectification at a Metal to Semiconductor Contact." *Physical Review*, Vol. 71 (May 15, 1947), pp. 717–727.
[46] Tamm, I., "Über Eine Mögliche Art der Elektronenbindung an Kristalloberflächen," *Physikalische Zeitschrift der Sowjetunion*, Vol. 1, No. 6 (1932), pp. 733–746.
[47] Shockley, W., "On the Surface States Associated with a Periodic Potential." *Physical Review*, Vol. 56 (August 15, 1939), pp. 317–323.
[48] Brattain, W. H., and Shockley, W., "Density of Surface States Deduced from Contact Potential Measurements." *Physical Review*, Vol. 72 (August 15, 1948), p. 345.
[49] Brattain, W. H., "Evidence for Surface States from Change in Contact Potential on Illumination." *Physical Review*, Vol. 72 (August 15, 1948), p. 345.
[50] Bardeen, J., and Brattain, W. H., "The Transistor, a Semiconductor Triode." *Physical Review*, Vol. 74 (July 15, 1948), pp. 230–231; "Physical Principles Involved in Transistor Action." *Physical Review*, Vol. 75 (July 15, 1949), pp. 203–231.
[51] Brattain, W. H., and Bardeen, J., "Nature of the Forward Current in Germanium Point Contacts." *Physical Review*, Vol. 74 (July 15, 1948), pp. 231–232.
[52] Shockley, W., Pearson, G. L., and Haynes, J. R., "Hole Injection in Germanium-Quantitative Studies and Filamentary Transistors." *Bell System Technical Journal*, Vol. 28 (July, 1949), pp. 344–366.
[53] Bardeen, J., "Theory of Relation Between Hole Concentrations and Characteristics of Germanium Point Contacts." *Bell System Technical Journal*, Vol. 29 (October, 1950), pp. 469–495.
[54] Shockley, W., "The Theory of *p-n* Junctions in Semiconductors and *p-n* Junction Transistors." *Bell System Technical Journal*, Vol. 28 (July, 1949), pp. 435–489.
[55] Shockley, W., Sparks, M., and Teal, G. K., "*p-n* Junction Transistors." *Physical Review*, Vol. 83 (July 1, 1951), pp. 151–162.
[56] Wallace, R. L., Jr., and Pietenpol, W. J., "Some Circuit Properties and Applications of *n-p-n* Transistors." *Bell System Technical Journal*, Vol. 30 (July, 1951), pp. 530–563.
[57] Hall, R. N., and Dunlap, W. C., Jr., "*p-n* Junctions Prepared by Impurity Diffusion." *Physical Review*, Vol. 80 (November 1, 1950), pp. 867–868.
[58] Saby, J. E., "Fused Impurity *p-n-p* Transistors." PROCEEDINGS OF THE IRE, Vol. 40 (November, 1952), pp. 1358–1360.
[59] Hall, R. N., "Power Rectifiers and Transistors." PROCEEDINGS OF THE IRE, Vol. 40 (November, 1952), pp. 1512–1518.
[60] Pearson, G. L., and Sawyer, B., "Silicon *p-n* Junction Alloy Diodes." PROCEEDINGS OF THE IRE, Vol. 40 (November, 1952), pp. 1348–1351.
[61] Scaff, J. H., and Theuerer, H. C., U. S. Patent 2,567,970, September 18, 1951.
[62] Fuller, C. S., and Ditzenberger, J. A., "The Diffusion of Boron and Phosphorus into Silicon." *Journal of Applied Physics*, Vol. 25 (November, 1954), pp. 1439–1440.
[63] Pearson, G. L., and Fuller, C. S., "Silicon *p-n* Junction Power Rectifiers and Lightning Protectors." PROCEEDINGS OF THE IRE, Vol. 42 (April, 1954), p. 760.
[64] Chapin, D. M., Fuller, C. S., and Pearson, G. L., "A New Silicon *p-n* Junction Photocell for Converting Solar Radiation into Electrical Power." *Journal of Applied Physics*, Vol. 25 (May, 1954), pp. 676–677.
[65] Brattain, W. H., and Garrett, C. G. B., "Surface Properties of Semiconductors." *Physica*, Vol. 20 (November, 1954), pp. 885–892.
[66] Lark-Horovitz, K., Middleton, A. E., Miller, E. P., and Walerstein, I., "Electrical Properties of Germanium Alloys." *Physical Review*, Vol. 69 (March, 1946), p. 258.
[67] Lark-Horovitz, K., and Johnson, V. A., "Theory of Resistivity in Germanium Alloys." *Physical Review*, Vol. 69 (March, 1946), p. 258.
[68] Pearson, G. L., and Bardeen, J., "Electrical Properties of Pure

Silicon and Silicon Alloys Containing Boron and Phosphorus." *Physical Review*, Vol. 75 (March 1, 1949), pp. 865–883.
[69] Conwell, E., and Weisskopf, V. F., "Theory of Impurity Scattering in Semiconductors." *Physical Review*, Vol. 77 (February 1, 1950), pp. 388–390.
[70] Pearson, G. L., Struthers, J. D., and Theuerer, H. C., "Correlation of Geiger Counter and Hall Effect Measurements in Alloys Containing Germanium and Radioactive Antimony 124." *Physical Review*, Vol. 77 (March 15, 1950), pp. 809–813.
[71] Teal, G. K., and Little, J. B., "Growth of Germanium Single Crystals." *Physical Review*, Vol. 78 (June 1, 1950), p. 647.
[72] Teal, G. K., and Buehler, E., "Growth of Silicon Single Crystals and of Single Crystal Silicon *p-n* Junctions." *Physical Review*, Vol. 87 (July 1, 1952), p. 190.
[73] Pfann, W. G., "Principles of Zone Refining." *Transactions of the American Institute of Mining and Metallurgical Engineers*, Vol. 194 (July, 1952), pp. 747–753.
[74] Burton, J. A., "Impurity Centers in Ge and Si," *Physica*, Vol. 20 (November, 1954), pp. 845–853.
[75] Haynes, J. R., and Shockley, W., "The Mobility and Life of Injected Holes and Electrons in Germanium." *Physical Review*, Vol. 81 (March 1, 1951), pp. 835–843.
[76] Becker, M., and Fan, H. Y., "Optical Properties of Semiconductors, II, Infra-Red Transmission of Germanium." *Physical Review*, Vol. 76 (November 15, 1949), pp. 1530–1531, "III, Infra-Red Transmission of Silicon." *Physical Review*, Vol. 76 (November 15, 1949).
[77] Briggs, H. B., "Optical Effects in Bulk Silicon and Germanium." *Physical Review*, Vol. 77 (January 15, 1950), p. 287.
[78] Goucher, F. S., "The Photon Yield of Electron-Hole Pairs in Germanium." *Physical Review*, Vol. 78 (June 15, 1950), p. 816.
[79] Debye, P. P., and Conwell, E. M., "Electrical Properties of *N*-type Germanium." *Physical Review*, Vol. 93 (February 15, 1954), pp. 693–706.
[80] Morin, F. J., and Maita, J. P., "Conductivity and Hall Effect in the Intrinsic Range of Germanium." *Physical Review*, Vol. 94 (June 15, 1954), pp. 1525–1529.
[81] Morin, F. J., and Maita, J. P., "Electrical Properties of Silicon Containing Arsenic and Boron." *Physical Review*, Vol. 96 (October 1, 1954), pp. 28–35.
[82] van Roosbroeck, W., "The Transport of Added Current Carriers in a Homogeneous Semiconductor." *Physical Review*, Vol. 91 (July 15, 1953), pp. 282–289.
[83] Bardeen, J., "Flow of Electrons and Holes in Semiconductors." See [117] page 128.
[84] Herring, C., "Theory of Transient Phenomena in the Transport of Holes in an Excess Semiconductor." *Bell System Technical Journal*, Vol. 28 (July, 1949), pp. 401–427.
[85] Prim, R. C., III, "Some Results Concerning the Partial Differential Equations Describing the Flow of Holes and Electrons in Semiconductors." *Bell System Technical Journal*, Vol. 30, (October, 1951), pp. 1174–1213.
[86] Brooks, H., "Theory of Diffusion of Injected Carriers in a Semiconductor of Mixed Conductivity Type." *Physical Review*, Vol. 90 (April 15, 1953), p. 336.
[87] Hall, R. N., "Electron-Hole Recombination in Germanium." *Physical Review*, Vol. 87 (July 15, 1952), p. 387.
[88] Shockley, W., and Read, W. T., Jr., "Statistics of the Recombination of Holes and Electrons." *Physical Review*, Vol. 87 (September 1, 1952), pp. 835–842.
[89] Pearson, G. L., and Suhl, H., "The Magneto-Resistance Effect in Oriented Single Crystals of Germanium." *Physical Review*, Vol. 83 (August 15, 1951), pp. 768–776.
[90] Pearson, G. L., and Herring, C., "Magneto-Resistance Effect and the Band Structure of Single Crystal Silicon." *Physica*, Vol. 20 (November, 1954), pp. 975–978.
[91] Smith, C. S., "Piezoresistance Effect in Germanium and Silicon." *Physical Review*, Vol. 94 (April 1, 1954), pp. 42–49.
[92] Dresselhaus, G., Kip, A. F., and Kittel, C., "Cyclotron Resonance of Electrons and Holes in Silicon and Germanium Crystals." *Physical Review*, Vol. 98 (April 15, 1955), pp. 368–384.
[93] Lax, B., Zeiger, H. J., and Dexter, R. N., "Anistropy of Cyclotron Resonance in Germanium." *Physica*, Vol. 20 (November, 1954), pp. 818–828.
[94] Herman, F., "Calculation of the Energy Band Structures of the Diamond and Germanium Crystals by the Method of Orthogonalized Plane Waves." *Physical Review*, Vol. 93 (March 15, 1954), pp. 1214–1225.
[95] Herring, C., "Transport Properties of a Many-Valley Semiconductor." *Bell System Technical Journal*, Vol. 34 (March, 1955), pp. 237–290.
[96] Fletcher, R. C., Yager, W. A., Pearson, G. L., and Merritt, F. R., "Hyperfine Splitting in Spin Resonance of Group V Donors in Silicon." *Physical Review*, Vol. 95 (August 1, 1955), pp. 844–845.
[97] Kohn, W., and Luttinger, J. M., "Hyperfine Splitting of Donor States in Silicon." *Physical Review*, Vol. 97 (February 15, 1955), pp. 883–888.
[98] Herring, C., "Theory of the Thermoelectric Power of Semiconductors." *Physical Review*, Vol. 96 (December 1, 1954), pp. 1163–1187.
[99] Lark-Horovitz, K., "Nucleon-Irradiated Semiconductors." See [114], page 47.
[100] Brown, W. L., Fletcher, R. C., and Wright, K. A., "Annealing of Bombardment Damage in Germanium." *Physical Review*, Vol. 92 (November 1, 1953), pp. 591–596.
[101] Gallagher, C. J., "Plastic Deformation of Germanium and Silicon." *Physical Review*, Vol. 88 (November 15, 1952), pp. 721–722.
[102] Vogel, F. L., Pfann, W. G., Corey, H. E., and Thomas, E. E., "Observations of Dislocations in Lineage Boundaries in Germanium." *Physical Review*, Vol. 90 (May 1, 1953), p. 489.
[103] Pearson, G. L., Read, W. T., Jr., and Morin, F. J., "Dislocations in Plastically Deformed Germanium." *Physical Review*, Vol. 93 (February, 1954), pp. 666–667.
[104] Read, W. T., Jr., "Theory of Dislocations in Germanium." *Philosophical Magazine*, Vol. 45 (August, 1954), pp. 775–796; Vol. 45 (November, 1954), pp. 1119–1128; Vol. 46 (February, 1955), pp. 111–131.
[105] Smith, R. A., "The Electronic and Optical Properties of the Lead Sulfide Group of Semiconductors." *Physica*, Vol. 20 (November, 1954), pp. 910–929.
[106] Smith, R. W., and Rose, A., "Space-Charge-Limited Currents in Single Crystals of Cadmium Sulfide." *Physical Review*, Vol. 97 (March 15, 1955), pp. 1531–1537.
[107] Welker, H., "Über Neue Halbleitende Verbendungen." *Zeitschrift für Naturforschung*, Vol. 7A (November, 1952), pp. 744–749; Vol. 8A (April, 1953), pp. 248–251.

BOOKS

[108] Wilson, A. H., *Semiconductors and Metals*. London, Cambridge University Press, 1939.
[109] Mott, N. F., and Gurney, R. W., *Electronic Processes in Ionic Crystals*. Oxford, Clarendon Press, 1940.
[110] Seitz, F., *The Modern Theory of Solids*. New York, McGraw-Hill Book Company, 1940.
[111] Torrey, H. C., and Whitmer, C. A., *Crystal Rectifiers*. New York, McGraw-Hill Book Company, 1948.
[112] Wright, D. A., *Semiconductors*. New York, John Wiley & Sons, Inc., 1950.
[113] Shockley, W., *Electrons and Holes in Semiconductors*. New York, D. Van Nostrand Company, 1950.
[114] Henisch, H. K., ed., *Semiconducting Materials*. London, Thornton Butterworth, Ltd., 1951.
[115] Moss, T. S., *Photoconductivity in the Elements*. London, Thornton Butterworth, Ltd., 1952.
[116] Kittel, C., *Introduction to Solid State Physics*. New York, John Wiley & Sons, Inc., 1953.
[117] Brackett, F. S., ed., *The Present State of Physics*. Washington, D. C., American Association for the Advancement of Science, 1954.
[118] Spenke, E., *Elektronische Halbleiter*. Berlin, Springer, 1955

Junction Transistor Electronics*

J. L. MOLL†, ASSOCIATE MEMBER, IRE

Summary—**The density and kind of charge carriers in semiconductors are controlled by the density and nature of added impurities. The charge carriers (electrons and holes) move about in the crystal simultaneously by diffusion and drift. Rectification occurs at *p-n* boundaries because of the difference in the carrier types on either side. Junction transistors, capable of amplification, consist of two junctions placed close to each other with a thin middle region or base layer. As a switch, the junction transistor is capable of switching "on" to voltage of the order of 10 millivolts and slope resistances of a few ohms. In transmission applications experimental junction transistors have been built to give useful power output at 250 mc and with maximum frequency of oscillation in the range from 500 to 1,000 mc. Reliability experience to date has been encouraging; some systems are operating with replacement rate as low as one per 150,000 transistor hours.**

INTRODUCTION

THE DISCOVERY of the point contact transistor[1] provided a powerful stimulus to the search for semiconductor amplifiers. The point contact transistor is able to operate on very low power and to perform useful transmission and switching circuit functions to moderately high frequencies. The potential use for such a small, rugged, reliable, efficient amplifier or switch spurred expanded research into the nature of the semiconductor materials as well as the electronics of semiconductor amplifiers.

The resulting extended knowledge of the nature of current carriers in semiconductors, the transport of carriers, and the nature of surfaces between opposite types of semiconductors facilitated development of commercially feasible junction transistor triodes.[2] At the present time the junction transistor appears to be the most useful of the family of semiconductor amplifiers.

There is too great a variety of semiconductor amplifiers to discuss them all thoroughly in one article so only the junction triode will be considered here. The literature is so extensive that complete references are clearly impossible, but an attempt will be made to give a representative sample. First an account will be given of some of the electrical properties of semiconductors as well as of boundary conditions that prevail at junctions. The properties of junction diodes, including the breakdown mechanism, will then be described. Finally, some of the more important junction transistor parameters will be obtained, together with examples of the range of values presently attained.

CHARGE CARRIERS—ELECTRONS AND HOLES IN SEMICONDUCTORS[3]

In a perfect single crystal of pure germanium or silicon, the atoms are arranged as in the diamond crystal lattice. There are four valence bonds per atom to be filled and four electrons per atom available for valence bonds. The bonded electrons are not free to move under the influence of an electric field because in the quantum mechanical sense there are no empty energy levels adjacent to the levels occupied by electrons. To remove an electron from the valence band to an energy state where it is free to move in an electric field requires 0.75 electron volts for germanium and 1.1 electron volts for silicon.[4] At any finite temperature there is a probability that any electron will be freed from the valence bond by the thermal vibrations. A freed electron results in an empty valence bond, or hole. At each temperature there is a statistical rate of thermal generation of electron-hole pairs. Thermal generation together with the finite probability of a free electron combining with an empty valence bond results in a thermal equilibrium number of electron-hole pairs. The electrons of course are free to move about the lattice and contribute to conductivity. The holes or empty valence bonds can be treated in a way similar to the free electrons; *i.e.*, the holes act like positively charged particles with a positive effective mass.

In free space, charged particles move with a velocity determined by the integral of the applied force so that in a sense an electron remembers the field that has been previously applied. In distinction to an electron in free space, electrons and holes in semiconductors move about in the crystal with thermal velocity of about 10^7 cm/sec and experience collisions with the crystal lattice about every 10^{-5} cm. The collisions result in a uniform random scattering, so that any motion imparted to the electron or hole before a collision occurs is lost. If an electric field is applied to the crystal, the free charged particles move under the influence of the field; but the frequent collisions with the crystal lattice result in proportionality between electric field and velocity [see (2)] rather than the proportionality between electric field and acceleration as is obtained in the case of a particle in free space.

In addition to the carrier motion induced by applied fields, carrier transport can occur by diffusion, again as a result of the thermal energy of the electrons and holes

* Original manuscript received by the IRE, October 10, 1955.
† Bell Telephone Labs., Inc., Murray Hill, N. J.
[1] J. Bardeen and W. H. Brattain, "The transistor, a semiconductor triode," *Phys. Rev.*, vol. 74, pp. 230–231; July, 1948; "Physical principles involved in transistor action," *Phys. Rev.*, vol. 75, pp. 1208–1225; April, 1949.
[2] W. Shockley, "The theory of *p-n* junctions in semiconductors and *p-n* junction transistors," *Bell Sys. Tech. Jour.*, vol. 28, pp. 435–489; July, 1949.
[3] For a more complete discussion of the properties of electrons and holes in semiconductors see W. Shockley, "Electrons and Holes in Semiconductors," D. Van Nostrand Co., Inc., New York, N. Y.; 1950. Also W. Shockley, "Transistor electronics: imperfections, unipolar transistors," PROC. IRE, vol. 40, pp. 1289–1313; Nov., 1952.
[4] E. M. Conwell, "Properties of silicon and germanium," PROC. IRE, vol. 40, pp. 1327–1337; November, 1952.

and the random collisions with the lattice. If, for example, a bunch of electron-hole pairs was introduced into a crystal of germanium near one crystal surface (say by shining a light on the surface and thus supplying the necessary energy for the electrons to break their valence bonds) the resulting electrons and holes would move deeper into the crystal through their diffusive motion. The electrons and holes tend, through diffusion, to even out their distribution whenever there is a difference in density in different parts of a crystal. The strength of the resulting diffusion current is porportional to the carrier gradient [see (3)]. Diffusion in a density gradient and drift in an electric field can, and often does, take place simultaneously in the semiconductor. The currents resulting from drift and diffusion may flow in the same or in opposite directions. The total transport current is the sum of the drift and diffusion currents [see (4), (5)].

The rate of diffusion is characterized by the diffusion constant D, which is defined as the ratio of diffusion current (expressed as number of carriers per second) to density gradient. The larger is D the faster the carriers diffuse. Drift is characterized by the mobility μ, which is defined as the ratio of velocity to electric field. For a given carrier and semiconductor the diffusion constant D and mobility μ are proportional to each other, with the proportionality given by the Einstein relation of (6). The mobilities and diffusion constants for both types of carriers are very important to transistor operation since the attainable frequency response increases with increasing mobility. Table I lists diffusion constants and mobilities for electrons and holes in high purity silicon[5] and germanium.

TABLE I

	Mobility $cm^2 V^{-1} sec^{-1}$	Diffusion constant $cm^2 sec^{-1}$
Germanium		
Electrons	3600	90
Holes	1800	45
Silicon		
Electrons	1500	38
Holes	500	13

The electrical conductivity of the semiconductor is proportional to the numbers of charge carriers and their mobility [see (7)]. For pure germanium in equilibrium at 300°K the number of electron hole pairs[6] (called intrinsic carrier density) is $2.5\times10^{13}/cm^3$, and the intrinsic conductivity is approximately .02 mho/cm. For silicon, the intrinsic carrier density at 300°K is $6.8\times10^{10}/cm^3$ and the intrinsic conductivity is approximately $.02\times10^{-3}$ mho/cm. The difference in intrinsic conductivity is essentially due to the difference in the number of intrinsic electrons and holes which in turn depends primarily on the previously mentioned energy gap between the conduction and valence bands.

Addition of the order of as much as a part per million of Group III or Group V chemical impurity to the germanium or silicon crystal results in drastic changes in the electrical properties. Group III elements result in a built-in vacant valence bond per impurity atom. This vacant bond can be made to leave the impurity atom with an energy of .01 to .05 electron volts, and results in a negatively charged immobile atom and a positively charged mobile hole. Similarly, Group V elements result in an extra electron per atom which can be moved away from the impurity atom with .01 to .05 electron volts, resulting in a positively charged immobile atom and a negatively charged mobile electron. The binding energy of the carrier to the impurity atom is so small that at room temperature, nearly all of the carriers are ionized and free to move.

Germanium or silicon doped with Group III chemical impurities conducts electrical current essentially by means of the resulting holes and is called p-type. When doped with Group V elements the conduction is essentially by means of electrons and the semiconductor is n-type. In any actual crystal both Group III and Group V impurities are simultaneously present and the effect is of cancellation so that only the net difference of impurities is significant. Thus, in an n-type crystal, the number of Group V (donors) minus the number of Group III (acceptor) atoms determines the number of conduction electrons.

The carrier which is most numerous is called the majority carrier. Thus electrons (holes) are the majority carriers in n-type (p-type) semiconductors. The opposite type of carrier (*i.e.*, holes in n-type and electrons in p-type) is called the minority carrier and exists at ther mal equilibrium in relatively very small quantities for doping in the range of $10^{15}/cm^3$ or more net impurities. The increased carrier density resulting from the addition of chemical impurities increases the probability of recombination of the thermally generated carriers. In n-type semiconductor, for example, the increased electron density results in an equilibrium hole density which is less than the density in the pure semiconductor. In either n or p-type germanium or silicon, the carrier densities obey the law that the product of carrier densities (at equilibrium) is a constant independent of the amount of impurity but dependent on temperature [see (8)]. For no added impurities, the electron and hole densities are equal to the intrinsic carrier density which has been previously discussed. If, for example, $2.5\times10^{16}/cm^3$ excess donors (or acceptors) are added to germanium, the minority carrier density will be only $2.5\times10^{10}/cm^3$ at 300°K, resulting in a ratio of majority to minority carrier density of a million to one. Even in this small ratio the minority carriers play an important role in determining junction properties.

[5] M. B. Prince, "Drift mobilities in semiconductors. I. Germanium, II. Silicon," *Phys. Rev.*, vol. 92, pp. 681–687; November, 1953, *Phys. Rev.*, vol. 93, pp. 1204–1206; March, 1954.
[6] Conwell, *op. cit*

p-n Boundaries, Junction Diodes

Probably the most clearly defined and best understood surface known to modern physics is the p-n boundary within a single crystal of semiconductor. For the sake of simplicity assume a planar boundary with uniform density of impurities on either side as in Fig. 1.

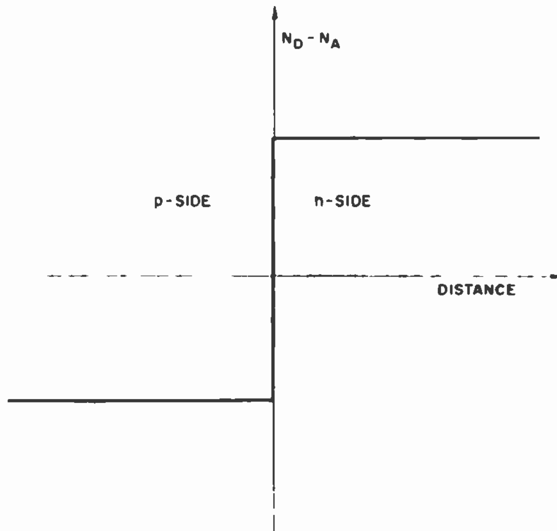


Fig. 1—Impurity concentration near a step junction.

Holes tend to diffuse to the right and electrons to diffuse to the left across the boundary. Since at equilibrium the current must be zero, an electrostatic contact potential is built up which opposes the diffusion of majority carriers across the boundary. This contact potential of course accelerates minority carriers from either side across the junction. In the equilibrium condition, the tendency of majority carriers from one side to diffuse across the junction is exactly balanced by the tendency of minority carriers from the other side to diffuse to the junction and drift across. Thus, the total hole current arising from the diffusion of holes from the p side is opposed by the current arising from minority holes on the n side which diffuse to the junction and drift across. Similarly, for electrons within the junction, a diffusion current from the majority side is opposed by a drift current from the minority side.

If electrical contact is made to the p- and n-type regions on either side of the boundary and if an external potential is applied (neglecting series resistance), the electrostatic voltage across the boundary is the sum of the built-in voltage and the applied voltage if the positive sense of the applied voltage is from n to p type. Inside the n- or p-type region, the electric field and the net space charge density are nearly zero. Near the p-n boundary an electric field exists, and there is a finite space charge density. A fairly accurate assumption is that near the p-n boundary almost all of the carriers have been swept out by the electrostatic fields, and the space charge density is equal to the density of ionized immobile impurity centers. Fig. 2 illustrates this assumption together with the fields as calculated from classical electrostatic field theory. The space charge layer width X_m is related to the voltage V by means of electrostatic theory; for a step junction, as illustrated in Fig. 1, the width varies with $V^{1/2}$, while for a graded junction (impurity density proportional to distance from junction), it varies as $V^{1/3}$ [see (9), (10)]. Since the space-charge layer width varies with voltage, when the voltage changes, the edges of the layer must be charged or discharged. Consequently, the junction has a "barrier capacitance" C_T. It so happens that C_T is given in terms of the junction width and dielectric constant by the ordinary condenser formula [see (11)].

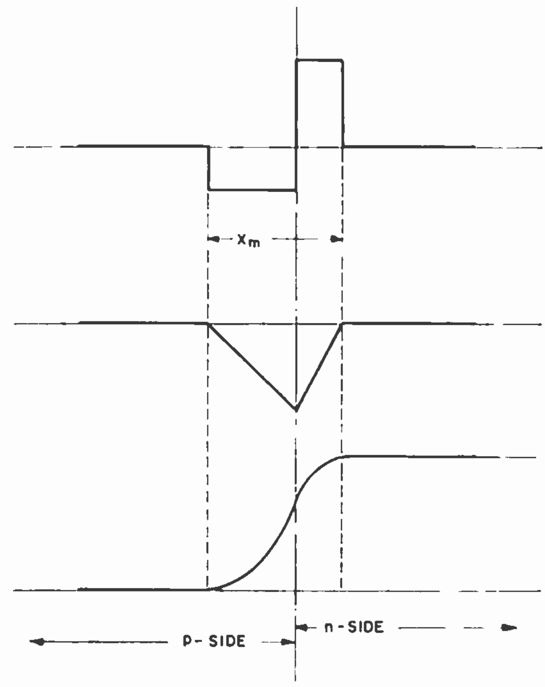


Fig. 2— Space charge density near a step p-n boundary; resulting electric field; resulting electrical potential.

Since the fields outside the junction are small, the net charge in the space charge region is negligible. The negative charge on the p side is balanced by an equal positive charge on the n side. A type of junction encountered very often has very low resistivity on one side compared to the other (say $N_P \gg N_n$). In this case the space-charge layer must extend almost entirely into the high resistivity (low net impurity density) side in order to preserve the over-all neutrality, so that the junction width Xm is independent of the low resistivity side, and varies inversely as the square root of the smaller net impurity density.

The charging current for the transition region capacitance consists of *majority* carriers. However, the junction current of most interest consists of *minority* carriers which have been either collected by the junction or injected through it, as explained below.

When voltage is applied across the junction, it has very little effect on the flow of minority carriers into the

junction since these carriers are aided by the field of the junction and limited in numbers by the rate of generation and diffusion of minority carriers to the junction. In marked contrast, the effect on the diffusion current (majority carriers) which is trying to surmount the potential barrier at the junction, is drastic. In mathematical terms, the minority carrier densities in either side of the junction (just outside the space charge region) are related exponentially to the applied voltage—thus applied voltage disturbs the thermal equilibrium conditions [see (12), (13)]. The current that flows with an applied voltage can be calculated from the boundary conditions, the equations for current density [see (4), (5)] and the equations of continuity [see (14), (15)]. Eq. (16), (17), and (18) give the resulting electron, hole, and total currents.

If the applied voltage is in the forward direction (p-type positive), the potential barrier to diffusion of majority carriers across the junction is reduced and copious current flows. In a good junction, recombination within the space-charge layer is negligible, and the current which consists of majority carriers as it enters the junction must of necessity be minority carrier current when it leaves the junction. For forward bias the minority carrier densities on either side of the junction are increased above the thermal equilibrium value. The added minority carriers tend to diffuse into the body of the semiconductor and to combine with majority carriers to restore thermal equilibrium. The recombination of minority carriers with majority carriers is a random process, with the excess carriers living on the average a time τ (lifetime) before recombining. (Values of τ from 10^{-7} to 10^{-3} seconds have been observed.) In the lifetime τ the carriers diffuse an average distance L (diffusion length) into the body of the crystal. ($L=\sqrt{D\tau}$, values of L in the range from 10^{-3} cm to 10^{-1} cm have been observed.) The effect of lifetime and diffusion constant on the junction current is expressible in terms of the diffusion length. For a given applied voltage, the current is inversely proportional to the diffusion length, or average distance the carriers must diffuse before recombining. For applied voltage in the reverse direction (n-side positive), the potential barrier to diffusion is increased and the minority carrier densities are decreased from the thermal value. The current consists essentially of minority carriers which diffuse out of the body of the semiconductor and drift across the junction. The reverse current density is substantially independent of reverse voltage for voltages not too large, and for resistivities commonly used is of the order of 10 ma/cm^2 for germanium junctions and 1μa/cm^2 for silicon junctions at 300°K. This difference in reverse current, or saturation current [see (19)] is due mainly to the difference in density of intrinsic carriers between the two materials. Fig. 3 shows the minority carrier densities near a junction for the conditions of both forward and reverse applied voltage.

Eq. (18) for the rectifier characteristic applies under conditions of both forward and reverse bias, and results in a degree of rectification which could not be verified experimentally until the advent of high-quality p-n junctions contained inside single crystals of germanium or silicon. Under proper conditions of minimum surface leakage and good crystal perfection the V-I characteristic of a p-n boundary conforms very closely to (18) at moderate voltages. There are still, however, some unexplained departures from the ideal rectifier equations.[7-9]

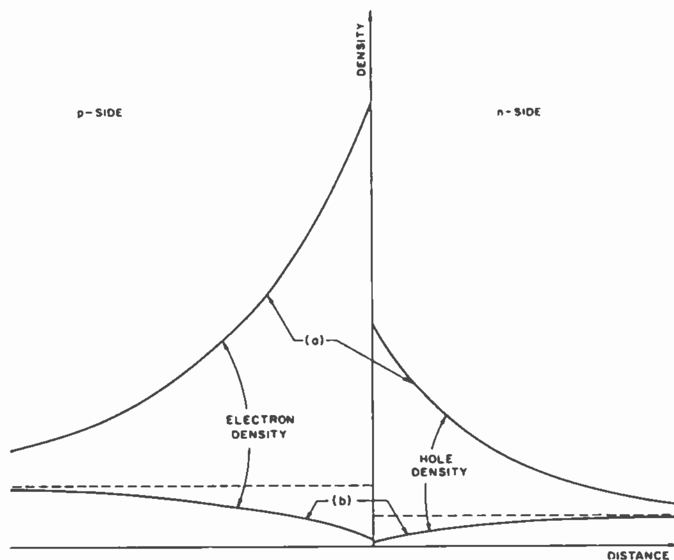


Fig. 3—Minority carrier densities near a p-n junction under the condition (a) forward bias, and (b) reverse bias. The transition region is not shown.

As the reverse voltage (*i.e.*, voltage in the low conducting direction) is increased, the barrier layer or space-charge layer increases in width and the maximum fields increase. Finally, at a critical voltage, the junction begins to draw copious current in the reverse direction and the phenomenon of breakdown takes place. Definitive proof of the nature of breakdown awaited the attainment of near-perfect p-n junctions in single crystals. Experiments of McKay and McAfee have shown that in the high fields encountered in a junction near breakdown the carriers attain enough energy to ionize valence electrons by collison.[10] For each valence electron ionized there has been created a new pair of carriers, *viz.*, the valence electron has become a conduction electron and it has left behind a free hole in the valence structure. When the field is high enough so that on the average each carrier pair so created can produce one more pair,

[7] M. B. Prince, "Silicon Conductivity Modulated Diffused p-n Junction Rectifiers," Talk given at American Phys. Soc. Meeting, Washington, D. C.; April 28–30, 1955.

[8] R. N. Hall, "Power rectifiers and transistors," Proc. IRE, vol. 40, pp. 1512–1518; November, 1952.

[9] H. Kleinknecht and K. Seiler, "Einkristalle and p-n Schichtkristalle aus Silizium," *Zeit. Phys.*, vol. 139, pp. 599–618; December, 1954.

[10] K. G. McKay and K. B. McAfee, "Electron multiplication in silicon and germanium," *Phys. Rev.*, vol. 91, p. 1079, September, 1953.

the discharge is self-maintaining. Since the current can then be quite large, and voltage substantially independent of current, the junction is said to be "broken down." The mathematical model proposed by Townsend for the avalanche breakdown in gases fits the semiconductor case almost perfectly.[11]

At voltages less than the breakdown voltage, there is a finite chance that any minority carrier moving across the junction will cause a secondary ionization. The secondary ionization can conveniently be characterized by a current avalanche multiplication factor M. The total junction current is M times the incident primary minority carrier current. Typical plots of M vs V/V_B are shown in Fig. 4. The shape of this plot is dependent on the value of V_B as well as the type of carrier in germanium,[12] but the measurements in silicon are not yet precise enough to show any such dependence. Eqs. (20), (21), and (22) give empirical analytical approximations to the variation of M with V/V_B. The theory of the avalanche process makes the breakdown voltage V_B roughly inversely proportional to the donor or acceptor density adjacent to the junction. Fig. 5 is a plot of V_B vs $N_D - N_A$ on the high-resistivity side of a step junction for silicon and germanium. It is assumed here that the low-resistivity side has such a high concentration of impurities that it does not affect the field distribution. There are significant departures from the inverse relationship: both plots have approximately a -0.7 slope on the log-log scale, rather than the -1 slope of a true inverse relation.

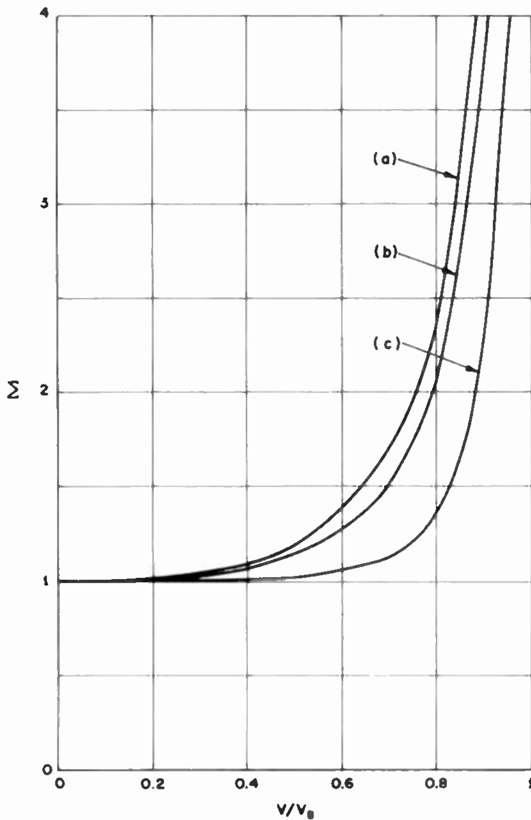


Fig. 4—Avalanche multiplication M vs V/V_B. (a) Electrons and holes in silicon,[11] (b) Holes in germanium $15 < V_B < 150$,[12] (c) Electrons in germanium $15 < V_B < 150$.[12]

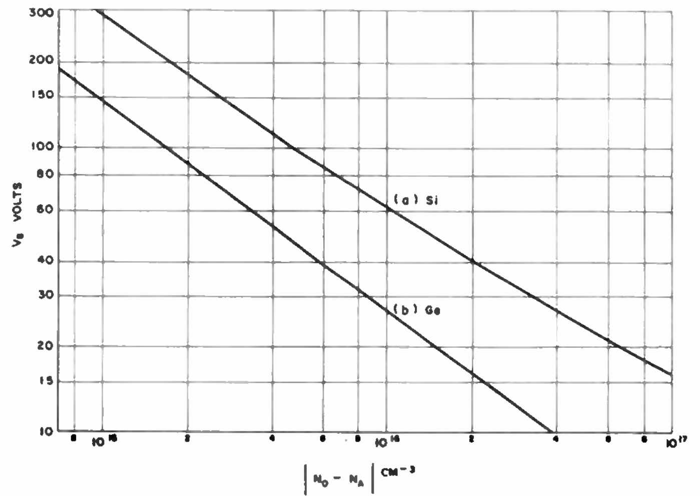


Fig. 5—V_B vs $|N_D - N_A|$ for step junction. (a) Silicon data obtained from D. K. Wilson), (b) Germanium.[12]

Fig. 6 shows a complete V-I characteristic for a p-n junction or diode. At low voltages and currents, the theoretical diode characteristic of (18) is followed. At high voltages, avalanche multiplication causes departures from this characteristic; and at high currents, the series resistance of the semiconductor regions may result in an appreciable voltage drop.

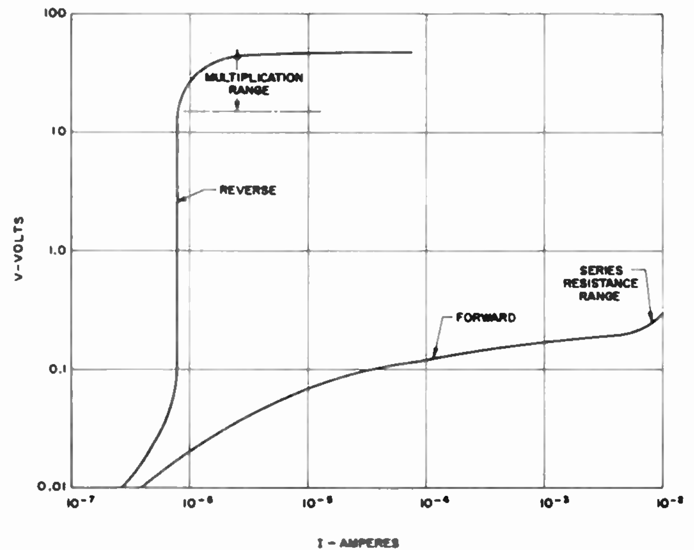


Fig. 6—p-n junction $V-I$ characteristic.

p-n-p AND n-p-n TRANSISTORS

A junction transistor consists of two junctions placed in close proximity to each other, as illustrated for the p-n-p transistor of Fig. 7 (next page). Discussion will be

[11] K. G. McKay, "Avalanche breakdown in silicon," *Phys. Rev.*, vol. 94, pp. 877–884; May, 1954.

[12] S. Miller, "Avalanche breakdown in germanium," *Phys. Rev.*, vol. 99, pp. 1234–1241; August, 1955.

centered on *p-n-p* structure but applies equally well to the *n-p-n* with the proper exchange of polarities.

In Fig. 7 the collector junction to the right is "reverse" biased so that normally very little current would flow—this due to thermally generated minority carriers in the body of the semiconductor. For typical germanium transistors this current—called I_{c0} or collector current with zero emitter current—is a few microamperes at room temperature and is millimicroamperes for silicon.

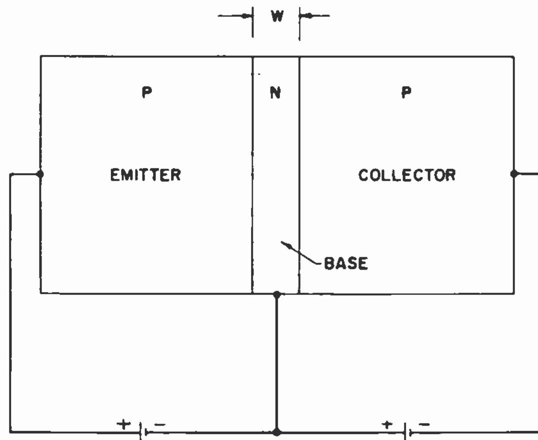


Fig. 7—*p-n-p* junction transistor with applied bias.

The emitter junction on the left is forward biased. This results in electrons entering the *p* region on the left, or emitter region, and in holes flowing from the emitter into the base layer. The holes that enter the base layer are "minority" carriers and, after diffusing across the base layer, are accelerated across the collector junction, thus contributing to the collector current. The fraction of emitter current that flows across the collector junction is called the α of the transistor, and may be as high as 0.99 even at low collector voltages. This type of operation results in amplification in a manner somewhat analogous to vacuum-tube amplification. The emitter, base, and collector of the transistor are roughly analogous to the cathode, grid, and plate of a vacuum tube. A small amount of control power applied to the emitter-base diode of a transistor controls a large amount of power in the collector circuit; just as a small amount of power applied to the grid circuit of a vacuum tube controls a large amount of power in the plate circuit. The control of large amounts of power by small amounts of power is, of course, the essence of amplification.

The current gain, or α, is defined as the incremental ratio of collector current to emitter current. The use of the current density, continuity equations [(4), (5), (14), (15)] and the junction boundary conditions [(12), (13)] makes possible the straightforward calculation of α for small minority carrier densities.[13] The problem of calculation of α is most easily solved by dividing it into three parts: (1) The calculation of emission efficiency, γ, or ratio of hole current at the emitter to total emitter current (for *p-n-p*). The hole current at the emitter is minority current in the base layer and hence can contribute to collector current. (2) The calculation of transport efficiency, β, or ratio of minority hole current at the collector to minority hole current at the emitter. Since the collector is assumed reverse biased, the hole density in the base near the collector is maintained near zero, resulting in a diffusion gradient across the base. The carriers require a finite time to diffuse across the base and so have a finite chance of recombining during the transit. (3) The calculation of the collection efficiency α^*, or ratio of total collector current to minority current reaching the collector. The current gain α is the product of emission, transport, and collection efficiency.

The electrons that are injected from the base layer into the emitter region must diffuse one electron diffusion length into the emitter before recombination. The holes that are injected into the base layer are collected after diffusing across the base layer. At a given forward bias on the emitter, the electron current is proportional to the ratio of the thermal equilibrium density in the emitter to the diffusion length for electrons. [Eq. (17) applies.] The hole current, however, is dependent on the ratio of thermal equilibrium hole density in the base layer to base layer width W. (Eq. (16) applies with L_p replaced by W.] If the emitter resistivity is much lower than base layer resistivity ($n_p \ll p_n$), and base-layer width of the same order of magnitude as the electron diffusion length in the emitter, the emission efficiency approaches unity [see (23)].

The transport efficiency also approaches unity if the base width W is much less than the diffusion length for holes in the base layer [see (24)]. Since, on the average, the holes will diffuse one diffusion length into the semiconductor before recombining, they have a very high probability of diffusing a much shorter distance without recombining. Emission efficiency γ and transport efficiency β are always less than unity.

The collection efficiency can be greater than unity if the collector region resistivity is sufficiently high that minority carriers contribute significantly to the conductivity. In this case, the holes entering the collector *p* region require appreciable field to move them away from the junction. This field accelerates minority electrons in the collector region towards the junction and causes an extra current to flow in addition to the hole current [see (25)]. If the collector resistivity is sufficiently low that the collector conductivity is almost entirely due to the majority hole density, this effect is negligible. An additional effect which can make the collection efficiency greater than unity is M, the avalanche multiplication factor. If the collector voltage is much less than the breakdown voltage, the avalanche multiplication factor is unity; but for high alpha transistors may be significant at voltages as low as $\frac{1}{3}$ the breakdown value. In its usual application the junction transistor has alpha less than unity. Many transistor circuits that are stable with alpha less than unity become unstable

[13] W. Shockley, M. Sparks, and G. K. Teal, "The *p-n* junction transistors" *Phys. Rev.*, vol. 83, pp. 151–162; July, 1951.

if alpha becomes greater than unity. For this reason, the effects which make collection efficiency greater than unity (and hence can make alpha greater than unity) place some design restrictions on transistors.

Since the transport of carriers occurs by diffusion, there is a finite delay between emitter current and collector current as well as a dispersion of carriers. The effect of the delay and dispersion can most easily be expressed as the frequency cutoff of β, or f_β, which is defined as the frequency where the dispersion is so great that the current transport factor has fallen to $1/\sqrt{2}$ of its low frequency value. The β frequency cutoff (26) is very nearly proportional to the inverse of the transit time for carriers across the base layer. Transit time is proportional to the square of the base layer width and inversely proportional to the diffusion constant. If the emitter transition region capacitance is very high, the emitter current may be shunted through this capacitance at high frequencies, resulting in a decrease in γ with frequency. The frequency cutoff of alpha, or frequency at which α has fallen to $1/\sqrt{2}$ of its low-frequency value, is approximately the smallest of the transport and emission cutoff frequencies.

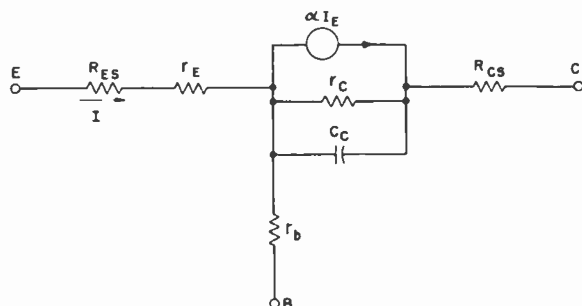


Fig. 8—Transistor equivalent circuit.

The circuit behavior of the transistor can be approximately represented by the equivalent circuit of Fig. 8. This circuit cannot, of course, give a perfect representation of a transistor. It can give a reasonably good approximation to the behavior over limited ranges of current, voltage and frequency. In Fig. 8 r_E is the slope resistance of the forward biased emitter-base diode and is about 25 ohms at 1 ma I_E.r_B is the effective spreading resistance from the center of the base layer to the base contact, and in many transistor designs is in the range of 50 to 200 ohms. C_c is the collector junction transition region capacitance and may be as low as 1 $\mu\mu$f for very small areas or as high as several hundred $\mu\mu$f for large transistors designed for handling high power. R_{Cs} and R_{Es} are the series resistances introduced by the semiconductor material in the collector and emitter regions and are a few milliohms for transistors prepared by alloying or plating the emitter and collector. For grown-junction transistors the range is up to a few tens of ohms for R_{Es} and a few hundreds of ohms for R_{Cs}.r_c is the collector leakage resistance and arises because α increases with collector voltage[14–16] due to collector space charge layer widening and collector current multiplication. Typical values of r_c at a few milliamperes collector current are of the order of a megohm. The α current generator is of course frequency dependent, decreasing in magnitude with increasing frequency.

Application as a Switch

The significance of many of the transistor parameters can be seen in terms of the behavior of the transistor as a switch. In its simplest form, switching means opening or closing an electrical circuit, or the changing of the state of an electrical circuit from high conduction to low conduction or vice versa. In more complicated relationships, all computing and control functions are forms of switching. Switching devices are naturally divided into two classes, regenerative and nonregenerative. A regenerative switch holds itself on or off while a nonregenerative switch must be held on or off. Our discussion is limited to the nonregenerative type.

The general requirements of a good switching device are:

1. The device shall have two states of conduction, namely, low conductance or "off" and high conductance or "on."
2. The "off" state should approximate an open circuit.
3. The "on" state should approximate a short circuit.
4. The amount of drive to hold the switch open or closed should be small (for nonregenerative switches).
5. The switch should open or close very rapidly.

Any amplifier can operate as a switch. The transistor as a switch is small, rugged and efficient; it can be designed to give a range of switching parameters unattainable by any other known device.

For switching applications, we must know the electrical properties of the transistor over all possible operating ranges. These properties are most easily explained in terms of a graphical representation of the variables. Actually, there are four variables to be considered; namely, collector current and voltage, and emitter current and voltage. The emitter-base voltage is so small that we will neglect it for the present and consider collector current and voltage as a function of emitter current, as in Fig. 9, next page. When $I_E=0$, the collector acts like a reverse biased diode and only I_{c0} flows; *i.e.*, a few microamperes for a typical germanium transistor, and a few millimicroamperes for silicon at room temperature. When the voltage approaches V_B, avalanche

[14] J. M. Early, "Effects of space charge layer widening in junction transistors," Proc. IRE, vol. 40, pp. 1401–1406; November, 1952.
[15] S. L. Miller and J. J. Ebers, "Alloyed junction avalanche transistors," *Bell Sys. Tech. Jour.*, vol. 34, pp. 883–902; September, 1955.
[16] M. C. Kidd, W. Hassenberg, and W. M. Webster, "Delayed collector conduction, a new effect in junction transistors," *RCA Rev.*, vol. 16, pp. 16–33; March, 1955.

multiplication sets in and the collector current increases rapidly. At I_E greater than zero and constant, the collector current is essentially a constant fraction of the emitter current from zero collector voltage up to voltage close to V_0 where the multiplication phenomenon becomes significant. In Fig. 9, V_0 is taken as that voltage where the multiplication is sufficient to make α reach one. When the collector voltage becomes positive, the collector junction becomes forward biased and begins to emit holes back into the base. The collector current decreases to zero at a forward collector voltage approximately equal to the emitter voltage. The collector characteristic for $I_E=0$ is the nonconducting or "off" state of the transistor. The essentially horizontal characteristics which occur when the collector voltage becomes positive correspond to the conducting or "on" state of the transistor switch.

Fig. 10 shows some of the simplest possible switching circuits. These circuits are all nonregenerative; *i.e.*, the switch must be held "on" or "off" by means of the control signal. In each case the transistor is turned off by applying a reverse bias to the emitter-base diode, and turned on by making the control current so large that

$$|I_E| > \left|\frac{V_{cc}}{\alpha R_L}\right|. \qquad (1)$$

Table II summarizes the drive conditions required to hold each of these simple circuits on or off. Table III lists the maximum available current and voltage gain

TABLE II

CONTROL REQUIREMENTS FOR NONREGENERATIVE TRANSISTOR SWITCHES

	Current to hold ON	Approximate voltage to hold ON	Voltage to hold OFF
Common base	$\frac{V_{cc}}{\alpha R_L}$	1/4 volt for Ge 3/4 volt for Si	0 volts
Common emitter	$(1-\alpha)\frac{V_{CC}}{\alpha R_L}$	−1/4 volt for Ge −3/4 volt for Si	from 0 to 1/10 volt
Common collector	$\frac{(1-\alpha)\ V_{cc}}{R_L}$	−1/4 volt for Ge −3/4 volt for Si	$\geq V_{CC}$

TABLE III

MAXIMUM AVAILABLE TRANSISTOR SWITCH GAIN

	Current gain	Voltage gain
Common base	α	$\frac{V_{cc}}{V_{E-B}}$
Common emitter	$\frac{\alpha}{1-\alpha}$	$\frac{V_{cc}}{V_{E-B}}$
Common collector	$\frac{1}{1-\alpha}$	≤ 1

through the switch. By switch gain we mean the ratio of controlled current, voltage or power to current, voltage or power required in the control circuit.

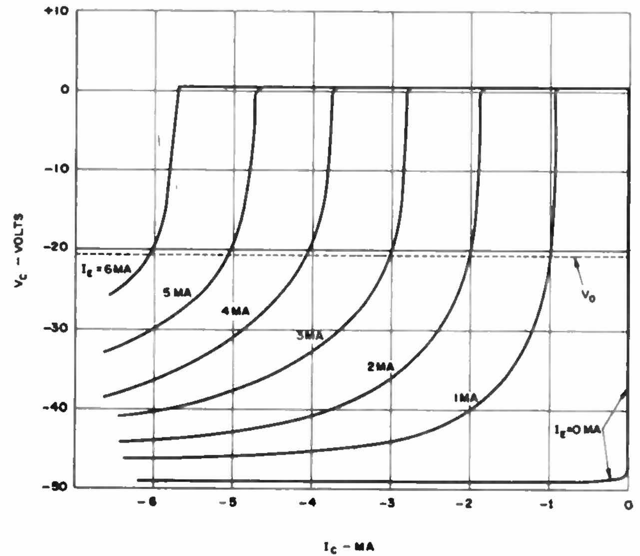


Fig. 9—Collector characteristics for a *p-n-p* transistor.

For each of the three circuits the power gain is greater than unity, but only the common emitter switch has both current and voltage gain. The maximum voltage gain is dependent on the maximum allowable voltage V_0 as well as the emitter-base voltage when the transistor is turned on. For germanium, the emitter-base voltage is of the order of $\frac{1}{4}$ to $\frac{1}{2}$ volt, and for silicon transistors it is $\frac{3}{4}$ to 1 volt.

The common emitter or common collector switch has the lowest on-resistance, since in this connection the base resistance is not in the load circuit, and the "on" junction voltages tend to cancel out. The collector-emitter voltage drop has been calculated for the common emitter and common collector circuits[17] and is given in (27) and (28) respectively. The series resistance introduced by the resistances of the emitter and collector regions, R_{Es} and R_{Cs} adds directly to the impedance between emitter and collector so that fabrication techniques that result in large series resistances seriously limit the applicability of the transistor as a switch. For transistors prepared by the method of alloying or by plating surface barrier contacts, the series resistances R_{Es} and R_{Cs} are very nearly zero. In this case the V_{CE}, I_C curve intersects the $I_C=0$ line at a voltage which is of the order of tens of millivolts and has a slope resistance of the order of ohms. The very good closed circuit thus attained has been used to advantage in low level sampling and logic circuits.[18,19]

[17] J. L. Moll and J. J. Ebers, "Large signal behavior of junction transistors," PROC. IRE, vol. 42, pp. 1761–1772; December, 1954.
[18] R. L. Bright, "Junction transistors used as switches," *AIEE Transactions*, vol. 74, pt. I, pp. 111–121; March, 1955.
[19] R. H. Beter, W. E. Bradley, R. B. Brown, M. Rubinoff, "Surface Barrier Transistor Switching Circuits," Abstract IRE Convention Record, 1955 National Convention Part 4, p. 139.

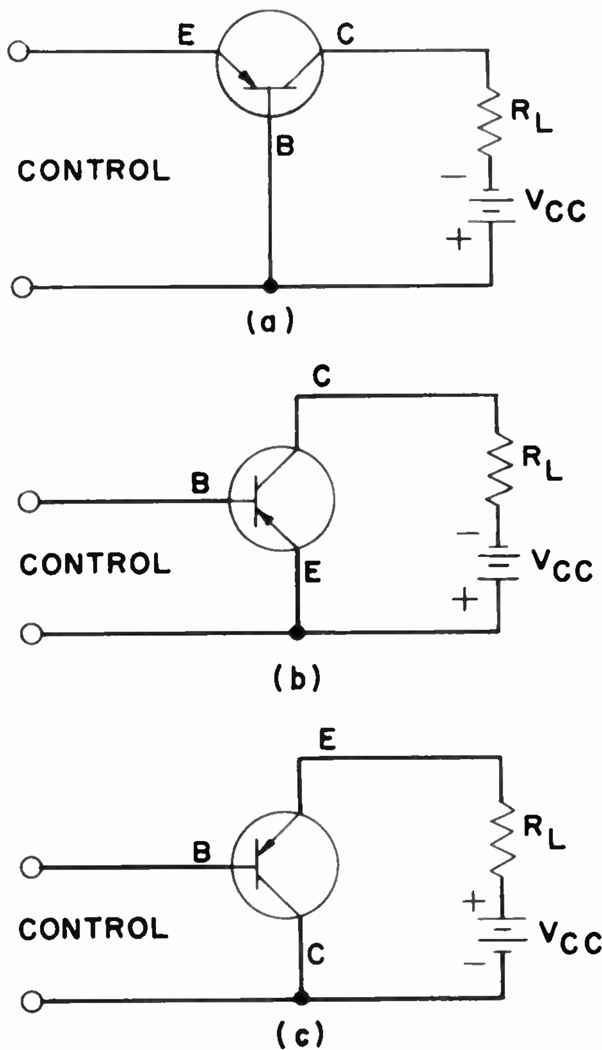


Fig. 10—Nonregenerative switches: (a) Common base, (b) Common emitter, (c) Common collector.

The off impedance of the switch is essentially the parallel combination of collector leakage and collector junction capacitance. In the common emitter and common collector circuits this leakage and capacitance may actually be multiplied by the gain of the transistor unless the base-emitter diode is slightly reversed biased—hence the requirement of about 1/10 volt to hold the switch off. If the emitter-base diode is reverse biased, the I_{c0} current divides between the collector and emitter, so that the collector conducts less than I_{c0}. The static off current can be as low as a few microamperes in germanium, and a few millimicroamperes in silicon at room temperature. The collector capacitance depends on collector voltage as well as base material resistivity and collector area, but may be as low as a few $\mu\mu$f for very small areas and up to hundreds of $\mu\mu$f for transistors designed to switch many watts of power.

The speed of switching is proportional to the amount of drive and frequency response of α or f_α.[20] High-frequency response requires thin base layers, resulting in a fundamental conflict between large power handling capability and fast response in the alloy type transistor.[21] The power handling capability of a transistor switch is proportional to the product of the maximum voltage it can withstand when in the off state and the maximum current it can safely conduct when in the on state. If the base layer is made very thin, the collector space charge region may extend across the entire base layer at a low voltage, and "punch-through" results. At punch-through, the transistor behaves as though the collector were connected internally to the emitter through a battery. A transistor is not capable of switching off a voltage greater than the punch-through voltage. The punch-through voltage for a given base layer thickness can be increased by lowering the resistivity of the base region, but this lowers the voltage at which avalanche multiplication occurs. Many transistor circuits become unstable if α becomes greater than unity so that collector multiplication places a further restraint on the maximum voltage. Designs capable of withstanding 10–20 volts and with f_α of the order of 50–100 mc are possible in germanium alloy type transistors. If silicon is used, the attainable frequency response is decreased by a factor of from 2 to 3 because of the lower mobility of the charge carriers. Transistors with frequency response to 50 or 100 mc will switch on or off in times of the order of 10 to 30 millimicroseconds.

Transmission Applications

Transmission, as its name implies, is the sending of an electrical signal from one point to another. Amplifiers are required to make transmission practical in the communications sense. Some of the attributes of a good transmission amplifier include:

1. Large power gain.
2. High-frequency response
3. High power handling capability.
4. Linearity and stability.
5. Low noise.

It is true that all of the desirable properties of the transistor are never maximized in the same design. In particular, a design for high power handling capability will in general sacrifice frequency response. We will limit the discussion to design for high frequency. A figure of merit which is a measure of the applicability of the transistor at any given frequency is the maximum frequency of oscillation[22] (29). The maximum frequency of operation is proportional to the geometric mean of the frequency cutoff of α and the reciprocal of the r_bC_c time constant. A good high-frequency design must therefore maximize f_α and at the same time minimize the product r_bC_c. There have been several distinct approaches to transistor design to maximize the frequency response. A few of them are described on the following pages.

[20] J. L. Moll, "Large signal transient response of junction transistors," Proc. IRE, vol. 42, pp. 1773–1784; December, 1954.

[21] J. J. Ebers and S. L. Miller, "Design of alloyed junction germanium transistors for high speed switching," *Bell Sys. Tech. Jour.*, vol. 34, pp. 761–781; July, 1955.

[22] R. L. Pritchard, "High-frequency power gain of junction transistors," Proc. IRE, vol. 43, pp. 1075–1085; September, 1955.

The tetrode transistor[23] is an *n-p-n* transistor grown within a single crystal of germanium and has two base contacts on opposite sides of the *p* layer, as shown in Fig. 11(a). A bias voltage is applied between the two base leads, and all of the transistor action is squeezed toward the positively biased lead. This results in a very low value of base resistance r_b. To attain high f_α, the base layer in the grown crystal must be thin. For low C_c, the cross section of the bar becomes small. Experimental tetrodes have oscillated at frequencies greater than 1,000 mc. The tetrode design tends to sacrifice power handling capability to attain frequency response.

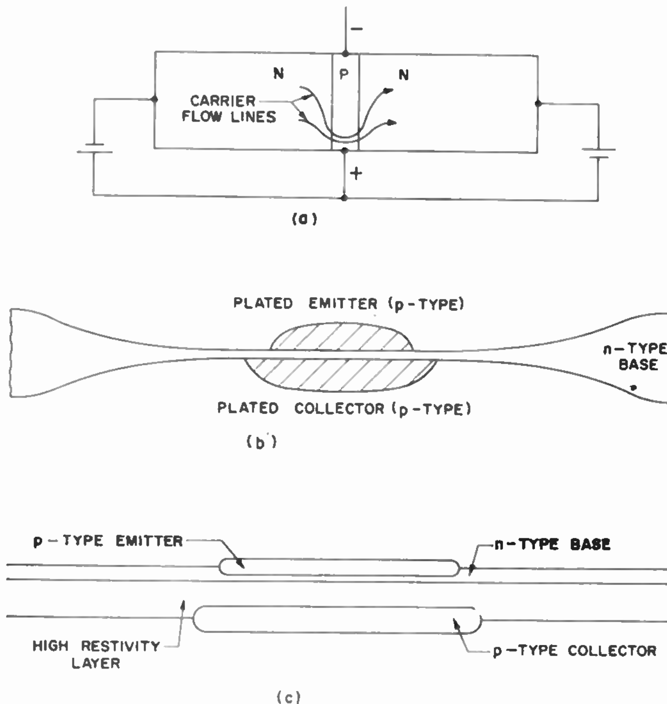


Fig. 11—High frequency transistor designs. (a) Tetrode, (b) Surface barrier, (c) *p-n-i-p*.

The surface barrier[24] transistor is made by jet etching of a germanium wafer to the order of 0.0001 inch thick and jet plating emitter and collector contacts on opposite faces, as in Fig. 11(b). This transistor uses a thin base layer to attain high f_α, small area to attain low C_c, and low resistivity base layers to attain low r_b. The surface barrier design attains frequency response at the expense of power handling capability, but retains the feature of being able to switch to very low emitter to collector voltage, making this design useful as a low level high-speed switch, or low level high-frequency amplifier.[25]

The *p-n-i-p*[26] transistor again uses a thin base layer to attain high frequency but adds an extra layer of high resistivity material between the base layer and collector region as in Fig. 11(c) to increase the collector-base transition layer thickness and hence decrease collector capacitance per unit area. The base layer is again made of low resistivity material to reduce r_b. By reducing capacitance per unit area and at the same time increasing the avalanche breakdown voltage of the collector junction, the *p-n-i-p* design attains high-frequency response with a greater power handling ability than in the case of the tetrode or surface barrier transistor. Experimental results with *p-n-i-p* structures include power output of 20 milliwatts at 250 mc with 160 milliwatts dc input power.[27]

The addition of a resistivity gradient in the base layer[28,29] with resistivity increasing from emitter to collector, results in "built-in" fields that cause minority carriers to drift from emitter to collector. In practical designs, the built-in field allows the base layer to be about twice as thick as for a transistor with uniform base layer resistivity and equal frequency cutoff. By using diffusion techniques it is possible to construct transistors with very thin, low resistivity base layers. Transistors fabricated in this way have a resistivity gradient in the base layer, with resistivity increasing from emitter to collector. The collector region resistivity can be much higher than the base layer resistivity, so that the collector space charge layer extends into the collector body, thus avoiding punch-through. Transistors have been fabricated with the diffusion techniques which have f_α of over 600 mc. These transistors have been operated in video amplifiers giving 20-db power gain per stage and flat to 20 mc within $\frac{1}{2}$ db.[30]

The transistor noise figure, as well as stability, are less easily designable figures than frequency response or power capability. The transistor stability is intimately related to the chemistry and physics of the free-space surfaces of the germanium or silicon. The experience with transistors in systems has been very encouraging as to stability. A transistor carrier repeater system has a replacement rate corresponding to 150,000 transistor hours per failure,[31] with comparable figures for other systems. This reliability compares quite favorably with present-day vacuum tubes, but the increasing complexity of systems will certainly place increasing reliability standards on electronic devices.

The noise figure of the transistor is also undoubtedly

[23] R. L. Wallace, Jr., L. G. Schimpf, and E. Dickten, "A junction transistor tetrode for high-frequency use," Proc. IRE, vol. 40, pp. 1395–1400; November, 1952.

[24] W. E. Bradley, J. W. Tiley, R. A. Williams, J. B. Angell, F. P. Keiper, R. Kansas, R. F. Schwarz, and J. F. Walsh, "The surface-barrier transistor," Proc. IRE, vol. 41, pp. 1702–1720; December, 1953.

[25] R. J. Turner, "Surface barrier transistor measurements and applications, *TeleTech.*, vol. 12, pp. 78–80, August, 1954.

[26] J. M. Early, "PNIP and NPIN junction transistors triodes," *Bell Sys. Tech. Jour.*, vol. 33, pp. 517–533; May, 1954.

[27] J. M. Early, "Very High Frequency Junction Transistors," Talk at AIEE Meeting, Columbus, Ohio; May 4–6, 1955.

[28] H. Krömer, "Der Drifttransistor," Die Naturwissenschaften, vol. 40, pp. 578–579; November, 1953.

[29] H. Krömer, "Zur Theorie des Diffusions- und des Drifttransistors," I, II, and III. *Arch. Elect. Übertragung*, vol. 8, pp. 223–228; 363–369, 499–504; May, August, November, 1954.

[30] C. A. Lee and D. E. Thomas, "A new high frequency *p-n-p* transistor," Talk at IRE Transistor Research Conference, Philadelphia, Pa.; June 20–22, 1955.

[31] Personal communication from R. M. Ryder.

related to the surface chemistry as well as to the internal structure of the device. However, in spite of the absence of a definitive understanding of the sources of noise, junction transistors are currently available which have noise figures comparable to present day vacuum tubes.

Conclusions

Increased knowledge of the nature of semiconductors and surfaces bounded by semiconductors of different type, as well as quantitative control of minute quantities of impurities in single crystal germanium and silicon has made the transistor possible. The fundamental understanding of semiconductors has resulted in almost complete understanding of the electronic behavior of junction transistors. The greatest gaps in our present-day knowledge relate to the physics and chemistry of the free surfaces.

Transistors with switching parameters in a range unattainable by any other known device are currently available. In addition, experimental transistors have been built which perform as well or better than vacuum tubes in video and IF applications, but the transistors are relatively much smaller and consume much less power. The reliability and noise figure of transistors compare quite well with the vacuum tube counterpart and both parameters should improve rapidly with the increasing understanding of surfaces.

Appendix

Formulas Relating to Semiconductor Electronics Definitions of Symbols

Drift velocity

$$v = \mu E. \quad (2)$$

Electron and hole drift current:

$$J_n = q\mu_n nE; \qquad J_p = q\mu_p pE. \quad (2a)$$

Electron and hole diffusion current:

$$J_n = qD_n \nabla n; \qquad J_p = -qD_p \nabla p. \quad (3)$$

Total transport current of holes (4) and electrons (5):

$$J_p = q\mu_p pE - qD_p \nabla p \quad (4)$$

$$J_n = +q\mu_n nE + qD_n \nabla n. \quad (5)$$

Einstein relation:

$$D = \frac{kT\mu}{q} \frac{kT}{q} = .026 \text{ volts at } 300°\text{K}. \quad (6)$$

Conductivity:

$$\sigma = \frac{J}{E} = q\mu_n n + q\mu_p p. \quad (7)$$

Thermal equilibrium condition: n_i^2 is given by (8a) and (8b) for *Ge* and *Si*,

$$n_0 p_0 = n_i^2 \quad (8)$$

$$n_i^2 = 9.3 \times 10^{31} T^3 e^{-8700/T} \text{ (Ge)} \quad (8a)$$

$$n_i^2 = 7.8 \times 10^{32} T^3 e^{-12,900/T} \text{ (Si)}. \quad (8b)$$

Formulas for space charge layer widths:

$$X_m = \left\{\frac{2\epsilon}{q}(V + V_i)\left(\frac{1}{N_p} + \frac{1}{N_n}\right)\right\}^{1/2} \text{(step)} \quad (9)$$

$$X_m = \left\{\frac{2\epsilon}{qa}(V + V_i)\right\}^{1/3} \text{(graded)} \quad (10)$$

$V_i \cong 0.3$ volts for germanium junctions at 300°K

$V_i \cong 0.8$ volts for silicon junctions at 300°K.

Junction capacitance:

$$C_T = A\frac{\epsilon}{X_m}. \quad (11)$$

Boundary conditions at junctions:

$$n_p = n_{p0} e^{q\Phi/kT} \quad (12)$$

$$p_n = p_{n0} e^{q\Phi/kT}. \quad (13)$$

Continuity equations for electrons (14) and holes (15)

$$\frac{dn}{dt} = -\frac{n - n_0}{\tau_n} + \frac{1}{q}\nabla \cdot J_n \quad (14)$$

$$\frac{dp}{dt} = -\frac{p - p_0}{\tau_p} - \frac{1}{q}\nabla \cdot J_p. \quad (15)$$

Hole, electron, and total transport currents at a planar p-n junction (rectifier):

$$I_p = A\frac{qD_p p_{n0}}{L_p}(e^{q\Phi/kT - 1}) \quad (16)$$

$$I_n = A\frac{qD_n n_{p0}}{L_n}(e^{q\Phi/kT} - 1) \quad (17)$$

$$I = I_p + I_n = I_s(e^{q\Phi/kT} - 1). \quad (18)$$

Saturation current for planar p-n junction

$$I_s = \left(\frac{qD_p p_{n0}}{L_p} + \frac{qD_n n_{p0}}{L_n}\right)A \quad (19)$$

$$= \left(\frac{qD_p}{L_p N_n} + \frac{qD_n}{L_n N_p}\right) n_i^2 A.$$

Multiplication factors for electrons and holes at step junctions; electrons or holes in silicon (20), electrons in germanium (21), holes in germanium (22):

$$M \cong \frac{1}{1 - \left(\frac{V}{V_B}\right)^{2.6}} \quad (20)$$

$$M_n \cong \frac{1}{1 - \left(\frac{V}{V_B}\right)^{6}} \quad (15 < V_B < 150) \quad (21)$$

$$M_p \cong \frac{1}{1-\left(\frac{V}{V_B}\right)^3} \quad (15 < V_B < 150). \tag{22}$$

Emission efficiency (p-n-p):

$$\gamma = \frac{1}{1+\frac{\sigma_B W}{\sigma_E L_{nE}}}. \tag{23}$$

Transport factor (p-n-p):

$$\beta = \frac{1}{\cosh \frac{W}{L_p}}. \tag{24}$$

Collection efficiency (p-n-p):

$$a^* = M_p\left(1+\frac{\sigma_{pc}}{\sigma_{nc}}\right). \tag{25}$$

Frequency cutoff (p-n-p):

$$f_\beta = \frac{1.2 D_p}{\pi W^2}. \tag{26}$$

"On" state voltage drop, common emitter (27), common collector (28):

$$V_{CE} = \frac{kT}{q}\ln\frac{1+\frac{I_c}{I_b}(1+\alpha_I)}{\alpha_I\left(1-\frac{I_c}{I_B}\frac{1-\alpha_N}{\alpha_N}\right)} - |I_c|(R_{C_s}+R_{E_s}) - |I_B| R_{E_s} \tag{27}$$

$$V_{EC} = +\frac{kT}{q}\ln\frac{\alpha_N\left(1+\frac{1-\alpha_I}{\alpha_I}\left|\frac{I_E}{I_B}\right|\right)}{1-(1-\alpha_N)\left|\frac{I_E}{I_B}\right|} + |I_E|(R_{C_s}+R_{E_s}) - |I_B| R_{C_s}. \tag{28}$$

Maximum frequency of oscillation:

$$f_{max} = \sqrt{\frac{f_\alpha}{25 r_b C_c}}. \tag{29}$$

A = junction area, cm^2
a = gradient of impurity density, cm^{-4}
C_c = collector capacitance, farads
C_T = transition region capacitance, farads
D = diffusion constant, $cm^2\ sec^{-1}$
D_p = diffusion constant for holes
D_n = diffusion constant for electrons
E = electric field, volt cm^{-1}
f_β = cutoff frequency of β
f_α = cutoff frequency of α
f_{max} = maximum frequency of oscillation
I = current, amperes
I_p = hole current
I_n = electron current
I_s = saturation current
I_c = collector current
I_E = emitter current
I_B = base current
J = current density, amperes cm^{-2}
J_p = hole current density
J_n = electron current density
k = Boltzman constant
L_p = diffusion length for holes, $L_p = \sqrt{D_p \tau_p}$
L_n = diffusion length for electrons, $L_n = \sqrt{D_n \tau_n}$
L_{nE} = diffusion length for electrons in emitter region
M = avalanche multiplication factor
M_p = avalanche multiplication factor for holes
M_n = avalanche multiplication factor for electrons
n = electron density, cm^{-3}
n_0 = thermal equilibrium electron density
n_i = thermal equilibrium electron density in pure (intrinsic) germanium or silicon ($n_0 = p_0 = n_i$ in the intrinsic semiconductor)
n_{p0} = thermal equilibrium electron density in p-type semiconductor
N_D = donor density, cm^{-3}
N_A = acceptor density
N_p = net impurity density in p-type semiconductor
N_n = net impurity density in n-type semiconductor
p = hole density
p_0 = thermal equilibrium hole density
p_{n0} = thermal equilibrium hole density in n-type semiconductor
q = electronic charge $= 1.6 \times 10^{-19}$ coulombs
r_E = differential resistance of emitter junction:

$$r_E \cong \frac{.026}{I_E} \text{ ohms at } T = 300°K$$

r_β = base layer spreading resistance, ohms
r_C = collector junction leakage resistance
R_{C_s} = series ohmic body and contact resistance in the collector body
R_{E_s} = series ohmic body and contact resistance in the emitter body
T = temperature, °K
v = velocity, cm sec^{-1}
V = applied voltage, volts
V_i = contact potential
V_B = avalanche breakdown voltage
V_{CE} = collector-emitter voltage
V_{EC} = emitter-collector voltage
W = base layer width, cm
X_m = width of transition region, cm
α = differential ratio of collector current to emitter current
α^* = collection efficiency
α_n = ratio of collector current to emitter current at zero collector volts, emitter forward biased

α_i = ratio of emitter current to collector current at zero emitter volts, collector forward biased
β = transport factor across base layer
γ = emitter emission efficiency
ϵ = dielectric permitivity, farads/cm
ϵ_0 = dielectric permitivity of free space
$\epsilon_0 = 8.85 \times 10^{-14}$ farads/cm
$\epsilon = \kappa\epsilon_0$ $\kappa(Ge) = 16$; $\kappa(Si) = 12$
μ = mobility of carriers cm^2 volt^{-1} sec^{-1}
μ_p = mobility of holes
μ_n = mobility of electrons
σ = conductivity
σ_B = conductivity of base layer
σ_E = conductivity of emitter region
σ_{pC} = conductivity due to hole density in collector region
σ_{nC} = conductivity due to electron density in collector region
τ_n = lifetime of excess electrons
τ_p = lifetime of excess holes
Φ = applied potential across p-n boundary, with the positive sense taken as p-type positive with respect to n-type

ACKNOWLEDGEMENT

The author wishes to thank his associates for many helpful discussions. R. M. Ryder gave considerable help and encouragement in writing this paper.

Photoconduction in Germanium and Silicon*

M. L. SCHULTZ† AND G. A. MORTON†, FELLOW IRE

Summary—**Germanium and silicon are probably the two best known of all of the semiconductors. Their electronic properties have been studied in great detail both theoretically and experimentally. The reason for this is that these two materials are of such great practical importance for transistors and rectifiers. Perhaps less generally appreciated is the fact that both of these materials are excellent photoconductors. Their intrinsic response extends through the entire visible spectrum into the near infrared. When doped with appropriate foreign atoms and operated at low temperatures, these materials are very effective impurity photoconductors with response extending into the far infrared portions of the spectrum.**

INTRODUCTION

An ADEQUATE understanding of the behavior of any semiconductor which exhibits a photoconductive response when subjected to a radiation stimulus is most readily obtained by considering separately the several processes which together constitute the photoconductive process. These include (a) production, when the material is exposed to radiation, of current carriers in excess of those present in thermal equilibrium in the dark; (b) motion of carriers under the influence of an electric field; and (c) temporary or permanent removal of excess carriers by trapping and recombination.

The excitation of current carriers, either electrons or holes or both, occurs from states to which the carriers are normally bound into states in which they are free to move in an electric field. Possible excitation processes are represented in the schematic energy level diagram of Fig. 1. An ideal semiconductor crystal may be described by an energy level diagram that includes a

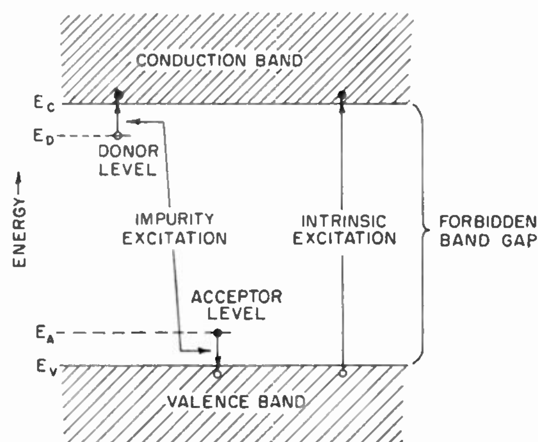


Fig. 1—Schematic energy level diagram illustrating the excitation processes which lead to the production of free charge carriers.

lower band of allowed energy states, the valence band, which is completely filled with electrons at absolute zero; a band of energies whose occupancy is forbidden; and a higher band of allowed states, the conduction band, which is empty of electrons at absolute zero. Actual semiconductor crystals always contain foreign atom impurities (either residual or deliberately added), and lattice imperfections. These introduce localized energy levels which may lie within the forbidden band gap of the material. Such levels are donors, if they possess an ionizable electron, or acceptors, if they can accept an electron; i.e., if they possess an ionizable hole.

Intrinsic excitation involves the transition of an electron from a state in the valence band to one in the conduction band. Each such excitation produces one free electron and one free hole. An excitation of an electron

* Original manuscript received by the IRE, August 29, 1955.
† RCA Laboratories, Princeton, N. J.

from a donor level into the conduction band produces one free electron and one bound hole, while excitation of a hole from an acceptor level into the valence band (i.e., excitation of an electron from the valence band into the acceptor level) produces one free hole and one bound electron. These excitations may be produced either by absorption of thermal energy from the crystal lattice or by absorption of photons of radiation of appropriate energy. If the excitation is produced by absorption of photons, photoconductivity, either intrinsic or impurity, is possible.

The energy associated with a photon of wavelength λ is given by

$$E = \frac{hc}{\lambda} \tag{1}$$

where h is Planck's constant and c is the velocity of light. The visible spectrum includes photon energy range from about 1.6 ev in the red to about 3.2 ev in the blue. The long wave-length threshold of photoconductivity associated with a given excitation process is given by

$$\lambda_{Th} = \frac{1.2395}{\Delta E} \text{ microns} \tag{2}$$

where ΔE is the energy difference in electron volts between the edges of the conduction and valence bands, or the energy difference between an impurity state and the edge of either the conduction or valence band, whichever is appropriate to the particular excitation process in question.

If optical excitation of carriers from impurity levels is to be possible, the semiconductor sample must be at a temperature low enough so that, in the absence of radiation, carriers are bound to the impurity levels involved. Since the impurity ionization energies E_D and E_A may be very small compared with the forbidden band gap E_G, cooling to very low temperatures; e.g., in liquid hydrogen or liquid helium, may be required to satisfy this condition. However, photoconductive response at very long wavelengths can then be observed.

The carrier concentration in the case of thermal excitation, for a given semiconductor and for a given set of impurity levels, is determined by the temperature of the crystal. The conductivity determined by this carrier concentration, or more precisely, the noise associated with random fluctuations in the conductivity, imposes a limitation upon the sensitivity of the material when used as a photoconductor.

The time rate of increase of the concentration of carriers excited by monochromatic radiation of wavelength λ in a layer of photoconductor of thickness dx at a depth x below the irradiated surface is given by

$$\left(\frac{dn}{dt}\right) = \eta(\lambda) f(\lambda)\alpha(\lambda) \exp(-\alpha(\lambda)x) \tag{3}$$

where n is the concentration of carriers of a given type, $\eta(\lambda)$ is the quantum efficiency, $f(\lambda)$ is the number of photons per sq. cm. per second entering the sample, and $\alpha(\lambda)$ is the absorption contant. At wavelengths for which the absorption constant is high, as in the intrinsic region, the excess carrier density is nonuniform in the x direction through the sample. If the absorption constant is sufficiently small, the above expression reduces to

$$\left(\frac{dn}{dt}\right) = \eta(\lambda) f(\lambda)\alpha(\lambda), \tag{4}$$

and a uniform density of carriers throughout the sample results. This condition usually applies for absorption in the impurity range. The impurity absorption constant may be written as

$$\alpha(\lambda) = \sigma(\lambda)n_i \tag{5}$$

where $\sigma(\lambda)$ is the absorption cross section and n_i is the concentration of un-ionized impurities.

The next consideration is the motion of current carriers under the influence of a field. In practice the velocity imparted to a carrier by the field is very small compared with the thermal motion of the carrier. Therefore, the photoconductive current is a small drift in the direction of the field superimposed upon a large random motion. The drift velocity can be shown to be proportional to the electric field, the constant of proportionality being known as the mobility μ. In general, the mobility μ_n for electrons will be different from mobility μ_p for holes.

The magnitude of the mobility in a given semiconductor is determined by the scattering of charge carriers by lattice vibrations, ionized impurity centers, neutral impurity centers, and lattice imperfections. These scattering processes are temperature dependent and different processes are dominant over different temperature ranges. At temperatures in the neighborhood of room temperature, the principal scattering is due to thermal vibrations of the lattice. If the band structure of the semiconductor could be represented by simple spherical surfaces in momentum space, it can be shown [1] that the mobility will vary with the inverse three-halves power of the absolute temperature. However, for silicon and germanium, the band structure in momentum space is more complicated so that the temperature variation for holes and electrons is somewhat greater. At low temperatures, the scattering is primarily due to impurities and imperfections in the lattice. In this temperature range, mobility decreases with decreasing temperature. Observed mobilities for germanium and silicon will be discussed in greater detail in a later section of this paper.

The excitation of carriers to the conduction and valence bands has been described above. In addition, there is a draining away of charge carriers by recombination processes. These include direct recombination, recombination at the surface of the specimen, and volume recombination at centers distributed throughout the

bulk of the material. The direct recombination of holes and electrons is a relatively rare occurrence. The time constant for this process in germanium has been calculated [2] to be of the order of one second. Since observed lifetimes of excess carriers are rarely greater than of the order of 10^{-3} second, the other recombination processes must be much more important. That direct recombination does occur in germanium and silicon, however, has been demonstrated [3, 4] by the observation of the emission of recombination radiation in wavelength ranges corresponding to the intrinsic gap widths.

The contribution due to surface recombination depends greatly upon the physical and chemical state of the surface. Where excitation occurs throughout the volume of the crystal, surface recombination will be small compared with volume recombination. For intrinsic photoconductivity, however, where excitation occurs close to the surface, surface recombination may be important.

Volume recombination involves the capture of a charge carrier of one sign by a deep-lying vacant level followed by the capture of a charge carrier of the opposite sign by the same level. This process is illustrated in Fig. 2. It is evident that two capture cross sections are involved here. First, the capture cross section of the level for the first type of charge carrier; second, the cross section of the occupied level for the capture of a charge carrier of the opposite sign. These cross sections

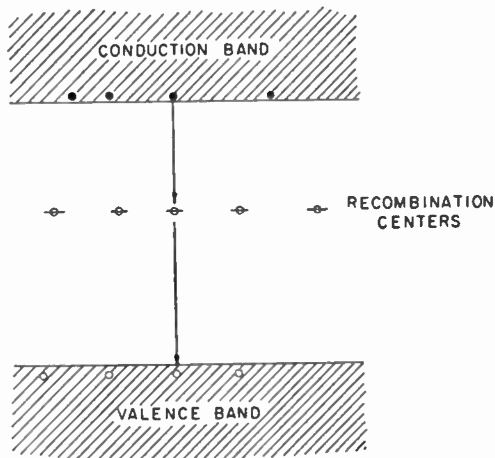


Fig. 2—Schematic energy level diagram illustrating the process of electron-hole recombination at centers.

may be very different in size. This recombination mechanism has been treated theoretically [5, 6]. Experimental verification of the theory has been obtained [7] by measurement of minority carrier lifetime as function of the concentration of recombination centers. Copper and nickel impurities have been shown to act as recombination centers of this type in germanium. It has also been shown [8] that at least a fraction of the recombinations occur with emission of radiation of wavelengths which may correspond to the location of the recombination centers in the forbidden band gap.

For steady state intrinsic photoconductivity, the rate of capture of holes and electrons must be equal to one another and to the rate of generation. Therefore,

$$n\sigma_{ec}\bar{v}_e(n_c - n_c') = p\sigma_{pc}\bar{v}_p n_c' = \left(\frac{dn}{dt}\right)_e = \left(\frac{dp}{dt}\right)_e \tag{6}$$

where n and p are the concentrations of free electrons and holes; $\bar{v}_e$ and $\bar{v}_p$ are the average thermal velocities of electrons and holes; σ_{ec} and σ_{pc} are the capture cross sections of the recombination centers for electrons and holes; n_c is the total concentration of recombination centers; and n_c' is the concentration of centers occupied by electrons. Expressions for the response time, and for the steady state photocurrent-intensity characteristics for intrinsic photoconductivity based upon equations for the rates of excitation and recombination similar to those given above, have been published by many authors; (see [9, 10]).

Where impurity photoconductivity is involved, recombination occurs when the center from which a charge carrier was excited captures a replacement charge. If the only source of carriers is these centers, the condition to be satisfied for equilibrium is

$$\left(\frac{dn}{dt}\right)_e = \sigma_{ec}\bar{v}_e n^2$$

or

$$\left(\frac{dp}{dt}\right)_e = \sigma_{pc}\bar{v}_p p^2. \tag{7}$$

A detailed discussion of the magnitude of response of an impurity photoconductor, including the effects of temperature, background radiation, compensating impurities, and nonphotoionizing absorption processes, has been given [11]. Sufficient experimental data is not yet available to provide an adequate test of the validity of such a treatment.

Centers or states in the forbidden energy band may also act as trapping centers. This occurs when the activation energy of the level is small enough so that the probability of thermal re-excitation of the captured carrier is greater than the probability of recombination by the capture of a carrier of the opposite sign. A general treatment of recombination and trapping processes applicable to any photoconductor has been given [10]. Some experimental evidence for the existence of trapping in germanium and silicon, usually obtained by study of the time rate of decay of photoconductivity or of quenching of photoconductivity, is now available [12–17].

The photocurrent produced in a photoconductor can be expressed by the formal relationship

$$i = e\frac{\tau_0}{T}F \tag{8}$$

where F is the number of photons absorbed per second, T is the time required for a carrier to cross the specimen, and τ_0 is the effective carrier lifetime. If the specimen

length is L, the applied voltage V and the mobility μ, the factor T is given by

$$T = \frac{L^2}{\mu V}. \tag{8a}$$

The factor τ_0 depends upon the recombination cross section and the carrier density. In the absence of traps, this lifetime is given by

$$\tau_0 = \frac{1}{v\sigma_c n} \tag{8b}$$

where v is the thermal velocity of carriers, σ_c the recombination cross section, and n the carrier density.

Photoconductivity of Intrinsic Germanium and Silicon

If a determination is made of the concentration of holes or electrons as a function of temperature by measurement of the Hall coefficient and conductivity, it is found that the activation energy of germanium in the neighborhood of room temperature is 0.68 ev [18]. This means that as far as thermal excitation is concerned, the energy required to raise an electron from the valence band to the conduction band is 0.68 ev. This is the quantity which has been called the forbidden band gap in Fig. 1. A more detailed knowledge of the structure of both the valence and conduction bands is required, however, in order to discuss optical transitions between them. A complete description of the energy band structure would require representation in a four dimensional manifold which includes the three crystal momentum coordinates and energy. If, however, appropriate crystallographic directions are chosen, two dimensional plots of energy vs crystal momentum k for such directions contain the essential information. Fig. 3(a) gives such plots for germanium [19] for the (100) and (111) crystallographic directions and depicts the dependence of electron energy upon k for the top-most valence bands and for the bottom-most conduction bands.

The thermal activation energy is the smallest distance between curves of the conduction band and the valence band indicated by the dotted lines on the figure. It will be observed that to make this transition an electron must change its momentum. The energy required to excite an electron from the valence band to the conduction band without change of momentum at $k=0$ is approximately 0.8 ev [4, 20]. When optical excitation involves an electron and a photon only, the principle of conservation of momentum requires that the excitation be vertical. Under these cirsumstances, the longest wavelength which can cause a transition is approximately 1.5 microns. However, if a phonon is either generated or absorbed in the course of excitation, there can be a change of momentum and the transition need not be vertical on the diagram.

The intrinsic absorption in germanium is illustrated in Fig. 4 [9]. The portion of the spectrum corresponding

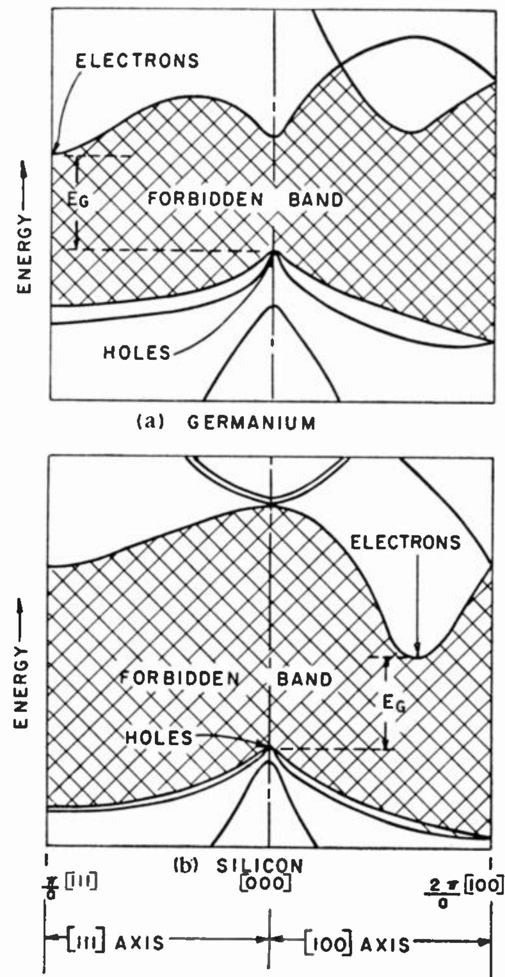


Fig. 3—Schematic diagrams of energy band contours in germanium and silicon crystals along (111) and (100) axes in the reduced zone.

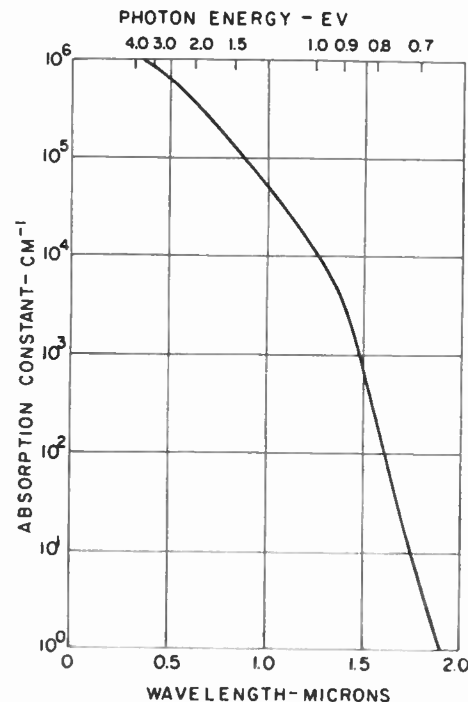


Fig. 4—Intrinsic optical absorption in germanium.

to photon energies greater than about 0.8 ev is due principally to vertical transitions, and is characterized by absorption constants in the range from about 10^3 cm^{-1} to about 10^6 cm^{-1}. The portion of the spectrum at lower energies is due to nonvertical transitions for which the transition probabilities are smaller ($\alpha < 10^3$ cm^{-1}). Recent detailed measurements [20, 21] of optical absorption have clearly demonstrated the existence of nonvertical transitions. Careful measurements [22] at the extreme tail of the absorption band permit the separate identification of transitions in which a phonon is absorbed, and of those in which a phonon is emitted. A theoretical treatment of vertical and nonvertical transitions has recently been given [23, 24].

The photoconductivity of germanium in the intrinsic range has not yet been studied in the same detail as has the optical absorption. A typical photoconductive response curve for germanium at room temperature, in which the response is expressed in arbitrary units for equal numbers of incident photons at all wavelengths, is given in Fig. 5 [9]. Proceeding from short to long wavelengths, the response increases somewhat until a maximum is reached at about 1.5 microns. Beyond about 1.6 microns, the response falls rapidly to a long wavelength threshold in the neighborhood of 1.8 microns which

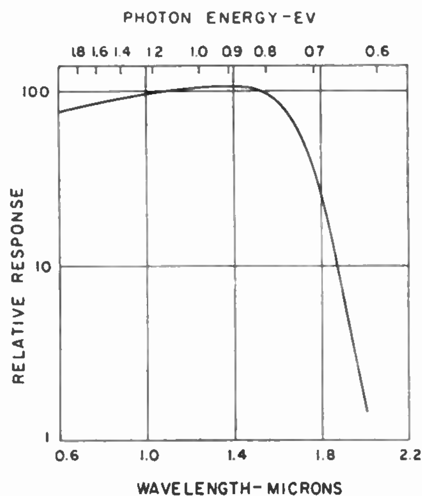


Fig. 5—Intrinsic photoconductivity of germanium.

corresponds approximately to the 0.68 ev of the minimum band separation. It may be noted that the maximum of photoconductivity occurs at approximately the wavelength at which the absorption constant begins to decrease rapidly. At short wavelengths where the absorption coefficient is very high, all of the photo conductive effect is confined to a very shallow layer on the side of the specimen upon which the radiation is incident. The high concentration of electrons and holes near the surface and the presence of additional recombination centers at the surface combine to produce rapid recombination and a smaller photoconductive current. When the wavelength gets longer than the minimum vertical transition so that the excitation is nonvertical involving the production or absorption of a phonon, the absorption coefficient decreases, and the radiation penetrates further into the material. As a result, surface recombination becomes less and less important. Therefore, although the rate of excitation of carriers remains constant as long as all of the radiation is absorbed in the material, the lifetime of the carriers becomes longer, and the photoconductive current becomes larger. This effect may account for the rise which has been noted in the photoconductive response curve.[1] The height of the maximum above the short wavelength portion of the curve varies considerably from sample to sample. These differences are probably due to the state of the surface and, therefore, to the magnitude of the surface recombination coefficient. The quantum efficiency, expressed as the number of electrons or holes produced per quantum absorbed, has been demonstrated to be unity over the wavelength range from 1.0 microns to the threshold within the experimental uncertainty of about 10 per cent [25].

The photoconductive threshold shifts to shorter wavelengths as the temperature is lowered below room temperature. This occurs because the width of the bandgap changes with temperature. The temperature coefficient of E_G as determined from thermal measurements at high temperatures [18] is given by:

$$E_G(T) = 0.785 - 3.5 \times 10^{-4} T \text{ electrons volts.} \quad (9)$$

Optical absorption measurements [22] give a quadratic dependence of E_G upon temperature at low temperatures.

In the long wavelength portion of the intrinsic response where the excitation involves phonon interactions to converse momentum, it has been pointed out that the excitation may involve either the absorption or emission of a phonon. At higher temperatures the obsorption of a phonon occurs frequently and, of course, the phonon contributes energy to the excitation. This energy amounts to about 0.02 or 0.03 ev. Therefore, the long wavelength threshold at these higher temperatures occurs at photon energies smaller than that corresponding to the minimum gap width by this amount. At low temperatures, phonon absorption is less probable. Therefore, the optical transition usually occurs with the emission of a phonon. Since the phonon energy must come from the photon, the long wavelength photoconductive limit will occur at photon energies slightly greater than that corresponding to the gap width.

Like germanium, the energy band structure of silicon in momentum space (k space) is complex. Fig. 3(b) is a diagram of energy plotted against the momentum vector k for two crystallographic directions in silicon [19]. Again, the minimum activation energy involves a change of momentum. At room temperature, the mini-

[1] H. B. DeVore has made a thorough analysis of this problem to be published.

mum band gap is 1.10 ev [26]. The minimum gap for vertical transitions is not less than about 1.5 ev [20]. As in the case of germanium, a detailed study of the long wavelength tail of the intrinsic absorption yields evidence [27] for the existence of transitions in which phonons are absorbed or are emitted. The temperature dependence of E_G as determined from thermal measurements [26] is given by:

$$E_G(T) = 1.21 - 3.6 \times 10^{-4}T \text{ electron volts.} \quad (10)$$

The photoconductive response of silicon resembles in its broader aspects that of germanium. Fig. 6 shows an early experimentally-measured spectral response curve

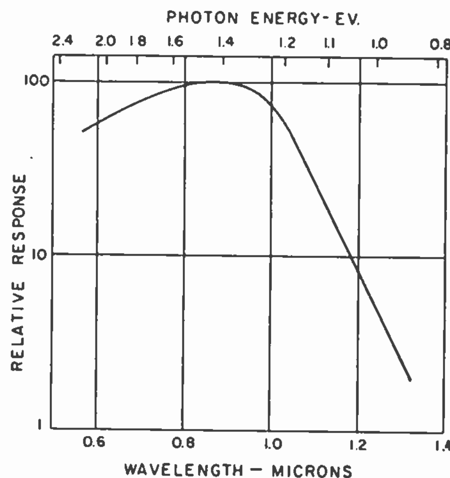


Fig. 6—Intrinsic photoconductivity of silicon.

of polycrystalline silicon [9]. The thermal activation energy, absorption edge, and position of intrinsic photoconductive response are consistent with one another. The photoconductivity of silicon is, in general, lower than that of germanium partly because the hole and electron mobility is lower, and partly because it has not been possible to produce silicon with as low a concentration of recombination centers as it has been for germanium. Therefore, the carrier lifetimes for silicon are smaller.

The Impurity Photoconductivity of Germanium and Silicon

A great many different kinds of impurity atoms can be introduced into silicon and germanium to produce energy levels lying in the forbidden gap. These levels may be acceptor or donor centers depending upon the particular atoms introduced. As such, they can give rise to impurity photoconductivity. Information concerning the location of such levels within the forbidden gap may be obtained experimentally by determination of carrier concentration as a function of temperature, which yields thermal ionization energies, or from meassurements of optical absorption or photoconductivity as a function of wavelength to give optical ionization energies. For those impurities which have been investigated, there is usually good agreement between the thermal and optical ionization energies.

The first impurity atoms to be considered are those in Column V of the Periodic Table. These atoms have one more valence electron than is required for the bonding when the atom is introduced substitutionally. The additional electron is very loosely bound to its parent atom and has a small ionization energy. It is convenient although not strictly correct to consider this electron as being attracted to the impurity center by a coulomb potential. This so-called hydrogenic model leads to the expectation that the ionization energy of the impurity be smaller than that of a free hydrogen atom by the factor $1/K^2 \cdot m^*/m$ where K is the dielectric constant of the crystal and m^* is an effective electron mass. It is possible to choose values of m^* such that approximate agreement is obtained with experimentally-observed values of ionization energies of these impurities. The values thus obtained are 0.01 and 0.04 ev for germanium and silicon respectively. The experimental values for impurity ionization energies are given in Tables I and II (next page). Detailed theoretical treatments [28–30] of impurities in germanium and silicon, which take into account the complicated band structures, have been carried to the point where reasonably good agreement with experiment has been obtained.

Impurities from Column III have one too few electrons to satisfy the bonding when they are present substitutionally in germanium or silicon. Therefore, there will be one hole bound to such a center. The ionization energies of the hole for impurities of this type in germanium are given approximately by the hydrogenic model. This model, however, fails for these impurities in silicon. There is a marked upward trend in the ionization energy from 0.045 ev for boron to 0.16 ev for indium.

The ionization energies are so small for these two classes of impurities either in germanium or in silicon that, at room temperature or even at liquid nitrogen temperature (excepting for indium in silicon), the conductivity is too high to permit impurity photoconductivity. When these materials are cooled to sufficiently low temperatures, that is, in the liquid hydrogen or liquid helium range, thermal excitation of carriers is sufficiently improbable so that the concentration of unionized centers is essentially equal to the total concentration. Under these circumstances the photoconductive response is large. The long wavelength limit for an impurity with an ionization energy of 0.01 ev should be about 120 microns.

Several elements from columns of the Periodic Table other than III and V have been investigated as impurities in germanium and silicon. Reasoning similar to that which leads to the hydrogenic model for Column III and V impurities suggests that for other impurities there will be more than one ionization energy corresponding to successive states of ionization and that multiple ionization might be expected to involve deeper

TABLE I

IMPURITY IONIZATION ENERGIES IN GERMANIUM

Periodic Table Column	Element	Donor or Acceptor	E_c-E_i ev	E_i-E_v ev	Photoconductive Threshold Microns
I	Li	D	0.01 (43)		
	Cu	A		0.040 (44) 0.31 (45)	29 (31, 34)
	Au	A; (D)	0.2 (32, 46)	0.053 (32, 33) 0.15 (32, 46)	>15 (32) 9 (16, 32, 34) 5.5 (16, 32)
II	Zn	A		0.029 (47)	>38 (31)
III	B	A		0.0104 (48)	
	Al	A		0.0102 (48)	
	Ga	A		0.0108 (48)	
	In	A		0.0112 (48)	>38 (31)
	Tl	A		0.014 (49)	
V	P	D	0.0120 (48)		
	As	D	0.0127 (48)		
	Sb	D	0.0097 (48)		
	Bi	D	0.012 (49)		
VII	Mn	A	0.35 (50)	0.16 (50)	
VIII	Fe	A	0.27 (51)	0.34 (51)	4.6 (17) 4.1 (17)
	Co	A	0.31 (52)	0.25 (52)	4.6 (52) 5.5 (52)
	Ni	A	0.30 (53)	0.22 (7, 53, 54)	4.6 (53) 5.6 (53)
	Pt	A	0.2 (47)	0.04 (47)	

TABLE II

IMPURITY IONIZATION ENERGIES IN SILICON

Periodic Table Column	Element	Donor or Acceptor	E_c-E_i ev	E_i-E_v ev	Photoconductive Threshold Microns
I	Li	D	0.033 (55)		
	Au	D	0.30 (58)	{0.33 (56) 0.39 (55)}	3.8 (57)
II	Zn	A		0.092 (58) 0.30 (58)	
III	B	A		0.045 (55)	30 (11)
	Al	A		0.057 (55)	
	Ga	A		0.065 (55)	
	In	A		0.16 (55)	9 (11)
V	P	D	0.039 (55)		
	As	D	0.049 (55)		
	Sb	D	0.039 (55)		

lying levels. These expectations are, in general, realized experimentally. The impurities which have been investigated are listed in Tables I and II along with ionization energies for the levels which have been observed and with photoconductive thresholds in those cases for which they have been determined. The photoconductive response curves for indium, zinc and copper doped germanium measured at liquid helium temperature shown in Fig. 7 [31] are typical of the behavior of low ionization energy impurity photoconductors.

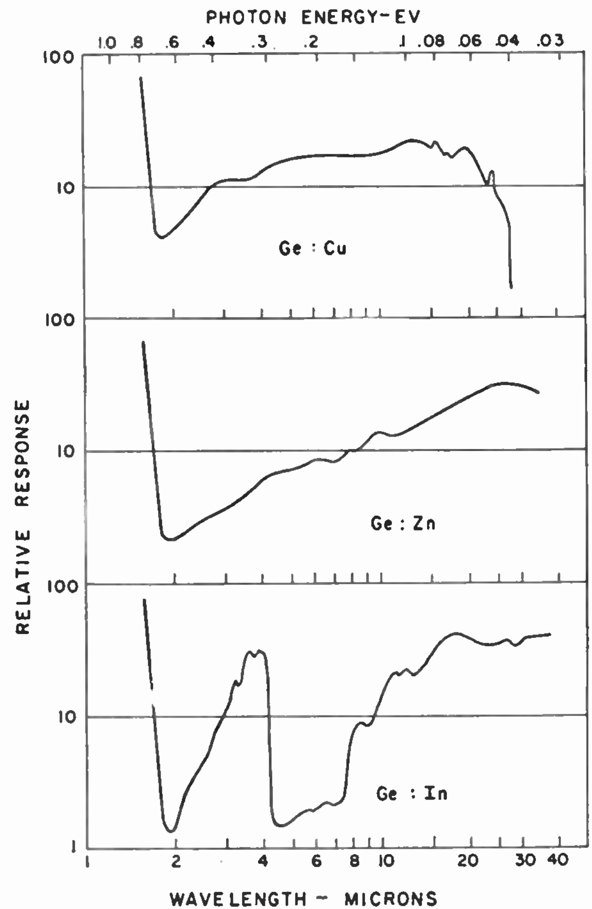


Fig. 7—Impurity photoconductivity of copper, zinc and indium doped germanium at liquid helium temperature.

For those impurities which exhibit multiple level behavior, the location of the shallowest level only can be determined if no compensating impurity is present. In order to determine the ionization energy of a second, deeper-lying level, an impurity of opposite conductivity type must be added in an amount sufficient to remove all of the charge carriers from the shallowest level of the multivalent impurity. A third, still deeper-lying level, if it exists, can be detected by compensating both shallower levels.

Gold, as a substitutional impurity in germanium, is the only example known to date of an impurity which exhibits three energy levels [32, 33]. The shallowest of these levels has an activation energy of about 0.05ev and when pure gold is introduced into germanium, the

material is p type.[2] If an appropriate amount of an n type impurity is introduced into gold doped germanium, the electrons from these donor centers fill the lowest gold levels. The material thereupon exhibits an ionization energy corresponding to the second impurity level, which is located at approximately 0.15 ev above the valence band. With this large an activation energy, the material is only slightly conducting at liquid air temperature and makes a very effective photoconductor. The spectral response of this type of photoconductor is shown in Fig. 8 [16, 32, 34].

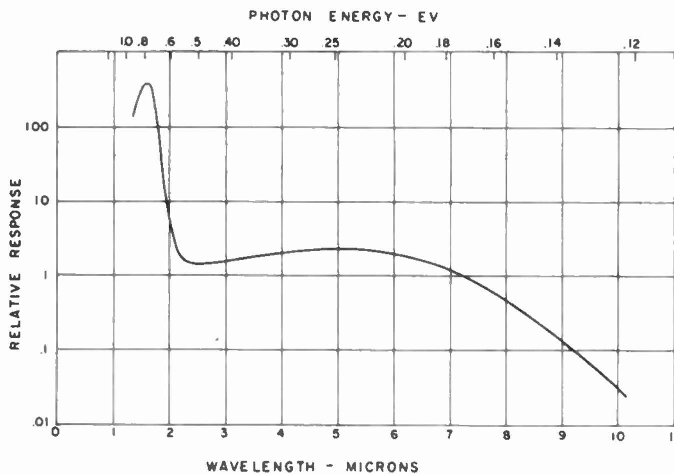


Fig. 8—Photoconductivity of partially compensated p type gold doped germanium at 77°K.

The second gold level can likewise be compensated by the addition of more n type impurity. With the addition of sufficient n-type impurity, the material becomes n type and exhibits a thermal ionization energy of about 0.2 ev. This behavior is due to the deepest-lying acceptor level which is located at approximately 0.5 ev above the valence band. It is interesting to note that gold atoms can be introduced into the material in either of two ways. One is by adding the foreign atom to the molten germanium from which a crystal is to be grown. The second is to grow an intrinsic crystal and then introduce gold by diffusion. Identical results are obtained with samples prepared in either of these ways.

Irrespective of whether the current carriers are due to excitation from impurity levels or excitation across the forbidden gap, their transport depends upon their mobility in the semiconductor. For high purity germanium at room temperature, the mobility for electrons is 3,900 cm/sec/volt/cm while that for holes is about 1,900 cm/sec/volt/cm [35]. The mobility for silicon is somewhat lower being in the neighborhood of 1,200 cm/sec/volt/cm for electrons and 500 cm/sec/volt/cm for holes [36]. The mobility in both these materials is a function of temperature. Fig. 9 is the measured mobility for a series of typical n type germanium samples containing different amounts of impurities [37]. As has been pointed out in an earlier section, the temperature dependence of mobility for germanium and silicon is somewhat different from what would be expected on the basis of the simplest semiconductor model. Experimentally it is found that in germanium the lattice scattering mobility for electrons varies as $T^{-1.66}$ and for holes as $T^{-2.33}$ [18]. In silicon the corresponding quantities vary as $T^{-1.5}$ and $T^{-2.3}$, respectively [36].

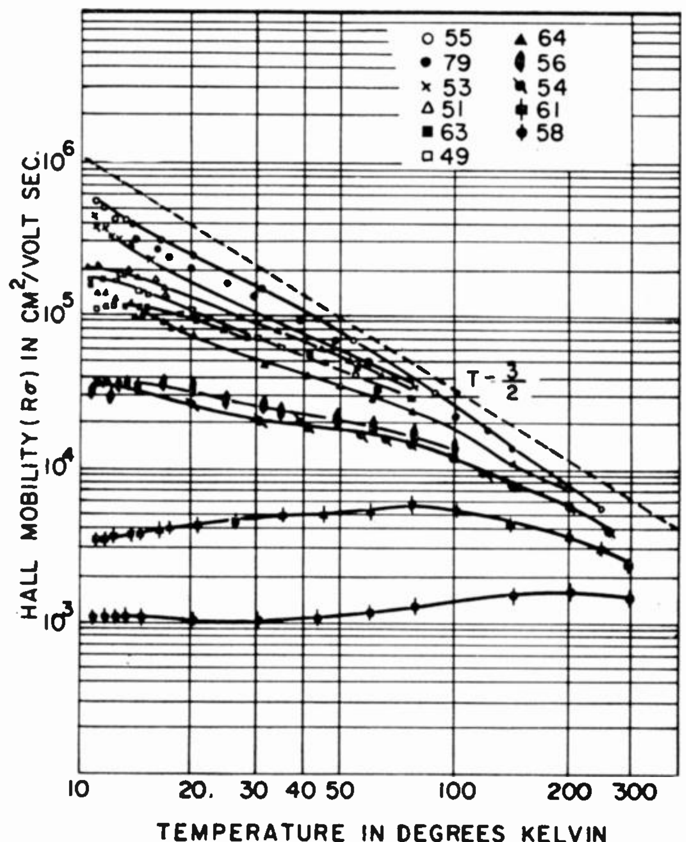


Fig. 9—Hall mobility of a set of arsenic doped germanium samples as a function of absolute temperature.

At low temperatures the impurity scattering introduces a temperature dependence which is in the opposite direction, that is, the mobility decreases with decreasing temperature as is seen from the experimental curves. The theory of scattering by ionized impurities gives a mobility which varies approximatley as $T^{3/2}$ and inversely as the concentration of ionized impurities [38, 39]. Scattering by neutral impurities should be temperature independent except for the dependence of the concentration of unionized impurities upon temperature [40]. Little is known experimentally concerning scattering by lattice imperfections. The scattering by dislocations has been treated theoretically [41, 42]. Such scattering is predicted to be anisotropic and to give a mobility which is proportional to T.

In order to obtain an approximate value for the photoconductive current resulting from an incident photon flux the following form of (8) may be used:

[2] Dunlap, W. C., Jr., "Gold, an Amphoteric Impurity Element in Germanium," *Bull. Am. Phys. Soc.*, vol. 30, p. 12, 1955, suggests that this is a donor level located 0.05 ev above the valance band.

$$i = \frac{eV\mu TF'}{L^2} \quad (11)$$

where F' is the number of photons per second producing excitation, $\mu = \mu_e + \mu_n$ the effective mobility which is the sum of the mobilities of the two carriers involved, L the length of the photoconductive crystal, and T the effective lifetime of the carriers. The effective lifetime of the carriers is the time during which carriers are free to move through the lattice. In other words, it is the recombination time minus the length of time carriers are trapped. Lifetimes in germanium of 1,000 microseconds can be achieved without too much difficulty. For silicon, practical lifetimes are one or two orders of magnitude shorter.

Junction Photoconductive Cells

Intrinsic germanium at room temperature has a specific resistivity of about 50 ohm-cm. The change of resistance due to amounts of light corresponding to those normally used with photoconductive cells is quite small. Expressed in other terms, the change in carrier density is often a very small percentage of the large number of carriers thermally excited in the material. Even with a coupling circuit having optimum impedance match, the sensitivity is low because of the current noise in the thermally excited current. When the material can be operated at temperatures well below room temperature, both intrinsic and suitably doped germanium are extremely practical materials for high sensitivity photoconductive cells.

Where the radiation to be detected by these cells lies in the portion of the spectrum at wavelengths smaller than 1.8 microns, it is possible to greatly increase the impedance of the cell at room temperature by introducing a *p-n* junction. Fig. 10 shows a schematic energy

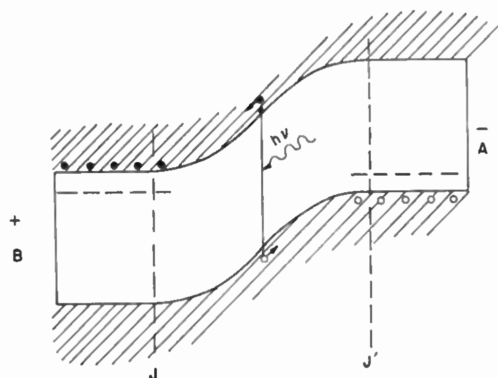


Fig. 10—Schematic energy diagram of a *p-n* junction photoconductive cell.

level diagram of a *p-n* junction photoconductive cell. in this type of photocell, almost the entire potential drop occurs across the *p-n* junction. On the *n* side of the junction electrons are thermally excited from impurity levels into the conduction band, while on the other side of the junction the acceptor levels contribute holes to the valence band. The applied potential tends to draw holes away from the barrier toward electrode A and electrons away from the barrier toward electrode B. If a photoconductive excitation occurs, for example, in the region marked JJ′ within a diffusion length of the barrier, the hole moves toward electrode A while the electron is drawn through the potential barrier and over to electrode B. This, of course, results in a photoconductive current. The junction cell is finding many practical applications.

Bibliography

[1] Bardeen, J., and Shockley, W., "Deformation Potentials and Mobilities in Non-Polar Crystals." *Physical Review*, Vol. 80 (1950), pp. 72–80.

[2] van Roosbroeck, W., and Shockley, W., "Photon-Radiative Recombination of Electrons and Holes in Germanium." *Physical Review*, Vol. 94 (1954), pp. 1558–60.

[3] Haynes, J. R., and Briggs, H. B., "Radiation Produced in Germanium and Silicon by Electron-Hole Recombination." *Physical Review*, Vol. 86 (1952), p. 647.

[4] Haynes, J. R., "New Radiation Resulting from Recombination of Holes and Electrons in Germanium." *Physical Review*, Vol. 98 (1955), pp. 1866–7.

[5] Hall, R. N., "Germanium Rectifier Characteristics." *Physical Review*, Vol. 83 (1951), p. 228. Also, "Electron-Hole Recombination in Germanium," *Physical Review*, Vol. 87 (1952), p. 387.

[6] Shockley, W., and Read, W. T., Jr., "Statistics of the Recombination of Holes and Electrons." *Physical Review*, Vol. 87 (1952), pp. 835–42.

[7] Burton, J. A., Hull, G. W., Morin, F. J., and Severiens, J. C., "Effect of Nickel and Copper Impurities on the Recombination of Holes and Electrons in Germanium." *Journal of Physical Chemistry*, Vol. 57 (1953), pp. 853–9.

[8] Aigrain, P., "Light Emission from Injecting Contacts on Germanium in the 2μ to 6μ Band." *Physica*, Vol. 20 (1954), pp. 1010–13.

[9] Moss, T. S., *Photoconductivity in the Elements*. Academic Press, New York, N. Y., 1952.

[10] Rose, A., "Recombination Processes in Insulators and Semiconductors." *Physical Review*, Vol. 97 (1955), pp. 322–33.

[11] Burstein, E., Picus, G., and Sclar, N., "Optical and Photoconductive Properties of Silicon and Germanium." Presented at the Conference on Photoconductivity, Atlantic City, N. J., 1954.

[12] Haynes, J. R., and Hornbeck, J. A., "Decay in Photoconductivity Associated with Hole Traps in N-type Silicon." *Physical Review*, Vol. 94 (1954), p. 1438.

[13] Hornbeck, J. A., and Haynes, J. R., "Trapping of Minority Carriers in Silicon. I. P-type Silicon." *Physical Review*, Vol. 97 (1955), pp. 311–21.

[14] Gebbie, H. A., Nisenoff, M., and Fan, H. Y., "Photoconductivity and Carrier Traps in Germanium." *Physical Review*, Vol. 91 (1953), p. 230.

[15] Hannay, N. B., Haynes, J. R., and Shulman, R. G., "The Interaction of Traps and Heat Treatments in Silicon." *Physical Review*, Vol. 96 (1954), p. 833.

[16] Newman, R., "Photoconductivity in Gold-Germanium Alloys." *Physical Review*, Vol. 94 (1954), pp. 278–85.

[17] Newman, R., and Tyler, W. W., "Properties of Germanium Doped With Iron. II. Photoconductivity." *Physical Review*, Vol. 96 (1954), pp. 882–6.

[18] Morin, F. J., and Maita, J. P., "Conductivity and Hall Effect in the Intrinsic Range of Germanium." *Physical Review*, Vol. 94 (1954), pp. 1525–29.

[19] Herman, F., "The Electron Energy Band Structure of Silicon and Germanium." Proceedings of the IRE, p. 1703, this issue. Also, "Speculations on the Energy Band Structure of Ge-Si Alloys." *Physical Review*, Vol. 95 (1954), pp. 847–8.

[20] Dash, W. C., Newman, R., and Taft, E. A., "Optical Absorption in Single Crystals of Ge and Si." *Bulletin of the American Physical Society*, Vol. 30, No. 1 (1955), p. 53.

[21] Fan, H. Y., Shepherd, M. L., and Spitzer, W., "Infrared Absorption and Energy Band Structure of Germanium and Silicon." Presented at the Conference on Photoconductivity, Atlantic City, N. J.; (1954).

[22] Macfarlane, G. G., and Roberts, V., "Infrared Absorption of Germanium Near the Lattice Edge." *Physical Review*, Vol. 97 (1955), pp. 1714–16.

[23] Hall, L. H., Bardeen, J., and Blatt, F. J., "Infrared Absorption Spectrum of Germanium." *Physical Review*, Vol. 95 (1954), pp. 559–60.

[24] Bardeen, J., Blatt, F. J., and Hall, L. H., "Indirect Transitions From the Valence to the Conduction Bands." Presented at the Conference on Photoconductivity, Atlantic City, N. J.; (1954).
[25] Goucher, F. S., "The Photon Yield of Electron-Hole Pairs in Germanium." *Physical Review*, Vol. 78 (1950), p. 816.
[26] Morin, F. J., and Maita, J. P., "Electrical Properties of Silicon Containing Arsenic and Boron." *Physical Review*, Vol. 96 (1954), pp. 28–35.
[27] Macfarlane, G. G., and Roberts, V., "Infrared Absorption of Silicon Near the Lattice Edge." *Physical Review*, Vol. 98 (1955), p. 1865.
[28] Kittel, C., and Mitchell, A. H., "Theory of Donor and Acceptor States in Silicon and Germanium." *Physical Review*, Vol. 96 (1954), pp. 1488–93.
[29] Kohn, W., and Luttinger, J. M., "Theory of Donor States in Silicon." *Physical Review*, Vol. 98 (1955), pp. 915–22.
[30] Lampert, M. A., "Ground State of Impurity Atoms in Semiconductors Having Anisotropic Energy Surfaces." *Physical Review*, Vol. 97 (1955), pp. 352–3.
[31] Burstein, E., Davisson, J. W., Bell, E. E., Turner, W. J., and Lipson, H. G., "Infrared Photoconductivity Due to Neutral Impurities in Germanium." *Physical Review*, Vol. 93 (1954), pp. 65–8.
[32] Morton, G. A., Hahn, E. E., and Schultz, M. L., "Impurity Photoconduction in Germanium." Presented at the Conference on Photoconductivity, Atlantic City, N. J.; (1954).
[33] Dunlap, W. C., Jr., "Gold, an Amphoteric Impurity Element in Germanium." *Bulletin of the American Physical Society*, Vol. 30, No. 2, (1955), p. 12.
[34] Kaiser, W., and Fan, H. Y., "Infrared Absorption, Photoconductivity, and Impurity States in Germanium." *Physical Review*, Vol. 93 (1954), pp. 977–80.
[35] Prince, M. B., "Drift Mobilities in Semiconductors. I. Germanium." *Physical Review*, Vol. 92 (1953), pp. 681–7.
[36] Prince, M. B., "Drift Mobilities in Semiconductors. II. Silicon." *Physical Review*, Vol. 93 (1954), pp. 1204–6.
[37] Debye, P. P., and Conwell, E. M., "Electrical Properties of *n*-type Germanium." *Physical Review*, Vol. 93 (1954), pp. 693–706.
[38] Conwell, E. M., and Weiskopf, V. F., "Theory of Impurity Scattering in Semiconductors." *Physical Review*, Vol. 77 (1950), pp. 388–90.
[39] Brooks, H., "Scattering by Ionized Impurities in Semiconductors." *Physical Review*, Vol. 83 (1951), p. 879.
[40] Erginsoy, C., "Neutral Impurity Scattering in Semiconductors." *Physical Review*, Vol. 79 (1950), pp. 1013–14.
[41] Dexter, D. L., and Seitz, F., "Effects of Dislocations on Mobilities in Semiconductors." *Physical Review*, Vol. 86 (1952), pp. 964–5.
[42] Read, W. T., Jr., "Scattering of Electrons by Charged Dislocations in Semiconductors." *Philosophical Magazine*, Vol. 46 (1955), pp. 111–31.
[43] Fuller, C. S., and Ditzenberger, J. A., "Diffusion of Lithium into Germanium and Silicon." *Physical Review*, Vol. 91 (1953), p. 193.
[44] Morin, F. J., and Maita, J. P., "Comparison of Copper-doped Germanium with Heat-treated Germanium." *Physical Review*, Vol. 90 (1953), p. 337.
[45] Battey, J. F., and Baum, R. M., "Energy of the High-lying Acceptor Level in Copper-doped Germanium." *Physical Review*, Vol. 94 (1954), p. 1393.
[46] Dunlap, W. C., Jr., "Electrical Properties of Gold-Germanium Alloys." *Physical Review*, Vol. 91 (1953), p. 1282.
[47] Dunlap, W. C., Jr., "Properties of Zinc-, Copper-, and Platinum-doped Germanium." *Physical Review*, Vol. 96 (1954), pp. 40–45.
[48] Geballe, T. H., and Morin, F. J., "Ionization Energies of Groups III and V Elements in Germanium." *Physical Review*, Vol. 95 (1954), pp. 1085–6.
[49] Morton, G. A., and Schultz, M. L., unpublished results.
[50] Woodbury, H. H., and Tyler, W. W., "Some Properties of Germanium Doped With Manganese." *Bulletin of the American Physical Society*, Vol. 30, No. 2 (1955), pp. 11–12.
[51] Tyler, W. W., and Woodbury, H. H., "Properties of Germanium Doped With Iron. I. Electrical Conductivity." *Physical Review*, Vol. 96 (1954), pp. 874–82.
[52] Tyler, W. W., Newman, R., and Woodbury, H. H., "Properties of Germanium Doped With Cobalt." *Physical Review*, Vol. 97 (1955), pp. 669–72.
[53] Tyler, W. W., Newman, R., and Woodbury, H. H., "Properties of Germanium Doped With Nickel." *Physical Review*, Vol. 98 (1955), pp. 461–6.
[54] van der Maesen, F., and Brenkman, J. A., "The Solid Solubility and the Diffusion of Nickel in Germanium." *Philips Research Reports*, Vol. 9 (1954), pp. 225–30.
[55] Morin, F. J., Maita, J. P., Shulman, R. G., and Hannay, N. B., "Impurity Levels in Silicon." *Physical Review*, Vol. 96 (1954), p. 833.
[56] Taft, E. A., and Horn, F. H., "Gold as a Donor in Silicon." *Physical Review*, Vol. 93 (1954), p. 64.
[57] Newman, R., "Photoconductivity in Gold Doped Silicon." *Physical Review*, Vol. 94 (1954), pp. 1530–1.
[58] Morin, F. J., Maita, J. P., Shulman, R. G., and Hannay, N. B., "Impurity Levels in Silicon." Presented at American Physical Society meeting, Minneapolis, Minn.; 1954.

Photoeffects in Intermetallic Compounds*

H. P. R. FREDERIKSE† AND R. F. BLUNT†

Summary—**The intermetallic semiconductors are classified with respect to their crystal structure and to the place of the component elements in the periodic system. A survey is given of the properties of compounds with the zincblende and fluorite lattice. Photoeffects of individual members of these two groups are discussed. Such phenomena include photoconduction, photovoltage, and the photo-electro-magnetic effect.**

THE SYSTEMATIC investigation of intermetallic compounds is a very recent development in a field of physics—semiconductor research—which in itself is only 15 or 20 years old. Although many binary compounds have been studied in the past, it is only since 1952 that these materials have been approached in the light of the new knowledge of semiconductors obtained from the work on germanium and silicon. The name "intermetallic" indicates that these compounds consist of two or more elements with a metallic character. For reasons of classification, reference will also be made to compounds containing sulphur, selenium, and tellurium, in spite of the fact that these elements can hardly be considered as metals.

* Original manuscript received by the IRE, August 5, 1955.
† National Bureau of Standards, Washington, D. C.

The intermetallic semiconductors can be categorized according to their crystal structure:

1. Compounds with the zincblende lattice. These binaries consist of elements from columns at opposite sides of the periodic table: III–V-, II–VI-, and I–VII-compounds. Examples are: AlSb, InP, CdS, HgTe, AgI, and CuBr.

2. Compounds with the calcium fluorite structure. To this class belong the II–IV-compounds like Mg_2Si and Mg_2Sn, and possibly the Ca-series: Ca_2Si-Ca_2Pb. The crystal structure of the latter group is, however, not certain.

3. Compounds possessing the sodium chloride lattice. These materials are combinations of elements from the 4th and 6th or from the 2nd and 6th columns. like PbS, SnTe, MgSe, and CaTe.

4. A few other compounds with complex structures like CbSb, ZnSb, and Mg_3Sb_2.

5. Ternary compounds with the diamond structure, known as chalcopyrites.[1] Examples are $CuFeS_2$ and $AgTlTe_2$.

A few words may be said about the binding character in intermetallic compounds. The proportions of the constituents obey the ordinary valence rules with only very few exceptions; the total number of valence electrons in the compound is nearly always 8 or a multiple of 8. The binding forces are predominantly covalent and increase from ionic to homopolar towards the center and the bottom of the periodic table; i.e., the heteropolar bond character increases in the following order from Ge via GaAs and ZnSe to CuBr and in the sequence MgTe, MgSe, MgS, MgO from the telluride to the oxide.

For a discussion of photoeffects in intermetallics we have chosen the compounds of the first two classes: compounds with the zincblende and those with the fluorite structure. Photoconductivity has been observed in a few other intermetallics; some of these are listed in Table I.

TABLE I

Compound	Energy gap from photoresponse at 300°K
Mg_3Sb_2[2,3]	0.8 ev
ZnSb[2]	0.55 ev
Cd_3As_2[2]	0.6 ev

All the photoeffects which will be discussed in this paper are observed in single crystals or slightly polycrystalline bulk material. It is well-known (e.g., from the very extensive work on the lead compounds) that the photosensitivity of thin films is sometimes several orders of magnitude higher.

The III–V-Compounds

This is the group of intermetallic semiconductors which has been most intensively studied in recent years, probably because of their close resemblance to the elements of the fourth column. The similarity shows up most strikingly in the lattice constants of these materials: Sn (6.46 Å), InSb (6.45 Å); Ge (5.65 Å), GaAS (5.64 Å). Many other properties also show a remarkable analogy to those of the corresponding elements of column IV.

Investigations of these materials include determination of conductivity and Hall coefficient,[4–8] rectification,[4] magneto-resistive effect,[9,10] optical absorption,[7,11–14] reflectivity,[12] photoconductivity,[15,16] thermoelectric power,[17] thermal conductivity,[17,18] and cyclotron resonance.[19]

Some of the results from optical absorption measurements[11] are shown in Fig. 1 (next page).

From all these data it has been possible to derive many important constants, such as, the forbidden energy gap, the Hall mobilities, and the effective masses

[1] C. H. L. Goodman and R. W. Douglas, "New semiconducting compounds of diamond type structure, *Physica*, vol. 20, pp. 1107–1110; November, 1954.

[2] V. P. Zhuze, I. V. Mochan and S. M. Ryvkin, "Photoconductivity of some intermetallic compounds," *Jour. Tech. Phys.* (U.S.S.R.), vol. 18, pp. 1494–1497; 1948.

[3] T. S. Moss, "Photoconductivity in magnesium antimonide layers," *Proc. Phys. Soc. B* (London), vol. 63, pp. 982–989; December, 1950.

[4] H. Welker, "Über neue halbleitende verbindungen, I u II," *Z. Naturf.*, vol. 7a, pp. 744–749, November, 1952; vol. 8a, pp. 248–251, April, 1953; *Physica*, vol. 20, pp. 893–909; November, 1954.

[5] H. Weiss, "Über die elektrischen eigenschaften von InSb," *Z. Naturf.*, vol. 8a, pp. 463–469; August, 1953.

[6] M. Tannenbaum and I. P. Maita, "Hall effect and conductivity in InSb single crystals," *Phys. Rev.*, vol. 91, pp. 1009–1010; August 15, 1953.

[7] R. G. Breckenridge, R. F. Blunt, W. R. Hosler, H. P. R. Frederikse, J. H. Becker, and W. Oshinsky, "Electrical and optical properties of intermetallic compounds. I. Indium Antimonide," *Phys. Rev.*, vol. 96, pp. 571–575; November 1, 1954.

[8] R. K. Willardson, A. C. Beer, and A. E. Middleton, "Electrical properties of semiconducting InSb," *Jour. Electrochem. Soc.*, vol. 101, pp. 354–358; July, 1954.

[9] G. L. Pearson and M. Tanenbaum, "The magnetoresistive effect in InSb," *Phys. Rev.*, vol. 90, p. 153; April 1, 1953.

[10] T. C. Harman, R. K. Willardson and A. C. Beer, "Analysis of magnetoresistance and Hall coefficient in *p*-type indium antimonide and *p*-type germanium," *Phys. Rev.*, vol. 95, pp. 699–702; August 1, 1954.

[11] F. Oswald and R. Schade, "Über die bestimmung der optischen Konstanten von Halbleitern des typus $A^{III}B^{V}$ im infraroten," *Z. Naturf.*, vol. 9a, pp. 611–617; July/August, 1954.

[12] H. B. Briggs, R. F. Cummings, H. J. Hrostowski, and M. Tanenbaum, "Optical properties of some group III-group V compounds," *Phys. Rev.*, vol. 93, p. 912; February 15, 1954.

[13] F. A. Cunnell, J. T. Edmonds, and J. L. Richards, "Measurements on some semiconducting compounds with the zinc-blende structure," *Proc. Phys. Soc. B* (London), vol. 67, pp. 848–849; November 1, 1954.

[14] M. Tanenbaum and H. B. Briggs, "Optical properties of indium antimonide," *Phys. Rev.*, vol. 91, pp. 1561–1562; September 15, 1953.

[15] H. P. R. Frederikse and R. F. Blunt, "Photoconductivity and photovoltaic effect in intermetallic compounds," to be published in "Photoconductivity" (John Wiley and Sons, Inc., New York; 1955).

[16] G. R. Mitchell, A. E. Goldberg, and S. W. Kurnick, "InSb photovoltaic cell," *Phys. Rev.*, vol. 97, pp. 239–240; January 1, 1955.

[17] H. P. R. Frederikse and E. V. Mielczarek, "Thermoelectric power of indium antimonide," *Phys. Rev.* (In press).

[18] G. Busch and M. Schneider, "Heat conduction in semiconductors," *Physica*, vol. 20, pp. 1084–1086; November, 1954.

[19] G. Dresselhaus, A. F. Kip, C. Kittel, and G. Wagoner, "Cyclotron and spin resonance in indium antimonide," *Phys. Rev.*, vol. 98, pp. 556–557; April 15, 1955.

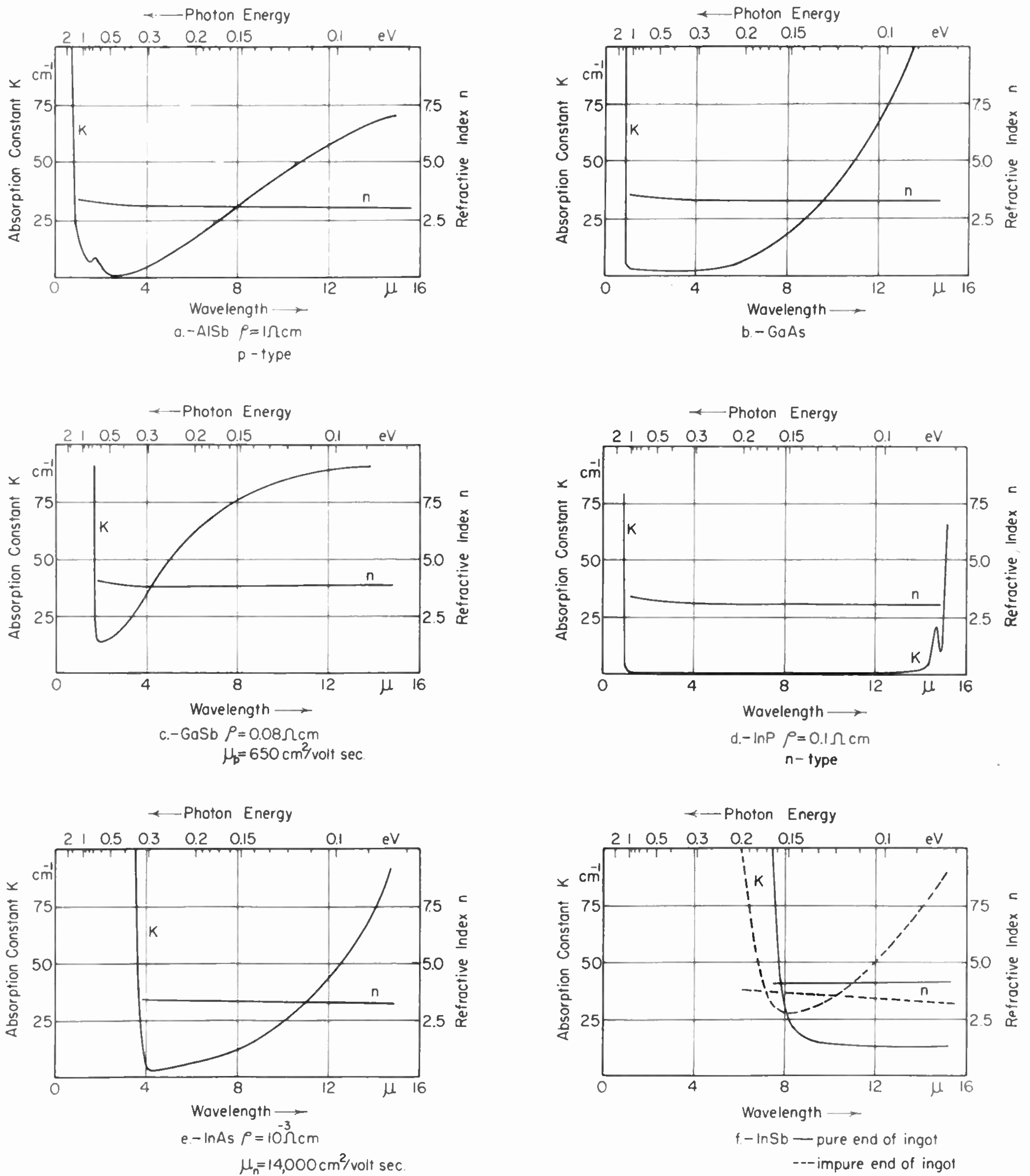


Fig. 1—Optical constants of several intermetallic semiconductors with the zincblende structure.

of the charge carriers. Values for these parameters have been compiled in Table II (next page).

Indium Antimonide

Preparation. The compound may be prepared by melting stoichiometric proportions of the components together in a graphite crucible under a protective atmosphere. The melting point is 523°C. Purification can be achieved by means of the zone-melting technique. The p-type and n-type impurities collect at opposite ends of the ingot due to the fact that the segregation coefficient for acceptor impurities is 1.2, while it appears

TABLE II
PROPERTIES OF III–V INTERMETALLIC COMPOUNDS

	Energy gap at 300°K (*ev*)	Hall mobility (cm^2/volt-sec) at 300°K		Effective masses		Rectification at 300°K	Transistor action at 300°K
		μ_e	μ_p	m_e^*	m_p^*		
InSb	0.16	65,000	700	0.015	0.18		
InAs	0.35	25,000	200	0.03	0.3		
InP	1.25	3,400	650			+	+
GaSb	0.67	4,000	650	0.20	0.39	+	
GaAs	1.35	3,500		0.03		+	
GaP	2.4						
AlSb	1.6		200			+	
AlAs	2.4?						
AlP	—						

to be about 0.5 for donor impurities.[20]

Optical Absorption. One of the most remarkable effects in InSb is the anomalous optical absorption[11,14] shown in Fig. 1(f). The absorption edge of *n*-type samples appears to move towards shorter wavelengths with increasing impurity content. This phenomenon has been explained by Burstein[21] on the basis of the very small effective mass of electrons. The small effective mass implies a small density of states at the bottom of the conduction band and also an unusually low impurity concentration at which degeneracy sets in. Consequently electrons, thermally excited from impurity levels, will rapidly fill up the lower part of the conduction band and any optical excitation will have to take place to a level considerably above the bottom of this band.

Photoconductivity. Measurements of photoresponse are usually made in the following manner.[15,22] Samples are ground to a thickness between 0.2 and 0.6 mm and either mechanically polished or electropolished. Wires are soldered to the ends with indium and the specimen is mounted on a copper backing, separated from the metal by a thin sheet of mica. Appropriate optical cells[23] are used when measurements above or below room temperature are undertaken. Monochromatic radiation from an infrared spectrometer is focused on the specimen. The light beam is chopped at various frequencies (up to 1,000 or 2,000 cps) producing an alternating change in electrical conductivity. This signal is then measured with a tuned amplifier across a load resistor in series with the specimen and a battery. For low impedance samples a step-up transformer is inserted before the amplifier.

[20] T. C. Harman, R. K. Willardson, H. L. Goering, and A. C. Beer, "Effect of zone refinining variables on the segregation of impurities in InSb," Abstract No. 97, Spring Meeting Electrochem. Soc., Cincinnati, Ohio; May, 1955.

[21] E. Burstein, "Anomalous optical absorption limit in InSb," *Phys. Rev.*, vol. 93, pp. 632–633; February 1, 1954.

[22] T. S. Moss, "Photoconductivity in the Elements," Butterworths Scientific Publications, London, p. 81; 1952.

[23] e.g. See W. H. Duehrig and I. L. Mador, "An optical cell for use with liquid helium," *Rev. Sci. Instr.*, vol. 23, pp. 421–424; August, 1952.

Results of measurements on a high-purity *n*-type sample of InSb are given in Fig. 2.[15] Hall effect and resistivity data show that this sample has an impurity content of 3.7×10^{13} cm^{-3} and an electron mobility of 37,000 cm^2/volt-second. It is probable that this low mobility is due to partial compensation by *p*-type impurities. The photoconductive signal is barely measureable at room temperature but increases considerably on cooling to liquid nitrogen and liquid helium temperatures. No generally accepted theory exists to determine the threshold of photoconductivity. Moss[22] identifies the width of the energy gap with that wavelength for which the response has fallen to one-half of its maximum value, while others report good agreement with the wavelength of maximum response itself.[24] The values for the gap at helium temperature are 0.23 *ev* on the basis of Moss' rule and 0.24 *ev* according to the second criterion. The difference is very small due to the steepness of the photoconductive cut-off and both values are in good agreement with results from electrical and absorption experiments.

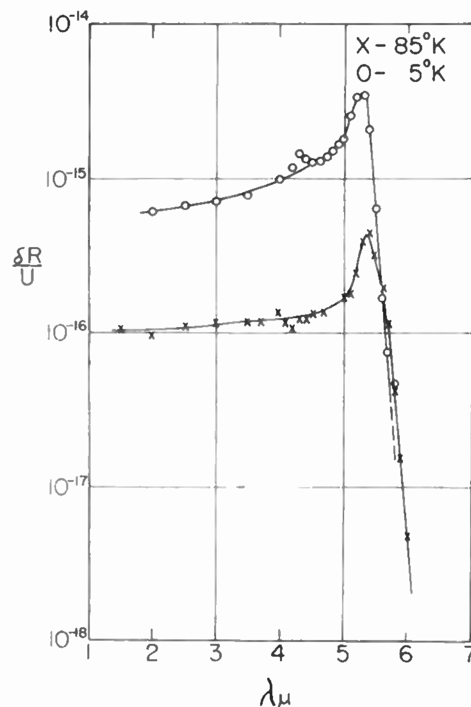


Fig. 2—Photoconductive response of *n*-type InSb. (After Frederikse and Blunt, *Photoconductivity*, John Wiley and Sons, Inc., New York; 1955).

Photovoltaic Effect. Sharp *p-n* junctions can be produced by adding a slight amount of acceptor impurity (e.g., zinc) to an *n*-type melt during the crystal-pulling process. Such a junction shows good rectification characteristics when cooled to liquid nitrogen temperature.[25]

[24] J. J. Loferski, "Infrared optical properties of single crystals of tellurium," *Phys. Rev.*, vol. 93, pp. 707–716; February 15, 1954.

[25] S. W. Kurnick, A. E. Goldberg, G. M. Mitchell and R. N. Zitter, "Photoeffects in InSb," Abstract No. 98, Spring Meeting Electrochem. Soc., Cincinnati, Ohio; May, 1955.

The spectral distribution of the photovoltage at 77°K is plotted in Fig. 3. Such photovoltaic cells have sensitivities of the order of 10 volts/watt/cm^2.

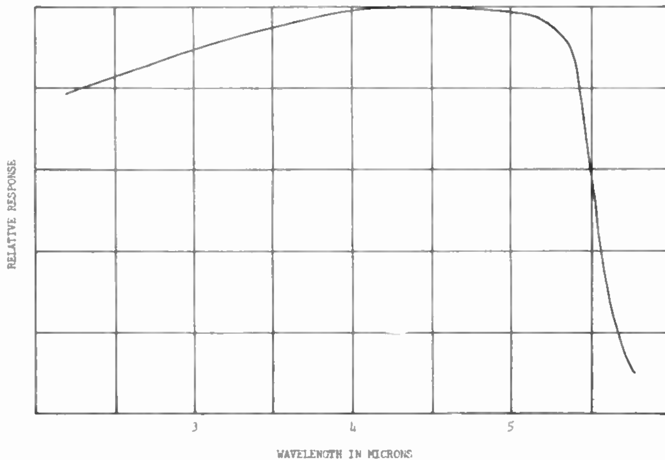


Fig. 3—Spectral distribution of the photovoltaic effect in InSb at 77°K.

Photoelectromagnetic (PEM) Effect.[25] If a rectangular sample is illuminated (by I quanta/sec/cm^2) on one of the surfaces, electron-hole pairs will be created in the surface layer and then diffuse into the bulk of the specimen. If the sample is now placed in a transverse magnetic field, holes and electrons will be deflected in opposite directions through the Hall angle $\tan^{-1}(\mu B)$, where μ is the mobility and B the magnetic induction (see Fig. 4). This produces a voltage across the sample perpendicular to the magnetic field and the direction of illumination. The short-circuit current i_s is then given by the following expression[25,26] in which the contribution of the holes is neglected due to the relatively small mobility of these carriers:

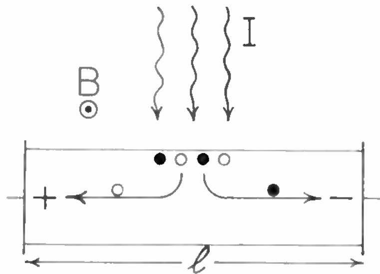


Fig. 4—Photo-electro-magnetic (PEM) effect.

$$i_s = \frac{eI\mu BL_D}{\sqrt{1+\mu^2B^2}} \frac{1}{1+\frac{\tau s}{L_D}\sqrt{1+\mu^2B^2}}; \qquad (1)$$

in which L_D = diffusion length = $(D\tau)^{1/2}$, D = diffusion constant, τ = lifetime of a charge carrier, and s = surface recombination velocity. It is clear from this formula that the plot of i_s vs B will show saturation for small s, while i_s will reach a maximum and then decrease with increasing B for large recombination velocities.

Simultaneous measurements of the PEM-effect and the photoconductive response on the same sample make a direct determination of τ possible. The photoconductive short circuit current is given by

$$i_{pc} = \frac{eI\mu F\tau}{1+(\tau s/L_D)}, \qquad (2)$$

where F is the electric field. It follows from (1) and (2) that the lifetime is then

$$\tau = D\left(\frac{i_{pc}}{F}\right)^2\left(\frac{B}{i_s}\right)_o^2. \qquad (3)$$

The index o indicates that the value of B/i_s has been taken for sufficiently small magnetic fields. Experiments like those described here have been carried out on a p-type sample of InSb which was electro-polished and contained 7×10^{14} impurities/cc. Results indicate lifetimes of the order of 10^{-7} seconds at room temperature and 10^{-6} seconds at 77°K. These values have been confirmed by direct measurements with a xenon spark source.[25] Moss[27] believes that development of better crystals will eventually make it possible to increase the lifetime to 10^{-6} seconds at room temperature. On this basis he predicts a limiting energy sensitivity of 10^{-10} watts at 300°K in the 7 μ spectral range for a 1 cycle per second bandpass.

Indium Arsenide

This compound is usually prepared in a sealed "vycor" tube because of the strong evaporation of the arsenic at high temperatures. The melting point of InAs is 936°C.[28] Zone melting improves the purity, and impurity concentrations as small as 10^{15} cm^{-3} have been reached. Electrical and optical measurements indicate that the properties of InAs are very similar to those of InSb.[29] Results of resistivity and Hall effect experiments show that the width of the energy gap E is 0.47 ev at absolute zero, while the optical absorption data yield a value of about 0.35 ev at 300°K[11,12] The electron mobility appears to be very high and the effective mass of electrons is smaller than 0.1 m_e. The optical absorption edge shifts as a function of the impurity content similar to the behavior of InSb.

Photovoltaic Effect. This effect has been observed in p-n junctions of InAs.[28] Fig. 5 (next page) presents the photovoltage of three cells at liquid nitrogen temperature. The time constant τ of these cells has been investigated with a light-pulse technique; the value for τ appears to be 7 μsec at 78°K.

[26] T. S. Moss, "The photo-electro-magnetic effect in germanium and lead sulphide," *Physica*, vol. 20, pp. 989–993; November, 1954.

[27] T. S. Moss, "Absorption and photoconductivity in Indium antimonide," to be published in *Photoconductivity* (John Wiley and Sons, Inc., New York, 1955).

[28] R. M. Talley and D. P. Enright, "Photovoltaic effect in InAs," *Phys. Rev.*, vol. 95, pp. 1092–1093; August 15, 1954.

[29] O. G. Folberth, O. Madelung, and H. Weiss, "Die elektrischen eigenschaften von indium arsenid. II.," *Z. Naturf.*, vol. 9a, pp. 954–958; November, 1954.

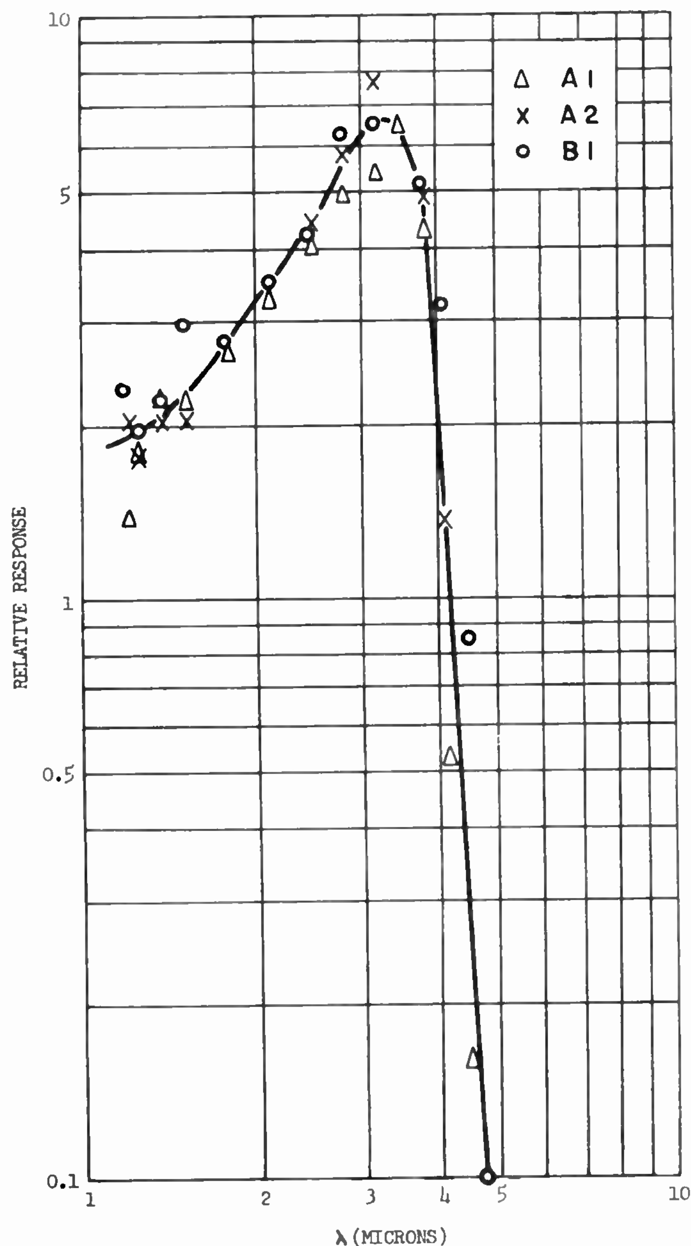


Fig. 5—Spectral response (normalized at 3.5 μ) for InAs photovoltaic cells at 78°K.

Gallium Antimonide

The properties of this compound are quite similar to those of germanium. Measurements of conductivity and Hall effect[30–32] as well as transmission measurements[11,12,31] have been carried out on this material. Results are given in Table II.

Photoconductivity. The photoresponse has been measured on one p-type sample.[15] The charge carrier concentration of this specimen is 10^{17} cm^{-3} at room temperature and about one-tenth as much at 85°K. Fig. 6 shows the spectral distribution at those two temperatures. The signal is relatively small, probably due to the high impurity content of the specimen. The smallness of the increase on cooling is surprising. No dependence on frequency was observed between 10 and 510 cps, indicating a lifetime smaller than 10^{-3} seconds. The energy gap derived from the response curves (on the basis of Moss' criterion[22]) are in good agreement with results from electrical and transmission experiments.

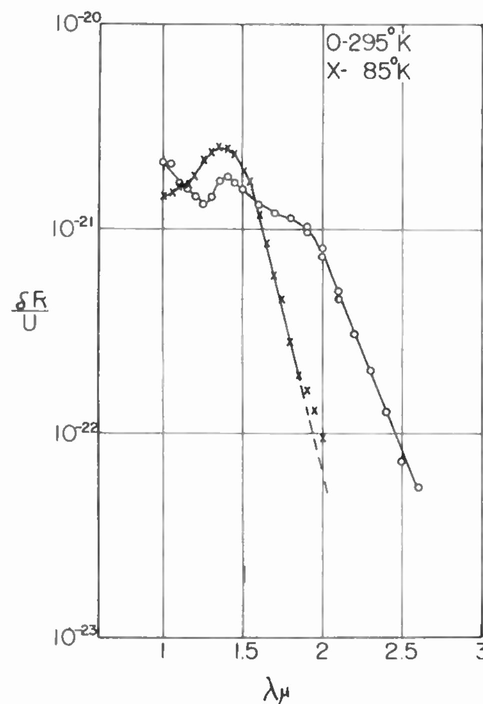


Fig. 6—Photoconductivity in GaSb as a function of wavelength. (After Frederikse and Blunt, *Photoconductivity*, John Wiley and Sons, Inc., New York, 1955).

Gallium Arsenide

This compound has an energy gap slightly larger than that of silicon. Until now only n-type material has been produced. Several properties of the compound have been measured such as conductivity, Hall coefficient, thermoelectric power, absorption, and reflection coefficient, as well as photoconductivity and rectifying characteristics.[33]

Photoeffects. Barrie *et al.*[33] report a high photosensitivity for polycrystalline samples; changes up to 5 per cent under room light illumination are observed. This response appears to be mainly associated with grain boundaries. Spectral response at room temperature of a GaAs specimen is in Fig. 7 (next page); the radiation is chopped at 800 cps in this case. The maximum of this curve corresponds to 1.30 ev. If the light beam

[30] H. N. Leifer and W. C. Dunlap, Jr., "Some properties of p-type gallium antimonide between 15°K and 925°K," *Phys. Rev.*, vol. 95, pp. 51–56; July 1, 1954.

[31] R. F. Blunt, W. R. Hosler, and H. P. R. Frederikse, "Electrical and optical properties of intermetallic compounds. II. Gallium antimonide," *Phys. Rev.*, vol. 96, pp. 576–577; November 1, 1954.

[32] D. P. Detweiler, "Electrical properties of gallium antimonide," *Phys. Rev.*, vol. 97, pp. 1575–1578; March 15, 1955.

[33] R. Barrie, F. A. Cunnell, J. T. Edmond, and I. M. Ross, "Some properties of gallium arsenide," *Physica*, vol. 20, pp. 1087–1090; November, 1954.

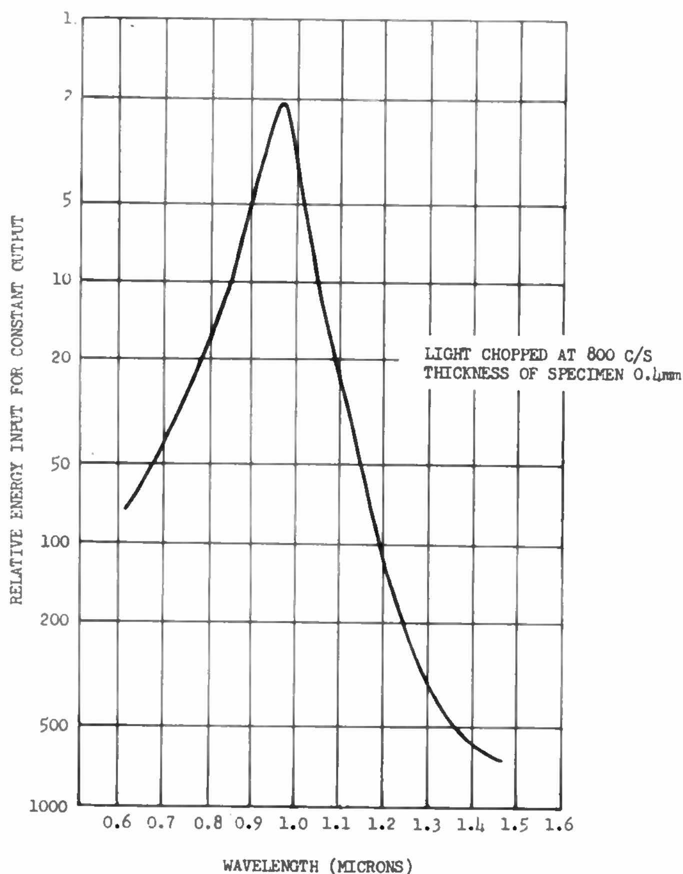


Fig. 7—Photoconduction in GaAs.

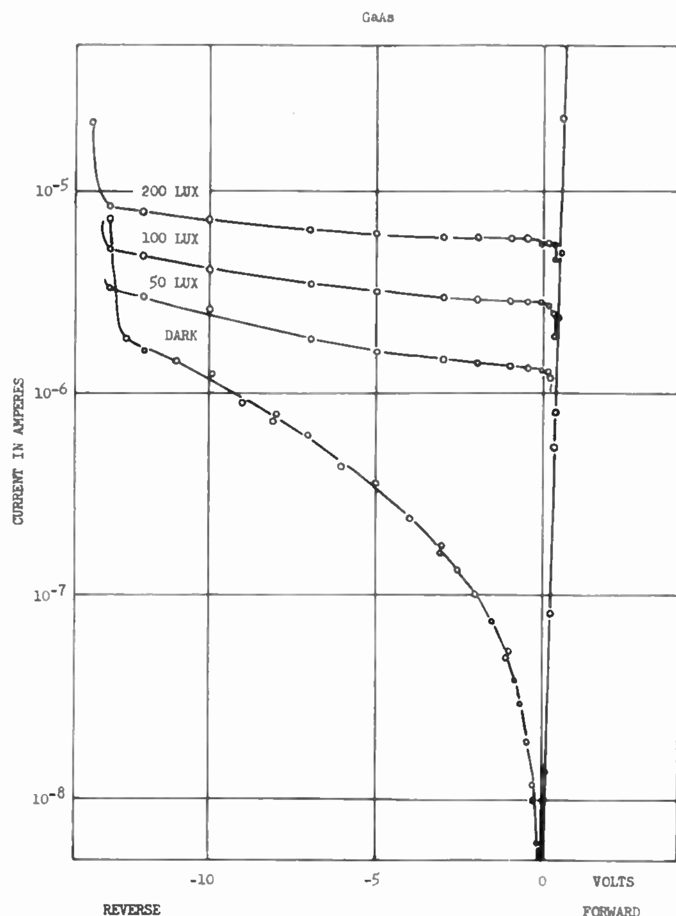


Fig. 8—Current-voltage characteristics of a GaAs-photodiode for different illuminations.

is chopped at a rate of 100 cps, the signal increases by a factor of 10, indicating a lifetime of about 0.01 second.

The compound appears to be a good rectifier. A photodiode illuminated with different light intensities shows current-voltage characteristics[4] as in Fig. 8.

The II–IV-Compounds

The best known members of this group are the compounds of Mg with the elements of column IV: Mg_2Si, Mg_2Ge, Mg_2Sn and Mg_2Pb. Electrical properties of these materials have been measured by several investigators.[15,34–36] Results are given in Table III.

TABLE III
Properties of II–IV-Compounds

	Energy gap (ev)	Hall Mobility (cm²/volt-sec) at 300°K μ_e	μ_p
Mg_2Si	0.77 (at 0°K)		
Mg_2Ge	$0.74-9\times10^{-4}\,T$	530	106
Mg_2Sn	$0.33-3.5\times10^{-4}$	320	260

Results on the compound Mg_2Pb indicate that this material is either metallic or a highly degenerate semiconductor.

Busch and Winkler[37] have also prepared mixed crystals, $Mg_2Ge_ySn_{1-y}$, where y can have all values between 0 and 1. These mixed crystals show a continuous change of energy gap between those of Mg_2Sn and Mg_2Ge.

Magnesium Stannide

Preparation.[38] This compound is prepared in the same way as InSb. The crystals are bluish and brittle and cleave very easily; the surface corrodes rapidly when in contact with water or moist air. Deviations from perfect stoichiometry are very small in spite of the strong evaporation of the magnesium during preparation of the compound. Most crystals appear to be n-type; small amounts of column-I metals act as acceptor impurities.

Resistivity and Hall Effect, Optical Absorption. Some results from measurements of these quantities are given in Table III.[38] The value for the forbidden energy gap

[34] W. D. Robertson and H. H. Uhlig, "Electrical properties of the intermetallic compounds Mg_2Sn and Mg_2Pb," *Trans. A.I.M.M.E.*, vol. 180, pp. 345–355; 1949.

[35] B. I. Boltaks, "The nature of the electric properties and magnetic permeability of the intermetallic compound Mg_2Sn," *Jour. Tech. Phys.* (U.S.S.R.), vol. 20, pp. 180–186; 1950 (in Russian).

[36] G. Busch and U. Winkler, "Elektrische Eigenschaften der intermetallischen verbindungen Mg_2Si, Mg_2Ge, Mg_2Sn und Mg_2Pb," *Physica*, 20, pp. 1067–1072; November, 1954.

[37] G. Busch and U. Winkler, "Elektrische leitfähigkeit von mischkristallen intermetallischer verbindungen," *Helv. Phys. Acta*, vol. 26, pp. 579–583, fasc. 6; 1953.

[38] R. F. Blunt, H. P. R. Frederikse, and W. R. Hosler, "Electrical and optical properties of intermetallic compounds. IV. Magnesium stannide," to be published (*Physical Review*).

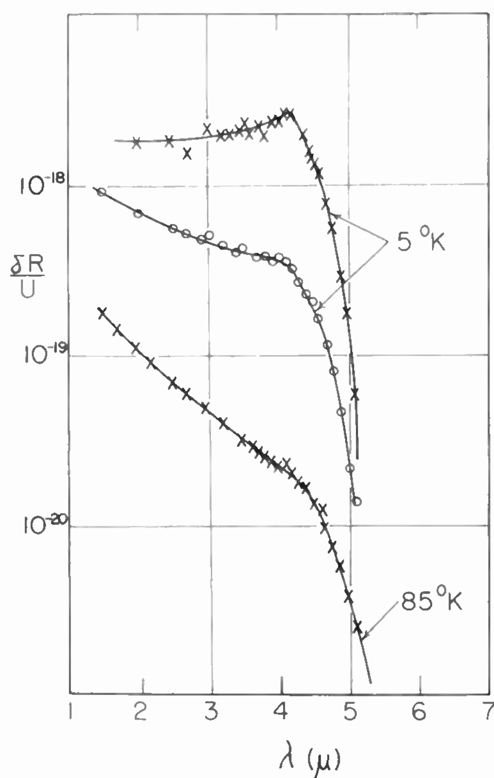


Fig. 9—Photoconductive response of Mg_2Sn as a function of wavelength. × = n-type, 0 = p-type.

E_0 is subject to doubt. Most of the values which have been reported in the literature range from 0.26 to 0.36 ev;[34,36,38] one interpretation of transmission measurements, however, based on the possibility of indirect transitions, yields a value for E_0 as low as 0.12 ev.[39]

Photoconductivity. Photoconduction has been observed in both n- and p-type samples;[38] the spectral response at 85°K and at 5°K is shown in Fig. 9. No signal could be detected at room temperature.

Assuming that the quantum efficiency is equal to one, it is possible to calculate the lifetime τ of the charge carriers.[40] The result of this calculation for an n-type sample at 85°K is $\tau = 2.5 \times 10^{-4}$ seconds (for 4 μ radiation).

Correlation of the photoconductive threshold with the width of the energy gap presents many difficulties. Moss' rule[22] does not seem applicable at all. Identifying the gap width with either the maximum or the "break" in the response curve, one obtains values for the energy gap which are 0.02 ev or 0.03 ev smaller than results from electrical or optical transmission experiments would indicate. This discrepancy might be connected with the very gradual decrease of absorption constant at the foot of the edge, which is far less steep than for most other crystals. As yet no unequivocal explanation has been offered for these observations.

Conclusions

The intense investigation of intermetallic compounds in recent years has led to the discovery of several new infrared-sensitive photoconductors. Simultaneous studies of other properties of the same materials, such as conductivity, Hall effect, optical transmission, etc., have greatly contributed to our understanding of the electronic structure of these solids. Results of experiments on spectral response, PEM-effect and other photoeffects in pure samples of InSb, InAs and Mg_2Sn classify them as potential detectors for the far infrared region.

Acknowledgment

Figures in this paper are reproduced with permission of The Electrochemical Society and the Editors of *Zeitschrift für Naturforschung*, *Physica*, and *The Physical Review*.

[39] G. G. Macfarlane, private communication (compare: G. G. Macfarlane and V. Roberts, "Infrared absorption of germanium near the lattice edge," *Phys. Rev.*, vol. 97, pp. 1714–1716; March 15, 1955.

[40] A. Rose, "Performance of photoconductors," to be published in *Photoconductivity* (John Wiley and Sons, Inc., New York, 1955).

Photoconductivity of the Sulfide, Selenide, and Telluride of Zinc or Cadmium*

R. H. BUBE†

Summary—**The highlights of recent photoconductivity research on evaporated, powder, and sintered layers, and on single crystals of photoconductors in the group ZnS, CdS, ZnSe, CdSe, ZnTe, and CdTe are reviewed. The following aspects of photoconductivity are discussed: spectral response, impurity sensitization, electrode contact problems, conductivity, mobility, speed of response, photocurrent vs light intensity, photocurrent vs temperature, infrared quenching, thermally stimulated current and trapping, space-charge limited current, photoconductivity and luminescence, photovoltaic effect, photoemissive effect, and surface photoconductivity.**

The utility of the concepts of the Fermi-level and demarcation-level in the description of many photoconductivity phenomena is briefly discussed.

INTRODUCTION

THE USE OF photoconductors for the performance of various tasks in research and industry is undergoing a broad expansion at the present time [1–6]. Photoconductors from the group discussed in this paper are currently in use, or being considered and developed for use, in such widely diverse applications as photocells for measurement of light intensity, detectors for X-rays, alpha- and beta-particles, and gamma rays, automatic automobile headlight dimmers, streetlight control, oximeters, "noiseless" switches, smoke and fire control, solar generators, detectors in computers, and miscellaneous toys and novelties. Fig. 1 shows several large CdS crystals and typical crystal and sintered-layer commercial cells.

Photoconductivity in these materials is characterized by the fact that the "gain" may be much larger than unity. The term "gain" is used to mean the ratio between the number of charges passed through the crystal per unit time and the number of photons absorbed per unit time. Excitation produces a free electron and a free hole; in most of these materials, electrons are the majority carriers and the holes are rapidly trapped. The free electron moves through the crystal under the influence of the field to the positive electrode, and hence out of the crystal. At the time that an electron moves out of the crystal at the positive electrode, another electron enters the crystal at the negative electrode. Charge continues to flow through the crystal until recombination occurs between the trapped hole and a free electron. Values of the "gain" as high as 10^4 have been observed in CdS. Note that if the negative electrode were of such a type that electrons were not free to enter the crystal from it, then the "gain" would be limited to a value of unity or less. The gain may be simply expressed in terms of the lifetime of a free electron, τ, and the transit time of an electron between electrodes, τ:

$$\text{Gain} = \tau/t. \quad (1)$$

When t is expressed in terms of the mobility, μ, the applied voltage, V, and the distance between electrodes, L, (1) for the gain may be converted to:

$$\text{Gain} = (\tau\mu V)/L^2. \quad (2)$$

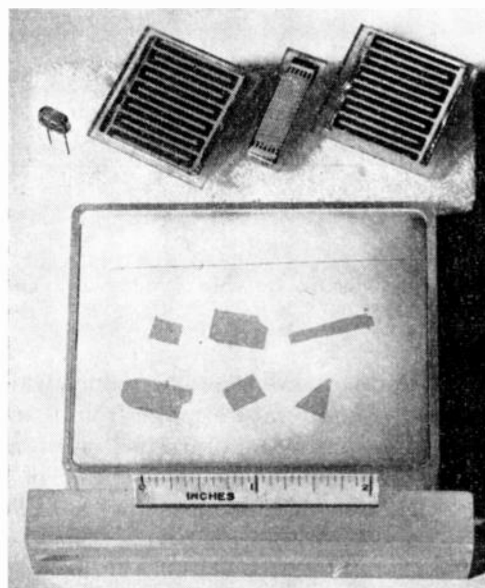


Fig. 1—Photograph of large plate-like CdS crystals and of commercial crystal and sintered-layer photocells.

Research in the photoconductivity of zinc sulfide holds an historic position in the field, dating back at least to the work of Gudden and Pohl [7] in 1920. It is only in the last decade or two, however, that appreciable advances have been made in the understanding of the photoconductivity processes involving quantum gains larger than unity in such materials. The development of the techniques for the growth of single crystals, given impetus by the publication of Frerichs [8] in 1946 on the vapor-phase growth of cadmium sulfide crystals, has played an important role in obtaining this understanding. In this paper we shall review the highlights of research in the past ten years on photoconductivity in the sulfide, selenide, and telluride of zinc or cadmium: six materials with many similar properties. We have not attempted to provide an exhaustive bibliography of the subject, but have sought instead to mention key publications which will lead the reader into the heart of the recent photoconductivity literature.

* Original manuscript received by the IRE, August 5, 1955.
† RCA Laboratories, Princeton, N. J.

The topics to be discussed have been divided into two main categories. The first concerns the preparation of photoconductors by the different methods which have been used: as evaporated, powder, or sintered layers, and as single crystals. The second concerns a number of photoconductivity phenomena, a comprehension of which is essential for an understanding of the nature of the photon and electron processes in these photoconductors.

Preparation of Photoconductors

Evaporated Layers

The use of evaporation in vacuum for the preparation of photoconductor layers has received wide use, principally because of the relative ease with which uniform layers of fairly large area can be produced without the presence of graininess. Evaporation methods are not difficult to develop using the normal vacuum techniques, and usually may be sufficiently controlled to yield reproducible results. The major disadvantage of evaporated layers is that the crystallinity of the layer is much less perfect than that of a single crystal; effects associated with crystal imperfections (such as trapping) may determine the performance of the layer. It is also difficult with evaporated layers to make a systematic study of impurity effects under controlled conditions.

Evaporated layers of CdS have been prepared by many investigators [9–13]. It is generally reported that choice of an optimum pressure and rate of evaporation enables the preparation of CdS evaporated layers with properties which approach those of single crystals. In particular, the optical properties of the evaporated layer are very similar to those of the single crystal, whereas the electrical properties are much more sensitive to the exact method of preparation and to the type of post-preparation treatment. The optical properties of thin evaporated layers of CdS (1,200–6,000A thick) are described by Gottesman [14].

Schwarz [15, 16] describes several variations in the evaporation procedure which have proven useful for the preparation of layers of CdS, CdSe, and CdTe; such variations include cathode sputtering and the use of a low-voltage arc to achieve evaporation.

Görlich and Heyne [17] have prepared evaporated layers of CdSe and CdTe, the latter being formed by vaporizing Cd and Te together in vacuum. Braithwaite [18] obtained layers of ZnTe by evaporating the material formed by fusing Zn and Te.

Powder Layers

In the years before the techniques for the growth of single crystals had become well developed, many measurements of photoconductivity were made on powder samples. Powders have been useful for the gross detection of the existence of photoconductivity [19], the spectral response of photoconductivity [20] and observations of such photoconductivity phenomena as growth and decay of photocurrent and infrared quenching [21]; such measurements, however, cannot be generally used to establish quantitatively reliable data on conductivities of materials because of the effect of the particle-to-particle contacts on the relationship between current and voltage.

Kuwabara [22] has measured the optical properties of very small particles of CdS in glass. He found that the absorption edge lies at shorter wavelengths than that for single crystals, and shifts to longer wavelength with heat treatment of the glass.

The photosensitivity which could be obtained from such powders has been, until recently, generally much lower than that which could be obtained from properly-prepared single crystals. Kolomiets [23] reported an increase in the photosensitivity of a CdS powder when the powder was heated in an atmosphere of oxygen; it was believed that the oxygen converted some of the CdS to CdO at the surface of the powder particles, thus increasing the conductivity and improving the particle-to-particle contact.

Thomsen and Bube [24] have found it possible to prepare CdS powder photoconductors which exhibit many of the desirable properties of single crystals. Powder layers of large area may be prepared by spreading onto a suitable surface a prepared mixture of microcrystalline photoconductor powder (prepared by a special phosphor technique, involving the incorporation of chlorine and copper impurity), in a plastic solution, and allowing the layer to harden. A two-inch square cell of such a powder layer can pass an ampere for an illumination of a few foot-candles. Photocells made from powder layers are described by Nicoll and Kazan. [25]

Sintered Layers

A variation of the powder layer which approaches even closer to having the same characteristics as single crystals, is the sintered layer. [24] Sintered layers of CdS or CdSe are made by spraying a "paint" of CdS or CdSe (a water mixture of sulfide or selenide with chloride and copper) onto a suitable surface, and firing the surface with its dried layer to form a polycrystalline sintered layer. This method is very useful in producing large-area photoconductors with crystal-like characteristics without involving the difficulties met in growing large-area single crystals or in producing large-area photoconductor layers by evaporation in vacuum.

Single Crystals

There are three essentially different methods for the preparation of single crystals of photoconductors: vapor phase reaction of the elements, sublimation of the powder and recrystallization, and growth from the melt.

The report by Frerichs [8, 26] of the growth of single crystals of CdS, CdSe, and CdTe by the vapor phase reaction of Cd and H_2S, H_2Se, and H_2Te provided considerable impetus to the development of this technique for the growth of single crystals. Many other investiga-

tors have elaborated on this basic method to grow crystals of CdS [27–30] and crystals of ZnS. [31–33]

The sublimation of ZnS powder with subsequent recrystallization from the vapor phase has been used for the growth of single crystals of ZnS by Kremheller [33] and Piper [34]. Sublimation of ZnS and CdS powder under a low pressure of a few psi of H_2S for periods of the order of days has been used by Reynolds and Czyzak [35–37] to grow crystals of CdS and ZnS as large as $13 \times 8 \times 5$ mm³.

CdTe has the lowest melting point of all the photoconductors under consideration in this paper (1041°C) and therefore is suitable for growing crystals from the melt as well as from the vapor phase. [38, 39]

Photoconductivity Phenomena

Phenomenological Photoconductivity Theory

An absolute theory of photoconductivity does not exist; although it is possible to predict the approximate spectral response of a material from a knowledge of its absorption characteristics, it is not possible to predict the "gain," usually called the photosensitivity, from a knowledge of the constituent elements only. The sensitivity, the temperature dependence, the speed of response, and many other characteristics of the photoconductivity of a material depend on the capture cross sections for electrons and holes of imperfections in the crystal. Recombination of electrons and holes takes place predominantly through these imperfections; hence the location, the concentration, and the nature of these imperfections determine most of the photoconductivity properties of the crystal. Such imperfections may be associated with crystal defects or with incorporated impurities; in CdS and CdSe it seems that crystal defects play a major role in determining the photoconductivity properties and in determining what effects are caused by incorporated impurities. (A model may be constructed to explain photoconductivity with quantum gain greater than unity in terms of distributed barriers in the crystal which are reduced by irradiation by light. [49, 50] Although such barriers may be present and give rise to such effects, they are not necessary for the occurrence of photoconductivity with gain greater than unity.)

To prepare a sensitive photoconductor, therefore, it is necessary to make certain that centers with large recombination cross section are absent, and that whatever recombination centers are present have as small a cross section as possible. For CdS and CdSe crystals, and for other n-type photoconductors in the group under consideration, the free hole formed by excitation across the band-gap is captured quickly at an imperfection; the "sensitizing" centers must therefore be of the type that have a small capture cross section for free electrons, once they have captured a hole. Values for this capture cross section for a free electron have been derived from measurements of photoconductivity for ZnS and CdS by Smith [40] Bube [20, 41], Fassbender [42, 43], Schoen, [47], and Broser and Warminsky [44], the values lie between 10^{-20} and 10^{-24} cm². Such cross sections are some five to nine orders of magnitude smaller than atomic dimensions, and indicate that the high photosensitivity of these materials is associated with the presence of centers which are effectively surrounded by a potential barrier for the capture of an electron.

In attempting the theoretical description of photoconductivity, there have been essentially two modes of approach. The first method of approach consists in setting up a fairly simple model: simple in the sense that only a small number of different types of discrete levels are assumed present in the forbidden gap. Then a system of differential equations is set up for the rates of change of the concentration of the various filled and empty states. The solution of these equations, usually in severely complicated form, is applied to the experimental data, a fit being attempted by the adjustment of suitable parameters. Examples of such calculations are given by Broser and Warminsky [44, 45] Frerichs [46], Schoen, [47] and Kallmann and Kramer [48]. In many cases such calculations are useful, but in general they are sufficiently far from reflecting the actual complexity of affairs in the photoconductor so that they provide only a partial and approximate analysis of the phenomena.

The second method of approach consists in attempting to treat the photoconductivity properties of a material in a collective and somewhat statistical way to obtain a semiquantitative description of phenomena and to retain a direct concept of the physical nature of the processes involved. This method has been largely developed by Rose [51–53] and is characterized by the use of the quasi or steady-state Fermi level, and by the general assumption that there is usually a fairly high concentration of levels with a broad energy distribution in the forbidden gap. It will prove convenient in the discussion of the photoconductivity phenomena in the following sections to use the method of Rose; at this point we shall briefly present its more important features.

The distance of the steady-state electron Fermi level from the conduction band, E_{fn}, (see Fig. 2, next page) may be calculated from the conductivity and the temperature according to the equation:

$$E_{fn} = kT \ln (N_c e\mu/\sigma) \tag{3}$$

where T is the absolute temperature, N_c is the concentration of states in the lowest kT-wide part of the conduction band, e is the electronic charge, μ is the mobility, and σ the measured conductivity. A similar hole Fermi level can be defined in terms of the concentration of free holes. In describing electronic phenomena, it is convenient to consider these Fermi levels as indicating the approximate boundary between shallow trapping levels, lying above the electron Fermi level or below the hole Fermi level, and recombination levels, lying between the two Fermi levels. If a level does not lie between the two Fermi levels, it is assumed that it is in thermal

equilibrium with the nearest allowed band, and does not play an appreciable role in recombination processes.

Actually the boundary between trapping levels and recombination levels is not given exactly by the location of the Fermi level, but by the location of a level called the demarcation level. An electron at the electron demarcation level has equal probability of being thermally excited into the conduction band and of recombining with a hole in the filled band. A hole at the hole demarcation level has equal probability of being thermally excited to the filled band and of recombining with an electron in the conduction band. The demarcation level is shifted from the Fermi level by an energy-difference which is usually small: kT times the natural logarithm of the ratio of the concentration of filled recombination levels, n_g, to empty recombination levels, p_g.

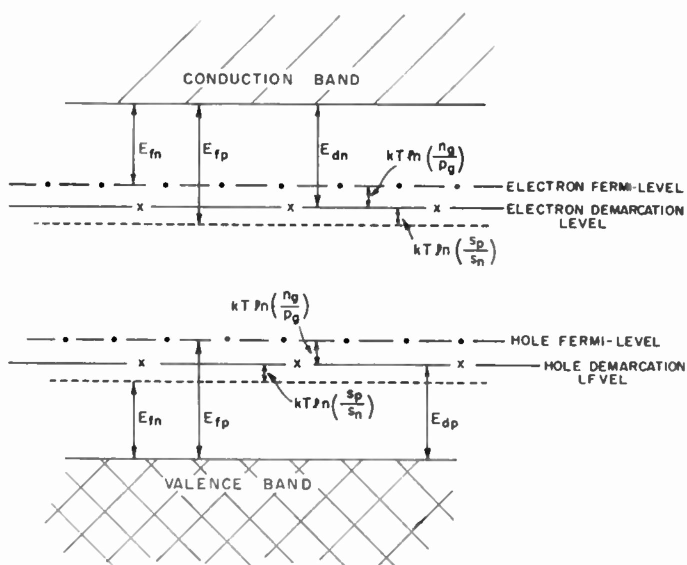


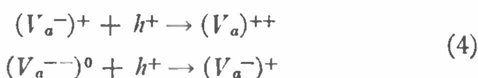
Fig. 2—Schematic representation of the electron and hole Fermi levels and demarcation levels.

A useful relationship exists between the electron Fermi level and the hole demarcation level, and between the hole Fermi level and the electron demarcation level. The hole demarcation level will lie at a distance above the filled band which is equal to the distance of the electron Fermi level below the conduction band plus a correction equal to kT times the natural logarithm of the ratio of the capture cross section of the levels for holes, s_p, to that for electrons, s_n. A similar relation connects electron demarcation level and hole Fermi level. Naturally a separate demarcation level must be defined for each class of states characterized by a different ratio of capture cross sections for electrons and holes.

Rose [53] has pointed out that usually a number of models can be constructed to explain a given photoconductivity observation and that the purpose of further experimentation is to eliminate some of the possible models. There are, however, three photoconductivity phenomena, which will be discussed in more detail in the following sections, which require the general assumption that there are two different classes of recombination centers for their explanation. These three phenomena are an increase in photosensitivity caused by the incorporation of impurities, in spite of the concurrent increase in the concentration of recombination centers; superlinear photoconductivity (photo-current increases as a power of light intensity greater than unity); and infrared quenching of photo-conductivity. These three phenomena are related in that they may be considered respectively as sensitization by impurity, sensitization by light, and desensitization by light.

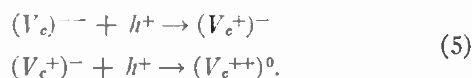
The essence of Rose's hypothesis is that there are two classes of recombination centers distributed through the forbidden gap. One possible example is that (1) Class I centers have a capture cross section for holes equal to or less than that of Class II centers; (2) Class I centers have a capture cross section for electrons greater than that of Class II centers, once a hole has been captured; (3) Class II centers are filled with electrons in the dark. Crystal defects that have the capture cross section properties stipulated can easily be conceived, purely on electrostatic grounds:

Class I centers:



$$(V_a^-)^+ + h^+ \rightarrow (V_a)^{++}$$
$$(V_a^{--})^0 + h^+ \rightarrow (V_a^-)^+ \quad (4)$$

Class II centers:



$$(V_c)^{--} + h^+ \rightarrow (V_c^+)^-$$
$$(V_c^+)^- + h^+ \rightarrow (V_c^{++})^0. \quad (5)$$

V_c represents a cation vacancy and V_a an anion vacancy. The sign inside the bracket represents the number of trapped electrons or holes; the sign outside the bracket represents the effective charge of the defect with respect to the rest of the crystal.

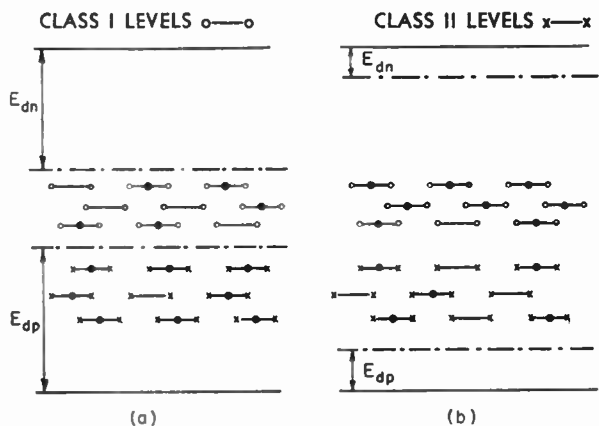


Fig. 3—Schematic representation of sensitization of a crystal with increasing excitation intensity. (a) Low excitation "desensitized." (b) High excitation "sensitized."

We shall see in more detail in the following sections how these concepts can be applied to describe photoconductivity phenomena in CdS and CdSe crystals. Fig. 3 illustrates the dependence of photosensitivity on

TABLE I

WIDTH OF THE BAND GAP, *ev*

(At room temperature except where specified)

Material	Absorption Spectrum	Photoconductivity Excitation Spectrum	Luminescence Excitation Spectrum	Luminescence Edge Emission	Conductivity vs Temperature	Bibliography
cub-ZnS	3.64					56
				3.60		57
hex-ZnS	3.64		3.70			20
	3.70					35, 58
	3.70					56
	3.55					59
	3.65	3.68			3.77	60
	3.60					61
cub-ZnSe		2.58				63
				2.66		57
cub-ZnTe	2.15					63
hex-CdS		2.41				62
	2.38	2.38	2.52 (77°K)			64
				2.45		57
	2.42					59
		2.47				65
	2.44					66
		2.44				67
		2.42–2.48				68
hex-CdSe		1.74				62
cub-CdTe		1.41				62
	1.42					38
		1.47				39
					1.43–1.57	69

the location of the demarcation levels in schematic fashion. The crystal is in a desensitized state when only Class I centers are able to function as recombination centers; the crystal becomes sensitized as Class II centers become able to function as recombination centers because of a shift in the demarcation level. Holes formed by excitation become concentrated in small capture cross section, Class II centers, and electrons formerly in Class II centers are transferred via the conduction and valence bands to fill and hence make unavailable for recombination the large capture cross section Class I centers. (In the remainder of the paper, the term "capture cross section" will mean the capture cross section for free electrons by a center which has previously captured a hole.)

Band Gap and Spectral Response

For a pure photoconductor crystal, the dependence of photosensitivity on wavelength will be very similar to the dependence of the absorption on wavelength. The photosensitivity will be high and fairly constant for irradiation with photons with energy greater than the forbidden gap (for exceptions caused by surface effects, see section on surface photoconductivity), and the photosensitivity will decrease very rapidly as the energy of the photons irradiating the crystal is made less than the forbidden gap.

The most common methods for determining the width of the band gap involve measurements of absorption, excitation spectrum for photoconductivity, excitation spectrum for luminescence, luminescence edge emission spectrum, and dark conductivity as a function of temperature. Small but definite differences exist between the values of band gap determined by different methods; this is largely because the theoretical concepts of the band edges in relation to these various measurements have not been clearly established.

Table I presents a summary of the results of measurements of band-gaps by many investigators. Table II presents a summary of the results of measurements of the temperature variation of the band-gap.

Hoehler [54] has measured the variation in the band-gap of CdS as a function of pressure between 1 and 330 atmospheres; he reports a change in the absorption edge of about −0.01A/atmosphere. From a comparison between the pressure-caused and the temperature-caused shift in the absorption edge, Hoehler concludes that the shift in the edge caused simply by thermal dilatation of the crystal accounts for only about 20 per cent of the total shift in the edge found with changing temperature.

Moss [55] has discovered that for many photoconductors, the ratio of the fourth power of the index of refraction to the wavelength corresponding to the absorption edge is approximately a constant; this constant is about 77 if the wavelength is expressed in microns. If, for example, a value of 2.45 is taken for the index of refraction of CdS, and a value of 0.51 micron for the edge wavelength, the ratio is 71.

TABLE II
TEMPERATURE DEPENDENCE OF THE BAND GAP (ev/degree)×10^{-5}
(Band Gap Decreases with Increasing Temp.)

Material	Absorption Spectrum	Temperature Range	Photoconductivity Excitation Spectrum	Temperature Range	Bibliography
hex-ZnS	46	At 77°K			61
	85	At 800°K			61
cub-ZnSe			72	90°–400°K	62
hex-CdS	41	4°–77°K			64
	49	77°–300°K			64
	50	290°–980°K			66
			52	90°–400°K	62
			46–58	300°–425°K	67
			26–37	At 115°K	68
			42–58	At 265°K	68
hex-CdSe			46	90°–400°K	62
cub-CdTe			36	90°–400°K	62

Impurity Sensitization

The incorporation of impurities in these photoconductors may increase the dark conductivity and the photosensitivity, decrease the dark conductivity and photosensitivity, or provide a new photosensitivity for photons with energy less than the width of the bandgap, corresponding to direct excitation from the impurity centers.

In ZnS, for example, the incorporation of silver impurity extends the spectral response to longer wavelengths by providing an absorption at about 3.42 ev, and copper extends the response even further by providing an absorption at about 3.30 ev [20].

In CdS, the incorporation of a halide or a trivalent cation increases both the dark conductivity and the photosensitivity without appreciably affecting the spectral response [24, 28, 29, 70]. The incorporation of silver or copper decreases the dark conductivity and the photosensitivity, and also adds considerable new photosensitivity in the red portion of the spectrum. [10, 11, 24, 28] The effect of these impurities on CdSe is very similar to that for CdS; in this case, the new spectral response associated with the incorporation of copper is in the near infrared.

The mechanism by which the incorporation of a halide, chloride for example (or of a trivalent cation), increases the photosensitivity of CdS merits further consideration [28]. When chloride ions are incorporated in CdS, they may have two effects: they may alter the effective valence of the ions of the host crystals, or they may cause crystal defects. The relative importance of these two effects will be determined by the atmosphere present during the growth of the crystals [29]. When a chloride ion substitutes for a sulphur ion in the crystal in accordance with the first of these effects, the results may be expressed by saying either for each chloride incorporated, a monovalent cadmium ion is formed; or for each chloride incorporated, a loosely-bound electron exists in an orbit about the chloride ion.

Pure cadmium sulfide crystals have a very low dark conductivity (about 10^{-12} mho/cm) and a low photosensitivity. The incorporation of only about 2 parts per million of chloride increases the dark conductivity to about 1 mho/cm. This is because the extra electron, which is loosely bound when chloride substitutes for sulfide, has a binding energy of about 0.04 ev [41]; the electron is therefore free at room temperature. But as the conductivity is increased from 10^{-12} mho/cm to 1 mho/cm by the incorporation of chloride, the photosensitivity is also increased by a factor between 10^3 and 10^6. [28] It is this increase in photosensitivity in spite of the fact that the incorporation of impurities potentially provides more recombination centers, which is easily explained by the concepts of Rose, previously discussed.

As the electrons provided by the incorporation of chloride increase the dark conductivity, they will also shift the Fermi level and bring new centers into the role of recombination centers. These new recombination centers will be filled in the dark and hence, if their nature is appropriate, they will be able to function as the small capture cross section centers of Class II type previously described. Holes formed by excitation will aggregate in the small capture cross section Class II centers and the electrons formerly in Class II centers will be transferred to the large capture cross section Class I centers which were responsible for the low sensitivity of the pure crystals. Thus, the sensitization process decreases the rate of capture of electrons by Class I centers, and hence, increases the lifetime of free electrons.

The effect of copper (and silver) in decreasing the dark conductivity and photosensitivity can be explained by considering the copper to play the role of a low-lying acceptor center; the copper centers accept the electrons provided by the incorporation of the chloride or other donor impurities, decrease the conductivity, and hence remove the sensitization by removing Class

II centers from their role of recombination centers. In addition, the copper centers now provide an additional sensitivity in the red caused by direct excitation from the copper centers to the conduction band. When high proportions of copper are incorporated in evaporated, powder, or sintered layers, the spectral response may remain high out to 9,000A. [10, 11, 24]

Bube and Thomsen [28] have reported specifically on the effects of Cl, Br, I, Al, Ga, In, Cu, and Ag impurity on the conductivity and photosensitivity in CdS and CdSe. Veith [10] and Goercke [11] have reported on the effects of excess Cd, and Cu and Ag impurity in CdS. Kroeger et al [29] have described the effects of Ga, In, Sb, Cl, and Ag impurity on the conductivity of CdS.

A number of authors have reported on the effect of oxygen impurity in CdS, CdSe, and CdTe, particularly on the properties of evaporated layers. Oxygen impurity has been reported to have the following effects: decrease in dark conductivity and slight decrease in photosensitivity of CdS after exposure to oxygen at room temperature (reversible in vacuum), but a permanent increase in dark conductivity by prolonged heating in oxygen [71], improved sensitivity, particularly for CdSe, if prepared in an atmosphere containing oxygen [15–17], extension of spectral response of CdS to longer wavelengths for up to 2 per cent oxygen impurity [37], and increase in concentration of trapping centers in CdS [72–74].

Electrode Contact Problems

For most applications, both practical and theoretical, it is desirable that the current through a photoconductor be proportional to the voltage applied to the photoconductor; i.e., that the material obey Ohm's law. An ohmic contact is one made with an electrode which is able to supply to the crystal an excess or a reservoir of carriers ready to enter the photoconductor as needed. Such an ohmic contact to an n-type insulating crystal is obtained by using a metal with work function smaller than that of the insulator. The nature of the effects of electrode material on the electrical properties of CdS has been investigated by Broser and Warminsky [75] and by Buttler and Muscheid [76]. By various treatments, Buttler and Muscheid were able to make ohmic contacts to CdS using both gold and aluminum electrodes; they found that by providing ohmic contacts to the crystals, undesirable effects such as rectification, instability, irreversible "forming" changes, and noise, were either eliminated or greatly reduced.

Following the approach of using materials with low work functions for the electrodes, Smith and Rose [77, 78] have shown that ohmic contacts to CdS can be made with indium or gallium electrodes (either melted or evaporated). In such crystals, the photovoltaic effect is absent, and Shulman et al. [79] have shown that the noise is much less than for crystals with silver electrodes.

Sintered layers of CdS or CdSe give a photocurrent which obeys Ohm's law for applied fields as small as 2 mv/cm, even when silver electrodes are used. [24]

Conductivity

The photoconductors discussed in this paper are n-type conductors, with the exception of CdTe which may exhibit either n-type or p-type conductivity, and of Znte which has been found only p-type to date.

The experiments of Smith [80] on electroluminescence in pure CdS crystals suggest that the emission is the result of the recombination of electrons and holes after injection of both charge carriers from the electrodes; in such pure crystals, hole conductivity may play a role in the electronic processes, but in most sensitive photoconducting crystals, it is almost certain that hole conductivity is completely negligible. Sommers [81] has shown from measurements of the photo-electro-magnetic effect on the CdS crystals used by Smith that the hole lifetime in these crystals is about 0.1 μsec., about 1/10 of that of the electron lifetime. In sensitive crystals, where the electron lifetime is 10^4 times or more longer than in pure crystals, it is likely that the hole lifetime is considerably shorter than in pure crystals.

Impurities from Groups III and VII of the periodic table act as n-type impurities in CdS, CdSe, and CdTe; impurities from Groups I and V act as p-type impurities. The activation energies for n-type impurities vary from a few hundredths of a volt in CdS [41, 70] to a few thousandths of a volt in CdTe [38]. Activation energies for p-type impurities have been obtained only for CdTe; they lie between 0.3 and 0.5 ev. It is probable that the activation energies of p-type impurities are even larger in CdS or CdSe, for which the band-gap is larger than in CdTe. The failure to detect any p-type CdS or CdSe when these same impurities are used may possibly be explained, therefore, by the large activation energies of p-type impurities which would be involved.

For high concentrations of Cl and Ga impurity in CdS, Kroeger et. al. [70] have reported that the activation energy becomes negligible for impurity concentrations of $10^{18}/cm^3$ and higher, and that conduction then takes place in an impurity band.

Hauffe [82] has measured the conductivity of CdSe layers in a partial pressure of Cd and Se as a function of temperature. For pressures greater than 10^{-1} mm Hg, he found that the conductivity increased as the 0.2 power of the pressure of the Cd vapor, or decreased as the 0.5 power of the pressure of the Se vapor.

Appel [69, 83] has reported on measurements of the temperature dependence of the conductivity of CdTe between 25° and 600° C.

Mobility

Values for the mobility of electrons in CdS have been reported by many investigators [40, 41, 43, 44, 70, 84]; they lie between about 1 and 200 cm^2/volt sec., with a probable average value of about 100 cm^2/volt sec. The mobility in CdSe is of the same order of magnitude.

Jenny and Bube [38] have reported that the electron mobility in CdTe is about 300 cm^2/volt sec. and that the hole mobility is about 30 cm^2/volt sec.

Measurements of the temperature variation of the electron mobility in CdS have been reported by Kroeger et al. [70]

Speed of Response

The growth and decay of both photoconductivity and luminescence are easy to measure, but difficult to interpret. Detailed measurements on the growth and decay of photoconductivity in ZnS and CdS have been made by many investigators [21, 73, 85–87], but the relative simplicity in fitting such curves with equations derived from many different theoretical models makes such fittings usually of little value.

It is generally true that the decay of photoconductivity consists of two parts: an initial rapid portion corresponding to the direct recombination of free electrons, and a subsequent, much slower portion corresponding to the recombination of electrons which have been thermally freed from traps. If high intensity excitation is used to give a high concentration of free electrons, the major portion of the decay curve will be of the first type, and the observed decay time will be equal to the true lifetime of a free electron needed to satisfy (2). If a low intensity excitation is used, the decay curve will be mainly of the second type and the observed decay time will be longer than the true lifetime of a free electron. The complex distribution of trapping levels and recombination levels in a real crystal makes theoretical calculations based in detail on a model containing only one or two sets of discrete levels of limited value only. Decay times vary from a few μsec. for insensitive crystals at high excitation intensity to a few seconds for sensitive crystals at low excitation intensity. CdSe photoconductors have a generally faster speed of response than CdS photoconductors.

The performance of a photoconductor can conveniently be described in terms of (2) by a comparison between the observed decay time and the calculated lifetime using the measured values of applied field and gain. For measurements on single crystals at high light intensities the observed decay time will be equal to the calculated lifetime, but at low light intensities, the observed decay time is frequently found to be as much as a hundred or a thousand times longer than the calculated lifetime. Under normal operating conditions, the observed decay times for powder layers are between 10^3 and 10^5 times longer than the lifetime calculated from (2), and for sintered layers, the observed decay times are about 10 to 10^3 times longer than the calculated lifetimes. The comparison of decay time with calculated lifetime for powder layers is strongly dependent on the applied field, because of the non ohmic current-voltage relationship found with powder layers.

There is a phenomenon involved in the growth of photoconductivity in CdS and CdSe which merits further brief consideration. The growth of photoconductivity in CdS crystals at room temperature depends strongly on the length of time which has elapsed since the previous excitation. After periods of darkness of several days, a pronounced S shaped growth curve is obtained, more than 20 seconds being required in one case, for example, for the photocurrent to rise the first 1 per cent of its equilibrium value. It has been shown by Bube [88] that such slow S shape growth curves occur only when the electron Fermi level passes through that portion of the forbidden gap which is between 0.6 and 0.8 ev from the conduction band in CdS, and between 0.3 and 0.6 ev from the conduction band in CdSe. The slow growth corresponds to the time required for readjustment of the occupancy of recombination levels. The holes become located at centers with small recombination cross section (Class II centers), and electrons are transferred from Class II centers to decrease the rate of recombination at large recombination cross section Class I centers.

The Class II levels postulated in this process probably lie slightly below the middle of the forbidden gap in both CdS and SdSe, and, as will be shown in the following sections, play an important role in many of the photoconductivity phenomena found in these materials. When the electron Fermi level is raised by the application of light through the critical range, the hole demarcation level corresponding to this electron Fermi level is lowered through that part of the forbidden gap over which these Class II levels are distributed. The crystal, therefore, is undergoing a process of sensitization in the growth of photoconductivity, and it is the time required to achieve the equilibrium sensitization that gives the slow S shape growth observed.

Photocurrent vs Light Intensitiy

Most simple considerations involving one class of recombination centers predict that the photocurrent should vary with a power of the light intensity, n, between 0.5 and 1.0, depending on the particular conditions assumed. Rose [51], for example, has shown how particular values of n of 0.5 and 1.0 can be obtained by considering a uniform distribution of trapping levels, and how any value between 0.5 and 1.0 can be obtained by considering an exponential distribution of trapping levels. Values of n in the range between 0.7 and 0.9 are very common in measurements on CdS crystals at room temperature.

In CdSe at room temperature, however, the photocurrent usually varies with a power of the light intensity greater than unity [40, 89], this phenomenon is called superlinear photoconductivity. Extensive measurements of superlinear photoconductivity in CdSe and also in CdS were made as a function of temperature by Bube [90]. Fig. 4 (next page) gives photocurrent vs light intensity data for a crystal of CdSe. A slope equal to or less than unity, S_1, is found at low temperatures, and at intermediate temperatures for high excitation intensity. A slope greater than unity, S_2, is found at intermediate temperatures for low excitation intensity, and at high temperatures for high excitation intensity.

A slope equal to or less than unity, S_3, is again found at high temperatures for low excitation intensity. An analysis of such data for many crystals of CdSe (all of which showed superlinearity, except for a very few crystals with very low sensitivity), shows that superlinearity occurs when and only when the electron Fermi level varies between 0.3 and 0.6 ev below the conduction band. Similar measurements on CdS at elevated temperatures show that superlinearity in CdS occurs when the electron Fermi level varies between 0.6 and 0.8 ev from the conduction band.

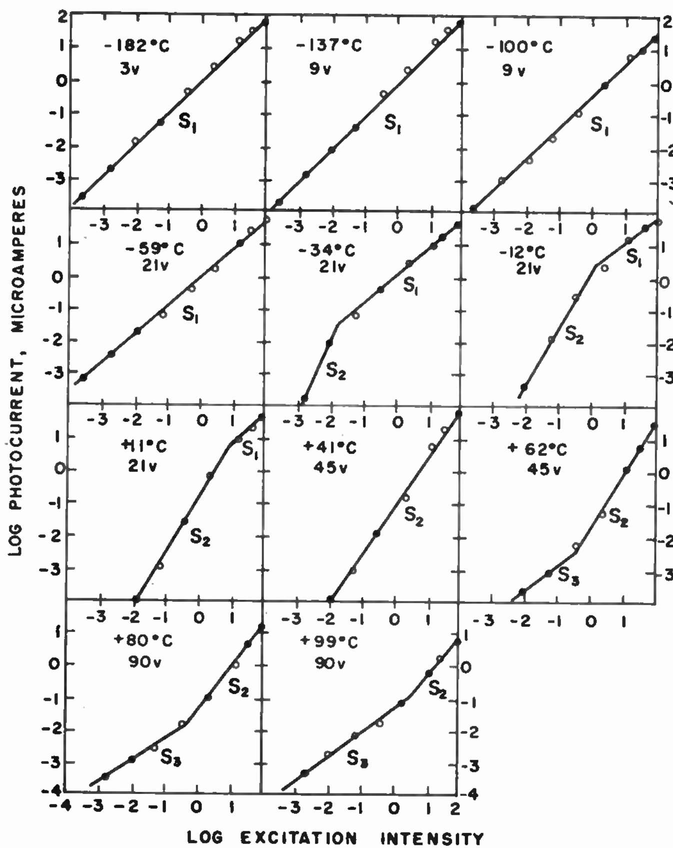


Fig. 4—Photocurrent of a crystal of CdSe as a function of excitation intensity for several different temperatures. The temperatures and the applied voltages are indicated on the figure. An excitation intensity of 100 is equivalent to 900 foot-candles.

According to Rose, superlinearity occurs when the hole demarcation level, the location of which is associated with the location of the electron Fermi level, moves down to include Class II crystal defect centers as recombination centers with small recombination cross section. Sensitization of the crystal—manifesting itself in the occurrence of superlinearity—occurs as long as the hole demarcation level moves down to include more centers of Class II type. When the Class II centers are not functioning as recombination centers, a linear or sublinear variation of photocurrent vs light intensity for the crystal in a desensitized state is found; when all of the Class II centers are acting as recombination centers, a linear or sublinear variation of photocurrent vs light intensity is again found, but this time for the crystal in a sensitized state. The location of the Class II levels is such that superlinear photoconductivity, for a normal range of excitation intensities (10^{-3} to 10^3 foot-candles), is found for CdSe between $-40°$ and 120°C, and for CdS between about 100° and 200°C.

In a practical sense, this means that, although CdS and CdSe may have equal sensitivities for high excitation intensities at room temperature, CdS will be considerably more sensitive than CdSe at low excitation intensities.

In evaporated, powder, or sintered layers in which high proportions of Cu are incorporated (up to 1,000 parts per million), superlinear photoconductivity is found for CdS at room temperature, probably associated in this case with levels caused by the Cu rather than with crystal defect levels [10, 24].

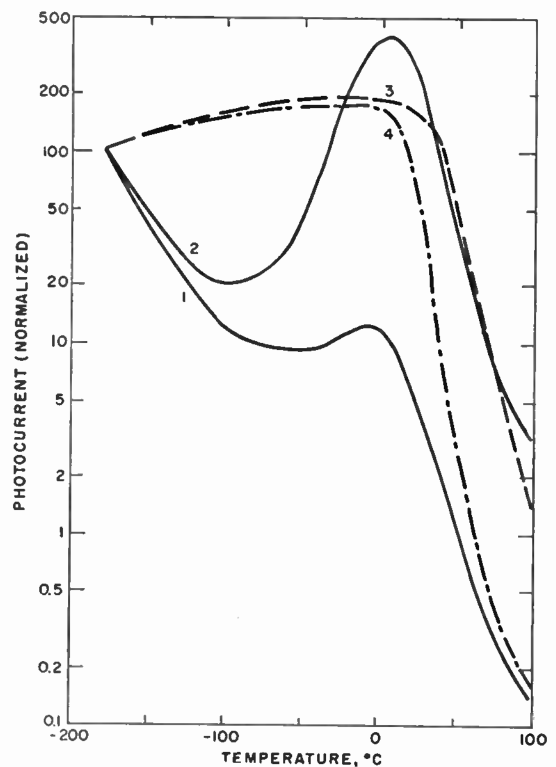


Fig. 5—Photocurrent as a function of temperature, normalized at $-180°$C, for an excitation intensity of 130 foot-candles, for 1) a crystal of CdSe prepared from vapor reaction of Cd and Se; 2) a crystal of CdSe prepared from sublimation of CdSe powder; 3) crystal of type (1) annealed 16 hours at 400°C under 20,000 psi argon; and 4) crystal of type (1) heated for 5 minutes at 900°C in Se vapor and quenched.

Photocurrent vs Temperature

Relatively small variations in photosensitivity with temperature can and do occur because of changes in the density and nature of levels in the neighborhood of the Fermi levels as they move with changing temperature. Such variations occur for most crystals below room temperature, and vary greatly with the preparation and treatment of the crystals. [91] Fig. 5 presents the normalized photocurrent vs temperature for four crystals of CdSe prepared by different techniques and sub-

jected to different treatments. [90] The variation of photocurrent with temperature below room temperature is dependent on the past history of the crystal.

The important thing to note, however, is that Fig. 5 also shows that all of these crystals of CdSe show a large and rapid decrease in photosensitivity when the temperature is raised above room temperature. Other measurements demonstrate that crystals of CdS show such a decrease in photosensitivity when the temperature is raised above about 100°C.

An analysis of measurements on the temperature break-point for the photosensitivity of CdSe and CdS crystals [90] indicates that the photosensitivity break-point is associated with a location of the Fermi level of 0.6 ev below the conduction band in CdS and of 0.3 ev below the conduction band in CdSe. A rapid decrease in photosensitivity occurs when the Fermi level drops further from the conduction band than this value. The actual temperature breakpoint for the photosensitivity is very sensitive to the excitation intensity, lying at lower temperatures for lower excitation intensities, in such a way as to give a constant location of the Fermi level for the breakpoint. Such a decrease in the photosensitivity occurs with increasing temperature, when the lowering of the electron Fermi level and the raising of the associated hole demarcation level start to remove Class II centers from acting as recombination centers and hence cause the crystal to revert to its desensitized condition.

A confirmation of the fact that the same Class II centers are responsible for both the temperature dependence and the superlinearity of photoconductivity is given by an examination of those few crystals of CdSe which do not show superlinear photoconductivity at room temperature. The photo-sensitivity of such crystals does not decrease even when the temperature is raised to 100°C.

The observation that certain photochemical reactions occur in CdS at temperatures over 100°C, caused by atomic displacements and rearrangements, has been reported by Boer. [92]

Infrared Quenching

The occurrence of infrared quenching of photoconductivity in CdS and ZnS has been reported by several investigators [8, 21, 48, 93, 94]. Taft and Hebb [95] found two quenching bands at 0.9 and 1.5 ev, which they associated with excitation of trapped holes in two different types of centers.

A more recent investigation of infrared quenching in CdS [88] has revealed the following details: three quenching "bands" with maxima at 0.89, 1.35, and 1.65 ev are found; insensitive crystals show only the 1.65 ev quenching; the 0.89 and 1.35 ev quenching "bands" are intimately connected, a linear relationship existing between their magnitudes in different crystals; quenching in the 0.89 ev "band" disappears for temperatures below about −50°C; all quenching disappears for temperatures greater than about 100°C. Fig. 6 shows infrared quenching spectra for several typical CdS crystals. Similar investigation of infrared quenching in CdSe crystals [88] reveals that infrared quenching in CdSe occurs only below −50°C; three quenching maxima, not well resolved, are found at 1.20, 1.05, and 0.79 ev.

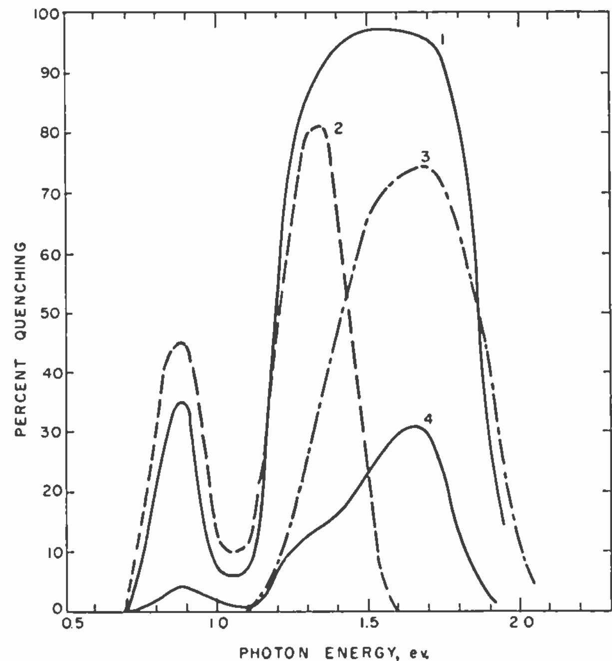


Fig. 6—Infrared quenching spectra at room temperature for four typical selected crystals of CdS. Relative sensitivities of the crystals are 2, 54, 0.2 and 0.8.

An analysis of these data shows that infrared quenching in CdS and CdSe can occur only when the Class II centers are acting as recombination centers; i.e., when the electron Fermi level lies above 0.6 ev from the conduction band in CdSe, and above 0.8 ev in CdS, and when the hole demarcation level therefore lies below the Class II levels. Infrared quenching, then, occurs when electrons are raised from the filled band to Class II centers, freeing a hole from a Class II center to be transferred to a Class I center where recombination is much more probable. Apparently two different types of Class II centers exist; in CdS, for example, one type is associated with the 1.65 ev quenching in insensitive crystals, and the other type is associated with the 1.35 and 0.89 ev quenching in sensitive crystals. A schematic representation of transitions involved in infrared quenching of CdS is given in Fig. 7 (next page). This transition scheme is based on assumption that the "band-shape" of the high-photon-energy quenching spectrum is not intrinsic to the process of quenching, but is really caused by a competition between excitation and quenching by the high-energy photons, excitation becoming much more prominent with increasing photon energy. Such a competition does exist over a range of energies slightly smaller than the band-gap. On this basis, the significant energy in the quenching spectrum is not the

energy for maximum quenching, but the lowest photon energy which will give any quenching at all. This lowest energy represents a threshold energy (the actual distance of the Class II defect levels above the valence band); for photons of higher energy than the threshold energy, the quenching arises from transitions from lower in the valence band.

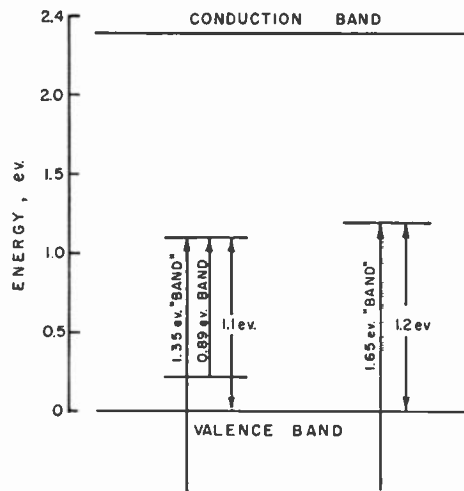


Fig. 7—Schematic representation of the infrared quenching transitions in CdS.

Thermally Stimulated Current

If a crystal is cooled to a low temperature, excited, and then heated in the dark at a constant rate, the difference between the current measured during the heating and the normal dark current at the corresponding temperature is the thermally-stimulated current. The thermally-stimulated current (analogous to the glow curve for measurement of luminescence) is contributed by the thermal emptying of filled energy levels near the conduction band. In the interpretation of measurements of thermally-stimulated current, it must be remembered that the magnitude of the current at a given temperature depends both on the concentration of levels emptying at that temperature and on the lifetime of a free electron at that temperature.

Fig. 8 shows typical curves of thermally stimulated current obtained for four crystals of CdS [88]. Such crystals show a remarkably reproducible trapping distribution from one crystal to another, as determined by these measurements; the curves of Fig. 8 indicate the presence of possibly as many as seven different current peaks which are common to all four crystals. The approximate depth of the trapping level can be determined by calculating the location of the electron Fermi level from the temperature and the conductivity of the current maximum. Such values are listed on Fig. 8. The dotted curve, which is for a more sensitive crystal (the ordinates having been reduced by a factor of 10 before plotting), has its peaks shifted to higher temperatures relative to the curves for the less sensitive crystals; the shift in the peaks is exactly that which will give the same trap depths as calculated from the Fermi level for the peak magnitudes and peak temperature locations for all four curves. A detailed study of the thermally-stimulated current curves for the crystals of CdS and CdSe for which superlinearity and infrared quenching have been observed, indicate that trapping levels lying about 0.4 and 0.7 ev below the conduction band seem to be characteristic of CdS crystals, and levels lying about 0.4 ev below the conduction band seem to be characteristic of CdSe crystals. [41, 88, 90]

Other investigations of thermally stimulated currents have been reported by Herman and Meyer [96], Muscheid [72], and Jensen. [97] The concentration of trapping levels in CdS has been reported by Bube [41] to vary between about $10^{13}/cm^3$ for "pure" insensitive CdS to $10^{17}/cm^3$ or higher for conducting CdS:Cl crystals. When only one thermally-stimulated current peak is prominent, an estimate can also be made of the "attempt-to-escape" frequency from the trap levels and hence of the capture cross section of these levels for free electrons; values of 10^6 sec^{-1} for the frequency and 10^{-20} cm^2 for the cross section have been obtained for pure CdS. [41]

Other investigations on the nature of trapping, the distribution of traps, and the effect of trapping on the decay time of CdS photoconductor have been made by Thielemann [87], Broser and Warminsky [98], and Niekisch [99].

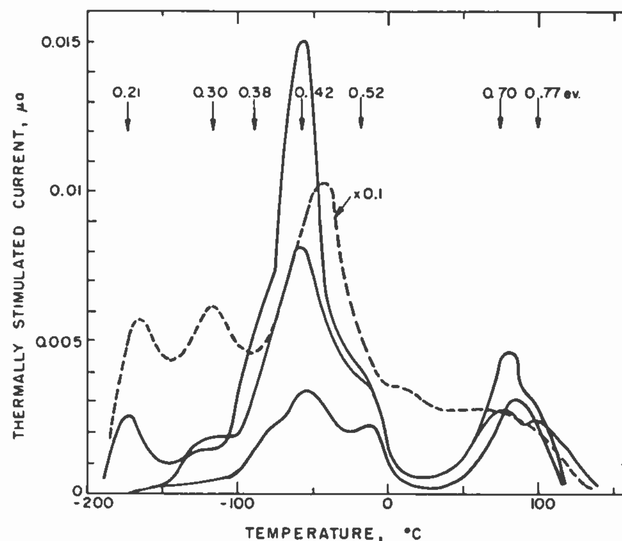


Fig. 8—Thermally-stimulated current curves for three "pure" insensitive CdS crystals (solid curves) and a sensitive CdS:Al crystal (dotted curve). Heating rate of 0.77 degree per sec.

Space-Charge Limited Current

Although it might be expected from the nature of an insulator that carriers could move freely through the solid if they could be injected into the insulator, the magnitude of the current being limited only by the space charge of the carriers themselves, little experimental evidence for such currents was reported until the work of Rose and Smith. [100–102] They showed

that large, steady space-charge limited currents (as high as 20 amp/cm²) could be drawn through a thin insulating crystal of CdS fitted with ohmic electrodes. Low fields (10^3–10^{15} volts/cm) were used to eliminate the possibility of collision ionization or field emission from traps. The variation of the current as a high power of the voltage (the fourth power or higher) was explained in terms of the effect of traps on the build-up of the space-charge limited current. The dynamics of the space-charge limited current build-up with varying field support this description.

Direct detection of the space charge in a crystal was made by dropping a crystal, which had been subjected to an applied field, onto the pan of an electrometer. A negative charge was found regardless of the polarity of the initially applied field, and the magnitude of the charge agreed fairly well with the estimated value.

Fig. 9 shows the data obtained with a crystal of CdS, 2.5×10^{-3} cm thick and with an electrode area of 5×10^{-4} cm². Curve I_0 is for the current through the crystal in the dark after a time sufficient to establish thermal equilibrium. Curves F_1, F_2, and F_3 were taken with the crystal exposed to three different intensities of light, increasing in the order given. When the volume-generated carriers exceed the space-charge injected carriers, the behavior is ohmic. The initial space-charge limited current after application of the field, before trapping affected its magnitude, was detected using a pulse technique with the results shown in the upper curve of Fig. 9. The magnitudes of these currents are very close to the theoretical values of space-charge limited currents calculated for a trap-free solid.

Curve $I^{0\prime}$ of Fig. 9 is similar to curve I_0, but it was taken within a few minutes after the crystal had been exposed to light. The much higher currents observed under these conditions can still be simply explained in terms of space-charge limited currents; an alternate explanation based on field emission from traps was proposed by Boer and Kuemmel. [103]

The measurement of space-charge limited currents can be used as an additional tool for the determination of the concentration of trapping levels in insulators, especially useful when the concentration of such levels is small.

Photoconductivity and Luminescence

It has been only natural that investigators in the field should have sought for a correlation between photoconductivity and luminescence in the same material. Although there have been some statements on both extremes; i.e., either that photoconductivity is completely unrelated to luminescence, or that photoconductivity is exactly analogous to luminescence, the present consensus seems to be that photoconductivity and luminescence share many features in common, but that there are also some marked differences. The most striking difference is the dissimilarity in the decay times; luminescence emission generally decays much faster

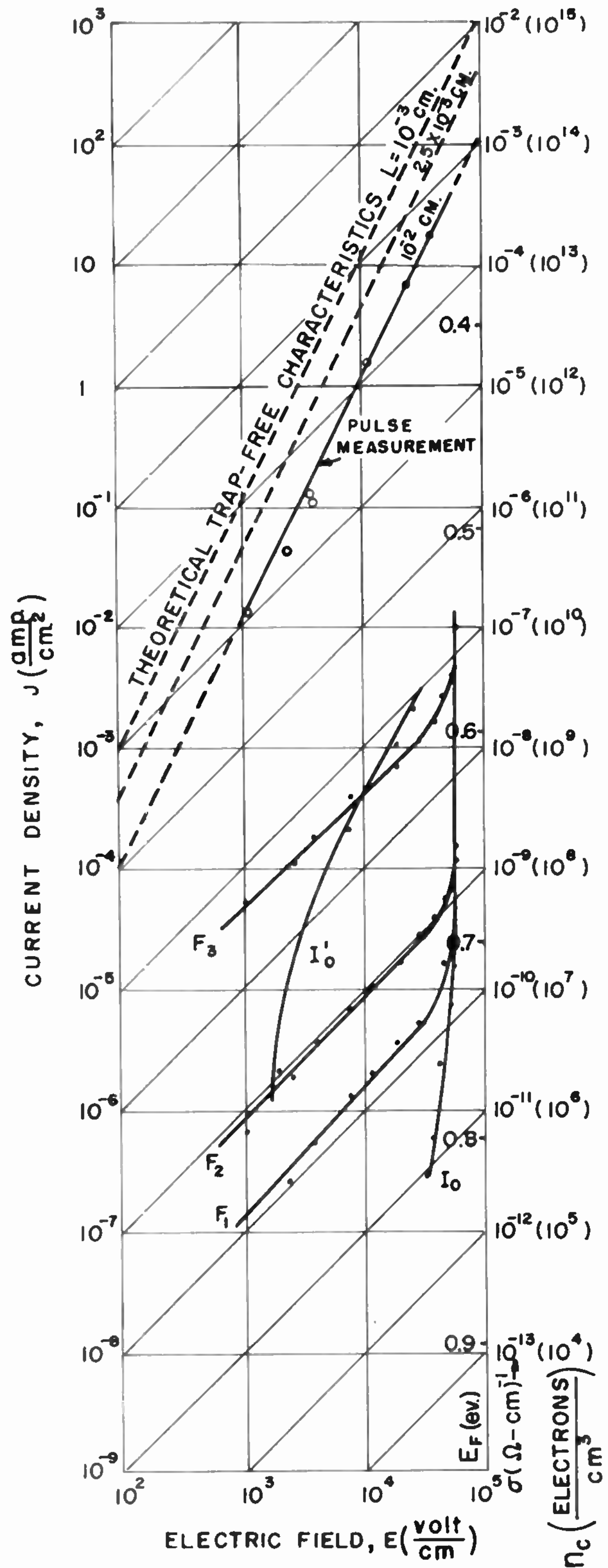


Fig. 9—Space-charge-limited current in an insulator. I_0' is initial dark current curve after exposure to light; I_0 is the thermal equilibrium dark current curve. F_1, F_2, and F_3 curves obtained with different light levels on the crystal, increasing in the order given. The current obtained from pulse measurements is compared with the theoretical space-charge limited current in a trap-free solid (calculated for $\mu=100$ cm²/volt sec and dielectric constant of 10). The conductivity, σ, the density of carriers in the conduction band, n_c, and the calculated electron Fermi level are also plotted on the right hand axis.

after the cessation of excitation than the photocurrent. [104, 105] A qualitative picture can be given as follows: as a result of excitation, many holes are captured at luminescence centers, and many others are captured at other centers throughout the crystal associated with defects and imperfections; at the cessation of excitation, free electrons are captured most rapidly by the luminescence centers, the concentration in the conduction band being partially maintained by the emptying of shallow trapping levels; after the major portion of the luminescence has decayed, the major portion of the photoconductivity still remains to undergo a slow recombination at those levels which have a smaller recombination cross section than the luminescence centers.

Many investigators have set up theoretical models for comparing photoconductivity and luminescence, both in the steady state [44, 46–48, 106, 107] and during the decay. [85, 96, 98]

Although Wilkins and Garlick [108] have argued for an intimate relationship between luminescence centers and trapping centers, so that an electron freed from a trap need not contribute to the photocurrent in returning to the luminescence center, the similarity of excitation spectra for luminescence emission, trapping, and photoconductivity for ZnS, ZnS:Ag, and ZnS:Cu phosphors indicates that, at least in these materials, luminescence centers and trapping centers are separate, and photoconductivity does result from the return to a luminescence center of an electron freed from a trap [20].

Smith [80] has reported on the electroluminescence of CdS crystals at low fields, associated with the recombination of injected electrons and holes, and Diemer [109] has reported on the light emission observed near dc breakdown of CdS:Cl crystals for which the current-voltage relationship is very similar to that of discharge in gases. Herforth and Krumbiegel [110] describe the effects of ultrasonics on the conductivity and the luminescence of ZnS and CdS.

Photovoltaic Effect

Using crystals of CdS with one ohmic and one non-ohmic contact, Reynolds *et al.* [111, 112] have developed a photovoltaic cell which gives an open circuit voltage of 0.4 volt in direct sunlight, and a short circuit current of 15 ma/cm^2. At 150°C, the open circuit voltage is one-half of its room temperature value, and the short circuit current is one-fifth of its room temperature value. The spectral response for the photovoltaic effect of this cell has maxima at 5,000A and at 7000A; authors propose existence of an intermediate band in forbidden gap of CdS to explain red photovoltaic response.

Photoemissive Effect

Apker and Taft [113] have described the field emission obtained from sharp needles of CdS and CdSe. A large increase in the field emission current was observed when the crystals were irradiated with light of such a wavelength range as to excite photoconductivity.

Surface Photoconductivity

When the spectral response of photoconductivity is measured, it is quite frequently found that the sensitivity is very high in a narrow region of wavelengths near the absorption edge; for shorter wavelengths the sensitivity drops to a constant value smaller than the peak by an order of magnitude or more, and for longer wavelengths the sensitivity decreases rapidly to negligible values as the energy of the photons becomes much smaller than the band gap of the photoconductor [114]. Reason for decrease in photosensitivity for excitation by photons of greater energy than band-gap, is sought in difference between surface sensitivity, the important feature for these highly-absorbed photons, and volume sensitivity, the important feature for those photons which penetrate most of the crystal thickness.

An investigation of the effect of water vapor on the surface photoconductivity of CdS and ZnS crystals has shown that adsorbed water vapor decreases the surface sensitivity by increasing the recombination rate at the surface. [104, 115, 116]

The difference between surface and volume sensitivity of pure CdS crystals as determined by measurements of the spectral response of photoconductivity, becomes much more marked the greater the volume sensitivity of the crystal [117]. Measurements by Klick [64] on CdS indicate this difference disappears with decreasing temperature, being practically absent at 4°K.

Tanaka and Aoki [50] deduce the presence of photosensitive surface barriers in CdS crystals from the dependence of surface potentials on wavelength. Wlerick [118] has measured the variation of the contact potential of CdS with illumination.

Conclusion

The future should see the continuance of the multiplication of applications for photoconductors, branching out from a variety of simple control and detection uses to such elaborate utilization as in light amplification and picture reproduction, television camera pickup, electrophotography, and conversion of solar energy.

It would appear that an ideal photoconductor would have the following characteristics:

1) A sufficiently large band-gap to provide the required resistivity for needs of specific application.

2) A sufficiently small band-gap to insure response over the desired spectral range.

3) A sufficiently perfect crystal structure with a minimum of trapping levels to permit the theoretical speed of response to be obtained at low light intensities.

4) A sufficient control over the concentration and nature of defects to permit the exclusion of centers with a large recombination cross section.

Bibliography

[1] Frerichs, R., and Warminsky, R., *Naturwissenschaften*, Vol. 33 (1946), p. 251.
[2] Kallmann, H., and Warminsky, R., *Research (London)*, Vol. 2, Supp. 389 (1949).
[3] Jacobs, J. E., *Electrical Engineering*, Vol. 70 (1951), p. 667.
[4] Frohnmeyer, Glocker, and Messner, *Naturwissenschaften*, Vol. 40 (1953), p. 338.
[5] Simon, H., *Annalen der Physik*, Vol. 12 (1953), p. 45.
[6] Jacobs, J. E., *General Electric Review*, Vol. 57 (1954), p. 28.
[7] Gudden, B., and Pohl, R. W., *Zeitschrift für Physik*, Vol. 2 (1920), pp. 181, 361; Vol. 3 (1920), p. 98.
[8] Frerichs, R., *Naturwissenschaften*, Vol. 33 (1946), p. 281.
[9] Weiss, K., *Zeitschrift für Naturforschung*, Vol. 2A (1947), p. 650.
[10] Veith, W., *Comptes Rendus*, Vol. 230 (1950), p. 947; *Zeitschrift für Angewandte Physik*, Vol. 7 (1955), p. 1.
[11] Goercke, P., *Annalen der Telecommunications*, Vol. 6 (1951), p. 325.
[12] Aitchison, R. E., *Nature*, Vol. 167 (1951), p. 812.
[13] Kuwabara, G., *Journal of Physical Society of Japan*, Vol. 9 (1954), p. 97.
[14] Gottesman, J., *Journal of the Optical Society of America*, Vol. 44 (1954), p. 368.
[15] Schwarz, E., *Nature*, Vol. 162 (1948), p. 614.
[16] Schwarz, E., *Proceedings of the Physical Society*, Vol. 62A (1949), p. 530; Vol. 63B (1950), p. 624; Vol. 64B (1951), p. 821.
[17] Görlich, P., and Heyne, J., *Optik*, Vol. 4 (1948), p. 206.
[18] Braithwaite, J. G. N., *Proceedings of the Physical Society*, Vol. 64B (1951), p. 274.
[19] Randall, J. T., and Wilkins, M. H. F., *Proceedings of the Royal Society*, Vol. 184A (1945), p. 347.
[20] Bube, R. H., *Physical Review*, Vol. 90 (1953), p. 70.
[21] Hardy, A. E., *Transactions of the Electrochemical Society*, Vol. 87 (1945), p. 355.
[22] Kuwabara, G., *Journal of the Physical Society of Japan*, Vol. 9 (1954), p. 992.
[23] Kolomiets, B. T., *Doklady Akad. Nauk.* S.S.S.R., Vol. 83 (1952), p. 561.
[24] Thomsen, S. M., and Bube, R. H., *The Review of Scientific Instruments*, Vol. 26 (1955), p. 664.
[25] Nicoll, F. H., and Kazan, B., *Journal of Optical Society of America*, Vol. 45 (1955), p. 647.
[26] Frerichs, R., *Physical Review*, Vol. 72 (1947), p. 594.
[27] Schossberger, F., *Journal of Electrochemical Society*, Vol. 102 (1955), p. 22.
[28] Bube, R. H., and Thomsen, S. M., *Journal of Chemical Physics*, Vol. 23 (1955), p. 15.
[29] Kroeger, F. A., Vink, H. J., and van den Boomgaard, J., *Zeitschrift für Physikalische Chemie*, Vol. 203 (1954), p. 1.
[30] Hashimoto, K., *Proceedings of the Physical Society of Japan*, Vol. 7 (1952), p. 276.
[31] Krumbiegel, J., *Naturwissenschaften*, Vol. 39 (1952), p. 447.
[32] Krumbiegel, J., *Zeitschrift für Naturforschung*, Vol. 9A (1954), p. 903.
[33] Kremheller, A., *Sylvania Technologist*, Vol. 8 (1955), p. 11.
[34] Piper, W. W., *Journal of Chemical Physics*, Vol. 20 (1952), p. 1343.
[35] Reynolds, D. C., and Czyzak, S. J., *Physical Review*, Vol. 79 (1950), p. 543.
[36] Czyzak, Craig, McCain, and Reynolds, *Journal of Applied Physics*, Vol. 23 (1952), p. 932.
[37] Reynolds, Czyzak, Allen, and Reynolds, *Journal of the Optical Society of America*, Vol. 45 (1955), p. 136.
[38] Jenny, D. A., and Bube, R. H., *Physical Review*, Vol. 96 (1954), p. 1190.
[39] Miyasawa, H., *Journal of Physical Society of Japan*, Vol. 9 (1954), p. 648.
[40] Smith, R. W., *RCA Review*, Vol. 12 (1951), p. 350.
[41] Bube, R. H., *Journal of Chemical Physics*, Vol. 23 (1955), p. 18.
[42] Fassbender, J., *Annalen der Physik*, Vol. 5 (1949), p. 33.
[43] Fassbender, J., and Lehmann, H., *Annalen der Physik*, Vol. 6 (1949), p. 215.
[44] Broser, I., and Warminsky, R., *Annalen der Physik*, Vol. 7 (1950), p. 289.
[45] Broser, Kallmann, and Warminsky, *Zeitschrift für Naturforschung*, Vol. 4A (1949), p. 631.
[46] Frerichs, R., *Physical Review*, Vol. 76 (1949), p. 1869.
[47] Schoen, M., *Physica*, Vol. 20 (1954), p. 930.
[48] Kallmann, H., and Kramer, B., *Physical Review*, Vol. 87 (1952), p. 91.
[49] Rose, Weimer, and Forgue, *Physical Review*, Vol. 76 (1949), p. 179.
[50] Tanaka, S., and Aoki, M., *Oyo Butsuri*, Vol. 22 (1953), p. 311.
[51] Rose, A., *RCA Review*, Vol. 12 (1951), p. 362.
[52] Rose, A., Paper in "*Proceedings of the Conference on Photoconductivity.*" November, 1954, John Wiley & Son. In press.
[53] Rose, A., *Physical Review*, Vol. 97 (1955), p. 322.
[54] Hoehler, G., *Annalen der Physik*, Vol. 4 (1949), p. 371.
[55] Moss, T. S., *Proceedings of the Physical Society*, Vol. 63B (1950), p. 167.
[56] Kroeger, F. A., and Hellingman, J. E., *Journal of the Electrochemical Society*, Vol. 93 (1948), p. 156.
[57] Shrader, R. E., unpublished data.
[58] Czyzak, Reynolds, Allen, and Reynolds, *Journal of the Optical Society of America*, Vol. 44 (1954), p. 864.
[59] Mollwo, E., *Reichsber. Physik*, Vol. 1 (1944), p. 1.
[60] Piper, W. W., *Physical Review*, Vol. 92 (1953), p. 23.
[61] VanDoorn, C. Z., *Physica*, Vol. 20 (1954), p. 1155.
[62] Bube, R. H., *Physical Review*, Vol. 98 (1955), p. 431.
[63] DeVore, H. B., unpublished data.
[64] Klick, C. C., *Physical Review*, Vol. 89 (1953), p. 274.
[65] Gobrecht, H., and Bartschat, A., *Zeitschrift für Physik*, Vol. 136 (1953), p. 224.
[66] Seiwert, R., *Annalen der Physik*, Vol. 6 (1949), p. 241.
[67] Caspary, R., and Mueser, H., *Zeitschrift für Physik*, Vol. 134 (1952), p. 101.
[68] Niekisch, E. A., *Annalen der Physik*, Vol. 8 (1951), p. 291.
[69] Appel, J., *Physica*, Vol. 20 (1954), p. 1110.
[70] Kroeger, Vink, and Volger, *Philips Research Reports*, Vol. 10 (1955), p. 39.
[71] Muscheid, W., *Annalen der Physik*, Vol. 13 (1953), p. 305.
[72] Muscheid, W., *Annalen der Physik*, Vol. 13 (1953), p. 322.
[73] Seraphin, B., *Annalen der Physik*, Vol. 13 (1953), p. 198.
[74] Hoehler, G., *Annalen der Physik*, Vol. 14 (1954), p. 426.
[75] Broser, I., and Warminsky, R., *Z. Naturforsch*, Vol. 5A (1950), p. 62.
[76] Buttler, W. M., and Muscheid, W., *Annalen der Physik*, Vol. 14 (1954), p. 215; Vol. 15 (1954), p. 82.
[77] Smith, R. W., and Rose, A., *Physical Review*, Vol. 92 (1953), p. 857(A).
[78] Smith, R. W., *Physical Review*, Vol. 97 (1955), p. 1525.
[79] Schulman, C., Smith, R. W., and Rose, A., *Physical Review*, Vol. 92 (1953), p. 857(A); Vol. 98 (1955), p. 384.
[80] Smith, R. W., *Bulletin of the American Physical Society*, Vol. 30 (1955), p. LA7.
[81] Sommers, H. S., Jr., unpublished data.
[82] Hauffe, K., *Annalen der Physik*, Vol. 15 (1954), p. 141.
[83] Appel, J., *Zeitschrift für Naturforschung*, Vol. 9A (1954), p. 265.
[84] Gildart, L., and Ewald, A. W., *Physical Review*, Vol. 83 (1951), p. 359.
[85] Gurevich, D. B., et. al., *Zhur. Ekspti. Teoret. Fiz.*, Vol. 20 (1950), p. 769.
[86] Fassbender, J., and Seraphin, B., *Annalen der Physik*, Vol. 10 (1952), p. 374.
[87] Thielemann, W., *Wiss. Z. Univ. Leipzig, Math-Naturw. Reihe*, No. 2 (1951–1952).
[88] Bube, R. H., *Physical Review*, Vol. 99 (1955), p. 1105.
[89] Dirksen, H. J., and Memelink, O. W., *Applied Scientific Research*, Vol. 4B (1954), p. 205.
[90] Bube, R. H., Paper in "*Proceedings of the Conference on Photo-Conductivity.*" Atlantic City, 1954, John Wiley & Son. In press.
[91] Niekisch, E. A., and Rompe, R., *Zeitschrift für Physikalische Chemie*, Vol. 198 (1951), p. 200.
[92] Boer, K. W., *Physica*, Vol. 20 (1954), p. 1103.
[93] Leverenz, H. W., "*An Introduction to Luminescence of Solids.*" John Wiley & Son (1950), pp. 164, 302–304, 396.
[94] Liebson, S. H., Paper at the May, 1955, Electrochemical Society Meeting in Cincinnati.
[95] Taft, E. A., and Hebb, M. H., *Journal of the Optical Society of America*, Vol. 42 (1952), p. 249.
[96] Herman, R. C., and Meyer, C. F., *Journal of Applied Physics*, Vol. 17 (1946), p. 743.
[97] Jensen, H. C., *Physical Review*, Vol. 96 (1954), p. 798.
[98] Broser, I., and Warminsky, R., *Zeitschrift für Physik*, Vol. 133 (1952), p. 340.
[99] Niekisch, E. A., *Zeitschrift für Naturforschung*, Vol. 9A (1954), p. 700; *Annalen der Physik*, Vol. 15 (1955), pp. 279, 288.
[100] Rose, A., and Smith, R. W., *Physical Review*, Vol. 92 (1953), p. 857A.
[101] Smith, R. W., and Rose, A., *Physical Review*, Vol. 97 (1955), p. 1531.
[102] Rose, A., *Physical Review*, Vol. 97 (1955), p. 1538.
[103] Boer, K. W., and Kuemmel, U., *Zeitschrift für Naturforschung*, Vol. 9A (1954), p. 177.
[104] Bube, R. H., *Physical Review*, Vol. 83 (1951), p. 393.
[105] Lambe, J., and Klick, C. C., *Physical Review*, Vol. 98 (1955), pp. 909, 985.

[106] Broser, I., and Warminsky, R., *Zeitschrift für Naturforschung*, Vol. 6A (1951), p. 85.
[107] Gurevich, D. B., et. al., *Doklady Akad. Nauk. S.S.S.R.*, Vol. 71 (1950), p. 29.
[108] Wilkins, M. H. F., and Garlick, G. F. J., *Nature*, Vol. 161 (1948), p. 565.
[109] Diemer, G., *Philips Research Reports*, Vol. 9 (1954), p. 109.
[110] Herforth, L., and Krumbiegel, J., *Naturwissenschaften*, Vol. 42 (1955), p. 39.
[111] Reynolds, Leies, Antes, and Marburger, *Physical Review*, Vol. 96 (1954), p. 533.
[112] Reynolds, D. C., and Czyzak, S. J., *Physical Review*, Vol. 96 (1954), p. 1705.
[113] Apker, L., and Taft, E., *Physical Review*, Vol. 88 (1952), p. 1037.
[114] Fassbender, J., *Naturwissenschaften*, Vol. 34 (1947), p. 212.
[115] Bube, R. H., *Journal of Chemical Physics*, Vol. 21 (1953) p. 1409.
[116] Liebson, S. H., *Journal of Electrochemical Society*, Vol. 101 (1954), p. 362.
[117] Bube, R. H., unpublished data.
[118] Wlerick, G., *Physica*, Vol. 20 (1954), p. 1099.

Performance of Photoconductors*

ALBERT ROSE†, FELLOW, IRE

Summary—All semi-conductors and insulators are photoconductors. The characteristic parameter of a photoconductor is the life time of a free carrier. The photo-electron current is equal to the product of absorbed photon current and the ratio of life time to transit time of a free carrier. This relation holds as well for the commonly known types of barriers as for uniform photoconductors. The photo-electron current may range from a small fraction of the photon current to many powers of ten greater than the photon current depending on the ratio of life time to transit time. There is good evidence for life times in different photoconductors extending from 10^{-12} seconds to values approaching a second. The life times of free electrons and free holes are in general independent of each other and markedly different. Only at sufficiently high excitation rates for which the free carrier densities exceed the densities of bound states do the electron and hole life times necessarily become equal.

The usual termination of the life time of a free carrier is by recombination with a deep lying bound state in the forbidden zone. These recombination processes are structure sensitive, complex, and generally not amenable to exact solution. A useful insight can be gained by dividing the bound states into deep-lying states, called ground states, that govern the rates of recombination and thereby the free carrier densities; and shallow lying bound states, called traps, that are responsible for observed response times exceeding the carrier life times. The ratio of observed response time to free carrier life time is of the order of the ratio of shallow trapped to free carriers. The ground states can be identified, accurately in some cases and approximately in others, as those states lying between the steady-state Fermi-levels for electrons and holes. The number of ground states increases with increasing light intensity and decreasing temperature since the energy separation of the steady-state Fermi-levels increases and embraces more ground states under these conditions. This point of view can readily account for the large variety of dependencies of photocurrent on light intensity and temperature. The special problems of infra red quenching, super-linearity, and impurity activation require the consideration of at least two kinds or classes of ground states.

A photoconductor can be a noiseless, or substantially noiseless, transducer of photon currents into electron currents. A very few measurements have confirmed this possibility. The bulk of all observations of noise are dominated by noise currents far in excess of the minimum attainable and noise currents that have at low frequencies a $1/f$ spectrum instead of a flat spectrum. The most likely source of these excess noise currents appears to be at barrier-type surfaces either at the metal-photoconductor contacts or at internal surfaces. Logically a collection of noise sources having a distribution of time constants and a high local gain are needed to account for the excess noise. Surfaces, normal to the flow of current, can provide the set of independent time constants needed, whereas a set of noise sources finely distributed in the volume of a photoconductor are likely to be averaged by the wandering motions of the free carriers into a *single* time constant. Barrier-type surfaces also provide the high local gains needed by virtue of their large ratios of carrier life time to carrier transit time. The transit time is by definition a small quantity since a barrier is a thin element. The noise currents associated with thermally generated carriers can be different from and smaller than those associated with optically generated carriers if the thermally generated carriers originate from different centers lying closer to the band edges and/or have shorter life times. A simple heuristic measure of the noise currents of thermally generated carriers is proposed whereby the noise is a maximum when the Fermi-level is invariant with temperature and decreases toward zero as dE_f/dT approaches E_f/T.

Introduction

THIS paper is concerned with a phenomenological analysis of the processes whereby the conductivity of a material is increased by exposure to radiation. The processes include the excitation of electrons from low lying free or bound states to higher lying free or bound states; the recombination of the excited electrons into their ground states and the steady state increase[1] in numbers of free electrons and free holes consistent with the rates of excitation and recombination.

By material is meant almost all semi-conductors and insulators. The "almost" is included only to cover the possibility that convincing evidence may in the future be found that the conductivity of some semi-conductor or insulator cannot be increased by any of the electron-exciting radiations. The preponderance of evidence is thus far, as one would expect from simple considerations, that every semi-conductor and insulator can be made more conducting by absorbing electron-exciting radiation. The materials include not only the well known photoconductors [1] such as the metal

* Original manuscript received by the IRE August 5, 1955.
† RCA Laboratories, Princeton, N. J.

[1] There is some evidence that radiation can lower the "dark" conductivity of a material. These are rare occurrences and are not discussed here.

sulphides, selenides, oxides and halides and the semiconductors used in transistor studies such as germanium, silicon and the inter-metallic compounds but also some highly insulating, non-crystalline materials such as amorphous selenium [2] evaporated layers of silica [3] and organic compounds like polythene [4].

By radiation is meant any radiation that results in the direct excitation of electrons. Thermal radiation, when it is absorbed by the lattice vibrations of a solid and passed on to the electrons through these lattice vibrations is not included. While a useful continuity of concepts can be established between photoconductivity and the increase in conductivity by simple heating, that continuity will not be pursued here. Thermal radiation when it is absorbed directly by electrons is included. So also is the gamut of photon radiation (infra-red, visible, ultra-violet, x-ray and γ-rays). Since photoconductivity is concerned chiefly with the rate of excitation and recombination of electrons, it would be an artificial separation to exclude excitation by particles such as electrons, α-particles, β-rays or other nuclear radiation. These are here included even though they have normally been described by the term "bombardment induced conductivity" rather than photoconductivity.

A phenomenological framework for photoconductivity is needed to identify the significant parameters and to promote and guide more detailed theoretical studies of these parameters. The parameters are chiefly the bound states in the forbidden zone of a solid. A body of knowledge is needed on the chemical or physical origin of these states, their location in the energy scale and their capture cross sections for free electrons and for free holes. With these data, theories of the origin and properties of defect structures can be more profitably proposed and sifted.

One does not need to argue but need only give passing mention to the fruitful connection between phenomenological analysis and the more basic theoretical studies. It is accordingly a striking fact that photoconductivity had been known for well over half a century and that many volumes of data had been published before any consistent framework was accepted for designing experiments or interpreting data. Prior to 1945 one would have considerable difficulty in finding data on photoconductivity presented in terms of the life time of a free carrier or the capture cross section of bound states for free carriers. Yet these are the central terms defining the photoconductive process and the terms upon which more detailed theoretical studies must be based. Not only was a consistent framework lacking, but the scientific value of the bulk of the data on photoconductivity was not recognized by some leaders in the field until 1937. Prior to this period, experiments in which a stream of photons gave rise to a numerically greater stream or current of electrons were ignored as being of little scientific value ([1], p. 174). In 1937 Hilsch and Pohl [5] proposed "for the first time" (in their words) a theory for such currents. Even then, it took at least the following ten years for these so called "secondary currents" to become accepted as a common and legitimate property of photoconductors. One bit of evidence in this direction is that Pohl and Stockmann [6] published a second paper on the same subject in 1947 because little attention had been paid to their first paper.

These historical remarks are necessary to provide a reasonable perspective in the field of photoconductivity. There are good reasons why an understanding did not come early and in simple form. Photoconductivity is a structure-sensitive phenomenon. Depending on its impurity content and structural defects, a given material can, and does, show a baffling variety of behavior. Other than that radiation usually increases the conductivity of a solid, the data do not fall into any easy pattern. The photocurrent may increase as some fractional power of the light intensity, linearly with light intensity or as some higher non-integer power of the light intensity. Sometimes, the addition of a second source of radiation causes the photocurrent to actually decrease. The photocurrent may increase with increasing temperature, decrease with increasing temperature or be invariant with change in temperature. Usually the photocurrent increases linearly with the applied voltage, although in the case of powders it is likely to increase as a high power of the voltage. Sometimes, the photocurrent saturates beyond a certain voltage, and, in a few instances, it has been observed to decrease beyond some critical voltage. If one adds to this almost amorphous behavior the frequent observations of irreversible fatigue effects, it is no wonder that much of the early speculation on photoconductivity revolved around whether it was a photochemical reaction or reversible electronic process, whether it was confined to minute barrier effects between crystal grains, whether it was confined to the surface of a solid and whether water vapor was essential to the process. In brief, it is no wonder that a simple understanding of the photoconductive process has had to wait on an improved understanding of the behavior of electrons in solids as well as on improved sources of materials.

The qualification "simple understanding" deserves some emphasis. Because the behavior of photoconductors is likely to be complex, it is all the more desirable that one's method of appraisal be simple. Several references [7] are cited here to show how complex a *complete* solution of even a simple model becomes. If one has confidence that materials can be fabricated with the purity demanded by these models, the complete solutions are of course justified even though they are complex. If, on the other hand, the materials one actually deals with are more complex than these simple models,

then complete solutions are practically ruled out and one seeks a simple point of view from which to interpret the varied behavior of many and diverse materials. Most photoconductors appear to fall within the second "if."

The Characteristic Relation

The simplest and most general relation characterizing photoconductivity is:

$$n = f\tau \tag{1}$$

where n is the steady state increase in the density of free carriers generated by f excitations per second per unit volume and τ is the life time of these carriers in the free states. A corresponding relation

$$N = F\tau \tag{2}$$

may be written where N is the *total* increase in free carriers in a given specimen and F is the *total* number of excitations per second occurring in the specimen.

Eqs. (1) and (2) are logical relations rather than physical laws. They merely state that the population of a given category is proportional to the rate of influx into that category and the life time or residence time in the category. (The total population of the world for example, is the product of birth rate and average life expectancy). Eqs. (1) and (2) also state that to achieve a steady state the rate of disappearance of free carriers must be equal to their rate of generation. That is,

$$\frac{n}{\tau} = f \tag{3a}$$

and

$$\frac{N}{\tau} = F. \tag{3b}$$

Usually the rate of disappearance of free carriers is determined by recombination of the carriers into ground states. These are, by definition, states from which the carriers have to be re-excited by the incoming radiation in order to reappear in the free states. The ground states are to be contrasted with shallow trapping states from which the carriers are thermally re-excited.

Under some conditions to be discussed later the disappearance of carriers is caused by diffusion out of the specimen or by the application of a field sufficiently strong to drain the carriers out at an electrode in a time shorter then their recombination life time. Because a material tries to remain electrically neutral, it will try to draw in as many carriers from one electrode as are drawn out at another electrode. For this reason, special electrodes from which carriers cannot be drawn in are required before an applied field can effectively drain them.

Eq. (1) states the essence of photoconductivity. The additional density n of free carriers can be converted into an additional conductivity σ by the relation

$$\sigma = ne\mu \tag{4}$$

where μ is the mobility of a free carrier and e is the electronic charge. Having the conductivity produced by light (in brief, the photoconductivity), one can then by straightforward relations compute the photocurrent to be expected from a given size of sample and a given applied voltage. It is the recognition that photoconductivity is indeed a *conductivity* caused by light that allows the process to be so simply and generally defined as in (1).[2] Much of the early work by Hilsch, Gudden and Pohl in the 1920's emphasized the parallel between photoconductivity and photo-emission in a vacuum. This is a highly special case of photoconductivity in which the life time is determined by the applied field and in which special electrodes are required. This parallel was used to good advantage by the early workers to demonstrate that the excitation of electrons in a solid is a quantum process involving one excited electron per absorbed photon. The quantum nature of photoemission had been recognized at that time but not the quantum nature of photoconductivity. The photocurrents in which the current of electrons equaled numerically the current of absorbed photons were called "primary photocurrents" to differentiate them from the photocurrents which exceeded the current of absorbed photons. The latter were called "secondary currents."

While the emphasis on primary photocurrents served its purpose of identifying the quantum nature of the absorption process, it was not a happy choice for the understanding of the usual and more normal photoconductive processes. These are contained in what was called the "secondary currents."

When photoconductivity is recognized as a change in *conductivity* the question of whether the electron current exceeds the current of photons takes on no special significance. The transition from electron currents less than the photon currents to electron currents greater than the photon currents can be brought about by a continuous increase in applied voltage or a continuous decrease in electrode spacing. It is interesting, by contrast, that the increase in currents caused by heating a semiconductor has always been regarded as a simple change in conductivity. One would indeed be affronted or, at least surprised, by the question of whether the electron current exceeded the rate of thermal excitation. Yet photoconductivity happened to grow up with this question guiding much of the experimental work.

It is worth re-stating that the rate of generation, whether it is photo or thermal, does become important for the current-voltage curve when the electrodes are so chosen that carriers cannot be brought in to the sample fast enough to maintain a constant density of carriers or a constant conductivity. Photo-emission

[2] Only those cases are excluded from (1) in which the electric field is high enough to allow free electrons to be multiplied by collision ionization. The fields required are of course very high and near the critical field for breakdown. Photoconductors normally are not operated in this range. The work of K. G. McKay, "Avalanche breakdown in silicon," *Phys. Rev.*, vol. 94, p. 1437; 1954, is a good example of controlled multiplication in a *p-n* junction.

in a vacuum diode, temperature-limited electron emission, photo-currents in the back direction of a *p-n* junction or a rectifier and the saturated back current in these same structures are all examples of such electrodes. As the applied voltage is increased in the range in which the currents are saturated, the density of carriers in the space between electrodes decreases inversely with increase in field. The product of field and conductivity remains constant and equal to the saturated current.

Concerning τ, the Life Time of a Free Carrier

To return to (1), a knowledge of τ the lifetime of a free carrier almost completely characterizes a photoconductor. The choice of rate of excitation is of course arbitrary and the choice of sample dimensions and applied voltage are similarly incidental to the photoconductive process itself. Even the forbidden gap energy may be regarded as an incidental parameter chosen more or less freely to give the desired spectral response. Only the life time of a free carrier describes the essential performance and contains the essential physics of a photoconductor.

The life time is determined either by the rate of recombination of carriers with their ground states or by the rate at which they are extracted by an applied field. For the most part, this discussion will be concerned with recombination lifetimes since they are the more commonly encountered. The details of the recombination process will be described more fully in the section on recombination. At this point it is sufficient to state that recombination takes place via bound states in the forbidden zone.[3] These bound states are composed of impurities, vacant lattice cites, intersticial atoms and crystal defects. The details of the recombination process may be exceedingly complex and varied. It is normally not easy to determine the recombination time given the state structure in the forbidden zone, and, conversely, it is normally impossible to define the state structure given the recombination time. What can be determined are certain patterns of behavior to be expected from certain energy distributions of states in the forbidden zone [8, 9].

It almost goes without saying that the recombination life time is not a material constant but varies with the mode of preparation and impurity content of the material. Even for a given material prepared in a given way, the life time is not a constant. If it were, the photocurrent would increase linearly with increasing light intensity and would be temperature independent. Neither is normally true [10–12]. The lifetime usually decreases with increasing light intensity giving rise to photocurrents that increase as a fractional power of the light intensity. In some cases the life time actually increases with increasing light intensity so that the photocurrent varies as a power greater than unity of the light intensity. This is called superlinearity. The variation of life time with temperature depends on the temperature range as well as on the light intensity [12, 13]. It is equally common to find photocurrents increasing, decreasing or remaining invariant with temperature.

[3] Recombination of free electrons with free holes becomes significant only at very high densities of both signs of carriers.

Some Orders of Magnitude

Since the essence of photoconductivity is contained in the life time of a free carrier it is instructive to estimate what reasonable range of values might be expected The recombination life time is given by an expression of the form:

$$\tau = \frac{1}{vsn} \tag{5}$$

where v is the thermal velocity of a free carrier, s the capture cross section of the capturing center and n is the number of these centers. Descriptively, if one attaches the capture cross section to the free carrier, τ is the time required for the free carrier to trace out a volume of space equal to the reciprocal density of capturing centers. At room temperature, v is 10^7 cm/sec. and varies only as the square root of the absolute temperature. The largest value of the capture cross section for a carrier having a thermal energy of kT is about 10^{-13} cm^2 [8, 14]. This is based on the distance that a coulomb field of a charged capturing center will reach out to depress space potential kT below its surrounding value. When the free carrier drops kT below the "lip" of the coulomb field it is likely, especially in low mobility materials, to be drawn in the rest of the way to the capturing center. The smallest values reported for capture cross sections lie around 10^{-22} cm^2 [8, 10, 15]. If one regards the small values of capture cross sections to be caused by a repulsive coulomb field around a capturing center, still smaller values should be possible especially at low temperatures. A commonly reported value for capture cross section is 10^{-15} cm^2 and is the magnitude to be expected from the physical dimension of an atom or ion.

The number of capturing centers is likely to be of the same order as the number of bound states in the forbidden zone. Values as low as 10^{11}/cm^3 have been reported for highly purified single crystals of germanium [16]. An upper limit for the density of bound states may be taken as approximately 10^{19}/cm^3. State densities higher than this are likely to lose their "bound" character and begin to form significant band structure. Bound state densities as high as 10^{19}/cm^3 can be expected in polycrystalline and amorphous materials even though they may be chemically pure. Evaporated layers are a common example. Since 10^{19}/cm^3 is still only one atom site out of 10^4 it is not unreasonable to expect chemical impurities also to supply such densities. An estimate of bound state densities in purified, single crystals is still likely to be as high as 10^{14} to 10^{16}/cm^3. Such values have been observed for example in single crystals of CdS [11, 17].

To summarize, capture cross sections extend from 10^{-13} cm² to 10^{-22}cm² with likely values around 10^{-15} cm²; densities of capturing centers extend from 10^{11} to 10^{19}/cm³ with likely values in single crystals around 10^{15}/cm³. If these values are combined in (5) to give range of life times, one gets life times as long as 10^4 and as short as 10^{-12} seconds. Median values lie around 10^{-7} seconds.

Evidence for lifetimes of minutes or hours has been reported by Broser and Warminsky [7] for highly conducting crystals of CdS at room temperature, and by Haynes and Hornbeck [18] for silicon. P. K. Weimer in some unpublished measurements on CdS crystals at liquid air temperature has observed *large* photocurrents that persisted for hours. (It is now common experience to expect *small* photocurrents in insulating photoconductors to persist for hours or days even at room temperature [4, 10, 17]. These long time effects are believed not to be a direct measure of carrier life times but rather of the slow emptying of traps [8]. This effect will be discussed in more detail in the sections on recombination processes.)

Lifetimes of 10^{-10} to 10^{-13} seconds are needed to account for the small photocurrents observed in many evaporated insulating materials. These currents are often so small that normal measuring techniques are likely to overlook them and classify the materials as non-photoconductors. Many evaporated metal sulphides, selenides and oxides examined for operation in the Vidicon type of television pick up tube are in this category [19]. Amorphous selenium, for example, that was classified by Gudden [1] as a non-photoconducting insulator was found by Weimer [2] to give electron currents approaching the incident photon currents when used in the Vidicon. The Vidicon, by virtue of its small electrode spacing or target thickness (10^{-3} cms), can easily detect photocurrents in materials with lifetimes of 10^{-8} to 10^{-13} seconds. Life times of 10^{-10} to 10^{-13} seconds are needed to fit the data of Pensak [3] and Anspacher [20] on bombardment induced conductivity in thin evaporated insulting films of SiO_2, MgF, As_2O_3 and As_2S_3. Finally it is interesting to note the observations of photoconductivity reported by Farmer [4] in polystyrene excited by x-rays. These results are noteworthy because the results obtained in an organic compound are no different operationally from those obtained in inorganic single crystals. The small conductivity (10^{-13} ohm^{-1} cm^{-1}) produced by the x-rays is consistent with the very short life times computed above and would normally be overlooked. On the other hand the conductivity took hours to decay when the x-rays were turned off. This is an exaggerated case of the trap-provoked disparity between observed speed of response and life time to be discussed below.

The median life times of 10^{-7} seconds are frequently observed in not too well crystallized or purified samples of germanium and silicon used in transistor studies.

Life times in some commonly known sensitive photoconductors range from 10^{-3}–10^{-5} seconds in PbS [21] and Tl_2S [22] to 10^{-2}- 10^{-3} seconds in CdS [10, 15] and CdSe [12, 23].

A Second Characteristic Relation

From (1) and Ohm's law, one may obtain the following simple expression for the photocurrent in a given photoconductor [8, 22, 24]:

$$I = eF\frac{\tau}{T_r}\text{ amperes} \tag{6}$$

where T_r is the transit time of a carrier between electrodes and is given by:

$$T_r = \frac{L}{E\mu} = \frac{L^2}{V\mu} \tag{7}$$

where L is the electrode spacing in cm, E and V are the field and applied voltage respectively in volts/cm and volts and μ is the mobility of a free carrier in cm²/volt sec. One may define a gain factor

$$G = \frac{\tau}{T_r}$$

as the number of electron charges passed through the photoconductor per absorbed photon (or per excitation). The term "gain" avoids the ever present ambiguity of the term "quantum efficiency." The latter term should be confined to the primary excitation process and should refer to the ratio of excitations to incident photons. It is seen here immediately that the gain can range from values below unity to values above unity without any abrupt or even definable change in the physics of the process. The only implied condition for gains greater than unity is that the electrodes are able to supply carriers freely to the photoconductor as they are needed. This is another way of defining an ohmic contact. But an ohmic contact is needed merely to observe ohms law even in the absence of light.

Equation (6) Applied to Barriers

Eq. (6) could have been written out of hand as a logical relation without reference to ohms law. It says quite generally that the current observed between two points in a circuit is equal to the rate of generating carriers multiplied by the ratio of their life time in the region between the two points to their transit time between these points. By life time is meant the statistical life time of the generated carriers. That is, if the originally generated carriers move out of the space between electrodes and others take their place, the carriers are considered to be still alive. The life time of a carrier is terminated when it moves out of the space between electrodes without being replaced by conduction from the opposite electrode or when it recombines with a deep-lying bound state from which it has to be re-excited by the incident radiation in order to appear as a free carrier again.

Because (6) is not based on any special model one might expect it to apply not only to a uniform photoconductor but also to rectifying junctions (metal-semiconductor and *p-n*) operated in the back direction, to *p-n-p* or *n-p-n* junctions, to insulating or high resistance barriers both thick and thin and even to vacuum photocells. And so it does. In the case of a vacuum photocell, for example, the life time of a free carrier (photo electron) is identically equal to its transit time from cathode to anode since it is not replaced at the cathode when it leaves at the anode. The gain remains identically unity independent of the applied voltage—that is, one electron charge passed between electrodes for each excitation at the cathode. The same argument may be applied to rectifying junctions operated in the back direction. Because the carriers are not replaced, the gain is identically unity. A caution must be observed here, however. In the case of the vacuum photocell the electron was excited at the cathode; the hole it left behind was already at the cathode and did not have to migrate there. In the case of a rectifying barrier, such as a metal-semi-conductor barrier, the electron-hole pair can be generated either in the volume of the barrier or at either electrode. If the pair is generated in the volume of the barrier *both* electron and hole must be able to leave the barrier at the electrodes in order to avoid a continuous building up of charge in the barrier. If the pair is generated at either electrode, only one of the carriers need cross the barrier. These considerations become important when, as will be discussed later, the life times of the electron and hole are widely different so that the electron lifetime, for example, may be sufficient for it to be drawn out of the barrier but the hole life time too short. In this case, no photocurrent is observed from pairs generated in the volume of the barrier but only from those generated at or near the cathode where the hole can easily escape.

The case of the *n-p-n* barrier is an interesting one in several respects. The photoelectric gain one computes for pairs generated in the *p*-section (or for pairs generated elsewhere and diffusing to the *p*-section) is also the ratio of collector current to base current called α in transistor terminology. Given an electron-hole pair in the *p*-section, G electrons will flow through the *p*-section in the time it takes the hole to diffuse out of the *p*-section. Here G is the photoelectric gain defined by the ratio of the life time of the extra electron to its transit time through the *p*-section. The life time of the extra electron is determined through the charge neutrality condition by the life time of the extra hole. But the hole is partially trapped in the potential well formed by the *p*-section. It escapes only slowly into the emitter. The electron transit time on the other hand is quite short. It is determined by the diffusion time of an electron across the *p*-section. Reasonable values for these times are 10^{-8} seconds for the electron transit time and 10^{-6} seconds for the hole transit time (and, therefore, also for the life time of the extra electron). These values give a photoelectric gain of 10^2. Gains as high as 10^3 have been observed. For the above argument to hold, the recombination life time of the electron (this is also the life time of a free pair in the *p*-section) must be greater than 10^{-6} seconds.

A final comment on the *n-p-n* junction is in order. At applied voltages greater than a few kT the photocurrent is saturated. Higher voltages add neither more current nor more speed of response.[4] In the case of a vacuum photocell or a rectifying junction the photocurrent is also saturated at a low voltage. Higher voltages, however, do improve the speed of response by reducing the transit time (life time) of the free carriers. In the case of the usual uniform photoconductor for which the recombination life time is the controlling time, higher voltages result in proportionately higher currents with no change in speed of response. In summary, the performance of an *n-p-n* junction reaches its maximum at an applied voltage of a few kT while the performance of other photoconductors continues to increase with increasing voltage.

It is immediately clear that (6) is applicable to an insulating or high resistance barrier between two conducting sections of semi-conductor by considering the barrier thickness to be the electrode spacing since substantially the full applied voltage is across the barrier. One is now primarily concerned with excitations in the barrier, and with carrier life times and transit times appropriate to the barrier. Excitations occurring outside the barrier are wasted unless the electron-hole pair can diffuse to the barrier. In materials other than germanium or silicon, this diffusion distance is likely to be a micron or less owing to the short *pair* life times.

The suggestion has frequently been made that high resistance barriers control the photoconductivity of evaporated films of lead sulphide. It is pointed out in the section on noise that if such barriers are present in lead sulphide, the diffusion of *all* generated pairs to the barriers is required in order to satisfy the observed signal to noise ratios.

This section may be summarized by saying that (6) is equally valid for uniform photoconductors, for any of the commonly known types of barriers and for vacuum photocells. Barriers do not offer any improvement in sensitivity or speed of response over what may be obtained with a simple uniform photoconductor. On the contrary, barriers are likely to give less performance than a uniform photoconductor by not being able to make full use of higher voltages, as in an *n-p-n* junction, or by not making use of all of the incident light, as in an insulating barrier where light absorbed further than a diffusion length from the barrier is not effective. While barriers offer no fundamental advantage over a uniform photoconductor, they may in the present undeveloped state of the art offer temporary technical advantages.

[4] A slow improvement in speed and sensitivity can result from the decrease in width of the *p*-section at higher voltages.

WRH

One cannot yet prepare uniform photoconductors at will to cover all of the ranges of life times and spectral responses desired.

Some Remarks on Recombination Processes

Detailed discussions of free carrier life times as determined by recombination have been presented [8, 9]. Those discussions will not be repeated here. What will be outlined are the methods of analysis, the reasons for choosing such a method and some of the results.

It was pointed out earlier in a footnote[3] that the recombination of a free electron with a free hole was an improbable occurrence. Since both free electron and free hole densities increase with increasing light intensity, the total rate of such recombinations increases as the square of the carrier density. At high carrier densities, then, the direct recombination of free pairs can become comparable with the recombination via bound states since the latter at very low or very high lights tends towards a linear dependence on carrier densities. Quantitatively an upper limit for the life time as determined by recombination of free pairs can be computed from detailed balancing between the rates of recombination of free carriers and Planck's radiation formula.

Consider radiation whose $h\nu$ lies within kT of the two edges of the forbidden zone. This radiation will be absorbed within some depth d at the surface of a material. The number of recombinations between free pairs within this depth d must at least be equal to the number of "edge emission" photons radiated according to Planck's law. Higher rates of recombination would result from the addition of "radiationless" recombinations. This type of analysis was carried out by Roosebroeck and Shockley [25] to obtain a capture cross section of $2.9\times 10^{-21}/\text{cm}^2$ for a free hole capturing a free electron in germanium. More generally the capture cross section for photon recombinations in a slice kT wide near the band edges can be written

$$s = \frac{1\cdot 1\times 10^{-23}}{d} \cdot \left(\frac{k\nu}{h\nu_0}\right)^2 \left(\frac{300}{T}\right)^{\frac{5}{2}} cm^2 \tag{8}$$

where

$h\nu_0$ = 1 electron volt

T = absolute temperature

$h\nu$ = width of the forbidden gap.

This form is chosen for ease of computation. The capture cross section increases as the square of the forbidden gap energy $h\nu$ and as the 2.5 power of the reciprocal of the absolute temperature.

The life time of a free pair in the highly purified and crystallized transistor materials, germanium and silicon, may range as high as 100 to 1,000 microseconds. In no other material have free pair life times been reliably reported as large as a microsecond. These pair life times are determined by recombination via bound states in the forbidden zone. If direct photon-emitting recombination of free carriers is to compete with bound state recombination, the free carrier densities would need to be in the order of $10^{17}/\text{cm}^3$ to achieve a life time of 10^{-4} seconds and $10^{19}/\text{cm}^3$ for a life time of 10^{-6} secs. The latter density is the maximum that can be contained within kT of the band edges. In the remainder of the discussion, the direct recombination of free carriers will be ignored as probably negligible for the carrier concentrations normally encountered.

If one ignores the direct recombination of free carriers, the simplest alternative is to introduce a single class of bound states all lying at the same energy level. A class of bound states is characterized by the capture cross sections of its members for free electrons and for free holes. This problem has been treated by Hall [26], by Shockley and Read [7], and by Isay [7]. It is instructive to look at the last of these three references to get a sense of proportion of the complexity of this problem. Isay obtains a complete solution of the problem without approximations. The result is expressed as an eighth degree polynomial occupying most of one page. The other authors break the problem down into separate ranges of conditions within which simplifying approximations can be made. Even so, the complexity is still evident.

What has just been described are the solutions to the simplest of the bound state recombination problems—a single class of states at a single level. It is possible, though not yet established, that germanium can be prepared sufficiently free from impurities and defects to satisfy this model.[5] The writer has little hesitation in believing that all other materials need a more complex model to match their recombination properties. The more complex model will have more than one class of bound states and more than one level (more likely, a quasi continuous distribution of levels) at which these states are found.

This preamble has been inserted to emphasize the need for a point of view that would allow relatively rapid qualitative appraisals to be made of the fitness of various models and, with some additional effort, even order of magnitude checks. The point of view and method to be described here has, in many respects, been used independently by Broser and Warminsky [28] in their more recent papers and by Schon [29] with more emphasis on luminescence.

Some passing comments on the meanings of life time are brought in here to help clarify the later discussion.

Some Meanings of Life Time

The literature on transistors has come to dominate so much of the thinking on semi-conductors that some of the properties of the less tractable materials (meaning all semi-conductors not called germanium or silicon) tend to be ignored. The subject of life times is a good example.

[5] For the variety of bound states found in germanium see bibliographical reference [27] and F. J. Morin et al., *Minneapolis APS Mtg.*, June 28, 1954, Paper R-5.

In transistor writings the term life time is customarily a short form for "life time of a free pair." Now, by definition, the life time of a free pair is terminated when either member of the pair is captured in a bound state. It is possible, though not likely, that the majority sign of carrier will be captured first. It is much more likely that the minority sign of carrier will be the first to be captured. For this reason "minority carrier life time" is another common phrase in transistor writings.

When the minority sign of carrier is captured, its opposite number amongst the majority carriers may or may *not* be captured within the same time. If the densities of both the free electrons and the free holes is large compared with the density of the bound states through which recombination can take place, then, by reasons of logic rather than physics, the life time of the electron, the life time of the hole and the life time of the pair are one and the same life time. This is the usual situation in transistors. If, on the other hand, the density of either the free electrons or the free holes is less than the density of the bound states through which recombination can take place, the life time of the free electron and the life time of the free hole are independent parameters and likely to be several or more orders of magnitude apart. The life time of the free pair is by definition the shorter of the two free carrier life times. What has just been described is the usual situation for sensitive photoconductors other than transistor-quality germanium and silicon. An example is a sensitive crystal of CdS in which the electron life time has been measured to be 10^{-3} seconds and the hole life time estimated (but not yet measured) to be less than 10^{-10} seconds.

If the electron and hole life times are not equal, the photoconductive effect measures the *longer* of the two life times while the PEM effect [30], and other phenomena that depend on the diffusion of free pairs, measures the *shorter* of the two life times. In the usual photoconductor this discrepancy is many powers of ten; it can even show up in measurements on not too well purified germanium and silicon at low or moderate light intensities.

Another aspect of life time is mentioned here and will be repeated later. If, for example an electron is excited into the conduction band, then beomes temporarily trapped, then is thermally re-excited into the conduction band and the cycle repeated many times before it is finally captured by a deep lying bound state from which thermal excitation is less likely than capture of a free hole, one can identify a total time of excitation. Part of this time is spent in the conduction band, part is spent in traps. *Only the time spent in the conduction band is counted as the life time of the free electron.*

A Point of View

In Fig. 1 is shown a quasi-continuous distribution of bound states in the forbidden zone of a solid. Free electrons and free holes are being generated at identical rates by light absorbed by the host crystal.[6] The free electrons and the free holes recombine via the bound states. The free electrons pour into the empty bound states at the same rate as they are generated. The free holes similarly pour into the filled bound states at the same rate as they are generated. The terms "empty" and "filled" mean "not occupied by an electron" or "occupied by an electron." Empty bound states that capture an electron are then ready to capture a hole and so on.

Direct transitions of electrons and holes between the bound states is assumed to be negligible because the bound states are physically remote from each other.[7] Exchange of electrons and holes between bound states is possible only through the intermediary of the free states.

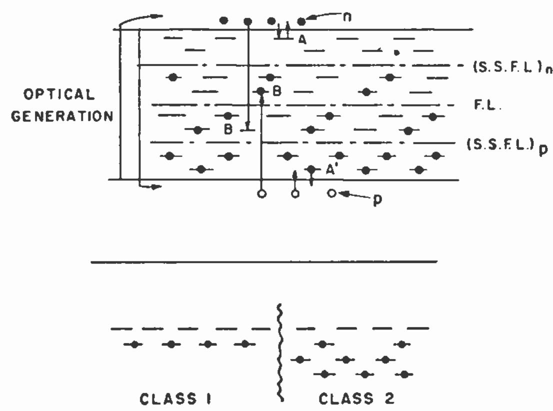


Fig. 1—Continuum of bound states in the forbidden zone showing the distinction between shallow trapping states and ground states. Recombination takes place through the ground states.

Consider the life history of an excited electron. Once in the conduction band it wanders around with thermal velocity, colliding frequently (about every 10^{-13} seconds) with the lattice. After a number of these lattice collisions the electron is likely to be captured by one of the empty bound states. Let the bound state lie close to the conduction band. In a relatively short time, depending on the depth of the bound state, the electron will be thermally re-excited into the conduction band. After a number of lattice collisions it will again be captured by an empty bound state. Let the bound state this time lie about a volt below the conduction band. From this depth, it would take at least some hours at room temperature before the electron would be thermally re-excited into the conduction band. Long before

[6] If the electrons are not excited from the filled band but from bound states very close to the filled band, a free hole density will still be built up by thermal excitation of the holes from the bound states into the filled band. Much of the same physical processes will take place as are discussed for excitation from the filled band. If the electrons are excited out of bound states too far from the filled band to permit thermal exchange, these bound states then become by definition the ground states for determining the life time of free electrons and the problem reduces to that discussed extensively in bibliographical reference [8].

[7] At sufficiently high densities of bound states, tunneling between bound states needs to be considered. See for example, D. L. Dexter, "X-ray coloration of alkali halides," *Phys. Rev.*, vol. 93, p. 985; 1954.

thermal re-excitation occurs, the electron in the deep lying bound state is likely to capture a free hole. This event definitely terminates the life history of the excited electron. A new optical excitation is required to return it to the conduction band.[8] Of the total life history, only the time spent in the conduction band is counted as the life time of a free electron—the central parameter that determines the photosensitivity of the material.

Several features of this life history are significant. The two types of experience after being captured by bound states suggest that the bound states might be divided into two classes: bound states near the conduction band from which an electron is more likely to be thermally re-excited into the conduction band than to capture a free hole, and bound states more remote from the conduction band in which an electron is more likely to capture a free hole (and terminate its life history) than to be thermally re-excited into the conduction band. The bound states near the conduction band will be called "shallow trapping states"; the bound states more remote from the conduction band will be called "ground states." The connotation of "ground state" is that the optically excited electron has completed its free-life history when it is captured by a ground state.

The occupation of the shallow trapping states by electrons or holes is determined by thermal exchange with free carriers at their nearest band edge. These states tend to be in thermal equilibrium with their free carriers. The occupancy of ground states, on the other hand, is determined almost entirely by the kinetics of recombination.

The precise demarcation line between shallow trapping and ground states will be discussed in a later section. While the energy depth below the conduction band is the most powerful parameter in making this determination, other parameters such as the capture cross sections of the bound states, the number of free and empty bound states and the number of free electrons and holes also enter in. Even without a precise determination of this demarcation line, a number of important qualitative conclusions can be drawn.

The demarcation line frequently lies close enough to the steady-state Fermi-level to allow a qualitative discussion to be carried out in terms of the latter. The steady-state Fermi-level is defined by the relation:

$$n = N_c e^{-E_{fn}/kT}$$

where n is the free electron density, N_c the number of effective states in the conduction band ($\sim 10^{19}/\text{cm}^3$ at room temperature) and E_{fn} is the distance from the conduction band to the steady-state Fermi-level for electrons. Similarly, a steady-state Fermi-level for holes E_{fp}, measured from the filled band, can be defined in terms of the free hole density. Other reasons for using the steady-state Fermi-level in these preliminary discussions are that it has an immediate connotative value in terms of the carrier densities and that its dependence on light intensity and temperature strictly parallels the dependence of the demarcation lines on these parameters when a single class of bound states is involved.

[8] The language here is a bit free since an electron in the filled band loses its identity.

Under thermal equilibrium conditions in the "dark" the two steady-state Fermi-levels may be thought of as coincident with the normal thermal equilibrium position of the Fermi-level. When the material is exposed to radiation, the two levels E_{fn} and E_{fp} move apart toward the conduction and filled bands respectively.

Returning to Fig. 1 and recalling the life history of an optically excited free carrier, one may conclude that the recombination of free electrons and holes is determined by the bound states that lie between the two steady-state Fermi-levels E_{fn} and E_{fp}. These states have been called ground states. This means that the life times of the free-electrons and free holes are determined by the ground states. The ground states also determine, through the life times by (6), the photocurrents. The shallow trapping states lying outside the steady-state Fermi levels do not effect significantly the recombination of free carriers and therefore do not affect the *steady-state* photocurrents.

Observed Rise and Decay Times

While the shallow trapping states do not affect the steady-state photocurrents they do affect the time required to set up the steady-state photocurrents and the time required for these photocurrents to decay when the optical excitation is interrupted. The rise time of the photocurrents is increased because in order to increase the density of free carriers the density of carriers in the shallow trapping states must be increased in the same proportion. The rise time is thus increased by the ratio of shallow-trapped to free carriers. Similarly when the optical excitation is interrupted, the shallow trapped carriers must be emptied into the ground states via the free states. The decay time, then, will also be extended by the ratio of shallow-trapped to free carriers. The observed rise time or decay time of the photocurrents of a photoconductor are thus not necessarily the same as the life time of a free carrier. Only at high lights, when the densities of free carriers are equal to or greater than the density of shallow trapped carriers will the response time of a photoconductor approach the life time of a free carrier. These densities are likely to be $10^{15}/\text{cm}^3$ and greater. At low light intensities and for photoconductive insulators the free carrier densities may be as low as $10^7/\text{cm}^3$. Under these conditions, if the shallow trapped carrier density is $10^{15}/\text{cm}^3$, the rise and decay time of the photocurrents will be 10^8 times larger than the life time of a free carrier. Such slow response times have been observed in many evaporated photoconductors [19] used in the Vidicon. Response times in these materials approached a second while the life times of free carriers determined from (6) were of the order of 10^{-8} seconds. In the case of the observations

of photoconductivity in polystyrene [4] the free carrier density was of the order of $10^5/cm^3$. The observed time of response was about an hour and must have been at least 10^{13} times longer than the life time of the free carriers.

It is this strikingly large disparity between the observed rise or decay times of photocurrents on the one hand and the life time of free carrier on the other hand that has probably discouraged in the past the simple interpretation of photocurrents according to (6). One's instinctive reaction is to regard the rise or decay time of the photocurrent as a direct measure of life time and to insert these observed rise or decay times in (6). The resultant discrepancies of many powers of ten would indeed make one suspect the validity of this relation. A major contribution of much of the work on photoconductive pick up tubes of the Vidicon type was a forced recognition that the life time of a free carrier and the decay time of a photocurrent were different times. This has paved the way to the confident use of (6) as a simple interpretation of photoconductive currents.

Sub-Linearity and Super-Linearity

It was pointed out that frequently the photo-current increases as a fractional power of the light intensity, the fraction lying between 0.5 and 1.0. Such a dependence on light intensities means that the life time of a free carrier becomes shorter at higher light intensities. Fig. 1 gives an immediate picture of how this occurs. As the light intensity is increased, the steady-state Fermi-levels, E_{fn} and E_{fp}, pull apart toward their respective band edges. As they pull apart they embrace more ground states. More ground states means that the carrier life times will be shorter. The particular fractional power dependence on light intensity depends on the energy distribution of the bound states. A uniform distribution of bound states in energy leads to powers close to unity. Powers between 0.5 and unity are obtained from bound states distributions that decrease rapidly with distance from the band edges [8].

The above remarks are true when the bound states are all of one class or predominantly so. Being of one class means that of all the bound states have the same capture cross sections for free carriers. If there are significant numbers of bound states having different capture cross sections, other dependencies of photocurrent on light intensity are possible [9, 12]. In particular, superlinearity becomes readily understandable. Let it be assumed that the addition of a second class of bound states can make a photoconductor more sensitive—that is, increase the life time of one sign of free carrier. If, at low lights, this second class of states lies outside or partially outside the two steady-state Fermi-levels (see Fig. 2), it will have little effect on the photosensitivity already determined by the first class of states lying within the two Fermi-levels. As the light intensity is increased, the steady-state Fermi-levels pull apart and embrace more and more of the second or sensitizing class of states. In this way the photoconductor becomes more sensitive as the light intensity is increased—in short, the current-light curve is super-linear.

A brief discussion of the way in which a second class of bound states can sensitize a photoconductor is in order at this point.

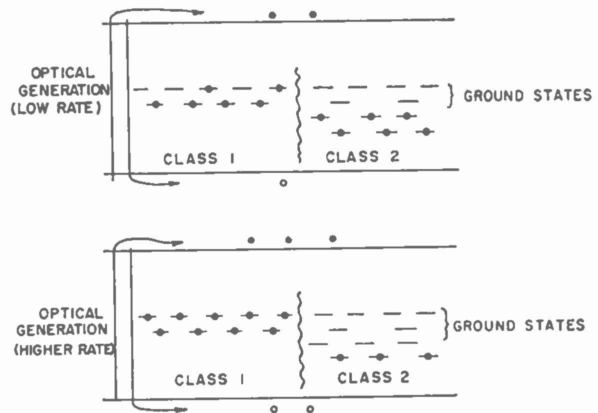


Fig. 2.—Two classes of ground states showing qualitatively the origin of superlinearity.

Sensitization and Saturation

Sensitizing a photoconductor means by (6) increasing the life time of a free carrier. The life time of a free carrier is determined by recombination with a ground state. A photoconductor, in practice, can be sensitized by the addition of chemical impurities (ground states in the present terminology). It requires explanation to show how addition of recombination centers (ground states) can actually slow down recombination of a free carrier and thereby sensitize photoconductor.

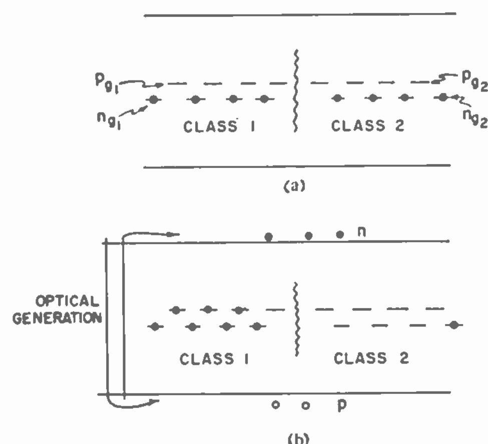


Fig. 3.—Two classes of ground states showing qualitatively the origin of sensitization.

Fig. 3 shows schematically two classes of bound states in the forbidden zone of a photoconductor. The two classes are shown separately here; actually. they are physically interspersed. Class I , *by itself*, is such as to make the photoconductor fast and insensitive. To be specific, let the capture cross sections of the states in class I be 10^{-15} cm^2 both for free electrons and for free holes. The class II states are assumed to have a capture

cross section of 10^{-20} cm^2 for free electrons and 10^{-15} cm^2 for free holes. In the dark, both class I and class II states are approximately half filled with electrons. When free carriers are generated, the occupation numbers of the ground states change radically. For example, free holes are captured at about the same rate by both class I and class II states. But whereas the holes in class I states rapidly capture free electrons, those in class II states do not, owing to their small capture cross section for electrons. The result is that the class II states tend to become filled with holes, the electrons from the class II states being transferred to class I states. After this redistribution, a free-electron has difficulty in being captured by class I states because they are already nearly filled with electrons and there are few empty places; it also has difficulty in being captured by class II states because these have a small capture cross section for electrons. The final result is that the life time of the free electron has been increased some 10^5 times by the addition of class II states. The life time of the free holes, meantime, has been decreased by a factor of two by the addition of class II states.

The transition from the distribution in Fig. 3(a) to that in Fig. 3(b) requires time. This time is accentuated if the photoconductor has been in the dark long enough to insure the thermal equilibrium distribution of Fig. 3(a) and if only a small amount of excitation is introduced. A long latent period is observed during which the transfer is taking place and the photocurrents are quite low. Near the end of this period the photocurrents rise rapidly. Frerichs [30a] reports data of this form using γ-radiation on CdS crystals.

If the class II states had been close to the filled band they would have had little effect on photosensitivity at low light since they would be outside the steady-state Fermi-levels. At high lights they would be incorporated into the ground states and would sensitize the photoconductor in the manner just described. One would then expect a low-light linear current-light curve, a super-linear range at intermediate lights followed by another linear section at higher light intensities.

At still higher light intensities, the free electron density becomes comparable with the density of ground states and holes begin to re-appear in the class I states. In this range the current vs light curve will appear almost to "saturate" or "flatten off." In the limit of very high light intensities (carrier densities larger than ground state densities) the sensitivity of the photoconductor will recede by a factor of 10^5 approximately to the sensitivity determined by the class I states alone. The "flattening off" of highly sensitized photoconductors at very high light intensities has been noted frequently and informally by S. V. Forgue, R. W. Smith and R. H. Bube of these laboratories.

Infra Red Quenching

Photocurrents generated by visible light may often be "quenched" or reduced by the addition of longer wavelengths of light. This is an old observation. Some recent quantitative results for infra red quenching in CdS crystals were reported by Taft and Hebb [31]. There it is shown that the same infra red has both a positive and a negative effect, the negative effect having the longer time constant.

The quenching effect of infra red can be understood from Fig. 3 as a shifting of the favorable distribution of electrons and holes in the ground states, shown in Fig. 3(b), to the unfavorable distribution shown in Fig. 3(a). This is done by exciting electrons from the filled band into class II states. This shift in distribution is likely to be a relatively slow process simply because of the large numbers of ground states. At the same time the infra red, by exciting some electrons from ground states of class I into the conduction band can show a more rapid positive contribution to the current. As Taft and Hebb [31] show, the final steady state effect depends on the wave length of the quenching light.

Temperature Dependence of Photoconductivity

Photocurrents have been observed both to increase and to decrease with temperature depending on the temperature range and light intensity [13]. From (5) the temperature can be expected to influence the life time of a free carrier through the capture cross sections and through the number of ground states. (The temperature variation of thermal velocity is too weak to account for most observations). If the very small capture cross sections in sensitive photoconductors are a result of a repulsive coulomb field around the capturing center, an increase in temperature can be expected to increase the capture cross section and decrease the photocurrents.[9]

Since the distance of the steady-state Fermi-levels from their respective band edges is proportional to the absolute temperature, higher temperatures tend to squeeze the two levels together. If there were predominantly one class of bound states, this squeezing together would reduce the number of ground states and result in an increase in photocurrent with increasing temperature. On the other hand, when there is more than one class of bound states, the behavior can become more varied since a change in temperature can shift a class of bound states in or out of the ground state category [12]. In particular, if one has a super-linear photoconductor, its photosensitivity should decrease with increasing temperature. The increasing radiation is pulling the two Fermi-levels apart to include more of the sensitive states; the increasing temperature would squeeze the two levels together to exclude the sensitive states. CdSe crystals, for example, are super-linear at room temperature and their photocurrent decreases rapidly with increasing temperature [10, 12, 23].

[9] In a recent conversation with W. Schottky he pointed out that the temperature dependence of capture cross section could be avoided by allowing the electron to tunnel through the potential barrier.

In general the interpretation of the temperature dependence of photocurrents in a given photoconductor is hazardous because more than one parameter can change with temperature.

Current-Light Curves in the Low, Intermediate and High-Light Ranges

The low-light range is defined by the conditions that both free carrier densities are less than the densities of electrons and holes in the ground states. Under these conditions, the life times and, by (1), also the densities of free electrons and free holes are independent. If there is predominantly one class of ground states, the free carrier densities increase as a fractional power of the light intensity, the fraction lying between 0.5 and 1.0. The spreading of the two steady-state Fermi-levels to include more ground states at higher light intensities accounts for the less than linear rise of photocurrent with light intensity.

If the photoconductor has been sensitized by the addition of other classes of bound states, two consequences are likely. First, the density of one sign of carrier will be several or more powers of ten higher than the density of the other sign of carrier. The addition of sensitizing states can increase the life time of one sign of carrier, but not of both signs of carriers. Second, by the mechanism already described the current-light curve may now be super-linear.

To summarize, the photocurrent in the low-light range of a sensitized photoconductor will be that of one sign of carrier and will increase as a power of the light intensity greater than 0.5.

The high-light range is defined by the conditions that both free carrier densities are greater than the densities of ground states. In the high light range, the life times and densities of both signs of free carrier are equal. The equal densities follow logically from the need to maintain charge neutrality, and the equal life times follow from the equal densities and (1).

A significant characteristic of the high light range is that the addition of ground states of any character can have only one effect, namely, to *shorten* the life times of *both* signs of carriers. This is to be contrasted with the low-light range in which some types of ground states may be added which can increase the life time of one sign of carrier while at the same time, decreasing the life time of the other sign of carrier. In the low-light range, the ground states can interact in the sense of exchanging electrons and holes, via the free states, as already described. In the high-light range, all ground states act independently of each other. Each ground state is occupied $s_n/(s_n+s_p)$ of the time by an electron and $s_p/(s_n+s_p)$ of the time by a hole. s_n and s_p are the capture cross sections of the ground state for free electrons and holes respectively. The addition of new ground states does not affect the occupancy of the already present ground states but simply adds more recombination paths for both electrons and holes.

Because the same ground states that were added in the low-light range to sensitize the photoconductor actually desensitized the photoconductor in the high-light range, the sensitivity of the photoconductor in the high-light range will be many powers of ten lower than its sensitivity in the low-light range. The current-light curve in the high-light range is strictly linear if the direct recombination of free pairs is neglected and if the capture process itself occupies a negligible time.

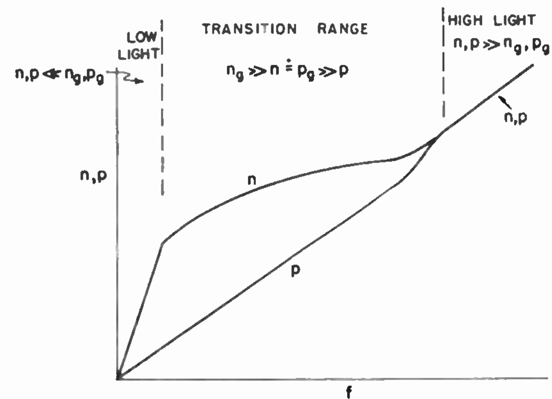


Fig. 4.—Carrier densities (and photocurrents) in the low intermediate and high-light ranges.

In Fig. 4 is shown schematically the carrier densities, or their corresponding photocurrents, in the low, intermediate and high-light ranges. The example chosen is one for which the free electron life time is much larger than the free hole life time in the low-light range. Since the free electron density matches the free hole density in the high light range, there must be an intermediate range in which the free electron density curve bends down to meet the free hole density curve. If there is predominantly one class of ground states, the curvature of the free electron curve in the intermediate range is approximately a half power. Such behavior has been observed for CdS crystals by Smith [10] and by Fassbender [32]. If the photoconductor has been sensitized by the addition of other classes of ground states, the free electron density curve in the intermediate range may indeed have such a low power as to appear completely saturated or flat. The sensitizing effect of the added ground states is being "undone" in this range by decoupling their interaction with the less sensitive states.

While the previous remarks have made free reference to a photoconductor in the low, intermediate, and high-light ranges, only the first two ranges have been observed or reported. A simple computation shows why it is difficult if not impossible to excite optically the usual photoconductor into the high light range. If one takes an optimistic figure of $10^{15}/cm^3$ for the ground state density, the free carrier densities must exceed this value. An optimistic capture cross section for the ground states is 10^{-15} cm^2. The life time of a free carrier by (5) is then 10^{-7} seconds. The rate of optical generation re-

quired to maintain a density of 10^{15} carriers/cm^3 having a life time of 10^{-7} seconds is by (3a), 10^{22} excitations/cm^3. An extremely high light intensity is about 10 lumens/cm^2 corresponding to about 10^{17} photons/cm^2/sec. If these photons are strongly absorbed in a layer 10^{-5} cm thick, the optical generation rate will be 10^{22}/cm^3/sec or barely sufficient to reach the high light range under optimistic assumptions. While optical excitation is unlikely to reach the high light range, electron excitation is more likely to. For example, a one milliampere, 20 K.V. beam of electrons focussed into a one millimeter spot and absorbed in a layer 10^{-4} cms thick corresponds to about 10^{24} excitations/cm^3/sec. For this reason, one sees qualitative evidence of high-light conditions in luminescent materials under high current bombardment. The emission frequently shifts to shorter wavelengths indicating that the steady-state Fermi-levels are pulling apart to include states closer to the band edges. In some cases, as in the ultra violet emission from ZnO [33], the emission corresponds closely to the energy gap itself. This would mean free carrier densities approaching 10^{19}/cm^3. Another well known bit of evidence from luminescence is that the luminescence decay time at high current densities as in ZnS [33] approaches values less than a microsecond. At lower excitation densities the decay time may approach a second owing to the preponderance of bound states over free carriers.

While it is difficult to excite optically the usual photoconductor into the high light range, this is not true for transistor quality germanium and silicon. Here the density of ground-states is likely to be 10^{11}/cm^3 or at least four powers of ten below that of the usual photoconductor. Since the free carrier density in silicon and germanium in the dark already exceeds the ground state density by two or more powers of ten it is actually difficult to make observations of photoconductivity in the low light range. Any observation of photocurrents approaching the dark currents is already in the high-light range.[10]

Quantitative Determination of the Demarcation Lines (Fig. 5)

Much of the utility of the concept of steady-state Fermi-levels marking the separation between the shallow trapping states and ground states can be demonstrated in the qualitative discussion of the preceding sections. The utility of trying to make close quantitative checks is yet to be established. Some preliminary and encouraging attempts have been reported by R. H. Bube [11, 12] in work on the photoconductivity of CdS and CdSe crystals. For quantitative checks, the demarcation lines discussed earlier should be used together

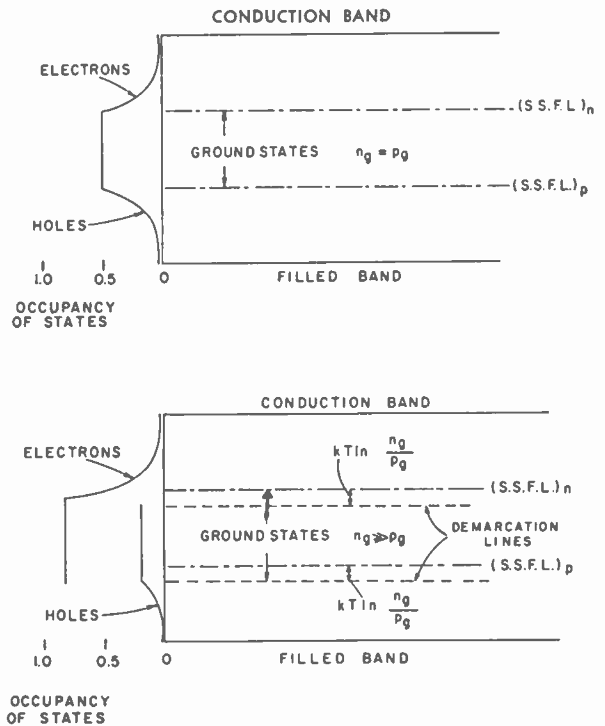


Fig. 5.—Ground states, steady-state Fermi-levels and demarcation lines.

with the steady-state Fermi-levels to determine the number of ground states.[11]

At the demarcation line for electrons, a trapped electron is equally likely to be thermally excited into the conduction band and to capture a free hole. States above this level are dominated by thermal exchange of electrons with the conduction band and states below this level are dominated by the kinetics of recombination. At the demarcation line the following equality holds:

$$\nu^* e^{-E/kT} = v s_p p. \tag{9}$$

The left side of (9) is the rate of thermal excitation of an electron into the conduction band from a bound state E electron volts below the conduction band. The factor ν^*, is the product of the number of chances per second

[10] A. R. Moore has informed the writer of his measurements of pair life times in silicon which appear to have been made at sufficiently low excitation intensities that the measured decay time of photoconductivity was not the life time of a free pair but rather of the longer lived of the two signs of carriers. Higher light intensities readily removed the discrepancy. It is likely here that the transition from "low light" to "high light" conditions was actually being observed. I am informed by A. R. Moore also that similar observations have been made on germanium at low temperatures. This would be consistent with the concept that the two steady-state Fermi-levels approach closer to the band edges and embrace a larger number of ground states as the temperature is decreased.

[11] In the qualitative discussion of insulators it was advantageous and sufficiently accurate to take the ground states as those states lying between the two steady-state Fermi-levels. In reference [9], I stated that a more accurate designation of ground states would be those states lying between the two demarcation lines. This is not correct. I am indebted to M. A. Lampert for pointing out that when $n_g > p_g$, the ground states should be taken between the demarcation line for holes and the steady-state Fermi-level for electrons (see Fig. 5). When $p_g > n_g$, the ground states should be taken between the demarcation line for electrons and the steady state Fermi-level for holes. In treating recombination in semi-conductors, it is especially important to make the proper choice since the steady-state Fermi-levels and the demarcation lines are likely to be well separated. In Fig. 10 of reference [9], showing a semi-conductor, the proper choice was made.

an electron has to jump and a transition probability factor. The right hand side of (9) is simply the reciprocal of the time required for a free hole to be captured by the electron in the bound state. s_p is the capture cross section of the electron for a free hole and p is the density of free holes.

In order to replace s_p and p by corresponding quantities relating to electrons we make use of the formal relation

$$f = ps_p v n_g = n s_n v p_g \qquad (10)$$

which states that the rate of optical generation f of free electrons and holes is equal to the rate at which free holes pour into the ground states and the rate at which free electrons pour into the ground states. n_g is the number of ground states occupied by electrons and p_g the number of ground states occupied by holes.

From (10) we get:

$$ps_p = ns_n \frac{p_g}{n_g}. \qquad (11)$$

Eq. (11) may be inserted in (9) to obtain:

$$\nu^* e^{-E/kT} = nvs_n \frac{p_g}{n_g} \qquad (12)$$

or

$$E = kT \ln \frac{\nu^*}{nvs_n} \frac{n_g}{p_g}. \qquad (13)$$

Now, from detailed balancing arguments (see appendix of article described in reference (8), $s_n/\nu^* = 10^{-26}$ at room temperature. Also, if v is taken as 10^7 cm/sec., (13) may be re-written:

$$E = kT \ln \frac{10^{19}}{n} \frac{n_g}{p_g}$$

$$= E_{fn} + kT \ln \frac{n_g}{p_g}. \qquad (14)$$

When the numbers of electrons and holes in the ground states are equal ($n_g = p_g$), (14) states that the demarcation line coincides with the steady-state Fermi-level for electrons. Similarly, the other demarcation line coincides with the steady-state Fermi-level for holes. If n_g/p_g is not unity, the demarcation line for electrons is shifted $kT \ln n_g/p_g$ farther from the conduction band, and the demarcation line for holes is shifted the same distance closer to the filled band. At room temperature, a ratio of 100 to one of electrons and holes in the ground states will shift the demarcation lines 0.1 volt from the steady-state Fermi-levels.

The above remarks are concerned with a single class of ground states in the low, intermediate or high-light ranges. If more than one class of ground states is present, the same relations hold for each class separately, so that depending on the ratios n_g/p_g within each class, their demarcation lines will be somewhat shifted. One caution, however, needs to be observed. Because, in the low light range, a redistribution of electrons and holes takes place among the several classes of ground states, one must independently estimate this redistribution and its effect on n_g/p_g for each class before applying (14). In the high light range, the classes of ground states act independently and the value of n_g/p_g for each class is equal to the ratio s_n/s_p for that class.

It remains to locate the demarcation lines for a semiconductor in which the photocurrent is less than the dark current. Consider, for example, an n-type semiconductor in which the optically generated electron and hole densities are less than the thermally generated electron and hole densities. The steady-state Fermi-level for electrons is, by assumption, within kT of the "dark" or thermal equilibrium value of the Fermi-level. The demarcation line for holes on the other hand is determined by the relation:

$$\nu^* e^{-E/kT} = vs_n(n_0 + n). \qquad (15)$$

This is (9) written for holes instead of for electrons with n_0 the thermal density of electrons added to n, the optically generated density. Since n was assumed small compared with n_0 (15) may be re-written as:

$$E = -kT \ln v \frac{s_n}{s_p} \frac{s_p}{\nu^*} n_0. \qquad (16)$$

By previous arguments (16) becomes

$$E = kT \ln \frac{10^{19}}{n_0} \frac{s_p}{s_n}$$

$$= E_{fn} + kT \ln \frac{s_p}{s_n}. \qquad (17)$$

That is, the more conducting the semi-conductor is in the dark the closer is the demarcation line for holes to the filled band. For $s_p = s_n$, the demarcation line for holes is the same distance from the filled band as the "dark" Fermi-level is from the conduction band. The ground states[12] lie between the dark Fermi-level and the demarcation line for holes.

The problem of recombination via a single level of bound states treated by Shockley and Read [7] can be analyzed rather simply in terms of the demarcation lines for a semi-conductor. Since only a single level of bound states is assumed, these states are counted in full as ground states when they are embraced by the demarcation line and steady-state Fermi-level but weighted only as $e^{-E/kT}$ when they lie E electron volts outside either of these lines.

Photoconducting Powders

This section on photoconducting powders is inserted not because of any close connection with the preceding discussion but because data on powders have frequently been reported and because these data have certain common characteristics that can be accounted for simply.

[12] See footnote reference 11.

The common characteristics are that the photocurrent increases faster than linearly with voltage but linearly (or nearly so) with light intensity.

The model may be outlined conveniently in terms of the dark current. If one assumes that the granules are appreciably more conducting than the contacts between granules, the field and potential drop within a granule will be negligibly small compared with the field and potential drop across the boundaries between granules. The current will be limited by the poor contacts between granules. This current is field dependent and increases as a high power of the voltage. The field dependence may be the result either of lowering the potential barrier between granules or of tunneling through the barriers. In either event, only a fraction of the free carriers within a granule that strike the barrier will get through (or over) the barrier. This fraction increases rapidly with applied field giving rise to the high power dependence on voltage. On the other hand if the number of free carriers is increased the current will increase in the same proportion since the number of carriers passing the barrier will now be the same fraction of the larger density of carriers.

This model accounts for the near-linear dependence of currents on light intensity and for the high power dependence on voltage. The model suggests also that the dependence on voltage should be similar for both the dark current and photocurrent. Such dependence has been observed by Smith[13] for photoconductive powders imbedded in a transparent insulator. (It is essential for this model that the photosensitivity of the contacts between granules be negligible compared with the photosensitivity of the granules themselves.) Another consequence, observed by Nicoll and Kazan [34] is that at sufficiently high voltage the contact resistance is broken down so that the dependence on voltage approaches the linear form.

Noise Currents

Ultimately the sensitivity of a photoconductor is measured not by "micro-amperes per lumen" or by the "number of electrons per photon" from (6), but by the signal to noise ratio for a given amount of incident radiation. "Signal" is the average photocurrent; "noise" is the root mean square fluctuation in current about the average. It is true that there are many applications for photoconductors where a large number of microamperes per lumen, apart from noise considerations, is at least convenient if not necessary to the operation of the device. However, when a photoconductor is used as a radiation detector, the ultimate background noise currents and the signal to noise ratio for a given amount of radiation are the significant parameters.

The subject of noise currents, like the subject of the average photocurrent, is characterized on the one hand by simply definable expectations and, on the other hand, by large and complex departures from these expectations.

The expectations are that a photoconductor should be a noiseless transducer of photon currents into photoelectron currents. The signal to noise ratio of the photocurrent should match the signal to noise ratio of the absorbed photon stream. A photo-multiplier tube is a good example of such a transducer.

The realizations in most photoconductors, as in most semi-conductors, are noise currents far in excess of minimum expectations and noise currents having approximately a $1/f$ spectrum instead of a "flat" spectrum. Large noise currents observed in CdS crystals with gold contacts are reported by Buttler [35]. Here the signal to noise ratio of the photocurrents is measured to be 10^3 times lower than the signal to noise ratio of the incident photon stream. An extensive discussion of and bibliography on noise currents in semi-conductors is contained in a paper by Petritz [36]. A representative set of measurements on noise currents in semi-conductors is contained in a paper by Campbell and McLean [37].

At the outset the writer states his leaning towards the interpretation that the $1/f$ spectrum and the large noise currents in excess of simple expectations are generated at poor contacts either at the electrode-semiconductor interface or at internal "surfaces." This leaning is supported both by the experience of converting CdS crystals from noisy to noiselesss transducers by using ohmic contacts [38, 39] and by an appreciation of the ease with which a contact can depart from truly ohmic character. The latter statement needs elaboration.

It was pointed out earlier that the validity of (6) for the average photocurrent depended upon having ohmic contacts. An ohmic contact is one that can supply a reservoir or excess number of carrier ready to enter the semi-conductor as needed. A metal having a lower work function than the semi-conductor is a simple example of an ohmic contact for electrons. One must, however, insure that there is no high work function film or monolayer between metal and semi-conductor. Now, in the measurement of the average photocurrent, a non-ohmic contact may not be serious. For example, if a hundred volts is applied across a photoconductor, and if a fraction of a volt is abstracted from the hundred volts to "break through" the poor contact between metal and photoconductor, the average photocurrent may depart less than one per cent from that to be expected from an ohmic contact. Even here, if one chooses to use only a few volts across the photoconductor, an ohmic contact may still show a distinct improvement over a non-ohmic contact [38]. While the measurement of the average photocurrent is not particularly sensitive to the nature of contact, since usually a few volts at most is sufficient to break down the poor contact, the measurement of noise currents is exceedingly sensitive to even slight departures from an ohmic contact. The reason, roughly, is that the noise voltages corresponding to the noise currents are likely to be of the order of microvolts. Poor

[13] R. W. Smith, private communication.

contacts that involve as little as some microvolts to break them down can have a significant influence on the noise since breakdown is a noisy process. In the case of the gold contacts on CdS, fairly accurate average photocurrents are measurable at a hundred volts applied across the crystal; at the same time, the noise currents are several powers of ten higher than the minimum set by the photon stream. *The noise criterion is perhaps the most critical test of an ohmic contact.*

These remarks have emphasized the ease with which a poor contact can cause the magnitude of noise currents to exceed their minimum values. It will be argued later that the noise currents generated at a surface are also much more likely than volume generated noise currents to have a $1/f$ spectrum.

Some Formal Expressions for Noise Currents

There is a conceptual as well as numerical convenience in reducing the variety of noise currents (Johnson, shot, photon, current and contact) to the same formalism used in shot noise. There is a further convenience in dealing with signal-to-noise ratios rather than noise current itself. The well-known expression for shot noise is

$$I_n = [2(eG)I\Delta f]^{1/2} \tag{18}$$

where I_n is the rms noise current in the band width Δf arising from an average current I. (eG) is the effective electronic charge. For example, G may be the gain of a photo-multiplier. G may also be less than unity if the carriers move only a fraction of the distance between two plane parallel electrodes. The signal to noise ratio is

$$\frac{I}{I_n} = \left(\frac{I}{2eG\Delta f}\right)^{1/2}. \tag{19}$$

The interpretation of the parenthesis on the right side of (19) is quite simple. For a bandwidth, Δf, the smallest observable (resolvable) element of time is $\Delta t = 1/2\Delta f$. Accordingly $I/2eG\Delta f$ is the average number of effective charges contained in the observation time Δt. To generalize, one may say, as is well known, that the parenthesis of (19) is *the average number of random events in the observation time* Δt. This reduces the expression for signal to noise ratio to the well known statistical relation that the rms deviation associated with an average number of events is the square root of that number. Thus the signal noise ratio or ratio of average number to rms deviation from the average number is also equal to the square root of the average number, viz. (19).

Consider the noise implications of the characteristic (2), repeated here for convenience:

$$N = F\tau. \tag{2}$$

N is the steady-state number of free carriers having a life time τ and generated by a photon flux of F excitations/sec. Now the bandwidth of this photoconductor is determined by the life time τ through the relation

$$\tau = \frac{1}{2\Delta f}. \tag{20}$$

Thus, τ is the smallest observable element of time and N is the average number of random photon events occurring in this observation time. The signal to noise ratio of the absorbed photon stream referred to the bandwidth of (20) is therefore $N^{1/2}$. The signal to noise ratio of the photocurrent is by (19) and (20):

$$\frac{I}{I_n} = \left(\frac{I}{2eg\Delta f}\right)^{1/2} = \left(\frac{I\tau}{eG}\right)^{1/2}. \tag{21}$$

From (6) and (9), (21) may be rewritten:

$$\frac{I}{I_n} = \left(\frac{I\tau}{eG}\right)^{1/2} = (f\tau)^{1/2} = N^{1/2}. \tag{22}$$

Thus the number $N^{1/2}$ is at the same time the signal to noise ratio of the incident photon stream, the signal to noise ratio of the photocurrent and the "signal to noise ratio" of the steady-state number of optically excited free carriers.[14]

If the observation time τ is made artificially longer by using a narrow pass amplifier, the larger value of τ is inserted in (22) and the signal to noise ratio is improved in proportion to $\tau^{1/2}$. Also, if the rise and decay times of the photocurrent are larger than the life time of a free carrier the larger time is inserted in (22).

From (22), it is immediately clear that the smallest signal light that can be detected is equal to the square root of the number of Planckian photons having the same $h\nu$ values as the signal photons, received by the photoconductor in the observation time.

Comparison of Noise Currents

The shot noise relation of (18) is a convenient rallying point for comparing the magnitudes of the various noise currents encountered in semi-conductors and vacuum tubes. When G = unity, (18) describes the noise currents in a vacuum photocell, a temperature limited diode or a *p-n* junction photocell. When the value for G is inserted from (6), namely, $G = \tau/T_r$, (18) describes the noise currents in a photoconductor. G may of course be greater or less than unity. Particularly large values of G may be expected at thin barriers either at the metal contacts to a semi-conductor, at the boundaries between grains of a polycrystalline material, owing to the small values for transit time T_r. This is a likely mechanism for large values of contact noise.

Finally thermal agitation noise can be described (within a factor of two) by (17) if the following values are inserted for G and I:

$$G = \frac{L_d}{L} = \frac{\text{diffusion length of a carrier}}{\text{inter-electrode distance}}$$

[14] While this statement is accurate for a photoconductor in which the life time does not fluctuate (for example, a *p-n* junction or other structures in which the life time is determined by transit time) Dr. D. O. North has pointed out in a private communication that the fluctuations in life time, characteristic of life times determined by recombination, cause the photocurrents to be twice as noisy as the photon currents.

$$I = \frac{kT}{eL_d} \sigma A = \text{random diffusion current.}$$

Here, σ and A are the conductivity and cross section of a semi-conductor whose resistance is R and whose Johnson noise is given by

$$\overline{I^2_n} = \frac{4kT}{R} \Delta f.$$

The virtue of the above mode of expression is that one gets an immediate measure for the applied field at which what is called "current noise" will exceed thermal agitation noise. This field is kT/eL_d. Applied fields greater than this value cause the drift length of the carrier to exceed its diffusion length [40].

Sources of Excess Noise

The minimum noise in a photoconductor is given by (18) with the photoconductive gain inserted for G. It is an understatement to say that this minimum value may easily be exceeded. Noisy contacts are the most likely source. However, other non-uniformities of a less obvious nature can contribute as well. These include any non-uniform potential distribution along the photoconductor and non-uniform photosensitivity of the photoconductor either along the line of the current or along a line normal to the current. The latter source of excess noise has a familiar parallel in the problem of combining photomultipliers having different gains.

There is, in addition to the above sources of excess noise, a somewhat more fundamental possibility for observing photocurrents that are noisier than the photon currents generating them [9]. Under certain conditions it is possible to identify a definite fraction of the free carriers with one class of recombination centers and the remainder with another class of recombination centers. If the larger fraction of carriers is supported or generated by a small fraction of the incident light, the photocurrents will be noisier than the photon currents. This is again similar to the problem of paralleling photomultipliers with different gains.

Evidence for Shot Noise in Photoconductors

There are remarkably few published reports in which the signal to noise ratio of the photoconductor currents approach the signal to noise ratio of the incident photon currents. The fact that there are any at all is, however, sufficient encouragement to recognize the "noiseless photoconductor" as a legitimate, though difficult, goal. A discussion of the usual results in which large noise currents are observed having a $1/f$ spectrum is given by Jones [41].

P. J. Fellgett [42] has reported measurements on lead sulphide and lead telluride cells in which the limiting noise was the shot noise of the temperature radiation falling on the cells. Van der Ziel [43] has reported that a part of the noise spectrum of the current in a germanium filament is consistent with the shot noise to be expected from the density and life time of the minority carriers. While these are not photocurrents, the physical basis for expecting simple shot noise here is the same as that for a photoconductor. Slocum and Shive [44] reported measurements of shot noise in a p-n junction photocell that agreed with that to be expected from temperature limited currents of the same magnitude. Lummis and Petritz [45] report a close connection between the noise spectrum of PbS and its macroscopic photoconductive properties. Finally, Shulman [39] reported some of the most complete measurements of shot noise in the photocurrents through single crystals of CdS having ohmic contacts. The validity of the simple noise relations discussed in this paper was supported by the following: the photoconductor gain measured from the noise currents agreed with a direct measurement of gain; the noise spectrum flattened off toward low frequencies approximately at the frequency given by the reciprocal of the life time of a free carrier; and the rms noise current increased as the square root of the photocurrent when the photocurrent was varied by light intensity and as the first power of the photocurrent when it was varied by the applied voltage.

Another interesting example of the matching of gains computed from noise with those observed directly is reported by Cope [46] for a photoconductive television pick up tube (Vidicon) recording x-ray pictures.

The $1/f$ Noise Spectrum

The commonly observed noise spectrum, both for photoconductors and semi-conductors is one in which the mean squared noise current per unit bandwidth varies as the reciprocal of the frequency. This spectrum immediately suggests a collection of noise sources having different time constants. A number of analyses have been published in which such a spectrum was resolved in terms of ions migrating to and from a critical surface, the ions having a dispersion of activation energies for diffusion [47, 48]. It is not necessary, however, to restrict the argument to ions since trapped electrons or holes in the neighborhood of a potential barrier can modulate the height of the barrier and the flow of current through or over the barrier. This is the normal photoconductive effect on a small scale.

If one thinks in terms of electronic charges migrating in and out of a critical region at a barrier, the $1/f$ spectrum can be reconstructed if the number of trapped charges per unit life time in the critical region varies as the reciprocal life time. The photoelectric gain per trapped charge is proportional to its life time in the critical region. Further, if the life time of the trapped charges in the critical region is the usual exponential function of trap depth, the number of electron charges per unit trap-depth energy in the critical region necessary to give the $1/f$ spectrum is a *constant*. The last statement is particularly suggestive since there is ample evidence in the usual semi-conductor for a quasi-continuous distribution of traps in energy [49] and, in particular, the

distributions can be nearly enough constant [8, 17] to satisfy the requirements of the $1/f$ spectrum. Since the life time of electrons in deep traps may be as much as hours or days, these deep trapped charges can account for the $1/f$ spectrum extending as far as has been measured, namely, to 10^{-4} cycles/sec. [50]

The collection of noise sources, having different characteristic times, required to match the observed $1/f$ spectrum can readily be found at a surface contact or internal barrier since various elements of a surface are independent. By way of contrast, it is more difficult to establish such a collection of noise sources, having different characteristic times, uniformly distributed within the *volume* of a photoconductor or semi-conductor even with the aid of traps. A free electron in the volume of a photoconductor, by virtue of its random motions within the volume, tends to average over various capturing centers to establish a *single* characteristic life time.

Another argument against assigning the $1/f$ spectrum to volume sources may be stated as follows. Since the noise fluctuations in photocurrent may be regarded as generated by fluctuations in photon current (in the same way as any AC component in the light intensity should give rise to an AC component in the current), one can show that a $1/f$ noise spectrum implies a response curve to fluctuations in light intensity that rises as $\ln f$ at low frequencies. In so far as one commonly finds the response curve *flattening off* at low frequencies, and in so far as this response curve is a measure of the average volume properties of a photoconductor, the average volume properties cannot provide the distribution of time constants required to account for the $1/f$ spectrum in this low frequency range. By contrast, highly localized volume properties such as at an internal or a contact surface may have little effect on the average-current response curve but a large effect on fluctuations from the average. (This has already been discussed in the section on "noise currents.") One has, therefore, more freedom in identifying the $1/f$ spectrum with surfaces, internal or contact.

Petritz points out that a set of independent time constants needed to match the $1/f$ spectrum can be obtained from the assumption of local inhomogeneities but cannot be obtained from traps uniformly distributed throughout the volume [36]. He goes on however to reject the local inhomogeneity model for his particular problem (p-n junction noise) because the macroscopic properties do not give evidence of these local inhomogeneities. The argument in the present paper, however, is that the local inhomogeneities can have a negligible effect on the macroscopic (average) currents but a very large effect on the noise.

Photocurrent Noise in the Presence of Dark Current Noise

It is often necessary to observe small photocurrents in the presence of much larger thermally generated dark currents. It is not immediately true that the noise from the "dark" electrons is weighted the same as the noise from the photo electrons. If both "dark" and photo electrons originate from the same centers their mean squared noise currents will of course be proportional to their relative numbers. If on the other hand, the "dark" electrons originate from states lying between the Fermi-level and the conduction band whereas the photo electrons originate from states below the Fermi-level, the dark electrons can be far less noisy than the same number of photo electrons.

Van der Ziel [51], for example, has pointed out that free electrons arising from well ionized states near the conduction band contribute less noise than those originating from deeper lying states. An independent and more general argument has been carried out by D. O. North[15] in which this is one of the conclusions arrived at.

If one regards thermal generation of carriers as similar to optical generation an heuristic measure of the noise contribution of thermal electrons may be obtained by the following argument. Consider the fluctuation in density of thermal carriers due to fluctuations in temperature. The maximum fluctuation is obtained in general when the Fermi-level has zero temperature coefficient. For this case:

$$\left.\frac{\Delta \bar{n}}{\Delta T}\right]_1 \doteq \bar{n}\frac{E_f}{kT^2}. \tag{23}$$

If the Fermi-level has a finite temperature coefficient,

$$\left.\frac{\Delta \bar{n}}{\Delta T}\right]_2 \doteq \bar{n}\left(\frac{E_f}{kT^2} - \frac{1}{kT}\frac{dE_f}{dT}\right). \tag{24}$$

If, now, one takes $\bar{n}$ as the maximum mean squared fluctuation of n carriers (to be observed at a constant Fermi-level) then the mean squared fluctuation in general will be

$$\langle (n-\bar{n})^2\rangle = \bar{n}\left.\frac{\Delta \bar{n}}{\Delta T}\right]_2 \Big/ \left.\frac{\Delta \bar{n}}{\Delta T}\right]_1 = \bar{n}\left(1 - \frac{T}{E_f}\frac{dE_f}{dT}\right). \tag{25}$$

In the limit when

$$\frac{dE_f}{dT} \to \left(\frac{T}{E_f}\right)^{-1},$$

the carrier density becomes invariant with temperature and the fluctuations in numbers of carriers approach zero.

A further parallel with photocurrents would be to define a life time for the thermal carriers in terms of the operation of instantly dropping the temperature and observing the time required for the carrier density to drop to half its initial value. This life time would determine the frequency range of the noise spectrum appropriate to fluctuations in the carrier density.

[15] Private communication.

Recapitulation of Noise Currents

Fig. 6 shows schematically the noise spectra from various sources in a photoconductor. In a photoconductor having ohmic contacts and negligible dark current, only the photocurrent noise and thermal agitation noise would be present. And the photocurrent noise would be given by the incident photon noise. The frequencies f_1, f_2 and f_3 are taken here to be the reciprocal life times of trapped charges near a barrier contact. f_4 is the reciprocal life time of photo-generated carriers and f_5 of thermally generated carriers. f_6 is reciprocal RC of photoconductor. Approximating association of f and reciprocal life time is for simplicity of schematic presentation.

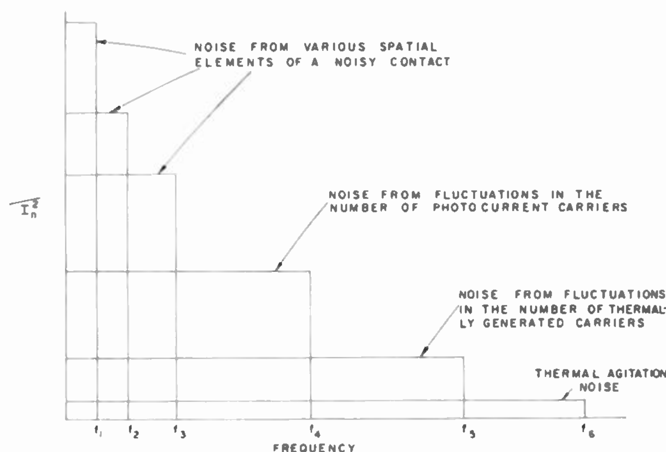


Fig. 6—Schematic representation of various noise currents in a photoconductor.

Acknowledgments

The ideas expressed in this paper have been generated over a number of years during which time the writer has profited from discussions with many of his colleagues. The writer is indebted to D. O. North, R. H. Bube, R. W. Smith, P. K. Weimer, S. V. Forgue, A. D. Cope, and C. I. Shulman among others.

Bibliography

[1] Gudden, B., "Lichtelectrische Erscheinungen." J. Springer, Berlin, Germany (1928).
[2] Weimer, P. K., and Cope, A. D., "Photoconductivity in Amorphous Selenium," *RCA Review*, Vol. 12 (1951), p. 314.
[3] Pensak, L., "Conductivity Induced by Electron Bombardment in Thin Insulating Films." *Physical Review*, Vol. 75 (1949), p. 472.
[4] (a) Farmer, F. T., "Electrical Properties of Polystyrene." *Nature*, Vol. 150 (1942), p. 521.
(b) Fowler, J. F., and Farmer, F. T., "Effect of Temperature on the Conductivity Induced in Insulators by X-Rays." *Nature*, Vol. 171 (1953), p. 1020.
[5] Hilsch, R., and Pohl, R., "Stationary Photoelectric Primary and Secondary Currents in Crystals, Particularly in KH-KBR Mixed Crystals." *Zeitschrift für Physik*, Vol. 108 (1937), p. 55.
[6] Pohl, R., and Stockman, F., "Die Rolle Sekundarer Elektronen Bei Der Lichtelektrischen Leitung." *Annalen der Physik*, Vol. 1 (1947), p. 275.
[7] (a) Broser, I., and Warminsky, R., "Zur Theorie der Huminessenz und der elektrischen Heitfähigkeit von Kadmiumsulfidkristallen." *Annalen der Physik*, Vol. 7 (1950), p. 289.
(b) Shockley, W., and Read, W. T., Jr., "Statistics of the Recombinations of Holes and Electrons." *Physical Review*, Vol. 87 (1952), p. 835.
(c) Isay, W. H., "Beitrag zur Leitfähigkeitstheorie der Halbleiter." *Annalen der Physik*, Vol. 13 (1953), p. 327.
[8] Rose, A., "An Outline of Some Photoconductive Processes." *RCA Review*, Vol. 12 (1951), p. 362.
[9] Rose, A., "Recombination Processes in Insulators and Semiconductors," *Physical Review*, Vol. 97 (1955), p. 322.
[10] Smith, R. W., "Some Aspects of the Photoconductivity of Cadmium Sulfide." *RCA Review*, Vol. 12 (1951), p. 350.
[11] Bube, R. H., and Thomsen, S. M., "Photoconductivity and Crystal Imperfections in Cadmium Sulfide Crystals. Part I. Effect of Impurities." *Journal of Chemical Physics*, Vol. 23 (1955), p. 15.
[12] Bube, R. H., "Superlinear Photoconductivity." Proceedings of Conference on Photoconductivity, Atlantic City, N. J., November, 1954. John Wiley and Sons, Inc., New York. In press.
[13] Niekisch, E. A., "Zur Photoleitfähigkeit von Cadmiumsulfid-Einkristallen Bei Tiefen Temperaturen." *Annalen der Physik*, Vol. 8 (1951), p. 291.
[14] Stockmann, F., "Strahlungslose elektronenübergange in Kristallen." *Zeitschrift für Physik*, Vol. 130 (1951), p. 477.
[15] Bube, R. H., "A Comparative Study of Photoconductivity and Luminescence." *Physical Review*, Vol. 83 (1951), p. 393.
[16] Dacey, G. C., "Space-Charge Limited Hole Current in Germanium." *Physical Review*, Vol. 90 (1953), p. 759.
[17] Smith, R. W., and Rose, A., "Space-Charge Limited Currents in Single Crystals of Cadmium Sulfide." *Physical Review*, Vol. 97 (1955), p. 1531.
[18] Haynes, J. R., and Hornbeck, J. A., "Decay in Photoconductivity Associated with Hole Traps in N-Type Silicon." *Physical Review*, Vol. 94 (1954), p. 1438.
[19] Forgue, S. V., Goodrich, R. R., and Cope, A. D., "Properties of Some Photoconductors, Principally Antimony Trisulfide." *RCA Review*, Vol. 12 (1951), p. 335.
[20] Anspacher, F., and Ehrenberg, W., "Electron-Bombardment Conductivity of Dielectric Films." *Proceedings of the Physical Society, A*, Vol. 64 (1951), p. 362.
[21] Pick, H., "Über Lichtelektrische Leitung am Bleisulfid." *Annalen der Physik*, Vol. 3 (1948), p. 255.
[22] von Hippel, A., and Rittner, E., "Thallous Sulfide Photoconductive Cells. II. Theoretical Discussion." *Journal of Chemical Physics*, Vol. 14 (1946), p. 370.
[23] Jacobs, J. E., "The Photoconductive Cell." *General Electric Review*, Vol. 57 (1954), p. 28.
[24] Pohl, R. W., and Stockman, F., "Zur Verstarkung von Elektronenstromen in Bestrahlten Kristallen." *Annalen der Physik*, Vol. 6 (1949), p. 89.
[25] van Roosbroeck, W., and Shockley, W., "Photon-Radiative Recombination of Electrons and Holes in Germanium." *Physical Review*, Vol. 94 (1954), p. 1558.
[26] Hall, R. N., "Electron-Hole Recombination in Germanium." *Physical Review*, Vol. 87 (1952), p. 387.
[27] Conwell, E. M., "Fundamental Properties of Semiconducting Materials." *Sylvania Technologist*, Vol. 7 (1954), p. 41.
[28] Broser-Warminsky, R., "Die Berechnung der energetischen Lag der Storstellen von Photohalbleitern (Kristallphosphoren) aus abkling-und ausheizkurven." *Physikalische Verhandlung*, Vol. 4 (1953), p. 191.
[29] Schon, M., "Über die Strahlungslosen Übergange der Elektronen im Gitter der Kristallphosphore." *Tech. Wiss. Abh. der Osram*, Vol. 6 (1953), p. 49.
[30] Aigrain, P., and Bulliard, H., "On the Photo-magnetoelectric Effect." *Comptes Rendus*, Vol. 236 (1953), pp. 595 and 672.
Moss, T. S., "Photoelectromagnetic and Photoconductive Effects in Lead Sulfide Single Crystals." *Proceedings of the Physical Society (London)*, Vol. 366 (1953), p. 993.
[30a] Frerichs, R., "On the Conductivity Produced in CdS Crystals by Irradiation with Gamma Rays." *Physical Review*, Vol. 76 (1949), p. 1869.
[31] Taft, E. A., and Hebb, M. H., "Note on Quenching of Photo-Conductivity in Cadmium Sulfide." *Journal of the Optical Society of America*, Vol. 42 (1952), p. 249.
[32] (a) Fassbender, J., "Über die Photoelektrischen Eigenschaften von Kadmiumsulfid-Einkristallen." *Annalen der Physik*, Vol. 5 (1949), p. 33.
(b) Fassbender, J., and Lehmann, H., "Berchnung von Elektronenbeweglichkeiten in Cadmiumsulfid-Einkristallen aus Wechsellichtmessungen." *Annalen der Physik*, Vol. 6 (1949), p. 215.
[33] Leverenz, H. W., "Luminescence of Solids." John Wiley and Sons, Inc., New York (1950), pp. 218 and 264.
[34] Nicoll, F. H., and Kazan, B., "Large Area High-Current Photo-Conductive Cells Using Cadmium Sulfide Powder." *Journal of the Optical Society of America*, Vol. 45 (1955), p. 647.
[35] Buttler, W. M., "Über das Randschichtrauschen von CdS-Einkristallen." *Annalen der Physik*, Vol. 11 (1953), p. 368.
[36] Petritz, R. L., "On the Theory of Noise in *p-n* Junctions and Related Devices." *Proc. IRE*, Vol. 40 (1952), p. 1440.
[37] Campbell, R. H., and McLean, W. S., "Noise from Current-Carrying Resistors 20 to 500 Kc." *Proc. IRE*, Vol. 37 (1949), p. 938.

[38] (a) Smith, R. W., and Rose, A., "Current-Voltage Characteristics of Cadmium Sulfide Crystals Using Ohmic Contacts." *Physical Review*, Vol. 92 (1953), p. 857.
[38] (b) Smith, R. W., "Properties of Ohmic Contacts to Cadmium Sulfide Single Crystals." *Physical Review*, Vol. 97 (1955), p. 1525.
[39] (a) Shulman, C. I., Smith, R. W., and Rose, A., "Some Noise Observations on CdS Crystal Photoconductors." *Physical Review*, Vol. 92 (1953), p. 857.
[39] (b) Shulman, C. I., "Measurement of Shot Noise in CdS Crystals." *Physical Review*, Vol. 98 (1955), p. 384.
[40] Davydov, B., and Gurevich, B., "Voltage Fluctuations in Semi-Conductors." *Journal of Physics (U.S.S.R.)*, Vol. 7 (1943), p. 138.
[41] Jones, R. C., "A Method of Describing the Detectivity of Photoconductive Cells." *Review of Scientific Instruments*, Vol. 24, (1953), p. 1035.
[42] Fellgett, P. J., "On the Ultimate Sensitivity and Practical Performance of Radiation Detectors." *Journal of the Optical Society of America*, Vol. 39 (1949), p. 970.
[43] van der Ziel, A., and Herzog, G. B., "Shot Noise in Germanium Single Crystals." *Physical Review*, Vol. 84 (1951), p. 1249.
[44] Slocum, A., and Shive, J. N., "Shot Dependence of *p-n* Junction Phototransistor Noise." *Journal of Applied Physics*, Vol. 25 (1954), p. 406.
[45] Lummis, F. L., and Petritz, R. L., "The Relation between Noise and Response Time in Photoconductors: II. Experiment." *Physical Review*, Vol. 86 (1952), p. 660.
[46] Cope, A. D., and Rose, A., "X-Ray Noise Observation Using a Photoconductive Pickup Tube." *Journal of Applied Physics*, Vol. 25 (1954), p. 240.
[47] Schottky, W., "Small-Shot Effect and Flicker Effect." *Physical Review*, Vol. 28 (1926), p. 74.
[48] (a) Du Pre, F. K., "A Suggestion Regarding the Spectral Density of Flicker Noise." *Physical Review*, Vol. 78 (1950), p. 615.
(b) van der Ziel, A., "On the Noise Spectra of Semiconductor Noise and of Flicker Effect." *Physica*, Vol. 16 (1950), p. 359.
[49] Bube, R. H., "Luminescence and Trapping in Zinc Sulfide Phosphors with and without Copper Activator." *Physical Review*, Vol. 80 (1950), p. 655.
[50] Rollin, B. V., and Templeton, I. M., "Noise in Semiconductors at Very Low Frequencies." *Proceedings of the Physical Society, B*, Vol. 66 (1953), p. 259.
[51] van der Ziel, A., "Shot Noise in Semiconductors." *Journal of Applied Physics*, Vol. 24 (1953), p. 222.

Some Additional References on the Characteristics of Bound States
[52] Tyler, W. W., and Woodbury, H. H., "Properties of Germanium Doped with Iron." *Physical Review*, Vol. 96 (1954), p. 874.
[53] Wlerick, G., "Equilibre Electronique d'un Semi-Conducteur Eclaire." *Journal de Physique et le Radium*, Vol. 15 (1954), p. 667.
[54] Burton, J. A., Hull, G. W., Morin, F. J., and Severeins, J. C., "Effect of Nickel and Copper Impurities on the Recombination of Holes and Electrons in Germanium." *Journal of Chemical Physics*, Vol. 57 (1953), p. 853.

Lead Salt Photoconductors*

T. S. MOSS†

***Summary*—This review article discusses the fundamental basis of photoconductivity in lead sulphide, telluride, and selenide, and its application in modern highly sensitive infrared detectors.**

The first part of the paper deals with the manufacture of cells and the processing of photosensitive layers; and with the characteristics of the final detectors, such as response time, sensitivity, and spectral distribution of sensitivity. The factors determining the maximum attainable sensitivity are discussed.

The second part covers fundamental semiconducting properties of the three materials, including absorption, activation energies, carrier lifetimes and effective masses; and the theory of photoeffects in layers. Some applications of these detectors are described.

An extensive list of references is given.

Introduction

THE LEAD SALT photoconductors, PbS, PbTe and PbSe, are of great technicological interest as highly sensitive and fast infrared red radiation detectors, and also of considerable scientific interest in the general field of photoeffects in semiconductors.

Until the end of World War II the only infrared detectors available for wavelengths greater than 1.5μ (15,000Å) were thermal detectors such as bolometers and thermopiles. Such devices operate by virtue of the fact that radiant energy is absorbed in the sensitive element, thus increasing its temperature. This temperature rise is then manifested as a change in resistance or generation of a thermoelectric voltage.

In the last few years there has been a great advance in the production and utilization of photoconductive infrared detectors, and in the fundamental investigation of the phenomenon of photoconductivity. In contrast to the heat detectors mentioned above, these devices operate by direct excitation of individual electrons by absorbed radiant quanta, so that they are quantum detectors closely allied to the familiar photoemissive cells. The fundamental difference between the photoemissive cell and the photoconductive cell is that in the former the absorbed quanta have sufficient energy to liberate electrons completely from the photocathode into the surrounding vacuum where they may be collected by a suitable anode. In photoconductive cells electrons are not emitted, but they absorb sufficient energy from the quanta to enable them to change from a *bound* state where they are incapable of contributing to conductivity to a *free* state where they are capable of carrying a current.

Photoconductivity thus appears as an increase in conductivity (due to an increase in the number of current carriers) in a partial or semiconductor under the action of irradiation.

* Original manuscript received by the IRE, August 5, 1955.
† Royal Aircraft Establishment, Farnborough, England.

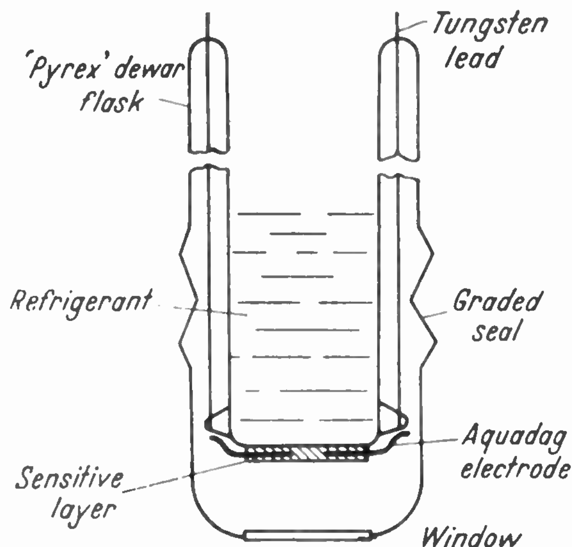


Fig. 1—Diagrammatic representation of a photoconductive cell.

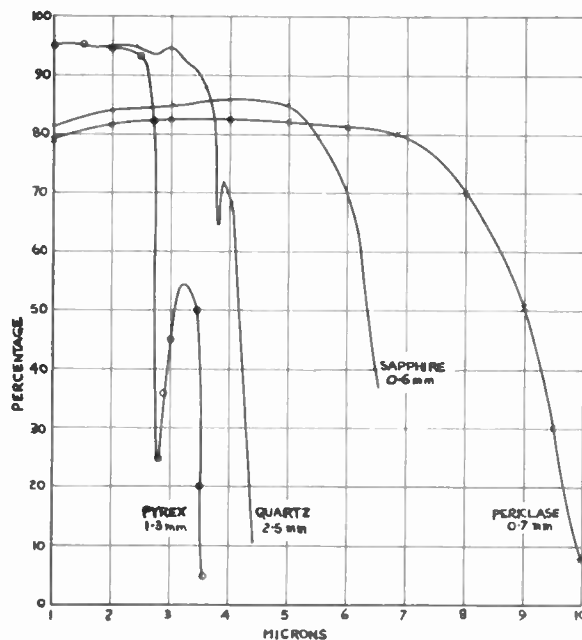


Fig. 2—Transmission of window materials.

In contrast to the heat detectors, photoconductors are not sensitive to all infrared wavelengths because for any material a certain minimum quantum energy is required to free the electrons. There exists, therefore, a fairly well defined maximum wavelength beyond which the sensitivity of a particular photoconductor falls off rapidly. A further fundamental difference between quantum and heat detectors is in the response time. For the latter, since the sensitive element has to warm up after the radiation is applied, the process is essentially rather slow. Even for so-called fast thermal detectors the response time is an appreciable fraction of a second. For the quantum detectors typical response times are one to a hundred microseconds.

The particular class of photoconductors considered here, namely lead sulphide, lead telluride, and lead selenide, is the one which at the present time is of the greatest practical use in the near infrared, and one on which intensive research work has been, and is still being carried out. Under conditions of operation most generally employed, these three types of lead salt photoconductor cover following parts of infrared spectrum:

PbS	1–3 microns
PbTe	2–5.5 microns
PbSe	3–7 microns.

It is worthy of emphasis that the long wavelength limit of sensitivity does not increase progressively with atomic weight as initially was expected. Over the wavelength region of 1.5μ–7μ these three materials are by far the most sensitive and most rapid infrared detectors at present available.

Manufacture and Properties of Photoconductive Cells

Lead salt photoconductive cells are made by two distinct processes, which affect not only the characteristics but also the actual physical form of the cells.

In the chemical method a thin layer is deposited from solution onto a suitable substrate—which is usually glass. These detectors are generally left open to the air—with possibily a protective coat of varnish—and can thus be made in a variety of shapes and sizes, and in particular may have very small volume. They cannot in general, however, be cooled by liquid air or other refrigerants.

In the second method the sensitive layer is produced by evaporation *in vacuo*. By definition such a cell must be made in a complete envelope, and provided with a suitable "window" of material transparent to the wavelengths involved. For convenience of manufacture, where large temperature differentials are required, and to facilitate cooling in use, the cells are usually made in the form of double-walled glass flasks. When sensitized the cell is sealed off in the evacuated state so that good thermal insulation is available. A typical cell of this type is shown diagrammatically in Fig. 1.

The body of the cell is usually made of low expansion borosilicate glass, such as "Pyrex." If the window material is other than the same glass, a multiglass graded seal will be required to match the expansion coefficient of the window to that of the Pyrex.

Fig. 2 shows the infrared transmission of various materials suitable for windows. Pyrex glass gives good transmission for wavelengths up to 2.7μ and is thus satisfactory for use with PbS cells when operated at room temperature. As will be discussed later, the range of sensitive wavelengths increases on cooling, and to obtain the benefit of this increased response from a cooled PbS detector, it is advisable to use a fused silica window, which is transparent to $\sim 4\mu$, or a "bubble" window of very thin glass.

For PbTe cells, windows of artificial sapphire give

adequate transmission. Such windows can now be obtained commercially, and may be sealed to suitable glasses by techniques described by Chasmar *et al.* [19]. For this window material about five glasses are required in the graded seal.

To cover the full range of sensitivity of the cooled PbSe cell the only suitable window material at present is crystalline MgO—periclase. Again this material may be sealed to a suitable glass. For both chemical and vacuum cells contacts are usually made to the sensitive layer by electrodes of aquadag or gold. In the case of chemical cells the electrodes are sometimes added *after* forming the sensitive layer.

Processing of Photosensitive Layers

The basic prodecure for production of layers by the vacuum evaporation method is as follows. Enough of the relevant compound to form a layer approximately 1μ thick is inserted into the cell which is then evacuated. The cell is heated in an oven to $\sim$600°C while a jet of cool air is directed onto the outer wall opposite the electrode surface where the material condenses as a mirror-like layer. The oven is now removed and the layer evaporated onto the electrode surface which is kept cool by water in the cell, an air blast, or possibly by liquid air. All techniques of making highly sensitive layers involve the use of oxygen at some stage in the process although it has been shown to be possible by Gibson [36] to make PbS cells without oxygen which show some sensitivity at low temperatures after illumination with a strong light. Various ways of introducing oxygen used by different workers are:

1. To include it in the starting material, usually by addition of PbO to the main compound.
2. To carry out one or more of the evaporations in a low pressure of oxygen.
3. To bake the final layer in a low pressure of oxygen.

Manufacture of the most sensitive layers is still very much an art rather than a science, and any or all of the above ways of introducing oxygen may be used for any one cell. Optimization of the times, temperatures, pressures, cell volumes, pumping speeds, quantities and composition of materials involved is the key to good layer production. Schwarz [109] has used an electric discharge during evaporation.

It may be noted here that the treatment required for a layer to have good sensitivity at room temperature differs from that required for a layer which is to give its best performance at liquid air temperature. Although the differences of technique are basically trivial they are so important that a worker who can make very good cells of one type may be incapable (at least without weeks or even months of trial and error work) of making even mediocre cells of the other type.

The chemical method of all cell manufacture has so far only been used for production of PbS and PbSe cells. The re-agents normally employed are lead acetate, sodium hydroxide, and thio- or seleno-urea. After precipitation of the layer and washing and drying it, sensitivity is optimized by baking in air or low pressure oxygen.

It is common practice to use a substrate layer—PbO for PbS cells and PbS for PbSe cells.

Details of techniques of manufacturing layers are given in the following references:

1. All types of cell—Lovell [64], Sutherland and Lee [126], Michellsen [69a], and Smith [117].
2. PbS (evaporated type)—Cashman [16], Sosnowski *et al.* [121], Frank [30], Frank and Raithel [31], Genzel and Muser [35], Gorlich [46], Milner and Watts [70], Oxley [83], Piwkowski [94], Ryvkin [104], and Schwarz [109–110].

PbS (chemical type)—Frank [30], Kicinski [55], and Pick [92].

3. PbTe—Bode and Levinstein [6], Moss [74], [77–78], and Simpson and Sutherland [114].
4. PbSe—Checinska [20], and Moss [77], [80].

In addition to the true photoconducting layers, it is possible to observe photovoltaic effects in all these types of materials and, hence, to make cells which generate voltages under the influence of IR radiation. Such effects or detectors have been discussed by the following:

1. PbS—Bose [7], Fischer *et al.* [29], Gibson [42], Gorlich [45–46], Hiller and Smolczyk [51], Lange [59], Lawrance [62], and Mitchell and Goldberg [72].
2. PbTe—Gibson [40].
3. PbSe—Gibson [41].

Some of the commercial firms manufacturing various types of cell are: General Electric Co. (USA); Continental Electric Co. (USA); Scientific Specialities Corp. (USA); Eastman Kodak Co. (USA); British Thompson Houston (England); Mullard (England); Plessey (England); Zeiss (Germany); Jean Turck (France).

Photographs of a variety of several types of cell are shown by Paulson [87].

Characteristics of Lead Salt Detectors

When examined under the electron microscope, the sensitive layers of these detectors are seen to consist of close-packed aggregates of microcrystals of sizes mainly between 1 and 0.1μ. Similar values for crystallite size have been deduced by Wilman [130] from electron diffraction measurements. A recent Photoswitch Report[1] shows that in chemical PbS layers the crystallites are about 0.5μ diameter before oxidation and 0.1μ after oxidation. The structure of PbS films has been studied by Doughty *et al.* [25] and the presence of lanarkite ($PbO \cdot PbSO_4$) on the surface of oxidized layers has been established by Roth *et al.* [102].

Uncooled PbS cells usually have a resistance of about $10^6\ \Omega$ if the sensitive layer is roughly square. Alternative electrode configurations may be used which give

[1] Report 54-1 of July, 1954, published at the Atlantic City Conference on Photoconductivity, November, 1954.

lower resistance; *e.g.*, slit-type cells where electrode separation is much less than electrode size. Such cells are widely used in spectroscopy. Uncooled PbSe cells are ten to fifty times lower in resistance than corresponding PbS types. Cooled PbSe and PbTe cells have resistances of many megohms. Often the resistance is so high that considerable care is needed to avoid leakage currents, particularly as moisture condensation is encouraged by the cooling procedure.

Cells are normally operated with a steady polarizing voltage applied via a series load resistor. For polarizing voltages between about 1 and 100 volts, the cell response is directly proportional to this voltage. The load resistor should preferably match the cell resistance approximately. In the case of high resistance cells the resistance is made as large as possible. In order to avoid excess noise when a current is passed, the load must be a wirewound resistance, and this fact limits the maximum value to about $2\times10^6\ \Omega$. Interrupted radiation is used, the ac voltage generated across the cell being passed via a condenser (which blocks off the polarizing voltage) to an amplifier tuned to the frequency of interruption. To ensure reaching the fundamental noise limits of the cell an amplifier with a gain $\sim10^7$ is needed. Such an amplifier is described by Brown [11].

Spectral Distribution of Photosensitivity

The most important characteristic of these detectors is the spectral response curve which shows the range of wavelengths over which they are useful detectors. These data are shown for the three types of cell at various temperatures between room temperature and 20°K in Figs. 3, 4 and 5.[2] It will be noted that at long wavelengths the sensitivity falls rapidly. As the fall is in general linear (on the logarithmic plot) it is clear that there is no absolute wavelength limit to the sensitivity, and some suitable criterion must be found for defining the range of sensitivity. That proposed by the author [76] is the wavelength where the sensitivity has fallen to 50 per cent of its maximum value, this point being designated $\lambda_{1/2}$. From the curves the values of this characteristic wavelength (in microns) are seen to be:

	295°K	195°K	90°K	20°K
PbS	2.9	3.3	3.8	4.1
PbTe	3.9		5.1	5.9
PbSe	5.0		7.1	8.2

It will be noted that there is a gradual fall in sensitivity at short wavelengths. This is mainly due to the fact that the curves are plotted on an equi-energy basis. If plotted on an equi-quantum basis, they show substantially constant sensitivity over a wide wavelength range, indicating that the quantum efficiency of the photoprocess is constant over this range.

All three types of cell show an increase in the long wavelength limit of sensitivity on cooling. The same temperature dependence is obtained for all three materials provided that the shift is expressed in terms of

[2] Measurements made by the author.

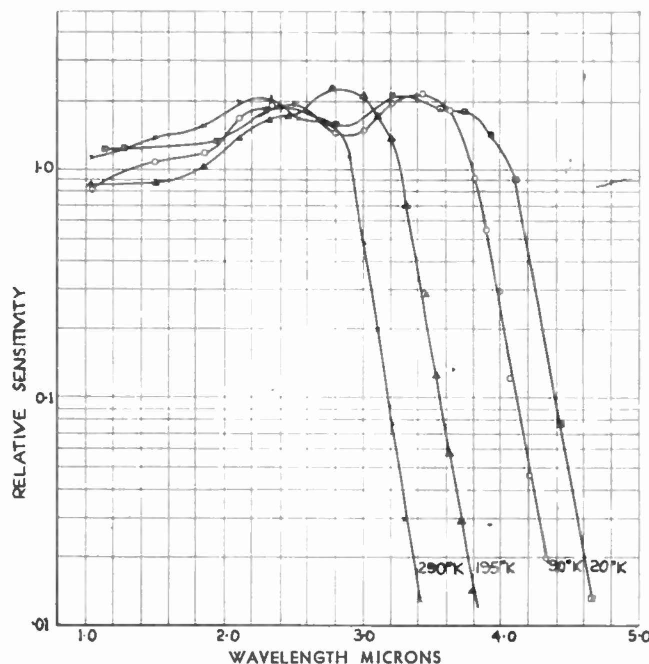


Fig. 3—Spectral sensitivity of a lead sulphide all at various temperatures.

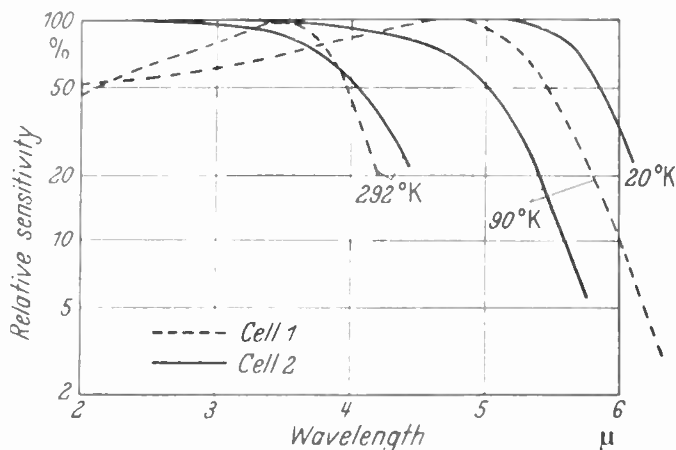


Fig. 4—Spectral sensitivity of lead telluride cells.

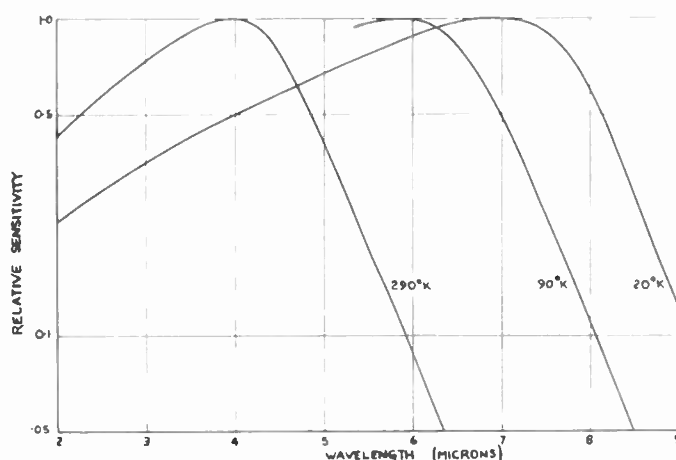


Fig. 5—Spectral sensitivity of lead selenide cell.

energy rather than wavelength, the value being 4×10^{-4}ev/°C. The curves have all been normalized to unity for the sake of convenience; in practice, the sensitivity varies considerably with temperature, particularly for PbTe where the room temperature sensitivity is usually immeasurably small.

Spectral sensitivity curves for PbS cells of various types which agree substantially with those of Fig. 3 have been published by various workers, notably: Fischer *et al.* [29], Sutherland *et al.* [125], Oxley [83], Moss [73], Sosnowski *et al.* [121], Lovell [64], Pick [92], Chasmar and Gibson [18], Gorlich [46], Kolomietz [56] Milner and Watts [70], Genzel and Muser [35], Mitchell and Goldberg [72], Mahlman *et al.* [68], Paulson [87].

Spectral data on PbTe cells have been published by Moss [74] and [77], Gibson [40], Lovell [64], Simpson *et al.* [115], Smith [116], Bode and Levinstein [6], and Clark and Cashman [22]. Some variation of the $\lambda_{1/2}$ values is observed depending on the processing, particularly on the degree of oxidation, of the layers. At 90°K optimum processing gives $\lambda_{1/2} \sim 5.5\mu$.

In the case of PbSe, spectral sensitivity curves by early workers in the field (Roth [102], Blackwell *et al.* [5], Moss and Chasmar [80], Starkiewicz [123], and Milner and Watts [71]) all indicated that the $\lambda_{1/2}$ value was less than for PbTe. More recent work by Gibson *et al.* [43], Moss [77], Gibson [41], and Roberts and Young [101] have shown that if the layers are correctly prepared, results similar to those of Fig. 5 can be obtained.

In addition, most of the firms listed as producing cells publish literature on their spectral characteristics.

Response Time and Frequency Response

It is an important characteristic of these detectors that the response time is very short compared with conventional thermal detectors. For PbTe and PbSe cells the response times are usually $<10^{-5}$ seconds even when the layers are operated at liquid air temperatures. For PbS cells at room temperature are usually $10^{-4} > \tau 10^{5}$ seconds; on cooling with solid CO_2 or liquid air the response time increases, occasionally to as much as 10^{-2} seconds.

The fundamental process of absorption of the radiation quantum and production of a photoelectron (or electron-hole pair) takes a negligible time, perhaps 10^{-14} seconds. However, a longer time is required to build up an equilibrium concentration of photoelectrons in the material. After the application of a steady level of radiation the continuous production of electrons builds up the carrier concentration until the rate of recombination of the photoelectrons (with photo-holes or trapping centers) equals the generation rate.

A simple statistical picture of the process may be given in the following way. Consider unit volume of photoconductor containing M trapping centers which is irradiated so that Q photoelectrons and photo-holes are produced per second. Then the equations for the build up of electron concentration (n) is

$$dn/dt = Q - Bn(n + M), \tag{1}$$

where B is the recombination coefficient. This equation represents the build up of the signal since the photocurrent for a given applied electric field is proportional to n. Integrating (1) and restricting its application to small signal conditions (*i.e.*, $n \ll M$) we obtain

$$n/n_\infty = 1 - \exp - \{B(M + 2n)t\}, \tag{2}$$

where n_∞ is the carrier concentration after infinite time. Thus the response time, defined as that required for the signal to reach $1 - 1/e$ of its saturation value, is given by $\tau = 1/B(M + 2n_\infty)$ or

$$\tau = 1/BM \tag{3}$$

for small signals.

Rough estimates of the values of τ to be expected may be made by estimating B and M. Now $B = cu$ where u is the average velocity of the photoelectrons and c is the cross section for capture. Usually c is of the order of lattice dimensions; *i.e.*, $\sim$1A° square, and u is generally the thermal velocity of quasifree electrons at room temperature, $\sim 10^7$ cm per second. Hence for densities of centers $M = 10^{12}$ to 10^{18} per cm³, the estimated τ values are 10^{-3} to 10^{-9} seconds. These values represent fairly well the range of lifetimes encountered in uncooled photoconductive layers and single crystals of the lead salts.

From (1) it will be observed that the saturation signal is given by

$$n_\infty = Q/B(M + n_\infty) = Q/BM$$

for small signals. Hence

$$n_\infty = Q\tau. \tag{4}$$

This is a general relation obtained equally from more complex theories of photoconductivity as well as from this simple treatment, namely that the photosensitivity is proportional to the response time and also to the intensity of irradiation, provided only relatively small signals are considered. More detailed theories of photoresponse are given by Pick [92], Hepner [50], Moss [76], and Gibson [38].

It is found, in general, that the signal generated by a cell by radiation interrupted at a frequency f is given simply by

$$S_f = S_0(1 + 4\pi^2\tau^2 f^2)^{-1/2} \tag{5}$$

although for some cells where two time-constants are found (see Gibson [38] on PbS, Scanlon *et al.* [108] on PbTe) the relation is naturally more complex. For modern high sensitivity evaporated PbS cells, response times lie within a factor of 3:1 of 50 μsec, and the corresponding frequency response curve is down to half its low-frequency value for a chopping rate $\sim$3,000 cps. For high sensitivity chemical layers, such as those produced by Eastman Kodak Co. (USA), time constants are longer, usually several hundreds of microseconds and the signal may be down to half-value at 300 cps.

For PbTe and PbSe cells, the primary time constants of correctly sensitized cells are so short that they are difficult to measure. For the same reason frequncy response curves are seldom plotted, as the cutoff frequency is usually too high to be generated conveniently. The scanty data available indicates that cooled PbSe and PbTe cells show little fall in sensitivity up to 20 kc, and should generally be usable up to 10^5 cps. For a PbTe cell Moss [77] quotes a measured time constant of 5 μsec. For uncooled PbSe cells, values of 0.5–1.5 μsec have been observed by the author.

As indicated by (4) it is to be expected that the signal from a photoconductor will be proportional to the response time. A general correlation of this type has been reported for a large number of PbS cells by McAlister (see Jones [54]). Specifically, McAlister found that the product of the sensitivity, measured on a signal/noise basis, and the response time was constant within a factor of 3 for cells with a range of time constants between 3 and 3,000 μsec. The correlation was equally good if the response time of a particular cell was changed by cooling as if it was varied from cell to cell. On this basis, PbS cells are classified by Jones [54] as Class II radiation detectors.

If the cells are subjected to a strong background illumination in addition to the small signal being used to measure τ, then the value of n is effectively increased, and as expected from (3) τ decreases. From (4), it follows of course, that the sensitivity also decreases in the presence of strong background illumination. The influence of strong background radiation has been shown clearly in the work of Pick [92], and Paulson [87].

Cell Sensitivities

The practical sensitivity of a detector is determined equally by the responsivity of the device and by the inevitable noise present in the system; *i.e.*, by the signal/noise ratio. The limit of detection is conventionally specified by the condition that signal = noise. Avoidable sources of noise, such as microfony, hum, pickup, contact noise, etc., are excluded from the present consideration. For photoconductive cells this leaves three main contributions to the noise:

1. Johnson or Nyquist noise—*i.e.*, fundamental noise present in any resistance in thermal equilibrium with its surroundings.
2. Current noise—*i.e.*, additional noise which appears in almost all semiconducting and photoconducting devices when a steady current is passed through them.
3. Radiation noise—it can be shown theoretically that the random arrival of radiation quanta at the sensitive layer generate additional noise, which may be important in cells of the highest sensitivity.

In general Johnson noise is relatively small in cells (or is arranged to be so by the simple expedient of increasing the polarizing current through the cell, thus increasing the signal in proportion, until current noise is predominant). Only in the very best cells is radiation noise significant, so that normally current, or flicker noise is the main type to be considered. A typical noise-frequency distribution for a PbTe cell is shown in Fig. 6. It will be noted that the amount of noise in a given bandwidth increases rapidly with decreasing frequency.

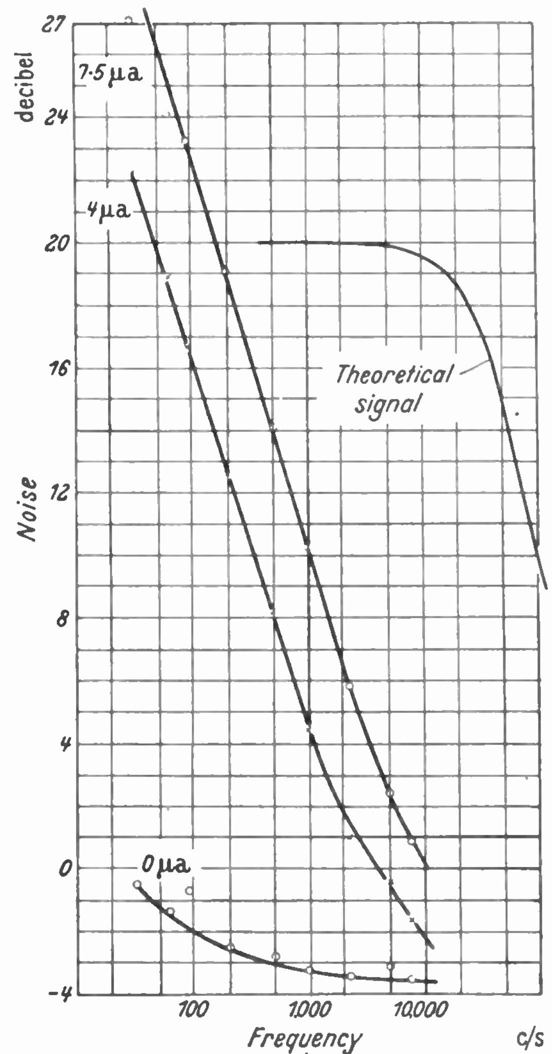


Fig. 6—Noise-frequency curves for lead telluride cell.

It is found that over a wide frequency range the noise power is universely proportional to the frequency, or alternatively the rms noise voltage $N \propto f^{1/2}$. Using the expression for the signal given in (5), we obtain

$$S/N \propto f^{1/2}(1 + 4\pi^2 f^2 \tau^2)^{-1/2}. \qquad (6)$$

This expression has its maximum value at $2\pi f\tau = 1$, so that the optimum frequency of operation of a current-noise limited cell is simply $f = \frac{1}{2}\pi\tau$. For the cell of Fig. 6 this would be about 30,000 cps.

A series of curves for signal, noise and signal/noise as functions of operating frequency and cell current for PbS cells are given by Sutherland and Lee [126], and Paulson [87]. The underlying causes of current noise are not yet well established, but various possibilities have been considered by Macfarlane [66], [67], Petritz and Siegert [91], Gorlich [48], and van der Ziel [127a]. Eckart [27] has shown that the current noise is the same for a given photocurrent as for the same dark current.

As, at present, there is no apparent necessity for the presence of current noise on theoretical grounds, then if only this form of noise was considered, one could visualize detectors of infinite sensitivity. However, the presence of radiation noise prevents this, and it is quite inevitable that this latter type of noise will occur.

The detector is normally surrounded by material at room temperature such as the cell walls, enclosing box, etc. These surfaces will be radiating and, in consequence, the sensitive layer will be subject to a steady stream of room temperature quanta. As the layers are sensitive at relatively long wavelengths, an appreciable fraction of the quanta will be capable of exciting photoelectrons and producing a photocurrent. As there will be statistical fluctuations in the rate of arrival of the quanta, so this photocurrent will fluctuate. These fluctuations are the manifestation of radiation noise. It will be clear that once a detector is sufficiently sensitive to register this radiation noise, any increase in its responsivity will increase the radiation noise as much as it increases the wanted signal, and thus no further increase in signal/noise can be obtained.

The relative proportions of quanta from a black body at 20°C which are effective for various types of cell are:

1. Infrared photo emissive cell (sensitive to $\sim 1.1\mu$) 10^{-15} per cent.
2. PbS cell 10^{-3} per cent.
3. PbTe cell (at 90°K) 0.3 per cent.
4. PbSe cell (at 90°K) 3 per cent.

The presence of this room temperature radiation is shown by the fact that the conductivity of a cooled PbTe layer decreases considerably when the surroundings of the layer are cooled (see Simpson [113]). Also the responsivity of a PbTe cell can be increased by hundreds of times by such cooling (Simpson and Sutherland [114]). Arrangements used for cooling the surroundings of cells are shown in Fig. 7.

Calculations by the author [75] show that for a long wavelength sensitive PbTe layer of 1 mm² cooled with liquid air, the limiting sensitivity imposed by radiation from surroundings at 0°C is such that signal = noise in 1 cps bandwidth should be given by 5×10^{-13} watts of radiation of wavelength 4μ. For a temperature of surroundings of 17°C the corresponding figure would be 7×10^{-13} watt. As half the energy of the radiation is lost in the chopping system, and as there are reflection losses at the cell window and at the layer itself, the best performance which can be expected in practice is to detect about 2×10^{-12} watts incident on the chopper. The limitations to the attainable sensitivity of photoconductive cells have also been considered by Fellgett [28] and Muser [81]. Though there is little published data on limiting sensitivities of PbTe cells, it is clear from the high resolution spectroscopy which has been carried out in recent years with these cells (see Thompson and Williams [127], Boyd and Thompson [8], Callomon and Thompson [14]) that sensitivities of this order are being obtained. Fellgett has measured a PbTe cell which was only a factor of 1.9 worse than the theoretical limit imposed by radiation noise, and he has recently stated[3] that the best modern PbTe cells are only 1.5:1 worse than the radiation limit.

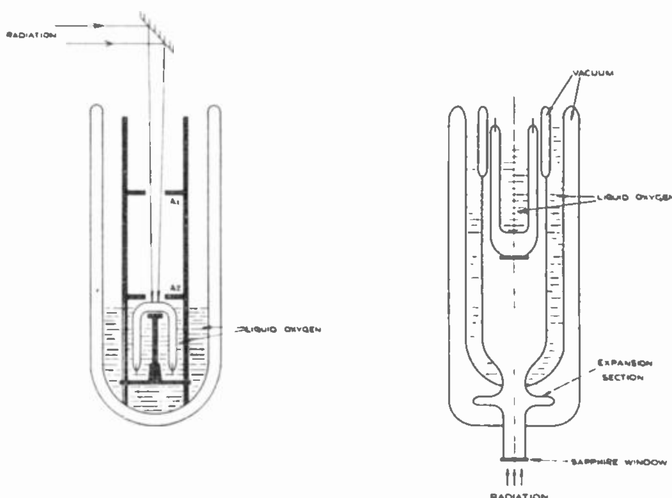


Fig. 7—Arrangements for screening radiation from cells.

For PbS layers 1 mm² square, Moss [75] has calculated the limiting sensitivity to be 5×10^{-14} watts for a layer at 0°C in surroundings at 17°C, and 8 to 15×10^{-14} watts for layers (with different spectral response curves) at 90°K in surroundings at 17°C. Fellgett[4] estimates that the best PbS cells, when cooled, are within 1.5:1 of the radiation limit; and about 20:1 above this limit when operated at room temperature. Figures of merit for all types of cell have been calculated by Jones [54].

Milner and Watts [70] quote a limiting sensitivity of 2×10^{-11} watts of 2.2μ radiation for an uncooled PbS cell of 0.1 cm² area, and 4×10^{-12} watts at solid CO_2 temperature. Advertised data for Eastman Kodak cells of this area give a sensitivity of 3×10^{-8} watts of black-body radiation of 500°K. Analysis of data given by Sutherland and Lee [126] shows that at optimum cell current and chopping frequency the best uncooled cell (which was probably of ~ 0.1 cm² area) would detect 6×10^{-8} watts for 1 cps bandwidth at 800 cps. Assuming the cell to have constant quantum sensitivity for all wavelengths up to 2.8μ gives an equivalent limiting sensitivity of 10^{-11} watts for 2.5μ radiation. On this same basis the limiting sensitivity of the best cell quoted by Lovell [64] is 8×10^{-12} watts at room temperature and 3×10^{-12} watts at -80°C. For a photovoltaic cell Lawrance [62] finds a sensitivity of 8×10^{-11} watts at 9 cps chopping frequency. Watts [129] found that for a liquid air cooled layer of 0.1 cm² area, cooling the surroundings of the cell with liquid air improved the sensitivity by a factor of 10–20 to 2×10^{-13} watt. As this figure is better than the radiation limit for a layer exposed to room temperature, it is clear that the sensitivity of this cell was in fact limited by radiation noise. Petritz [88] has shown that the noise from such a radiation lim-

[3] P. Fellgett, Lecture presented at the Institute of Physics, London, England; September, 1954.
[4] Fellgett, *loc. cit.*

ited cell should have the same frequency variation as the signal, and by observing this type of behavior in a cooled cell Lummis and Petritz [65] have concluded that the cell was "seeing" radiation noise. From a consideration of actual cell time constant and theoretical radiation lifetimes, Petritz [89] has estimated that PbS cells are about a factor of 2:1 worse than the radiation limit. Carlisle and Alderton [15] find that, near room temperature, cell sensitivity varies by about 5 per cent per °C, and McFee [69] has observed that for PbS cells operated at 90 cps chopping frequency, optimum signal/noise is obtained at a temperature of about 180°K.

Information on the sensitivity of PbSe cells is meager. For a cooled cell Roberts and Young [101] found the sensitivity to be about 6 times better than a high sensitivity Schwarz thermopile over the wavelength range of 2–6μ. As such a thermopile will have a sensitivity 2×10^{-10} watts (see Gebbie *et al.* [34]) the limiting sensitivity of the cell should be $\sim 3\times10^{-11}$ watt. For an uncooled cell Starkiewicz [123] quotes a minimum detectable energy of 10^{-8} watts for a layer 10 mm^2 using 30 cps bandwidth. The corresponding figure for 1 mm^2 and 1 cps bandwidth would be 6×10^{-10} watts. Checinska [20] gives a detection limit of 5×10^{-8} watt for 1 cm^2 area. The dependence of the signal/noise ratio on the source temperature has been studied by Crooker and Dorling [23].

Fundamental Properties

The lead salts are semiconductors which have a relatively low concentration of free current carriers, about 10^{-4} or 10^{-7} times less than in a typical metal at room temperature. Ideally the materials would be insulators at very low temperatures. However, the density of carriers *potentially* available is similar to that in a metal, and provided these can be liberated from their bound state they can move freely through the lattice and give conductivity. Freeing of these carriers by thermal energy gives semiconduction, by radiant energy gives photoconduction.

Absorption

For photoconductivity to occur at a given wavelength the prime requirement is for the material to be absorbing at that wavelength. Absorption data obtained by direct transmission measurements on single crystals of these materials have been given by Paul *et al.* [86], Gibson [41], Clark and Cashman [22], and Paul and Jones [85], and from analysis of reflection measurements with polarized light to obtain both refractive index and absorption index values by Avery [1–3]. The most characteristic feature of the results is a clearly defined absorption edge which corresponds well with the $\lambda_{1/2}$ values for photosensitive layers at any corresponding temperatures. The positions of the three absorption edges at room temperature are shown in Fig. 8. The presence of similar absorption edges in thin layers has been established for PbS by Vernier [128] and for PbTe by Lasser and Levinstein [61], although early work by Gibson [37] showed no sign of such edges.

The fall in photosensitivity at long wavelengths thus arises from the fall in absorption, which in turn is consequent on the fact that the radiation quanta at such long wavelengths have insufficient energy to free photoelectrons.

Activation Energies

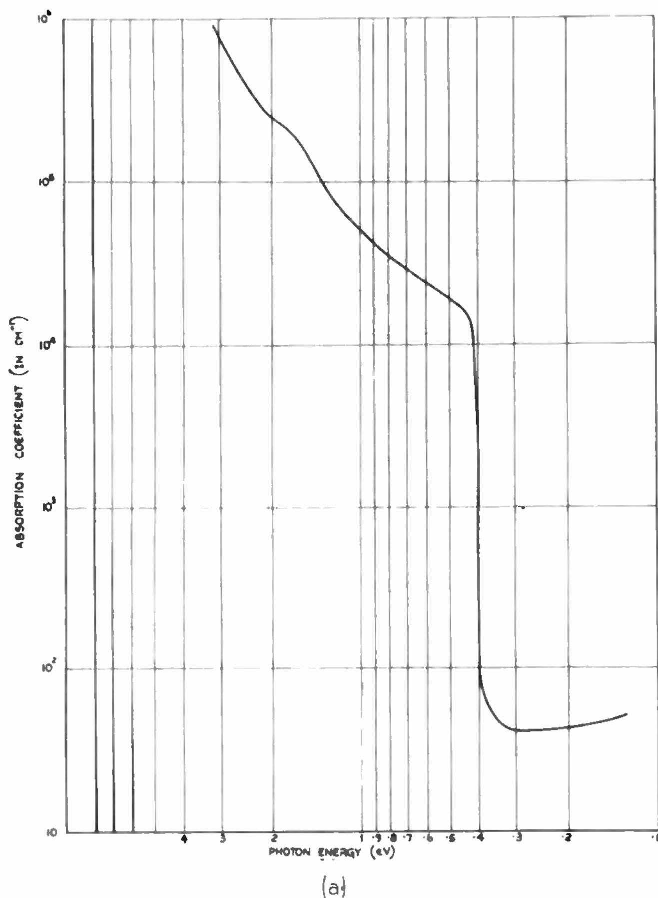
Many of the properties of semiconductors and photoconductors can be explained in a qualitative way from the energy band diagram for the materials. When atoms are "put together" to form a solid, the originally sharp allowed energy levels of the atoms are broadened by mutual interactions in *bands* of allowed levels separated by forbidden zones. If, as may well happen, there are just sufficient electrons available in a crystal to occupy *all* the states in a certain number of the lower bands (leaving the next higher band empty), the material is basically a nonconductor. No net current can be carried by electrons in a full band, because to carry a current in an applied field electrons must be accelerated by the field; *i.e.*, their energy must increase slightly. As there are no slightly higher energy states available to the electrons it follows that no current will flow. However, if by use of thermal or optical energies an electron is excited across the forbidden zone into the empty band above, it may be freely accelerated by an applied field and so gives conductivity. It should be noted that the "positive hole" left in the full band by the removal of the electron enables conductivity to occur in this band by the process of successive electrons moving into the vacancy. In general this positive hole is almost as mobile a current carrier as the electron, and this hole current adds to the electron current, although the hole itself moves in the opposite direction to the electron.

The width of this forbidden zone is a most important characteristic of a semiconductor, and it is termed the activation energy of the material. It is determined with least possible ambiguity (but not necessarily with highest accuracy) from the position of the absorption edge in single crystals, and from the $\lambda_{1/2}$ values of the photosensitivity curves. The best present day values for the activation energy for the three materials at 20°C are considered by the author to be:

PbS 0.40 ev
PbTe 0.31 ev
PbSe 0.25 ev.

All values decrease by about 4×10^{-4} ev per °C of cooling. Rather lower values are quoted by Smith [117].

It is also possible to measure this activation energy



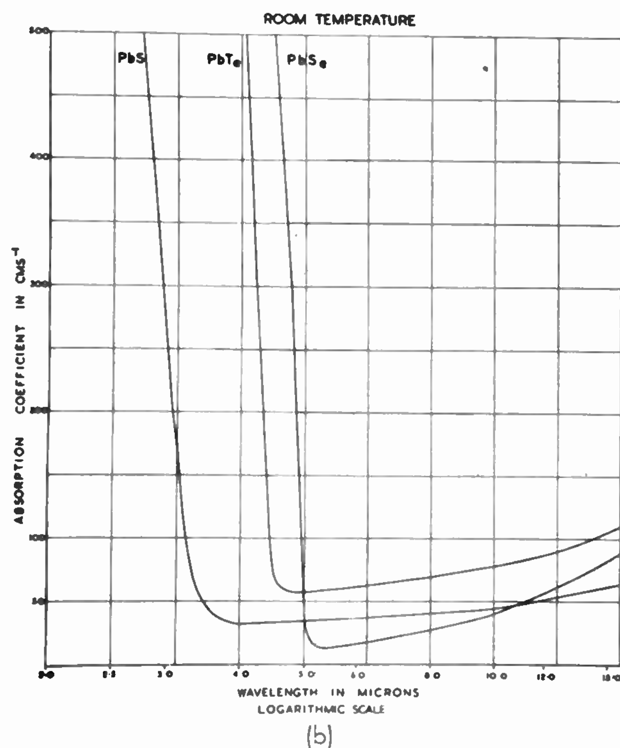


Fig. 8—(a) Absorption coefficient/energy of radiation for PbS at room temperature. (b) Absorption spectra of lead sulphide selenide and telluride crystals; room temperature (data from Gibson).

by thermal means by using the Hall effect, for example, to determine the temperature dependence of the carrier concentration in the unilluminated material. In the past very high values of activation energy have been deduced from such data (Putley and Arthur [97], Putley [95], and Smith [116]). There is now no doubt that these high values are in error, and the recent determination by Scanlon [105] of the activation energy of PbS from thermal measurements of 0.37 ev is in good agreement with the optically determined values above. Using the photoelectromagnetic effects it has recently been possible to observe photoeffects in single crystals of PbS (Moss [78–79]). These results lead to an activation energy at room temperature of 0.41 ev. As, at the same time, these results show the quantum efficiency of the process to be near unity, it is established that the effect cannot be caused by impurities but arises from intrinsic band-to-band excitation. Photoconductivity in bulk PbS has also been observed by Soule and Cashman [122], and spectral sensitivity measurements on a *p-n* junction made by Mitchell and Goldberg [72] confirm an activation energy of 0.4 ev.

An alternative method of estimating the thermal activation energy from rectification characteristics of natural PbS *p-n* junctions has now been put forward by the author.[5] This gives a room temperature energy of 0.40 ev. An attempt to calculate the band structure of PbS has been made by Bell *et al.* [4].

Carrier Lifetimes

Carrier lifetimes in bulk PbS have been estimated in a variety of ways for PbS crystals by Moss [77]. Using the photo-electro-magnetic effect, values between 6×10^{-10} and 9×10^{-6} seconds have been calculated. The larger values were confirmed by diffusion length, and by direct pulse response, measurements. It was observed in this work that the response time varied inversely as the square of the carrier concentration, implying that the Auger effect was the predominant recombination mechanism. On purely theoretical grounds Pincherle [93] has shown that this is to be expected.

The maximum lifetime possible in the three materials has now been calculated on the basis of radiative recombination by Mackintosh[6] who gies the following values: PbS, 40 μsec; PbTe, 0.8 μsec; PbSe, 0.6 μsec. That this theory is still somewhat tentative is shown by the fact that Petritz [89] obtains: PbS, 2,800 μsec.; PbTe, 310 μsec.

The actual recombination radiation from PbS has been detected by Galkin and Korolev [32] and by Garlick and Dumbleton [33]. It has its peak intensity near the absorption edge.

[5] T. S. Moss, "Measurements on *p-n* junctions of PbS," *Proc. Phys. Soc.*, at press.

[6] Unpublished work, quoted in reference [118]. At press for *Proc. Phys. Soc.*

Carrier Mobilities and Effective Masses

Data obtained by Putley [95] have been analyzed by Macfarlane and Pincherle[7] to give the following average room temperature values:

	PbS	PbTe	PbSe
Electrons	~600	1200	1175 cm^2/volt sec
Positive Holes	~250	475	868

For PbS, Moss [77] has used an average *sum* of electron and hole mobilities of 800 cm^2/volt sec, which is similar to the value given by the recent data of Scanlon and Brebrick [106], and Petritz and Scanlon [90]. These latter workers find an electron/hole mobility ratio for PbS of 1.4. For PbSe, Hirahara and Murakami [52] deduce a mobility ratio of 1.4. For all three materials the analysis of Macfarlane and Pincherle[7] shows a $(\text{temperature})^{-5/2}$ dependence of electron mobility. For holes the index is somewhat higher. For PbSe, Devyatkova *et al.* [24] finds a $(\text{temperature})^{-3}$ mobility law.

Macfarlane and Pincherle deduce the following values for the mean mass product; *i.e.*,

$$\frac{(\text{electron mass} \times \text{holes mass})^{1/2}}{\text{free electron mass}} : \text{PbS } 0.34, \text{ PbTe } 0.20\text{–}0.25, \text{ PbSe } 0.31.$$

Putley [96] gives values of 0.2 to 0.36 for PbSe. From fundamental quantum-mechanical theory Bell *et al.* [4] estimate the mass of holes in PbS to be ~0.5 free electron masses. Scanlon *et al.* [107] deduce an effective electron mass of 0.29 of a free electron for PbS from analysis of mobility data.

Theory of Photoeffects in Layers

Two theories of the photoconductive mechanism have been proposed (see Moss [76]): (I) Single crystal, recombination theory. (II) Barrier modulation theory.

In the first case the equilibrium increase in carrier density is determined by rates of generation and recombination of photoelectrons and photo-holes, and the photo-current is directly proportional to this carrier increase, the mobility of the photo-carriers being the same as those carrying the dark current. On the barrier theory it is postulated that rectifying contacts are set up between crystallites during the processing of the layer, and that generation of a few photoelectrons or holes in the immediate neighborhood of these space charge barriers so lowers the barrier height that relatively large numbers of either electrons or holes can cross the barrier. Possibilities of discriminating between these two theories have been discussed by Gibson [38], and Muser [81].

There are two essential differences in the consequences of these theories.

1. In Theory (I) the *numbers* of carriers increase, mobility remaining constant; while in Theory (II) numbers change little and effective *mobility* increases.

2. In Theory (I) the quantum efficiency will be near unit, but never greater, while in Theory (II) only a small proportion of absorbed quanta (*i.e.*, those in the barrier regions) will produce useful photoelectrons, but these will permit the flow of many electrons across the photoconductor. In this case the *apparent* quantum efficiency can be much greater than unity, although the primary quantum efficiency will be less than unity. In either theory it is now postulated that initial photoeffect is a band-to-band transition producing a free electron and a free hole.

[7] Unpublished work quoted in reference [118].

It was initially thought that the measurement of the resistance of layers at high frequencies would show the presence of intercrystallite capacities and hence give information on the presence of barriers. Such measurements have been made by Chasmar [17], Rittner and Grace [100], Humphrey *et al.* [53], and Broudy and Levinstein [10]. The latter workers, however, now conclude that such measurements can neither prove or disprove the presence of barriers.

For cooled PbTe layers the author (see Moss [77]) has found that the numbers theory gives the correct magnitude for the photoresponse with quantum efficiencies near—but less than—unity. Simpson and Sutherland [114] in their extensive study of conductivity and photoconductivity concluded that there were no potential barriers in the layers. Bode and Levinstein [6] postulate trapping centers in their single crystal theory, but do not regard potential barirers as necessities. Valuable information has recently been obtained from Hall effect studies by Levy [63], who has shown that under strong illumination an increase in carrier numbers of as much as 25:1 can be obtained with a variation in mobility of less than 2:1. The mobilities were in the range 0.2 to 2 cm^2/volt sec, and sometimes *decreased* slightly on illumination. Silverman and Levinstein [111] found film mobilities in PbSe and PbTe as great as $\frac{1}{5}$ of those in bulk material in annealed films, but the values varied markedly with processing. Studies of the interrelation of response time, resistivity and sensitivity of PbSe layers have been made by Sosnowski and Chmielewski [119] and for PbS layers by Lashkarev *et al.* [60], and Ryvkin [104].

Petritz [89] points out that any gain in apparent quantum efficiency resulting from a barrier modulation system cannot improve the limiting sensitivity of a cell, because this limit is set by fluctuations in background radiation which will be magnified just as the wanted signal is.

In order to obtain the maximum sensitivity when limited by radiation noise it is necessary for *each* absorbed photon to produce a photoelectron. This is unlikely to occur in barrier model system since only photons absorbed near the barriers are effective. The sensitivity figures quoted earlier show that the attainable sensitivity is so near the calculated value that the primary quantum efficiency must be at least 50 per cent, which tends to favor the no-barrier model.

The recent treatment by Rittner [99] may prove the best compromise. He postulates that the sensitizing has the function of compensating exactly the p and n type

impurities, and that when this occurs a space-change layer extends throughout the material. The whole layer then takes on quasi-intrinsic properties, and the internal barriers are photo-insensitive and merely serve to reduce mobility to low values observed in layers.

Summarizing the present state of the theory of photoconductive layers, it is the opinion of the author that the following conclusions may be reached:

1. Potential barriers are not theoretically necessary to achieve the sensitivities actually attained in the best cells so far; *i.e.*, to reach the theoretical radiation noise limit. In fact they are probably a hindrance to this goal.

2. The majority of recent evidence on cooled PbTe layers supports the numbers theory, with barrier effects of minor importance.

3. In highly oxidized uncooled PbS cells, barriers may well be necessary to explain the resistance/temperature characteristics (see Mahlman *et al.* [68]), and they are clearly present to a sufficient degree to give detectable intercrystallite photovoltaic effects, (Sosnowski *et al.* [121] and Dutton [26]), but even so they are probably not of paramount importance in the production of photosensitivity [118].

Applications

The two most important properties of these detectors which govern their applications are their high sensitivities over the spectral range 1–7μ and their rapid response times of a few microseconds. Within this spectral region their superiority in both respects over conventional heat detectors renders them the best detectors currently available for many applications.

As a guide to the usefulness of the cells in detecting heat radiation, the following gives the percentage of the power radiated by a black-body source, which is effective in operating each type of cell:

Source Temperature 0°3	PbS at 20°C	PbTe at 90°K	PbSe at 90°K
300	2.5	25	48
150	0.25	8	26
50	0.01	2	11
20	0.004	1	7
0	0.001	0.6	5

Radiation pyrometers using uncooled PbS cells have been described by Milner and Watts [70], Gibson [38], Pyatt [98], Bracewell [9] and Harman and Watts [49]. In the first system the radiation is balanced against a standard radiation source inside the pyrometer, while the second is essentially a "two-color" system as it measures the ratio of the radiant intensities in two wavelength bands.

As these pyrometers have been usable down to 150°C (with accuracy ±4°C) one can conclude from the above table that using a cooled PbTe cell one could do pyrometry down to 50°C or even 20°C, while with cooled PbSe cells, 0°C and below become measurable. Milner and Watts [70] have also described the use of such cells as heat sensitive relays for use in control or warning systems, while Cashman [16] has discussed their use for reproduction of sound from film tracks.

The most striking application of the cells so far has been in infrared spectroscopy. Making use of the high sensitivity of cooled PbTe cells, Boyd and Thompson [8] and Thompson and Williams [127] have achieved resolution of 0.5 cm^{-1} in the 5 region. More recent results by these and other workers (see Callomon and Thompson [14]) show the resolution in this region to be <0.15 cm^{-1}. Similar resolution has long been obtainable (at shorter wavelengths) using PbS cells (see Sutherland *et al.* [125]).

With a cooled PbSe cell Roberts and Young [101] have obtained high resolution in the 5–7μ region using a small prism spectrometer. Fig. 9 shows the results

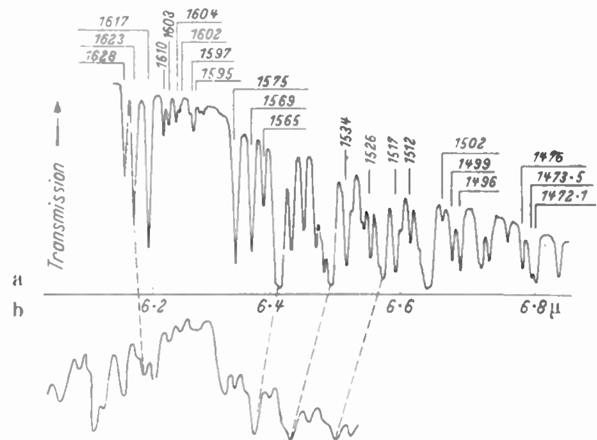


Fig. 9—Six micron water vapor band taken with a lead selenide detector at 90°K. (a) Chart recording. (b) Taken at 50 scans/sec.

thus obtainable. This figure also shows how the rapid response time of cells may be utilized to do high speed spectra. The system used was similar to that described by Brown and Roberts [12] and it is seen that resolution of 6 cm^{-1} can be obtained at a scanning rate of 50 spectra/sec. Bullock and Silverman [13] have also made a high speed spectrometer, with which they have been able to resolve the 4.3μ CO_2 doublet at 120 spectra/sec, and they have shown that such an instrument can be used to study explosions and gas discharges.

The rapid response time may be advantageously employed to study thermal transients such as generated in high speed braking (Parker and Marshall [84]). Lovell [64] describes the use of PbS cells by the German Army to detect ships and aircraft, and in speech transmission systems, while Kuiper [58] has used such cells for astronomical measurements.

Acknowledgment

Acknowledgment is due to the Chief Scientist, British Ministry of Supply, for permission to publish this article, to Drs. Mackintosh and Pincherle for use of their information prior to publication, to the Physical Society of London, for permission to reproduce Fig. 3, and to *Research* for permission to reproduce Figs. 4, 6, and 9.

Bibliography

[1] Avery, D. G., "The Optical Constants of PbS and PbTe in the Region 0.5–3μ." *Proceedings of the Physical Society*, Vol. B64 (1951), p. 1087.
[2] Avery, D. G., "The Optical Constants of PbS, PbSe and PbTe in the 0.5–3μ Region of the Spectrum." *Proceedings of the Physical Society*, Vol. B66 (1953), p. 134.
[3] Avery, D. G., "Further Measurements on the Optical Properties of PbS, PbSe and PbTe." *Proceedings of the Physical Society*, Vol. B67 (1954), p. 2.
[4] Bell, D. G., Hum, D. M., Pincherle, L., Sciama, D. W., and Woodward, P. M., "The Electronic Band Structure of PbS." *Proceedings of the Royal Society*, Vol. A217 (1953), p. 71.
[5] Blackwell, D. E., Simpson, O., and Sutherland, G. B. B. M., "PbSe Cells for Infra-Red Spectroscopy." *Nature* (*London*), Vol. 160 (1947), p. 793.
[6] Bode, D. E., and Levinstein, H., "Effect of Oxygen on the Optical Properties of PbTe Films." *Physical Review*, Vol. 96 (1954), p. 259.
[7] Bose, J. C., U. S. Patent No. 755840, 1904.
[8] Boyd, D. R. J., and Thompson, H. W., "Infra-red Spectrum of Methylacetylene," *Transactions of the Faraday Society*, Vol. 48 (1952), p. 493.
[9] Bracewell, R. A., "An Infra-red Radiation Pyrometer." *Electronic Engineering*, Vol. 27 (1955), p. 238.
[10] Broudy, R., and Levinstein, H., "High Frequency Resistance of Thin Films, and High Frequency Resistance of Photoconducting Films." *Physical Review*, Vol. 94 (1954), pp. 285, 290.
[11] Brown, D. A. H., "A Photocell Amplifier for Infra-red Spectroscopy." *Journal of Scientific Instruments*, Vol. 29 (1952), p. 292.
[12] Brown, D. A. H., and Roberts, V., "A Simple High Speed Spectrometer for the Infra-red." *Journal of Scientific Instruments*, Vol. 30 (1953), p. 5.
[13] Bullock, B. W., and Silverman, S. J., "A Rapid Scanning Spectrometer for Oscillographic Presentation in the Near Infra-red." *Journal of the Optical Society of America*, Vol. 40 (1950), p. 608.
[14] Callomon, H. J., and Thompson, H. W., "Vibration-Rotation Bands and Molecular Constants of Carbonyl Sulphide." *Proceedings of the Royal Society*, Vol. 222 (1954), p. 431.
[15] Carlisle, S. S., and Alderton, G., "Temperature Coefficient of Sensitivity of PbS Photoconductive Cells at Room Temperature." *Nature* (*London*), Vol. 163 (1949), p. 529.
[16] Cashman, R. J., "New Photoconductive Cells: and Photodetectors for U.V., Visible and I.R. Radiation." *Journal of the Optical Society of America*, Vol. 36 (1946), p. 356. Also in the *Proceedings of the National Electronics Conference*, Vol. 2 (1946), p. 171.
[17] Chasmar, R. P., "High Frequency Characteristics of PbS and PbSe Layers." *Nature* (*London*), Vol. 161 (1948), p. 281.
[18] Chasmar, R. P., and Gibson, A. F., "Long Period Photo-Effects in PbS." *Proceedings of the Physical Society*, Vol. B64 (1951), p. 595.
[19] Chasmar, R. P., Craston, J. L., Isaacs, G., and Young, A. S., "A Method of Sealing Sapphire to Glass and its Application to Infra-red Photocells." *Journal of Scientific Instruments*, Vol. 28 (1951), p. 206.
[20] Checinska, H., "Photoconductive and Photovoltaic Layers of PbSe." Bulletin of the Academy of Science, Polon., Cl. 3, Vol. 1 (1953), p. 123.
[21] Checinska, H., "Photoconductive and Photovoltaic PbSe Layers." *Acta Physica*, Polon., Vol. 12 (1953), p. 194.
[22] Clark, M. A., and Cashman, R. J., "Transmission and Spectral Response of PbS and PbTe." *Physical Review*, Vol. 85 (1952), p. 1043.
[23] Crooker, A. M., and Dorling, E. B., "Dependence of Signal-noise Ratio in PbSe Layers on the Black Body Source Temperature." *Physical Review*, Vol. 96 (1954), p. 846.
[24] Devyatkova, E. D., Maslakovets, Y. P., Stilbans, L. S., and Staviskaya, T. S., "Temperature Dependence of Carrier Mobility in Semiconductors." *Doklady Akad. Nauk. S.S.S.R.*, Vol. 84 (1952), p. 681.
[25] Doughty, J., Lark-Horovitz, K., Roth, L. M., and Shapiro, B., "The Structure of PbS Films." *Physical Review*, Vol. 79 (1950), p. 203.
[26] Dutton, D., "Fine Spot Scanning of PbS Films." Paper read at the Atlantic City Conference on Photoconductivity, November, 1954.
[27] Eckart, F., "Noise Measurements on Thin PbS Layers." *Annalen der Physik*, Vol. 11 (1952), p. 166.
[28] Fellgett, P., "On the Ultimate Sensitivity and Practical Performance of Radiation Detectors." *Journal of the Optical Society of America*, Vol. 39 (1949), p. 970.
[29] Fischer, F., Gudden, B., and Treu, M., "On the PbS Light Detector." *Physikalische Zeitschrift für Physik*, Vol. 39 (1938), p. 127.
[30] Frank, H., "Photoelectric Conductivity of PbS." *Slaboproudy Obzor*, Vol. 14 (1953), p. 243.
[31] Frank, K., and Raithel, D., "Manufacture of Very High Resistance Photosensitive PbS Layers. *Zeitschrift für Physik*, Vol. 126 (1949), p. 389.
[32] Galkin, L. N., and Korolev, N. V., "Infra-red Luminescene Spectrum of PbS." *Doklady Akad. Nauk. S.S.S.R.*, Vol. 92 (1953), p. 529.
[33] Garlick, G. F. J., and Dumbleton, M. J., "Phosphors Emitting Infra-red Radiation." *Proceedings of the Physical Society*, Vol. B67 (1954), p. 442.
[34] Gebbie, H. A., Harding, W. R., Hilsum, C., Pryce, A. W., and Roberts, V., "Atmospheric Transmission in the 1 to 14 Region." *Proceedings of the Royal Society*, Vol. A206 (1951), p. 87.
[35] Genzel, L., and Muser, H., "On the Optical and Photoelectric Properties of a New Type of PbS Layer." *Zeitschrift für Physik*, Vol. 134 (1953), p. 419.
[36] Gibson, A. F., "Optical Sensitization of PbS Layers Made Without Oxygen." *Nature* (*London*), Vol. 163 (1949), p. 321.
[37] Gibson, A. F., "The Absorption Spectra of Solid PbS, PbSe, and PbTe." *Proceedings Physical Society*, Vol. B63 (1950), p. 756.
[38] Gibson, A. F., "Sensitivity and Response Time of PbS Photoconductive Cells." *Proceedings of the Physical Society*, Vol. B64 (1951), p. 603.
[39] Gibson, A. F., "A Two-Colour Infra-red Radiation Pyrometer." *Journal of Scientific Instruments*, Vol. 28 (1951), p. 153.
[40] Gibson, A. F., "Single Contact PbTe Photocells." *Proceedings of the Physical Society*, Vol. B65 (1952a), p. 196.
[41] Gibson, A. F., "The Absorption Spectra of Single Crystals of PbS, PbSe, and PbTe." *Proceedings of the Physical Society*, Vol. B65 (1952b), p. 378.
[42] Gibson, A. F., "PbS Rectifier Photocells." *Proceedings of the Physical Society*, Vol. B65 (1952c), p. 214.
[43] Gibson, A. F., Lawson, W. D., and Moss, T. S., "The Long Wave Limit of Infra-red Photoconductivity in PbSe. *Proceedings of the Physical Society*, Vol. A64 (1951), p. 1054.
[44] Goldberg, G. R., and Mitchell, A. E., "Occurrence of Natural *P-N* Junctions in PbSe." *Physical Review*, Vol. 93 (1953), p. 1421.
[45] Gorlich, P., "Influences of Gases on the Properties of PbS Layers." *Zeitschrift für Naturforschung*, Vol. 2a (1947), p. 47.
[46] Gorlich, P., "The Application of Photo-Resistors." *Zeitschrift für Naturforschung*, Vol. 5a (1950), p. 563.
[47] Gorlich, P., *Die Photosellen*, Leipzig, Germany, Akademische Verlagsgesellschaft Geest and Portig, 1951a.
[48] Gorlich, P., "On the Question of Noise in Photo-sensitive Semiconductors." *Optik*, Vol. 8 (1951b), p. 512.
[49] Harman, J. D., and Watts, B. N., "An Infra-red Radiation Pyrometer." *Journal of Scientific Instruments*, Vol. 32 (1955), p. 167.
[50] Hepner, W. A., "Response of Photoconductive PbS Cells." *Nature, London*, Vol. 159 (1947), p. 96.
[51] Hiller, J. E., and Smolczyk, H. G., "Antimony Content and Photoelectric Effect in Different PbS Specimens." *Naturwissenschaften*, Vol. 39 (1952), p 208.
[52] Hirahara, E., and Murakami, M., "Hall Effect and Electrical Conductivity of PbSe." *Journal of the Physical Society of Japan*, Vol. 9 (1954), p. 671.
[53] Humphrey, J. N., Lummis, F. L., and Scanlon, W. W., "Capacitance Effects in Thin Conductive Films." *Physical Review*, Vol. 90 (1953), p. 111.
[54] Jones, R. C., "Performance of Detectors for Visible and Infrared radiation, and the Relation Between the Speed of Response and the Detectivity of PbS Cells." *Advances in Electronics*, Vol. 5 (1953), p. 2. Also in *Journal of the Optical Society of America*, Vol. 43, p. 1008.
[55] Kicinski, F., "Preparation of Photoconductive Cells by Chemicals Deposition of PbS." *Chemistry and Industry*, Vol. 17 (1948), p. 54.
[56] Kolomietz, B. T., "Characteristics and Properties of PbS Photoresistances." *Zhurnal tekhnickeskoi Fiziki*, Vol. 21 (1951), p. 1.
[57] Krentzian, O., "Preparation of PbTe Photoconductors." *Zeitschrift für Physik*, Vol. 126 (1949), p. 666.
[58] Kuiper, G. P., "Infra-red Spectra of Planets." *Astroptrys. Journal*, Vol. 106 (1947), p. 251.
[59] Lange, B., "A New Photoelectric cell." *Zeitschrift Physikalische* Vol. 31 (1930), p. 964.
[60] Lashkarev, V. E., Pomanenko, I. R., and Fedorus, G. A., "Non-linear Photoconductivity of PbS Photoresistances." *Zhurnal eksperimentalnoi i teoreticheskoi Fiziki*, Vol. 19 (1949), p. 887.
[61] Lasser, M. E., and Levinstein, H., "Optical Properties of PbTe." *Physical Review*, Vol. 96 (1954), p. 47.
[62] Lawrance, R., "The Photovoltaic Effect in Natural PbS." *Australian Journal of Scientific Research*, Vol. A4 (1951), p. 569.
[63] Levy, J. L., "Sensitive Hall Measurements on NaCl and Photoconductive PbTe." *Physical Review*, Vol. 92 (1953), p. 215.
[64] Lovell, B., *Electronics*, London, Chapman and Hall, 1951; p. 97 and thereafter.

[65] Lummis, F. L., and Petritz, R. L., "The Relation Between Noise and Response Time in Photoconductors." *Physical Review*, Vol. 86 (1952), p. 660.
[66] Macfarlane, G. G., "A Theory of Flicker Noise in Valves and Impurity Semiconductors." *Proceedings of the Physical Society*, Vol. 59 (1947), p. 366 and p. 403.
[67] Macfarlane, G. G., "A Theory of Contact Noise in Semiconductors." *Proceedings Physical Society*, Vol. B63 (1950), p. 807.
[68] Mahlman, G. W., Nottingham, W. B., and Slater, J. C., "Productivity of PbS." Paper read at the Atlantic City Conference on Photoconductivity, November, 1954.
[69] McFee, R. H., "Temperature Variation of Properties of Photosensitive PbS Films," *Physical Review*, Vol. 79 (1950), p. 203.
[69a] Michellsen, F., "Notes on Development of PbS PhSe and PbTe Photocells." *Optik*, Vol. 8 (1951), p. 75.
[70] Milner, C. J., and Watts, B. N., "PbS Photo-cells," *Research*, Vol. 5 (1952), p. 267.
[71] Milner, C. J., and Watts, B. N., "Lead Selenide Photoconductive Cells," *Nature*, Vol. 163 (1949), p. 322.
[72] Mitchell, G. R., and Goldberg, A. E., "Further Evidence for the Energy Gap of PbS." *Physical Review*, Vol. 93 (1953), p. 1421.
[73] Moss, T. S., "Spectral Sensitivity of PbS Layers." *Nature*, Vol. 159 (1947), p. 476.
[74] Moss, T. S., "Spectral Sensitivity of PbS Layers." *Nature*, Vol. 161 (1948), p. 766.
[75] Moss, T. S., "The Ultimate Limits of Sensitivity of PbS and PbTe Photoconductive Detectors." *Journal of the Optical Society of America*, Vol. 40 (1950), p. 603.
[76] Moss, T. S., *Photoconductivity*, Academic Press, New York, 1952.
[77] Moss, T. S., "PbTe and PbSe Infra-red Detectors," *Research*, Vol. 6 (1953a), p. 258.
[78] Moss, T. S., "Photoelectromagnetic and Photoconductive Effects in PbS Crystals." *Proceedings of the Physical Society*, vol. B66 (1953b), p. 993.
[79] Moss, T. S., "The Photoelectromagnetic Effect in Ge and PbS." *Physica*, Vol. 20 (1954), p. 989.
[80] Moss, T. S., and Chasmar, R. P., "Spectral Response of PbSe." *Nature*, Vol. 161 (1948), p. 244.
[81] Muser, H., "Sensitivity Limits of Photoresistors." *Zeitschrift für Physik*, Vol. 129 (1951a), p. 514.
[82] Muser, H., "On Localized Decreases of Resistance in Photoconductive Layers." *Zeitschrift Physik Chem.*, Vol. 198 (1915b), p. 52.
[83] Oxley, C. L., "Characteristics of Cooled PbS Cells." *Journal of the Optical Society of America*, Vol. 36 (1946), p. 356.
[84] Parker, R. C., and Marshall, P. R., "The Measurement of the Temperature of Sliding Surfaces." *Proceedings of the Institute of Mechanical Engineering*, London, Vol. 158 (1948), p. 209.
[85] Paul, W., and Jones, R. V., "Absorption Spectra of PbS at Different Temperatures." *Proceedings of the Physical Society*, Vol. B66 (1953), p. 194.
[86] Paul, W., Jones, D. A., and Jones, R. V., "Infra-red Transmission of Galena." *Proceedings of the Physical Society*, Vol. B64 (1951), p. 528.
[87] Paulson, R., "Photoconductive Detectors for Infra-red Systems." *Electrical Engineering*, Vol. 74 (1955), p. 214.
[88] Petritz, R. L., "The Relation Between Noise and Response Time in Photoconductors." *Physical Review*, Vol. 86 (1952), p. 660.
[89] Petritz, R. L., "Relation Between Lifetime, Limit of Sensitivity and Information Rate in Photo-conductors." Paper read at the Atlantic City Conference on Photoconductivity, Nov. 1954.
[90] Petritz, R. L., and Scanlon, W. W., "Mobility of Electrons and Holes in the Polar Crystal PbS." *Physical Review*, Vol. 97 (1955), p. 1620.
[91] Petritz, R. L., and Seigert, A. J. F., "On the Theory of Noise in Semi-conductors." *Physical Review*, Vol. 79 (1950), p. 215.
[92] Pick, H., "On Photoconductivity in PbS." *Annalen der Physik*, Vol. 3 (1948), p. 253.
[93] Pincherle, L., "Anger Effect in Semiconductors." *Proceedings of the Physical Society*, Vol. B68 (1955), p. 319.
[94] Piwkowski, T., "Photosensitive PbS Layers With New Properties." Bulletin of the Academy of Science, Polon., Cl. 3, Vol. 1 (1953), p. 185.
[95] Putley, E. H., "Electrical Conductivity in PbS, PbSe and PbTe: Conductivity and Hall Coefficient of Sintered PbS: and Intrinsic Conduction in PbS, PbSe and PbTe." *Proceedings of the Physical Society*, Vol. B65 (1952), pp. 388, p. 736, and 992.
[96] Putley, E. H., "Thermoelectric and Galvanoelectric Effects in PbSe and PbTe." *Proceedings of the Physisal Society*, Vol. B68 (1955), p. 35.
[97] Putley, E. H., and Arthur, J. B., "PbS—An Intrinsic Semiconductor." *Proceedings of the Physical Society*, Vol. B64 (1951), p. 616.
[98] Pyatt, E. C., "Simplified Brightness Temperature Pyrometer Using Photoconductive Cell." *Journal of Scientific Instruments*, Vol. 29 (1952), p. 125.
[99] Rittner, E. S., "Electron Processes in Photoconductors." Paper read at the Atlantic City Conference on Photoconductivity, November, 1954.
[100] Rittner, E. S., and Grace, F., "Impedance Measurements on PbS Photoconductive Cells. *Physical Review*, Vol. 86 (1952), p. 955.
[101] Roberts, V., and Young, A. S., "Application of PbSe Photoconductive Cells to Infra-red Spectroscopy." *Journal of Scientific Instruments*, Vol. 30 (1953), p. 199.
[102] Roth, L., Dissertation. Erlangen, 1938.
[103] Roth, L. M., Meissner, K. W., and Lark-Horovitz, K., "Optical and Electrical Behaviour of Lanarkite." *Physical Review*, Vol. 85 (1952), p. 724.
[104] Ryvkin, S. M., "Fundamental Investigation of the Effort of Oxygen on PbS Photoresistances." *Physical Review*, Vol. 92 (1953), p. 1573. *Journal of Technical Physics*, U.S.S.R., Vol. 22 (1952), p. 1930.
[105] Scanlon, W. W., "Interpretation of Hall Effect and Resistivity Data in PbS and Similar Binary Compound Semiconductors." *Physical Review*, Vol. 92 (1953), p. 1573.
[106] Scanlon, W. W., and Brebrick, R. F., "Preparation of PbS Crystals of Controlled Properties." *Physical Review*, Vol. 94 (1954), p. 1430.
[107] Scanlon, W. W., Brebrick, R. F., and Petritz, R. L., "Chemical and Physical Properties of PbS Single Crystals." Paper read at the Atlantic City Conference on Photoconductivity, November, 1954.
[108] Scanlon, W. W., Petritz, R. L., and Lummis, F. L., "Multiple Time Constants in Photoconductivity." *Physical Review*, Vol. 86 (1952), p. 659.
[109] Schwarz, E., "New Photoconductive Cells." *Nature* (*London*), Vol. 162 (1948), p. 614.
[110] Schwarz, E., "Theory of Photoconductivity of Layers of Semiconducting Substance." *Proceedings of the Royal Physical Society*, Vol. A62 (1949), p. 530.
[111] Silverman, S. J., and Levinstein, H., "Electrical Properties of Single Crystals and Thin Films of PbSe and PbTe." *Physisal Review*, Vol. 94 (1954), p. 871.
[112] Simpson, O., "Conductivity of Evaporated Films of PbSe." *Nature*, London, Vol. 160 (1947), p. 791.
[113] Simpson, O., "Effect of Room Radiation on the Response of PbTe Photoconductors." *Proceedings of the Physical Society*, Vol. 61 (1948), p. 486.
[114] Simpson, O., and Sutherland, G. B. B. M., "Photoconductivity in the Infra-red Region of the Spectrum." *Transactions of the Royal Society*, Vol. A243, p. 547.
[115] Simpson, O., Sutherland, G. B. B. M., and Blackwell, D. E., "PbTe Cells for Spectroscopy." *Nature*, London, Vol. 161 (1948), p. 281.
[116] Smith, R. A., "Infra-red Photoconductors." *Philosophical Magazine Supplement*, Vol. 2 (1953), p. 321.
[117] Smith, R. A., "The Electronic and Optical Properties of the PbS Group of Semiconductors." *Physica*, Vol. 20 (1954), p. 910.
[118] Smollett, M., and Pratt, R. G., "The Effect of Absorbed Air on Photoconductive PbS Layers." *Proceedings of the Physical Society*, Vol. B68 (1955), p. 390.
[119] Sosnowski, L., and Chmielewski, M., "Response Time of Photoconductivity of PbS." Bulletin of the Academy of Science, Polon., Cl. 3, Vol. 1 (1953), p. 199.
[120] Sosnowski, L., Soule, B. W., and Starkiewicz, J., "Occurrence of Random Photovoltaic Barriers in Photoconductive Layers." *Nature*, London, Vol. 160 (1947), p. 471.
[121] Sosnowski, L., Starkiewicz, J., and Simpson, O., "PbS Photoconductive Cells." *Nature*, London, Vol. 159 (1947), p. 818.
[122] Soule, D. E., and Cashman, R. J., "Bulk Photoconductivity in PbS." *Physical Review*, Vol. 93 (1953), p. 635.
[123] Starkiewicz, J., "PbSe Photoconductive Cells." *Journal of the Optical Society of America*, Vol. 38 (1948), p. 481.
[124] Starkiewicz, J., Sosnowski, L., and Simpson, O., "Photovoltaic Effects Exhibited in High Resistance Semiconducting Films." *Nature*, London, Vol. 158 (1946), p. 28.
[125] Sutherland, G. B. B. M., Blackwell, D. E., and Fellgett, P., "Use of PbS Cells in Infra-red Spectroscopy, *Nature*, London, Vol. 158 (1946), p. 873.
[126] Sutherland, G. B. B. M., and Lee, E., "Developments in the Infra-red Region of the Spectrum," *Reports on Progress in Physics*, Vol. 11 (1946–7), p. 144.
[127] Thompson, H. W., and Williams, R. L., "Infra-red Spectra of Methyl Cyanide and Methyl Isocyanide," *Transactions of the Faraday Society*, Vol. 48 (1952), p. 502.
[127a] van der Ziel, A., *Noise*. New York, Prentice Hall, 1954.
[128] Vernier, P., "Optical and Photoelectric Properties of Microcrystalline Layers of PbS." *Journal de Physique et le Radium*, Vol. 14 (1953), p. 175.
[129] Watts, B. N., "Increased Sensitivity of Infra-red Photoconductive Receivers." *Proceedings of the Physical Society*, Vol. A62 (1949), p. 456.
[130] Wilman, H., "Structure of PbS and PbSe Deposits and Effects of Oxygen." *Proceedings of the Physical Society*, Vol. A60 (1948), p. 117.

WRH

Theory and Experiments on a Basic Element of a Storage Light Amplifier*

J. E. ROSENTHAL†

Summary—**A simple theory is developed for the performance of the basic element of a storage light amplifier. This theory permits the determination of the optimum operating frequency and of restrictions, which have to be imposed on operating characteristics, such as the feedback ratio, in order to make storage possible. The theory is compared with a set of experimental data obtained for a basic element of a storage light amplifier. The comparison leads to numerical values for a set of parameters appearing in the theoretical analysis. In view of the approximate nature of the assumptions made, the agreement between theory and experiment is considered to be very satisfactory. The information obtained from the analysis of this basic element is used to design a large display area model of a storage light amplifier. The processing techniques involved are described briefly in an Appendix.**

Introduction

THE LIGHT amplifier is a device which changes the radiation of a given intensity and wavelength distribution impressed upon it into radiation of higher intensity and possibly different spectral distribution. With the advent of moderately efficient electroluminescent phosphors and high sensitivity photoluminescent mechanisms, a solid state light amplifier became a definite possibility. A storage light amplifier is a device, in the form of a panel, which stores and reradiates a black and white pattern impressed upon it as long as electrical power is supplied to it. The intensity of the re-emitted pattern may be equal to or even lower than that of the original pattern. Thus, strictly speaking, this device is not an amplifier at all. However, the name of storage light amplifier has been commonly adopted for it because the mechanism is similar to that for a true light amplifier. Once the light pattern has been impressed on the storage light amplifier, the sustaining intensity, which is obtained by feedback, is only a small fraction of the total radiation emitted by the panel.

In the last two years data have been published on several types of light amplifiers.[1-3] In all these devices the photoconductor is used to control the light output from an electroluminescent element by controlling the voltage applied across it. The purpose of this paper is to present a simple theoretical calculation on the operation and performance of the storage light amplifier, to compare the theoretical performance with the experimental data available, and to base on this theory the design of a practical unit. The experiments were performed at this laboratory by Dr. A. Bramley.

Basic Theory

The basic operating principle of a storage light amplifier is as follows:

In their simplest embodiments all the devices described in the literature are electrically equivalent and can be represented by a resistor (the photoconductor) in series with a condenser (the electroluminescent element). An alternating potential of voltage V and frequency $\omega/2\pi$ is applied across this circuit. Let V_p be the resultant voltage across the electroluminescent element. If V_0 is the minimum voltage which can activate the element, then for $V_p \geqq V_0$ the brightness B (in footlamberts) of the light produced under these circumstances is

$$B = k\omega^a V_p^m \tag{1}$$

where k, a, and m are the constants depending on the constitution of the electroluminescent element. (When $V_p < V_0$, $B=0$.) If I is the illumination incident on the photoconductor, its resistance R may have the form

$$1/R = \gamma_0 + \gamma I^n \tag{2}$$

where γ_0, γ, and n are constants depending on the photoconductor. The dark conductivity γ_0 may frequently be neglected and is not taken into account in the expressions below. In terms of these quantities and of the capacity C of the electroluminescent element, the voltage across the element is

$$V_p = V/(1 + \omega^2C^2R^2)^{1/2} = V/[1 + (\omega C/\gamma I^n)^2]^{1/2}. \tag{3}$$

The brightness B can therefore be expressed as

$$B = \frac{k\omega^a V^m}{(1 + \omega^2C^2R^2)^{m/2}} = \frac{k\omega^a V^m}{[1 + (\omega C/\gamma I^n)^2]^{m/2}}. \tag{4}$$

Since the device operates only if $V_p \geqq V_0$, this imposes a condition on R or I

$$\omega CR \leqq [(V/V_0)^2 - 1]^{1/2} \tag{5}$$

$$I^n \geqq \frac{C\omega}{\gamma[(V/V_0)^2 - 1]^{1/2}}. \tag{6}$$

The conditions for the optimum operation of the circuit may be derived from (4). Two cases are of particular practical importance:

1) R is constant—this is the mode of operation of the device which permits the determination of various circuit parameters associated with the electrolumines-

* Original manuscript received by the IRE, August 15, 1955; revised manuscript received, September 27, 1955.

† A. B. Du Mont Laboratories, Inc., Passaic, N. J.

[1] W. White, "X-ray image intensification and method," U. S. Patent No. 2,650,310; August 25, 1953.

[2] R. K. Orthuber and L. R. Ullery, "A solid state image intensifier," *Jour. Opt. Soc. Amer.*, vol. 44, pp. 297–299; April, 1954.

[3] A. Bramley and J. E. Rosenthal, "Transient voltage indicator and information display panel," *Rev. Sci. Instr.*, vol. 24, p. 471; June, 1953.

cent element. (In this case the discussion obviously refers to the operation of the electroluminescent element in series with a fixed resistance.)

2) The feedback ratio is constant—this is the mode of operation of a storage light amplifier. Let I_f be the illumination on the photoconductor resulting from a brightness B of the radiation emitted by the electroluminescent element, and let I_0 be the illumination due to other causes. The total intensity I of the light incident on the photoconductor is therefore

$$I = I_0 + I_f.$$

The feedback ratio is defined in the usual way as

$$\beta = I_f/B.$$

Eq. (4) therefore takes the form

$$B = \frac{k\omega^a V^m}{\{1 + [\omega C/\gamma(I_0 + \beta B)^n]^2\}^{m/2}}. \quad (7)$$

When the storage light amplifier is triggered off, $I_0 \gg B$ since initially $B=0$. The triggering signal, however, may last only a few milliseconds. During the steady state operation, after the triggering signal has passed, the ambient light level I_0 is very low for the experiments to be discussed here (of the order of 0.1 foot-lambert) so that it can be ignored in (7) as compared with βB.

For a given set of exponents m and n (7) may not admit a solution for B for all values of β. For example, let $m=2$, which, as will be seen below, corresponds to the experimental set-up. Eq. (7) takes the form

$$B[1 + (C\omega/\gamma\beta^n B^n)^2] = k\omega^a V^2. \quad (8)$$

When $n=1$, this is a quadratic equation in B, which will have a real positive root if, and only if,

$$C\omega \leqq \tfrac{1}{2}k\gamma V^2\omega^a\beta. \quad (9)$$

If $n=\frac{1}{2}$, (8) is linear in B, it gives a positive value for B if

$$C\omega \leqq V\gamma(k\omega^a\beta)^{1/2}. \quad (10)$$

On the other hand, if $n=\frac{1}{4}$, (8) is quadratic in $\sqrt{B}$, but has a real solution for all values of β. The only restriction imposed is the one given by (6) for I.

To find the maximum value of B, B_M, and the corresponding value of ω, (3) is differentiated with respect to ω and the derivative $dB/d\omega$ is set equal to zero. The resulting values for cases 1 and 2 are:

1) (R constant)

$$B_M = \frac{kV^m a^{a/2}(m-a)^{(m-a)/2}}{m^{m/2}(CR)^a}, \quad (11)$$

which occurs at a frequency given by

$$C\omega R = [a/(m-a)]^{1/2}. \quad (12)$$

Eq. (12) shows that for optimum operation of the circuit there must be suitable matching of impedances.

2) (β constant)

$$(B_M)^{1-na} = \frac{kV^m a^{a/2}(m-a)^{(m-a)/2}\gamma^a\beta^{na}}{m^{m/2}C^a}, \quad (13)$$

which occurs at a frequency given by

$$C\omega = \gamma\beta^n B_M{}^n[a/(m-a)]^{1/2}. \quad (14)$$

Experimental Set-up

The experimental measurements were carried out on a structure which may be considered a simplified form of the storage light amplifier. Its cross section is shown in Fig. 1, which should explain the fundamental principle

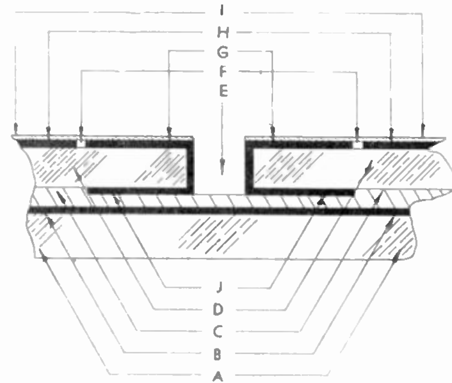


Fig. 1—Cross section of elementary storage light amplifier (not to scale): (A) and (D) glass plates; (B), (G), (H), and (J) NESA coating; (C) electroluminescent layer; (E) perforation in glass plate; (F) gap in conducting coating; (I) CdS layer.

of its operation. The basic components shown are as follows:

A glass plate (A), 2 inches square, coated with a conducting layer (B), *e.g.*, NESA, has deposited on it a thin electroluminescent layer (C) about 0.0025 inch thick, consisting of a resin impregnated with electroluminescent phosphor powder. (A Sylvania green phosphor was used in these experiments.) This layer (C), frequently called a phosphor-resin matrix, is covered by another glass plate (D), 0.015 inch thick with a fine hole (E), 0.017 inch in diameter in its center. The purpose of (D) is to carry the response mechanism coordinating the light output of the electroluminescent layer with the intensity of the light incident on the device. It makes intimate contact with the top surface of the electroluminescent layer (C) to achieve light output which is uniform with uniform conditions of excitation. To make the response mechanism, the glass plate is coated with the conducting coating in a concentric arrangement consisting of a ring (G) surrounding the hole (E) and a continuous coating (H) surrounding (G). The outside diameter of the ring (G) is 0.20 inch, and the insulating gap (F) between (G) and (H) is 0.010 inch wide. A photoconducting CdS film (I) is evaporated onto the top surface of the glass plate (D) over the NESA coating at (G) and (H), which serves as the electrode for it. When the CdS film is exposed to different light levels, its resistance across the gap (F) in the NESA coating changes.

One terminal of the ac voltage is applied to the NESA coating (H), the other to the continuous NESA coating (B) under the electroluminescent layer (C). To complete the circuit, a contact must be provided on top of the electroluminescent layer (C). This is accomplished by extending the NESA coating (G) through the perforation (E) along its wall. On the other side of the perforation the extended conducting coating forms a ring (J) of 0.20 inch outside diameter, which surrounds the perforation. The combination of electroluminescent layer (C) and conducting coatings (B) and (H) constitutes the electroluminescent element discussed in the section on Basic Theory.

Fig. 2 illustrates the electrode system of the NESA ring (G) and continuous coating (H) on the upper side of the glass plate (D).

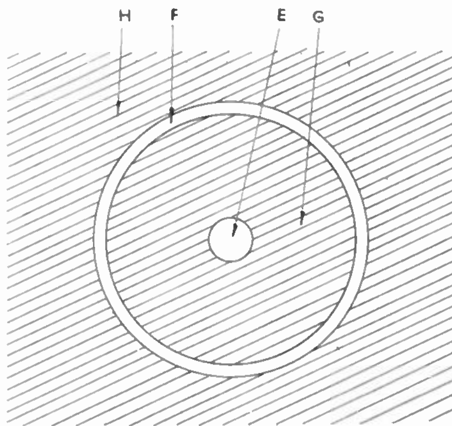


Fig. 2—Distribution of conducting coating on the top glass surface of the elementary storage light amplifier under the CdS layer (not to scale): (E) perforation in glass plate; (F) gap in conducting coating; (G) ring-shaped conducting coating; (H) conducting coating continuous on top surface except for region bounded by (F).

Determination of Characteristics

As shown in the section on Basic Theory, a number of parameters must be determined experimentally to permit the theoretical calculation of the optimum frequency of operation. Since these parameters depend largely on the processing, the techniques used are described briefly in the Appendix.

To operate the storage light amplifier, an alternating voltage of 250–500 v in the audio frequency range (20–2,000 cps) is applied between the NESA coatings (H) and (B). The first experiments were directed toward the determination of the capacity C of the electroluminescent element and of the exponent m, which gives the dependence of light output on the voltage. They involved finding the ratio of the electroluminescent light output when the gap (F) between the portions (G) and (H) of the conducting coating is bridged by an 8 megohm resistance and when it is shorted by a probe. The measurements were performed at 400 v. Considering the brightness $B(R)$ as a function of the resistance R we can write this ratio as $B(0)/B(8\times10^6)$. Eq. (3) gives

$$B(0)/B(R) = (1 + \omega^2C^2R^2)^{m/2}. \qquad (15)$$

The experimental ratios obtained at different frequencies are:

$f = \omega/2\pi$	<100	100	200	1,000
$B(0)/B(8 \times 10^6)$	1	1	2	15

No difference was observed up to a frequency of 100 cps in the brightness of the circular spot on the electroluminescent element in the two cases.

Eq. (15) permits the determination of C and m from the experimental values of the brightness ratio at 200 and 1,000 cps. If the experimental error were negligible, the only possible solutions would be $m=1.465$ and $C=125$ $\mu\mu f$. However, this value of m is too low to agree with the data available on the dependence of the electroluminescent radiation on the applied voltage.[4] The lowest value of m consistent with these other results is $m=2$. With this choice of m, the value of C which gives the best agreement with the experimental values for the brightness ratios is $C=84$ $\mu\mu f$. Corresponding to these values of C and m, the calculated values of $B(0)/B(8\times10^6)$ are 1.7 at $f=200$ and 18.5 at $f=1,000$, well within the limits of experimental error. The capacity of the electroluminescent element as computed from its geometry (.20 inch diameter, .0025 inch spacing between the plates, see above) for a dielectric constant of 8 is 22 $\mu\mu f$. The use of this calculated value for C in (15) gives, however, radically different values for m from data taken at 200 and at 1,000 cycles. We are faced therefore with a fairly large variation between the calculated capacity and the value deduced from light output readings. The following two sets of experiments help to account for such a wide variation.

First the capacity was determined on a bridge operating at 1 mc. (The unit used was the Kay Lab Dynamic Micro Miker.) The reading obtained was 14 $\mu\mu f$. In this measurement the applied voltage was, of course, far below the cutoff voltage of the panel.

The second set of experiments was directed to ascertaining what changes in polarization occur when the phosphor particles become conducting as a result of acting as electroluminescent radiators. Such polarization changes would obviously affect the capacity of the dielectric element. Conductivity in the phosphor particles may also be induced by exposure to uv radiation. The latter method is convenient for estimating the magnitude of the polarization currents which are induced in an electroluminescent element by its own radiation. The experiments consisted of measuring the current through the electroluminescent element due to a dc voltage of 60 v (applied across the element) when intense radiation of the phosphor was induced by a uv emitting lamp. The uv intensity was adjusted so that the light emitted by the phosphor was comparable in intensity with electroluminescent radiation at 300 v and 500 cycles.

[4] Mellon Institute of Industrial Research, Quarterly Rep. No. 4, second series of the Computer Components Fellowship No. 347, pp. V1–V7; 1954.

With the lamp on, the current through the element was increased 40 per cent over the dark current. (These currents were in the 0.1 μa range.) If the radiation from the uv lamp was interrupted, a large current pulse was observed, which changed sign depending on whether the uv radiation was flashed on or blacked out.

The results of these two sets of experiments are in agreement with Roberts[5] who observed a rapid decrease in the dielectric constant of a panel as the ac frequency was increased. On this basis the agreement between the calculated value of 22 $\mu\mu$f for dc operation and the measured value of 14 $\mu\mu$f at 1 mc can be considered good. Furthermore, Roberts found that the dielectric constant of the electroluminescent panel at low frequencies (*i.e.*, at about 100 cycles) was about 15. This is approximately twice the value of the dielectric constant used in the calculations. With this correction the calculated value of the capacity of the electroluminescent element is 41 $\mu\mu$f. While this is still below the expected value of 84 $\mu\mu$f, in view of the complex polarization currents through the electroluminescent element, this difference is not excessive.

The value of β is determined from the geometry of the device as shown in Figs. 1 and 2. It has been calculated that 50 per cent of the light emitted from one side of the electroluminescent layer (C) is incident on the gap (F) constituting the photoconducting area. This means that $\beta=0.5$. The loss of light can be accounted for as follows: 10 per cent are lost in its passage through the NESA coating (J), which is only 90 per cent transparent; the remaining 40 per cent of the light loss occurs at the interface between the glass plate (D) and the photoconductor (I) in the ring-shaped gap (F) as a result of the processing technique used in obtaining the insulating gap (F). In the experimental model, Photoresist was used, which required the surface of the glass plate (D) to be etched with ensuing loss in transparency. However, in the final model, as opposed to the preliminary one used for these measurements, the glass plates will be sprayed with aquedag to form the gap. This type of processing does not change the transparency of the glass interface at the gap.

Since the electroluminescent element used in these experiments emits green light, the response of the photoconductor was also studied by green light. For convenience, a Wratten No. 58 filter was used in front of an incandescent light illuminating the photoconductor. It is well known that the resistance of a photoconductor is a function of the voltage maintained across it; *i.e.*, the "constants" n and γ in Eq. (2) are really functions of the voltage. To determine the extent of the variation of n and γ with voltage, the response of the photoconductor to green light was measured at 25, 50, and 100 volts ac. As seen below, the variation is not excessive, and we may assume that one set of n, γ values will satisfactorily represent the resistance of the photoconductor at various voltages for the range of illumination selected. The value of 0.2 foot-lambert was chosen as the minimum value of feedback illumination because it is believed that the brightness of the storage device would have to be at least 0.4 foot-lambert to be of any practical value.

The values of the resistance can be suitably well represented by (2) with $n=0.4$ and $\gamma=1.8\times10^{-7}$. Such calculated values are also included in Table I below.

TABLE I

Illumination (in ft.-lambert)	Resistance (in megohms)			
	at 25 V	at 50 V	at 100 V	Calculated
0.2	11	8.2	7.7	10.6
0.5	7.4	5.9	6.3	7.3
1.0	6.4	4.6	5.0	5.6
5.0	4.3	2.0	2.0	2.9
10.0	3.2	1.6	1.6	2.2

Excessive variations with voltage of the parameters n and γ might well tend to counterbalance partially the change in resistance from the feedback radiation impinging on the photoconductor.

The coefficient of proportionality k of Eq. (1) was determined as a function of a by measuring the brightness of the panel when the gap (F) in the conducting coating was shorted by a probe. At 500 cps and at a voltage $V=500$ volts rms, $B(0)=kV^2\omega^a=6$ foot-lambert, giving $k=24\times10^{-6}(10^3\pi)^{-a}$.

Knowing the experimental values for the various parameters in Eq. (7), it is possible to compute the storage effect for any value of the feedback ratio β.

Footnote [4] gives—for a sine wave input—$a=0.38$ for the Sylvania green 600 v panel and 0.45 for the Sylvania green 120 v panel. Our results indicate that a is very closely 1. In view of this discrepancy, we computed the brightness of the storage light amplifier as a function of the frequency for the two values $a=0.40$ and $a=1.00$ for a feedback ratio $\beta=0.5$. Fig. 3 (next page) shows the ratio B/B_M plotted against the frequency f for these two values of a as well as a set of experimental values discussed below. The theoretical curves show that as a increases, the maximum is shifted toward higher frequencies. While the shapes of the two theoretical curves are quite similar, the actual values of B_M differ considerably, being $B_M=2.00$ for $a=0.40$ and $B_M=1.00$ for $a=1.00$. The latter value is in good agreement with the experimental value $B_M=0.9$ foot-lambert.

It is interesting to note that Eqs. (13) and (14) give

$$B_M = \tfrac{1}{2}k\omega V^2, \qquad (a = 1);$$

i.e., it is one-half the light output that would be expected if the voltage V of frequency $\omega/2\pi$ were applied across the electroluminescent element only without any additional resistances.

[5] S. Roberts, "Dielectric changes of electroluminescent phosphor during illumination," *Jour. Opt. Soc. Amer.*, vol. 43, pp. 590–592; July, 1953.

It is found experimentally that a storage effect does occur for this case where the computed value of the feedback ratio equals one-half. This effect is observed over the frequency range from 20 to 1,000 cps. In this experiment the CdS photoconductor is excited by external radiation and its impedance is kept low by feedback from the electroluminescent element. At a frequency below 1,000 cps, external radiation will trigger the photoconductor so that the element becomes and remains luminous until the voltage is decreased below 300 v. To produce the triggering action, the photoconductor needs to be irradiated by an external source for only a millisecond. Once triggered, the radiation from the electroluminescent panel with 400 v across the combination electroluminescent panel and photoconductor varies as follows with frequency:

f	20	200	250	1,000
B/B_M	1/10	5/6	1	0

These experimental values are marked as crosses in Fig. 3.

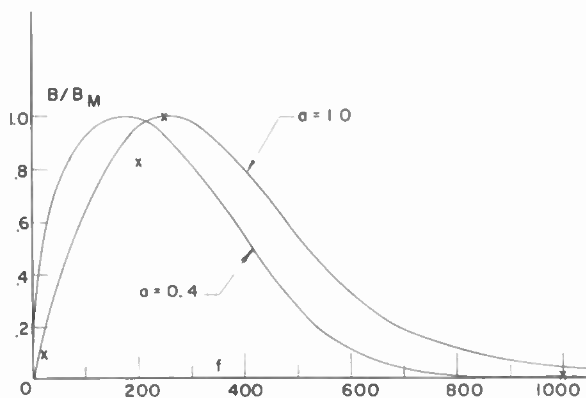


Fig. 3—Brightness distribution of the elementary storage light amplifier as a function of the frequency for a feedback ratio of one-half. The theoretical curves are computed for two different types of frequency dependence of the electroluminescent element. The crosses represent experimental values.

While these data by themselves are insufficient to permit any definite conclusion, they confirm that $a = 0.4$ is too low a value of a for the electroluminescent element used and that $a = 1.0$ is a better selection. The disagreement with the results in reference 4 for the same Sylvania phosphor presumably indicates that processing plays a vital part in determining the values of the parameters.

Model with Large Display Area

A storage light amplifier with a large display area could be made to consist of a multiplicity of elements described above. The main requirement is that the basic design be repeated in an appropriate pattern. Except for a larger total area, the glass plate (A), conducting layer (B), and electroluminescent layer (C) would be the same as before. However, the glass plate (D) would have a multiplicity of fine holes (E) arranged in a rectangular array. (Glass sheets perforated in this fashion were kindly furnished us by the Corning Glass Co.) Fig. 4 represents a cross section of such a regular model. The conducting coating is extended as before along the wall of the perforations (E) to form the conducting rings (J) on the underside of plate (D). However, since there is now a multiplicity of such rings (J), they must be separated at their closest point of approach by a distance

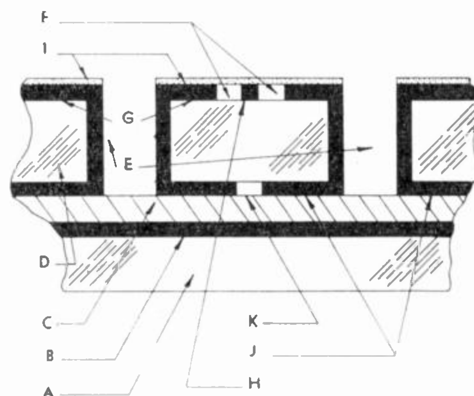


Fig. 4—Cross section through a large display area model of a storage light amplifier (not to scale): (A) and (D) glass plates; (B), (G), (H) and (J) NESA coating; (C) electroluminescent layer; (E) perforations in glass plate; (F) and (K) gaps in the conducting coating; (I) CdS layer.

several thousandths of an inch greater than 0.0025 inch, which is the thickness of the electroluminescent layer, to minimize the possibility of arcing between different sets of array. This is shown in Fig. 4 as the gap (K) between the NESA rings (J). Fig. 5 is analogous to Fig. 2 for the single element. It shows the electrode system of the NESA rings (G) and latticed coating (H) on the upper side of the glass plate (D).

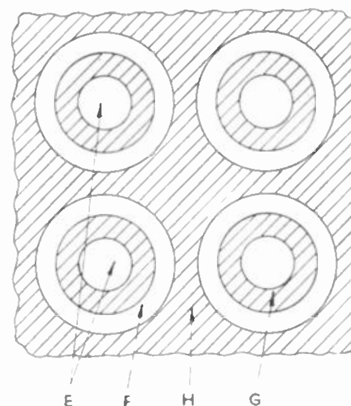


Fig. 5—Distribution of conducting coating on the top glass surface of a large display area model of a storage light amplifier (not to scale): (E) perforations in glass plate; (F) ring-shaped gaps in the conducting coating; (G) conducting rings; (H) conducting coating in a latticed pattern.

In the ultimate design of the storage device, a center-to-center spacing of the holes of 0.080 inch is anticipated to give an operable device. For the equivalent resolution of a 525 line raster, the storage panel would require

over-all dimensions of 42 inches. This is slightly larger than the present TV raster presentation.

It is obvious from the geometry of the device that the sensitive photoconductor area; *i.e.*, the gap (F) is only a small fraction of the total area of the device illuminated by the incident light. For a storage tube this causes no difficulty. The efficiency of the device is controlled by the impedance match between the photoconductor and electroluminescent layer and by the feedback of the light produced in the electroluminescent layer. The requirements are as follows: the brightness levels of the incident light energizing the unit need not be in excess of 1 foot-lambert, while the lowest brightness level should be greater than 0.2 foot-lambert for reliable response. The triggering light flux should be incident on the photoconductor for a few milliseconds, a time interval greater than the minimum response time of the photoconductor.

The situation is different in the case of a light amplifier. Here the incident light flux extends from a black level (*i.e.*, one below 0.2 foot-lambert) to a maximum of no more than 2 foot-lambert. To make full use of the light flux incident on each elementary area, it is necessary to devise some means of concentrating it all on the photoconducting gaps (F). A possible solution is to place over the glass plate (D) in Fig. 4 an additional glass plate whose outer surface is composed of minute cylindrical lenses such that the major part of the incident light falls on the sensitive area of the photoconductor. While such an array of cylindrical lenses sounds formidable, it is not difficult to manufacture. Actually many cheap glass plates are decorated in this way by proper dies in the molding.

Conclusions

In comparing this light amplifier with other models, it should be pointed out that for the design proposed in reference 2 the simple basic theory needs to be amended to include a condenser in parallel with the photoconductive resistance. As a result, some of the current will bypass the photoconductive resistance thus decreasing the over-all sensitivity of the device. It is believed that the type of structure for a light amplifier described in this paper has the following inherent advantages, which enhance its sensitivity:

The capacitive reactance across the gap (F) is so high as compared to the resistivity of the photoconductor that it can be ignored as a possible bypass for the current.

It is possible to use uv or X-ray without attenuation by passing the radiation through a transparent plate; *e.g.*, glass or quartz.

Evaporated CdS films permit the use of higher voltage.

A solid state structure of this type can also be incorporated as a target in a cathode-ray tube replacing a screen, provided the photoconductor is chosen so that its impedance can be lowered by electron bombardment. Here the information to be stored is impressed upon the target not by illumination, but by an electron beam scanning across the target assembly.

Appendix

Phosphor Matrix

The dielectric used was polyvinyl chloride resin. The composition of the electroluminescent phosphor-resin matrix was approximately 22.5 per cent (by weight) of phosphor, 15 per cent of VYNS resin, the remaining 62.5 per cent consisting primarily of the solvent isobutyl ketone with a few per cent of plasticizer DOPE. (The resin and plasticizer were kindly supplied by the American Cyanamid Co.) This mixture was ballmilled for 48 hours before being applied to the lower glass plate (A) by a draw plate. Finally the phosphor-resin matrix was heated to about 110–200°C with the top plate in position and was subjected for a few minutes to a pressure of a few psi. On cooling the resulting panel was firmly bound.

Photoconductor

As indicated in the section on Basic Theory, the impedance of the photoconductor in the dark must be matched against that of the electroluminescent element. The reactance of the basic element being of the order of 25 $\mu\mu$, its impedance at 200 cps is

$$1/(25 \times 10^{-12} \times 2\pi \times 200) = 10^8/\pi = 32 \text{ megohms}.$$

This is the order of magnitude required of the dark resistance of the CdS layer across the gap (F) since values from 10^8 to 10^9 ohms for the specific dark resistance of CdS correspond to its range of maximum sensitivity.

The following procedures can be used to form the photoconducting layer: evaporation, dusting finely divided crystalline CdS particles over the electroluminescent layer, flow coating them, and settling them from an appropriate liquid. For these last three methods, the particle size should be less than 0.001 inch; *i.e.*, less than the thickness of the electroluminescent layer. However, it was found that layers of very fine CdS particles have a low sensitivity. This may have been due to failure to obtain particles of uniformly high sensitivity.

Evaporation was selected as the most suitable method of deposition. Combined with suitable processing it gives a CdS layer with good photoconducting properties, which is sensitive not only to visible radiation in the green and blue but also extremely sensitive to uv radiation or to X-rays. The method of construction of the storage device is such that these short wavelength radiations can be utilized.

Since both electrodes are NESA coated, the element obviously shows no photovoltaic effect. The breakdown voltage of the evaporated CdS film for a 0.012 inch gap is well over 300 v dc. (For most commercial CdS crystal photoconductors, the breakdown voltage is well below 200 v dc.)

Another advantage of these evaporated CdS films is their high dark resistance.

Although for the particular photoconductor used in this experimental storage light amplifier $n=0.4$, it is possible to produce photoconductors with either a lower or higher value for n, which may even exceed $n=1$.

A description of the processing of these evaporated films was given by Arthur Bramley at the Chicago meeting of the American Physical Society.[6] It may be added that the CdS was evaporated from a tantalum boat onto the prepared NESA coating. This preparation usually consisted in depositing a very thin film of Ag on the NESA coating and the gap in it. At the high temperature, at which the NESA coated glass plate was maintained during the evaporation and which was still further raised during the aging, the Ag diffuses into the CdS film evaporated over the Ag film. On the other hand since the temperature of the tantalum boat was not raised beyond 850°C, the temperature of evaporation, the amount of impurity Cd evaporated should be less than 1 part to 10^5 parts of evaporated CdS. As mentioned in footnote [6], the evaporated CdS films were aged by baking to 350°C in air for 10 minutes.

[6] A. Bramley, "Deposition of CdS in vacuum," *Phys. Rev.*, vol. 98, p. 246, abstract V11; April 1, 1955.

An Electroluminescent Light-Amplifying Picture Panel*

B. KAZAN†, SENIOR MEMBER, IRE, AND F. H. NICOLL†, SENIOR MEMBER, IRE

Summary—**This paper describes an experimental amplifier for light which uses a photoconductive layer electrically in series with an electroluminescent phosphor layer. With the series combination excited by an alternating voltage of audio frequency; e.g., 400 cps, a low light level impinging on the photoconductor decreases its resistance sufficiently to cause a much larger light output from the phosphor. Large-area light-amplifying panels, 12 inches square, using a newly developed photoconductive CdS powder have been built. These panels are capable of producing intensified half-tone images with high resolution. The response time of these panels, determined basically by the photoconductive material, varies from 0.1 second to several seconds. New photoconductive materials offer promise of greatly increased speed. Since the photoconductors can be made sensitive to X-rays and infrared, the principles of the device may be useful for image conversion and intensification with these radiation sources.**

Introduction

The possibility of a panel type of light amplifier was initially demonstrated by the operation of an electroluminescent panel[1] in series with a small photoconductive crystal of CdS as shown in Fig. 1. In the dark the high impedance of the CdS crystal permitted only a negligible fraction of the alternating supply voltage to be applied across the electroluminescent layer. As the incident light on the crystal was increased, its impedance decreased and greater voltage appeared across the phosphor layer. It was further demonstrated that, with equal areas, the lumens emitted by the phosphor could be many times greater (*e.g.*, 50 times) than the lumens incident on the crystal.

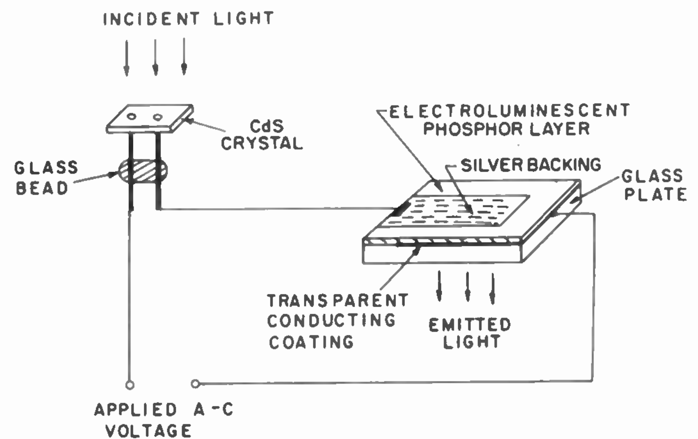


Fig. 1—Elemental light amplifier.

The experiment described used small single crystals of CdS. Subsequently, however, a powder was developed which could be used for making large-area photoconductive cells.[2] This material enabled the practical construction of a thin large-area light-intensifying panel capable of picture reproduction.

* Original manuscript received by the IRE, September 1, 1955.
† RCA Laboratories, Princeton, N. J.
[1] E. C. Payne, E. L. Mager, and C. W. Jerome, "Electroluminescence, a new method of producing light," *Ill. Eng.*, vol. 45, pp. 688–693; November, 1950.

[2] F. H. Nicoll and B. Kazan, "Large-area high-current photoconductive cells using CdS powder," *Jour. Opt. Soc. Amer.*, vol. 45, pp. 647–650; August, 1955.

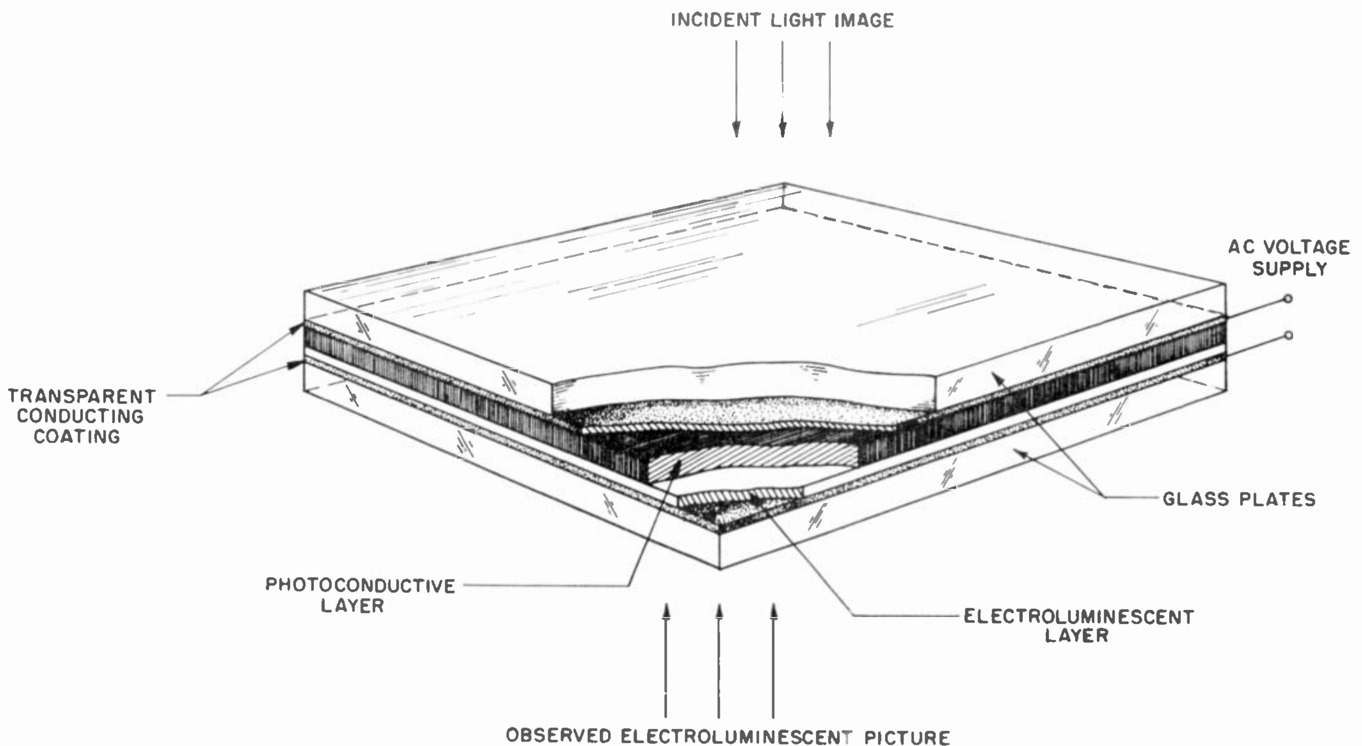


Fig. 2—Sandwich type light amplifier.

The structure initially selected was sandwich-like, using a thin photoconductive layer and a layer of electroluminescent material held between two pieces of glass as shown in Fig. 2. A similar sandwich-like structure was also suggested by another laboratory.[3] Extensive experiments with such structures showed numerous shortcomings, in particular the severe limitation in gain caused by the opacity of the thick photoconductive layer required to control the electroluminescent layer. The development of new structures was thus required.

Interdigital-Electrode Type Amplifier

To illuminate the photoconductor more efficiently, structure of Fig. 3 (next page) was devised. A set of fine interdigital transparent conducting electrodes on a glass plate was covered with a thin continuous layer of electroluminescent phosphor. Above phosphor, another thin layer of photoconductive powder was deposited. With ac voltage applied between two sets of electrodes in the dark, there is relatively little ac current flow between them due to the high impedance of the materials between the conducting lines. When portions of the photoconductive layer are illuminated, however, relatively low impedance paths are provided at these areas between the electrodes. Since the resulting increase in ac currents must flow through the phosphor layer, which is between the photoconductor and the electrodes, electroluminescent light is emitted adjacent to the corresponding illuminated areas of the photoconductor.

In operation, it was found that most of the photocurrents flowed only through the regions of the phosphor layer adjacent to the edges of the electrodes. This reduced the sensitivity and made it difficult to make the impedance of the phosphor elements low enough with respect to the illuminated photoconductive elements. Although the principles were sound, it appeared possible to obtain greater light gains with other structures.

Mesh-Supported-Photoconductor Type Amplifier

One such structure is in Fig. 4 (p. 1891). A transparent conducting coating on glass was sprayed with a thin phosphor layer. To prevent light feedback, a thin insulating opaque layer of lampblack in binder was then sprayed on the phosphor layer. A fine metallic mesh[4] was then stretched approximately 15 mils above the surface of the opaque layer. The photoconductive powder in a binder was then spread over the mesh, forming a curved surface within each mesh hole, to improve light penetration. The above device was relatively effcient, but was difficult to make uniform.

Ridged-Photoconductor Type Amplifier

Another arrangement is in Fig. 5 (p. 1891). A grooved lucite plate, with many fine parallel "V" grooves, has a fine conducting line of silver paint at the bottom of

[3] R. K. Orthuber and L. R. Ullery, "A solid-state image intensifier," *Jour. Opt. Soc. Amer.*, vol. 44, pp. 297–299; April, 1954.

[4] "Lektromesh," 25 square holes to the linear inch, made of nickel, with about 50 per cent transmission.

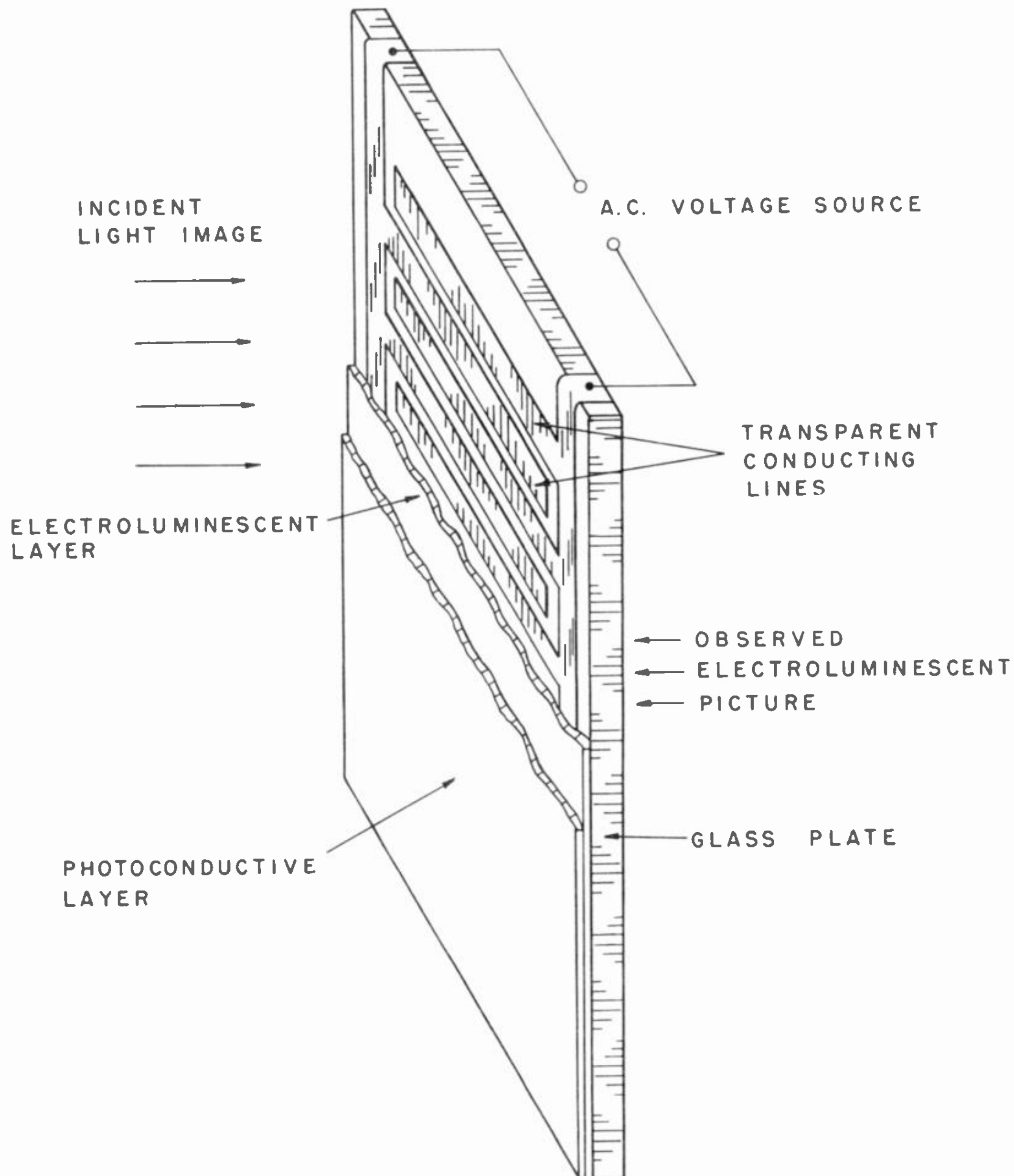


Fig. 3—Interdigital type light amplifier.

each groove. These grooves are filled with a bonded photoconductive powder. An opaque current-diffusing layer (consisting of a conducting form of CdS powder) is interposed between the grooved plate and the electroluminescent phosphor which is bonded to the transparent conducting coating of a glass plate. The ridges of the photoconductor provide continuous conducting paths for the photocurrents from the electrode at the apex of the photoconductor to the bottom of the photoconductor grooves. By means of the auxiliary current-diffusing layer, photocurrents which otherwise would enter the electroluminescent layer at restricted regions near the groove bottoms are caused to spread or diffuse slightly before entering the phosphor layer. Since the amount of diffusion is limited to about the width of a single photoconductive groove, essentially all of the phosphor layer can be excited by the photoconductive layer and at the same time resolution of device, basically determined by distance between grooves, is not significantly affected. Current-diffusing layer, being opaque, serves additional function of preventing feedback.

This arrangement enables efficient illumination of the photoconductor, and efficient excitation of the phosphor. Devices up to 12 inches×12 inches in area operated with high gain and good resolution but conductivity variations in the dry CdS powder used for the diffusing layer were sometimes evident.

Grooved-Photoconductor Type Amplifier

Structure of Fig. 6 (p. 1892) had the advantageous features of previous structure, but did not require a grooved lucite plate and used a current-diffusing powder

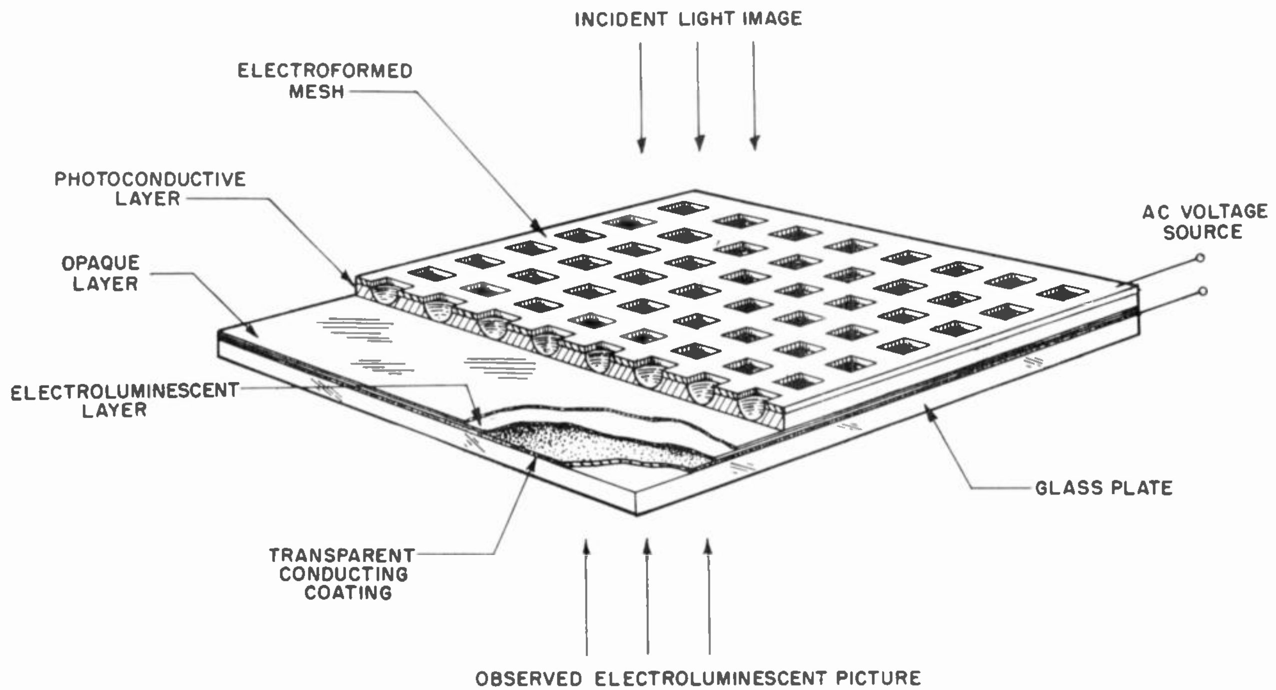


Fig. 4—Mesh-supported-photoconductor type light amplifier.

in bonded form. A transparent conducting coating on a glass plate is sprayed with a thin (about 1 mil thick) phosphor layer. After covering the phosphor layer with an opaque layer (lampblack in Araldite, a fraction of a mil thick), a heavier current-diffusing layer of conducting CdS powder (bonded with Araldite and initially about 20 mils thick) is spread on the opaque layer and machined flat (to about 10 mils in thickness). A heavy layer of bonded photoconductive CdS powder (bonded in Araldite approximately 20 mils thick) is spread on the conducting CdS and again machined flat (to about 14 mils). After spraying the photoconductor surface with air-drying silver paint, fine "V" grooves (of about 60 degrees included angle) are cut into the photoconductor, (15 mils deep, and 25 mils between centers). The bottom of the "V" grooves cuts slightly into the conducting CdS layer and the tops of the grooves are left with narrow conducting silver lines, several mils wide, which, connected to a common terminal, act as one electrode for the device.[5]

Structures such as described in Fig. 6 were made up to 12 inches × 12inches in area. The gain and resolution equalled those of the grooved lucite plate structure of Fig. 5. In terms of resolution the output pictures were comparable in quality with commerical tv pictures and had very good uniformity.

[5] The structure of Fig. 6 required machinable plastic-bonded layers of photoconductive powder which retained full photosensitivity. This was accomplished by using CN-502 Araldite suitably diluted with diacetone alcohol. It was noted that the plastic binder had marked effects on the response time of certain batches of CdS photoconductive powder. Such changes in response time were progressively acquired as the samples cured over a period of several days, with the photosensitivity also changing. These phenomena are not yet completely understood.

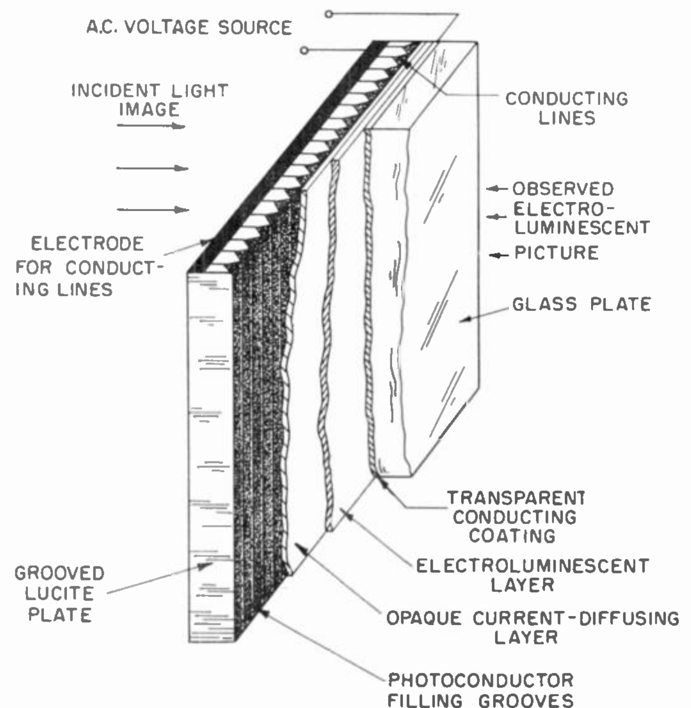


Fig. 5—Ridged-photoconductor type light amplifier.

Fig. 7 (next page) is a photograph of an early 12-inch grooved photoconductor-type amplifier with an amplified output picture produced by projecting a low level image on photoconductor side. Small dark spots are attributed to imperfect bonding of the photoconductive and current-diffusing layers. The few narrow dark lines visible in the picture are caused by disconnected conducting lines.

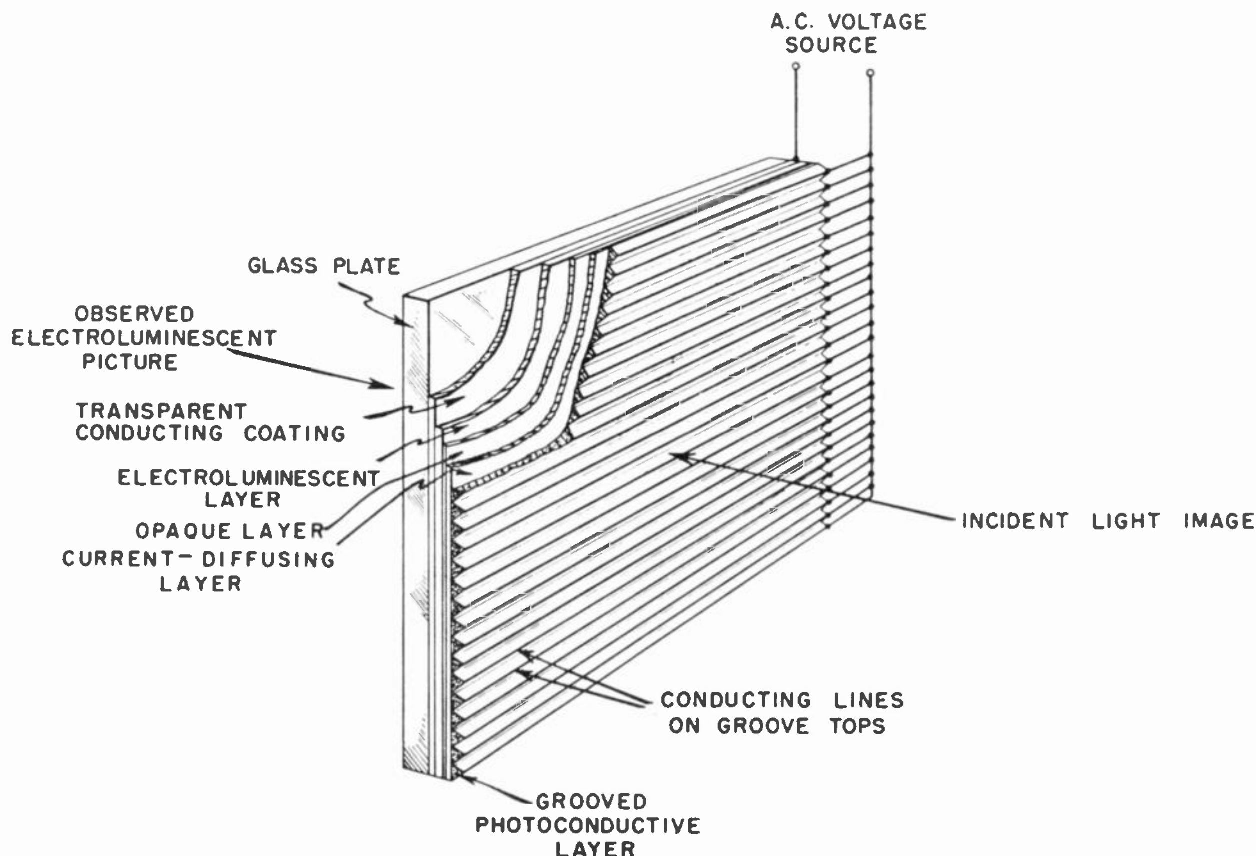


Fig. 6—Grooved-photoconductor type light amplifier.

Fig. 7—Photograph of grooved-photoconductor type light amplifier showing amplified picture resulting from a projected image.

Component Characteristics

Three major components of ridged and grooved photoconductor-type amplifiers are electroluminescent layer, photoconductive layer and current diffusing layer.

The electroluminescent layer is a capacitor with negligible losses at the audio frequencies. For layers 1 mil thick as used, the capacitance is about 200 $\mu\mu$f per cm². When excited with ac voltage, phosphor used emits light whose average value is given approximately by:

$$L_0 = k_1 V^{(3.3)} \omega^{(.7)}, \tag{1}$$

where k_1 is approximately 3×10^{-10}, V is the rms voltage applied, $\omega/2\pi$ is the frequency, and L_0 is luminance in foot lamberts. Fig. 8 (next page) shows a curve of measured values of electroluminescent light output as a function of applied ac voltage at a fixed frequency of 400 cps. Decay time of phosphor upon removal of exciting voltage is relatively short (approximately 1 milli-second). The phosphor response time, however, upon sudden application of voltage depends on its previous excitation and on the audio frequency used, being in the range of one to 30 milliseconds.

The photoconductive powder with a given dc field across it has approximately linear conductivity with incident light intensity. On the other hand, for a given light intensity, the conductivity of a cell varies as a relatively high power of the field. The measured dc current may be approximately written:

$$I_{dc} = k_2 L_i V^n + I_d, \tag{2}$$

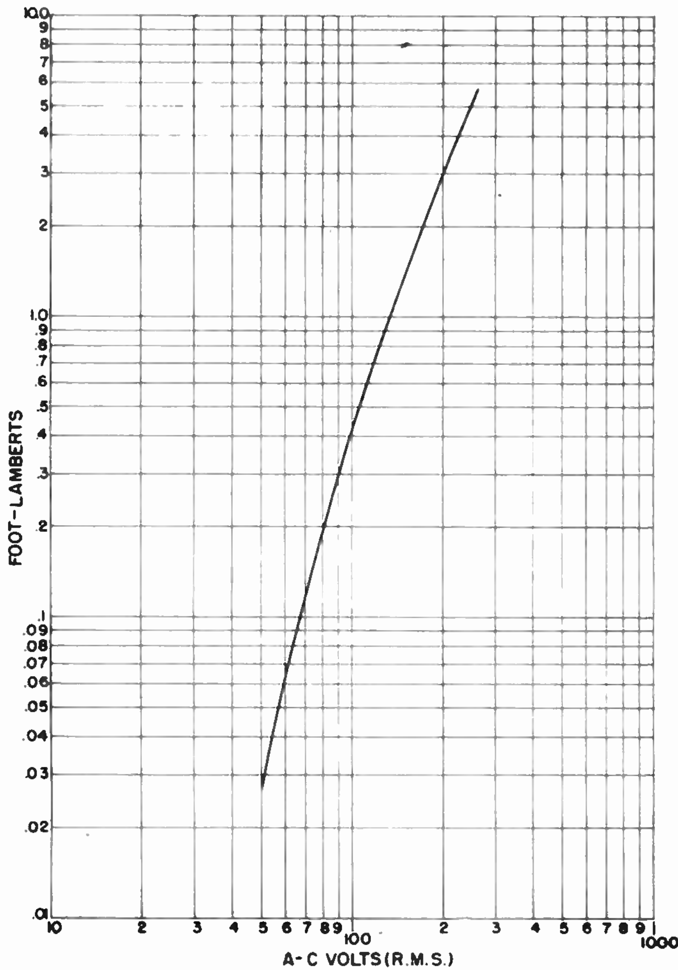


Fig. 8—Curve of electroluminescent light vs applied ac voltage.

where k_2 is a constant, L_i is the input illuminance in foot candles, V is the dc voltage applied, n is approximately 4 or greater, and I_d is the dark current. The dark current, I_d, varies as a high power of the applied voltage, but is usually small compared with the first term.

If an ac voltage is applied to the photoconductor, the instantaneous photocurrent may be represented by the expression:

$$I_{ac} = \frac{k_2}{10} L_i(V_p \sin \omega t)^n - j\omega C V_p \sin \omega t \qquad (3)$$

where $V_p \sin \omega t$ is the applied ac voltage, C is the capacitance of the layer, and $\omega/2\pi$ is the frequency. In general, the dark current under ac conditions is very much smaller than the capacitive current, $\omega C V_p \sin \omega t$, and can be neglected. The constant k_2 for the dc expression (2) has been changed to $k_2/10$ for the ac expression (3). This factor of 10, experimentally determined, is unchanged in the range of 300 to 4,000 cps.

If the measured dc photocurrents are plotted on a linear scale, a curve as shown in Fig. 9 is produced. Over the range of dc voltage used, the photocurrent per cm², neglecting the dark current can be approximately expressed as follows instead of by (2):

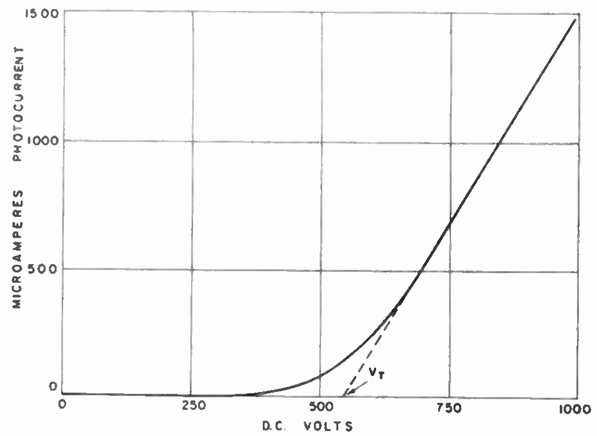


Fig. 9—Curve of photocurrent vs applied dc voltage.

$$I_{dc} = (1/R)(V - V_T) \quad \text{when } |V| > VT$$
$$\text{and } I_{dc} = 0 \quad \text{when } |V| < V_T. \qquad (4)$$

In the above expression V_T represents the threshold voltage; *i.e.*, the point of intersection of the straight-line sloping portion of the curve with the voltage axis and R the incremental resistance of the photoconductor along the sloping portion. Since $1/R$ of the photoconductor varies linearly with the input light intensity, L_i, the above expression can be rewritten:

$$I_{dc} = k_3 L_i(V - V_T) \quad \text{when } |V| > V_T$$
$$\text{and } I_{dc} = 0 \quad \text{when } |V| < V_T. \qquad (4a)$$

For the case of ac applied voltage we obtain:

$$I_{ac} = (k_3/10)L_i(V_p \sin \omega t - V_T) - j\omega C V_p \sin \omega t \quad \text{when } |V_p \sin \omega t| > V_T$$
$$\text{and } I_{ac} = 0 \quad \text{when } |V_p \sin \omega t| < V_T. \qquad (5)$$

A calculated value of $k_3 = 1.25 \times 10^{-4}$ is obtained using the dc curve of Fig. 9. The photoconductive cell used had electrodes of 20 mil spacing, an effective area of photoconductor of 0.75 cm², and was illuminated with 0.035 foot candles (2,870°K Tungsten source).

In the dark, the capacitance per cm² of a photoconductive layer in the grooved lucite structure of Fig. 5 is about 8 $\mu\mu$f measured at 1,000 cps. This capacitance is measured between the set of conducting lines at the bottom of the grooves and an electrode covering the top surface of the photoconductor, thus including the capacitance of the lucite.

The photoconductor used has a response time between 0.1 and several seconds with the shortest response time at the high light levels. Since the photoconductor is much slower in response than the electroluminescent material, it controls the amplifier operation.

Current-diffusing layer, consisting of specially prepared CdS powder, is a nonphotoconductive resistive material. Unlike photoconductor its conductivity is not lowered by ac voltage. Fig. 10 (next page) shows a typical curve of dc current vs dc voltage for a 1 cm² bonded

sample of this material 10 mils thick, with electrodes on opposite surfaces. Since operation of the light amplifier requires currents below one milliampere/cm^2, the voltage drop going through the layer is relatively small (<30 volts).

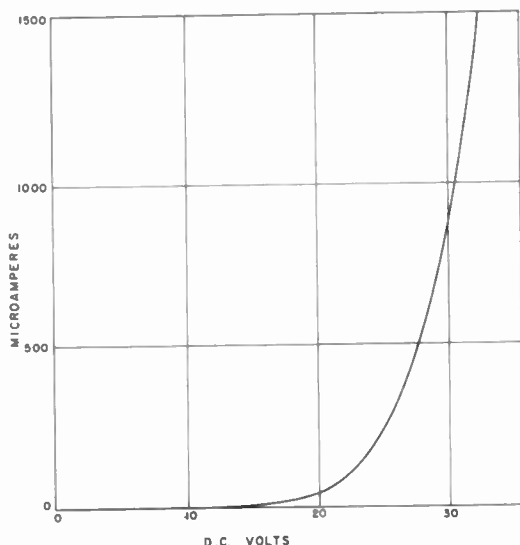


Fig. 10—Curve of current vs applied voltage for current-diffusing layer.

As shown by the curve, the powder has a rapidly increasing impedance with a lowering of the voltage below 30 volts, which is useful for maintaining picture resolution and small-area contrast.

Analysis of the Complete Amplifier

The dark current and the capacitive current through the photoconductor layer are small and can be neglected. The effects of the current-diffusing layer can also be neglected since, as previously indicated, a relatively small voltage builds up across this layer at any time. With these assumptions, the problem simplifies to a series combination of a nonlinear resistor (controlled by the incident light) and a capacitor with ac voltage applied across the two. Using (3) for the photoconductor, the following differential equation applies.

$$dq/dt = \frac{k_2}{10} L_i [V_p \sin \omega t - q/C]^n \tag{6}$$

where C = capacity per cm^2 of phosphor layer, q = instantaneous charge on phosphor layer, and V_p = peak value of applied ac voltage. Because (6) is awkward to solve, it is more convenient to use the approximation, (4). This leads to:

$$\frac{dq}{dt} = \frac{V_p \sin \omega t - V_T - q/C}{R} \qquad |V_p \sin \omega t - q/C| > V_T$$

$$\frac{dq}{dt} = 0 \qquad |V_p \sin \omega t - q/C| < V_T \tag{7}$$

where V_T = threshold voltage, R = ac resistance of photoconductor ($=10R_{dc}$). Solution of this equation[6] gives for the charge, q, on the condenser:

$$q = -V_T C + \frac{V_p C(\sin \omega t - RC\omega \cos \omega t)}{1 + R^2C^2\omega^2}$$
$$+ \left\{ V_T C + q_1 - \frac{V_p C}{1 - R^2C^2\omega^2} \left[\frac{V_T}{V_p} + \frac{q_1}{CV_p} - RC\omega \sqrt{1 - \left(\frac{V_T}{V_p} + \frac{q_1}{CV_p} \right)^2} \right] \right\}$$
$$\cdot \left\{ e - \frac{1}{RC} \left[t - \frac{1}{\omega} \arcsin \left(\frac{V_T}{V_p} + \frac{q_1}{CV_p} \right) \right] \right\} \tag{8}$$

where q_1 = charge on condenser at the start of a conduction cycle.

Of interest is the maximum value of q which determines the peak voltage on the capacitor (phosphor). Tests have shown that the light output of the electroluminescent layer with the nonsinusoidal wave shapes across it in the light amplifier, is approximately the same as that obtained with a sine wave of equal peak value. Since q reaches a maximum at the end of each conduction period, a solution is desired for its value at this time under cyclic conditions. However, under cyclic conditions the charge at the end of a conduction cycle must be minus that at the beginning of the cycle. Suppose q_1 is the charge at the end of the last conducting cycle so that at the end of the new conducting cycle $q = -q_1$. The time t, corresponding to the end of the new cycle can be determined from the relation given with (7) for the conducting condition: $|V_p \sin \omega t - q/C| > V_T$. This relation gives, at the end of the conducting cycle,

$$V_p \sin \omega t + q_1/C = V_T \text{ or } t = \frac{1}{\omega} \arcsin \left(\frac{V_T}{V_p} - \frac{q_1}{CV_p} \right).$$

Substituting this value of t into (8) and using $-q_1$ for q leads to the condition:

$$-x + b - a\sqrt{1-(b-x)^2} = [x + b + a\sqrt{1-(b-x)^2}] \cdot e^{-a[\arcsin(b-x) - \arcsin(b+x)]} \tag{9}$$

where

$$x = \frac{q_1}{V_p C},$$

$$a = \frac{1}{R\omega C}$$

and

$$b = \frac{V_T}{V_p}.$$

[6] Solution of the equation and formulation of expression (9) were provided by E. G. Ramberg.

For very large a,

$$x \cong b - \frac{a}{\sqrt{a^2 + 1}}$$

for very small a,

$$x \cong \frac{-a(1 - b^2)}{\sqrt{1 - b^2 + ab}}.$$

In (9), x represents the fraction of the peak ac supply voltage which appears on the capacitor when it has its maximum charge. Eq. (9) can be solved by trial and error giving a curve such as shown in Fig. 11 of x vs a. This particular curve was plotted with an assumed value of $b = V_T/V_p$ of 0.65. Assuming a given peak ac voltage, V_p, actual values of peak voltage on capacitor can be determined from Fig. 11 as factor, a, is varied.

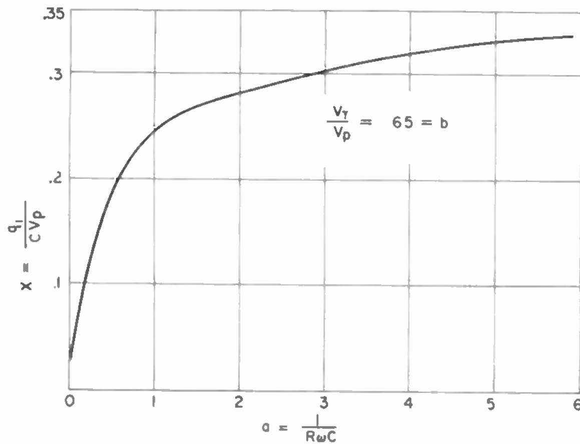


Fig. 11—Curve of fraction of peak voltage across electroluminescent layer vs $a(=1/R\omega C)$.

Converting the values of x in Fig. 11 to light output, using the curve of Fig. 8, produces the calculated curve of Fig. 12, assuming a 1,000 volt peak ac supply. The values of a are indicated at the top of Fig. 12.

Referring now to (5) for the photoconductor with $k_3 = 1.25 \times 10^{-4}$, the factor, a, can be written in terms of the input light on the photoconductor, [assuming $C = 200$ $\mu\mu$f per cm^2 and $\omega = 2\pi(400)$]:

$$L_i = .04a.$$

The corresponding input light levels in lumens per square foot are shown at the bottom edge of Fig. 12. Also shown in Fig. 12 is a curve of measured values of output and input light using a grooved lucite plate amplifier, such as shown in Fig. 5, operated at 400 cycles with an applied ac voltage of 1,000 volts peak.

Fig. 13 shows the curves of light gain as a function of input light level, derived from the curves of Fig. 12. Although Figs. 12 and 13 indicate fairly good correlation between the measured and calculated values, such correlation was obtained by using a value of $b = 0.65$. This assumes a threshold voltage for the photoconductor of 650 volts as compared to the threshold voltage of about 550 volts shown in Fig. 9 for the photoconductor sample measured. The discrepancy is believed to be partially due to the voltage drop across the current-diffusing layer being somewhat greater than assumed, since the current flow into it from the photoconductor is concentrated along narrow lines rather than spread

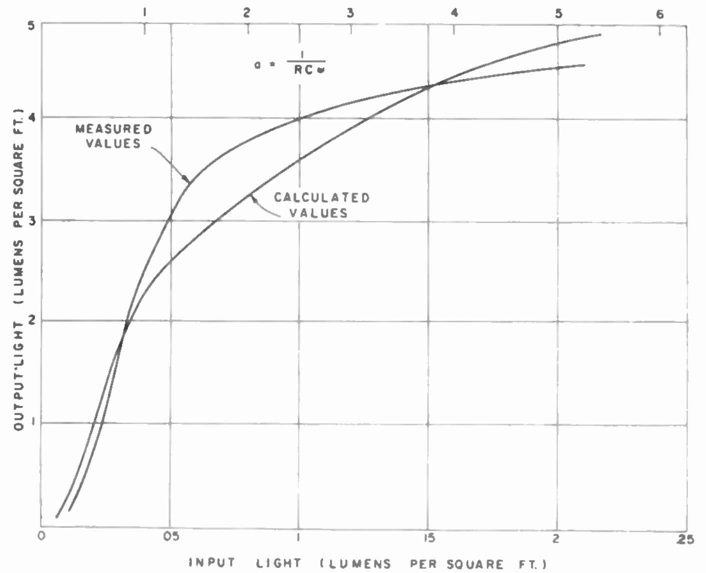


Fig. 12—Curve of output lumens vs input lumens for light amplifier.

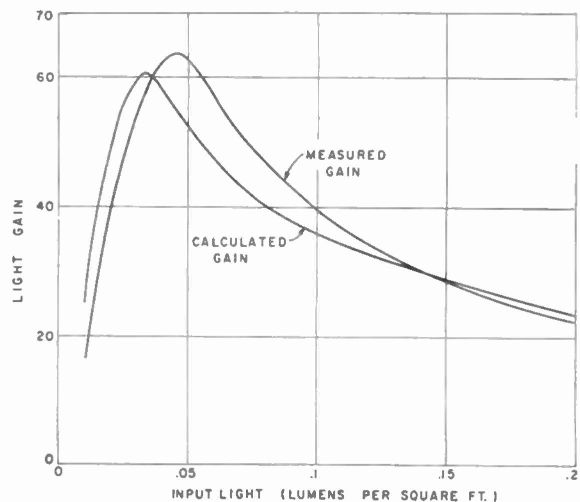


Fig. 13—Curve of gain vs input light for light amplifier.

over the surface as is the case when solid electrodes are used in measurements. In addition, correlation of photoconductor resistance with light level is somewhat inaccurate due to the time constants and slow drift of the photoconductor at low light levels. Although the calculated operation of the light amplifier is approximate, the model assumed for the photoconductor is reasonably useful.

Both the measured and calculated curves of Figs. 12 and 13 are based on the use of a tungsten light source operating at 2,870°K. Such a light source emits a considerable amount of infrared radiation to which the photoconductor is sensitive, but which is not measured by a foot candle meter. To measure the "intensity gain" of the light amplifier, *i.e.*, the gain using input light whose spectral distribution is identical to the output light, a test was made using an electroluminescent source panel to provide the input light. This source had a spectral distribution identical to that of the amplifier output, having been made with the same type of phosphor. Under these conditions, the maximum energy gain was about 14.

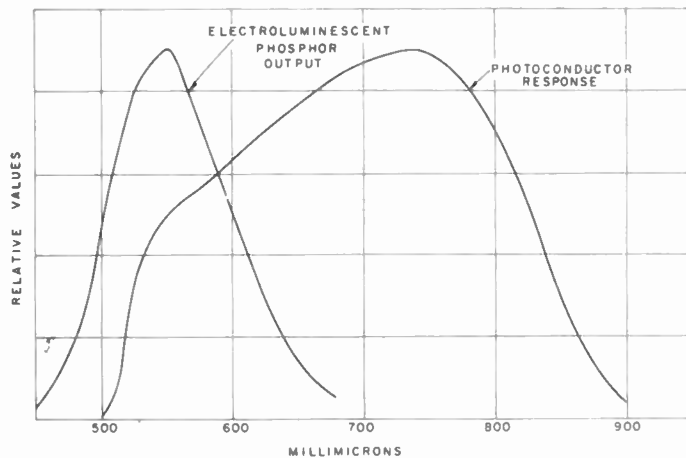


Fig. 14—Spectral curves for photoconductor and electroluminescent phosphor.

Fig. 14 shows in relative values the spectral response curve of the photoconductor and the spectral distribution curve for the electroluminescent phosphor used in the light amplifier. These curves indicate that the operation of the light amplifier with the yellow electroluminescent input light reduces the gain to about 50 per cent of the gain expected with an input light source having its peak coinciding with that of the photoconductor. With such matching conditions, the maximum energy gain is about 30.

Gain and Light Output as Affected by Supply Frequency

Both the measured and calculated gains are based on the use of an ac voltage of 400 cycles. The effects of varying the frequency can be seen by reference to Fig. 12. Increasing ω and decreasing R by a factor, K, causes no change in a $(=1/R\omega C)$ and in the peak voltage applied to the phosphor layer. Such an increase in the values of $1/R$ for the photoconductor, however, requires that the input light levels be multiplied by this same factor. Since the voltage on the phosphor layer is changed in frequency by a factor, K, the light output is increased by the factor $K^{(.7)}$ as indicated by (1). If an audio frequency, f, other than 400 cycles is used, the amplifier will have new gain curves of the same shape as in Fig. 14, but changed in level by the factor given below:

$$\gamma = \left(\frac{400}{f}\right)^{.3} \tag{10}$$

Also, the input light levels will be multiplied by the factor $f/400$ for corresponding points on the curves.

As an example of the effects of a frequency change, assume that the operating frequency is 10,000 cycles instead of 400. From (10) the gain at all points of the curve, including the maximum gain will be reduced to 38 per cent of the gain at 400 cycles. At the point of maximum gain, the input light will be 1.1 foot candles and the output light 27.5 foot lamberts. Similar considerations show that operation below 400 cps will increase the gain and lower the levels of input and output light at the point of maximum gain. At the high audio frequencies, considerable heat is generated in the photoconductor, for example, at 10,000 cps the power dissipation is about 1 watt per cm^2.

The Use of Feedback

The ridged photoconductor panel (Fig. 5) and the grooved photoconductor panel (Fig. 6) amplifiers have been designed to operate without feedback. However, a fraction of the output light may be permitted to excite the photoconductor, for example, by eliminating the opaque layer and making the current-diffusing layer thin. In determining the gain with feedback, the fact that the feedback light from the phosphor is of a different spectral distribution than the input light must be considered (factor k) and also the fact that the photoconductive layer may not be illuminated from the phosphor side with the same efficiency as from the input side when, for example, the photoconductive layer is grooved (factor B). With these considerations, the gain with feedfack, G_f, is given by:

$$G_f = \frac{G}{1 - kBG} \tag{11}$$

where G is the gain without feedback, k is the ratio of photoconductor responses to output and input light respectively due to spectral differences, and B is the effective fraction of the output light capable of exciting the photoconductor. Since the gain, G, of the amplifier varies with the total input (or output) light (11) is valid only if the output is maintained constant.

The effect of limited amounts of feedback is thus to increase the gain as well as to make the gain curve more peaked. If the factor kBG is made greater than unity, a picture element, once excited by an input signal, will remain on regeneratively. Two of the primary problems with such storage devices are: (1) providing efficient feedback illumination of the photoconductor, and (2) preventing adjacent elements from exciting each other.

Operation with Infrared and X-Ray Input Pictures

Although much of the work on the light amplifier as described above has been concerned with its operation using input pictures of visible light, some tests have been made using both infrared and X-ray input images. Inspection of the spectral sensitivity curve (Fig. 14) for the photoconductor shows that in the near infrared (700–900 millimicrons) the powder has sensitivities comparable with those for wave-lengths in the visible region. For wavelengths further in the infrared the amplifier requires new photoconductive materials.

Tests with X rays indicate that some amplification can be achieved with direct X-ray excitation. It is also possible to operate the amplifier with a fluoroscope screen close to the input side so that the visible image formed on the fluoroscope screen will be intensified by the amplifier.

Future Improvements

Further development of electroluminescent phosphors will directly benefit the light amplifier. The sensitivity of the photoconductive powder is considerably lower when operated with ac voltages as compared to dc. Since this effect is not inherent in photoconductors but a property of the particular powder, substantial increases in light gain can be expected with improved photoconductors. In addition, the nonlinearity of the photoconductive powder further limits the effectiveness of the present amplifier. Increase of the photo-conductor breakdown voltage should serve to produce substantial increases in gain because of the rapidly increasing light output of the phosphor with voltage. The response time of the amplifier is also subject to improvement with new photoconductive materials.

Acknowledgment

We wish to thank Dr. V. K. Zworykin for his active support of the work and to acknowledge the interest shown by E. W. Herold in the project. Discussions with D. W. Epstein, E. G. Ramberg, and A. Rose during the course of the work are appreciated. The project could not have succeeded without the contributions of S. M. Thomsen in developing the photoconductive powder and S. Larach in developing the electroluminescent phosphor. The assistance of P. J. Messineo is also acknowledged in the fabrication of the electroluminescent layers.

Opto-Electronic Devices and Networks*

E. E. LOEBNER†, SENIOR MEMBER, IRE

Summary—**The transducing properties of electroluminescent and photo-responsive cells are described. The light amplifying and spectrum-converting characteristics of a circuit consisting of an electric power supply and a series combination of the two types of cells whose impedances have been matched are discussed. The construction and operation of an opto-electronic bistable device—the "optron"—employing positive radiation feedback, is reported. The optron is both a storage cell and a switch with dual (optical and electrical) signal input and output. Numerous logic networks, composed of electric series and parallel combinations of electroluminescent and photoconductive cells with selected optical couplings have been designed, constructed, and operated.**

Introduction

THE DISCOVERIES of electroluminescent phenomena and recent remarkable advancements in the manufacture of electroluminescent materials and cells [1], together with the progress currently being made in the photoconductive branch of solid state science [2–5], have opened up practical possibilities for a new class of solid state devices. The solid state analogs of cathode ray imaging and pick-up tubes, and light amplifying [6, 7] as well as color and radiation converting [8] screens are only some of the simpler embodiments of possible opto-electronic (optronic) devices and systems.

Proper matching of electrical and optical characteristics of electroluminescent and photoconductive cells and their incorporation into circuits with electrical and optical branches can give rise to active optronic devices, such as radiation amplifiers, bistable elements for switching and storage, optronic flip-flops, counters, shift-registers, etc. A remarkable similarity to magnetic amplifier circuits is exhibited by these systems, especially when their operation is based on ac electroluminescent and photoconductive phenomena. The optical branches of the systems enable a complete isolation of various electrical subsystems, as well as a simultaneous input and output to or from a very great number of elements. The most attractive features seem to be the possible microminiaturization (which, in the present state of the art, is the ultimate in miniaturiza-

* Original manuscript received by the IRE, October 5, 1955.
† David Sarnoff Research Center, RCA Laboratories Division, Princeton, N. J.

tion) resulting in small weight and extremely low power consumption, and the optical input and output which invite photographic techniques. Presently, the most serious limitation appears to be the speed of response of the photoconductors, especially its dependence upon the intensity of illumination.

This paper reviews briefly the phenomena of electroluminescence, photoconduction, and allied effects, such as infra-red quenching. A discussion is given of the characteristics of a number of opto-electronic transducers, amplifying devices, especially their spectral response and spectral emittance, inclusive effects upon the radiant and luminous power gains of the devices. The optical spectrum characteristics of the input and output become of paramount interest in bistable devices which utilize positive radiation feedback, or trans-radiance. Various circuit applications of optronic devices are discussed in the section on networks.

Fundamentals of Solid State Opto-Electronic Phenomena

Solid state electronic phenomena can be conveniently described using the simplified one-dimensional energy band model. In this model, imperfect semiconducting and insulating crystals are depicted to have an energy band gap, whose width in electron volts is determined by the atomic and electronic configurations of the host lattice, while the density and energy distribution of localized levels (some of them in the forbidden energy gap region) depend upon the density and the type of numerous crystal imperfections. Electrochemical inhomogeneities, which extend over regions that are large compared with the interatomic distance, give rise to buckling of the energy bands, since the potential energy-displacement relationship is modulated by local space charge. The energy variations of the band edges hinder the transport of mobile charges in applied electric fields and are one of the causes of non-ohmic crystal behavior. The present lack of knowledge of detailed structure of the energy band model applicable to the sulfide family of materials, which are used in the majority of these devices, does not preclude a discussion of opto-electronic phenomena. In this class of phenomena, optical transitions of electronic charge carriers are linked closely with the transport of these charges under the influence of electric fields.

It is well known that the distribution of the electronic population among the crystal energy states, associated with the bands and with the localized levels, is governed by the thermal and the irradiance condition of the substance as well as the probability for the occurrence of all possible radiative and nonradiative transitions between the various states. The approach to a steady-state energy distribution of the electronic population can, in the sulfides, take as long as several days. Admitting the possibility of space charge accumulations and shifts, it is seen that under the influence of varying applied fields and varying illumination, steady-state conditions can be assumed only for limiting cases.

Two of the more prominent opto-electronic phenomena are photoconduction and electroluminescence. They both combine the simpler optical and electronic effects of photo- and thermo-luminescence, and of semiconduction and barrier-type rectification. Other opto-electronic effects could also be included, as long as the application of electrical fields or currents changes the optical behavior of the crystals.

Electroluminescence

Electroluminescence commonly designates "cold light" emission from a substance due to the action of an applied electric field. By cold light or luminescence, we mean the emission of visible light from substances whose temperature is insufficient to cause them to spontaneously emit light, or incandesce. Electroluminescence can be continuous or alternating, depending upon the applied electric field. Thus far only ac electroluminescence, discovered 20 years ago [9], has proved to be of more than purely laboratory interest. The more efficient electroluminophors (substances exhibiting the phenomena of electroluminescence) are compounds in the sulfide phosphor family and, in addition to conventional activation, usually have undergone some special heat treatment in various atmospheres [1]. The detailed mechanism by which energy stored in the crystal electrostatic field is converted into an emitted photon is still obscure. The number of rivaling theories is considerable [10–13], but there are not yet sufficient reliable experimental data to undertake a quantitative check of the theories.

Electroluminous emittance is not uniform on the microscopic scale, but occurs in many bright spots [14, 15] clustered in various patterns which show some relation to the crystal axis [13, 16]. Under alternating field excitation, the luminous emittance of the small electroluminescing spots pulsates at a rate equal to the frequency of the applied voltage [15, 16]. In single crystals under low field excitation, this light output is out of phase with the applied voltage. The integrated output of all light emitting spots excited with low voltages and at low frequencies can be seen as a function of time in Fig. 1 together with the applied voltage and the crystal current.

The time-averaged electroluminous emittance increases faster than linearly with increasing amplitude of applied sinusoidal voltage, as in Fig. 2. Many electroluminescent cells have been investigated and various empirical formulas have been fitted to the data. However, in no case could a satisfactory fit be obtained over the entire range of experimental values.

The spectral composition of the electroluminous emittance in zinc sulfide phosphors is characteristic of the particular activator systems present in the crystals. Thus relatively broad band emissions ranging from deep blue to orange have been obtained at room temperature [17].

It should be pointed out that the equivalent circuit of an electroluminescent cell resembles to a large extent that for crystal rectifiers. This is not very surprising since it is generally accepted that electroluminescence is localized in or near rectifying barriers. Complex elec-

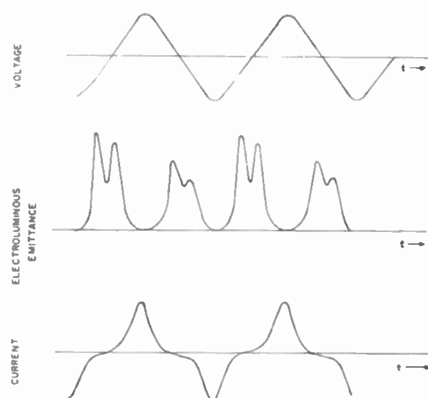


Fig. 1—Time dependence of three variables of an electroluminescing ZnS single crystal under low voltage, low frequency excitation.

trical and optical characteristics of electroluminescent cells preclude use of conventional ac theory in analyzing electronic circuits embodying these cells. Hence, only a qualitative treatment is given in this paper.

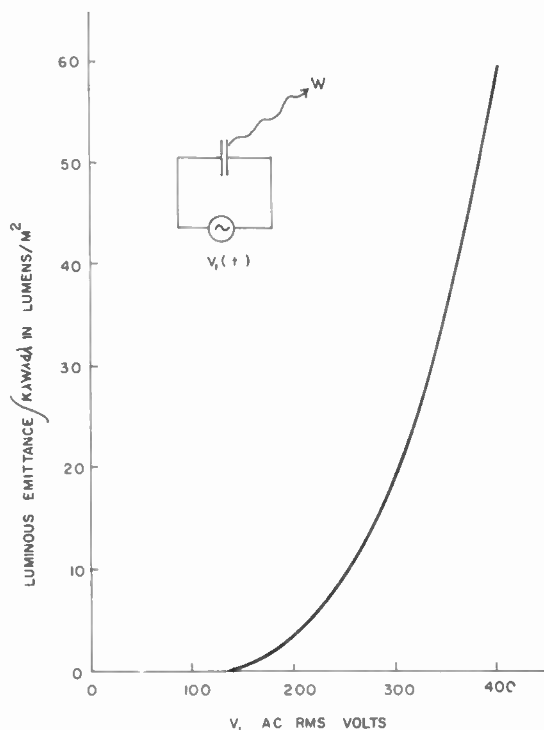


Fig. 2—Time averaged luminous emittance $\int K_\lambda W_\lambda d\lambda$ of an electroluminescent cell as a function of sinusoidal, 60 cps applied voltage V_1.

AC Photoconduction

It has been known for the last 80 years that resistance of certain substances can be lowered by exposure to visible or to nonvisible radiation. Such substances are known as photoconductors. Although there is considerably more known about the photoconductive processes than about electroluminescence, a quantitative theory of the detailed opto-electronic phenomena occuring in actual crystalline substances is not yet firmly established. Many photoconductors also show a non-ohmic voltage-current dependence. An illustration is given in Fig. 3, which shows the ac voltage, the photocurrent, and the dark current of a CdS crystal to which silver paint electrodes have been applied and which was exposed to steady low-level white illumination. An additional dc voltage bias can eliminate the photocurrent of one polarity and enhance the other. Similar effects can be obtained by illuminating the crystal non-uniformly. Thus it becomes very difficult to analyze the behavior of photocells in ac circuits if other non-linear rectifying and complex elements are present.

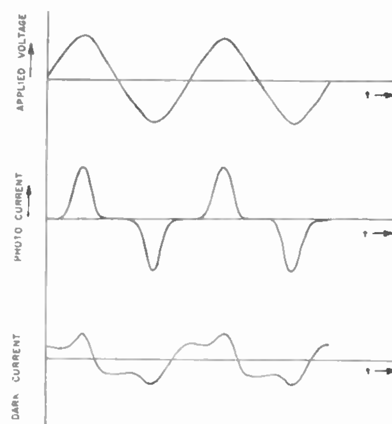


Fig. 3—Time dependence of three variables of a photoconducting CdS single crystal under an applied low frequency ac field and excited by constant radiation of low intensity.

If a photoconductive cell is placed in series with a constant impedance and an ac power supply, the amplitude of voltage appearing across cell will be a function of the irradiance of the cell. One possible relationship is shown in Fig. 4 (next page). V_S is the sum of the instantaneous voltages V_X and V_D, which are seldom in phase. The amplitude and shape of V_X depends upon the frequency of V_S, the impedance Z_D, and the thermal and irradiance condition of the photocell.

Other Radiation Effects

It should be realized that the behavior of electroluminescent and photoconductive cells is actually more complex than it might appear from the discussion above. In addition to pure electroluminescence, commercial cells exhibit a number of other effects which might considerably alter the functioning of systems in which they are utilized. If an electroluminescent cell has been exposed to radiation of shorter wavelengths prior to the application of the electric field, photoelectroluminescence can occur, that is, radiation is emitted at fields well below the threshold value of pure electroluminescence. However, the duration of this radiation extends over only a limited number of cycles. Certain

materials are known to exhibit a quenching of photoluminescence when electric fields are applied to them. The photoluminescence accompanying a steady irradiance can be strongly modulated by ac electric fields. Exposure to radiation of long wavelength can produce a strong and long-lasting quenching of electroluminescence.

Infra-red radiation incident on a photoconducting material can stimulate or quench photoconduction.

Usually all the effects described above are present to some extent. However, their time, voltage, intensity, and temperature dependencies are not necessarily the same, and thus they present a challenge to the designer of opto-electronic equipment.

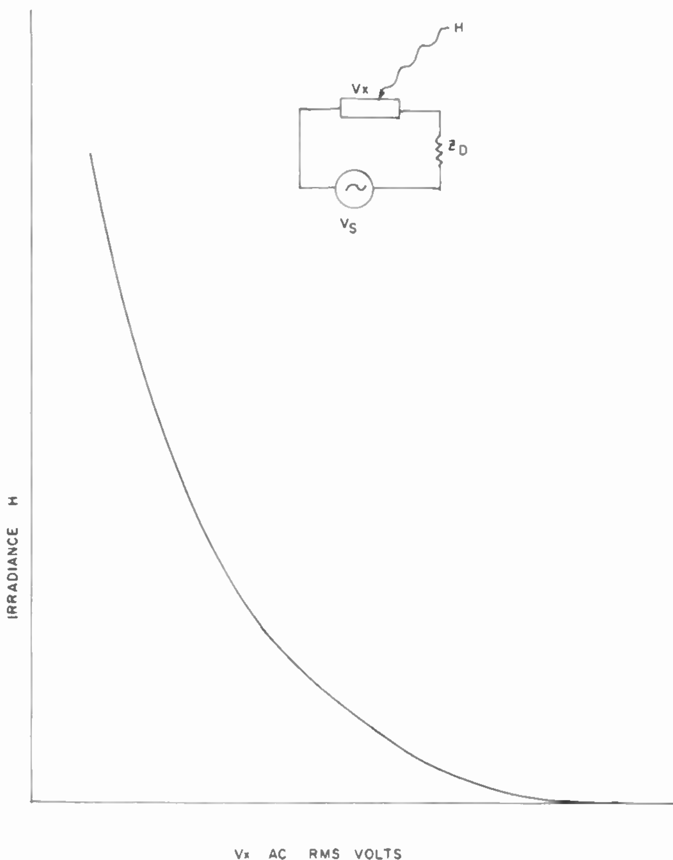


Fig. 4—Irradiance of photoconductor as a function of voltage across it was ac voltage applied to photoconductor and series impedance. Units of variables are arbitrary.

It is hoped that this description conveys the impression that even the most modest beginnings of device utilization of all these effects present a number of unconventional problems, and that further experimental and theoretical investigations into the fundamentals of opto-electronic solid state phenomena will be necessary before a wider device application can be achieved.

Light Amplifiers and Transducers

The recent progress in transistors has focused attention on solid state devices. The transistor is an amplifying device in which a current signal is transmitted at an increased magnitude. Similarly, it is possible to define a light amplifier as a device capable of being actuated by a light signal from external sources and of emitting a related signal at an increased magnitude. It is well to remember that the accepted definition of light is radiant energy evaluated in proportion to the luminous sensation produced by it [18]. Therefore, we find it convenient to distinguish between two kinds of gains: the *luminous* and the *radiant* power gain.

If the ultimate detecting device is a human eye, the luminous power gain is a very useful concept. If we are interested merely in power gain of the device, radiant power gain seems more appropriate. Luminous power gain is defined as the ratio of the related luminous emittance of a specified spectral power distribution to the activating illuminance of a specified spectral power distribution under specific operating conditions of the light amplifiers. Radiant power gain is defined similarly, substituting radiant for luminous and irradiance for illuminance. The radiant power gain is thus given by

$$G^R = \frac{\int_0^\infty W_\lambda d\lambda}{\int_0^\infty H_\lambda d\lambda} \tag{1}$$

where $W_\lambda d\lambda$ (watts/unit area) is the radiant emittance in a wavelength band extending from $\lambda - \frac{1}{2}d\lambda$ to $\lambda + \frac{1}{2}d\lambda$, and H_λ represents the spectral irradiance. The luminous gain is given by

$$G^L = \frac{\int_{\lambda_1}^{\lambda_2} K_\lambda W_\lambda d\lambda}{\int_{\lambda_1}^{\lambda_2} K_\lambda H_\lambda d\lambda}. \tag{2}$$

Here K_λ (lumens/watt) stands for the luminosity function, and care has to be taken to use the correct values when scotopic vision takes place.

Since the spectral response of light amplifiers can easily extend beyond the region of visible radiation, luminous gain ratios approaching infinity can result. Thus, it was found convenient to arbitrarily set the limits of the integral at $\lambda_1 = 380\ m\mu$ and $\lambda_2 = 760\ m\mu$.

Color conversion may result since the spectral response of the device is broad and usually different from its spectral emittance. If the spectral distribution of W differs materially from that of H, we speak of heterochromatic light amplification or of radiation conversion. The case of identical spectral distributions we have called homochromatic light amplification. It has the distinguishing feature that it assures the equivalence of luminous and radiant power gains.

The light amplifying or radiation converting devices consist of a radiation detecting and a radiation emitting element which are electrically in series with each other and with a dc or ac power supply. The most common arrangement for solid state devices is a photoconducting

cell in series with an electroluminescent cell, as shown in Fig. 5. To simplify our discussion, let us assume for the present that both the emitter and the detector are linear elements and that we can assign a complex impedance to them. Then, if the detector's impedance Z_D is very high in the dark and is predominantly resistive, only a small fraction of the applied voltage V_2 will appear across the emitter. If this voltage fraction is below the voltage threshold V_R of the electroluminescent cell, there is no radiant emittance from the device. But since the impedance Z_D can be controlled by incident radiation, suitable adjustment of the detector's irradiance H, will regulate the emitter's radiant output W within wide limits. The amplification arises from the fact that small changes in the mean irradiance of the detector can vary its impedance considerably, and thus will control relatively large changes in the mean radiant emittance. Radiant power gains of the order of twenty have been measured.

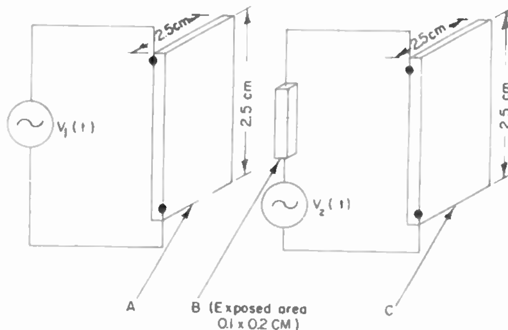


Fig. 5—Light amplifier experimental arrangement: *A* is an electroluminescent area source; *B* is a CdS photo-detecting crystal mounted 0.4 cm above the plane of *A*; *C* is an area electroluminescent source identical to *A*.

It is interesting to note that if the detector has a very high dark resistance, both a decrease in its resistance and the generation of a reactive impedance component will give rise to an increased voltage amplitude across the emitter if the amplifier is subjected to ac rather than dc fields. It should be kept in mind that the mean impedance of the emitter varies nearly inversely with rms ac voltage amplitude. This causes the gain to be considerably lower than would be the case if its impedance were voltage independent. Since the instantaneous impedances of both cells vary throughout a single excitation cycle, the above analysis is only approximate and should be carried out independently for each voltage cycle phase position. It has also been observed that it takes a considerable number of ac cycles for the amplifier to change its output level following activation. In order to prevent distortion of the amplified light signal, the variations in the light input signal should be slow relative to the frequency of the applied voltage. The frequencies used in our experiments varied from 15 to 3,000 cps. The upper limit of the useful frequency range depended mostly on the time constant of the CdS detectors, since the time constants of most electroluminescent devices are considerably shorter. For cases in which the response of the detectors varies throughout the excitation cycle and the light input signal has a frequency close to that of the excitation frequency or harmonics thereof, the relative phase of the light signal and of the excitation voltage affect the resultant power gain of the light amplifier.

Various designs of light amplifiers have been proposed. Both the emitter and the detector can be single crystals, relatively homogeneous thin films, or powders, bonded with or suspended in dielectric materials. If the spectral emittance band and the spectral response band of the device overlap, positive radiation feedback will result when a portion of the light output is allowed to reach the detector.

Solid state light amplifiers have been proposed independently by several workers. Luminous or radiant gains larger than one were reported by Orthuber and Ullery [6], Kazan and Nicoll [19], Cusano [8], and Loebner [7]. Cusano's device is a radiation converter, or light transducer, having a radiant power gain of about 5. Orthuber and Ullery report a luminous power gain of 24. However, their result cannot be converted into radiant gain since they did not specify the spectral distributions of the incident and the emitted light, and whether or not nonvisible radiation was coactuating their device.

At the Sylvania Waltham Laboratories the author has investigated homochromatic light amplification. The experimental arrangement used is depicted in Fig. 5. The electroluminescent panels (*A*) and (*C*) were ceramic matrix, metal base lamps [20], while the photoconducting CdS crystal (*B*) was obtained commercially.[1] Results of luminous gain (in this case equivalent to radiant gain) measurements, performed using an SEI photometer, are seen in Fig. 6 (next page).[2]

Fig. 6 indicates that gains larger than unity occur only if the illumination as well as the applied voltage exceed particular threshold values. The low illuminance threshold of the device could be decreased considerably by using steady illumination bias of about 0.3 lumens/m^2. The occurrence of a threshold field in electroluminescence is chiefly responsible for these amplifier thresholds. The approximate constancy of the gain throughout at least one order of magnitude of illuminance signifies a region of linear amplification, while the subsequent decrease is characteristic of the amplifier's saturation.

One of the most promising applications of light amplifiers are image intensifiers, which consist of screens composed of a multitude of light amplifying elements. By a suitable choice of spectral input and output characteristics of the individual elements, various light conversions of infra-red or of ultraviolet into visible radiation can be obtained.

[1] CdS cells were purchased from the Standard Piezo Co.

[2] For an explanation of the centibel scale see W. B. Nottingham, "Screens for Cathode-Ray Tubes," p. 623 in Soller, Starr, and Valley, "Cathode Ray Tube Displays," McGraw-Hill, N. Y., 1948.

Color television displays could be realized in a screen containing three types of elements distributed in a prescribed manner, by projecting a monochrome image from a small cathode-ray tube which is operated by color deflection circuits. The realization of these and other commercial and noncommercial applications awaits improvement in the response time of light amplifiers.

Bi-Stable Elements

Remembering the characteristics of the homochromatic light amplifier shown in Fig. 6 at an illuminance level somewhat above 0.3 lumens/m², we can see that an increase in the operating ac voltage $V_2(t)$ will result in an increased luminous gain. If the physical arrangement shown in Fig. 5 is modified by permitting light generated in electroluminescent cell (C) to reach photoconductive cell (B), then the operating points of the amplifier will not fall on a vertical line, but with increasing voltage will shift to higher illuminance values.

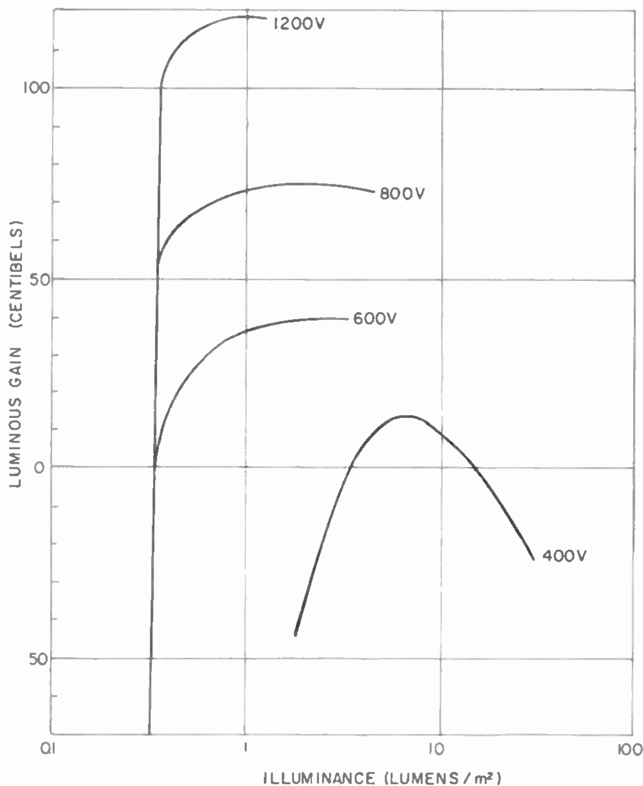


Fig. 6—Homochromatic light amplification. Gain vs illuminance curves obtained with arrangement seen in Fig. 5 at various operating rms 60 cps voltages V_2. A only illuminates B. Sinusoidal voltages V_1 and V_2 are in phase.

Such a light amplifier is regenerative. The radiance incident on the detector (B), and originating from electroluminescent cell (C), whose output in turn is controlled by detector (B), we have termed transradiance. Using standard optical methods, transradiance can be obtained with values varying from zero to a value equal to the radiant emittance of (C), if the areas of (B) and (C) are equal; the value can become even larger if the radiation emitting area of (C) is larger than the light detecting area of (B). In the experiments to be described below, we chose a geometry which resulted in a transradiance, or transluminance equal to the emittance of the device. A more schematic representation of nonregenerative and regenerative light amplifiers can be seen in Figs. 7 and 8, respectively.

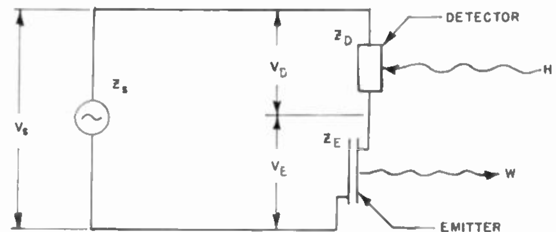


Fig. 7—Schematic of a nonregenerative light amplifier.

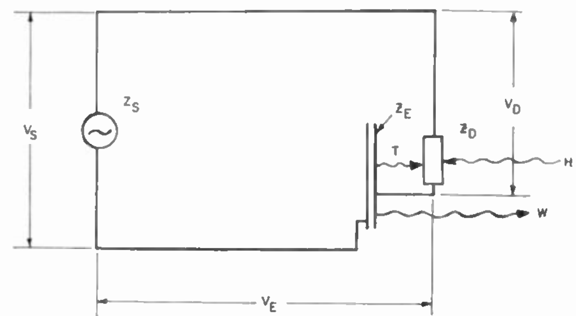


Fig. 8—Schematic of a regenerative light amplifier.

As the voltage across the regenerative device is increased, an operating point is reached for which the sum total of illuminance and transluminance has a value which on Fig. 6 corresponds to a gain close to unity. At this point, the device becomes unstable, the optical feedback exceeds unity, and the emittance, and therefore the transluminance, continue to rise without external influence until the saturation region of the device is reached. The light amplifier will stabilize at another operating point, as seen on Fig. 9 (opposite) where rise values of ac voltage $V_2(t)$ across device is plotted as a function of peak alternating current flowing through the device when biased with about 3 lumens/m².

The functional voltage-current relationship of this bistable opto-electronic device, which, because of characteristics to be discussed later, has been called the "optron," resembles that of certain transistors, double based diodes, etc., used in pulse circuits. However, Fig. 9 represents the dynamic characteristic rather than a static one. If the rms value of voltage is set below 270 volts, the electroluminescent cell (C) emits light very feebly, the optron as a whole has a high impedance and exhibits stable operation. If the voltage is increased to a value above 380 volts, the device will be stable along the voltage-current curve above point B; however, it will then be light emitting. The chosen illuminance bias determines a switching region in the V_2-I_2 plane (bounded by the curves $\overline{AB}$, $\widehat{CB}$, $\overline{DC}$ and $\widehat{DA}$).

If the ac voltage V_2 and the light bias are unchanged, the optron will stabilize at either curve $\widehat{CB}$ or $\widehat{DA}$. Since at a given light bias, and between two limiting ac voltages, the optron has two stable operating points, it is bistable. Nevertheless, it results in the following unstable operation within this region: If the optron is being operated at point E, and the ac voltage V_2 temporarily exceeds 380 volts, the device may stabilize at point E', and becomes light emitting instead of returning to point E. The criteria for stabilization at E' are the length of time that the ac voltage V_2 remains above 380 volts and the amount by which it exceeds this value. These will determine the build-up time and light emissive level of the emitter (C). Another important factor is the response time of the detector (B), which governs the build-up of the photocurrent amplitude and the duration of the phase shift, which will be discussed later.

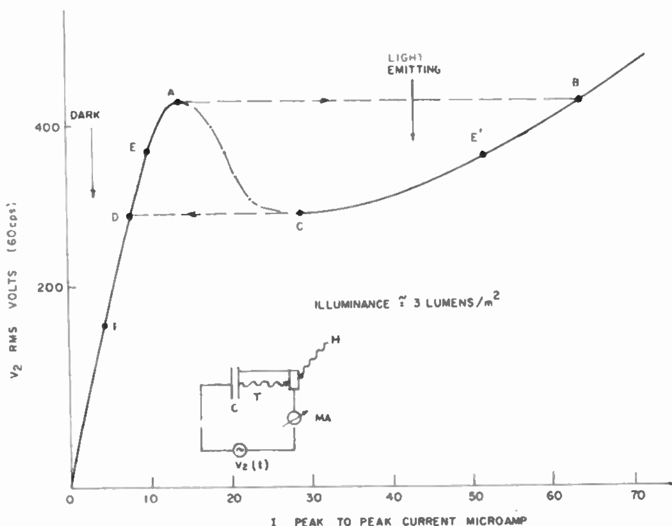


Fig. 9—The dynamic V_2-I_2 characteristic of light-biased optron #12 operated by voltage variation. Area $ABCD$ represents a hysteresis region in which the voltage-current relationship is not unique, but depends upon the operation of the device.

The inverse situation occurs when an optron is operated in the lit condition and the voltage V_2 is temporarily lowered below 270 volts. The optron will then switch back into its dark operating condition if time is allowed for the photoconductivity to decay to a sufficiently low level. The time needed for the device to fully recover its dark operating condition depends not only upon electroluminescence and photoconduction, but also upon other phenomena. Thus photoelectroluminescence, produced in the emitter when V_2 is returned to the level E, can cause the optron to revert back to the lit condition at operating point $E/$. On the other hand, relatively short exposure to infra-red radiation can disable the optron by lowering the photosensitivity of its detector as well as the emissivity of its emitter, so that it resists triggering into the lit condition for an extended period of time.

We have also investigated the current-voltage phase of a number of these devices. This relationship is seen in Fig. 10, which shows the shape of the I_2-V_2 Lissajous figures for optron 10-2. In the high impedance stable condition, the optron has a resistive, although very nonlinear, character. This general shape remains the same up to the triggering voltage, and during this triggering becomes characteristic of a reactive impedance.

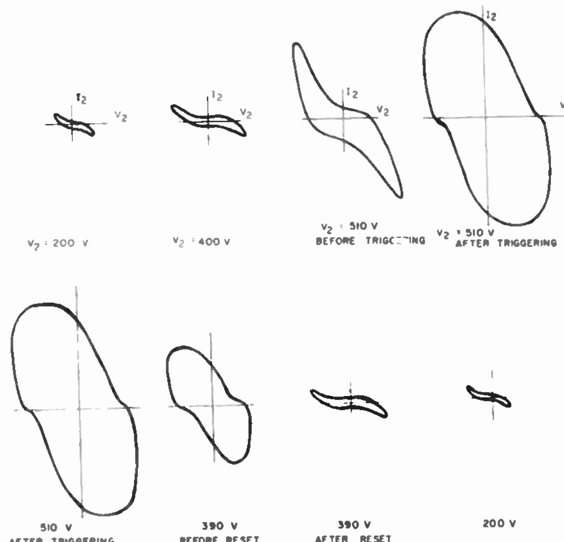


Fig. 10—The dynamic I_2-V_2 phase relationship of light-biased optron #10 operated by voltage variation.

When the voltage is decreased, the reactive character persists to considerably lower voltages, but ultimately returns to the resistive character and high impedance operation. These phase shifts seem to be the result of reactance changes in the detector elements. Since it was previously shown (see Fig. 3) that the detectors respond to light signals only within certain portions of the excitation cycle, a complete analysis of the switching operation should include the relative phase of the electroluminous emittance maxima and the photocurrent maxima.

This observation demonstrated that a simple record of peak to peak ac current through the optron provides insufficient information on the mean current and the effective impedance at various operating points. If a resistance R is placed in series with an optron, the negative impedance region AC on Fig. 9 can be determined approximately. This procedure is not rigorous enough to give accurate results, since the phase relationship between the various voltages in the circuit is very complicated. However, it does give a fair qualitative representation of the negative impedance characteristics of the device.

Fig. 11 shows a family of such voltage-current curves for optron #9 at various illuminance levels, obtained using an electroluminescent cell to provide the biasing light. The graphs show that the mean ac voltage V_2, as well as the interval ΔV_2, in which bistable operation

occurs, both decrease with increasing light bias. This characteristic can be utilized for an alternate triggering mode. If the illuminance of (B) is temporarily increased, the peak voltage, denoting the end of the high impedance stable region, is lowered; and, given sufficient time, the operating point of the device will shift into the low impedance region.

Bi-stable device can be returned to its high-impedance, dark operation by infra-red input light pulse, if its intensity and duration are properly selected.

For the majority of optrons which we constructed, voltage triggering without the use of an illuminance bias would have required voltages well in excess of breakdown in the photocell or the electroluminescent phospher layer. This is indicated by the curve for zero bias, seen in Fig. 11.

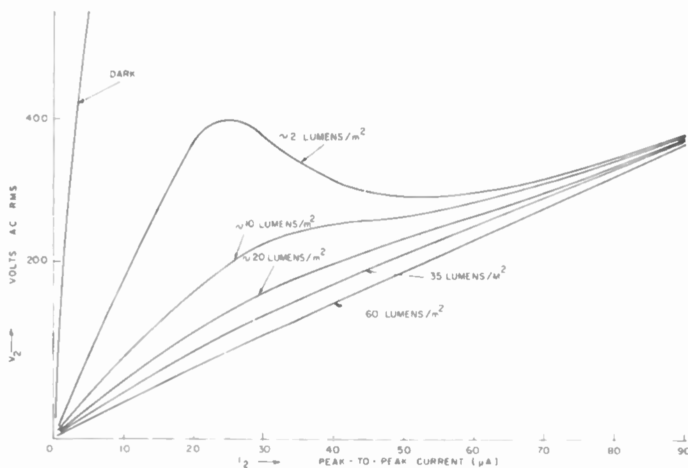
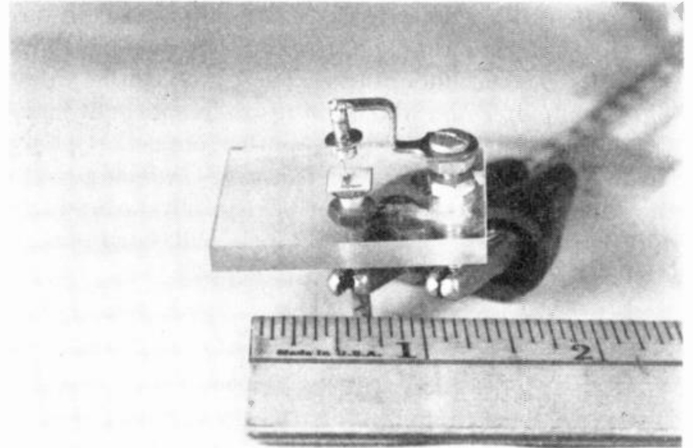


Fig. 11—Operating dynamic ac V_2-I_2 characteristics of optron #9 for various values of light bias.

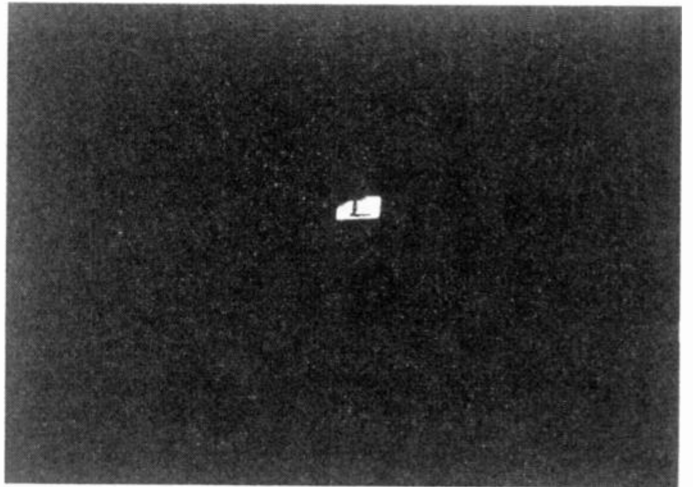
A small optron can be seen in the photograph in Fig. 12. A CdS crystal rests on the transparent electrode of a $\frac{1}{4}'' \times \frac{1}{4}''$ ceramic electroluminescent cell. A small patch of silver paint on the front side of the crystal assures good electrical contact between the two cells. The second contact to the CdS crystal is a cats-whisker coated with silver paint, and is one of the external electrical leads to the device. The other lead of the device is terminated in the metal base which is the second electrode of the electroluminescent cell. The second photograph is a picture of the same as the first, except without illumination. Optrons constructed in a similar manner and smaller than $\frac{1}{16}'' \times \frac{1}{16}''$ have been operated successfully.

Bi-stable elements with optical input and output, to be used in digital computers as memory cells and as switches, probably were first proposed by Marshall, Bowman, and Schwertz [21]. Although they considered electroluminescent and photoconductive phenomena, they did not report the construction of such elements [22]. A solid state optical storage element, for triggering by a voltage pulse, was mentioned by Bramley and Rosenthal [23], while Orthuber and Ullery [6] indicated the undesirability of optical feedback in their solid state image intensifier. Dual characteristics of optrons have not been described to date.[3]

The switching time of optically triggered optrons depends upon the time constant of the photodetector, and thus is an inverse function of illumination intensity.



(a)



(b)

Fig. 12—Photograph of optron #37. (a) illuminated, (b) triggered into electroluminescence.

The triggering interval of the investigated optrons varied from 10 seconds to 2 milliseconds. The development of suitable detector materials, with fast photocurrent decays, might cause the optron to become a serious rival to the magnetic storage core in computer applications. It is usable in its present form for indicator applications which require prolonged storage of

[3] It was called to the writer's attention recently that a working optron, which utilizers a photoconducting CdS film as the detector, has recently been constructed at the M.I.T. Lincoln Laboratory.

visual display information and in which recovery times of the order of hundreds of milliseconds can be tolerated.

As a regenerative amplifier, the optron is capable of luminous or radiant intensity gains of the order of 50. By proper choice of optical arrangement, as little as 10^{-5} of the emitting flux will trigger on optron. Since the limitations of electroluminescence and of photoconduction are not known at the present time, the physical and technological limitations of optrons cannot be estimated.

Logic Networks

The dual input-output characteristic as well as the light pulse integrating ability of optrons, enables the development of unconventional circuits, composed of separate and combined optical and electrical branches, and capable of a multitude of logic operations. We have succeeded in demonstrating that in principle it is possible to build computers having as the only elements electroluminescent and photoconductive phosphors.

The optically controlled storage-cell, composed of an optron whose emitter has been shunted by a second detector, is seen in Fig. 13. A light pulse incident on the series detector D_{11} will trigger emitter E_1 into the lit condition, while a light pulse incident on the shunt detector $D_{2\bar{1}}$ will trigger emitter E_1 into the dark condition. Should electrical input and output be desired, optical binary read-in is provided by grid emitters G_1 and G_2, while detector D_{01} enables optical read-out.

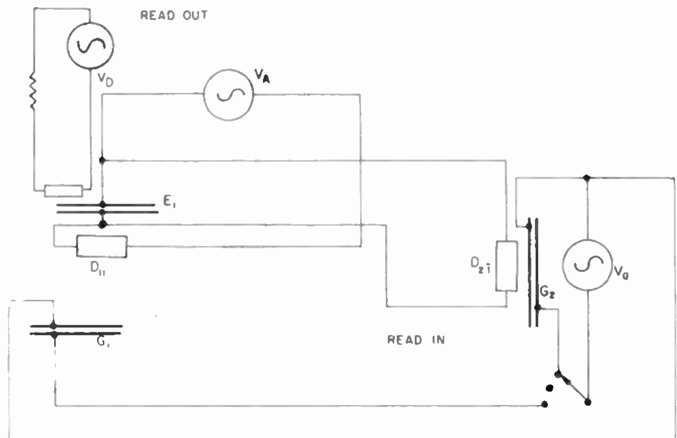


Fig. 13—Optronic storage cell schematic.

Note the complete electrical isolation and solely optical coupling of the read-in, storage, and read-out circuits.

Since the operation of a shift register necessitates the incorporation of several basic computer operations, the successful functioning of an optronic shift register should give an indication of the effectiveness of optrons in gating circuits. Consequently, a shift register was designed and constructed by W. Humphrey, J. Hayes and the writer, according to the schematic diagram shown in Fig. 14. This device consists of two emitter cells and eight detector cells, and is operated by means of three control emitters. Three of the detectors are in parallel with each other and in series with an emitter. Transradiances are adjusted so that it is necessary to illuminate at least two of the three detectors in order to trigger the appropriate half of the optronic shift register circuit into the lit or "on" condition. For this reason, these three detectors have been designated as "half-detectors." The notation for the detectors is as follows: subscript indices preceding the comma refer to transradiances from emitters with matching subscripts, while indices following the comma designate the electrical connection of the detector to the emitters. The bar over an index designates parallel connection, while indices not barred refer to series connections. The functioning of the shift register is as follows: Two emitters

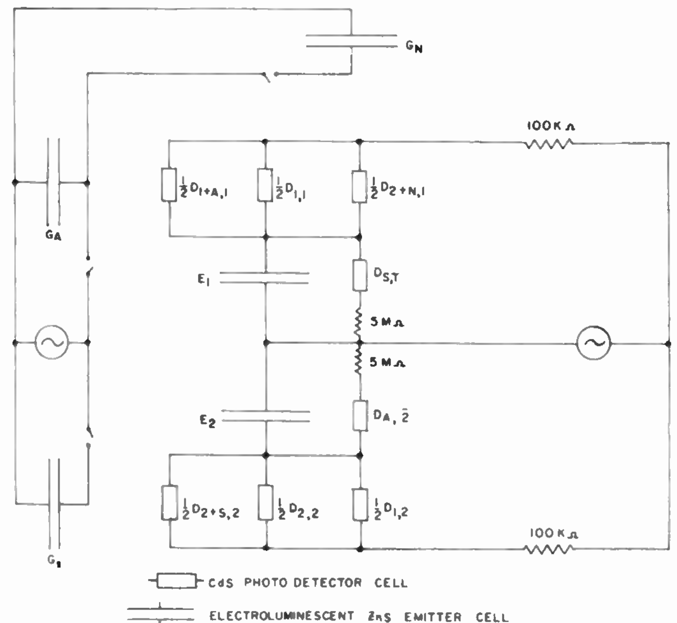


Fig. 14—Optronic shift register schematic. Circuit wiring is true representation, but light channels would have to be drawn in three dimensions. Optical coupling of detectors to emitters indicated by subscripts.

G_N and G_A must be energized in order to initiate operation of the register. These emitters will illuminate detectors $\frac{1}{2}D_{1+A,1}$ and $\frac{1}{2}D_{2+N,1}$, which then in turn will activate emitter E_1. Since emitter E_1 illuminates detectors $\frac{1}{2}D_{1+A,1}$ and $D_{1,1}$ its lit condition is stable and it will remain lit even after emitters G_A and G_N have been extinguished. To transfer the "on" position from E_1 to E_2 it is necessary only to energize temporarily emitter G_S. This will cause detectors $\frac{1}{2}D_{2+S,2}$ and $D_{S,\bar{1}}$ to be illuminated. The illumination of $\frac{1}{2}D_{2+S,2}$ by G_S and of $D_{1,2}$ by E_1 is sufficient to trigger E_2, while the illumination of $D_{S,\bar{1}}$ causes E_1 to be switched off. It is obvious that such an operation requires a faster build up of photoconductivity in $\frac{1}{2}D_{2+S,2}$ and $\frac{1}{2}D_{1,2}$ than in $D_{S,\bar{1}}$. This can be

achieved by judicious choice of a series resistor and of the transradiance from G_S to $D_{S,\bar{1}}$. The transfer of the "on" position is performed similarly by temporarily energizing G_A. (G_X is necessary only for the initial shift setting.) The shift operation consumes a large fraction of a second, determined mainly by the long response times of the CdS cells used.

Numerous other optronic circuits have been constructed. By permitting light originating from some optrons to reach other optrons in a predetermined fashion, cascades, delay lines, and screen-matrices capable of integration have been built and investigated.

Discussion

The opto-electronic characteristics of electroluminescent and photoconductive transducers make them suitable for the construction of devices and functional networks capable of light amplification, light switching, and light storage. The special case of ac homochromatic light amplification, which has been treated in detail, was found to provide data useful for the analysis of the operation of optrons, that is, bistable regenerative homochromatic light amplifiers. The triggering of optrons into either of these two stable operating conditions can be accomplished by electrical or optical or combined means. Logic networks, such as optical-switch/binary-storage cells, shift registers, etc., have been built using series and parallel combinations of electroluminescent and photoconductive cells and resistors. An attractive feature of these networks is the intrinsic potentiality of complete electrical isolation of the various parts which are coupled exclusively by optical means. The objectionally long switching intervals, caused by the slow decay of available photoconductors, might sometimes be considered tolerable for optron devices in view of their minute size, small weight, low power consumption, and optical accessibility.

Acknowledgment

The writer would like to acknowledge the help of Dr. R. Rulon in supplying specially constructed electroluminescent cells, the assistance of J. Hayes in the construction of light amplifying and bi-stable devices and circuits, and that of Miss B. A. Francis in some of the experimental work. Various stages of this investigation were carried out under direction of Drs. E.G. Schneider, S. C. Weiler, and L. S. Sheingold. The author wishes to thank them for their interest and encouragement.

Bibliography

[1] Henderson, S. T., "Electroluminescence." *Research*, Vol. 8 (June, 1955), pp. 219–225 (review article).
[2] Kröger, F. A., Vink, H. J., and vanden Boomgaard, J., "Controlled Conductivity in CdS Single Crystals." *Zeitschrift für physikalische Chemie*, Vol. 203 (1954), pp. 1–72; "Temperature Dependence of the Hall Effect and the Resistivity of CdS Single Crystals." *Philips Research Report*, Vol. 10 (1955), pp. 39–76.
[3] Smith, R. W., and Rose, A., "Space-Charge-Limited Currents in Single Crystals of Cadmium Sulfide." *Physical Review*, Vol. 97 (March 15, 1955), pp. 1531–1537.
[4] Lambe, J., "Recombination Processes in Cadmium Sulfide." *Physical Review*, Vol. 98 (May 15, 1955), pp. 985–992.
[5] Kallmann, H., Kramer, B., and Perlmutter, A., "Infra-red Stimulation and Quenching of Photoconductivity in Luminescent Powders." *Physical Review*, Vol. 49 (July 15, 1955), pp. 391–400.
[6] Orthuber, R. K., and Ullery, L. R., "A Solid State Image Intensifier." *Journal of the Optical Society of America*, Vol. 42 (1952), p. 850.
[7] Loebner, E. E., "Luminous and Radiant Power Gains of Light Amplifying Devices," submitted for publication in the *Journal of the Optical Society of America*.
[8] Cusano, D. A., "Field Enhanced Solid-State Luminescence." *Physical Review*, Vol. 98 (May 15, 1955), p. 1169.
[9] Destriau, G., "The New Phenomenon of Electroluminescence." *Philosophical Magazine*, Vol. 38 (1946), pp. 700–773.
[10] Piper, W. W., and Williams, F. E., "The Mechanism of Electroluminescence in Zinc Sulfide." *British Journal of Applied Physics*, Supplement No. 4 (1954), pp. 39–44.
[11] Lehovec, K., "New Photoelectric Devices Utilizing Carrier Injection." Proceedings of the IRE, Vol. 40 (November, 1952), pp. 1407–1409.
[12] Curie, D., "Sur le mécanisme de l'electroluminescence." *Journal de Physique et le Radium*, Vol. 14, Part II (February, 1953), pp. 135–136.
[13] Zalm, P., Diemer, G., and Klasens, H. A., "Some Aspects of the Voltage and Frequency Dependence of Electroluminescent Zinc Sulfide." *Philips Research Report*, Vol. 10 (June, 1955), pp. 205–215.
[14] Roberts, S., "Dielectric Changes of Electroluminescent Phosphor during Illumination." *Physical Review*, Vol. 90 (April 15, 1953), p. 364.
[15] Waymouth, J. F., and Bitter, F., "Experiments in Electroluminescence." *Physical Review*, Vol. 95 (August 15, 1954), pp. 941–949.
[16] Loebner, E. E., and Freund, H., "Stacked Barriers in an Electroluminescent Zinc Sulfide Crystal." *Physical Review*, Vol. 98 (1955), p. 1545.
[17] Loebner, E. E., "Time-Dependent Spectra of Electroluminescent Zinc Sulfide." *Physical Review*, Vol. 92 (November 1, 1953), p. 846.
[18] Committee on Colorimetry of the Optical Society of America, "The Science of Color." Corowell, New York; 1953.
[19] Kazan, B., and Nicoll, F. H., "An Electroluminescent Night-Amplifying Picture Panel." First Annual Technical Meeting on Electron Devices, PGED IRE, Washington, D. C., October, 1955.
[20] Rulon, R., "Advancements in the Panelescent Lamp." *Sylvania Technologist*, Vol. 8 (April, 1955), pp. 45–47.
[21] Marshall, B. O., Jr., Bowman, J. R., and Schwertz, F. A., "Optical Elements for Computers." Quarterly Report No. 6, Computer Components Fellowship No. 347, Mellon Institute of Industrial Research, University of Pittsburgh; June, 1952.
[22] "Final Report of the Computer Components Fellowship No. 347." Mellon Institute of Industrial Research, University of Pittsburgh; December, 1953.
[23] Bramley, A., and Rosenthal, J. E., "Transient Voltage Indicator and Information Display Panel." *Review of Scientific Instruments*, Vol. 24 (June, 1953), pp. 471–472.

Cathodoluminescence*

G. F. J. GARLICK†

Summary—A review of recent advances in fundamental knowledge of the mechanism of luminescence excitation by electrons is given. The development of the scintillation counter and studies of electron scattering together with progress in the study of photoluminescence mechanisms in phosphors have provided a new basis for establishing a sound theory of cathodoluminescence.

It is shown that the brightness-voltage relations for conventional screens result from the complex microcrystalline form of the phosphor. Single crystal studies with electron beam and single particle excitation reveal a more simple behavior. Ultimate screen efficiency is shown to be strongly dependent on loss of scattered primary and secondary electrons from the phosphor surface and on nonradiative electron-hole recombination in electron traps and other nonluminescent centers. Current saturation and electron burn are also discussed

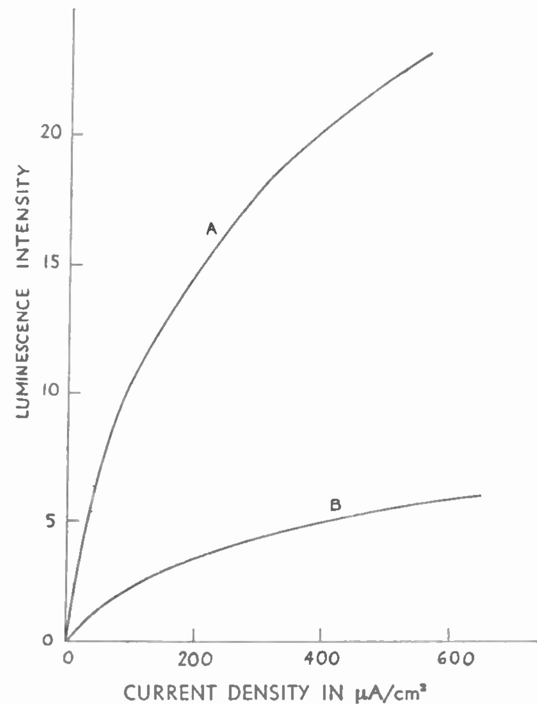
Introduction

UP TO THE year 1949 when the last American review of cathodoluminescence by the author was published[1] most of the studies of electron-excited luminescence were made on conventional cathode ray tube screens. For these screens, mostly of sulphide and silicate phosphors, empirical relations were established between the luminescence yield and voltage and current density of the exciting electron beam. In addition, many studies were made with raster excitation, thus adding a further complexity to the conditions of excitation. Very few measurements were made at other than room temperature. In 1947, publications by Strange and Henderson[2] gave results for the simplest conditions of excitation possible for the powder layer screen but added no basically new information to the empirical laws already found. From such studies it was found that the luminescence intensity (L) was related to voltage (V) and current density (i) of the cathode ray beam as follows:

$$L = f(i)[V - V_0]^n$$

where $f(i)$ was a linear function at low current densities but became nonlinear at higher densities because of saturation effects. The power of $(V - V_0)$, namely, $-n$, lay between 1 and 3 depending on the particular phosphor and on the specific conditions of preparation of the screen. The "dead voltage" V_0 appeared to be a function of screen contamination and was very small for carefully-prepared screens. Experimental results showing the above relation are given in Figs. 1(a) and 1(b).

For some manganese-activated zinc silicate phosphors



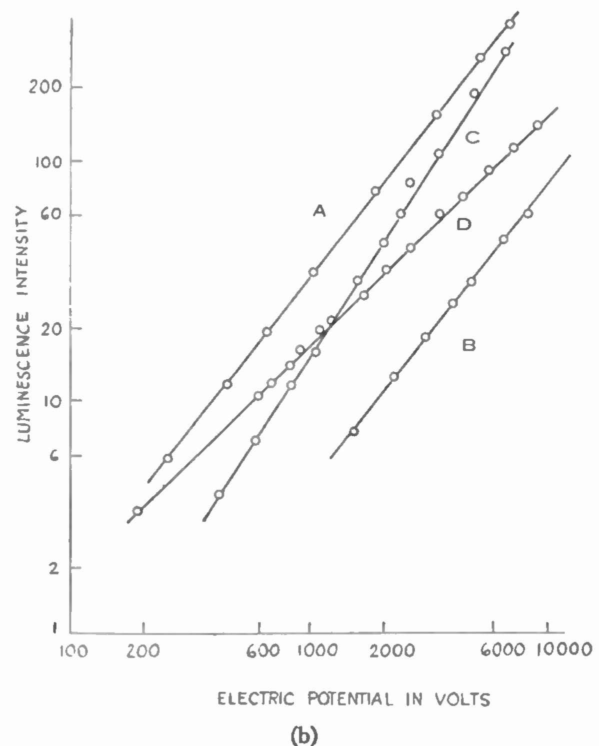


Fig. 1(a)—Variation of luminescence with beam current density for a Zn_2SiO_4-Mn phosphor. A) Tube potential 10 kv. B) Tube potential 4 kv. (After Strange and Henderson.) (b) Variation of luminescence with tube potential for various phosphors. A) ZnS-Ag (10μa/cm²). B) ZnS-Ag (550μa/cm²). C) Zn_2SiO_4-Mn (5μa/cm²). D) ZnS-Mn (10μa/cm²). (After Strange and Henderson.)

* Original manuscript received by the IRE, August 5, 1955.

† Department of Physics, University of Birmingham, Edgbaston, Birmingham, England.

[1] G. F. J. Garlick, "Cathodoluminescence," *Advances in Electronics*, vol. 2, p. 152, 1950; review.

[2] J. W. Strange and S. T. Henderson, "Cathodoluminescence," *Proc. Phys. Soc.* (London), vol. 58, p. 368, July, 1946.

n was equal to 2 which prompted Fano[3] to develop a diffusion theory to explain this particular power law. However, it is quite clear from later studies reviewed below that the value of n has no fundamental significance and generally alters with voltage, tending to unity as V increases. The author in a previous review[4] regarded n values greater than unity as an indication of a variation of the ratio of radiative to nonradiative transitions in the excited phosphor with changing depth of penetration of the exciting electrons.

Closely related to the process of cathodoluminescence are, of course, the processes of primary electron penetration into, and secondary electron emission from, the surface of solids. In the above studies it was tacitly assumed that the primary electrons had a range x in solids given by the Thomson-Whiddington law:

$$x = k(V_0 - V)^2$$

where V_0 is the incident beam potential and V that after penetration x which was obtained for the passage of electrons through foils. The validity of this law for electrons in the cathodoluminescence process is now seriously open to question as discussed below. Secondary electron emission constitutes a loss of primary beam energy to the luminescence emission process. However, it was not realized in earlier studies how serious this loss might be. Recent studies of electron scattering have modified the picture of secondary emission in relation to luminescence in the phosphor.

Perhaps one of the greatest advances has been made possible by the development of the scintillation counter (photomultiplier and single crystal phosphor as particle or radiation detector) for nuclear physics research. Now it is possible to study the effects of a single electron or other particle on a luminescent solid, and thus to obtain a clearer life history of the particle as it penetrates into the phosphor.

In this review the several modern studies in the fields mentioned are brought together and their importance in formulating a more comprehensive theory of cathodoluminescence is discussed. We may consider the process of cathodoluminescence as dependent on the following main stages:

1) The penetration of primary electrons into the phosphor and their elastic scattering resulting in the loss of some of them by reemission from the phosphor surface.

2) Inelastic scattering of primary electrons resulting in ionization and production of secondary electrons. The cascade collisions of the latter then result in a plasma of secondary electrons which lose energy until they finally reach thermal velocities. Some of these secondary electrons are lost by emission from the phosphor surface.

3) The secondary electrons return to their original energy states in the phosphor crystal, giving up their excess energy in a radiative transition (luminescence) or in a nonradiative transition (thermal degradation of energy). The ratio of these two kinds of transition obviously affects the efficiency of light production as also do the earlier processes where primary or secondary electrons are lost by their reemission from the phosphor. We now consider these stages in detail.

Penetration and Scattering of Primary Electrons

Two important recent sources of information on the penetration of electrons into solids are available. One is the study of electron paths in nuclear emulsions; the other, some very beautiful experiments of Ehrenberg and Franks[5] on the penetration of a 0.7μ wide "pencil" of electrons (up to 40k ev) into single crystal phosphors. Both fields of study show that very strong scattering of primary electrons occurs as soon as they enter a dense medium and they are then no longer able to be identified. Hence it is difficult to speak of the range of an electron in a solid. For very fast electrons in an emulsion it is possible to follow the track of the electron if there are no branches, but in the case of phosphors, the envelope of excitation resulting from the primary beam impact is measured; i.e., the envelope of secondary electron plasma. Fig. 2 shows schematically the results in each case. As shown by the photographs obtained by Ehrenberg and Franks, the excitation volume forms a capped sphere with its diameter many times the size of the original cross-sectional dimension of the primary beam.

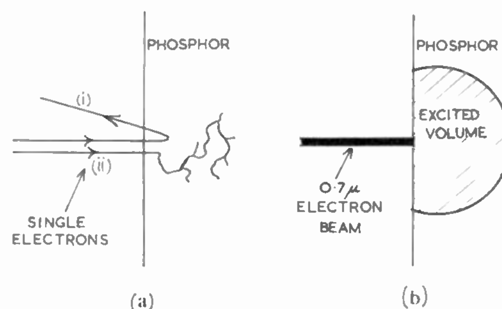


Fig. 2(a)—Schematic diagram of trajectories of single primary electrons incident on a phosphor crystal. i) Primary scattered out of phosphor. ii) Primary absorbed with strong scattering and production of secondary tracks. (b) Excitation envelope in phosphor excited by narrow pencil of electrons (10–40k ev). (After Ehrenberg and Franks.)

The sphere boundary is fairly sharp but this is probably emphasized by the logarithmic response of the film used to take the photographs. Range measurements deduced by Ehrenberg and Franks refer to the range for excitation and not for the primary beam. From absorption of electrons in foils very diffuse results are obtained for range. In the case of emulsions, the path length of the primary is obviously very different from its penetration depth into the emulsion. In a recent paper[6] the author

[3] U. Fano, "Theory of Cathodoluminescence," *Phys. Rev.*, vol. 58, p. 544; September, 1940.

[4] Garlick, *loc. cit.*

[5] W. Ehrenberg and J. Franks, "The penetration of electrons into luminescent material," *Proc. Phys. Soc.* (London) B, vol. 66, p. 1057, December, 1953.

[6] G. F. J. Garlick, "Excitation of phosphors by electrons," *Brit. Jour. Appl. Phys.*, suppl. 4, S 103; January, 1955.

has brought together the results from absorption studies for electrons in foils (Al)[7] and the results of Ehrenberg and Franks. Although the latter attempt to prevent the exclusion of the Thomson-Whiddington law they do not succeed and in general the range law is found to be of the form

$$n = C \cdot E^y$$

where y is a logarithmic function of E. A recent theoretical determination of the specific energy loss, i.e., the loss per unit path length, dE/dx for a primary electron penetrating into a solid, has been made by Dekker and van der Ziel[8] and gives

$$-dE/dx \propto \log E/E.$$

The Thomson-Whiddington law gives simply the inverse function of E without the logarithmic factor. The departure from the latter law will thus be most marked at low primary energies such as those obtained in the cathode ray tube where scattering of the primary electrons is very strong. An important consequence of the scattering is, as stated above, the loss of primary and secondary electrons from the phosphor by re-emission Each particle leaving represents energy lost to the luminescence process and so the problem of secondary emission must be considered when attempting to determine the limiting efficiency possible for a cathode ray tube screen.

Secondary Electron Emission

Excellent accounts of secondary emission characteristics for solids have been given recently by Bruining[9] and by McKay.[10] However, the problem of emission of scattered primary electrons in relation to luminescence efficiency has never been seriously considered until recently.[11] Very early experiments by Schonland[12] with electrons showed that strong back scattering of primary electrons occurs from solids and that the effect increases with atomic number. Fig. 3 shows a curve taken from a previous paper by the author and reconstructed from results of Palluel[13] for 20k ev electrons. It is seen that a fraction approaching 50 per cent of the primary electrons may be back-scattered. Each of these electrons carries away from the phosphor almost all its primary energy. Add to this the amount removed by lower energy secondary electrons and it is obvious that secondary emission is responsible for one limitation on the efficiency of phosphors in cathode ray tubes. From a theoretical viewpoint, Blanchard and Fano[14] show that scattering depends much more on the fractional energy loss of the primary electrons $(E_0-E)/E_0$ (E_0 is initial energy and E after some penetration) than on the absolute value of E_0, as borne out approximately in experiments over a very wide range of energies. Thus, at all cathode ray tube voltages, secondary emission and primary electron scattering and emission can constitute a serious loss of energy which seems at present to be unavoidable.

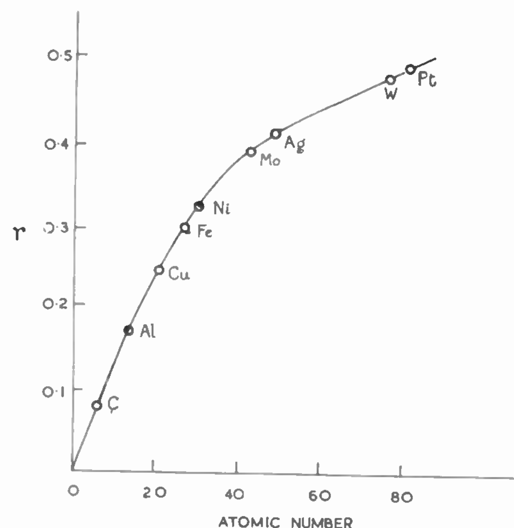


Fig. 3—Variation of ratio r of back scattered to incident primary electrons of 20k ev energy on targets of different atomic numbers. (Constructed from data by Palluel.)

Production of Luminescence

To understand the later stages of the cathodoluminescence process when the primary excitation has produced the plasma of secondary electrons of thermal velocities, we use the energy band model for a crystal phosphor containing impurity or other centers in which luminescence transitions can occur. The schematic energy diagram for this model is shown in Fig 4. Also

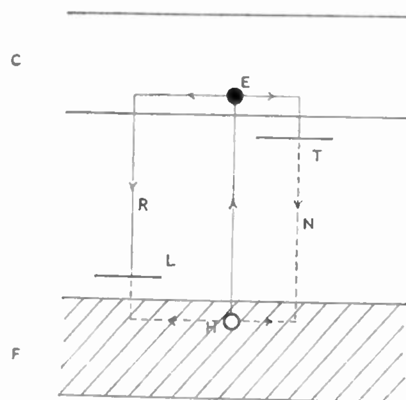


Fig. 4 Energy band scheme for crystal phosphor. C=Conduction band. F=Filled band. L=Luminescence centers. T=Electron traps. E=Electron. H=Positive hole. R=Radiative transition. N=Nonradiative transition.

included are centers in the crystal where nonradiative electron transitions can occur. After the primary processes above have occurred, the phosphor finally reaches a state where thermal velocity electrons in the conduc-

[7] L. Katz and A. S. Penfold, "Range-energy relations for electrons etc., *Rev. Mod. Phys.*, vol. 24, p. 28; January, 1952.

[8] A. J. Dekker and A. van der Ziel, "Theory of the production of secondary electrons in solids," *Phys. Rev.*, vol. 86, p. 755; June, 1952.

[9] H. Bruining, "Physics and Applications of Secondary Electron Emission," Pergamon Press, London, Eng.; 1954.

[10] K. G. McKay, "Secondary electron emission," *Advances in Electronics*, vol. 1, p. 65, 1949; review.

[11] Garlick, *Brit. Jour. Appl. Physics*, *loc. cit.*

[12] *Ibid.*

[13] *Ibid.*

[14] C. H. Blanchard and U. Fano, "A formula for multiple scattered electrons," *Phys. Rev.*, vol. 82, p. 767; June 1951.

tion levels C are recombining either with empty emission centers L or undergoing nonradiative transitions in the centers T. The relative number of the centers L to the centers T can be estimated from the efficiency of excitation of luminescence by ultra violet light of an optimum wavelength. Studies by Bril and Klasens[15] have shown that the maximum efficiency for a copper-activated zinc sulphide phosphor is about 50 per cent, although other workers have claimed higher values. For a value of 50 per cent it is concluded that the nonradiative processes outnumber radiative ones by about two to one. If we now allow for the loss of energy by primary and secondary electron emission from the phosphor, the optimum efficiency may be reduced to about 25 per cent, a value substantiated by experiments of various workers.[16,17] For those who have a practical interest in phosphor efficiencies the above figure of 25 per cent applies to screens of zinc sulphide or zinc cadmium sulphide with aluminum backing and operated in the region of 20 kv. In practical form it is equivalent to a light production efficiency of up to 100 lumens per watt. It seems from the above fundamental considerations to be doubtful whether this efficiency will ever be improved by more than a factor of two.

Dependence of Phosphor Efficiency on Energy and Intensity of Electrons

Experiments using scintillation counters with single crystal phosphors show that for single electrons the luminescence emission is proportional to the electron energy (if the electron is completely absorbed in the phosphor) over a very wide range. Only when the specific energy loss dE/dx (or density of ionization) is high does departure from proportional behavior occur. In the case of heavy particles such as alpha particles such departure is thus more noticeable.[18] Thus, for phosphor screens, the power law relations between cathode ray beam energy and luminescence must be regarded as due to the complex geometrical form of the screen. In some recent cases of carefully-prepared screens, linearity between luminescence and voltage is obtained over the wide range of 1 to 50 kv.[19] For uniform evaporated phosphor layers a similar behavior is reported.[20] The independence of the efficiency of electron energy must also arise from the fact that electron emission by back scattering (and hence energy loss by the process) is approximately independent of electron energy as shown above.

An outstanding problem in the fundamental interpretation of cathodoluminescence is that of current saturation in the luminescence output. In this respect the effect appears to be related to the type of phosphor since silicates show less saturation than sulphides.

A recent paper by Bril[21] attempts to provide an explanation of current saturation under steady and under raster excitation of the phosphor. Behavior appears to depend on the occupancy density of the available luminescence centers. In zinc silicates the concentration of centers is much higher than in the zinc sulphides, but the decay time of emission is much longer (≈ 10 msec) than in the sulphides (≈ 10 μsecs). Sulphides show marked saturation compared with silicates under raster excitation but show much less for steady excitation. For silicates the effect of excitation conditions is less marked.

The longer lifetime in the silicates limits the number of available centers for steady excitation, but for television raster conditions the pulsed excitation has a period comparable with the decay time and so the centers become available for excitation in the "dark" interval. In the sulphides the problem appears to be more complex since the nonradiative centers play a decisive part in determining efficiency of emission under any specified excitation conditions. At present, a series of experiments is needed, on current saturation in single crystals and uniformly-evaporated layers, and over a wide range of temperature and excitation conditions.

Phosphor Deterioration under Electron Bombardment

Before the introduction of aluminum-backed screens the usefulness of the phosphor screen was limited by the effects of prolonged bombardment on the "sticking" potential which decreased with time until the brightness fell off due to inadequate secondary emission.[22] It is interesting to note that Ehrenberg and Franks[23] found no evidence of sticking potentials for uncoated single crystals of inorganic type for potentials up to 40 kv. Thus the sticking potential would appear to be a function of the complex form of the conventional phosphor screen and, of course, has been overcome in modern tubes by the aluminum backing of the screen. Ion burn has been minimized by the use of ion traps and so the discussion of screen deterioration is reduced to a consideration of the effects of electrons on the long-term efficiency of the luminescence processes in the screen. In this respect few fundamental experimental studies are available.

In general electrons bombarding a phosphor in vacuums behave as reducing agents. In some phosphors a visible darkening of the phosphor base color occurs and

[15] A. Bril and H. A. Klasens, "Intrinsic efficiencies of phosphors under cathode ray excitation," *Philips Res. Rep.*, vol. 7, p. 401; December, 1952.
[16] C. G. A. Hill, "Applied cathode luminescence," *Brit. Jour. Appl. Phys.*, Suppl. 4, S6; January, 1955.
[17] Bril and Klasens, *loc. cit.*
[18] G. F. J. Garlick, "Luminescent materials for scintillation counters," *Prog. Nucl. Phys.*, part II, p. 51, 1952; review.
[19] Hill, *loc. cit.*
[20] L. R. Koller and E. D. Alden, "Electron penetration and scattering in phosphors," *Phys. Rev.*, vol. 83, p. 684; August, 1951.
[21] A. Bril, "On the saturation of fluorescence with cathode ray excitation," *Physica*, vol. 15, p. 361; May, 1949.
[22] Garlick, *Advances in Electronics*, loc. cit.
[23] *Ibid.*

some experiments with fluorides by Hill[24] show that a sensitization of ultra-violet excited luminescence occurs on prolonged electron excitation due to the formation of ultra-violet absorbing centers in the fluoride. This is analogous to the production of F centers in the alkali halides by particle bombardment.[25] Such centers may function as sites for nonradiative transitions of conduction electrons to their ground states, thus decreasing the luminescence efficiency. Again, experiments on single crystals would furnish much useful information on "electron burn." A practical point arises here. F center production by electrons can be inhibited in some crystals by introduction of specific impurities.[26] It might be worthwhile to make experiments of this type to determine the effect of additions of this kind on electron burn.

Conclusion

The above discussions are inevitably one-sided since they stress the fundamental aspects of cathodoluminescence. However, it will be realized that detailed processes such as electron scattering and nonradiative transitions affect the practical performance and indicate the limits to future advances in cathode ray tube screen efficiency and maintenance.

[24] *Ibid.*
[25] F. Seitz, "Colour centres in alkali halides," *Rev. Mod. Phys.*, vol. 26, p. 7; January, 1954.
[26] J. H. Schulman, R. J. Ginther, and C. C. Klick, "Enhancement of X-ray induced absorption bands in alkaline earth compounds," *Jour. Chem. Phys.*, vol. 20, p. 1966; December, 1952.

Electroluminescence and Related Topics*

G. DESTRIAU† AND H. F. IVEY‡, SENIOR MEMBER, IRE

Summary—**Light may be generated by the direct action of an electric field on various solid materials by a number of different mechanisms (electroluminescence). Electric fields may also influence luminescence phenomena excited by various types of radiation (electrophotoluminescence). The varied phenomena which have been observed, and their dependence on the operating parameters, are reviewed and theoretical explanations are given where possible. Possible practical applications of electroluminescence and electrophotoluminescence are also discussed.**

General Introduction

THE GENERAL designation of electroluminescence is now used to cover various effects which can occur when certain phosphors are subjected to an electric field. In some experiments electrodes are in contact with a phosphor layer or with a single crystal while in other experiments there are no electrode contacts; the crystallites, in the latter case, are embedded in an insulator.

Galvanoluminescence, *i.e.*, light emission arising from electrolytic action, has been known for a long time [1, 2], but will not be discussed here. The emission of light by single crystals, with electrodes in contact, was first reported on silicon carbide by Lossew [3] in 1923. Some disturbances of the after-glow of a damp phosphor powder squeezed between two electrodes were reported by Dechene [4] and a similar weakening of the after-glow, under the action of silent discharge, was observed by Coustal [5], but these last phenomena must be ascribed to the drop of potential which appears near one of the electrodes in the damp powder so that crystallites in this region are subjected to very high fields.

The action of electric fields upon microcrystals embedded in an insulator was first reported in 1920 by Gudden and Pohl [6], who pointed out a momentary enhancement of the after-glow from a phosphor previously irradiated by ultra-violet radiation. The sustained emission of light by a phosphor powder embedded in an insulator was first reported by Destriau [7] in 1936. Since that time many other phenomena have been discovered; these will be discussed in detail below.

We shall consider in the following sections the points listed below.

I. *Intrinsic electroluminescence* (Destriau effect), covering chiefly light emission by suitable phosphor powders embedded in an insulator and subjected only to the action of an alternating electric field.

II. *Carrier-injection electroluminescence*, covering light emitted by the sole action of an electric field on crystals or particles in contact with conducting electrodes so that current injection can occur.

III. *Electrophotoluminescence*, covering all phenomena which involve an electric field and a previous or simultaneous photo-excitation, that is to say, the following

* Original manuscript received by the IRE, August 5, 1955; revised manuscript received, September 14, 1955.
† Laboratoire de Luminescence, Faculté des Sciences, Paris, France. Consultant, Lamp Division, Westinghouse Electric Corporation, Bloomfield, N. J.
‡ Research Department, Lamp Division, Westinghouse Electric Corporation, Bloomfield, N. J.

effects: (1) The flash of momentary illumination in the after-glow region (Gudden and Pohl effect [6]), (2) The quenching [8] or enhancement [9] of photoluminescence, (3) The enhancment of infra-red quenching [10] or of infra-red stimulation [11], and

IV. *Applications* of electroluminescence and electrophotoluminescence. The principal unsolved problems or trends in the field will be briefly indicated.

An earlier review paper, covering only intrinsic electroluminescence and electrophotoluminescence, was published in 1947 [12]. Most papers published before this year are, therefore, omitted in the present bibliography.

I. Intrinsic Electroluminescence

A. Introduction

Before entering more deeply into the subject, some information must be given on the devices used as EL cells[1] (or EL condensers), as well as on the general behavior of the phenomenon itself.

The EL condenser is usually formed of a thin layer of solid (plastic or ceramic) or liquid insulator in which the EL phosphor powder is embedded. This mixture is placed between two electrodes, at least one of which is transparent, for example, conducting glass (glass whose surface has been treated, near its melting point, with an aqueous solution of tin or antimony chloride). The second electrode can be another conducting glass, but is usually a metal layer deposited by thermal vaporization. In order to deliver power to such a cell, it must be operated on alternating or pulsed voltages. In special cases the dielectric of the cell may be vacuum or even air; such cells may also be operated on dc, but properly speaking this constitutes carrier-injection EL rather than intrinsic EL.

In order to obtain very high fields in the phosphor with moderate applied voltages, the thickness of the EL cell must be very small. On the other hand, for uniformity of illumination, the thickness should be large compared to the particle size of the phosphor employed. In practice thicknesses of 25 to 100 microns (1 to 4 mils) are employed; the capacitance of the cell is then about 100 $\mu\mu$fd/cm^2 and the operating voltage may be between 100 and 600 volts.

With such a capacitance, a drop of potential and a shift in phase may occur along the conducting glass, particularly at high frequencies, if the resistance of the conducting coating is too high. For example, if the coating resistance were 10^4 ohms/square and a circular cell of 7.3-cm radius were employed (with contact only to the edge of the circular piece of conducting glass) then for $f = 5,000$ cps the potential at the center of the cell would be only 95 per cent of that at the periphery; furthermore, there would be a phase difference of $\pi/7$ between center and periphery. The same results would be obtained for $f = 50$ cps for a cell of 73 cm radius. Since, as will be shown below, the light output is very sensitive to the magnitude of the field, this would be an undesirable situation. Such a phase shift would also have to be taken into account in the study of the phase difference between the light emission and the applied voltage. However, since conducting coatings are available with resistances of 100 or 200 ohms/square (and optical transmittance of 80 per cent or more), these effects can generally be avoided.

For a uniform suspension of homogeneous phosphor spheres in a dielectric, the field acting in the crystals is given by the Maxwell-Wagner formula [13] reported again more recently by Roberts [14],

$$E_1 = E_0\left(\frac{3K_2}{2K_2 + K_1 - V(K_1 - K_2)}\right),$$

where E_1 is the field in the crystals, E_0 is the mean field applied to the medium, K_1 and K_2 are the dielectric constants of the crystals and the dielectric, respectively, and V is the ratio of the volume of the crystals to the total volume. This formula shows that a high field strength in the crystals requires an insulator of high dielectric constant.

If the phosphor is suspended in a liquid dielectric, such as castor oil, the particles will move under the action of the electric field, as is well known [15], and form more or less closed bridges extending from one electrode to the other. Under these conditions the equation given above no longer applies. This behavior of liquid-embedded materials must be taken into account in interpreting the results of measurements on such EL cells [16]. Furthermore, although such cells are very convenient for laboratory experiments, a thin film of solid insulator must be introduced into the cell if the possibility of carrier injection at the electrode contacts is to be definitely excluded.

When an EL cell is subjected to a direct potential, only a flash of light occurs when the potential is applied and another flash when it is removed. Waymouth and Bitter [17], indeed, report that under proper conditions emission is observed only at removal of the potential. In order to obtain sustained emission, the effects of polarization must be avoided. Usually sinusoidal potentials are used but other potential forms, such as square waves, and even more complicated waves have been tested [17, 18a] (see Fig. 1). Sustained emission can also be obtained with a direct potential if the crystals are embedded in castor oil and put in a thin glass tube which is rotated about its axis between two plane fixed electrodes so that the field direction in the crystals is continuously changing [12]. It has been shown recently [19], even in alternating fields, moreover, that the light emission of some phosphors is increased when the crystals are rotated.

The observation of electroluminescent crystallites with a microscope shows that the light emission comes chiefly from some scattered spots especially located near the points of individual elongated crystals or near crystal junctions [17, 20]. In the latter case, the strongest illumination occurs when the field direction is per-

[1] For brevity, the abbreviation EL will be used for "electroluminescence" or "electroluminescent."

pendicular to the junction surface. In the former case, the strongest illumination occurs when the field direction is parallel to the needle axis.

Electroluminescence is a primary effect; that is to say, electroluminescence is an emission of light directly excited by the field and not by a secondary process involving, for example, ultra-violet radiation or silent discharges in small cavities in the embedding material.

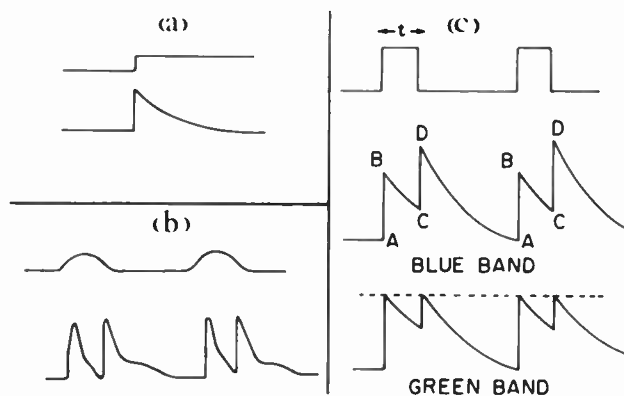


Fig. 1—Light emitted by EL phosphors (upper curves) when subjected to fields of different waveform (lower curves). (a) A direct potential step. A flash is excited at the application of the field and the brightness falls quickly to zero. (b) Isolated voltage pulses. Two flashes occur, one at the beginning and another at the end of each pulse. (c) Square-waves of potential (from Nudelman and Matossi [17]). Phosphor: ZnS:Cu, Pb. For $t>0.7$ millisec, $AB=CD$ for the blue band, while for the green band the two peak heights are equal.

Herwelly [21] has doubted this conclusion but many experiments have been recently made by other workers [22–25] in order to prove that EL is a primary effect; their results can be summarized as follows:

1. When a layer of air is left between one of the electrodes and the phosphor, the light emission is weakened (because, according to the formula given above, the field strength in the crystal is lowered) and nitrogen lines appear in the spectrum. If the air is replaced by oil, the light emission is greater than in the first case but nitrogen lines do not appear in the spectrum.

2. Emission due to ultra-violet radiation produced in air bubbles and light emission by electroluminescence are not in phase.

3. With the same phosphor the after-glow following photoexcitation is noticeable, while it is nearly zero in intrinsic electroluminescence.

4. A crystal placed outside the field, but still near another excited crystal, does not luminesce.

5. The color of photoluminescence and of EL is different for some phosphors.

6. A good photoluminescent material is not necessarily a good EL phosphor.

7. The voltage dependence of EL and of luminescence excited by gas discharges are different.

8. Although EL emission occurs only at localized points, the emission excited by ultra-violet light is uniform throughout the crystals.

B. Electroluminescent Phosphors

The general statement may be made that there are many more phosphors excitable by cathode-rays or by X-rays than by ultra-violet radiation. Further, the number of EL materials is much smaller than the number of photoluminescent materials. On the other hand, there seems to be no well-authenticated instance of a material which exhibits intrinsic EL which does not also exhibit photoluminescence (at least at low temperature), although the optimum activator concentration may be different in the two cases. EL phosphors can be found in many host lattices, such as Zn_2SiO_4 [26–28], ZnS, CdS, ZnSe, and ZnO, but in all cases the surface treatment is of very great importance.

Zalm, Diemer, and Klasens [20, 20a] have reported several possibilities of making EL some of the usual photoluminescent ZnS phosphors available commercially; either by a surface addition of copper sulfide in the amount of 10^{-5} to 10^{-4} atom of copper per atom of Zn in the host lattice, or by a direct surface addition of copper by evaporation in vacuum. Similarly, Waymouth and Bitter [17] and Burns [29] have suggested the formation of spots with a high conductivity.

In the same way, Gumlich, as well as Gobrecht, Hahn and Gumlich [25], have reported that a slight surface oxidation was able to make a normal phosphor El. And, as a matter of fact, an improvement of EL ZnS phosphors by reheating, the effect of which is to oxidize the crystal surfaces, was reported earlier [12]. Homer, Rulon and Butler [30], on the other hand, although they add zinc oxide to the raw materials, remove any free oxide present in the finished material; these workers, however, add lead to their phosphors and suggest that this may result in precipitation of conductive material, perhaps on the surface of the crystals.

A different surface treatment had been already reported by Saddy and Destriau [12], who suggested the formation of a surface layer with a great copper content; too much copper, however, is not desirable and the surplus of Cu may be removed by washing in ammonia. Froelich [31], in a similar way, indicated that only a limited amount of copper was retained in the final material; cyanide washing was used to remove the excess. Froelich [32] has also reported a sensitization of the yellow band of manganese by a small addition of copper.

In order to check the effects of surface conditions upon EL properties, the variation of the copper content of the crystals with depth has been recently studied [33]. Thin layers were successively removed by short acid attacks, the copper content in the solution was determined colorimetrically, and the EL properties of the remaining crystals were measured by the thresholds of visible EL for a dark-adapted eye. Two kinds of curves were obtained according to whether the samples were or were not oxidized on the surface. The copper content was lower in the bulk than upon the surface; however, with products strongly oxidized on the surface, the copper content was maximum in a thin layer

near the surface (Fig. 2). When thin layers were successively removed, the EL sensitivity first increased and then, after reaching a maximum, decreased progressively.

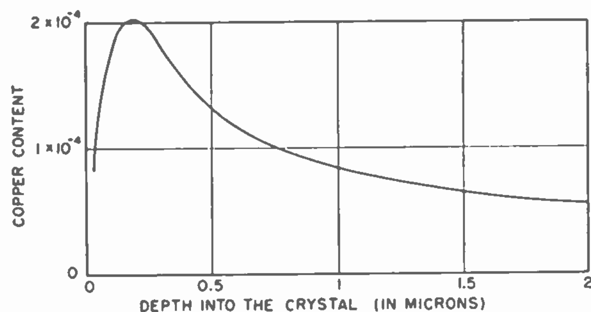


Fig. 2—Variation of the copper content with depth for crystals of oxidized ZnS:Cu of average particle size ten microns.

Still another indication of the effect of the condition of the surface of the phosphor particles on EL was obtained by Thorington [34] in his studies of materials subjected to alternating electric fields in vacuum rather than in air or a dielectric medium. It was found that many materials which were normally not EL became so when placed in a vacuum, and that readmission of air destroyed the EL thus produced. Impregnation of the phosphors in vacuum with various oils and plastic materials either greatly reduced or completely destroyed the vacuum-induced sensitivity, despite the fact that the field strength in the phosphor, for a given applied voltage, is increased upon introduction of a material with dielectric constant greater than unity. The photoluminescence was the same in vacuum, in air, or in oil. From these observations, and from others concerning the effects on EL of heating phosphors in air and in vacuum, Thorington concluded that adsorption of gases on the phosphor surfaces, and the resultant changes in surface conditions, greatly influences the fundamental EL excitation mechanism. Howard [35] has drawn parallels between the observations of Thorington and the now well-known effects of atmospheric surroundings on the potential barriers at the surfaces of semiconductors such as germanium or silicon.

EL materials have been found also among some organic compounds. Such a possibility was first indicated by Payne [36], but described more fully by Bernanose and his co-workers [37]. Films of cellophane or thin paper are immersed in a fluorescent solution of acridine derivatives, such as gonacrin or brilliant acridine orange E, or in carbazole, then dried and embedded in paraffin between two plane electrodes.

C. Dependence of Brightness on Applied Voltage

Only the case of sinusoidal applied voltages will be considered here. In this case the average brightness B increases very rapidly when the voltage V is increased, although in general the rate of increase is lower for higher V. It is generally held that there seems to be no real threshold voltage for EL and that the apparent thresholds depend simply on the sensitivity of the measuring device employed. One of the authors (Destriau) has recently extended the measurements to very low brightnesses, roughly 1/100 that just detectable by the dark-adapted eye, and finds no observable threshold or change in behavior of the usual type of EL cell with powdered phosphor. A definite threshold voltage has been reported, however, for single crystals of ZnS by Piper and Williams [38] and for individual particles of a phosphor powder by Waymouth and Bitter [17]. Visible EL is observed in ordinary EL cells at about 3,000 volts/cm with suitable phosphors, but applied field strengths of 20,000 to 40,000 volts/cm are required to give high EL brightness. Typical data for the voltage dependence of the brightness of intrinsic EL are shown in Fig. 3(a).

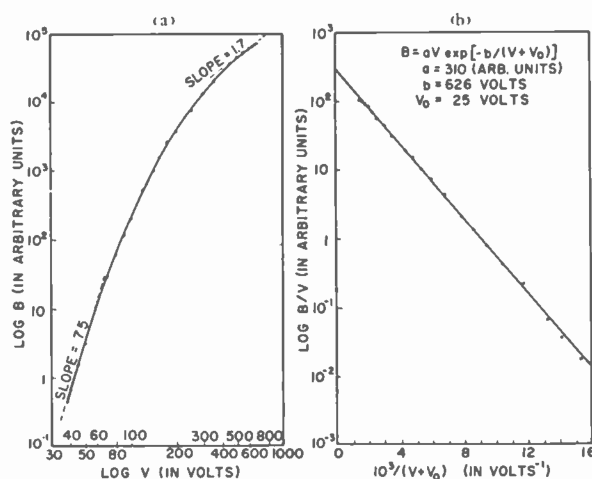


Fig. 3—(a) Typical data for voltage dependence of EL brightness for a green-emitting ZnS:Cu phosphor (data from W. Lehmann). (b) Fit of the empirical equation (6) to the above data.

An early study [7] in 1937 showd that, with a constant frequency of 50 cps, plots of the logarithm of the brightness as a function of $1/V$ were approximately straight lines so that

$$B = a \exp(-b/V), \tag{1}$$

where a and b are constants. This very simple formula was easily explained by rough theoretical considerations. However, in order to take into account the slight curvature of the plots a modification was introduced by writing

$$B = aV^n \exp(-b/V). \tag{2}$$

This equation has three constants, a, b, and n. The values of n were found, depending upon the phosphor, to be between 1 and 3, with 2 a typical value. A theoretical discussion of this equation will be given in a later section. Bernanose [37] reports that data on EL of organic materials also fit this relation.

Recently many other simpler empirical emission laws have been proposed. One [20, 39] gives B as proportional to a power of V,

$$B = aV^n, \tag{3}$$

with various values for the constant n. Another [17] gives B simply as a linear function of V, with a threshold V_0,

$$B = C(V - V_0). \tag{4}$$

It may be noted, however, that differentiation of (1) shows it to have an inflection point at a voltage equal to $b/2$. If (1) is equated to (4) at this point and also the two first derivatives are equated, then the two expressions may be related to each other over a limited range of voltage by setting

$$C = 4ae^{-2/b} \quad \text{and} \quad V_0 = b/4. \tag{5}$$

Another empirical law has been suggested by Howard, Ivey, and Lehmann [40],

$$B = aV \exp[-b/(V + V_0)]. \tag{6}$$

In many cases this equation is in good agreement with experimental results over a wide range of voltages (Fig. 3(b); slight deviations may occur at very high voltages, however. One of the authors (Ivey) earlier showed that a similar expression, but without the V before the exponential term, gives good fit except for very low voltages; in this case deviations obviously must occur because the equation predicts a finite value of B for $V=0$. Lehmann [41] has recently studied EL output over a very wide range of applied voltages. He finds that as V is increased to very high values the output tends to approach a finite value, in contrast to (2), (3), (4), and (6) above; this approach to saturation is best described by Destriau's original expression, (1).

Besides these empirical or experimental laws, many other more complicated forms have been proposed. Many of these equations are based on theoretical considerations and will be discussed in the appropriate section. It should be remarked that if observation is restricted only to one or two logarithmic decades of brightness values, the choice of an equation for empirical purposes is not critical, since any of those which have been proposed [with the exception of (4)] will suffice.

D. Brightness Waves

The light emitted with alternating field excitation is not continuous but flickers with twice the frequency of the applied voltage. The curves showing the light emission as a function of time during a cycle of the field variation are called "brightness waves." Even with sinusoidal fields the brightness wave forms are not regular. A disturbance, which may be either an inflection or sometimes a small secondary peak, often occurs later in time than the main peak; this disturbance can also occur before the main peak. The maximum of the light emission (the main peak) is not in phase with the potential applied and, moreover, of the two main peaks which occur in each cycle, one may be higher than the other.

The brightness wave forms can be characterized by [42] (see Fig. 4):

1. A shift of the brightness wave compared to the potential wave (phase angle ϕ).
2. A disturbance at M and M' when the brightness is either B_2 or B_2', where the prime sign indicates the half-cycle when the transparent electrode, through which the observation is made, is positive.
3. A difference between the brightnesses, B_1 and B_1' reached at the two successive main peaks at A and A'.

Let m designate the "modulation ratio,"

$$m = 2\frac{B_1 - B_1'}{B_1 + B_1'}, \tag{7}$$

(m can be positive or negative) and d designate the "disturbance ratio,"

$$d = B_2/B_1 \quad \text{or} \quad d' = B_2'/B_1'. \tag{8}$$

d_c and d are usually nearly equal so that here we shall consider their average value as the "disturbance ratio."

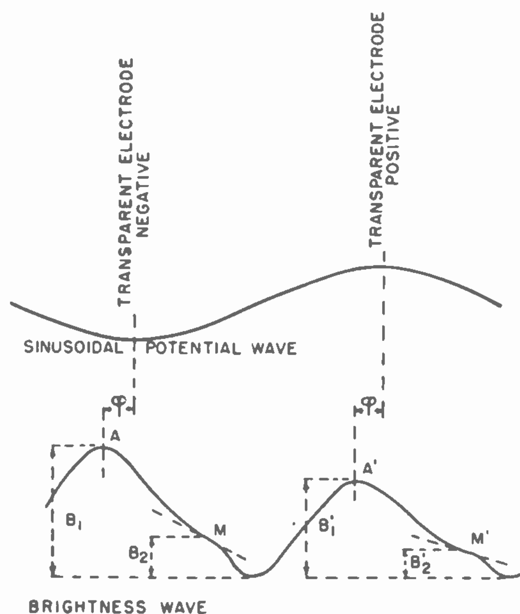


Fig. 4—The significant characteristics of a brightness wave, showing the primary and secondary peaks.

The brightness wave forms, as characterized by the three constants ϕ, m, and d, are dependent on many factors [43]:

1. The ratio a/b between the average thickness of the phosphor layer, a, and the average thickness of the insulator, b.
2. The thickness of the phosphor layer, a.
3. The applied voltage (or better, the field strength).
4. The metal backing or metal electrode.
5. The frequency.
6. The temperature.

The last two factors, frequency and temperature, will be considered in another section; here only the first four factors will be discussed.

The first characteristic of brightness waves is the phase angle ϕ. It was pointed out some time ago [12] that the field acting in the crystals was out of phase with the field in the insulator round the crystals, the first leading the second by an angle ψ such that

$$\tan \psi = 2/K_1\rho f, \tag{9}$$

where K_1 is the dielectric constant of the crystals, ρ is their specific resistance, and f is the frequency. But the field in the insulator also lags the potential applied to the EL cell by an angle γ such that [42]

$$\tan \gamma = \frac{2K_2\rho f}{4\frac{b}{a} + K_1\rho^2 f^2\left(K_2 + K_1\frac{b}{a}\right)}, \tag{10}$$

where K_2 is the dielectric constant of the insulator, a is the thickness of the phosphor layer, and b is the thickness of the insulating layer. Thus $\phi = \psi - \gamma$, and

$$\tan \phi = \frac{\tan \psi}{1 + kx}, \tag{11}$$

with

$$k = K_2/K_1 \quad \text{and} \quad x = a/b. \tag{11a}$$

Thus, $\phi = 0$ only if there is no insulator ($b = 0$, electrodes in contact with the crystal) and it increases towards ψ when a/b is decreased to very small values.

Furthermore, brightness waveforms are a function of the phosphor thickness, a. When a thin phosphor layer is embedded in an insulator between two sheets of mica or between two conducting glasses, the brightness waves usually assume a symmetrical form without any disturbance and with only a small difference between the heights of the two successive peaks. If the thickness of the phosphor layer is made great enough, however, a small disturbance can occur even with a symmetrical cell; this apparently arises from crystal-crystal contacts in the layer. Zalm, Diemer, and Klasens [20] attribute the secondary peaks to the action of the electric field on electrons and holes captured in traps.

Brightness wave forms are also dependent on voltage. One observes, first, an increase of ϕ when the voltage is increased. This variation, which is ascribed to a small increase of conductivity in the crystals under the action of high field strengths, is usually very small. Recently, however, a new phosphor has been found in which this increase is very noticeable (Fig. 5). The modulation ratio, as well as the disturbance ratio, is modified when the applied potential increases. In Fig. 5 one observes that the disturbance vanishes progressively from recording 1 (160 volts) to recording 4 (960 volts). Even with the same phosphor the disturbance can pass progressively from the right to the left side when potential is increased while, simultaneously, the highest peak passes from transparent electrode negative to transparent electrode positive (Fig. 6).

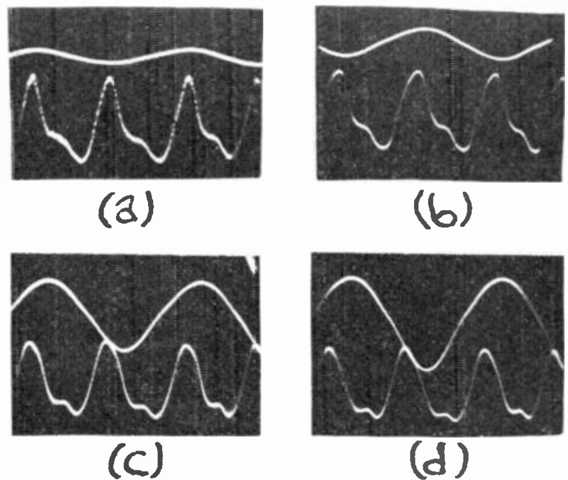


Fig. 5—Progressive shift of the brightness waves when the potential is increased. Phosphor: Lehmann green ZnS:Cu with a palladium electrode. Curve 1, 160 volts; curve 2, 320 volts; curve 3, 720 volts; curve 4, 960 volts.

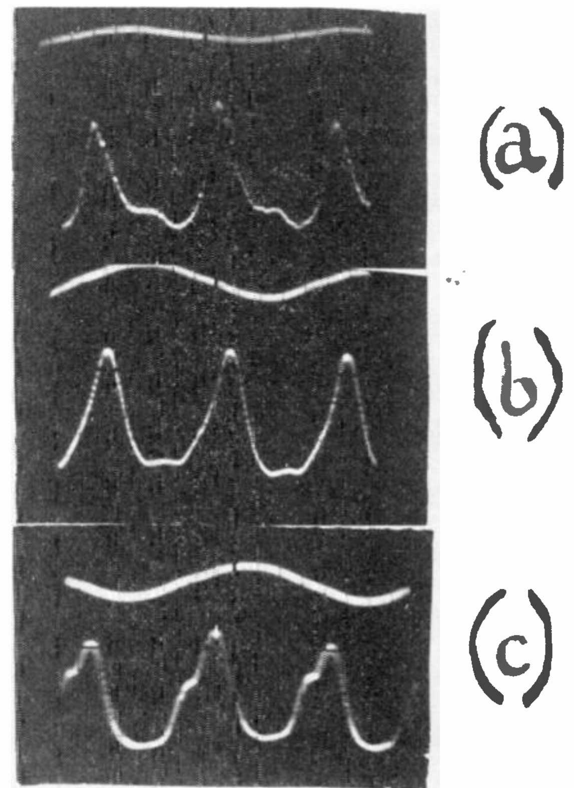


Fig. 6—Brightness waves obtained with the same sample of ZnS:Cu (Destriau, C10, blue). Metal backing, silver; frequency, 50 cps. The disturbed side changes progressively from the right (decreasing side) in (a) to the left (increasing side) in (c) while the polarity of the highest peak changes from metal backing positive in (a) to metal backing negative in (c).

Besides these variations, the brightness waveforms are dependent on the nature of the metal electrode if the electrode is in contact with the phosphor or separated from it by only a thin layer of insulator. The phase angle ϕ, modulation ratio m, and disturbance ratio d are all different according to the electronic work function of the metal backing [42]. The lower the work func-

tion, the lower the phase angle ϕ (Fig. 7). The behavior of the ratios d and m, however, is different for different phosphors. With a blue-emitting ZnS:Cu phosphor we find, for example:

Electrode	Work Function	Disturbance Ratio
	ev	
Mica	-	0.06
Palladium	4.9	0.09
Aluminum	4.0	0.17
Lithium	2.4	0.24

Furthermore, with some phosphor samples, the disturbance can be marked as a small true secondary peak while, with other samples, it is only an inflection.

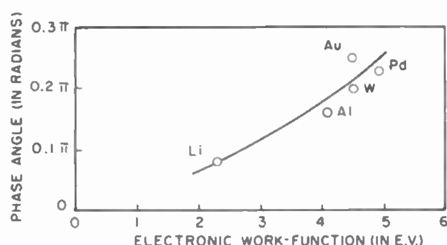


Fig. 7—For the same potential applied to the EL cell, the phase angle between the potential wave and the brightness wave is dependent on the electronic work function of the metal electrode.

The results discussed above assume a single activator. In a multiply activated phosphor the brightness wave is influenced by each activator as if this activator were present alone, and the different components can be separated by suitable optical filters. Brightness waves of EL organic compounds (Fig. 8) are more regular than

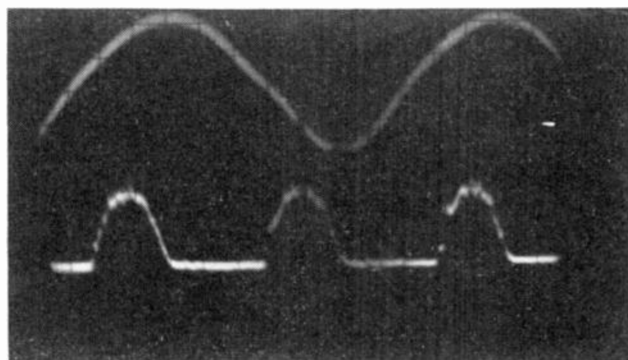


Fig. 8—Brightness wave of an organic compound.

brightness waves of inorganic phosphors; no disturbance occurs and the brightness wave is only shifted compared to the applied potential wave. However, with such compounds no experiments have been made thus far for different thicknesses or for different metal backings.

The occurrence of electrical polarization in the phosphor particles will obviously influence the shape of the brightness waves. Several workers have studied the effect of such polarization [17, 18, 18a, 44–46]. Furthermore, it is possible for carrier-injection to occur from the electrodes under certain conditions; this type of electroluminescence will be discussed more in detail in a later section. It is obvious, however, that if carrier-injection does occur, the shape of the brightness waves will be modified. Some of the effects discussed in this section undoubtedly occur because of carrier-injection. Gobrecht, Hahn, and Seeman [47] also have studied brightness waves recently.

E. Dependence of Brightness on Frequency of the Applied Voltage

The exact form of the dependence of EL brightness on frequency is different for different phosphors [26, 48–50]. In general, however, it may be said that for a material with a single emission band the brightness at low frequencies increases linearly with frequency or at a slightly lower rate. At higher frequencies there may be a tendency toward saturation or, for some materials, even a maximum output; in general the departure from linearity is greater the lower the applied voltage. Typical data are shown in Fig. 9. The only deviation from

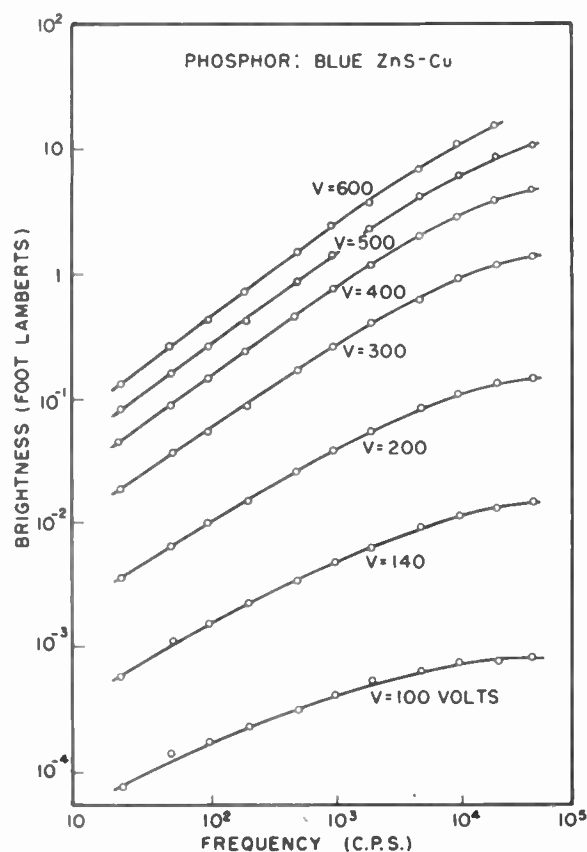


Fig. 9—Frequency dependence of the EL brightness of a blue-emitting ZnS:Cu for various applied voltages (from Lehmann).

this rule seems to be for materials with accidental or intentional additions of so-called "killers," e.g., Fe, Co, or Ni, for which the EL output may increase more rapidly than linearly with frequency [51]. It may be noted that these same materials also show "superlinearity" of brightness as a function of excitation intensity for photoluminescence at room temperature [52]. The characteristics of the embedding dielectric may in some cases also have an influence on the frequency dependence of the output.

With multiply activated phosphors, each characteristic emission band may vary with frequency at a different rate, so that the color of the emitted light changes

progressively with the frequency. The emission of ZnS:Cu, Pb [50, 53] consists of two bands, one green and the other blue; there is not complete agreement among various workers as to the origin of these two bands but this point is unimportant for the present discussion. The blue emission increases more or less linearly with frequency, but the green emission approaches saturation; at low frequency the intensity of the green band is greater than that of the blue band, but the opposite is true at high frequency. A similar behavior occurs with Cu and Mn activated ZnS [54]. At low frequency the yellow Mn band prevails, but the opposite occurs at high frequency so that the emission changes progressively from yellow to blue when the frequency is increased (Fig. 10). Because of such changes in color, the variation of brightness with frequency may be quite complicated.

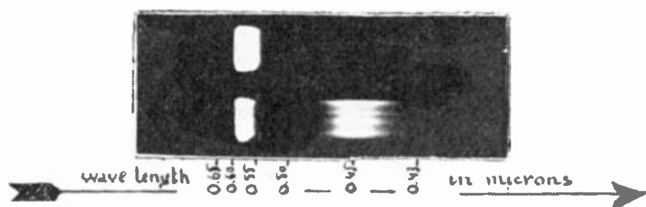


Fig. 10—Excitation of ZnS:Mn, Cu phosphor with the same potential and two different frequencies. *Above:* Frequency = 50 cps. Only the yellow band of Mn appears. *Below:* Frequency = 500 cps. Two bands, the yellow band of Mn and the blue band of copper, are emitted.

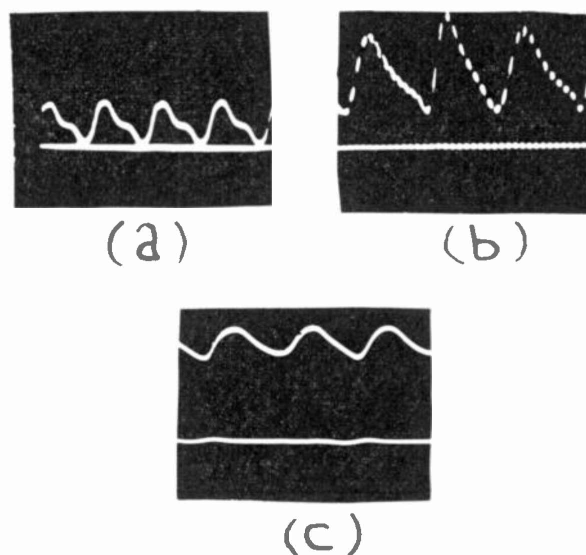


Fig. 11—Brightness waves for different frequencies, but always with the same potential applied to the EL cell. (a) $f=50$ cps, (b) $f=500$ cps, (c) $f=5{,}000$ cps.

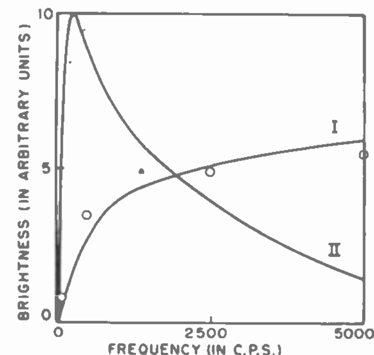


Fig. 12—Variation with frequency of the two components of the brightness wave. Curve 1, continuous component, B_0. Curve 2, periodic component, (B_1-B_0). Experimental points obtained with barrier layer photocell roughly fit curve 1 (continuous component).

The brightness of the cell, as stated, flickers with twice the applied frequency. Two peaks occur in each cycle and between two successive peaks the brightness does not, in general, fall to zero. Let B_0 be the residual brightness between two peaks and B_1 be the brightness reached at the peak. B_0 can be considered as the continuous component of the brightness wave and (B_1-B_0) as the periodic component, with frequency twice that of the applied field. The early experiments described above, made with an integrating detecting device, involved only the mean brightness; it is obvious that more accurate measurements must take into account the two components separately [55]. The higher the frequency, the higher the continuous component. The periodic component, however, first increases with the frequency, passes through a maximum, and then decreases. Some recordings are shown in Fig. 11 and the variations of B_0 and (B_1-B_0) are shown in curves 1 and 2, respectively, of Fig. 12. Experimental points obtained with an integrating device (barrier-layer photocell and galvanometer) are also shown in Fig. 12, and it is seen that they fall close to curve 1.

The relative importance of the two components depends on the mean lifetime of an excited center. The mean lifetime (10^{-4} to 10^{-3} second for a particular sample [55] can be found from the oscilloscopic trace of the afterglow which immediately follows removal of the field. Variations of the periodic component compared to the continuous components are shown in Fig. 13. As is to be expected, the shorter the lifetime, the greater the amplitude of the periodic component of the output.

As a consequence of the mean lifetime, which at high frequency becomes appreciable compared to the period of the applied field, the phase angle decreases progressively when the frequency is increased, passes through zero for a certain frequency, and then becomes negative at higher frequency. On the other hand, the brightness waves assume a more regular form at high frequency than at low frequency (Fig. 14).

F. Dependence of Brightness on Temperature

The first experiments on the temperature dependence of EL output were reported in the earlier review paper [12], but they were made with only one sample (ZnS:Cu,

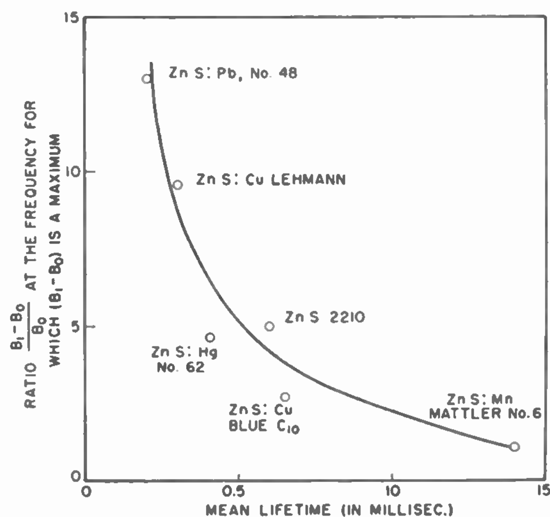


Fig. 13—Ratio of the periodic component (B_1-B_0) to the continuous component (B_0) compared with the mean lifetime of an excited luminescent center.

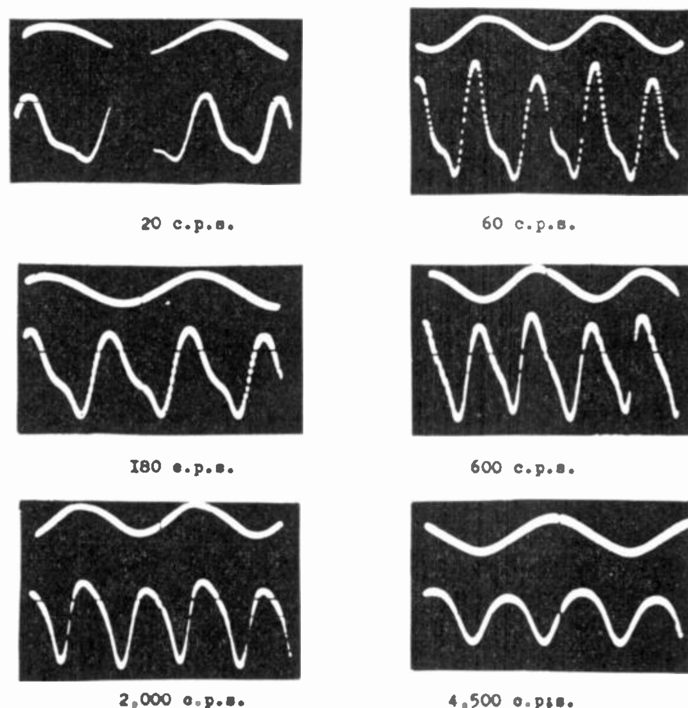


Fig. 14—Brightness waves at various frequencies. At low frequencies, from 20 to 600 cps, the potential wave (sinusoidal wave) leads the brightness wave. At 2,000 cps the potential wave and the brightness wave are in phase, while at higher frequency it is the potential wave which leads the brightness wave. At high frequency the shape of the brightness wave is less disturbed than at low frequency.

Guntz No. 13) and over a small range of temperature. Later experiments have been made with different samples over a larger range of temperature from -150 to 200°C [14, 39, 53, 56–61]. Such experiments require EL cells in which neither any change in the dielectric constant of the insulator nor any chemical interaction between the insulator and phosphor occurs.

Mattler [61] seems to have exercised great care to avoid all sources of error by a preliminary systematic study of insulators as well as of the electrostatic properties of EL condensers. In the range of temperature from -155 to $+155$°C Mattler suggests the use of EL cells in which the phosphor, embedded in a very thin layer of Araldite, is spread upon a sheet of mica, the opposite face of which is covered by a thin transparent conducting layer of tin oxide. The metal backing is directly deposited upon the phosphor layer. The dielectric constant of Araldite changes appreciably at about 120°C [62], but this change is not of too much consequence for the electrostatic properties of the EL cell, at least if the thickness of the Araldite is very small compared to the total thickness of the cell. The change in the capacity of such a condenser is roughly a few per cent between -155 and $+80$°C and about 35 per cent between $+80$ and $+155$°C.

The temperature dependence of the brightness is very different depending upon the phosphor. Some results of Mattler [61] are reported on Fig. 15. With some phosphors (for example, ZnS:Cu, ZnS:Ag, or ZnS:Cu, Pb) the brightness first increases with the temperature, passes through a maximum and then decreases. With other phosphors (for example, ZnS·ZnO:Cu and ZnS·ZnSe·ZnO:Cu) the brightness decreases progressively when the temperature is increased from -155°C. Other workers [57–59] have obtained curves with more

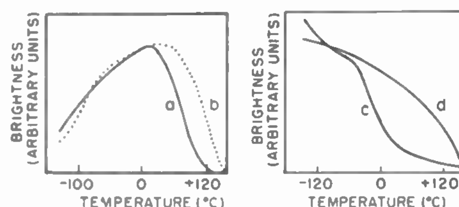


Fig. 15—Dependence of EL brightness on temperature (from Mattler [61]). Curve a—ZnS:Cu, blue. Curve b—ZnS:Cu, green. Curve c—ZnS·ZnSe·ZnO:Cu. Curve d—ZnS:Mn.

than one peak. A connection between thermoluminescence and temperature dependence of electroluminescence, indicating that electron traps play a role in both processes, has been suggested by various observers [56–60, 63].

Here also, as already indicated for the dependence of brightness on frequency, with multiply activated phosphors each characteristic spectral band may vary in a different manner so that the color of the emitted light can be changed greatly either by heating or by cooling. Thus, Mattler observed that in ZnS:Cu the blue band prevails at low temperature, but the green band at high temperature. Similarly, for ZnS:Cu, Mn the relative importance of the blue or green Cu band compared to the yellow Mn band is greater at low temperatures. Besides these variations, a progressive shift of the spectral band can occur even in phosphors with only one

activator; such variations will be reported later in the next section concerning the spectral emission and its variations. Brightness wave forms also change progressively with temperature [20, 58, 64]; Mattler [64] and Haake [58] have recently shown that the disturbance or secondary peak shifts progressively from the descending side of a main peak to the ascending side of the following main peak as the temperature is lowered (Fig. 16).

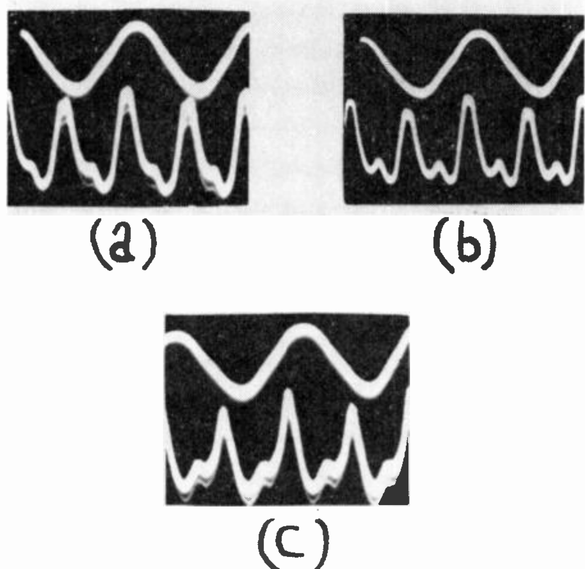


Fig. 16—Disturbances of the brightness waves of EL at low temperatures (from Mattler [64]). (a) 290°K, (b) 199°K, (c) 105°K. The secondary peak passes progressively from the descending side of a main peak to the ascending side of the following main peak as the temperature is lowered.

G. Dependence of Spectral Emission on Voltage, Frequency, and Temperature

The spectral emission of EL is in general dependent on the following factors: (1) voltage, (2) frequency, and (3) temperature. Two kinds of modifications, however, must be considered separately.

1. Spectral modifications which occur in phosphors with multiple activators; such modifications result primarily from a variation in the relative importance of the various bands. In case all the bands are not emitted at the same time in each cycle of the field variation, then periodic spectral changes can also occur with twice the applied frequency.
2. Intrinsic modifications which occur in a single band characteristic of one activator. Such effects are generally small compared to those involving more than one emission band.

Consider first phosphors with many activators. It is obvious that the characteristic spectral bands of different activators, or even two bands of the same activator, may not be equally sensitive to photoexcitation and to EL; one would therefore expect to obtain a different spectral emission depending on the kind of excitation [12]. Such behavior is shown schematically in Fig. 17 for a ZnS:Cu phosphor exhibiting the two bands (blue and green) of copper; the blue band, in this material, is more sensitive to EL than is the green band.

The changes in color with frequency and temperature of multiply activated phosphors have already been discussed. It may be remarked that color shifts produced by an increase in frequency are similar to those produced by a decrease in temperature, and that emission of shorter wavelength is promoted by high frequency or by low temperature. Changes in spectral emission with voltage are not generally as pronounced as those with frequency or temperature. One exception to this statement is ZnS:Cu, Mn, where an increase in voltage favors the yellow Mn emission at the expense of the blue (or green) Cu emission. This observation, however, is in agreement with the fact that the brightness of ZnS:Mn increases faster as the voltage is increased than does that of ZnS:Cu; in the doubly activated phosphor each band, therefore, retains its individual behavior. Waymouth and Bitter [64a] have also recently reported that the spectrum of the light emitted at the removal of a static voltage differs from that emitted at application of the voltage.

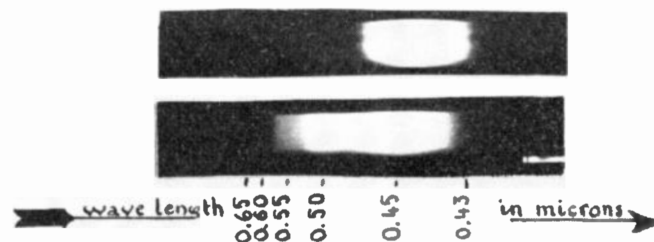


Fig. 17—The spectral output of a ZnS:Cu phosphor showing for photoexcitation (*below*) two bands (green and blue), but giving only the blue band for EL (*above*).

All the characteristic bands of each activator are not emitted simultaneously during a cycle of field vaiation. It has been shown that at low frequency the brightness waves lead the applied potential waves by a phase angle which (other experimental conditions being similar) is different according to the activator; for example the phase angle is greater for Cu than for Mn. Thus, with ZnS:Cu, Mn the maximum of the Cu band occurs a short time before the peak of the yellow Mn band [54] (about 10^{-3} seconds at a frequency of 50 cps). During this short time the color changes, passing quickly from blue to yellow, and this change is repeated each half-cycle. Even with some phosphors with only one activator, the different spectral regions are not emitted simultaneously. In this case the brightness waveforms are dependent on the spectral region observed [54, 65].

Besides the spectral variations discussed above intrinsic modifications can also be found for an individual emission band itself. If the temperature is lowered, an effect is found similar to the change which occurs in photoluminescence. Variations with frequency have not been reported, but small changes with voltage have been found [12]. In this last case the peak of the emission band can be slightly displaced either towards shorter wavelengths (as for example with the green band of $ZnS:Cu$) or towards the longer wavelengths (as for example with $ZnS:Mn$); these variations, however, are indeed of very small magnitude.

H. Experiments in High Magnetic Fields

A possible quenching of electroluminescence by magnetic fields is to be expected if the path length of the electrons in the electric field direction is of any consequence. The most recent effort to observe such an effect was by A. N. Ince [66] with magnetic field intensities as high as 130,000 oersteds, but no effect was found either on the light output or on the brightness waveform.

The path of a free electron crossing mutually perpendicular electric and magnetic fields is a cycloid [67]. The maximum energy (W_m) attained by the electron at the peak of this cycloidal curve is

$$W_m = 2m(E/\mu H)^2, \qquad (12)$$

where m is the electron mass, E is the electric field strength, H is the magnetic field strength, and μ is the magnetic permeability. Taking into account that in ZnS the mean depth of a luminescent center is 3 ev and assuming this value for W_m, Ince predicts from the above formula that a magnetic field of the order of magnitude of

$$\mu H = E(2m/W_m)^{1/2} \qquad (13)$$

(that is to say, $H = 20{,}000$ oersteds for $E = 10{,}000$ volts/cm), should give noticeable quenching. Since a magnetic field six times as great is found experimentally not to produce any change in the light output, one can assume that the light emission is produced in a barrier where the local electric field is much higher than the mean applied field.

Ince's calculation, however, does not take into account the mean free path length of the electrons. The energy gained by an electron between collisions is determined by the projection of the trajectory on the electric field direction. A possible quenching of EL under the action of a magnetic field must be ascribed to a diminution of this projection of the mean free path which, for a cycloidal path, is given by

$$l = l_0\left(1 - \frac{e}{m}\,\frac{l_0\mu^2H^2}{8E}\right), \qquad (14)$$

where l_0 is the distance traveled in the field direction without a magnetic field and l is the corresponding value in the presence of a magnetic field.

From this result, similar calculations can be made for all theories of EL which assume direct ionization of luminescent centers by accelerated electrons. For example, in (2) for the voltage dependence,

$$B = aE^2 \exp(-b/E), \qquad (15)$$

according to the theory to be discussed later, b should depend on the mean free path and the following expression has been given [12],[2]

$$b = b_0 + qH^2 = b_0 + (0.0220\, W/b_0E)H^2, \qquad (16)$$

where W is the energy (in ev) of the center to be ionized, E and H are given in practical units (E in volts/cm and H in oersteds), and b_0 is the value of b in the absence of a magnetic field. For a typical phosphor ($W = 3$ ev, $E = 12{,}500$ volts/cm, and $b_0 = 5 \times 10^4$ volts/cm) this equation predicts a decrease of 4 per cent in brightness for $H = 10{,}000$ oersteds, a value comparable in order of magnitude to that obtained by Ince. Both the above formula and that of Ince show that for the same effect on output, the required magnetic field intensity is lower, the lower the electric field strength. A noticeable effect could be then obtained more easily in very low electric fields.

An action of magnetic fields (50,000 oersteds) on luminescence was reported long ago by Rupp [68] who observed a momentary enhancement of phosphorescence similar to the Gudden and Pohl effect. However, Mattler and Destriau [69], working with a magnetic field of 60,000 oersteds and a large number of phosphors, were unable to verify this effect.

I. Changes of Electroluminescent Brightness with Time

It has been known for some time [12] that the brightness of an EL condenser increases slowly during the first few minutes following the field application. Vigean [70] has recently found, with a copper activated zinc sulfide, an increase of about 20 per cent between 5 seconds and 10 minutes after the field application.

In the same way, the study of brightness waves shows that with some EL phosphors, such as $ZnS:Mn$ or the blue band of $ZnS:Cu$, the first few peaks start from about zero and then increase progresssively while with other samples, such as $ZnS:Pb$ or the green band of $ZnS:Cu$, the first peak is already very noticeable. Frankl [71] has observed that equilibrium is attained much more rapidly if half-wave rectified alternating voltage is used than if both half-cycles are applied to a single crystal of ZnS.

Other modifications in time involving an effect of fatigue or mechanical or chemical modifications of insulator and phosphor crystallites will be described in the section devoted to technical applications of EL.

[2] The exact expression is $b = b_0(1 - \beta H^2)^{-1} \sim b_0(1 + \beta H^2)$.

J. Intrinsic Electroluminescence of Single Crystals

Electroluminescence of single crystals of zinc sulfide has been reported by several workers [71–74, 20a]. Electrical connection to such crystals is generally made by electrodes in direct contact so that the possibility of carrier-injection electroluminescence also exists. Indeed, brightness waves obtained on single crystals generally show two well resolved peaks for each half-cycle of the applied potential. One peak ("out-of-phase" component) is approximately 90 degrees out of phase with the applied potential and is generally believed to be similar in nature to the intrinsic electroluminescence of powders; the other ("in-phase" component) is in phase with the applied potential and presumably arises from carrier-injection from the electrodes. This explanation was first proposed by Piper and Williams [72]. Frankl [71] differs with this view, however, and believes that all the excitation occurs in phase with the applied voltage; the out-of-phase emission he attributes to electrons returning to the region of their origin when the field decreases later in the cycle. The in-phase component increases with voltage at a more rapid rate than does the out-of-phase component; hence, at low voltages the latter predominates while at higher voltages the former is more important [73]. The spectra of the two components are reported to be identical by Frankl [71], but this seems in disagreement with the results of Watson, Dropkin, and Halpin [73].

Observations on single crystals of ZnS may be complicated by the faulty crystal structure and the impurity content of even the best samples [20a]. Many specimens show nonuniformity of potential drop across the crystal indicating the presence of internal rectifying barriers [73, 74]. Piper and Williams [72] found that the two output pulses (of each kind) in each cycle originated at different electrodes; they found light to be emitted only when the electrode considered was becoming negative in the cycle. Waymouth and Bitter [18] and Zalm, Diemer, and Klasens [20] also arrived at the same conclusions concerning individual particles of a powder. Watson, Dropkin, and Halpin [73], however, observed that in their experiments the emitting region was in the interior of the crystal and was not affected by a change in phase of an applied alternating voltage or by reversal of polarity of a direct voltage. Frankl [71] has pointed out that if a crystal which seems to show emission at localized regions is immersed in a liquid with high optical index of refraction, the emission becomes much more uniform. This implies that the apparent localization may result from optical scattering and trapping effects in the crystal.

The light output from ZnS single crystals has been reported by Piper and Williams [72] to vary linearly with the current through the crystal; Frankl [71], however, finds that only the in-phase component is strictly proportional to current, while the out-of-phase component increases at a slower rate. Scarcity of quantitative data on single crystals precludes any extensive discussion at the present time. EL of single crystals of CdS and other materials has also been reported, but seems generally to arise from carrier-injection; discussion will therefore be reserved for a later section.

K. Theory of Intrinsic Electroluminescence[3]

The spectrum of intrinsic EL is generally similar to the photoluminescent spectrum of the same material in the sense that the same emission bands occur in both spectra [50], although they occur in different proportions in the two cases depending on temperature, frequency of the applied voltage, etc., as discussed above. For this reason it is usually assumed that the activator centers responsible for the emission are the same in the two cases and that only the excitation mechanism differs.

The predominant term in the empirical equation (2) for the voltage dependence of EL brightness is, of course, the exponential factor. Destriau [12, 75] early pointed out the similarity to the results of problems involving penetration of potential barriers in high electrostatic fields and suggested some unspecified process of internal field emission. The Zener effect, or direct electronic transition from the filled valence band to the conduction band, is now quite well known in semiconductors. In the case of materials emitting in the visible region, and which must therefore have a forbidden energy zone wider than two electron volts, however, the calculated field strengths for such a process are of the order of magnitude of 10^7 volts/cm, which is much greater than the fields experimentally used. A model based on direct field excitation of the electroluminescent centers meets with similar difficulty, although Ueta [27] suggests such a mechanism.

D. Curie [44, 76] studied the problem of EL extensively and proposed a three-step excitation process as follows:

1. Transfer of electrons from donor levels to the conduction band under the action of the applied electric field and/or temperature.
2. Acceleration of these electrons in the conduction band. Carrier multiplication or avalanche formation may occur.
3. Collision of these electrons with luminescent centers, which are thereby excited or ionized, or with the basic crystal lattice itself and the consequent production of electron-hole pairs.

Garlick [77] earlier had mentioned the possibility of such a mechanism, but gave no detailed discussion.

[3] Since this section was written W. W. Piper and F. E. Williams have published a paper on the "Theory of electroluminescence," *Phys. Rev.*, vol. 98, pp. 1809–1813; June 15, 1955. Williams also presented a paper on this topic at the Symposium on Luminescence, Polytechnic Institute of Brooklyn; September 9, 1955.

Curie was not very specific about step 1 of this process, but in any case he assumed that it is not the important rate-determining step. The occurrence of the exponential term in (2) then arises from consideration of the free path lengths of the accelerated electrons. It is required that the electron attain an energy equal to or greater than $W = h\nu$ from the electric field in traveling a path l, or

$$W = h\nu = elE. \tag{17}$$

It is well known, however, from kinetic theory, that in a gas the fraction of particles having path lengths which equal or exceed a certain value l is given simply by

$$\exp(-l/\bar{l}), \tag{18}$$

where $\bar{l}$ is the mean free path. Combination of these equations leads to

$$B \approx \exp(-W/e\bar{l}E) = \exp(-b/E). \tag{19}$$

It may be noted that this result was also given earlier by Destriau [12] and in a treatment of dielectric breakdown by Seitz. Curie gave no quantitative discussion of the pre-exponential factor of (2), but only considered it as proportional to the number of accelerated electrons. Although he placed the site of the excitation in the bulk of the phosphor crystals, he did indicate that surface potential barriers might play a role in the initial production of carriers (step 1).

Piper and Williams [38, 72, 78] also independently considered field acceleration of electrons as the exciting mechanism of EL. Their quantitative treatment differs from Curie's, however, in that the rate-determining mechanism also includes the initial production of carriers (step 1, above). They assume the presence of a surface potential barrier; this barrier is essentially of the Mott-Schottky exhaustion type and arises from ionization of shallow donor levels. The phosphor is also assumed to possess deep donor levels lying perhaps 0.5 ev below the conduction band. The concentration of deep donors is assumed to be much less than that of shallow donors, so that the latter alone determine the potential distribution in the barrier region. Most of the applied potential appears across the barrier and even for moderate applied electric field strengths the field in the barrier region will be sufficient to produce ionization of these deep donors, and subsequent acceleration and collision can lead to luminescence. The process is shown schematically in Fig. 18. They assume that the donors will be ionized completely if the local field is equal to or exceeds a critical value

$$E_c = 2.4 \times 10^6 \epsilon^2 \text{ volts/cm}, \tag{20}$$

Where ϵ is the depth of the donor below the conduction band, in electron volts; for $E < E_c$ no ionization is assumed to occur. These authors tried several simple assumptions as to the distribution in energy of the donors and found best agreement with their experimental results on single crystals for the case of a linear distribution; i.e.,

$$n(\epsilon)d\epsilon = n_0 \epsilon d\epsilon. \tag{21}$$

It is further assumed that the ionizing efficiency of the electrons in the conduction band is given by

$$\eta = c[E_c^2(\epsilon) - E_t^2], \tag{22}$$

where E_c is given by (20) and E_t is a threshold field strength necessary for electrons to be accelerated in the conduction band. In this case the integrated light output over a cycle of the applied voltage is given by

$$B = a\left[1 + 3\left(\frac{V_t}{V}\right)^{1/2}\right]\left[1 - \left(\frac{V_t}{V}\right)^{1/2}\right]^3 V^2, \tag{23}$$

where V_t is a threshold voltage corresponding to E_t. The fit to their experimental data, however, is not too good.

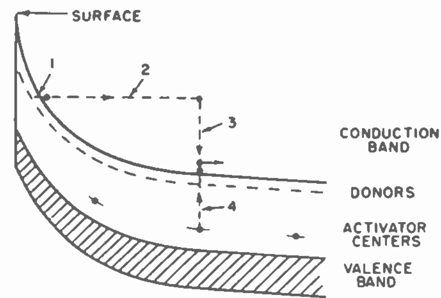


Fig. 18—Schematic representation of the collision mechanism of electroluminescence. Electrons from donor levels in the surface barrier region are liberated by the action of the high electric field and/or temperature (1), are then accelerated by the field (2) to acquire energies above the bottom of the conduction band, and collide with activator centers whereby they lose their energy (3) and the activator center is ionized (4) or excited. Emission may occur immediately, or later in the cycle when the applied potential is reversing.

Burns [29], Zalm, Diemer, and Klasens [20, 79], and Fischer [80] have also discussed the theory of EL along the same lines proposed by Piper and Williams. Howard [81] has extended the quantitative considerations pertaining to this model. The major difference from Piper and Williams concerns the dependence of the rate of ionization of the donor levels upon the field strength. For this purpose an expression derived by Franz [82] was used, that is,

$$p = aE \exp(-b\epsilon_g^{1/2}\epsilon^{3/2}/E), \tag{24}$$

where ϵ_g is the width of the forbidden band, ϵ is the depth of the donor below the conduction band, E is the field strength, p is the rate of the emptying process, and a and b are constants. The further assumptions which were found to give the best fit to experimental data were: (1) donors lying at a single depth below the conduction band and of density independent of position in the barrier region, and (2) an ionization efficiency independent of the energy of the accelerated electrons,

and therefore of the applied voltage. The last assumption implies that the potential drop across the surface barrier is very great compared to that across the bulk of the crystal; the result of the analysis is the same, however, if an exponential relation similar to (19) above is assumed for the ionization efficiency. The resulting expression for the light output is

$$B = a'\left[\frac{V^{1/2}}{C}\left(\frac{V^{1/2}}{C} - 1\right) \cdot \exp\left(-\frac{C}{V^{1/2}}\right) - Ei\left(-\frac{C}{V^{1/2}}\right)\right], \quad (25)$$

where a' and C are constants and Ei is the customary exponential integral. If this integral is expanded into an asymptotic series the expression simplifies to

$$B = aV^{3/2}\left(1 - \frac{3}{C}V^{1/2}\right)\exp\left(-\frac{C}{V^{1/2}}\right). \quad (26)$$

This result is sufficiently accurate for most purposes.

The fit of (26) to experimental data is shown in Fig. 19; it should be noted that the fit is good over five decades in brightness. The data of Fig. 19, however, are for a powdered phosphor embedded in a dielectric. In such a cell it is observed [17, 41, 79] that at very low voltages only a few particles are visible; as the voltage is increased the number of emitting particles increases, and for very high voltages the brightness of all particles (regardless of whether they started to emit at low or at higher voltages) is practically the same. For this reason, data on an assembly of particles probably should not be taken as representative of the process actually occurring in individual particles. Waymouth and Bitter [17] found that the voltage dependence of the emission from a single particle was not the same as that of an entire EL cell; they attributed this to the fact that different particles have different sensitivities. Zalm, Diemer, and Klasens [79], on the other hand, report a similar behavior for individual particles and for an assembly of particles. Eq. (26), in any case, also gives good fit to the data of Piper and Williams on single crystals (although it should be noted that only six experimental points are available).

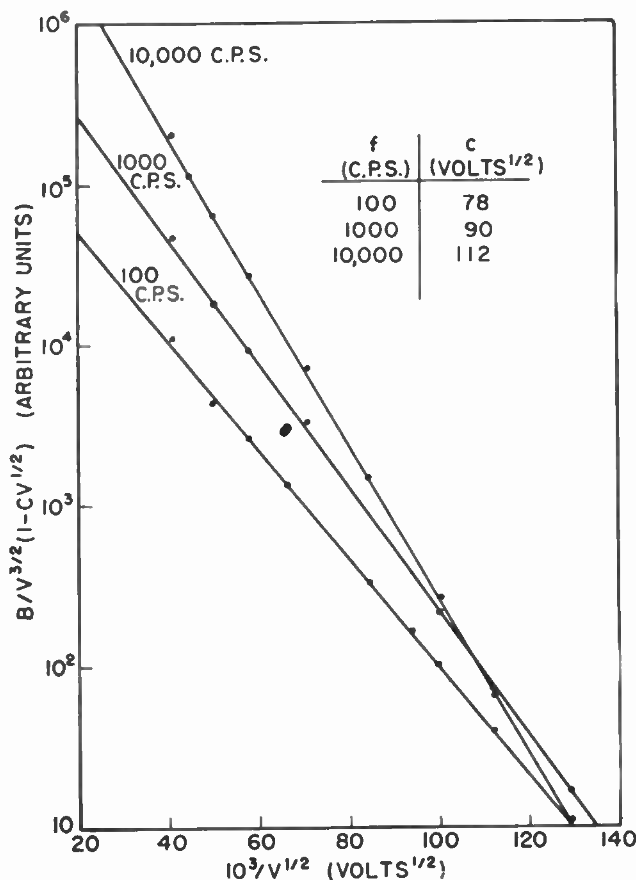


Fig. 19—Fit of the theoretical equation (26) to experimental data on the voltage dependence of EL brightness (from Howard [80]).

Howard's calculation actually does not take into account the sinusoidal variation of voltage usually employed and actually used to obtain the data for Fig. 19. Because of the rapid variation of EL output with voltage, however, most of the emission occurs at the peak of the voltage wave where the variation in V is small so that calculations based only on the peak value, or the corresponding rms value, of V are sufficient for practical purposes. In experimental support of this assumption, Waymouth and Bitter [17] found that the integrated light output resulting from application of voltage pulses was dependent only on the maximum voltage and independent of the rise time of the pulse for values between 50 μsec and 50 millisec. Similarly Piper and Williams [38] report that the integrated light output produced by a triangular voltage pulse depends only on the peak voltage and is independent of the width of the pulse if the latter is between 50 μsec and 0.5 sec. Furthermore, Nudelman and Matossi [18] found that brightness data obtained with square-wave excitation could be fitted by the empirical equation of Destriau, (2) with $n=2$, first proposed for sinusoidal excitation. All of the theories advanced thus far assume that the excitation is not so intense that the number of donors available for emptying by the field changes appreciably during a cycle of the applied voltage; they thus cannot be used to describe the approach to saturation.

Other modifications of the Piper-Williams theory have been proposed. Taylor and Alfrey [56] suggest replacing the original expression for ionization efficiency, (22), by an exponential expression similar to (19), with the result

$$B = aV^{7/4}\exp(-b/V^{1/2}). \quad (27)$$

They generalize this result by replacing the exponent 7/4 by n and indicate that excellent fit to the Piper and Williams data is obtained with $n=1$. This equation (with $n=1$) has also been proposed as an empirical result by Schwertz and his co-workers [83]; they claim better fit than for the empirical equation (5) discussed

earlier, but it should be noted that their data extends only over about one and half decades in brightness. Taylor and Alfrey [56] also developed a theory involving thermal, rather than field, ionization of the donors and obtain simply

$$B = a \exp(-b/V^{1/2}). \tag{28}$$

Zalm, Diemer, and Klasens [79, 84] have also discussed the voltage dependence of electroluminescence. Although their various theoretical assumptions as to donor distribution, etc., predict a variety of pre-exponential factors, they conclude that the simplest form, (28), is to be preferred. In view of the similarity of (26), (27), and (28), and the predominating role of the exponential factor, any of these expressions is suitable if the observations do not extend over too wide a range. Zalm, Diemer, and Klasens [79], however, found that their data fit [28] over more than seven decades in brightness; with such a wide range of data the choice of a pre-exponential factor should be fairly definite. It will be noted that (28) predicts a saturation in output as the voltage is increased.

The assumption that a surface potential barrier plays an important part in the mechanism of electroluminescence is, of course, consistent with the many examples cited in Section IB concerning the sensitivity of EL phosphors to surface conditions and surface treatment. In such a surface-dominated process, the possible influence of localized electronic levels at the surface, i.e., "surface states," in determining the barrier height or in serving as sources of electrons cannot be excluded. O'Neill [85], on the other hand, has assumed that surface states are involved in the light emission process itself; in view of the correspondence, as noted above, between the emission bands in photoluminescence and in electroluminescence, such a view does not seem tenable unless one is also prepared to abandon all luminescence theories now in vogue. Much work remains to be done, however, on refinement of the calculations pertaining to the barrier mechanism. Correlation of experimental data on EL with the physical constants of the barrier as obtained from separate and independent types of measurement will also be necessary to establish firmly the theoretical interpretation.

A complete theoretical treatment of the frequency dependence of EL has not yet been satisfactorily accomplished. This problem is, of course, confused by the sometimes contradictory behavior of different materials. Howard [80] has considered that, for the same voltage applied to the phosphor, the voltage appearing across the surface barrier (which will be the effective voltage), will still be a function of frequency. His equivalent electrical circuit for the phosphor particle is simply a capacitor representing the barrier in series with a resistor representing the bulk of the particle, and he obtains

$$V' = V^{3/2}[V + gf^x(1 + gf^x/V)^{1/2}]^{-1/2}, \tag{29}$$

where V is the applied voltage, V' is the effective voltage, f is the frequency, and g and x are constants. This result, in combination with an *a priori* assumption that for a given voltage across the barrier the ouput will vary linearly with frequency (i.e., equal output at each reversal of field), seems to explain experimental data in many cases. Zalm, Diemer, and Klasens have also discussed the equivalent circuit of an EL cell and its consequences [20], as well as other considerations pertaining to the frequency dependence of electroluminescence [79].

It is obvious that any full explanation of the frequency dependence of EL brightness and the related color shifts must also consider the lifetime of the excited luminescent centers [55]. One determinant of this lifetime, according to the commonly accepted Schön-Klasens theory of luminescence, will be the behavior of holes in the valence band of the phosphor. Zalm, Diemer, and Klasens [20] have used the diagram of Fig. 20

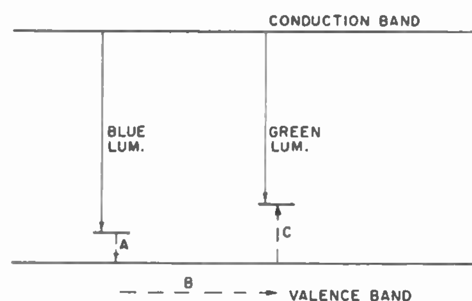


Fig. 20—Energy diagram of a ZnS phosphor containing blue and green luminescent centers. The mechanism for the transport of energy from the blue to the green centers by hole migration is shown by the arrows A, B, and C (from Zalm, Diemer, and Klasens [20]).

to explain the observed shift in emission of ZnS:Cu from green to blue as the frequency is increased. An empty blue center will be filled by an electron from a filled green center if sufficient time and activation energy for the transition is available (the arrows in the diagram refer to the path of the hole which actually is responsible for the transition). At high frequencies or low temperatures this will not be the case and the blue emission will reach its full value. The same model also explains the fact that the blue emission decays at a faster rate than does the green emission [18,20]. Spatial separation of the electrons and the centers from which they came, *i.e.*, polarization or trapping, will also play a role in the variation of brightness with frequency [17, 18, 29, 44] and with temperature. Haake [58, 86] has emphasized the fact that the temperature dependence and the frequency dependence of EL must be considered simultaneously rather than separately.

D. Curie [44] has obtained a theoretical expression for the frequency dependence of EL output on the assumption that the electrons recombine with luminescent centers according to a bimolecular law, which is generally true for sulfide-type phosphors. His result is

$$B = \frac{Kn_0^2\alpha}{1 + (n_0\alpha/2f)}, \tag{30}$$

where K is a constant of proportionality, n_0 is the initial concentration of electrons in the conduction band, and α is the recombination coefficient defined by $dn/dt = \alpha n^2$. If α is large compared to f (i.e., if the lifetime of excited centers is short compared to the period of the applied field), then the brightness should be proportional to f, as is generally observed. If the frequency is made high enough ($n_0\alpha/2f \ll 1$), a saturation output should be observed, again in agreement with experiment. The result as given above is based on very elementary considerations, and Curie has pointed out that n_0 may not be independent of the frequency.

The theoretical treatment of Taylor and Alfrey [56] also yields an expression for the temperature and frequency dependence of EL brightness which they find in good agreement with their data for single crystals of ZnS:Cu. They obtain for the factor a in (28),

$$a = \gamma \exp(-\gamma/4f), \tag{31}$$

where γ is the usual expression for the rate of thermal release from traps,

$$\gamma = s \exp(-E/kT). \tag{31a}$$

Here E is the energy separation of the trap from the conduction band and s is the "attempt-to-escape" frequency. Eq. (28) applies only if $\gamma V/2\pi fb > 1$. Johnson, Piper, and Williams [59] also have considered the effect of electron trapping on the temperature dependence of EL and obtain the relation

$$B = 1 - C[1 - \exp(-\gamma_2/2f)] \exp(-\gamma_1/2f), \tag{32}$$

where γ_2 and γ_1 are the temperature-dependent rates of removal of electrons from, and return to, the excited activator centers, as given by expressions similar to (31a). Eq. (31) predicts a maximum in B as T is varied, while (32) predicts a minimum. In either case, however, the reciprocal of the temperature of this minimum or maximum should be proportional to the negative logarithm of the frequency. Both the theories just discussed assume that the release of electrons from traps to the conduction band is of importance. Haake [58] has discussed the effect of such release on the shape of the brightness waves. He has also emphasized the necessity of differentiating between the temperature variation of the light emission process (which will also be evidenced in photoluminescence, for example) and that of the electroluminescence excitation process.

Goffaux [87] has recently discussed the mechanism of EL from a slightly different point of view. In a manner analogous to one of the theories of dielectric breakdown (that of Fröhlich), he considers that the equivalent temperature of the electrons may be raised by the electric field to a value above the thermodynamic temperature to the crystal lattice. The consequences of this method of attack, as applied to EL, have not yet been given in detail.

II. Carrier-Injection Electroluminescence

A. Experimetnal Observations

As mentioned earlier, the phenomenon now denoted as carrier-injection EL was first observed by Lossew [3, 88, 89] on silicon carbide crystals while he was investigating their properties as radio detectors, a fact which may be of historical interest to readers of this journal. The EL of silicon carbide has since been studied by a number of workers [90–94]. Experiments on this material are complicated by its refractory nature; crystals are generally not very perfect and are of variable impurity content. It is well known that silicon carbide crystals exist with both p-type and n-type electrical conductivity [95] and exhibit rectifying properties.

Lossew [88] observed two different types of EL in silicon carbide:

1. "Luminescence I" is blue-green in color and occurs in localized spots at the electrode when the crystal is biased in the reverse direction.
2. "Luminescence II" is generally yellow or orange in color and extends over the surface of the crystal near the electrode contact when the crystal is biased in the forward direction.

Some crystals were found to exhibit only Luminescence I, while others could exhibit either type under certain conditions, or sometimes both kinds simultaneously. Lossew and Claus [90] differ as to the electrode polarity for the two effects, but this is presumably due to the fact that they used crystals of different type conductivity. In general Luminescence I is found at the anode and Luminescence II at the cathode. Lossew commented on the fast response of the EL of SiC.

The Luminescence II of silicon carbide has been investigated most extensively. According to Lehovec, Accardo, and Jamgochian [93], the spectrum of the yellow emission consists of two overlapping bands with peaks at 5,500A and 6,100A; in purer crystals, however, a green emission with peaks at 4,750A and 5,250A was observed. The spectrum of the green emission is independent of temperatures, while in the case of the yellow-emitting samples, the 5,500A peak became greater relative to the 6,100A peak as the temperature was decreased to −150°C. Szigeti, Bauer, and Weiszburg [94], however, with better resolution, found five peaks at 4,900A, 5,200A, 5,400A, 5,850A, and 6,100A in a single specimen. Such a spectrum must correspond to a very complicated activator system; the absorption edge of silicon carbide itself lies at 4,270A (2.9 ev). Although Lossew [88] observed a change in emission color as the current (or voltage) was increased, this was apparently due to heating of the specimen; Lehovec, Accardo, and Jamgochian [93] observed no effect of current on the spectrum.

Lehovec, Accardo, and Jamgochian [93] also find that the light output is related to the current by the expression

$$B = bi - a, \tag{33}$$

where a and b are constants and a is interpreted as due to leakage around the rectifying barrier or some such shunting process. Emission is observed at voltages as low as 1.8 volts. Although at room temperature only one quantum of light is emitted for 10^6 injected electrons, the efficiency increases rapidly as the temperature is lowered, and extrapolation indicates an efficiency of unity at about 80°K. This temperature dependence has been explained [93, 94] on the usual basis of a competing temperature-dependent non-radiating process. G. Curie and D. Curie [96] have commented on the simplicity of the brightness waves obtained with SiC compared to those for ZnS.

Radiation resulting from the injection of minority carriers across p-n junctions (biased in the forward direction) in germanium and silicon was first observed by Haynes and Briggs [97] and later studied by others [98–101]. In this case the position of the radiation peak is observed to agree very closely with the optical absorption of the material, i.e., with the width of the forbidden energy band between the valence and conduction bands. For germanium and silicon, therefore, the radiation is not visible; in germanium the spectral peak is at 1.77 microns (0.70 ev) and in silicon at 1.12 microns (1.10 ev). In very thin germanium specimens a second emission peak is also observed at 1.5 microns [98]; this has been correlated by Haynes with the known complexities of the optical absorption of germanium. The emission is quite localized at the p-n junction; emission has also been obtained with point contacts. The output is proportional to the injected current. For the particular specimens and operating conditions used by Haynes and Briggs [97], one quantum was emitted for about 5,000 injected minority carriers at room temperature; this efficiency was increased by a factor of roughly five at the temperature of liquid nitrogen. Because of the high refractive index of these materials, light trapping may easily occur [100, 101]; the actual efficiencies of photon production may, therefore, be considerably higher than the figures given here. An apparently similar EL phenomenon has been observed by Wolff and Hebert [102] in gallium phosphide; here, because of the wider energy gap, the radiation lies in the visible red and near infra-red regions. The emission was observed near the cathode; presumably, then, the injection occurred at a contact barrier rather than at a p-n junction. Similar carrier-injection EL is apparently to be expected in other so-called type III–V semiconductors.[4]

A different type of emission has been observed in silicon p-n junctions biased in the reverse direction by Newman and his co-workers [103]. Here the emission is yellowish-white and frequently occurs at localized spots. The output is very low and corresponds to an optical efficiency of 10^{-8} lumens/watt. The similarity to the Luminescence I observed in silicon carbide is quite obvious; correspondingly, the infra-red emission which occurs when such a junction is biased in the forward direction may be considered as analogous to Luminescence II in silicon carbide.

The EL of zinc sulfide crystals has been discussed in a previous section, where it was pointed out that with this material it is difficult to separate intrinsic effects and those resulting from carrier-injection. Piper and Williams [72] found that the magnitude of their in-phase component was dependent on the nature of the metallic cathode contact; metals with low electronic work function resulted in lower threshold voltages than those with high work functions. Frankl [71] reports that the efficiency of the in-phase component is less than that of the out-of-phase component. He found that the efficiency drops rapidly as the voltage across the crystal was increased; at the lowest voltages employed he obtained one photon for roughly every five electrons passing through the crystal. It might be remarked that at the present time no way is known to make p-type ZnS.

Although cadmium sulfide is generally similar in its photoluminescent properties to zinc sulfide (all wavelengths being shifted to longer values because of the narrower band separation of CdS, which is 2.5 ev compared to 3.7 ev for ZnS), the EL properties of single crystals of the two materials [23, 104, 105] are quite different. This arises, at least in part, from the much higher electrical conductivity of CdS compared to that of ZnS. Kröger, Vink, and van den Boomgaard [106] varied the conductivity of CdS over wide limits by the addition of impurities, but were unable to accomplish conversion to p-type conductivity. The observations on EL of CdS crystals that have been made are confusing in many respects because the crystals used by different investigators apparently varied widely in conductivity.

Böer and Kümmel [23] studied CdS crystals at liquid air temperature and in static fields of the order of 10^5 volt/cm. They observed emission extending over the entire crystal in a narrow range of field strengths just below dielectric breakdown, which was thermal in nature. They state that the EL was identical in color to the photoluminescence (which may be green or red depending on the conditions of preparation and temperature). Diemer [104] used fields of the same order of magnitude and observed his crystals under a microscope. The electrical characteristics and the visual observations indicated strong similarity to a normal gas discharge; regions corresponding to a Townsend avalanche, to Meek streamers, and finally to an arc could be observed as the current was increased if breakdown was prevented by resistance ballasting. The emission was orange-red in color, showed slow build-up and decay, and was apparently thermal in nature (optical absorption measurements indicated a temperature of 600°C). Spots of green emission at the electrodes were also occasionally observed, although there was no influence of electrode material. Smith [105] has observed two different types of behavior in CdS. In crystals with very

[4] In a recent paper entitled "Radiative transitions in semi-conductors," *Phys. Rev.*, vol. 99, pp. 1892–1893; September 15, 1955, R. Braunstein discusses EL in GaSb, GaAs, InP, and Ge-Si alloys. At the Symposium on Luminescence, Polytech. Inst. of Brooklyn, September 9, 1955, G. A. Wolff, R. A. Hebert, and J. D. Broder also reported on emission from mixed crystals of GaP·InP and GaP·GaAs.

high conductivity, a red or straw-colored emission was observed at fields of only 150 volts/cm. Such fields were near thermal breakdown for these crystals, however, and this emission is therefore apparently similar to that observed by Diemer. In crystals of lower conductivity a green emission is observed at field strengths of roughly 10^3 volts/cm (probe measurements). The crystals which show the best EL have very poor photoconducting properties and must be "formed" by application of high voltage to produce EL sensitivity; this forming action apparently occurs at the anode electrode. The green emission may extend over the entire crystal but is most intense near the anode; the spectrum seems the same as the well-known "edge emission" of CdS, which is commonly accepted as arising from band-to-band transitions. The intensity of the green emission varies linearly with the current through the crystal. Prior to the onset of the emission the crystal operates under space-charge-limited conditions; the contacts to the crystal were ohmic. Spots of yellow emission were also observed at the anode before the appearance of the green emission.

The intrinsic EL of zinc oxide has been reported by Destriau [12]. This material is known to exist only with *n*-type conductivity; if prepared under proper conditions the resistivity can be quite low, as little as one ohm-cm [107]. The width of the forbidden band is 3.2 ev. Carrier-injection EL of ZnO was reported very early by Lossew [88], who stated that its output was very low compared to that of SiC. Fischer [108] used sintered samples of low conductivity and observed emission when a point contact was biased in the reverse direction; voltages as low as 2.5 volts were sufficient to give detectable ouput.

Carrier-injection EL has also been observed in several other materials. Fischer [108] mentions zinc silicate and several zinc-sulfide type phosphors. As early as 1934, Güntherschulze and Gerlach [109] found that the typical blue luminescence of calcium tungstate or the red emission of ruby (impure aluminum oxide) could be stimulated by application of direct voltages to the powdered material.

B. Theoretical Considerations

As discussed in a previous section, it is generally believed that in intrinsic EL the activator centers are emptied by impact of electrons in the conduction band. According to Piper and Williams [72] the only distinction between the two components of emission observed in ZnS single crystals in contact with electrodes is the origin of the electrons, the excitation and emission being the same in the two cases. The entire emission in ZnS, therefore, is attributed to this "collision" process (see Fig. 18). One feature of this process is the necessity of high electric field strengths so that the electrons may gain the required energy of several electron volts (above the bottom of the conduction band), either in the potential barrier itself or by acceleration in the conduction band.

Curie [44] and Lehovec and his co-workers [93] have suggested a similar collision mechanism for the Luminescence I observed in silicon carbide, and Fischer [108] has also suggested this mechanism for his observations on zinc oxide. As suggested above, the results of Newman and his collaborators [103] on reverse-biased silicon junctions also seems analogous to Luminescence I. McKay and his co-workers [110] also have correlated recently this emission in silicon with avalanche breakdown in such junctions. All these phenomena occur when rectifying junctions or contacts are biased in the reverse direction and the necessary high field strengths for the collision process are therefore available.

We must now consider EL observed when rectifying junctions or contacts are biased in the forward direction so that high field strengths are not developed. Lossew [89] explained his Luminescence II in silicon carbide as arising from the loss of energy by electrons as they are decelerated, and therefore similar to the "bremsstrahlung" which is responsible for the continuous X-ray spectrum; Claus [90] also accepted this theory. As pointed out by Tetzner [91], however, in this case the wavelength of the emitted radiation should shift toward shorter values as the voltage is increased, contrary to observation. Lehovec, Accardo, and Jamgochain [93] first proposed the presently accepted explanation that this emission is the result of recombination following injection of minority carriers. This process is illustrated in Fig. 21. It will be noted that in this mechanism it is not necessary that electrons gain energies above the bottom of the conduction band (compare Fig. 18). Once excess minority carriers have been injected, they can recombine with majority carriers without acquiring further energy from the field, and this mechanism will therefore be designated a "recombination" process to distinguish it from the "collision" process discussed above. It is obvious that in addition to injection over a *p-n* barrier, as shown in Fig. 21, which will introduce electrons into the *p*-type material and holes into the *n*-type material simultaneously, injection of a single kind of minority carrier can occur at appropriate metallic contacts.

The recombination of minority carriers with majority carriers may occur in a number of ways. First, they may recombine in a nonradiating manner; as is well known in semiconductor techniques such recombination may be greatly influenced by surface conditions and by impurity centers. Second, two methods of radiative recombination are possible: (1) by direct band-to-band transition (or perhaps through intermediate exciton states) and (2) through impurity (luminescence activator) states. In the case where the emission is characteristic of the host crystal lattice itself, the process is commonly referred to by workers in the field of luminescence as "edge emission," since it normally lies at the edge of

the optical absorption band. It is well known that the recombination of two free carriers is much facilitated by the presence of a third "body," which may be an impurity or the surface. For this reason activated emission normally predominates over edge emission.

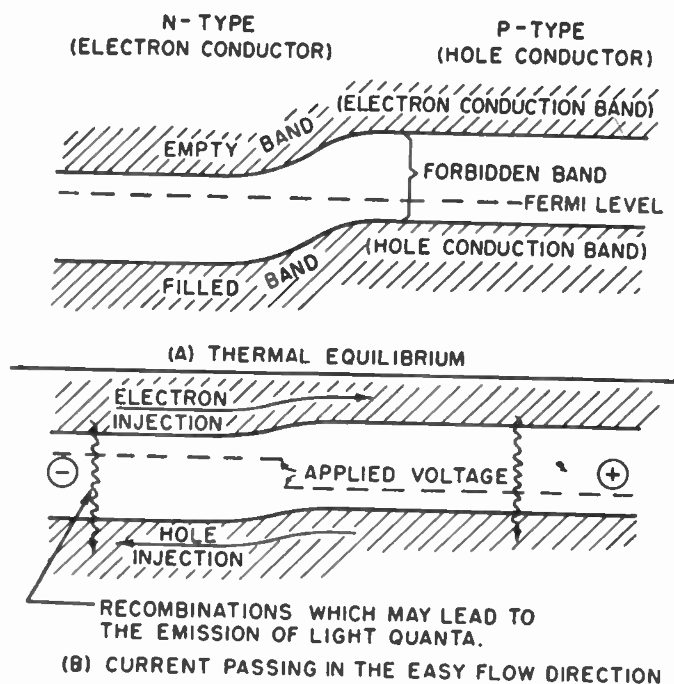


Fig. 21—Energy diagrams for a *p-n* barrier in the absence of an applied field (a) and for current flow in the forward direction with consequent injection of minority charge carriers (b). In addition to direct radiative recombination of holes and electrons as shown, recombination may also occur via luminescence activator centers or in a radiationless manner (from Lehovec, Accardo, and Jamgochian [93]).

It is commonly believed, because of the spectral distribution, that the EL recombination in silicon carbide occurs at an activator center. The exact nature of the activator, or activators, is not known, although Szigeti [94] suggests that it is excess carbon. Tetzner [91] reports that very pure SiC shows no EL. Schön [111] has extensively discussed the EL of silicon carbide along the lines proposed by Lehovec. In germanium and silicon, and in the case of the green emission of CdS observed by Smith [105], the emitted spectrum agrees well with the known optical absorption. It is natural, therefore, to interpret this as "edge emission" following minority carrier-injection, and indeed Haynes and Briggs [97] and Smith [105] did make this interpretation. In the case of CdS (*n*-type) hole injection occurs at the anode. The theory of recombination radiation has been considered by Van Roosbroeck and Shockley [112] and by Newman [100]. The low efficiency of this process, particularly at room temperature, arises from the competing radiationless transitions. Newman observed that the application of a magnetic field (20,000 oersteds) of appropriate direction to deflect carriers away from the germanium surface increased the efficiency by 30 per cent, while reversing the field so as to deflect carriers to the surface decreased the output by 40 per cent. In addition to the edge emission of germanium which has been discussed here, Aigrain [101] has also observed a far infra-red emission with a peak at wavelengths greater than 6 microns; this he attributes to recombination via traps rather than directly band-to-band.

Smith [105] attributed the yellow spots he observed at the anode of his CdS crystals before onset of the green emission to field emission. The reason for the difference in color, however, seems obscure. The orange-red emission observed by Smith in his high-conductivity crystals and by Diemer [104] apparently involves avalanche formation and collision processes in the bulk of the crystal, but as pointed out by Diemer the situation is made very complicated by the nearness to electrical breakdown and the high temperature involved. EL under such conditions will here be ascribed to a "breakdown" process.

All of the examples of EL discussed above may therefore be described as arising from one of three mechanisms: collision, recombination, or breakdown. Which type occurs in a certain material or a certain specimen of a given material may depend on a number of factors, such as the resistivity, the nature of the contacts, the direction of current passage, etc. For example, silicon and silicon carbide, and presumably other materials, exhibit collision EL when current flows in the reverse direction and recombination EL in the forward direction. High-conductivity cadmium sulfide exhibits emission of the breakdown type, while material of low conductivity shows recombination EL with ohmic contacts; there seems no reason why collision EL should not also be observed in this material with suitable rectifying electrodes and suitable resistivity. In general, emission can occur either through the medium of activators or by "edge emission"; the mechanism, and therefore, the color of the emission, may be different for the same material under different conditions. The collision process may also occur without carrier-injection (intrinsic EL).

Although older than intrinsic EL, carrier-injection EL has not thus far received the quantitative study accorded the other type. The variety of experimental phenomena possible and its close connection with normal semiconductor technique, however, make it a very interesting field for future research.

III. Electrophotoluminescence

A. The Momentary Illumination Produced in an Excited Phosphor by an Electric Field (Gudden and Pohl Effect)

1. General Introduction. When certain phosphors, principally of the zinc sulfide type, are excited (by violet or ultra-violet light, X-rays, or alpha particles) and then the exciting radiation is removed, a momentary flash of light can be obtained if an electric field is applied [6]. These materials may show negligible, or no, electroluminescence. If a constant field is applied, one can observe other flashes when the field is removed or

successively applied and removed many times. This behavior arises from the fact that the polarization field quickly cancels the internal field when the external field is maintained constant; when the external field is removed, the internal field rises suddently (polarization field) and decreases again quickly to zero.

In order to avoid such polarization effects, an alternating electric field may be used; this is, of course, equivalent to many successive field applications and removals. In this case, and with a constant effective field strength, one observes a noticeable flash only at the first field application, provided the field is applied a second time shortly after the first removal. In order to obtain a second flash at a subsequent field application, without other irradiation, the direction of the field must be changed, and this second flash is a maximum for a rotation of 90 degrees so that the second field direction is perpendicular to the first [12]. The second flash is zero if the rotation is 180 degrees.

On the other hand, a second flash occurs without any change in the field direction if the second field application follows the first one by an appropriate time. Then the emission at the second field application is a function of T, where T is the time between the two field applications; it is zero for $T=0$, passes through a maximum for a certain value of T, and then decreases towards zero again when T becomes longer and longer [12, 113].

The flash of momentary illumination can be observed not only during the after-glow as described above (after removal of the exciting beam) but also during the action of the exciting beam. In this case however, as will be shown later, the phenomenon can be disturbed by other effects such as quenching and enhancing effects. A similar flash of momentary illumination also occurs at field removal if irradiation by the exciting beam has been carried out in the presence of an electric field.

2. Theoretical Considerations. It seems convenient to ascribe this phenomenon to the emptying of electron traps either by electrons accelerated in the conduction band or by the direct action of the electric field on the traps; the effect is then similar to thermoluminescence. A direct action of the field on the energy structure of the phosphor, chiefly a disturbance of the impurity levels, is suggested by the observed spectral changes which will be described later. If these speculations are correct, thermoluminescence after prior field application should be weaker than the usual thermoluminescence without previous action of the field. As a matter of fact, this initial diminution is observed but one of the authors (Destriau) has found, surprisingly enough, with some phosphors two curves of thermoluminescence which cross (see Fig. 22).

It should be noted, however, that although thermoluminescence and even infra-red stimulation can occur after action of the field, application of a field after thermoluminescence or infra-red stimulation, on the other hand, does not produce any emission [12]. From this observation one might conclude that the field has an effect only on shallow traps; this conclusion is not certain, however, because the field obviously has a directional character, whereas temperature or infrared radiation do not. The directional action of the field is clearly demonstrated by the second flash obtained when the field is reapplied in a direction perpendicular to the original orientation. There are several possible theoretical explanations for the directional effect, but at the present time they must be regarded as purely tentative and will not, therefore, be discussed here. The flash of momentary illumination obtained at reapplication of the field, in the same direction, following a time of rest T after the first application, can be ascribed to electron exchange between shallow and deep traps according to a statistical mechanism as recently shown by D. Curie [114], or by electron redistribution over traps which would respond to fields oriented in different directions.

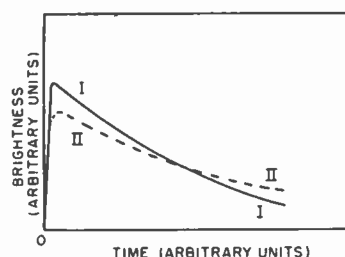


Fig. 22—Schematic representation of thermoluminescence, depending upon whether an electric field has been previously applied (curve 2, dashed line) or not (curve 1, full line). In both cases the phosphor has been brought immediately to a temperature T at time 0 and maintained at this temperature during the decay.

3. Dependence of the Light Sum on Field Strength, on Time, and on the Amount of Previous Excitation. Quantitatively the flash intensity is characterized by the total light energy liberated; in phosphor nomenclature the similar quantity in stimulation experiments is called the "light sum," and the same description will be used here. This light sum is lower when the field is applied a long interval of time after irradiation than if the field is applied immediately after irradiation. Furthermore, with some phosphors, chiefly certain mixtures of zinc and cadmium sulfides, the light sum obtained after two identical and successive excitations and field actions is not constant, but is lower at the second experiment than at the first; in this case a long time must be allowed between two successive experiments in order to obtain comparable results. The light sum also depends on voltage; the higher the applied voltage the greater the light sum.

The dependence of the light sum on the amount, r, of absorbed exciting radiation is more complicated. It first increases with r, passes through a maximum for a certain value, r_m, of r, and then decreases again. This behavior is shown in Fig. 23, where the excitation was made by X-rays and a copper X-ray filter, the thickness of which varied by steps, was placed against an EL cell.

Since the thickness of the filter was greater at the center than at the periphery, the exciting beam intensity falling upon the EL cell, and thus the energy absorbed by the phosphor per unit time, is greater near the periphery. After a short duration of irradiation, the flash obtained upon application of a field is greater near the periphery than near the center [Fig. 23(a)]. With a

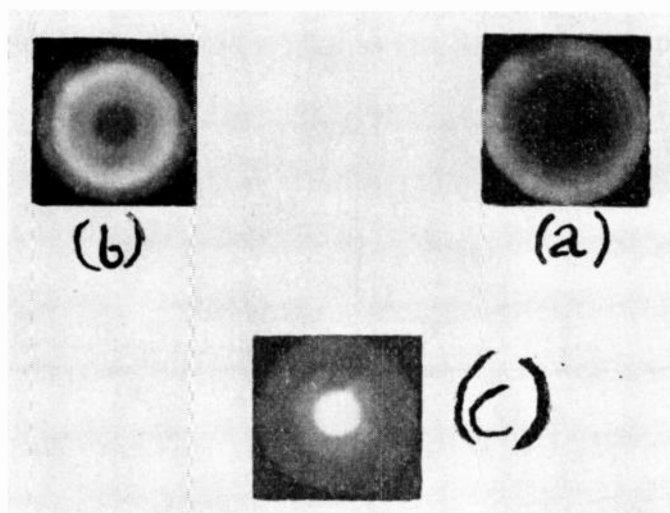


Fig. 23—Dependence of the light sum in the Gudden-Pohl effect on the amount of absorbed exciting X-radiation (stepped copper X-ray filter). (a) Previous irradiation of short duration. The flash is a maximum at the periphery under the thinnest step of the filter. (b) After longer irradiation the flash is a maximum under the intermediate step of the filter. (c) After very long previous irradiation the flash is a maximum under the central step, the one of greatest thickness.

greater duration of irradiation, the optimum intensity is reached under the second step where the flash is then maximum [Fig. 23(b)]. Lastly, if the duration of excitation is further increased, the optimum intensity is reached at the central step where the flash is then maximum [Fig. 23(c)]. Such a technique, which is able to emphasize a certain thickness of material, has been called "electrophotoradiography" [12].

4. Dependence of the Light Sum on Temperature. Experiments on temperature effects have been made only in a small range between 15 and 80°C. [115, 116]. The value, r_m, of the amount of exciting radiation which corresponds to the maximum light sum decreases linearly when the temperature is increased, and the light sum obtained for $r=r_m$ first increases with the temperature up to about 50°C (for the particular specimen studied) and then decreases for higher temperatures (Fig. 24).

Similar observations can be made for each of the flashes one obtains at application of the field (when irradiation is made without the field) or at removal of the field (when irradiation is made with the field). The maximum, however, is not reached at the same temperature in the two cases. Krautz [116a] has made measurements on both kinds of stimulation for several phosphors at the temperature of liquid nitrogen.

5. Dependence of the Light Sum on Frequency of the Applied Field. Nudelman and Matossi [117], using ZnS phosphors excited by continuous ultra-violet radiation, have recently shown that if the frequency is increased, the momentary flash at the application of the field first decreases until a frequency of about 20 kcps is reached, and then either remains constant at higher frequency or increases slightly up to 100 kcps. The momentary flash obtained at removal of the field increases with frequency over this entire range.

6. Quenching of the Light Sum by Irradiation with Long Wavelength Light. It has already been reported [12] that the flash of momentary illumination is quenched by exposure of the phosphor to infra-red irradiation before or during the excitation. This effect has recently been studied more accurately by Vigean [118]. She first showed that the maximum quenching was obtained with yellow light. Let S_0 be the light sum obtained with no yellow irradiation, and S_1 be the light sum with identical exciting irradiation and identical field strength but with an additional irradiation by the yellow light of sodium. The ratio $y=S_1/S_0$ gives, quantitatively, the quenching effect of the yellow light.

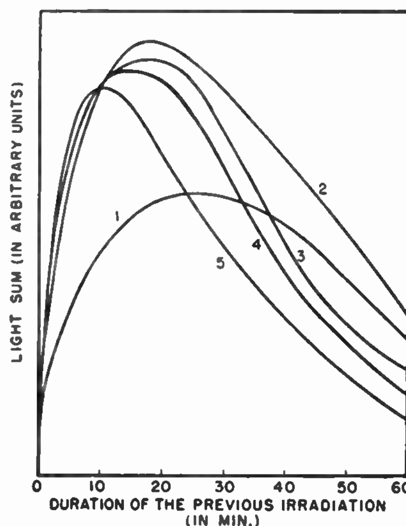


Fig. 24—Dependence of the light sum in the Gudden-Pohl effect on the duration of the previous irradiation at different temperatures. Curve 1, 20°C; curve 2, 36°C; curve 3, 48°C; curve 4, 52°C; curve 5, 61°C.

If this irradiation with yellow light is made before the exciting irradiation (X-rays or ultra-violet radiation) and the field is applied at a fixed time after excitation, the ratio y assumes a constant value, y_1, independent of the time when yellow light is applied, even if it is many hours before the excitation. If the irradiation with yellow light is made either during or after the exciting irradiation, however, the ratio y decreases, first quickly and then more slowly, the shorter the time between the application of the yellow light and the field.

The quenching action of long wavelength radiation on phosphorescence is well known but the effect of such

radiations on the flash of momentary illumination is greater than that on phosphorescence. For the same conditions of time, exciting beam intensity, and field strength, the light sum (S_1) emitted in the flash is a decreasing exponential function of the amount L of yellow light received by the phosphor,

$$S_1 = S_0 \exp(-kL).$$

Vigean and D. Curie [119] have suggested that this quenching can be ascribed to a mechanism involving two processes: (1) photodisturbance of sulfur atoms in the neighborhood of luminescent centers, as indicated by M. Curie [120] and by Garlick and Mason [121], and (2) upon application of the field, electrons from these disturbed atoms may fill some of the empty luminescent centers without radiation. D. Curie also points out that another process might be considered; this is emptying by the yellow radiation of the filled traps responsible for the Gudden-Phol phenomena. Such an explanation could apply only if the ultra-violet excitation preceded the yellow irradiation, and not in the reverse situation.

7. Brightness Waveforms. Some recordings of brightness waves of the momentary illumination are shown in Fig. 25. The emission lasts for several cycles of the alternating field, and the amplitudes of the successive maxima decrease quickly. As in the observations on intrinsic EL, a peak occurs in each half-cycle near the time when the field is maximum, and at low frequency the brightness wave leads the potential wave, while the opposite occurs at high frequency (higher than 2,000 cps) [122].

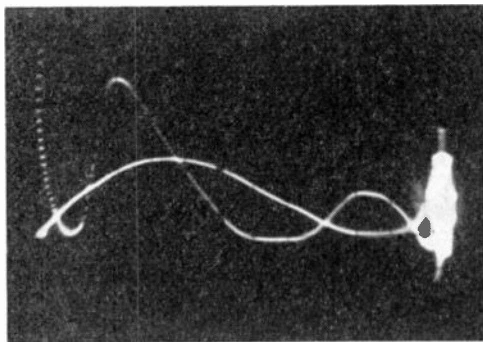


Fig. 25—Brightness waves in the Gudden-Pohl effect. The phase difference between the potential wave and the brightness wave decreases progressively in the successive half cycles.

We must also consider that in this phenomenon the shift of brightness wave compared to the sinusoidal potential wave is greater for the first peak than for the second, and that this shift decreases quickly in the following peaks. This fact must be ascribed to the conductivity of the phosphor, which is determined by the number of free electrons, and is greater at the time of the first half-cycle of field variation than in those following [122].

In some cases, similar to observations made in intrinsic EL, alternate peaks of the brightness wave are of different magnitude; the successive odd peaks and the successive even peaks fall upon two different decreasing curves (see Fig. 26). Such behavior must be ascribed to a polarity effect [122].

8. Emission Spectrum. The spectral band emitted at the flash of momentary illumination is different from the spectrum of phosphorescence. Because of the short duration of the flash, only rough experiments have been made with optical filters; in some cases the disturbance is great enough to be noticeable to the eye [123].

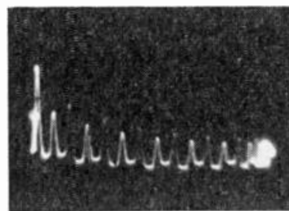


Fig. 26—Polarity effect in the Gudden-Pohl flash. Odd peaks and even peaks decrease progressively, following two different curves.

B. Quenching or Enhancing Effects of Electric Fields on Photoluminescence

1. General Introduction. When certain phosphors, irradiated by X-rays or by ultra-violet radiation, are simultaneously subjected to an electric field, the luminescence may be either quenched or enhanced, depending upon the phosphor. The quenching effect was first observed by G. Destriau [8]; the enhancing effect with X-ray excitation was first observed by G. Destriau and M. Destriau [9, 124], while that for ultra-violet excitation was first obtained by Cusano [125].

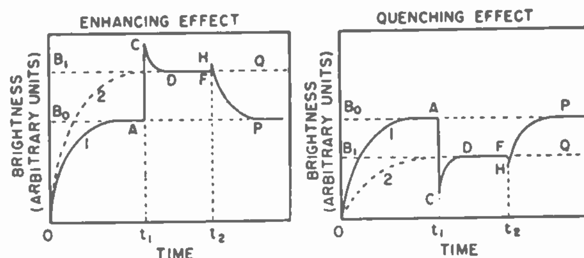


Fig. 27—General behavior of the quenching and enhancing effects of an electric field. The exciting radiation is applied at time zero, while the field is applied at t_1 and removed at t_2.

Fig. 27 shows schematically the variation of brightness with time; the brightness increases according to curve 1 (OAP) without a field and according to curve 2 ($ODFQ$) if a field is applied. Let B_0 be the equilibrium brightness reached at the plateau of curve 1 and B_1 be the plateau reached by curve 2. The ratio $z = B_1/B_0$ is either smaller than unity (quenching effect) or greater than unity (enhancing effect).

If the field is applied only between times t_1 and t_2, the brightness first raises progressively, following OA on curve 1; as soon as the field is applied (time t_1) the

brightness immediately falls (quenching effect) or rises (enhancing effect) from A to C and then it varies progressively, following CO towards the plateau of curve 2. At time t_2, when the field is removed, the brightness falls, or rises, again immediately to FH and then it varies progressively, following HP towards the plateau of curve 1.

In both cases, i.e., for either the quenching or the enhancing effect, the time of field application as well as the time of field removal are followed by a transient period of relatively short duration in which the brightness changes progressively. With some phosphors, which show simultaneously a Gudden and Pohl effect, the curves of Fig. 27 are disturbed during the short transient periods at times t_1 and t_2. On the other hand, quenching is also observed, but with a smaller intensity, when the field is applied and removed a certain time before irradiation. In this last case, the brightness during the first minutes of irradiation is lower than if no field had been applied [126].

2. Magnitude of the Quenching or Enhancing Ratio and Its Dependence on Field Strength. In the earliest experiments [9, 124], using powders, the enhancing effect was found in manganese-activated sulfide phosphors excited by X-rays; such materials under ultra-violet excitation, and all other phosphors tried under either type of excitation, showed only the quenching effect. Since the phosphor powder was embedded in a dielectric, it was necessary to use alternating potentials. With phosphors showing the quenching effect, the higher the electric field strength the higher is the quenching ratio. With phosphors showing the enhancing effect, however, as the field strength is increased the ratio z first increases and then approaches a plateau for some samples or passes through a very flat maximum for some others. Values of the enhancing ratio as high as 4.5 have been obtained; the quenching ratio may approach zero.

Recently, with a different experimental arrangement (electrodes in contact with a thin transparent ZnS: Mn, Cl film) and direct potentials, Cusano [125] has found an enhancing effect in which the ratio z can reach values as high as 90. In these experiments, the brightness is enhanced regardless of the type of excitation (X-rays or ultra-violet radiation). It is found that the enhancement ratio varies inversely as the square root of the exciting intensity.[5] For low exciting intensity 10 visible quanta are released for each incident ultra-violet quantum, so that despite the wavelength conversion there is still an increase in radiant energy; i.e., true "amplification"; this is, of course, not true for X-ray excitation because of the high energy of the incident quanta. Above a certain threshold field strength (about 1×10^4 volt/cm) the output for constant excitation intensity varies linearly with the applied field (up to 1×10^5 volt/cm). It may be noted that these films show negligible electroluminescence (for either ac or dc) and that the enhancement shows a polarity effect. The time constant for changes in the electric field is of the order of milliseconds, but is a few seconds for changes in the intensity of the exciting radiation.

[5] Matossi [161] has recently commented on this point.

3. Dependence of the Quenching or Enhancing Ratio on Temperature. It is well known that the intensity of fluorescence, B_0, is lowered when the temperature is raised. A similar behavior occurs for the fluorescence, B_1, when the phosphor is simultaneously subjected to an electrical field. The ratio z depends on temperature through both quantities B_0 and B_1. Some experiments have been described [116, 127] which show a strong diminution of the quenching ratio when the temperature is increased. With some phosphors the brightness when the field is applied falls nearly to zero for temperatures above 90°C. Experimentally it is found that curves giving the expression $(z^{-1}-1)$ is a function of the field strength, for a constant value of the temperature, can be superimposed by a simple change in the scale of ordinates (see Fig. 28).

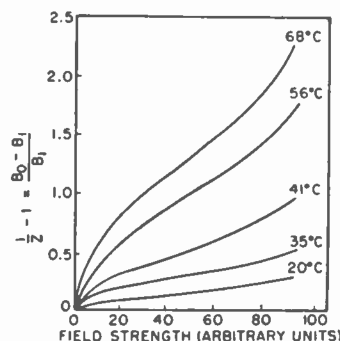


Fig. 28—Field dependence of $(z^{-1}-1)$ at various temperatures.

In contrast with the above results, the enhancing ratio is not dependent on temperature, at least within the range of 20 to 60°C [124].

4. Dependence of the Quenching or Enhancing Ratio on Frequency. The enhancing ratio is not dependent on the frequency of the applied alternating field [124], but the behavior of the quenching effect is different according to the frequency employed.

Matossi and Nudelman [128] have found that the first drop (at the application of the field) is greater at high frequency than at low frequency. It is the opposite, however, for the second drop (at removal of the field) and this second drop is even replaced by a flash of momentary illumination for elevated frequencies. Similar results have been indicated by Olson [129], who observed chiefly the action of the field upon the afterglow of a phosphorescent material. The quenching ratio increases with the frequency as does also the sudden brightness increase at field removal. Studies of electrical quenching have been made with frequencies up to 50 mcps by Miller [130]; quenching ratios as small as 0.5 were obtained and in some cases peaks were observed in the curves above 10 mcps.

5. *Spectral Changes in the Quenching and Enhancing Effects.* Spectral disturbances occur in the light emission when the field acts so that the quenching ratio [123] as well as enhancing ratio [131] are different according to the spectral region observed. With some phosphors the spectral change is great enough to be noticeable by direct visual observation. Sometimes, but rarely, a strong red flash occurs either (depending on the phosphor) at field application or at field removal. As a matter of fact, with such samples, the light emission is simultaneously quenched in a certain spectral region but enhanced in another.

For example, a phosphor prepared by Levy and West in 1937 for fluoroscopic screen applications (silver-activated ZnS with small additions of nickel) gives a noticeable red flash only at the field removal (see Fig. 29). On the other hand, some ZnS-CdS mixtures activated with silver and manganese show strong permanent spectral changes at field application [131]. However it is not known at the present time whether this results from a spectral change in a single band or from a different sensitivity to the electric field of two different and overlapping spectral bands.

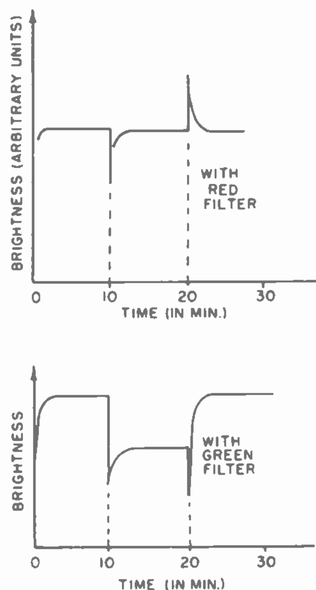


Fig. 29—Spectral disturbances in the quenching effect. Curves obtained with the same phosphor (Levy-West). The field was applied only between times 10 and 20 minutes after the beginning of irradiation. Note the difference in the quenching ratio when observed through red and green optical filters, and particularly the difference at removal of the field.

6. *Brightness Waves and Transient Disturbances at Application of the Field.* One must distinguish whether the irradiation is made by an exciting beam of constant intensity (X-ray generator or ultra-violet lamp operating on dc) or by an exciting beam giving pulses of light (X-ray generator or ultra-violet lamp operating on ac). In the latter case the brightness waves for the quenching effect are greatly perturbed, chiefly by a sharp secondary maximum which occurs near the time when the potential passes through zero in each half-cycle. This sharp maximum increases and then decreases quickly during the first few cycles which immediately follow the field application; during the same time the chief maximum is first lowered and then it increases again. After a certain time the brightness wave remains in a constant disturbed form [132] (see Fig. 30). In the enhancing effect of CdSZnS:Mn only a few of the brightness waves are disturbed; the sharp and momentary peak is very small and, for some samples, does not even occur, so that at the field application one observes only an increase in the height of the main peak [9, 124].

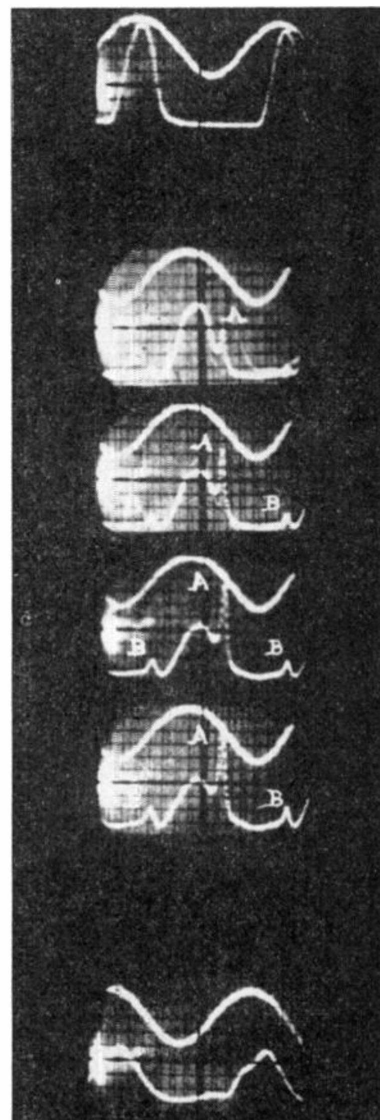


Fig. 30—Rapid change of the brightness waves in the quenching effect immediately after the field is applied. (a) Sinusoidal field wave and brightness wave excited by a half-wave X-ray generator (no field on the cell). (b) At application of the field, small secondary peaks occur at *A* and *B* while the main peak is lowered. (c), (d), and (e) show rapid changes of the sharp secondary peaks: (c) 0.06 second after field application, (d) 0.12 second after field application, (e) 0.18 second after field application, (f) brightness wave 2 seconds after field application.

When a direct potential is applied to the X-ray tube or to the ultra-violet lamp the light emission of the

phosphor is itself of a constant intensity; when an ac potential is then applied to the EL cell the light is modulated with a frequency twice the frequency of the field. The wave giving the light modulation changes during the first few cycles after application of the field and then tends towards a permanent form [133].

Steinberger, Low, and Alexander [134] have recently studied such disturbances in a more accurate manner. They consider separately pulses (designated A) which occur when the transparent electrode through which the irradiation is made is positive and pulses (designated B) which occur when this transparent electrode is negative. Type A pulses do not depend on the phosphor, but type B pulses do. The curves giving the changes in time of pulses A and B cross so that B is higher than A during the first cycles following the field application, but the opposite becomes true after a certain time. Lastly, when the field strength is raised the amplitude of the pulses first increases, reaches a maximum for a certain field strength, and then decreases when the field becomes still higher.

7. *Theoretical Considerations.* In order to explain the behavior of the quenching effect Matossi [135] has presented an extensive mathematical analysis which includes two assumed actions of the applied electric field. These are: (1) emptying of electron traps by the field (by a mechanism not specified in detail) and (2) a loss of electrons due to field-induced nonradiative transitions. The exact mechanism for the second effect is also not specified, but it is suggested that a possible process is migration of electrons to the surface under the action of the field, with resultant nonradiative combination at surface states or impurity levels. This analysis predicts results which seem in general accord with experiment. It does not however, explain the fact that for the same material one may obtain a quenching effect for ultra-violet excitation and an enhancing effect for X-ray excitation. In a study of the brightness waves of an EL phosphor with superimposed ultra-violet excitation, Matossi and Nudelman [117] have shown that the quenching and EL effects have different origins.

The theoretical basis for the light amplification effects observed by Cusano [125] has been presented by Williams [136], and is similar to that proposed by Piper and Williams [39] for EL and discussed above. Electrons liberated by incident radiation are accelerated by the high electric field existing in a potential barrier region in the same manner that electrons liberated from donor centers by the field are accelerated in EL; the collision excitation and emission process is the same in the two cases. The polarity effect observed is indicative, of course, of the presence of a potential barrier. Cusano also observed a 100A shift of the emission spectrum toward shorter wavelengths as the applied field is increased, and this is interpreted as a kind of Stark effect arising from the high barrier fields. Williams [136] also suggests the possibility of amplification effects in minority carrier-injection EL by the action of space-charge effects. As pointed out by von Hippel and his co-workers [137] the space charge produced by optically liberated carriers may also produce field emission of electrons from the electrodes; such a mechanism might also lead to light amplification, although no work along these lines has been reported.

One outstanding fact which is as yet unexplained is that enhancement effects have been found thus far only in manganese-activated sulfide phosphors; this is equally true for transparent films as for phosphor powders.[6]

C. Effect of Electric Fields on the Quenching or Stimulating Action of Infra-Red Radiation

It has been shown by Destriau [10] that with some phosphors the well-known quenching effect of infra-red radiation upon phosphorescence is increased when the phosphor is subjected to an alternating field during the excitation and infra-red irradiation. No other experiments have been made on this phenomenon since the earlier review paper [12]. In order to summarize briefly these first results, let: B_0 be the brightness obtained at a certain time t during the after-glow after excitation, B_1 be the brightness at the same time if infra-red irradiation is applied simultaneously with the exciting beam, B_2 be the brightness at the same time t if an electric field is applied during excitation and maintained during the after-glow, and B_3 be the brightness reached, also at the same time t, when the above infra-red irradiation as well as the above field action are superimposed.

The quenching ratio for the infra-red irradiation is

$$z_{IR} = B_1/B_0.$$

The quenching ratio for the electric field is

$$z_F = B_2/B_0.$$

Experiments have shown that ratio B_3/B_0 is smaller than the product $z_F \times z_{IR}$.

Many phosphors, such as some of the doubly-activated alkaline earth sulfides, show the phenomenon of infra-red stimulation. To demonstrate this effect the material is first excited (by ultra-violet light, for example) and then the emission allowed to decay after the exciting source is removed; if then infra-red radiation is allowed to fall on the phosphor a burst of visible emission is observed. Low, Steinberger, and Braun [11] have shown recently that with some phosphors (for exple, SrS:Eu, Sm) this stimulation is enhanced considerably if the phosphors are subjected to an alternating electric field simultaneously with the ultra-violet excitation. The ratio between the initial brightness in the two cases (with and without field during excitation) may be as high as 1.5.

One of the interesting aspects of this new phenomenon is the memory of the phosphors. Even several hours after

[6] Woods and Wright [162] have recently observed field enhancement of the cathodoluminescence of magnesium oxide.

the simultaneous action of field and ultra-violet radiation, the infra-red stimulated luminescence remains stronger than that obtained following normal excitation by ultra-violet radiation without field. Such a memory, pointed out also by Vigean [118] in the quenching by infra-red or yellow radiation of the Gudden and Pohl effect, seems to be a specific aspect of all infra-red action on luminescence.

IV. Applications of Electroluminescence and Electrophotoluminescence

A. Electroluminescent Lamps

Electroluminescence is obviously of great interest from the point of view of application to light sources. In devices which are now commercially available for the production of light by means of luminescence, the phosphor is excited by ultra-violet photons or by electrons which must themselves be produced by a gas discharge or by thermionic emission and subsequent acceleration in an electric field. In an EL device, however, the light would be generated by the direct action of the field on the phosphor, with no intermediate processes. The first suggestion of an EL lamp was apparently due to Bay and Szigeti [138] in 1939. They suggested use of the carrier-injection EL of silicon carbide or similar materials. Sherwood and Skinner [139] have also described a similar lamp.

The possibility of a practical light source by means of intrinsic EL was mentioned by Destriau [12] as early as 1947, at which time he also described EL cells in which the phosphor is embedded in a plastic dielectric rather than a liquid. The first practical interest in such light sources was that of Payne, Mager, and Jerome [140]; more recent work has been reported by several workers [50, 141–143]. One feature of EL lamps which is of great importance is the fact that they offer the possibility of a large area source of uniform illumination; sources of other geometry have been described, however [144, 145]. An additional advantage is that the intensity may be continuously controlled by varying the applied voltage; with sources involving gas discharges this is not as conveniently possible.

The four primary requirements of a light source for purposes of general illumination are (1) high brightness, (2) reasonable luminous efficiency, (3) satisfactory color rendition, and (4) adequate maintenance of these characteristics during life. EL phosphors are available in a wide variety of colors so that there is no apparent difficulty in producing light with a desired spectral distribution by blending as is done today with phosphors for fluorescent lamps and television picture tubes. Since the EL emission color generally depends on frequency (and also on temperature and sometimes on the applied voltage) it would be necessary to design such blends for operation at specified conditions, but for illumination purposes this would be of no great consequence.

Although presently available commercial EL lamps give brightnesses of only five or ten foot lamberts when operated at 60 cps [142, 143], higher output can be obtained, of course, by increasing the frequency. The highest figure reported in the literature is about 180 foot lamberts (at 4,000 cps) [144], but values of more than 1,700 foot lamberts may be obtained in the laboratory at a frequency of 20,000 cps [146]. At elevated frequencies, therefore, ample brightness is available for lighting a room by means of a luminous ceiling, which would require no more than 100 foot lamberts.

The brightness of EL lamps is observed to decrease during life of the lamp [20, 142, 143], although it may show an increase during the first few hours of operation. Plastic-embedded lamps may retaion 90 per cent of their initial (or peak) brightness after 1,000 hours of operation and 75 per cent after 4,000 hours [142, 143]; corresponding figures for ordinary fluorescent lamps are 95 per cent and 85 per cent (and 82 per cent at 8,000 hours). Maintenance of output is largely conditioned by the embedding material and particularly its interaction with the phosphor; acid-neutralizing additions to the plastic were found by Mager [141] to improve the maintenance. EL cells must also be protected from the action of atmospheric moisture; Jenkins [145] reports that hermetic sealing in an inert atmosphere gives 90 per cent maintenance at 3,000 hours. Rulon [147] has described EL lamps in which the phosphor powder is embedded in an inorganic material (glass or enamel) rather than an organic material. The maintenance of such lamps is reported to be essentially the same as for plastic-embedded lamps. These lamps are made with a thick metal backing and are quite rugged mechanically and simple to produce, even in complicated shapes, by techniques standard in the porcelain enameling industry. Bramley and Rosenthal [26] also described earlier an EL cell consisting of a silicate phosphor embedded in glass.

The efficiency of EL lamps at first increases as the voltage is increased, but then passes through a maximum and decreases again for higher voltages. The maximum luminous efficiency reported is between 4 and 5 lumens/watt [142]. This value is to be compared with about 16 lumens/watt for incandescent lamps or 65 lumens/watt for ordinary fluorescent lighting. The figure for EL quoted above, moreover, is not for a source blended to produce white light but rather for a greenish emission of higher luminous efficiency than a white source. Even under optimum conditions, therefore, the EL efficiencies now available leave much to be desired. Furthermore, even these efficiencies cannot be obtained simultaneously with maximum brightness. For these reasons, EL lamps are not applicable at the present for general illumination purposes but only for special or novelty applications. They are suitable for clock faces, radio dials, aircraft instrument lighting, house numbers, photographic dark room illumination, etc. Since EL lamps are essentially capactivie devices, they have a very small power factor. This is generally disadvantage-

ous, although it is conceivable that such a feature might be used to advantage for power factor correction in industrial applications, should this ever become feasible.

Although they are not now applicable for general lighting purposes, EL lamps are still in their infancy, having been introduced only five years ago, and it is still too early to predict their eventual development. The EL lamps of the future may not employ intrinsic EL as do the present designs, but rather carrier-injection EL. Low efficiency now seems to be the major impediment and this may be a difficulty inherent in intrinsic EL; it must be recalled, nevertheless, that although good efficiency may theoretically be possible with carrier-injection EL, it has not yet been demonstrated experimentally. Lehovec [148] has shown, however, how "graded-seal" *p-n* junctions might be constructed from two different materials so that light emitted at the junction would not be absorbed in the material through which it has to pass before emerging even though it were generated initially by "edge emission." If the emission is achieved via a luminescence activator center, on the other hand, then such self-absorption presents no problem. Halsted and Koller [39, 149] have described EL cells employing transparent phosphor films rather than powers. The problems associated with embedding materials are obviously avoided in this construction; no practical performance data is yet available, however.

B. Other Applications

Very early Lossew [88], noting the rapid response of the EL of silicon carbide to changes in voltage, suggested its use, in conjunction with a moving photographic film to record the output, as an oscillograph. This proposal, of course, antedated the modern cathode-ray oscilloscope. Schwertz, Haller, and Mazenko [150] have used EL, in conjunction with xerography, to produce self-luminous halftone reproductions. Dietz and Jordan [151] suggest the use of EL in an indicating electrical fuse.

Piper [1952] has proposed the use of EL in an essentially two-dimensional television display device which would occupy much less space, particularly for large-size pictures, than the conventional cathode-ray tube. If some method of information storage can be incorporated into such a scheme, currently available brightnesses are adequate, although eff'ciency might still present a problem. The greatest difficulty with this method at present probably lies in the necessary signal distribution system to replace the scanning mechanism of the cathode-ray tube.

The use of the Gudden-Pohl effect in "electrophotoradiography" has been mentioned in a previous section. Low [153] has suggested the application of an electric field to enhance the output of scintillation counters for radiation detection.

Bramley and Rosenthal [154], commenting on the rapid rise in EL brightness with increasing voltage, suggested its application as an overvoltage indicator. They also pointed out that if a photoconductor is connected in series with the EL cell and the voltage supply and so located that the emission of the EL cell falls upon it, then a "lock-in" action will be obtained and the emission will continue even after the overvoltage has been removed. Marshall, Schwertz, and Bowman [155] have shown that if a parallel resistor-capacitor combination is placed in series with such an EL cell-photoconductor arrangement, then a bistable circuit should be obtained which will be responsive to external pulses of light. These workers suggest application as an information storage device in digital computer systems: apparently, however, a working model has not yet been constructed.

If provision is made to prevent optical feedback from the EL cell to the photoconductor in the arrangement described above, then lock-in is not obtained, but light amplification results; the intensity of the EL output will be a continuous function of the intensity of the incident radiation. The thickness or electrical characteristics of the photoconducting and EL materials must be so chosen that with no incident radiation only a small fraction of the applied voltage appears across the phosphor. The excitant need not be visible, but rather radiation of any wavelength to which the photoconductor is sensitive. If a multiplicity of such elements is arranged in a planar array, then an image intensifier or converter is obtained [156–160]. Orthuber and Ullery [157] reported a brightness gain of 24 for a single-element device, but were able to obtain only a value of 2 for a multi-element arrangement. Gains as high as 50 have since been reported, however, by Kazan and Nicoll [158].

It is also possible, of course, to make a light intensifier or amplifier by utilizing field enhancement of photoluminescence (electrophotoluminescence) instead of electroluminescence. Destriau early suggested the application of this effect to X-ray fluoroscopic or intensifying screens; even with the phosphor powders which show the greatest enhancement, however, the resultant brightness is still below that of commercial X-ray screens because of the low photoluminescent efficiency of these materials. The use of transparent phosphor films in image amplifiers, as described by Cusano [125], offers many possible applications to television and imaging devices in general. The maximum brightness obtained thus far seems to be about 4 foot lamberts, with an efficiency of 0.2 lumen/watt. Both these figures and the optical time constant of the device will have to be improved before extensive use can be made of such amplifiers. Here again, as in the case of EL, one is concerned with a very new development and it is difficult at the present time to predict future improvements and possibilities.

In conclusion it may be said that just as solid-state devices such as the transistor have caused a revolution in the field of electrical amplification and control, so

too the direct generation and control of light in solids by electric fields may in the near future create a revolution in the field of luminescence and related subjects.

ACKNOWLEDGMENT

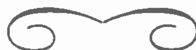
The authors are indebted to many persons for discussions and data which have been helpful in preparing the present paper or in evolving many of the ideas there expressed. Among these are Drs. J. Mattler and G. and D. Curie of the Faculté des Sciences, Paris, and Messrs. B. T. Howard, W. Lehmann, and C. H. Haake of the Research Department, Lamp Division, Westinghouse Electric Corporation.

BIBLIOGRAPHY

[1] Braun, F., "Luminosity of Immersed Electrodes." *Annalen der. Physik*, Vol. 65 (1898), p. 361.
[2] Dufford, R. T., "Luminescence Associated with Electrolysis." *Journal of the Optical Society of America*, Vol. 18 (1929), p. 17.
[3] Lossew, O. W., *Telegrafia i Telefonia*, No. 18 (1923), p. 61; No. 26 (1924), p. 403, No. 44 (1927), p. 485; and No. 53 (1929), p. 153; "Oscillating Crystals." *Wireless World*, Vol. 15 (1924), p. 93; *Zeitschrift fur Fernmeldetech*, No. 7 (1926), p. 97.
[4] Dechene, G., "Modification of the Brightness of Phosphorescent Zinc Sulfide under the Influence of an Electric Current." *Comptes Rendus*, Vol. 201 (1935), p. 139; *Journal de Physique et le Radium*, Vol. 9 (1938), p. 109.
[5] Coustal, R., "Action of Silent Discharges on Phosphorescent Sulphides." *Comptes Rendus*, Vol. 198 (1934), pp. 1403, 1596.
[6] Gudden, B., and Pohl, R. W., "Enhancement of Phosphorescence by Electric Fields." *Zeitschrift für Physik.*, Vol. 2 (1920) p. 192.
[7] Destriau, G., "Experimental Studies on the Action of an Electric Field on Phosphorescent Sulfides." *Journal de Chimie Physique*, Vol. 33 (1936), p. 620, Vol. 34 (1937), p. 117.
[8] Destriau, G. "Action of an Electric Field on the Luminescence Excited by Radiation of Short Wavelength." *Journal de Physique et le Radium*, Vol. 4 (1943), p. 32.
[9] Destriau, G., and Destriau, M., "The Increase of Sensitivity to X-Rays of Some Mixtures of ZnS and CdS under the Action of Electric Fields." Meeting of the Electrochemical Society, Chicago (May 2–6, 1954).
M. Destriau, *Comptes Rendus*, Vol. 238 (1954), p. 2298.
[10] Destriau, G., "Disturbance of Electrophotoluminescent Phenomena by Infra-Red Radiation." *Journal de Physique et le Radium*, Vol. 4 (1943), p. 77.
[11] Low, W., Steinberger, J. T., and Braun, E. A., "Effect of Alternating Electric Fields on the Excitation of a Strontium Sulfide Phosphor." *Journal of the Optical Society of America*, Vol. 44 (1954), p. 504.
[12] Destriau, G., "The New Phenomenon of Electrophotoluminescence and Its Possibilities for the Investigation of Crystal Lattices." *Philosophical Mag.*, Vol. 38 (1947), pp. 700, 774,880.
[13] Wagner, K. W., "Explanation of the Dielectric After-Effect According to Maxwell's Theory." *Archiv für Elektrotechnik*, Vol. 2 (1914), p. 371.
[14] Roberts, S., "Field Strength and Temperature Studies of Electroluminescent Powders in Dielectric Media." *Journal of the Optical Society of America*, Vol. 42 (1952), p. 850.
[15] Pearce, C. A. R., "The Mechanism of the Resolution of Water-Oil Emulsions by Electrical Treatment." *British Journal of Applied Physics*, Vol. 5 (1954), p. 136.
[16] Lehmann, W., "Dielectric Behavior of Electroluminescent Zinc Sulfide." Meeting of the Electrochemical Society, Cincinnati (May 3, 1955). To appear in *Journal of the Electrochemical Society*, January, 1956.
[17] Waymouth, J. F., and Bitter, F., "Experiments on Electroluminescence." *Physical Review*, Vol. 95 (1954), p. 941.
[18] Nudelman, S., and Matossi, F., "Electroluminescence with Nonsinusoidal Fields." *Journal of the Electrochemical Society*, Vol. 101 (1954), p. 546.
[18a] Matossi, F., and Nudelman, S., "Electroluminescence Excited by Short Field Pulses." *Physical Review*, Vol. 99 (1955), p. 1100.
[19] Destriau, G., *Comptes Rendus* (in press).
[20] Zalm, P., Diemer, G., and Klasens, H. A., "Electroluminescent Zinc Sulfide Phosphors." *Philips Res. Reps.*, Vol. 9 (1954) p. 81.
[20a] Diemer, G., "Light Patterns in Electroluminescent ZnS Single Crystals Activated by Diffusion of Cu." *Philips Research Reports*, Vol. 10 (1955), p. 194.
[21] Herwelly, A., "On the Luminescence of Phosphors in Strong Electric fields." *Acta Physica Austriaca*, Vol. 5 (1951), p. 30.
[22] Destriau, G., "Excitation of Luminescence by Varying Electric Fields. A Primary Effect." *Journal de Physique et le Radium*, Vol. 14 (1953), p. 307.
[23] Böer, K. M., and Kümmel, U., "On the Luminescence of Single Crystals of CdS in Strong Static Electric Fields," *Zeitschrift für Physikalische Chemie*, Vol. 200 (1952), p. 193.
[24] Wendel, G., "Proof of Electroluminescence with the Aid of Multilayer Color Film." *Annalen der Physik*, Vol. 12 (1953), p. 222.
[25] Gumlich, H. E., Diplomarbeit, Technische Universität, Berlin-Charlottenburg (September 8, 1953), and Gobrecht, H., Hahn, D., and Gumlich, H. E., "The Electroluminescence of Various Phosphors and its Dependence on the Strength and Frequency of the Alternating Electric Field." *Zeitschrift für Physik*, Vol. 136 (1954), p. 612.
[26] Bramley, A., and Rosenthal, J. E., "Field Dependent Fluorescence of Vitreous Zinc Silicate Phosphor." Report of the 12th Massachusetts Institute of Technology Conference on Physical Electronics, 1952, and *Physical Review*, Vol. 87 (1952), p. 1125.
[27] Ueta, M., "Field Emission of Crystal Phosphors." *Journal of the Physical Society of Japan*, Vol. 8 (1953), p. 429.
[28] Luyckx, A., and Stokkink, A. J., "Some Experimental Results on Electroluminescence." *British Journal of Applied Physics*, Supplement No. 4 (1955), p. 57.
[29] Burns, L., "Electroluminescence of Insulated Particles." *Journal of the Electrochemical Society*, Vol. 100 (1953), p. 572.
[30] Homer, H. H., Rulon, R. M., and Butler, K. H., "Electroluminescent Zinc Sulfide Phosphors," *Journal of the Electrochemical Society*, Vol. 100 (1953), p. 566.
[31] Froelich, H. C., "Electroluminescent Zinc Sulfide Phosphor." U. S. Patent 2,660,560 (11/24/53) and *Journal of the Electrochemical Society*, Vol. 100 (1953), p. 496.
[32] Froelich, H. C., "Sensitized Electroluminescent Response." *Journal of the Optical Society of America*, Vol. 43 (1953), p. 320.
[33] Destriau, G., "The Distribution of Copper Activator in Phosphor Crystals and its Effects on Electroluminescence." Symposium on Luminescence, Polytechnic Institute of Brooklyn (September 9, 1955.)
[34] Thorington, L., "Electroluminescence of Phosphors in Vacuo." Meeting of the Electrochemical Society, Philadelphia (May 4–8, 1952.)
[35] Howard, B. T., Private communication.
[36] Payne, E. C., "Electroluminescent Lamp." French Patent 1,022,100 (1952).
[37] Bernanose, A., Comte, M., and Vouaux, P., "A New Method of Light Emission by Certain Organic Compounds." *Journal de Chimie Physique*, Vol. 50 (1953), p. 64.
Bernanose, A., and Vouaux, P., "Organic Electroluminescence. Type of Emission." *Journal de Chimie Physique*, Vol. 50 (1953), p. 26.
Bernanose, A., and Vouaux, P., "The Spectrum of Acridine Orange Excited by Alternating Fields." *Bulletin de la Societé Chimique de France*, Vol. 20 (1953), p. 962.
Bernanose, A., "Electroluminescence of Organic Compounds." *British Journal of Applied Physics*, Supplement No. 4 (1955), p. 54.
Bernanose, A., and Marquet, G., "Electroluminescence of Carbazole by an Alternating Electric field. Characteristics of Organic Electroluminescence." *Journal de Chimie Physique*, Vol. 51 (1955), p. 255.
Bernanose, A., and Michon, F., "Organic Electroluminescence and Phosphorescence." *Journal de Chimie Physique*, Vol. 51 (1954), p. 622.
[38] Piper, W. W., and Williams, F. E., "The Mechanism of Electroluminescence of Zinc Sulfide," *British Journal of Applied Physics*, Supplement No. 4 (1955), p. 39.
[39] Halsted, R. E., and Koller, L. R., "Electroluminescence in Thin Films of ZnS:Mn." *Physical Review*, Vol. 93 (1954), p. 349.
[40] Howard, B. T., Ivey, H. F., and Lehmann, W., "Voltage Dependence of Electroluminescent Brightness." *Physical Review*, Vol. 96 (1954), p. 799.
[41] Lehmann, W., "Voltage Dependence of Electroluminescent Brightness." Symposium on Luminescence, Polytechnic Institute of Brooklyn (September 10, 1955).
[42] Destriau, G., "On the Effect of Various Metal Electrodes on Brightness Waves in Electroluminescence." Meeting of the Illuminating Engineering Society, Cleveland (September 12, 1955).
[43] Destriau, G., "Brightness Waveforms in Electroluminescence." *British Journal of Applied Physics*, Supplement No. 4 (1955), p. 49.
[44] Curie, D., "On the Mechanism of Electroluminescence." *Journal de Physique et le Radium*, Vol. 14 (1953), pp. 510–672.
[45] Kallmann, H., and Rosenberg, B., "Persistent Internal Polarization." *Physical Review*, Vol. 97 (1955), p. 1596.

[46] Matossi, F., "Polarization Effects on the Brightness Waves of Electroluminescence," *Physical Review*, Vol. 98 (1955), p. 434.
[47] Gobrecht, H., Hahn, D., and Seeman, F. W., "The Brightness Waves in the Luminescent Excitation of Phosphors by Alternating Electric Fields." *Zeitschrift für Physik.*, Vol. 140 (1955), p. 432.
[48] Schwiecker, W., "On Light Generation by Electroluminescence." *Lichttechnik*, Vol. 5 (1953), p. 152.
[49] Jerome, C. W., and Gungle, W. C., "Electrical Measurements on Electroluminescent Lamps with ZnS Phosphors." *Journal of the Electrochemical Society*, Vol. 100 (1953), p. 34.
[50] Waymouth, J. F., "Optical Measurements on Electroluminescent Zinc Sulfide." *Journal of the Electrochemical Society*, Vol. 100 (1953), p. 80.
[51] Lehmann, W., "Superlinear Frequency Dependence of Electroluminescent Brightness of Killed Phosphors." *Physical Review* (in press).
[52] Nail, N. R., Urbach, F., and Pearlman, D., "New Observations on Superlinear Luminescence." *Journal of the Optical Society of America*, Vol. 39 (1949), p. 690.
[53] Waymouth, J. F., Jerome, C. W., and Gungle, W. C., "Electroluminescence. Electrical and Optical Properties." *Sylvania Technologist*, Vol. 5 (1952), p. 54.
[54] Destriau, G., "Emission Spectra of Electroluminescent Materials with Multiple Activators." *Journal de Physique et le Radium*, Vol. 15 (1954), p. 13.
[55] Destriau, G., *Journal de Physique et le Radium*, (in press).
[56] Taylor, J. B., and Alfrey, G. F., *British Journal of Applied Physics*, Supplement No. 4 (1955), p. 44.
[57] Gobrecht, H., Hahn, D., and Gumlich, H. E., "On the Temperature Dependence of the Electroluminescence of ZnS and ZnO Phosphors." *Zeitschrift für Physik*, Vol. 136 (1954), p. 623.
[58] Haake, C. H., "Temperature Dependence of Electroluminescence." *Physical Review*, Vol. 98 (1955), p. 1544, and "The Temperature and Frequency Dependence of Electroluminescence," Symposium on Electroluminescence, Polytechnic Institute of Brooklyn (September 10, 1955).
[59] Johnson, P. D., Piper, W. W., and Williams, F. E., "The Effect of Electron Traps on Electroluminescence." Meeting of the Electrochemical Society, Cincinnati (May 2, 1955).
[60] Neumark, G., "Influence of an Electric Field on the Thermoluminescence of ZnS Crystals." *Physical Review*, Vol. 98 (1955), p. 1546.
[61] Mattler, J., "Modifications of Electroluminescence under the Action of Temperature." *Comptes Rendus*, Vol. 239 (1954), p. 1116.
[62] Andre, P., "Some Physical Constants of Araldite." *Le Vide*, Vol. 7 (1952), p. 1200.
[63] Mattler, J., private communication.
[64] Mattler, J., *Comptes Rendus*, (in press).
[64a] Waymouth, J. F., and Bitter, F., "Field-Induced Color Shift in Electroluminescent Zinc Sulfide." Symposium on Luminescence, Polytechnic Institute of Brooklyn, Sept., 10, 1955.
[65] Loebner, E. E., "Time Dependent Spectra of Electroluminescent Zinc Sulfide." *Physical Review*, Vol. 92 (1953), p. 846.
[66] Ince, A. N., "The Effect of Intense Magnetic Fields on Electroluminescent Powder Phosphors." *Proceedings of the Physical Society* (London), Vol. B67 (1954), p. 870.
[67] Millman J., and Seely, S., *Electronics*. New York, McGraw-Hill Book Co., Inc., 1941, p. 42.
[68] Rupp, E., "On Illumination of Phosphors by Magnetic and Electric Fields." *Annalen der Physik*, Vol. 75 (1924), p. 326.
[69] Mattler, J., and Destriau G., unpublished work.
[70] Vigean, F., "Electroluminescence—The Initial Brightness Produced by an Alternating Field." *Comptes Rendus*, Vol. 236 (1953), p. 1151.
[71] Frankl, D. R., "Electroluminescence in ZnS Single Crystals." Meeting of the Electrochemical Society, Cincinnati, (May 1–5, 1955).
[72] Piper, W. W., and Williams, F. E., "Electroluminescence of Single Crystals of ZnS:Cu." *Physical Review*, Vol. 87 (1952), p. 151.
[73] Watson, W. R., Dropkin, J. J., and Halpin, A. T., "Electroluminescence in ZnS Single Crystals." Meeting of the Electrochemical Society Chicago, (May 2–6, 1954).
[74] Loebner, E. E., and Freund, H., "Stacked Barriers in an Electroluminescent ZnS Crystal." *Physical Review*, Vol. 98 (1955), p. 1545, and Loebner, E. E., Symposium on Luminescence, Polytechnic Institute of Brooklyn (September 10, 1955).
[75] Destriau, G., "Luminescence in Electric Fields and Electronic Phenomena in Semi-Conductors." *Comptes Rendus*, Vol. 209 (1939), p. 36, and "Researches Upon Electrophotoluminescence." *Transactions of the Faraday Society*, Vol. 35 (1939), p. 227.
[76] Curie, D., "The Mean Free Paths of Electrons in Crystals, Electroluminescent Effects and Dielectric Breakdown." *Journal de Physique et le Radium*, Vol. 13 (1952), p. 317; Vol. 14 (1953), p. 135.
[77] Garlick, G. F. J., *Luminescent Materials*. Oxford University Press, 1949, p. 150.
[78] Williams, F. E., *Advances in Electronics*, Vol. V, *Solid State Luminescence*, L. Martin, Ed., New York, Academic Press, 1953, p. 137.
[79] Zalm, P., Diemer, G., and Klasens, H. A., "Some Aspects of the Voltage and Frequency Dependence of Electroluminescent Zinc Sulfide." *Philips Research Reports*, Vol. 10 (1955), p. 205.
[80] Fischer, A., "Explanation of the Destriau Effect." *Physikalische Verhandlungen*, Vol. 5 (1954), p. 64.
[81] Howard, B. T., "On the Theory of Electroluminescence." *Physical Review*, Vol. 98 (1955), p. 1544.
[82] Franz, W., "On Internal Field Emission from Localized Defect Centers in Insulating Crystals." *Annalen der Physik*, Vol. 11 (1952), p. 17. *Ergebnisse der exakten Naturwissenschaften*, Vol. 27 (1953), p. 1.
[83] Schwertz, F. A., Mazenko, J. J., Michalik, E. R., and Freund, R. E., "Voltage Dependence of Electroluminescent Brightness. I. Dielectric-Embedded Phosphors. II. Chemically Deposited Phosphor Films." *Physical Review*, Vol. 98 (1955), pp. 1133, 1134.
[84] Zalm, P., *British Journal of Applied Physics*, Supplement No. 4 (1955), p. 48.
[85] O'Neill, G. D., "Electroluminescence and Surface States." *Sylvania Technologist*, Vol. 6 (1953), p. 47.
[86] Haake, C. H., "Some Comments on the Frequency Dependence of Electroluminescent Brightness." *Physical Review* (in press.)
[87] Goffaux, R., "Remarks on the Mechanism of Electroluminescence." *Bulletin Acadamie Royale Belgique*, Vol. 40 (1954), p. 808.
[88] Lossew, O. W., "Luminescent Carborundum Detector and Detection Effect and Oscillations with Crystals." *Philosophical Magazine*, Vol. 6 (1928), p. 1024.
[89] Lossew, O. W., "Application of the Quantum Theory to the Luminescence of the Carborundum Detector." *Zeitschrift für Physik*, Vol. 30 (1929), p. 920; Vol. 32 (1931), p. 692.
[90] Claus, B., "Rectifying Effect and Luminescent Phenomena with Carborundum Crystals." *Zeitschrift fur Physik*, Vol. 31 (1930), p. 360; *Annalen der Physik*, Vol. 11 (1931), p. 331; Vol. 14 (1932), p. 644.
[91] Tetzner, H., "Investigation of Luminescent Phenomena in Synthetic Silicon Carbide Crystals." *Zeitschrift für Angewandte Physik*, Vol. 1 (1948), p. 153.
[92] Sillars, R. W., "Light Emission from Silicon Carbide." *Physica Review*, Vol. 85 (1952), p. 136.
[93] Lehovec, K., Accardo, C. A., and Jamgochian, E., "Injected Light Emission of Silicon Carbide Crystals." *Physical Review*, Vol. 83 (1951), p. 603; Vol. 89 (1953), p. 20.
[94] Szigeti, G., Bauer, G. T., and Weiszburg, J., "A Microphotometric Measurement of the Electroluminescence of Silicon Carbide Crystals." *Acta Physica Hungarica*, Vol. 4 (1954), p. 57.
Szigeti, G., "Electroluminescence Excited by Direct Currents in Silicon Carbide Crystals." *Acta Physica Hungarica*, Vol. 4 (1954), p. 64; *British Journal of Applied Physics*, Supplement No. 4 (1955), p. 56.
[95] Kendall, J. T., "Electronic Conduction in Silicon Carbide." *The Journal of Chemical Physics*, Vol. 21 (1953), p. 821.
[96] Curie, G., and Curie, D., "Brightness Waves in Electroluminescence and in the Luminescent Effect at Surfaces." *Journal de Physique et le Radium*, Vol. 15 (1954), p. 61.
[97] Haynes, J. R., and Briggs, H. B., "Radiation Produced in Germanium and Silicon by Electron-hole Recombination." *Physical Review*, Vol. 86 (1952), p. 647, 12th Physical Electronics Conference, M.I.T., 1952.
[98] Haynes, J. R., "New Radiation Resulting from Recombination of Holes and Electrons in Germanium." *Physical Review*, Vol. 98 (1955), p. 1866.
[99] Gunn, J. B., "Radiative Transitions in Germanium." *Proceedings of the Physical Society* (London), Vol. 66B (1953), p. 330.
[100] Newman, R., "Optical Studies of Injected Carriers. I. Recombination Radiation in Germanium." *Physical Review*, Vol. 91 (1953), p. 1313.
[101] Aigrain, P., "Light Emission from Injecting Contacts on Germanium in the 2μ to 6μ Band." *Physica*, Vol. 20 (1954), p. 1010.
[102] Wolff, G. A., and Hebert, R., "On the Luminescence of Gallium Phosphide." Meeting of the Electrochemical Society, Cincinnati (May 2–5, 1955).
[103] Newman, R., Dash, W. C., Hall, R. N., and Burch, W. E., "Visible Light from a Silicon *p-n* Junction." *Physical Review*, Vol. 98 (1955), p. 1536.
[104] Diemer, G., "Electrical Breakdown and Light Emission in CdS Single Crystals." *Philips Research Reports*, Vol. 9 (1954), p. 109.
[105] Smith, R. W., "Radiation from CdS Crystals Generated by dc Electric Fields." *Physical Review*, Vol. 93 (1954), p. 347; Vol. 98 (1955), p. 1169.

[106] Kröger, F. A., Vink, H. J., and van den Boomgaard, J., "Controlled Conductivity in CdS Single Crystals." *Zeitschrift fur Physikalische Chemie*, Vol. 203 (1954), p. 1.
[107] Hahn, E. E., "Some Electrical Properties of Zinc Oxide Semiconductor." *Journal of Applied Physics*, Vol. 22 (1951), p. 855.
[108] Fischer, A., "On the Electroluminescence of Semiconducting Phosphors." *Zeitschrift fur Naturforschung*, Vol. 8a (1953), p. 756.
[109] Güntherschulze, A., and Gerlach, M., "Luminescence Excited in Calcium Tungstate by an Electric Current." *Zeitschrift Physik*, Vol. 88 (1934), p. 355.
[110] McKay, K. G., and Chynoweth, A. G., "Optical Studies of Avalanche Breakdown in Silicon." *Physical Review*, Vol. 99 (1955), p. 1648.
Rose, D. J., and McKay, K. G., "Microplasmas in Silicon." *Physical Review*, Vol. 99 (1955), p. 1648.
[111] Schön, M., "On the Luminescence of Carborundum Crystals During Passage of Current." *Zeitschrift für Naturforschung*, Vol. 8a (1953), p. 442.
[112] Van Roosbroeck, W., and Shockley, W., "Photon-Radiative Recombination of Electrons and Holes in Germanium." *Physical Review*, Vol. 94 (1954), p. 1558.
[113] Mattler, J., and Curie, D., "Behavior of Electron Traps in the Phenomenon of Electrophotoluminescence." *Comptes Rendus*, Vol. 230 (1950), p. 2086.
[114] Curie, D., "On the Fundamental Law of the Emptying of Traps." *Comptes Rendus*, Vol. 240 (1955), p. 1614.
[115] Destriau, G., "Effect of Temperature on the Transient Illumination of Phosphorescent Sulfides Submitted to the Action of Electric Fields." *Comptes Rendus*, Vol. 230 (1950), p. 205.
[116] Destriau, G., and Mattler, J., "Effect of an Elevation in Temperature on the Phenomenon of Electrophotoluminescence." *Journal de Physique et le Radium*, Vol. 11 (1950), p. 529.
[116a] Krautz, E., "Phosphors in High Electric Fields." *Zeitschrift für Naturforschung*, Vol. 4a (1949), p. 284.
[117] Nudelman, S., and Matossi, F., "Effect of Electric Fields on Ultra-Violet-Excited Phosphors." *Physical Review*, Vol. 98 (1955), p. 1545, Meeting Electrochemical Society, Cincinnati (May 1–5, 1955).
[118] Vigean, F., "Quenching Action of Longwave Radiation in Electrophotoluminescence. Experimental Results." *Comptes Rendus*, Vol. 232 (1951), p. 819.
[119] Vigean, F., and Curie, D., "Quenching Action of Longwave Radiation in Electrophotoluminescence. Tentative Explanation." *Comptes Rendus*, Vol. 232 (1951), p. 955.
[120] Curie, M., *Fluorescence-Phosphorescence*. Paris, Hermann, 1946.
[121] Garlich, G. F. J., and Mason, D. E., "Electron Traps and Infrared Stimulation of Phosphors." *Journal of the Electrochemical Society*, Vol. 96 (1949), p. 112.
[122] Destriau, G., and Mattler, J., "Oscillographic Study of Transient Emission." *Journal de Physique et le Radium*, Vol. 13 (1952), p. 205.
[123] Destriau, G., and Mattler, J., "Action of Alternating Electric Fields on the Intensity of Luminescence of Sulfides Excited by X-Rays." *Journal de Physique et le Radium*, Vol. 9 (1948), p. 258.
[124] Destriau, G., Mattler, J., Destriau, M., and Gumlich, H. E., "The Enhancement Effect of Electric Fields on Some X-Ray Excited Phosphors." *Journal of the Electrochemical Society*, to appear in December, 1955 issue.
[125] Cusano, D. A., "Radiation-Controlled Electroluminescence and Light Amplification in Phosphor Films." *Physical Review*, Vol. 98 (1955), p. 546.
[126] Destriau, G., "Semipermanent Modifications of the Properties of Crystals Subjected to Alternating Current." *Comptes Rendus*, Vol. 230 (1950), p. 1061.
[127] Mattler, J., "The Effect of Temperature on the Extinction of Sulfide Luminescence by Electric Fields." *Comptes Rendus*, Vol. 230 (1950), p. 76.
[128] Matossi, F., and Nudelman, S., "Influence of Electric Fields on Luminescence." *Physical Review*, Vol. 89 (1953), p. 660.
[129] Olson, K. W., "Variations in the Decay of Phosphorescence with Frequency of Applied Electric Fields." *Physical Review*, Vol. 92 (1953), p. 1323.
[130] Miller, T., "Radiofrequency Field Quenching of Ultra-Violet-Excited Phosphors." *Journal of Applied Physics*, Vol. 23 (1952), p. 1289.
[131] Destriau, G., unpublished work.
[132] Destriau, G., "Brightness Waves and Transitory Phenomena in the Quenching of Luminescence by Alternating Electric Fields." *Journal of Applied Physics*, Vol. 25 (1954), p. 67.
[133] Gumlich, H. E., in press.
[134] Steinberger, J. J., Low, W., and Alexander, E., "Influence of Alternating Electric Fields on the Light Emission of Some Phosphors." *Physical Review*, Vol. 99 (1955), p. 1217.
[135] Matossi, F., "Interpretation of Electroluminescent Effects in an Excited Phosphor." *Naturwissenschaften*, Vol. 40 (1953), p. 239; *Physical Review*, Vol. 94 (1954), p. 1151.
[136] Williams, F. E., "Theoretical Basis for Light-Amplifying Phosphors." *Physical Review*, Vol. 98 (1955), p. 547.
[137] von Hippel, A., Gross, E. P., Jelatis, J. G., and Geller, M., "Photocurrent, Space-Charge Buildup, and Field Emission in Alkali Halide Crystals." *Physical Review*, Vol. 91 (1953), p. 568.
[138] Bay, Z., and Szigeti, G., "Electric Source of Light." U. S. Patent No. 2,254,957 (9/2/41) and Hungarian Patent 140,547.
[139] Sherwood, E. J., and Skinner, G. D., "Improvements in or Relating to Electric Light Sources." British Patent 567,184 (2/1/45).
[140] Payne, E. C., Mager, E. L., and Jerome, C. W., "Electroluminescence. A New Method of Producing Light." *Illuminating Engineering*, Vol. 45 (1950), p. 688; *Sylvania Technologist* (January, 1951), p. 2.
[141] Mager, E. L., "Electroluminescent Lamps." U. S. Patent 2,566,349 (9/4/51) and 2,624,857 (1/6/53).
[142] Butler, K. H., Jerome, C. W., and Waymouth, J. F., "The Electroluminescent Lamp. A New Light Source." *Electrical Engineering*, Vol. 73 (1954), p. 524.
[143] Butler, K. H., and Waymouth, J. F., "Electroluminescence of Zinc Sulfide Phosphors." *British Journal of Applied Physics*, Supplement No. 4 (1955), p. 33.
[144] Mager, E. L., and Lowry, E. F., "Electroluminescent Lamp." U. S. Patent 2,684,450 (7/20/54).
[145] Jenkins, H. G., "Improvements in or Relating to Electroluminescent Devices." British Patent No. 717,171 (10/20/54) and U. S. Patent 2,714,683 (8/2/55). See also *British Journal of Applied Physics*, Supplement No. 4 (1955), p. 38.
[146] Lehmann, W., private communication.
[147] Rulon, R. M., "Electroluminescent Lamps Using Glass Dielectrics." Meeting of the Electrochemical Society, Chicago (May 2–6, 1954), and "Advancements in Panelescent Lamps." *Sylvania Technologist*, Vol. 8 (1955), p. 45.
[148] Lehovec, K., "New Photoelectric Devices Utilizing Carrier Injection." PROCEEDINGS OF THE IRE, Vol. 40 (1952), p. 1407.
[149] Koller, L. R., "Electroluminescent Structure." U. S. Patent 2,709,765 (5/31/55).
[150] Schwertz, F. A., Haller, M. N., and Mazenko, J. J., "Self-Luminous Halftones." *Journal of the Optical Society of America*, Vol. 44 (1954), p. 745.
[151] Dietz, E., Jr., and Jordan, J. J., "Indicator Fuse Plug." U. S. Patent 2,702,329 (2/15/55).
[152] Piper, W. W., "Phosphor Screen." U. S. Patent 2,698,915 (1/4/55).
[153] Low, W., "Possibility of a Nuclear Radiation Detector by Means of Electroluminescent Phosphors." *Physical Review*, Vol. 98 (1955), p. 556.
[154] Bramley, A., and Rosenthal, J. E., "Transient Voltage Indicator and Information Display Board." *Review of Scientific Instruments*, Vol. 24 (1953), p. 471.
[155] Marshall, B. O., Jr., Schwertz, F. A., and Bowman, J. R., Quarterly Report No. 3, First Series, Computer Components Fellowship No. 347, Mellon Institute, 1951.
[156] White, W. C., "X-ray Intensification and Method." U. S. Patent 2,650,310 (8/25/53).
[157] Orthuber, R. K., and Ullery, L. R., "A Solid-State Image Intensifier," *Journal of the Optical Society of America*, Vol. 44 (1954), p. 297.
[158] Kazan, B., and Nicoll, F. H., in press.
[159] Amalgamated Wireless (Australasia) Ltd., "Light Amplifying Cell." British Patent 713,916 (8/18/54).
[160] Westinghouse Electric Corp., "Improvements in or Relating to Radiation Image Transformers." British Patent 724,977 (2/23/55).
[161] Matossi, F., "Dependence of Light Amplification in Phosphors on Light Intensity." *Physical Review*, Vol. 99 (1955), p. 1332.
[162] Woods, J., and Wright, D. A., "Field Enhanced Cathodo-Luminescence in Magnesium Oxide." *Proceedings of the Physical Society* (London), Vol. 68B (1955), p. 566.



The Physical Chemistry of Crystal Phosphors*

F. A. KRÖGER†

Summary—After a historical introduction, a survey is given of the present views regarding the constitution and the preparation of inorganic crystal phosphors. Particular attention is paid to the incorporation of atoms with a valency deviating from that of the atoms of the base material, and to the stabilization of atoms in a particular valency.

I. Introduction

PHOSPHORS in use in industry are for the greater part inorganic crystalline solids. In principle all crystalline solids show luminescence, provided the temperature is sufficiently low. Above a certain critical temperature, however, the luminescence is invariably quenched. The spectrum that is emitted in fluorescence may vary over a wide range. In engineering we are mainly interested in phosphors which show fluorescence up to a temperature well above room temperature and having a spectrum lying in an appropriate range. Dependent on the application, this may be the ultraviolet, the visible, or the infrared part of the spectrum.

Both the quenching temperature and the spectral distribution of the fluorescence have been found to be strongly dependent on the presence of a small concentration of "centers" in the crystals. Such centers may be formed by foreign atoms, or by regions in the crystals that are perturbed in some way or another, e.g. by the fact that atoms of the crystal that should be present according to the crystal structure are lacking (vacancies) or that sites which should not be occupied are occupied (interstitials). Dislocations might also play a rôle.

Centers may have favorable or unfavorable effects. The former are called activators, the latter killers or quenchers. In some cases the absence of killer centers is sufficient to get a good phosphor. This is the case, for instance, for the tungstates such as $CaWO_4$ and $MgWO_4$, and also for barium platin cyanure ($BaPt(CN)_4$), one of the oldest substances used for X-ray intensifier screens. In others the required properties only come into existence by the introduction of activators. It is the task of the chemist to prepare systems in which centers having unfavorable effects are absent and in which centers having favorable effects are present in the optimum concentration.

II. Historical

The first phosphors known are probably alkaline earth sulfides. Thus Livy (200 B.C.)[1] describes nightly orgies (Bacchanalia) in which women, initiated into the orgiastic proceedings, dipped torches into the waters of the Tiber without them being extinguished. These torches were probably covered with luminescent calcium sulphide. Recipes for the preparation of such phosphors have been re-invented several times (Casciarolo, 1602, Homberg, 1694, Canton, 1768, Balmain, 1881 and many others).[2]

The recognition of the rôle a particular element may play in causing luminescence, i.e., the function of the activator, seems to be due to Homberg[3] who found that stones from Bologna ($=BaSO_4$) became luminescent when heated with sulfur after being pulverized in a bronze mortar, but not when an iron mortar had been used. Copper proved to be the necessary ingredient for obtaining a phosphor.

A great many activators were found at the end of the 19th century: Verneuil[4]) showed that bismuth was the activator in a calcium sulphide phosphor prepared by Balmain. Becquerel[5] and Lecoq de Boisbaudran[6] found the activator action of manganese, chromium, and several rare earth atoms for various sulphates, carbonates, and oxides.

Many of these phosphors could be prepared by simply heating together the base compound and a small quantity of the activator. In some cases, however, other compounds must be added to get good results. Thus Becquerel[7] and de Visser[8] found that CaS-Bi phosphors only become good phosphors when salts of sodium, lithium, or rubidium were added in the preparation. A great many of such phosphors were later found and studied by Lenard and his co-workers.[9] Phosphors of this group are therefore often indicated by the name of Lenard phosphors. According to Lenard the extra substance added was required to enhance reaction at a low temperature by causing the formation of a melt. but every substance having this effect could not be used

* Original manuscript received by the IRE, August 5, 1955.
† Philips Research Labs., N. V. Philips Gloeilampenfabrieken, Eindhoven-Netherlands.

[1] W. P. Jorissen, "The use of phosphorescent calcium sulfide by the Bacchantes," *Jour. Chem. Ed.*, vol. 25, pp. 685–686; December, 1948.

[2] F. Fritz, "Leuchtfarben, Verlag Bodenbinder," Steglitz, Berlin, pp. 13–30; 1940.

[3] *Hist. Acad. Sci.* (Paris), Tome II, p. 133; 1733. See ref. 2, p. 21.

[4] A. Verneuil, "Sur la préparation du sulfure de calcium à phosphorescence violette," *Compt. Rend.* (Paris), vol. 103, p. 600; 1886. "Sur les causes déterminantes de la phosphorescence du sulfure de calcium," *Compt. Rend.* (Paris), vol. 104, p. 501; 1887.

[5] E. Becquerel, "Action du manganèse sur le pouvoir de phosphorescence du carbonate de chaux," *Compt. Rend.* (Paris), vol. 103, p. 1098; 1886.

[6] L. de Boisbaudran, "Fluorescences du manganèse et du bismuth," *Compt. Rend.* (Paris), vol. 103, p. 1064; 1886. "Sur la Fluorescence rouge de l'alumine," *Compt. Rend.* (Paris), vol. 103, p. 110; 1886.

[7] E. Becquerel, "Sur la préparation des sulfures de calcium et de strontium phosphorescents," *Compt. Rend.* (Paris), vol. 107, p. 892; 1888.

[8] L. E. O. de Visser, "Essai d'une théorie sur la phosphorescence de longue durée, spécialement sur celle des sulfures alcalinoterreux," *Rec. Trav. Chim. Pays Bays*, vol. 20, p. 435; 1901. "Sur la phosphorescence du sulfure de calcium bismuthifère, préparé en présence de traces de sodium," *Rec. Trav. Chim. Pays Bays*, vol. 22, p. 133; 1903.

[9] P. Lenard and V. Klatt, "Ueber die Phosphoreszenzen des Kupfers, Wismuths und Mangans in den Erdalkali sulfiden," *Ann. Phys.*, 3, vol. 38, p. 89; 1889.

WRH

with equal success, and each phosphor required its particular "flux." As shall be shown below it has proved possible to give a simple explanation for the specific effect of the flux.

Rather than following the ups and downs of former research and development in this field, we shall discuss in the following sections the preparation of phosphors in terms of the present knowledge of the process.

III. The Present Status of Phosphor Preparation

The preparation of a phosphor usually comprises the following steps: (1) purification of primary materials, (2) reaction of the primary materials, to give the required compound, and (3) the formation of well-formed flowless crystals of the appropriate size from the compound with addition of the proper amount of foreign atoms necessary for the appearance of luminescence.

The first step is often carried out in aqueous solution, using the well-established chemical techniques of purification, such as recrystallization, partial precipitation, absorption at active surfaces, etc. As nowadays more and more high-grade chemicals are available on the market it is often advisable to delegate the carrying out of this step to the suppliers of the primary materials.

The second step may also be carried out in solution. This method is used, for instance, for the formation of ZnS, which is mostly precipitated from a pure solution of a zinc salt with H_2S or $(NH_4)_2S$. This step may, however, also be carried out in the solid state. Thus $CaWO_4$ is formed by mixing calcium carbonate ($CaCO_3$) and tungstic oxide (WO_3) and heating the mixture at an appropriate temperature (1,000°C.).

The final step, the formation of the actual phosphor, must be carried out under such conditions that the components of the system are sufficiently mobile to allow recrystallization to take place. This may be achieved by heating at a high temperature. In general a temperature equal to or exceeding two-thirds of the absolute temperature of the melting point is sufficient. As this temperature may be unpleasantly high it is often advisable to add foreign compounds which, together with the phosphor base, form a system with a lower melting point (flux).

Thus pure ZnS may be crystallized by heating at $T \cong 1{,}200$°C.; when suitable fluxes like NaCl or $CaCl_2$ are added, however, crystallization becomes possible at temperatures as low as 800°C.

It must be expected that phosphors prepared with addition of a flux contain atoms of the flux in the crystals of the phosphor proper. For this reason the flux must not consist of atoms of a type that are harmful to the luminescence (killers). As we shall see, the atoms of the flux may also have a favorable effect in the phosphor. In fact in many cases this favorable effect is at least as important as its mineralizing action.

When activated phosphors are prepared with addition of fluxes, the activator will in general distribute itself over the phosphor phase and the flux. Equilibria of this kind have been studied in detail by R. Ward and his coworkers[10] for infrared sensitive SrS and SrSe phosphors activated by europium or cerium and samarium, prepared with $SrCl_2$ as a flux.

IV. The Incorporation of the Activator

Inorganic crystals used as a base for phosphors are mostly solids built up of ions. Foreign atoms built in such crystals are also present as ions, which take the place of one of the ions of the base. The activator action of such ions is closely related to the valency with which they are incorporated. Thus for manganese as an activator the luminescence due to divalent manganese (Mn^{2+}) differs markedly from that due to tetravalent manganese (Mn^{4+}), whereas trivalent manganese does not give rise to luminescence.

If the activator ions have the same valency as the ion which they replace, incorporation takes place without complications, provided the ionic radii do not differ too much; e.g.,

$$(1-\delta)ZnS + \delta MnS \rightarrow Zn_{(1-\delta)}Mn_{\delta}S. \quad (1)$$

When the valencies differ, however, complications arise because a molecule of the activator compound does not fit in the place of a removed molecule of the base. Two types of complications may occur: (1) atoms (or ions) that do not fit in the normal lattice are incorporated at sites that are normally not occupied (so-called interlattice sites), or (2) vacancies are formed. For incorporation of the silver activator in ZnS the latter possibility may be formulated as follows:

$$(1-2\delta)ZnS + \delta Ag_2S \rightarrow Zn_{(1-2\delta)}Ag_{2\delta}S_{(1-\delta)}(V_S)_{\delta}, \quad (2)$$

in which V_S stands for a sulfur vacancy. Similarly, upon incorporation of chlorine, zinc vacancies are formed:

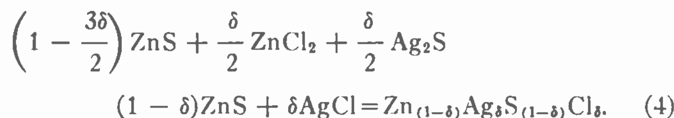
$$(1-2\delta)ZnS + \delta ZnCl_2 \rightarrow Zn_{(1-\delta)}(V_{Zn})_{\delta}S_{(1-2\delta)}Cl_{2\delta}. \quad (3)$$

The formation of vacancies (or interstitials) usually involves a large amount of energy. Therefore, in such cases, the concentration of foreign atoms that can be dissolved in the base may be extremely low and—as far as an activator is concerned—too low to give a good phosphor.

The solubility may be markedly increased if we manage to prevent formation of vacancies or interstitials. This can be achieved by incorporating atoms like silver,

[10] R. Ward, "Infrared-sensitive phosphors of the alkaline earth sulfides and selenides," *Jour. Electrochem. Soc.*, vol. 93, pp. 1–176; May, 1948.
R. W. Maison, C. F. Hiskey, and R. Ward, "The use of radioactive tracers in a study of activator distribution in infrared-sensitive phosphors," *Jour. Am. Chem. Soc.*, vol. 71, pp. 509–514; February, 1949.
J. Prener, R. W. Maison, and R. Ward, "The use of radioactive tracers in a study of activator distribution in infrared-sensitive phosphors; Effect of strontium oxide," *Jour. Am. Chem. Soc.*, vol. 71, pp. 1803–1805; May, 1949.
A. Dreeben and R. Ward, "Radioactive tracer study of activator distribution in infrared phosphor systems; the effect of oxide and calcium ions," *Jour. Am. Chem. Soc.*, vol. 73, pp. 4679–4680; October, 1951.

that tend to produce sulfur vacancies together with atoms like chlorine that tend to give zinc vacancies:[11]



$$\left(1-\frac{3\delta}{2}\right)ZnS+\frac{\delta}{2}ZnCl_2+\frac{\delta}{2}Ag_2S$$
$$(1-\delta)ZnS+\delta AgCl=Zn_{(1-\delta)}Ag_\delta S_{(1-\delta)}Cl_\delta. \quad (4)$$

If ZnS and AgCl are supposed to consist of Zn^{2+}, S^{2-}, Ag^+ and Cl^- ions, the effect of chlorine on the solubility of silver and vice versa can be described in terms of charge compensation: the lack of positive charge resulting from the replacement of Zn^{2+} by Ag^+ is compensated by the lack of negative charge due to the replacement of S^{2-} by Cl^-. As is shown by the reaction equation (4), however, the assumption of ions is not essential for the effect: it occurs quite generally, independent of the way in which charges are distributed over the atoms in the solid. This effect of chlorine (and also of similar atoms like bromine or iodine) now explains why, in the case of ZnS phosphors, chlorides and bromides, when used as flux, have been found to produce particularly good phosphors: the flux acts as a source of chlorine or bromine atoms required for the enhancement of the solubility of monovalent activator ions. It is easily seen that compensation similar to that achieved by chlorine is also affected by metal atoms having a valency higher than that of zinc; e.g., aluminum, gallium, praseodymium, etc:[12]

$$(1-4\delta)ZnS+\delta Ag_2S+\delta Al_2S_3\rightarrow Zn_{(1-4\delta)}Ag_{2\delta}Al_{2\delta}S. \quad (5)$$

As in this case the mineralizing action of the chloride is lacking, good phosphors are only obtained when the preparation is carried out at a higher temperature (~1,200°C). In the case of ZnS-Ag, Al the monovalent atom (silver) is the activator. It may also happen, however, that the activator is a trivalent atom. This is the case for instance in CaS-Bi. Now monovalent metal atoms are necessary to increase the solubility of the bismuth. This explains why in this case substances containing monovalent metal atoms like Na_2SO_4 have been found to be good fluxes (Section II).

From the fact that it is possible to give a satisfactory explanation of the rather complex experimental data on the basis of the simple considerations given above it may be concluded that the assumptions underlying these considerations are correct; viz., that (1) the activators are actually monovalent (in ZnS) or trivalent (in CaS), and that (2) they are present at normal lattice sites, replacing atoms of the base.

V. Electron Traps

So far we have only paid attention to the activator centers required for the actual luminescent electronic transition. If a phosphor shows a long persisting temperature-dependent afterglow, this is due to the presence of other centers which are able to bind free electrons and which release these only when vibrational collisions with the surrounding atoms provide the required energy. Such centers are called electron traps. It has been found that these traps, just as the fluorescence centers, may be due to the presence of foreign atoms. Thus in $Zn_2SiO_4-MnSiO_4$, arsenic forms traps.

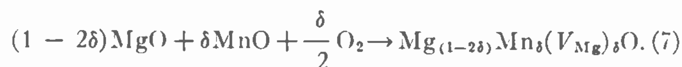
In ZnS phosphors the co-activators, like chlorine and aluminum, which make possible the incorporation of the monovalent activators, themselves act as traps. This can easily be understood: when sulfur is replaced by chlorine or when zinc is replaced by aluminum, centers are formed that have an effective charge $+/e/$ relative to the normal crystal, and electrons may be bound in the field of this effective-positive charge.

VI. Stabilization of the Activator in a Particular Valency

In the introduction we have seen that the activator action of a particular atom is closely related to the valency with which it is incorporated. In cases in which an activator may occur with various valencies the principle of compensation outlined above may be used to stabilize the activator in a particular valency.[13] Thus if manganese is incorporated in a substance like MgO, it dissolves mainly in the divalent form

$$(1-\delta)MgO+\delta MnO\rightarrow Mg_{(1-\delta)}Mn_\delta O. \quad (6)$$

Under strongly oxidizing conditions incorporation may take place in a higher valency, but now vacancies are formed:

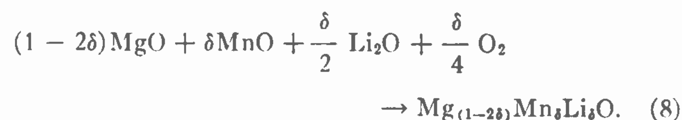


$$(1-2\delta)MgO+\delta MnO+\frac{\delta}{2}O_2\rightarrow Mg_{(1-2\delta)}Mn_\delta(V_{Mg})_\delta O. \quad (7)$$

As a consequence, only a small quantity of manganese can be incorporated, the rest remaining as a separate phase (MnO_2, Mn_3O_4, Mn_2O_3, etc.).

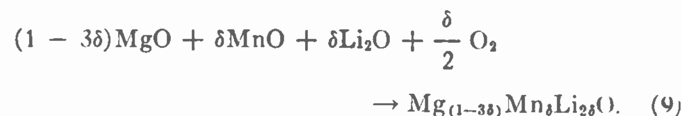
When atoms having a lower valency than magnesium; e.g., monovalent lithium, are added, however, the solubility of manganese in a valency higher than two may be increased.

Trivalent manganese is stabilized if one lithium is added for each manganese present:



$$(1-2\delta)MgO+\delta MnO+\frac{\delta}{2}Li_2O+\frac{\delta}{4}O_2$$
$$\rightarrow Mg_{(1-2\delta)}Mn_\delta Li_\delta O. \quad (8)$$

Tetravalent manganese, however, is stabilized if two or more lithium atoms are present for each manganese:



$$(1-3\delta)MgO+\delta MnO+\delta Li_2O+\frac{\delta}{2}O_2$$
$$\rightarrow Mg_{(1-3\delta)}Mn_\delta Li_{2\delta}O. \quad (9)$$

[11] F. A. Kröger, and J. E. Hellingman, "Chemical proof of the presence of chlorine in blue fluorescent zinc sulfide," *Jour. Electrochem. Soc.*, vol. 95, pp. 68–69; February, 1949.

[12] F. A. Kröger, and J. A. M. Dikhoff, "Trivalent cations in fluorescent zinc sulfide," *Physica*, vol. 16, pp. 297–316; March, 1950.

[13] J. S. Prener, "A magnesium oxide phosphor activated by tetravalent manganese," *Jour. Chem. Phys.*, vol. 21, pp. 160–161; January, 1953.

This principle has been used by Klasens[14] to make stable red luminescent arsenate phosphors activated by tetravalent manganese.

VII. Vacancies as Activators

In the discussion given so far the problems with which we have to deal have been simplified by assuming that the energy involved in forming vacancies was practically prohibitive for their formation. This is not quite correct.

Although application of the compensation principle leads to better phosphors in many cases, the effect is of a quantitative rather than of a qualitative nature. Thus vacancies may be present, albeit in comparatively low concentrations. It has even been found that some types of vacancies may form fluorescence centers, and thus act as activators, viz., zinc vacancies in zinc sulphide phosphors. Some of these vacancies may already be present in pure ZnS as a consequence of the gain of entropy.[15] The concentration is increased, however, in the presence of atoms like chlorine and aluminum, as the incorporation of such atoms is accompanied by the formation of vacancies (Section IV).[16] Actually the blue luminescence of so-called "self-activated" ZnS, which is due to zinc vacancies, is shown particularly by samples prepared with addition of $ZnCl_2$ or Al_2S_3.

VIII. A More General Theory; Future Developments

Although we have attempted to give some insight into the factors governing phosphor formation, the treatment here has been too limited and too simple to tell exactly what happens during the preparation of a phosphor. We might also put it like this: we have given a picture in black and white, whereas the reality shows a lot of gradation. In the case of the oxidation of manganese, for instance, it has been stated that lithium favors the incorporation of high valent manganese. From the reaction equations (8) and (9) it is seen, however, that oxygen from the atmosphere is involved. When there is no oxygen, oxidation cannot occur. Accordingly the degree of oxidation and therewith the properties of the phosphor must depend on the partial pressure of oxygen in the atmosphere. Similar effects exist in nearly all cases.

A quantitative theory for the way in which the composition of solids varies as a function of the atmosphere in which they are being prepared has been given a long time ago by Schottky and Wagner,[17] with special reference to their semiconductor properties.

In agreement with experiments on pure solids, the theory predicts a gradual variation of properties of the crystals when the composition of the atmosphere is changed. Verwey *et al.*,[18] and later also Hauffe, *et al.*,[19] have shown that the dependence changes markedly when foreign atoms with a valency different from that of the atoms of the base crystal, are present. In this case the properties are practically independent of the atmosphere over a wide range of atmospheres but they change outside this range.

Recently the Schottky-Wagner theory has been applied also in such complicated cases, and it has been found possible to give a complete description of the properties of solids containing foreign atoms over the entire range of atmospheres.[20] It is probable that this theory will also prove to be of value in the phosphor field.

Its application has two aspects: In the first place the theory may aid in interpreting experimental results. It can only be applied, however, if the experimental conditions are accurately known. Therefore the tendency to use the theory will stimulate the carrying out of experiments under well-defined—as well as simple—conditions.

It may be expected that this combination will prove to be extremely favorable for further developments in the physical chemistry of crystal phosphors.

[14] H. A. Klasens, "Temperature dependence of the luminescence and chemical stability of basic magnesium arsenate activated with tetravalent manganese," *Philips Res. Rep.*, vol. 9, pp. 377–390; October, 1954.

[15] R. H. Rube, "A luminescence and trapping in zinc sulfide phosphors with and without copper activator," *Phys. Rev.*, vol. 80, pp. 655; 1950. "The host crystal luminescence of zinc sulfide phosphors," *Jour. Chem. Phys.*, vol. 20, pp. 708–718; April, 1952.
R. H. Bube, and S. Larach, "Luminescence and trapping in phosphors containing gallium," *Jour. Chem. Phys.*, vol. 21, pp. 5–11; January, 1953.

[16] F. A. Kröger, and H. J. Vink, "The origin of the fluorescence in self-activated ZnS, CdS and ZnO." *Jour. Chem. Phys.*, vol. 22, pp. 250–252; February, 1954.

[17] C. Wagner, and W. Schottky, "Theorie der geordneten Mischphasen, I," *Z. physik. Chem.* B, vol. 11, p. 163; 1931.
C. Wagner, "Theorie der geordneten Mischphasen, II." *Z. physik. Chem. Bodenstein Festband*, pp. 177–186; 1931. "Theorie der geordneten Mischphasen, III," *Z. physik. Chem.* B, vol. 22, pp. 181–194; 1933.
C. Wagner, "Fehlordnungs-erscheinungen in kristallisierten polaren Verbindungen als Grundlage für Elektronen und Ionenleitung, *Z. Elektrochem.*, vol. 39, pp. 543–545; July, 1933.
W. Schottky, "Ueber den Mechanismus der Ionen-bewegung in festen Elektrolyten," *Z. physik. Chem.* B, vol. 29, pp. 335–355; 1935.

[18] E. J. W. Verwey, P. W. Haayman, and F. C. Romeyn, "Over de electrische eigenschappen van metaaloxyden met vreemde atomen in het kristalrooster," *Chem. Weekbl*, vol. 44, pp. 705–708; December 4, 1948.
E. J. W. Verwey, P. W. Haayman, F. C. Romeyn, and G. W. van Oosterhout, "Controlled-valence semi-conductors," *Philips Res. Rep.*, vol. 5, pp. 173–187; June, 9950.

[19] K. Hauffe, "Fehlordnungs-erscheinungen und Leitungsvorgänge in ionen und elektronen leitenden festen Stoffen," *Ergebn. exakt. Naturw.*, vol. 25, pp. 193–292 (particularly page 271), 1951.

[20] F. A. Kröger, H. J. Vink, and J. van den Boomgaard, "Controlled conductivity in CdS single crystals," *Z. physik. Chem.*, vol. 203, pp. 1–72; 1954.

Some Properties of Ferrites in Connection with Their Chemistry*

E. W. GORTER†

Summary—After an elementary introduction on the origin of magnetism, intended for the reader who is not acquainted with the subject, it is shown how the molecular-field hypothesis can account for the magnetic properties of ferromagnetics and antiferromagnetics, and for the noncompensated antiferromagnetics, with which latter materials we are concerned here. A brief description of the spinel lattice is given, and an account of the crystal chemistry of the spinels which is necessary to understand the experimental saturation magnetizations which are discussed. A short survey of methods of preparation is given. The second part discusses the anisotropies, and some of the magnetization processes which influence permeability, and the factors which influence high-frequency permeability and losses. Among these are the ferromagnetic resonance phenomenon and the dimensional resonance and relaxation phenomena. In this part the way in which these factors are influenced by chemical composition and preparation technique is indicated.

Finally, a short history of the development of the ferrites is given

I. Origin of Ferromagnetism of Oxides

A. Introduction

When a material is subjected to a magnetic field H, a magnetization I (*e.g.*, per cm^3) is produced. The susceptibility (*e.g.*, per cm^3) $\chi = I/H$ is either very small and negative (diamagnetism), or small and positive $1 \geqq \chi > 0$ (paramagnetism), or very large and positive $\chi \geqq 1$ (ferromagnetism).

The origin of these magnetizations lies in the orbital motion of the negatively charged electrons around the atomic nuclei, and in the spinning motions of the electrons around their own axes. These "circular" motions of an electrically charged particle may be likened to a flow of an electric current through a circular wire, which also produces a magnetic moment.

Let us first consider a free atom and see what spectra has taught us. We shall use the notation of the old quantum theory and introduce some corrections later. The electrons move in different orbits, the energy of which is determined mainly by the principal quantum number n, connected with the average distance from the nucleus: the principal axes of the elliptical orbits are proportional to n^2. An electron having a certain value of n may have different discrete values of the orbital angular momentum, (and therefore, different eccentricities of the elliptical orbits) equal to l units $\hbar = h/2\pi$ (h = Planck's constant), in which l may be any integral number between 0 and $n-1$. Orbits with $l=0$, 1, 2, 3, 4, 5 are indicated by s, p, d, f, g, h, etc.; *i.e.*, for $n=6$: by $6s$, $6p$, $6f$, $6g$, and $6h$, respectively.

The orbital motion of the electron causes an orbital magnetic moment

$$\mu_l = \frac{e}{2mc} \cdot p_l = l \text{ units } \frac{e}{2mc} \cdot \frac{h}{2\pi},$$

which are called Bohr magnetons, μ_B, (e = charge-, m = mass of the electron, c = velocity of light).

The angular momentum of the spinning motion is s units in which $s = \frac{1}{2}$.

The magnetic moment μ_s caused by the spinning motion appears not to be $s\mu_b$, but $\mu_s = (e/mc)\ p_s = 2\ s\mu_b$.

We shall first consider the case in which only one electron is present. When a strong magnetic field is applied at 0°K the angular momenta p_l and p_s are oriented in the direction of the field; at higher temperatures they will have any positions provided that their projection on the direction of the field is an integral number m_l, or a half-integral number m_s of units $\hbar$. m_s may be $+\frac{1}{2}$ or $-\frac{1}{2}$; m_l may be $l, l-1 \cdots 0 \cdots -l+1, l$. This means that p_l and p_s will process at certain angles with respect to H.

The electron can therefore occupy several different energy levels, characterized by the possible values of n, l, m_l, m_s.

The Pauli principle states that an energy state characterized by certain values of n, l, m_l, and m_s can only contain one electron. This law determines the maximum number of electrons that can occupy energy levels with a certain l value; viz., $2(2l+1)$. There can be two s electrons ($l=0$) in each principal orbit, six p electrons ($l=1$), ten d electrons ($l=2$), fourteen f electrons ($l=3$), etc. Thus in the orbit with $n=1$ there are two electrons; for $n=2$, eight electrons; for $n=3$, eighteen electrons; generally for $n=n$, $2n^2$ electrons (see Table I.)

Before we turn to the chemical consequences of this law; *i.e.*, the periodic table of elements, we shall see what happens when more than one electron is present per atom.

B. Free Atoms

When a number i of electrons are present in an atom the angular momenta interact with one another. In a weak magnetic field the orbital angular momenta, each of l_i units, may be added vectorially to give a total atomic orbital angular momentum of L units $\hbar$. The value of L may be $0 \rightarrow \sum l_i$. Similarly, a total atomic spin angular momentum of S units is produced by vectorial addition of the spin angular momenta s_i of each electron. $S = 0 \rightarrow \sum s_i$ (integral values) or $\frac{1}{2} \rightarrow \sum s_i$ (half-integral values) for an even and odd number of electrons respectively (Fig. 1(a) and (b) on page 1947).

The total atomic angular momentum obtained by vectorial addition of S and L must amount to J units $\hbar$ ($J = L+S,\ L+S-1 \cdots |L-S|$) (Fig. 1(c)). The laws governing the paramagnetic behavior of atoms is dependent on the strength of this spin-orbit coupling and may be mainly a function of J, or of S and L separately.

* Original manuscript received by the IRE, November 7, 1955.
† Philips Research Labs., N. V. Philips' Gloeilampenfabrieken, Eindhoven, Netherlands.

TABLE I

			Principal quantum number n / Orbital quantum number l / Name of shell	1 / 0 / 1s	2 / 0 / 2s	1 / 2p	3 / 0 / 3s	1 / 3p	2 / 3d	4 / 0 / 4s	1 / 4p	2 / 4d	3 / 4f	5 / 0 / 5s	1 / 5p	2 / 5d	3 / 5f	4 / 5g	6 / 0 / 6s
Z	Element	Symbol																	
1	hydrogen	H		1															
2	helium	He	noble gas	2	0														
4	beryllium	Be		2	↓ 2	0													
10	neon	Ne	noble gas	2	2	↓ 6	0												
12	magnesium	Mg		2	2	6	↓ 2	0											
18	argon	A	noble gas	2	2	6	2	↓ 6	0	0									
20	calcium	Ca	1st transition series	2	2	6	2	6	0	↓ 2									
30	zinc	Zn		2	2	6	2	6	↓ 10	2	0								
36	krypton	Kr	noble gas	2	2	6	2	6	10	2	↓ 6			0					
38	strontium	Sr	2d transition series	2	2	6	2	6	10	2	6	0		↓ 2					
48	cadmium	Cd		2	2	6	2	6	10	2	6	↓ 10		2	0				
54	xenon	X	noble gas	2	2	6	2	6	10	2	6	10		2	↓ 6				0
56	barium	Ba		2	2	6	2	6	10	2	6	10		2	6				↓ 2
57	lanthanum	La	rare earths	2	2	6	2	6	10	2	6	10	0	2	6	1			2
71	lutetium	Lu		2	2	6	2	6	10	2	6	10	↓ 14	2	6	1			2

We see that a great number of energy levels are possible through different possible vectorial additions of the momenta, and even more in a magnetic field. It will be of interest to know the state of lowest energy. This is given by Hund's rules:

1) Coupling of p_{l_i} to p_L and of p_{s_i} to p_S proceeds in such a way that p_L and p_S are maxima. If both cannot have their maximum value because of Pauli's law, p_S is maxima.
2) In each group with equal n and l, the angular momenta p_S and p_L are antiparallel in the first half of the group, and parallel in the second half.

For a strong $p_L - p_S$ coupling the resultant angular momentum p_J will precess around an applied field H so that m_J in the direction of H is again either an integral or a half-integral number of units $\hbar$. The magnetic moments connected with the angular momenta p_L and p_S are $L\mu_B$ and $2S\mu_B$ respectively; the magnetic moment connected with p_J does not lie on the same axis as p_J but precesses around it (see Fig. 1(d)). Fig. 1(d) also shows that $\mu_J = J + S\cos\alpha$ and application of the cosine rule gives

$$\mu_J = \left(1 + \frac{J^2 + S^2 - L^2}{2J}\right)\mu_B \text{ or } \mu_J = g\mu_B.$$

Quantum mechanics teach us that the above expression should be modified in such a way that J^2 must be replaced by $J(J+1)$ or J^{*2}, S^2 by S^{*2}, L^2 by L^{*2}.

As a result, *e.g.*, the expression $p_l = \hbar$ changes to $p_l = \sqrt{l(l+1)}\cdot\hbar$ or $l^*\hbar$, and the expressions for p_s, p_l, μ_l, μ_s, p_L, p_S, p_J, μ_L, μ_S, μ_J change similarly in all cases where there is a precession around a nonfixed axis. The projections of the mechanical and magnetic moments in a fixed direction, *e.g.*, in an applied magnetic field, are not changed. They are $m_J\hbar$ and $gm_J\mu_B$, respectively.

C. The Periodic Table of Elements

The Pauli principle is the basis for the periodic system of the elements; *i.e.*, for the chemical behavior of the elements. Each element is characterized by the positive charge Z of the nucleus, or by the number Z of electrons moving around it. The electron configurations of a few elements are given in Table I, above.

It is seen that up to the element argon ($Z=18$) the orbits are filled up in the order $n=1$, 2, 3, and in each of these, in the order $l=0$, 1, 2; in calcium, 2 electrons are present in the 4s shell, although the 3d shell is still empty. This shell is filled up in the elements between scandium (Sc, $Z=21$) and Zn, which are called the elements of the *first transition series*. A similar transition series occurs between the elements yttrium (Y, $Z=39$) and Cd.

In Ba, two electrons are present in the 6s shell, although the 4f, 5d and 5f shells are still empty. In La, one electron is added to the 5d shell and now only the 4f shell is completely filled up in the elements between Cerium (Ce, $Z=58$) and Lu, which are the rare earth metals.[1]

[1] These data are obtained from spectra; some irregularities found in the electron distributions among 4s and 3d, 5s and 4d, and 5p and 4f shells are disregarded here, because later we shall discuss not atoms but ions, in which these irregularities do not occur.

The electron configuration of the noble gas atoms with two s and six p electrons in the outer shell is generally very stable; therefore ions that have this configuration easily occur in an ionic lattice; *e.g.*, oxygen ($Z=8$) with double negative charge (with Ne configuration), lithium ($Z=3$) with single positive charge (Li^{1+}) ($\approx$He), Mg with double positive charge (Mg^{2+}) ($\approx$Ne), and Al with triple positive charge (Al^{3+}) ($\approx$Ne). The elements of the two transition series generally do not give ions with noble-gas electron configuration, but ions with lower charge. Often one transition element can form ions with different charge, as in the case of manganese: Mn^{2+}, Mn^{3+}, Mn^{4+}, Mn^{5+}, Mn^{6+} and Mn^{7+}.

D. Magnetism in Chemical Compounds

In atoms and ions having only completely filled l shells (the noble gas atoms, the alkali earth metal atoms Be, Ca, etc.) the orbital and spin angular momenta cancel each other, and therefore have $J=0$. They have no permanent magnetic moments but only induced magnetic moments, which are the cause of diamagnetism. Similarly, *e.g.*, in hydrogen molecules (H_2) in which the two electrons are likewise "paired," the resultant angular momentum is zero. The formation of covalent compounds from the atoms usually causes a decided change in the distribution of the electrons and therefore in the magnetic behavior.

In metals, a number of electrons are not present in the energy levels provided by the above picture for free atoms, but move freely through the lattice, and this may give rise to a weak paramagnetism in many cases.

In *ionic compounds*, which are the ones we shall deal with, the ions are subjected to strong electrical fields, but this does not seriously change their electron distribution. Thus sodium chloride, which consists of Na^+ and Cl^- ions, both with noble gas distribution, is diamagnetic. The ions that interest us here are those with incompletely filled $3d$, $4d$ or $4f$ orbits; *i.e.*, those of the first and second transition series and the rare earth metal ions.

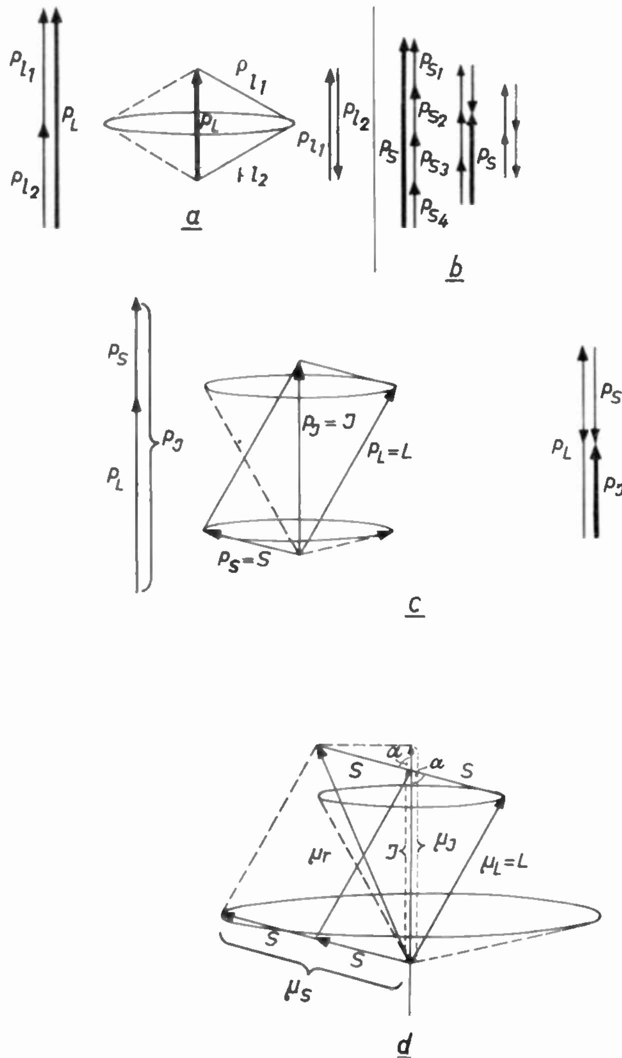


Fig. 1—(a) The three possibilities of vectorial addition of the orbital angular momenta of two electrons, each $p_l = \hbar$, to a total orbital momentum. (b) The three possibilities of vectorial addition of the spin angular momenta of four electrons to a total spin angular momentum. (c) Vectorial addition of the total orbital angular momentum p_L and the total spin angular momentum p_S to a total angular momentum p_J, in units $\hbar$. (d) Vectorial addition of total spin magnetic moment $\mu_S = 2S$ and total orbital magnetic moment μ_L to μ_r: the projection of μ_r on the axis of p_J (see central figure of (c)) is the total magnetic moment μ_J.

Hund's rules (see section *B*) determine L, S and J in these series. As an example the values for the first transition series are given in Table II. These ions have the configuration of the argon atom (see preceding section)

TABLE II

Number of 3d electrons		1	2	3	4	5	6	7	8
		L	S	J	J^*	g^*	g^*J^*	$2S^*$ μ param. in μ_B	$2S$ μ ferrom. in μ_B
0	Ca^{2+}, Sc^{3+}, Ti^{4+}	0	0	0	0	1	0	0	0
1	Ti^{3+}	2	1/2	3/2	1.94	0.80	1.55	1.73	1
2	V^{3+}	3	2/2	4/2	2.45	0.61	1.64	2.83	2
3	Cr^{3+}, Mn^{4+}	3	3/2	3/2	1.94	0.40	0.78	3.87	3
4	Mn^{3+}	2	4/2	0	0	1	0	4.90	4
5	Mn^{2+}, Fe^{3+}	0	5/2	5/2	2.96	2	5.91	5.92	5
6	Fe^{2+}	2	4/2	8/2	4.48	1.50	6.74	4.90	4
7	Co^{2+}	3	3/2	9/2	4.98	1.33	6.62	3.87	3
8	Ni^{2+}	3	2/2	8/2	4.48	1.25	5.60	2.83	2
9	Cu^{2+}	2	1/2	5/2	2.96	1.20	3.55	1.73	1
10	Zn^{2+}	0	0	0	0			0	0

and contain in addition the following number of $3d$ electrons (see Table II).

Surprisingly, the magnetic moments in Bohr magnetons derived from paramagnetic measurements on "dilute" compounds, such as alums $M^{3+}(NH_4)^+(SO_4)_2^{2-} \cdot 12\ H_2O$, do not agree with the g^*J^* values (nor with an expression for no spin-orbit coupling: $4S(S+1)+L(L+1)$), calculated for free ions, but agree fairly well with $2S^*$. In the series of the rare earth ions a similar table can be made. Here, after some necessary corrections, the paramagnetic measurements give values which do correspond with g^*J^*. (The ions with incomplete $4d$, $5d$ and $6d$ shells generally give very much lower magnetic moments than the assumptions made above indicate.)

The cause for this discrepancy is that in the first transition series the $3d$ electrons move near the surface of the ion so that the orbital motion is strongly influenced by the strong periodic electrical fields caused by the adjacent positive and negative charges of ions or water dipoles. In fact the orbital momentum is so strongly coupled to the lattice that practically only the spin momentum contributes to the magnetic moment. Since $\mu_J = g^*J^*\mu_B$ and $p_J = J^*\hbar$, and $\mu_B/\hbar = e/2\ mc$, we have $g^*e/2\ mc = \mu_J/p_J$. The ratio μ_J/p_J can be obtained from measurements of rotation by magnetization or magnetization by rotation. Such measurements give good agreement with the values calculated for compounds of the rare earth ions, but give values $\leqq 2$ for those of the first transition series ions; *i.e.*, ≈ 2 for the first half of the series, but slightly smaller for the second half. These latter values show that there still is some contribution of the orbital momentum, in agreement with the magnetic moments, which are slightly higher than $2S^*$. We shall now introduce a new g factor, $g' = (\mu_S + \mu_L)/p_S$ obtained from the experimental values of the magnetic moments, which are put equal to $g'S$. This latter g factor is slightly above 2 for the second half of the first transition series.

E. Paramagnetic Behavior as a Function of Temperature

We have seen that a magnetic moment μ in a magnetic field H precesses around the direction of H with certain angles α. It can easily be shown that the potential energy then is $E(\alpha) = -\mu H \cos \alpha$. Before the advent of quantum theory α was assumed to have no particular value. In that case the number of electrons in the solid angle $d\omega = 1/2 \sin \alpha d\alpha$ between the cones with angle α and $\alpha + d\alpha$ is

$$dN = Ae^{-E(\alpha)/kT}d\omega \quad (1)$$

(k = Boltzmann's constant).

This can be understood qualitatively according to ordinary Boltzmann statistics, because increasing H tends to decrease the average value of α, ($\bar{\alpha}$), but the thermal motion tends to make $\bar{\alpha}$ larger with increasing temperature T.

A is found by integrating over all values of α; *i.e.*, from

$$N = A2\pi \int_{\alpha=0}^{\alpha=\pi} e^{(\mu H/kT)\cos\alpha} \sin \alpha,$$

and the average moment in the direction of the field, $\mu \cos \alpha$, is found by multiplying both sides of (1) by $\mu \cos \alpha$, eliminating A, and integrating the resulting expression.

The result per mole is the Langevin function L:

$$I = N\mu \mathrm{L}\left(\frac{\mu H}{kT}\right) = N\mu\left(\coth\frac{\mu H}{kT} - \frac{kT}{\mu H}\right). \quad (2)$$

In the new quantum theory α may only have discrete values and the Langevin function changes to a Brillouin function B:

$$\bar{I} = N\mu \mathrm{B}\left(\frac{\mu H}{kT}\right) = N\mu \mathrm{B}\left(\frac{Jg^*\mu_B H}{kT}\right) = N\mu \mathrm{B}(y), \quad (3)$$

in which

$$\mathrm{B}(y) = \frac{2J+1}{2J}\coth\frac{2Jy+y}{2J} - \frac{1}{2J}\coth\frac{y}{2J}\cdot$$

This function may be developed into a series and for small values of H/T we may only use the first power terms in $g^*\mu_B H/kT$:

$$\bar{I} = Ng^*J\mu_B\left(\frac{J+1}{3kT}\cdot g^*\mu_B H\right)$$

$$= Ng^{*2}\mu_B^2\cdot\frac{J(J+1)H}{3kT} \equiv \frac{CH}{T} \quad (4)$$

This gives

$$\chi_{\mathrm{mole}} = \frac{I_{\mathrm{mole}}}{H} = \frac{C}{T}, \quad (5)$$

which is Curie's law found experimentally for many paramagnetic substances.

Many other substances at temperatures which are not too low follow a law $\chi_{\mathrm{mole}} = C_{\mathrm{mole}}/(T+\Delta)$ (in which Δ may be several tens of degrees) from which the magnetic moment may be similarly derived. From the expression (4) we may calculate the magnetic moment $J^*g^*\mu_B$ from the measured susceptibility.

F. Exchange Interaction

So far we have discussed materials in which the magnetic ions are widely separated. In other materials, such as the metals or oxides, the magnetic atoms or ions are very close together. Among these materials we find the ferromagnetics, in which all spin magnetic moments may be parallel, even without an external field. In such materials there is an interaction between the spin moments, "spins," of neighboring atoms or ions. This exchange interaction is connected with the overlap of the orbitals of the outer electrons of two adjacent atoms or ions. This exchange interaction is of the same type as the covalent bond formation in the

hydrogen molecule, H_2, and in many carbon compounds. For atoms with incomplete s and p shells it also gives antiparallel spin orientation. For $3d$ and $4f$ electrons the exchange interaction may be either negative (antiparallel spins) or positive (parallel spins), and according to Slater-Néel the sign is dependent on the distance between the two $3d$ (resp. $4f$) shells; *i.e.*, it changes from $-$ to $+$ for increasing distance. At still larger distances the interaction decreases rapidly so that only near neighbors have to be taken into account.

It may be mentioned that these exchange interactions are very much stronger than the magnetic interaction between the magnetic moments of the ions: the latter can easily be shown to be of the order of a thermal energy kT corresponding at the most to a few °K. We see, therefore, that it cannot explain ferromagnetism at room temperature. Moreover such a dipole-dipole interaction is different in magnitude and sign whether the relative orientation is "abreast" or "astern."

G. Weiss Molecular Field Approximation for Ferromagnetics and Antiferromagnetics

Let us look at a ferromagnetic like iron, the structure of which is given in part in Fig. 2 (disregarding the arrows). Even if the positive exchange interaction between nearest neighbors (empty and hatched spheres) is known, the calculation of the magnetic behavior of the total lattice is a much too complicated statistical problem.

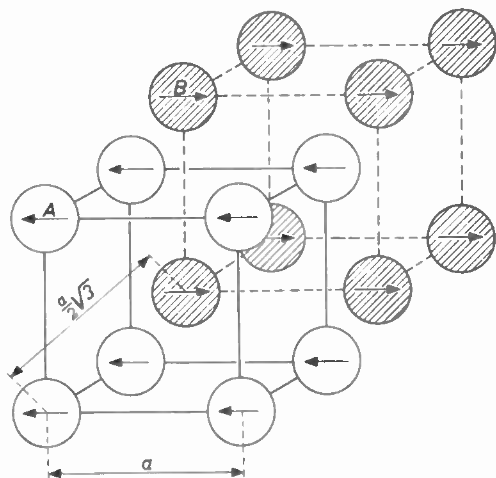


Fig. 2—Part of body-centered crystal lattice; each atom A has eight nearest neighbors B vice versa. Iron and chromium have this crystal structure. The arrows show antiparallel orientation of spins of nearest neighbors (perhaps present in chromium).

One therefore still makes use of an approximational method given by Pierre Weiss in 1907, in which the interactions of the surrounding atoms with a certain atom are represented by a "molecular" field h which is put proportional to the magnetization of the surrounding lattice. For positive interaction we have pro gramatom: $h=\gamma I$, in which γ is a positive constant, so that the effective field this ion feels is $\vec{H}_{\text{effective}}=\vec{H}+\gamma\vec{I}$.

This field is put in the Brillouin function, (section E), *i.e.*, in the Curie law for the paramagnetic behavior, which becomes

$$\vec{I}=C\left(\frac{\vec{H}-\gamma\vec{I}}{T}\right)$$

which gives

$$\frac{I}{H}=\chi_{\text{gramat.}}=\frac{C}{T-\gamma C}=\frac{C}{T-\Theta}. \tag{6}$$

It now appears that there is a temperature below which, even for $H=0$, there is a magnetization I, viz. the Curie temperature $\Theta=\gamma C$. Below this temperature there is a long-range parallel order of the spins, which gives a spontaneous magnetization, *i.e.*, ferromagnetism.

In the case of negative exchange interaction [1] between nearest neighbors in the same structure, we would have $I=C(H-\gamma I)/T$, in which $\gamma=$positive, which would give

$$\frac{I}{H}=\frac{C}{T+\Theta}. \tag{7}$$

Here also a temperature exists below which there is a long-range order of the spins; now however, the spins lie in opposite directions for every pair of neighbors, *i.e.* for the two "sublattices" A and B (empty and hatched), so that no spontaneous magnetization results. Such a material is called *antiferromagnetic*. The temperature at which this order sets in cannot be found without taking into account the interaction between the ions inside each sublattice.

For the paramagnetic behavior of such a material we write

$$\vec{I}_A=\frac{C'}{T}(\vec{H}-\alpha\vec{I}_A-\beta\vec{I}_B) \tag{8}$$

$$\vec{I}_B=\frac{C'}{T}(\vec{H}-\alpha\vec{I}_B-\beta\vec{I}_A) \tag{9}$$

in which I_A and I_B, are the magnetizations per gram-atom of the sublattices A and B, α is the interaction parameter between ions A and A, or B and B, and β between ions A and B.

Adding and rearranging the two equations we obtain for $\vec{I}=\vec{I}_A+\vec{I}_B$:

$$\frac{I}{H}=\frac{2C'}{T+C'(\alpha+\beta)}.$$

or for one gramion

$$\frac{C}{T+\frac{1}{2}C(\alpha+\beta)}=\frac{C}{T+\Theta}. \tag{10}$$

The inverse of (5), (6), and (7), are given by the straight lines in Fig. 3, as $1/\chi=f(T)$: Θ is called the asymptotic Curie temperature.

The temperature below which the spin order given in Fig. 2 (arrows) occurs is, however, not Θ, because

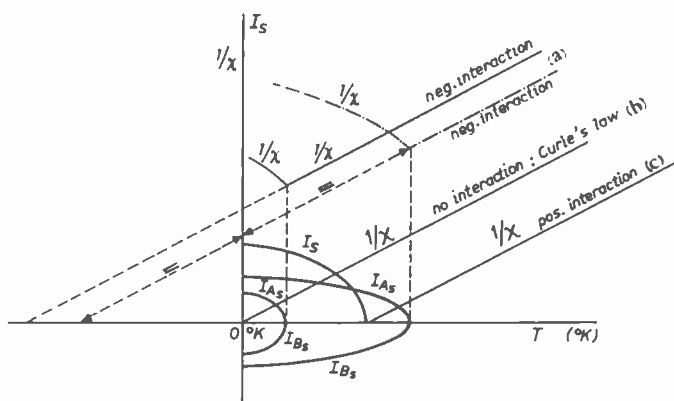


Fig. 3—The inverse of the susceptibility vs temperature (a) for two antiferromagnetics (negative interaction): dash-dot line is the special case of nearest-neighbor-interaction in a body-centered lattice; (b) for a dilute paramagnetic salt which follows Curie's law, and (c) for a ferromagnetic (positive interaction). The thick lines show the antiparallel sublattice spontaneous magnetizations I_{A_s} and I_{B_s} for the antiferromagnetics and the spontaneous magnetization I_s of the ferromagnetic.

we must now find T for which *both* (8) and (9) still have a finite I_A or I_B for $H=0$. Putting $H=0$ we can eliminate I_A and I_B, and find a positive solution $T=(\beta-\alpha)C'$ or $T_N=\frac{1}{2}(\beta-\alpha)C$ (see above).[2]

This point, the Néel temperature, is recognizable from the χ measurement by a break in the $1/\chi$ vs T curve (Fig. 3), and further by a specific heat peak, like that occurring at the ferromagnetic Curie temperature.

II. Behavior of Ferromagnetics and Antiferromagnetics Below the Curie (Néel) Temperature

Below the Curie or Néel points of ferromagnetics or antiferromagnetics the technically available field strength H may be neglected against the very large molecular field h. Only h appears in the Brillouin function (3) and the resulting shapes of the spontaneous magnetization I_s and sublattice magnetizations I_{A_s} and I_{B_s} are given schematically in Fig. 3. (The cause of the change of χ with T below T_N is not discussed here.)

At 0°K all spins in a ferromagnetic are rigorously parallel. Nevertheless, a piece of iron is not a permanent magnet. This is because each crystal is subdivided in *Weiss domains*, inside each of which the spins are parallel, but which have different domain orientations. These orientations are determined by the coupling of the domain magnetization to certain crystallographic directions, to directions of stress, and to the macroscopic shape of the piece of iron.

The energies connected with these anisotropies (to be discussed later) are not large with respect to the magnetization energy ($\frac{1}{2}HI$) in technically available magnetic fields ($\approx$30,000 oersteds), so that at 0°K it is possible to obtain a state of saturation in which all spins throughout the material are parallel. At higher temperatures the spins are not rigorously parallel, since they then precess with specific angles α with respect to H.

[2] Metallic chromium has this crystal structure and is antiferromagnetic; it may have the above spin order.

At higher fields the number of spins in states with small α will be larger, and in extremely high fields (*e.g.*, 10^6 oersteds), of the order of the molecular field h, the saturation magnetization would still be equal to that of 0°K.

In order to measure a value for the spontaneous magnetization I_s inside the Weiss domain in the absence of a magnetic field H, it is customary to measure I as a function of H until the I/H curve is a straight line, and to extrapolate towards $H=0$ along this line, since the paramagnetic susceptibility due to the alignment of the spins against thermal agitation will follow a Curie law. In order to find exact Curie temperatures other methods have to be used [2].

The interesting unit in which to express the saturation magnetization is the Bohr magneton per formula unit or "molecule."

The spontaneous magnetization is:

I_s	in cgs units per cm³,
$\frac{I_s}{d}$	per gram (d = specific weight),
$\frac{I_s}{d}\cdot A$	per mole or gramatom (A = molecular weight, or atomic weight respectively),
$\frac{I_s}{d}\cdot\frac{A}{N}$	in cgs units per formula unit or "molecule" (resp. atom) (N = Avogadro's number = number of formula units per mole, or atoms per gramatom), or
$\frac{I_s}{d}\cdot\frac{A}{N\mu_B}$	Bohr magnetons per formula unit ($N\mu_B=5585$). (10)

For the spontaneous magnetization in Bohr magnetons per atom or ion at 0°K, hereafter called the *saturation moment*, one should expect a value $2S$ if there is no orbital contribution, or gS if there is an orbital contribution: g is defined as

$$g=\frac{2mc}{e}\cdot\frac{\mu_S}{p_S} \tag{11}$$

Exchange interaction is to be expected in the first place in those materials where the magnetic atoms touch each other; *i.e.*, in the metals. It must be remembered that in the ferromagnetic metals with incomplete $3d$ shells (iron, cobalt and nickel) the metallic bond is formed by electrons which move freely through the lattice. For free atoms we would expect saturation moments for Fe: 4 μ_B, Co: 3 μ_B, Ni: 2 μ_B, (see Table II, column 8) but one actually finds: Fe: 2.22 μ_B, Co: 1.71 μ_B, Ni: 0.6 μ_B.

Although there are different theories to account for this discrepancy, a current picture for cobalt is that there are only 0.7 electrons in the conduction band, (an equal number with + and −spins) and an average of 8.3 electrons in the $3d$ band of which 5 have +spins and an average of 3.3. have −spins. Similar models have been proposed for Fe and Ni.

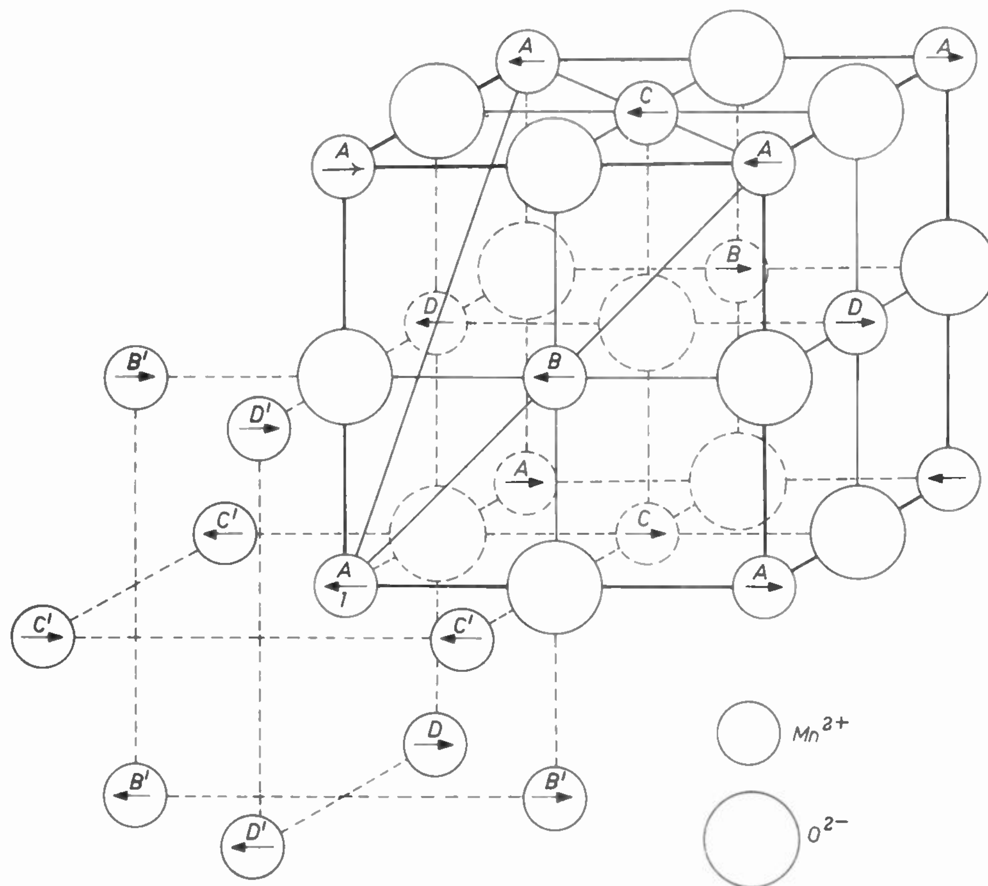


Fig. 4—Part of the crystal structure of MnO; the Mn^{2+} ions form a face-centered cubic lattice. Each A ion has four nearest neighbors belonging to each of the B, C, and D sublattices, and six next-nearest neighbors belonging to the A sublattice. The arrows indicate the arrangement of the spins deduced from neutron diffraction experiments.

I. Superexchange Interaction

As we have seen, ionic bond formation does not generally alter the electron configuration of the ions. Since Hund's rule gives us the quantum number S, we should therefore expect a much simpler behavior for ionic ferromagnetic compounds like the ferrites which we are going to discuss, which contain ions of the first transition elements. Thus for a material like the mineral magnetite, Fe_3O_4, which may be taken to consist of 1 Fe^{2+}, 2 Fe^{3+} and 4 O^{2-} ions, a saturation moment of $4+5+5 = 14\mu_B$ is expected.

The experimental value of Weiss and Forrer (1929) however, is, 4.08 μ_B. This baffling discrepancy has remained unexplained for 19 years. Néel in 1948 [3] showed that the exchange interactions in this material, and in all ferrites, are negative. In order to understand how this can result in a ferromagnetic material we shall study the crystal structure of these materials (Part II, section A) and show what magnetic properties were predicted by Néel by the application of the Weiss molecular field theory, and how these were confirmed by experiment.

Before doing so, however, we shall first turn our attention to another antiferromagnetic oxide, manganous oxide, MnO, the crystal structure of which X-ray diffraction[3] has shown to be that in Fig. 4 (without arrows), in which Mn^2 and O^{2-} ions alternate in endless rows in three perpendicular directions. Unlike the case of Fig. 2, here it is impossible for an ion (*e.g.*, $A1$ in Fig. 4) to have a spin direction antiparallel to that of all its nearest neighbors, since a third of its B, C and D neighbors (at the distance $\frac{1}{2}a\sqrt{2}$) must have parallel spins to give equal number of west and east spins in the lattice. Among all possible spin orders there is however one in which the spin of one ion (*e.g.*, $A1$) is antiparallel to those of all its next-nearest neighbors (A) at a distance a, and the spin order given by the arrows is actually that found by neutron diffraction [4].[4] The spin orientations of B, C, and D may be independent of that of A_1, but according to the neutron diffraction results are parallel in the oblique plane shown in

[3] X-rays have wavelengths of the same order of magnitude as the interatomic distances in crystals ($\approx 10^{-8}$ cm). Since the lattice is strictly periodic, in the diffracted beams sharp reflexions result from which the crystal structure can be obtained. X-rays are diffracted by the electrons, and will indirectly give the locations of the nuclei since the electron densities are greatest near the nuclei.

[4] Neutron beams having wavelengths of the order of the interatomic distances are scattered by the atomic nuclei, which gives the same diffraction pattern as X-rays, and by the magnetic moments of the unpaired electrons. These two scattering amplitudes are added so that the scattering power of a Mn^{2+} ion with spin pointing west is different from one with spin pointing east. Below the Néel temperature, therefore, antiferromagnetics show extra reflections, from which the spin orientation can be found. This possibility was first suggested by J. S. Smart [4].

Fig. 4.[5] But now the Mn^{2+} ions A are separated by the O^{2-} ions, and the question arises how the electron densities can possibly overlap so as to give exchange interaction.

Kramers [5] as early as 1934 showed theoretically the possibility of indirect exchange interaction via anions, and Anderson [6] in 1950 worked out a theory; the main results are briefly given here. To the ground state (state of lowest energy) of a pair of ions $Mn^{2+}O^{2-}$:

$$\underset{3d^5}{\overset{Mn^{2+}}{\rightarrow\rightarrow\rightarrow\rightarrow\rightarrow}}\ \underset{2p^2}{\overset{O^{2-}}{\rightleftarrows}}\ \underset{2p^2}{(\rightleftarrows)}\ \underset{2p^2}{(\rightleftarrows)}$$

is admixed a state of higher energy in which one of the p electrons of the O^{2-} has gone over to the Mn^{2+} ion:

$$\underset{3d^6}{\overset{Mn^{2+}}{\rightleftarrows\rightarrow\rightarrow\rightarrow\rightarrow}}\quad \underset{2p^2}{\rightarrow}\ \overset{O^{1-}}{\underset{2p^2}{(\rightleftarrows)}}\ \underset{2p^2}{(\rightleftarrows)}$$

in which the electron enters the $3d$ orbit with antiparallel spin according to Hund's rule. The electron density of a p orbit has dumbbell shape: ∞, and thus the unpaired p electron of the O^- ion can give negative exchange interaction (*e.g.*, give a covalent bond) with the Mn^{2+} ion on the other side of the dumbbell so that we obtain

$$\overset{Mn^{1+}}{\rightleftarrows\rightarrow\rightarrow\rightarrow\rightarrow}\quad \overset{O^{1-}}{\rightleftarrows}\quad \overset{Mn^{2+}}{\leftarrow\leftarrow\leftarrow\leftarrow\leftarrow}.$$

The mixture of the two states has lower energy than the original ground state according to quantum mechanics, and the above-mentioned valence bond lowers the energy even more; the latter decrease ϕ is the superexchange energy.

The net result was a negative indirect interaction in colinear $Mn^{2+}-O^{2-}-Mn^{2+}$ configurations. In Anderson's theory the indirect or *superexchange* interaction is strongest for such colinear configurations and weakest for right-angle configurations

$$\begin{matrix} Mn^{2+}\text{—}O \\ \quad\quad | \\ \quad\quad Mn^{2+} \end{matrix}$$

This agrees very well with the findings in ferrite-like materials, as we shall show, but not in MnO, where a molecular-field treatment similar to that given in Part I, Section G would deduce from the experimental ratio Θ/T_N a ratio between α (AA interaction parameter) and β (AB, AC or AD interaction parameter) equal to $\alpha/\beta=\frac{3}{4}$, instead of $\alpha/\beta\gg 1$ [6]. According to the above picture superexchange between ions with less than half-filled $3d$ shells should have a positive sign, because the p electron can now enter the $3d$ orbit with parallel spin (Hund's rule), but many oxides containing such ions are antiferromagnetic; *e.g.*, $Cr_2^{3+}O_3^{2-}$ [7] and $Ca^{2+}Mn^{4+}O_3^{2-}$ [8]. In solid solutions of $La^{3+}Mn^{3+}O_3^{2-}$ and $Ca^{2+}Mn^{4+}O_3^{2-}$ which have a (perovskite) structure in which only colinear configurations Mn—O—Mn occur, positive interaction, or real ferromagnetism, has indeed been found [8]. For such a case, where the same metal ion occurs with different valency states, (which according to de Boer and Verwey [9] is the condition for conductivity in ionic semiconductors), Zener [10] has proposed a different exchange mechanism, "double exchange," in which an electron jumps from a Mn^{3+} ion to a adjacent O^{2-} ion, while simultaneously an electron with parallel spin jumps from the O^{2-} ion to the Mn^{4+} ion. The energy is lower for parallel orientation of the ionic moments than for antiparallel orientation, the difference being twice the exchange energy. J. Smit has given another type of exchange for these materials [11].

[5] This makes the magnetic structure non-cubic and should make the crystal structure non-cubic; a small deviation from the cubic structure has indeed been found by means of X-ray diffraction.

Since no oxides are known which show positive exchange interaction, but no conductivity, it cannot be said with certainty that the Anderson superexchange can yield a positive interaction. We can still summarize the cases discussed above as follows:

Exchange Interaction	Positive	Negative
direct	Fe	Cr
indirect	$La_{0.65}^{3+}Ca_{0.35}^{2+}Mn_{0.65}^{3+}Mn_{0.35}^{4+}O_3^{2-}$	$Mn^{2+}O^{2-}$

Recently modified versions of Anderson's theory have been suggested [12].

II. Factors Determining the Saturation Magnetization of Spinels [14]

A. The Crystal Structure of Ferrites ("Spinel" Structure)

The crystal structure of the ferrites [13], *e.g.*, $Fe^{2+}Fe_2^{3+}O_4^{2-}$, is more complicated than the ones previously depicted (Figs. 2 and 4), and is shown in Fig. 5.

Fig. 5—The total cube represents the unit cell of the spinel lattice which consists of eight small cubes (octants). The positions of the ions are different in two octants sharing a cube face and are only shown for two such octants. A and B ions are metallic ions surrounded by four and six oxygen ions (large circles) respectively.

The total cube is the *unit cell*, the smallest single type "brick" from which the structure can be built by stacking them in three perpendicular dimensions. This unit

cell consists of 8 smaller cubes "octants," of which there are two different types. Two octants sharing a face are different, those sharing an edge are equal. Both types of octants contain 4 O^{2-} ions about half-way on the body diagonals and 4 metal ions A in alternate corners; the left octant contains one metal ion A in its center, the right one, 4 metal ions B half-way on the body diagonals. Considering that the ions lying in the faces of the unit cell belong to two adjoining cells, and those in the corners to eight cells, the total number of ions in the unit cell is 4 A ions, 8 B ions and 32 oxygen ions, corresponding to eight formula units of *e.g.*, $MgAl_2O_4$ (the mineral *spinel*, from which the structure derives its name).

The differentiation between A and B ions is made because the A ions are surrounded by 4 oxygen ions forming a regular tetrahedron, and the B ions by 6 oxygen ions forming a regular octahedron. This is important because:

a) some metallic ions will preferably occupy tetrahedrally surrounded sites, others octahedrally surrounded ones;
b) the magnetic ions in A and B sides have the strongest tendency to have mutually antiparallel spin orientation, which we shall explain now.

In Fig. 6 we show the most important configurations of A and B ions with respect to the oxygen ions. In (b),

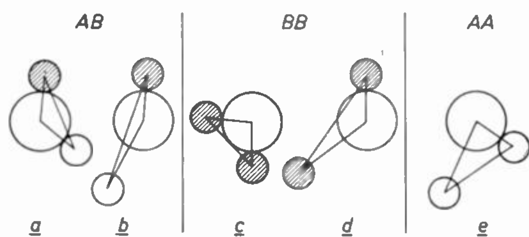


Fig. 6—Some of the angles metallic ion-oxygen ion-metallic ion occurring in spinels. Small open and hatched circles are metallic ions in A and B sites respectively.

(a) and (d) the angle M–O–M (M=metal ion) is favorable for superexchange, but in (b) and (d) one distance M–O is large, and since the exchange interaction depends on the overlap of the electron densities of the adjoining ions, this distance is unfavorable. In (c) both distances are favorable, but the angle is unfavorable, whereas in (e) both angle and one distance are unfavorable. Thus we see that the AB interaction will be stronger than the BB and AA interactions—a fact recognized by Néel before Anderson's theory had been worked out—and that the AA interaction will be weakest.

Since it influences the angles, it is of some interest to note that the oxygen ions are situated a little further away from the A ions than half-way on the body diagonals of the octants. As a result, the octahedron around the B ions is slightly distorted.

B. Solid Solutions of Spinels

The number of different metallic ions that can occur in the spinel lattice is very great, the primary condition being that they are small enough to keep the oxygen ions near to each other, since the building would become unstable otherwise. The radius of the oxygen ions is 1.32 Å (1 Å = 1 Angstrom unit = 10^{-8} cm), and the radii of the metallic ions occurring in spinels lie, with a few exceptions, between 0.5 and 1 Å. The charge of the ions may be anything between 1 and 4 and may be even 6. A few of these ions with their radii in Å are listed in Table III, below.

Spinels containing only two types of ions may have formulas like the following:

valencies metallic ions:

(a) 1, 3 $Li_{0.5}^{I}Fe_{2.5}^{III}O_4$; $Li_{0.5}^{I}Al_{2.5}^{III}O_4$.
(b) 1, 4 $Li_{4/3}^{I}Ti_{5/3}^{IV}O_4$.
(c) 1, 6 $Ag_2^{I}Mo^{VI}O_4$; $Na_2^{I}W^{VI}O_4$: Na^{I}, Ag^{I}, Mo^{VI} and W^{VI} occur only in such 1.6 compounds.
(d) 2, 3 $Mg^{II}Al_2^{III}O_4$; $Mg^{II}Fe_2^{III}O_4$; $Zn^{II}Fe_2^{III}O_4$, $Co^{II}Fe_2^{III}O_4$, $MnCr_2O_4$ and generally aluminates, chromites and ferrites of the divalent metals listed.
(e) 2, 4 $Mg_2^{II}Ti^{IV}O_4$ and other titanates.

When metal ions are present which can occur in different states of valency (Fe^{2+} and Fe^{3+}; Co^{2+}, Co^{3+}, Co^{4+}; Mn^{2+}, Mn^{3+}, Mn^{4+}; Ti^{3+}, Ti^{4+}) it is possible by chemical analysis to determine only the oxygen excess or shortage with respect to a certain valency state; *e.g.*, Fe^{3+}, Co^{2+}, Mn^{2+}, Ti^{4+}, but not whether Mn_3O_4 is $Mn^{II}Mn_2^{III}O_4$ or $Mn_2^{II}Mn^{IV}O_4$ (or likewise for Co_3O_4) or whether $MnFe_2O_4$ is $Mn^{II}Fe_2^{III}O_4$ or $Mn^{III}Fe^{II}Fe^{III}O_4$, or whether Fe_2TiO_4

TABLE III

Valency 1		2		3		4		6	
Li^+	(0.78)	Mg^{2+}	(0.78)	Al^{3+}	(0.57)	Ti^{4+}	(0.69)	Mo^{6+}	(0.62?)
Na^+	(0.98)								
Ag^+	(1.13)			Cr^{3+}	(0.64)				
		Mn^{2+}	(0.91)	Mn^{3+}	(0.70)	Mn^{4+}	(0.52)		
		Fe^{2+}	(0.83)	Fe^{3+}	(0.67)				
		Co^{2+}	(0.82)						
		Ni^{2+}	(0.78)						
		Cu^{2+}	(0.85)						
		Zn^{2+}	(0.82)						
		Cd^{2+}	(1.03)						

is $Fe_2^{II}Ti^{IV}O_4$ or $Fe^{II}Fe^{III}Ti^{III}O_4$. For $MnFe_2O_4$ the former formula may be deduced from the saturation moment (see Table IV), for Fe_2TiO_4 the former formula is more likely in view of the analogy with $Mg_2^{II}Ti^{IV}O_4$. It seems unlikely, however, that two metals in one compound will both be present in more than one valency state: *e.g.*, the states $Mn^{3+}+Fe^{2+}$ and $Mn^{2+}+Fe^{3+}$ in the lattice will have different energies, and because the change of the one ion pair into the other involves, roughly speaking, only the transfer of an electron, this will occur very easily. The spinels containing mono-, di-, tri-, and tetravalent metallic ions can form solid solutions with each other in any proportions, provided the number of ions is 3 [6] and the total number of positive charges is 8. Such solid solutions exist because, apart from the energy of the system, the probability of a mixture on an atomic scale (a solid solution) of two simple spinels is greater than that of a mixture of crystals of the simple spinel and crystals of the other simple spinel. In thermodynamic language: the free energy $F=U-TS$ tends to be a minimum. (U is the energy, and S the entropy, which is connected with the above probability.) At high temperature TS is predominant, at lower temperatures, U. At 0°K, therefore, no solid solution is really stable, but in practice they can exist down to 0°K, because the movement of cations through the lattice, necessary for separation into separate crystals of two simple spinels, is difficult. This is because the ions have to slip through narrow passages where they are subjected to strong repulsive forces, so that the diffusion rate, the number of ions changing position per second, is large only at high temperatures. Therefore for a given cooling rate, a temperature is reached at which we do not allow sufficient time for equilibrium (the state of lowest free energy) to be maintained.

For spinel solid solutions this temperature usually lies above that at which the solid solution becomes energetically unstable, but sometimes is below it. In the series $CoFe_2O_4$—Co_3O_4 and $MgFe_2O_4$—$MgAl_2O_4$ a "miscibility gap" occurs, which at higher temperatures becomes narrower and finally disappears.

Solid solutions may occur with "compounds" which either do not exist in the free state, or have a different crystal structure: $[Na_{0.5}Fe_{2.5}O_4]$ does not exist,[7] but may give solid solutions with $Li_{0.5}Fe_{2.5}O_4$, or in other words: part of the Li^+ ions ($\approx 40\%$) in this compound may be replaced by Na^+ ions. $CaFe_2O_4$ does exist, but has a completely different, unknown crystal structure. In $ZnFe_2O_4$ up to 35 per cent of the Zn^{2+} ions can be replaced by Ca^{2+} ions at high temperatures: separation into $ZnFe_2O_4$ (still containing some Ca) and $CaFe_2O_4$ crystals (containing some Zn) is prevented by rapid cooling ("quenching") from a high temperature (*e.g.*, 1,250°C [14].

Some spinels have a slightly elongated unit cell, with an axial ratio (height:width) of *e.g.*, 1.16 (Mn_3O_4), or 1.06 ($CuFe_2O_4$). These tetragonal spinels with cubic spinels form solid solutions in which the axial ratio decreases with the ratio tetragonal:cubic spinel and is unity for about a 1:1 ratio. In several cases a miscibility gap has been found.

What is usually called a non-stoichiometric ferrite is a solid solution $M^{II}Fe_2^{III}O_4$—Fe_2O_3 which might be regarded as a solid solution with a cubic form of $Fe_2^{III}O_3$, "γ-Fe_2O_3" which has the spinel structure and has been assigned a formula $Fe_{8/3}^{III}\square_{1/3}O_4$ [15] in which $\square$ is a cation vacancy. Such solid solutions again follow the general rule of greater solubility at higher temperature and on cooling slowly crystals of α-Fe_2O_3 with another, hexagonal, crystal structure are formed between the spinel crystals. Such a second phase may be detected by the appearance of its own specific reflections in the X-ray diffraction pattern, or by observing a polished surface of the material under a microscope in reflected light (Fig. 7(d)). It is difficult to decide whether a small

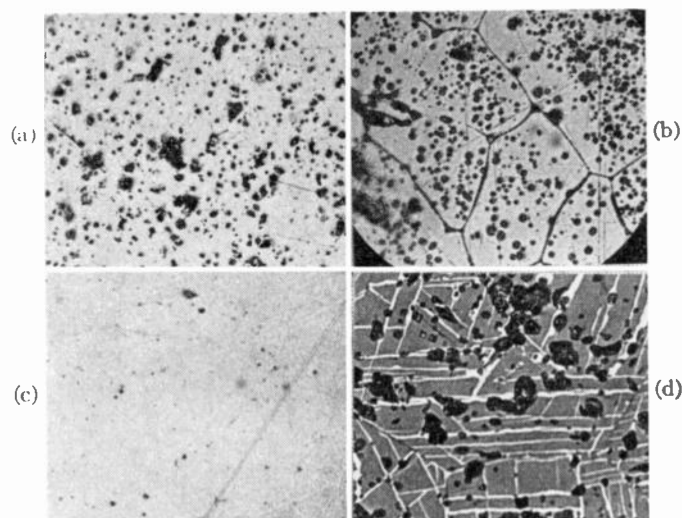


Fig. 7—Photomicrographs of polished surfaces of ferrites observed in reflected light. (400×) (a) Ordinary sintered NiZn ferrite; the dark areas are air pores. (b) A similar ferrite sintered at a high temperature; there are fewer pores, and they are more spherical. (c) Very dense ferrite for square-loop. (d) Ferrite showing a separation of thin plates of α-Fe_2O_3, which has stronger reflectivity.

excess of MO ($M=Mn^{2+}$, Ni^{2+} etc.) can dissolve in the spinel lattice of a ferrite MFe_2O_4 since all these oxides, except ZnO, have the structure given in Fig. 4. The edge of their unit cell in many cases is very nearly equal to half that of the spinel unit cell, so that the X-ray reflections practically coincide with some of the spinel reflections, and therefore only alter their intensities. Since simple spinels have unit cell edges of the order of 8 Å, which differ by several times 10^{-2} Å, whereas cell edges may be determined with an accuracy of some units 10^{-2} Å, a sensitive way to detect solid-solution formation is to look for a change of cell edge with composition.

In some cases no solid solution of MO in MFe_2O_4 was found (*e.g.*, in $ZnFe_2O_4$), but since the cell edge changes appreciably with heat treatment, even for one exact

[6] See, however, below, for solid solutions with Fe_2O_3.

[7] A mixture of $Na_2CO_3+5Fe_2O_3$ heated to complete reaction gives $2NaFeO_2$ and leaves $4Fe_2O_3$ unused.

composition, this method must be used with extreme care.[8] If the material is allowed to take up oxygen from the atmosphere in which it is heated, MnO may dissolve in $MnFe_2O_4$ as Mn_3O_4. (See Part III.)

C. Cation Distribution in Oxides with Spinel Structure

Eight formula units $MM_2'O_4$ of a binary spinel like $MgFe_2O_4$ must occupy the sites of a unit cell $A_8B_{16}O_{32}$. It appeared logical to place the M ions on *A* sites, the M_2' ions on *B* sites. We shall call this the "normal" arrangement and write the ions in *B* sites between brackets $Mn^{II}_{(A)}\,[Al^{III}]_{(B)}O_4$ or simply $Mn[Al_2]O_4$.

Among the ferrites, only $Zn^{II}[Fe_2^{III}]O_4$ and $Cd^{II}[Fe_2^{III}]O_4$ are normal, the others are all inverse: $Fe^{III}[Mn^{II}Fe^{III}]O_4$, $Fe^{III}[Fe^{II}Fe^{III}]O_4$, $Fe[Co\,Fe]O_4$, $Fe[Ni\,Fe]O_4$; $Fe[Cu\,Fe]O_4$[9] and $Fe(Mg\,Fe)O_4$. These results were first obtained directly from X-ray diffraction intensities for Zn-, Cd, Mg and Cu ferrites [16], and indirectly from a comparison of cell edges with those of corresponding chromites or aluminates for the others [17]. Although for every spinel $MM_2'O_4$ there will be an energy difference between the normal and inverse arrangements $M[M_2']O_4$ and $M'[MM']O_4$, at low temperatures either the one or the other is stable. Higher temperatures will tend to cause intermediate arrangements $M'_{1-x}M_x[M_{1-x}M'_{1+x}]O_4$ to be formed, with $x=1/3$ for a hypothetical completely random arrangement at extremely high temperature, and x changing towards 1 or 0 for normal or inverse spinels with decreasing temperature.

The values $x=0$ and $x=1$ are only obtained if equilibrium is maintained down to the lowest temperature; in practice, with practical cooling rates (down to $15°/h$) the diffusion between *A* and *B* sites stops at, say, 300–600°C. For $MgFe_2O_4$ the inversion parameter has been determined by X-ray [18] and magnetic [19] measurements down to about 400°C: it follows a Boltzmann distribution law:

$$\frac{x(1+x)}{(1-x)}2 = e^{-E/kT}.$$

D. Specific Preference of Cations for A and B Sites

We have seen that many ions may occur in both *A* and *B* sites. This is specially true for ions with noble-gas configuration: Li^+, Mg^{2+}, Al^{3+}, Ti^{4+}, and those with a half-filled *3d* orbit: Mn^{2+} and Fe^{3+}. In these cases the most stable cation distribution is determined by their size and charge, as can be calculated from an electrostatic model of charged spheres [20]. Apart from this there are specific effects: Zn^{2+} and Co^{2+} have equal radii but $Zn[Fe_2]O_4$ is normal, $Fe[CoFe]O_4$ inverse. Such specific effects are the formation of covalent bonds of Zn^{2+} and Cd^{2+} involving their two 4*s* or 5*s* electrons respectively and six 2*p* electrons of the oxygen, giving four bonds directed towards the four corners of a tetrahedron. This means they will tend to occur in *A* sites.

Among the other ions only Ni^{2+} and Cr^{3+} have a *strong* preference for octahedral surroundings; *i.e.*, *B* sites. This is connected with the electric fields created by the neighboring ions, which influence the energy levels and the spatial distribution of the *3d* electrons [21].

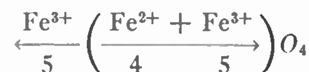
The other transition ions have weaker specific preferences for *A* or *B* sites, which have the same origin. In solid solutions containing three or more different ions, those ions which have specific preferences for either *A* or *B* sites, will still all be present in these sites; *e.g.*, Zn^{2+} in *A*, Ni^{2+} in *B* sites. Those without specific preferences will generally be found in both sites also when two of these are present, because a higher-temperature equilibrium is frozen in.

This distribution is influenced greatly by the presence of other ions which do have a specific preference, as Ni^{2+} and Cr^{3+}. In $Mg[MgTi]O_4$ all Ti^{4+} is reported to be in the octahedral *B* sites, and no substance with Ti^{4+} surrounded by 4 oxygen ions had yet been reported. In $Fe_{0.7}Ti_{0.3}[Ni_{1.5}Fe_{0.3}Ti_{0.2}]O_4$, however, the larger part of the Ti^{4+} ions is present in *A* sites, because of the presence of the Ni^{2+} ions in *B* sites [14]. In $Fe[Li_{0.5}Fe_{1.5}]O_4$ all Li^+ ions are in *B* sites; if we replace 1.5 Fe^{3+} by 1.5 Cr^{3+} we have a distribution $Fe_{0.8}Li_{0.2}[Li_{0.3}Fe_{0.2}Cr_{1.5}]O_4$ [14]. The possible causes for such "anomalous" distributions are discussed elsewhere [14].

The cation distribution may be deduced directly from magnetic saturation measurements combined with the effective *g* factor obtained from microwave resonance measurements and from X-ray diffraction patterns and neutron diffraction patterns.[10]

E. Some Results of Néel's Molecular Field Theory for Ferromagnetic Spinels

Using the rule for the angle-dependence of the superexchange interaction, we can understand that the *AB* interaction is larger than the *AA* and *BB* interactions. Using the experimental crystallographic data of section *C* we can now understand why



has 4 μ_B per formula unit. This behavior is called non-compensated antiferromagnetism or *ferrimagnetism*.

We shall now give the main results of molecular-field theory applied by Neél to the materials like spinels [3]. Let us take a spinel $Fe_{x_A}, M_{1-x_A}, [Fe_{x_B}, M_{1-x_B}]O_4$ in which M is a nonmagnetic ion and x_A and x_B are the

[8] This is due to a change in cation distribution (see section *F*).

[9] Mainly inverse, as has since been shown.

[10] The scattering power for X-rays depends on the number of electrons *Z*; for the different transition metals the difference in scattering power is therefore small. The scattering power of the atomic nuclei for neutrons may be widely different for metals with almost equal *Z*, so that for crystallographic problems involving materials which contain such metal ions, neutron diffraction is a powerful, although so far scarcely used, tool. The cation distribution in, *e.g.*, $MnFe_2O_4$, where all other methods fail, can be found by neutron diffraction.

number of ferric (Fe^{3+}) ions on A and B sites respectively. We now have to consider *three* exchange interactions; *i.e.*, three interaction parameters, which we call $-n$ for AB, αn for AA, and βn for BB interactions.

For the paramagnetic behavior we have

$$\vec{I} = x_A \vec{I}_A + x_B \vec{I}_B \tag{12}$$

$$\vec{h}_A = n(-x_B \vec{I}_B + \alpha\, x_A \vec{I}_A) \tag{13}$$

$$\vec{h}_B = n(-x_A \vec{I}_A + \beta x_B \vec{I}_B) \tag{14}$$

$$\vec{I}_A = \frac{C}{T}(\vec{H} + \vec{h}_A) \tag{15}$$

$$\vec{I}_B = \frac{C}{T}(\vec{H} + \vec{h}_B) \tag{16}$$

in which the molecular fields h_A and h_B and the sublattice magnetizations I_A and I_B have the same meaning as in Part I section G. Elimination of h_A, h_B, I_A and I_B from (13–16) gives a form

$$\frac{H}{I_{\text{mole}}} = \frac{1}{\chi_{\text{mole}}} = \frac{T}{C_{\text{mole}}} + \frac{1}{\chi_0} - \frac{s}{T-\theta} \tag{17}$$

in which $1/\chi_0$, s and θ are functions of x_A, x_B, α, β, and n. This expression (17) represents a hyperbolic curve for $1/\chi$ vs T, and the paramagnetic measurements on ferrites do indeed give such curves (Fig. 11).

The behavior below the Curie temperature is given by

$$I_s = x_B I_{B_s} - x_A I_{A_s} \tag{17a}$$

(the suffix s means saturation)

$$\vec{h}_A = n(-x_B \vec{I}_{B_s} + \alpha x_A \vec{I}_{A_s}) \tag{18}$$

$$\vec{h}_B = n(-x_A \vec{I}_{A_s} + \beta x_B \vec{I}_{B_s}) \tag{19}$$

$$I_{A_s} = NgS\mu_B \cdot \mathrm{B}_s \frac{(gS\mu_B h_A)}{kT} \tag{20}$$

$$I_{B_s} = NgS\mu_B \cdot \mathrm{B}_s \frac{(gS\mu_B h_B)}{kT} \tag{21}$$

in which B_s is the Brillouin function of Part I section E. In these Brillouin functions H may be neglected with respect to h_A and h_B.

Néel from these equations deduced the different possible I_s vs T curves when x_A/x_B is kept constant and α and β are varied. Since the author is a chemist and is interested in the behavior of a series of solid solutions in which x_A and x_B are varied, he gives a picture of Néel's results in such a way that x_A/x_B is varied and α and β are kept constant (Fig. 8). At 0°K (where $I_{A_s} = I_{B_s}$) $I_s = 0$ when $x_A = x_B$. At higher temperatures this is only true if $h_A = h_B$, and this is only true if $\alpha = \beta$. This would be a very rare coincidence, so that at higher temperatures $I_s \neq 0$, see curve (L) (Fig. 8). Near the absolute zero and

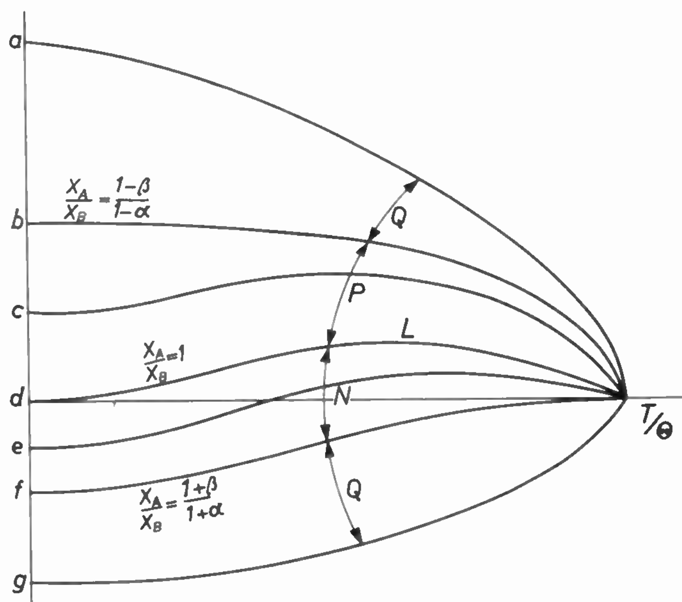


Fig. 8—Theoretical spontaneous-magnetization vs temperature curves for a varing ratio of ferric ions in A and B sites (x_A and x_B respectively). The capitals refer to Néel's nomenclature.

Curie temperatures simplified expressions may be used for the Brillouin curve, which give $dI_s/dT = 0$ near 0°K for

$$x_A/x_B = \frac{1-\beta}{1-\alpha},$$

and $dI_s/dT = 0$ near the Curie temperature for

$$x_A/x_B = \frac{1+\beta}{1+\alpha},$$

giving curves (b) and (f), respectively. Between the "normal" curves (a) and (g) a number of strange curves are seen among which (P) and (N) are the strangest of all. The series of solid solutions $NiFe_2O_4$—$NiFeAlO_4$ (Fig. 14) is an example of a series where all these strange curves indeed occur.

Néel found from his theory that a variation of the parameters α, β, x_A and x_B, gives four conditions for which the energy W of the molecular field is minima, even at 0°K.

a) Both A and B sublattices are saturated.
b) I_{A_s} and I_{B_s} are 0. ($W = 0$)
c) The A sublattice is saturated but the B sublattice is not. (from $dW/dI_{B_s} = 0$)
d) The B sublattice is saturated, but the A sublattice is not. (from $dW/dI_{A_s} = 0$).

A nonsaturated sublattice means that the spins inside the sublattice are not parallel. Yafet and Kittel [22] have shown theoretically that the lowest energy for nonparallel spins inside a sublattice occurs when these are ordered in 2 "face-centered" sublattices A' and A'', and four face-centered sublattices B^1, B^2, B^3, B^4, inside each of which the spins are parallel.

For the theory it is sufficient to use only two sub-

lattices B' and B''; *i.e.*, a plane projection of B^1, B^2, B^3 and B^4. The equilibrium between the tendency of I_{B_s}' and I_{B_s}'' to be both antiparallel to I_{A_s} (interaction parameter $=-n$), and the tendency to be mutually antiparallel because of the $B'B''$ interaction (interaction parameter $=n\gamma_2$) results in an angle between I_A and I_{B_s}' (and I_{B_s}'') (Fig. 9 (a)).

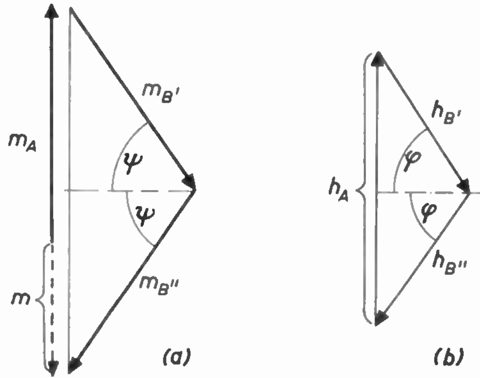


Fig. 9—Arrangement of the sublattice magnetizations (*a*) and of the molecular fields (*b*) in a ferrite with comparable *AB* and *BB* interactions.

We have $\vec{h_A}+\vec{h_B''}=\vec{h_B'}$, and since the molecular fields have the same directions as the magnetizations and h_B'' must be equal to h_B', we may substitute

$$n(x_A\vec{I_{A_s}} + \gamma_2 x_B \vec{I_{B_s}''}) = n\gamma_2 x_B \vec{I_{B_s}'}$$

and

$$\sin\psi = h_A/2h_B' = -\ x_A I_{A_s}/2x_B I_{B_s}'\gamma_2. \qquad \text{(Fig. 9(b))}$$

Since

$$I_s = 2x_B I_{B_s}' \sin\psi - x_A I_{A_s} \qquad \text{(Fig 9(a))}$$

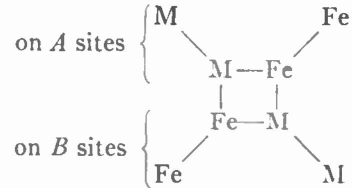
we have

$$I_s = -\ x_A I_{A_s}\left(1+\frac{1}{\gamma_2}\right). \qquad (22)$$

Thus, we find, surprisingly, that for the case of strong $B'B''$ interaction I_S is independent on I_{B_s} (and similarly for strong $A'A''$ interaction independent of I_{A_s}). The magnetizations for Néel's four cases of minimum energy, in Yafet and Kittel's interpretation, are shown in Fig. 10. These may, according to the theory, occur in one spinel for different temperatures.

In a series of solid solutions between materials having the arrangements (*a*1) and (*a*2) at *all* temperatures, the change of sign of I_s at 0°K should *not* be accompanied by the occurrence of the strange I_s vs T curves as depicted in Fig. 8. We have found such a behavior in the system $MnFe_2O_4$—$MnCr_2O_4$ [14].

For ferrites that contain, in addition to Fe^{3+} ions, one or more other magnetic ions M, there are already 10 different interactions represented by lines in the following diagram:



Using Néel's simple theory means assuming the M—M, M—Fe and Fe—Fe interactions to be equal, which is a very crude approximation. Nevertheless the m/T curve calculated in this way with the interaction constants obtained from $m_{T=0}$ and the shape of the $1/\chi$ vs T curve agrees remarkably well with experiment for various simple ferrites (Fig. 11) [19].

F. Some Experimental Data on Saturation Magnetization

The saturation moments of the single ferrites agree well with the assumption of completely antiparallel spins in the *A* and *B* sublattices (17a). The saturation moment pro formula unit MFe_2O_4 is

$$m = \sum_i (x_i g_i S_i)_A - \sum_i (x_i g_i S_i)_B \qquad (23)$$

in which x_i, g_i and S_i are the number of ions, the g factor and the total spin quantum number of the ith magnetic ion in *A* and *B* sites respectively. For a completely inverse ferrite $Fe[M^{II}Fe]O_4$, for Fe^{3+} xgS is 5 on both sides, so that $m=gS_{(M^{II})}$.

In Table IV, p. 1958, the average experimental values of m obtained by Guillaud, Pauthenet and author are compared with the values of $2S_{(M^{II})}$. (A detailed

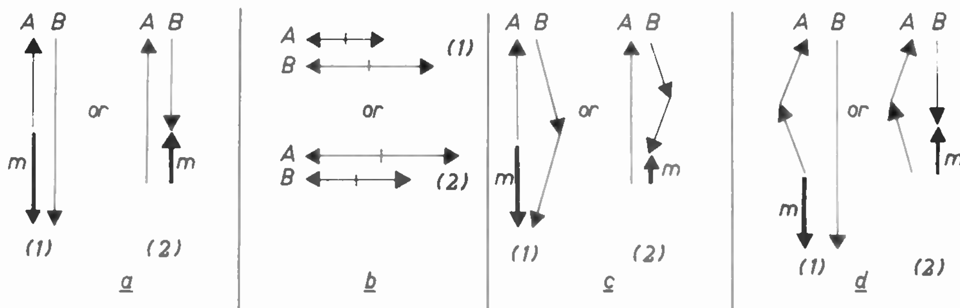


Fig. 10—Four different theoretically possible orientations of the moments in the sublattices *A* and *B* for different ratios between *AA*, *AB* and *BB* interactions. (a) Interactions *AA* and *BB* both small with respect to *AB*. (b) Interactions *AA* and *BB* large with respect to *AB*. (c) Interaction *BB* comparable to *AB*, and *AA* small with respect to *AB*. (d) Interaction *AA* comparable to *AB*, and *BB* small with respect to *AB*.

TABLE IV

	$2S(M^{II})$	m_{exp}	Curie temp. °C.	g_{exp}	Cell edge a in Å units	X-ray density in g/cm³
$MnFe_2O_4$	5	4.6–5.0	300	2.00	8.50	5.00
$Fe^{II}Fe_2O_4$	4	4.1	585	2.06	8.39	5.24
$CoFe_2O_4$	3	3.7	520	2.9 at 300° C.	8.38	5.29
$NiFe_2O_4$	2	2.3	585	2.25	8.34	5.38
$CuFe_2O_4$	1	1.3	455	2.15	8.68/8.24 (tetragonal)	5.38
$MgFe_2O_4$	0	1.1	440	2.05	8.36	4.52
$Li_{0.5}Fe_{2.5}O_4$	2.5	2.5–2.6	670	2.08	8.33	4.75

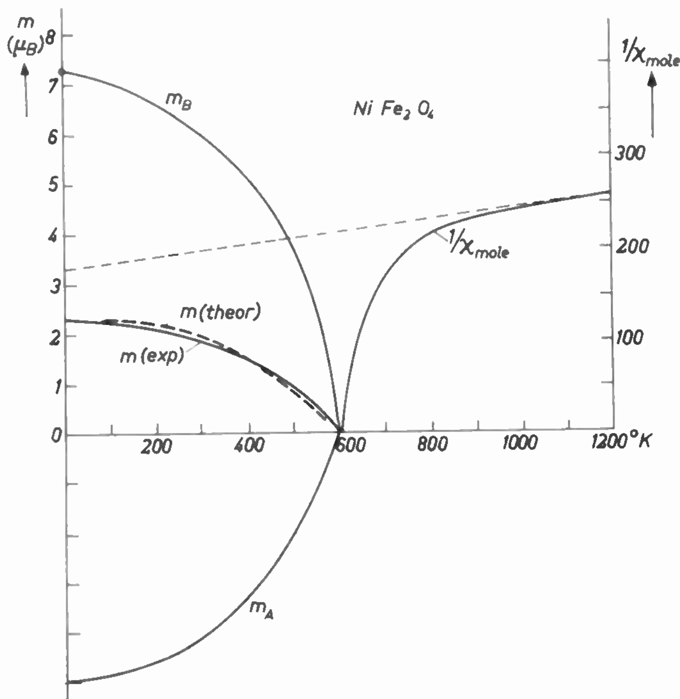


Fig. 11—The inverse of the susceptiblity χ and the saturation magnetization per formula unit m in Bohr magnetons for nickel ferrite (after Pauthenet [19]. It is seen that the theoretical curve (m_{theor}) agrees well with the experimental curve (m_{exp}). m_A and m_B are the antiparallel sublattice magnetizations, drawn schematically.

discussion of these values and literature references are given in [14].)

The discrepancies between the first two columns are explained as follows:

For $MnFe_2O_4$ no explanation can be given by us.

For $CuFe_2O_4$ and $MgFe_2O_4$ the arrangement is not completely inverse but only mainly so. This is proved by the fact that for samples quenched from a high temperature m is higher (see Part II, section C).

For $CoFe_2O_4$ and $NiFe_2O_4$ m is the same after quenching from a high temperature so that this explanation is not valid here. The discrepancies must therefore be ascribed to a contribution of the orbital moment; *i.e.*, to g being >2. This has been confirmed for $NiFe_2O_4$ by measurements of the effective g factor,

$$g_{eff} = \frac{\sum_i (x_i g_i S_i)_A - \sum_i (x_i g_i S_i)_B}{\sum_i (x_i S_i)_A - \sum_i (x_i S_i)_B}. \qquad (24)$$

From a comparison of the moments found for paramagnetic salts with $2S^*$ in Table II it follows that there the orbital contribution is largest for the same two ions Co^{2+} and Ni^{2+} and that it also plays a part, although a smaller one, in salts of Fe^{2+} and Cu^{2+}. We assume, therefore, that for $FeFe_2O_4$ and $CuFe_2O_4$ also, $g>2$. The antiparallel spin arrangement has been confirmed for several ferrites by neutron diffraction.

The saturation moments of solid solutions between the inverse ferrites and zinc ferrite all show a similar behavior which we shall illustrate for nickel-zinc ferrites $Zn_aFe_{1-a}[Ni_{1-a}Fe_{1+a}]O_4$ (Fig. 12).

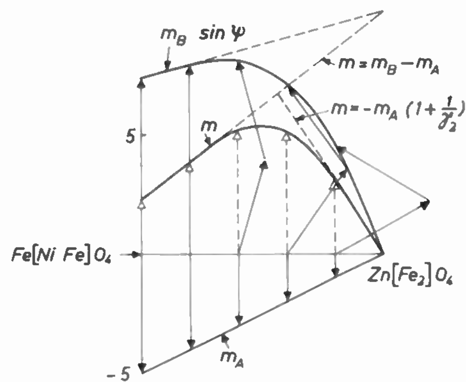


Fig. 12—Experimental saturation moment m in Bohr magnetons obtained from low-temperature measurements, for solid solutions $NiFe_2O_4$—$ZnFe_2O_4$. It is seen that $m = m_B \sin\psi - m_a$. Broken line $m = m_B - m_A$: theoretical for case (a) of Fig. 10. Broken line $m = -m_A[1+(1/\gamma_2)]$: theoretical for case (c) of Fig. 10.

For $a<0.4$ the saturation moments agree with those calculated with (17a); *i.e.*, $m=10a+(1-a)x2.3$. At $a>0.7$ the saturation moments agree with a formula like (22). For $0.4<a<0.7$ the curvature may be due to fluctuations in the ratio of Zn^{2+} and Fe^{3+} ions on A sites around each B site [23]

For $a=1$ ($ZnFe_2O_4$) there is complete antiferromagnetism, but the spin order seems to be very complicated and must be regarded as an unsolved problem. The Néel temperature is 10°K [24].

The Curie temperatures in all MZn ferrites decrease with increasing Zn content, which can also be accounted for by the theory.

In Fig. 13 a three-dimensional diagram is given of the system $Zn_aMn_{1-a}Fe_2O_4$ in which in three different planes we see the m vs T curves for different a, the curves giving m vs the composition parameter a for different tem-

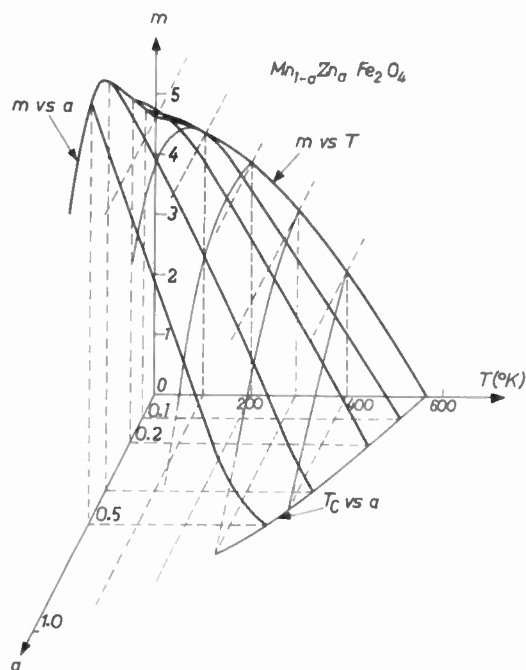


Fig. 13—Three-dimensional graph showing m vs T curves of different compositions a in the MnZn ferrite series of solid solutions $Mn_{1-a}Zn_aFe_2O_4$, and m vs composition at different temperatures. At 0°K introduction of $ZnFe_2O_4$ (*i.e.*, increasing a from $a=0$) increases m, but at 300°K it lowers m.

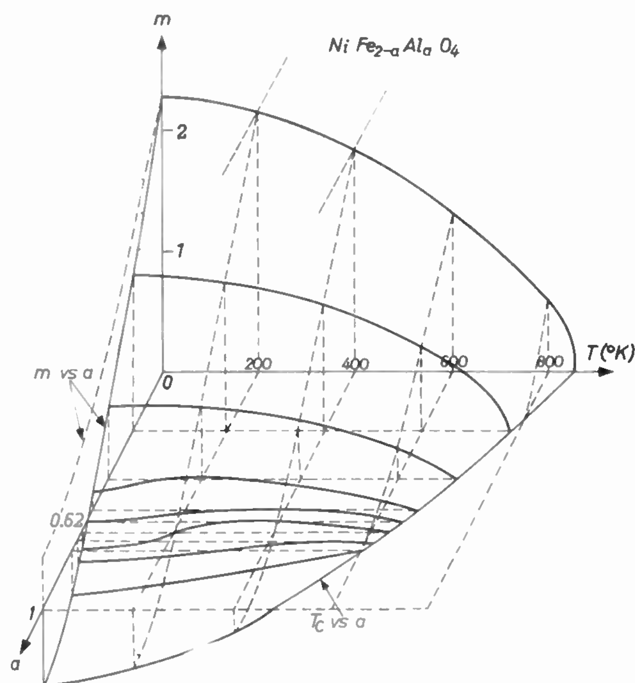


Fig. 14—Three-dimensional graph showing m vs T curves of different compositions a in the series of solid solutions $NiFe_{2-a}Al_aO_4$, very slowly cooled. The shapes of the curves are all those predicted from Néel's theory (see Fig. 8). Introduction of Al^{3+} (*i.e.*, increasing a from $a=0$) lowers m at all temperatures. The broken m vs a curve at 0°K is that for quenched samples.

peratures, and the Curie temperature vs a. Guillaud's [25] data have been used. It is seen that although m increases when a is increased from 0 to, *e.g.*, 0.2 at 0°K, this is not the case any more at 300°K or room temperature. At room temperature $MnFe_2O_4$ has the highest m *i.e.*, the highest I_s (see Part I section *H*) in this system, contrary to the NiZn ferrite system in which the maximum of I_s at about $a=0.4$ persists up to above room temperature.

Finally in Fig. 14 the same type of diagram is given for the system $NiFe_2O_4$—$NiFeAlO_4$ [14, 26] (for the complete diagram $NiFe_2O_4$—$NiAl_2O_4$ see [26]). It is seen that the Curie temperature also decreases when Al^{3+} ions are introduced, which occur mainly, but not completely [26a], on B sites. For slowly cooled materials $m_{T=0}$ changes sign at $a=0.62$. For slightly higher a, m changes sign with temperature.[11] When these materials are quenched $m_{T=0}$ remains positive (broken line), and the m/T curves (not shown) are of the normal type Q of Fig. 8.

Generally substitution of nonmagnetic metal ions which go mainly into A sites (Zn^{2+}, Cd^{+}, Ga^{3+} [26], In^{3+} [27] at first increases the saturation moment, whereas substitution of metal ions which go mainly into B sites (Al^{3+}, Sc^{3+} [27], $Ni^{2+}+Ti^{4+}$ [14] but also Cr^{3+}) lowers the saturation moment even at low concentrations.

In all cases, however, the Curie temperature is decreased by these substitutions, also in the case of substitution of Fe^{3+} by magnetic ions like Cr^{3+}, although slight anomalies can occur [14].

[11] This curve is not found by saturation magnetization vs temperature measurements, which only give the absolute value of the spontaneous magnetization. The point $m=0$ is not found since the temperature at which it occurs is strongly dependent on the distribution of Al^{3+} ions amongst A and B sites, and one does not succeed in making this distribution homogeneous throughout the material. The curve N shown is obtained from remanent magnetization vs T measurements. In the system $Fe[Li_{0.5}Fe_{1.5}]O_4-Fe_{0.5}Li_{0.5}[Cr_2]O_4$ such curves of type N occur in a wide range of compositions, so that they are found also from saturation magnetization measurements. Moreover for $Li_{0.5}Fe_{1.25}Cr_{1.25}O_4$ the point $m=0$ lies at $\approx 38°$ C, so that it can serve for a demonstration of ferrimagnetism [28, 14].

III. The Preparation of Ferrites

The ferrites are prepared by a ceramic technique. The raw materials, *e.g.*, oxides or carbonates, are milled, usually in steel ball-mills, and the mixture of very fine particles obtained is dried and prefired in order to obtain a homogeneous endproduct. This prefiring process may be repeated if necessary. After the prefiring the material is milled again.

Snoek [29] has shown in extensive experiments that the useful magnetic properties are very sensitive to homogeneity and to the oxygen content of the material. The homogeneity is influenced by the prefiring and milling treatments, and the necessary treatments are prescribed by:

a) The state of reactivity of the raw materials, which in turn depends on their method of manufacture,
b) on the composition of the final material, because the presence of some ions (*e.g.*, Cd^{2+}, Li^{+}) enhances sintering at much lower temperatures than that of others (*e.g.*, Cr^{3+}, Al^{3+}),
c) on the temperature adopted for the subsequent final sintering.

After the treatments mentioned, the powders are pressed in steel dies to the final shape, according to usual ceramic technique, which may involve the use of a preferably organic binder, which is burned out during the subsequent sintering process. The shapes are sintered in a stationary or a push-through furnace at temperatures between 1,100 and 1,450°C. During this sintering process the pressed piece shrinks (but equally in all dimensions if the process is well controlled) and the degree of shrinkage can be controlled within a few per cents by varying the prefiring and final sintering temperatures and durations. The finished parts may be treated by grinding and polishing of the surfaces.

Since there is a limit to the shrinkage control, the question may be asked why ferrites are not cast, like metals. The answer is that at the melting points, which usually lie above 1,600°C, the equilibrium oxygen pressures are very high, so that very serious reduction of Fe^{3+} to Fe^{2+} ions occurs. Reduction may occur but to a lesser degree, at the higher sintering temperatures, but in this case the FeO content usually remains below 1 per cent.

Using a sintering process we have to take a certain porosity into the bargain. This porosity depends on composition, pressing technique, and temperature and duration of the heat-treatments.

Examples of photomicrographs, made in reflected light of polished surfaces of some ferrites are given in Fig. 7. The number of pores is important for the magnetic properties as is discussed in Part IV, section *D*.

The shape of the pores also is important for the magnetic properties; when the sintering is good, as at a high sintering temperature, the pores of irregular shape coalesce to fewer, larger, and more sperical pores (Fig. 7(b)).

Another property that is influenced by the manufacturing technique is the internal strain in the material. Rapid cooling after sintering introduces large strains in various directions since the flow of material, which tends to relieve strains set up by the unavoidably unequal cooling rates for different parts of the piece, has a rate strongly decreasing with decreasing temperature, so that it cannot maintain equilibrium down to room temperature. Very slow cooling and long annealing treatments decrease the strains but it is impossible to get rid of them altogether, so that one must reckon with a minimum internal strain of the order of 10^9 dynes/cm². The presence of a second phase as α-Fe_2O_3 (Fig. 7(d)), apart from acting in the same way as air pores do, presumably increases such strains because of the unequal thermal expansion coefficients of the two phases.

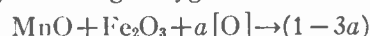
We have seen that the oxygen content is important for the magnetic properties. Apart from the firing temperature, the oxygen content of the atmosphere in which the material is sintered influences this to a great extent.

Let us for example assume that we wish to prepare $Mn^{II}Fe_2^{III}O_4$. Depending on the oxygen content of the atmosphere we may obtain the following results at a certain temperature high enough to give equilibrium inside the time during which it is maintained.

a) Much too large oxygen content:

$MnO+Fe_2O_3+1/3[O]\rightarrow 1/3Mn_3O_4+1Fe_2O_3$, in which at 1000°C at least part of the Fe_2O_3 is present as a second phase. Such products and those with even higher oxygen content, such as $Mn_2O_3\cdot 2Fe_2O_3$ solid solutions and $MnO_2+Fe_2O_3$, can only exist in atmospheric oxygen at temperatures much lower than those at which ferrites are manufactured, so that they may be disregarded here.

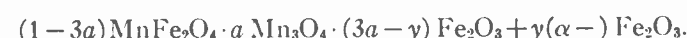
b) Too large oxygen content:



$MnO+Fe_2O_3+a[O]\rightarrow(1-3a)$

$MnFe_2O_4\cdot a\ Mn_3O_4\cdot 3a\ Fe_2O_3$,

which, as we have seen in Part II section *B*, is a solid solution for small values of *a* and not too slow cooling. For larger values of *a* and slow cooling part of the Fe_2O_3 separates, we obtain:



$(1-3a)MnFe_2O_4\cdot a\ Mn_3O_4\cdot(3a-y)\ Fe_2O_3+y(\alpha-)\ Fe_2O_3$.

c) Correct oxygen content:

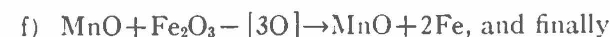
$MnO+Fe_2O_3\rightarrow MnFe_2O_4$.

d) Too low oxygen content:

$MnO+Fe_2O_3-a[O]\rightarrow(1-a)\ MnFe_2^{III}O_4+a(MnO\cdot FeO)$ in which however, the Mn^{2+} and Fe^{2+} ions give an equilibrium among the two phases, so that the spinel phase shows conductivity.

Further reduction leads to:

e) $MnO+Fe_2O_3-[O]\rightarrow MnO\cdot 2FeO$,



f) $MnO+Fe_2O_3-[3O]\rightarrow MnO+2Fe$, and finally

g) $MnO+Fe_2O_3-[4O]\rightarrow Mn\cdot 2Fe$ in this order.

In principle the state of oxidation after the firing is independent of that before the firing, so that the atmosphere during the presintering is generally unimportant.

A firing in air at 1,300°C will for $MnFe_2O_4$ give a single-phase product of type b; for $MgFe_2O_4$, a product of type d. Products of type b exist for $CoFe_2O_4$, but obviously do not exist for $MgFe_2O_4$ and $ZnFe_2O_4$, not even at lower temperatures, because Mg and Zn have no higher valency states. Ni^{3+} and Fe^{4+} ions exist in other compounds, but are so unstable that they do not occur in spinels at "normal" firing temperatures.

Products of type d will reoxidize on cooling in the same atmosphere in which they were fired, since the equilibrium oxygen pressure of the materials decreases strongly with temperature. Since the oxidation is sufficiently rapid only at the surface, the reoxidation will be the more complete the more porous the structure [29].

If we wish to preserve a material containing ferrous ions, a very dense material, or a rapid cooling, or an increasingly reducing atmosphere during cooling must be used.

TABLE V

SOME PROPERTIES OF FERRITES AT ROOM TEMPERATURE

ferrite	$K_1 \times 10^{-5}$	$\lambda_s \times 10^6$	$\lambda_{111} \times 10^6$	$\lambda_{100} \times 10^6$	$4\pi I_S$ for theor. density
$MnFe_2O_4$	-0.03	-5			5,200
$Fe^{II}Fe_2O_4$	-1.1	$+41$	$+81$	-19	6,000
$CoFe_2O_4$	$+38$		$+120$	-590	5,000
$NiFe_2O_4$	0.62	-26			3,400
$CuFe_2O_4$		-10			1,700
$MgFe_2O_4$		-6			1,400
$Li_{0.5}Fe_{2.5}O_4$		-1			3,900

IV. Physical Properties of Ferrites in Relation to the Chemistry [11]

A. Permeability and the Basic Properties and Processes on Which it Depends

We have discussed the susceptibility χ and the magnetization I: $\chi = I/H$, in which I may be expressed in cgs units/cm³. In ferromagnetism it is customary to use the induction B and the permeability $\mu = B/H$: $B = H + 4\pi I$ and $\mu = 1 + 4\pi\chi$, using Gaussian units. In high-frequency technique we are interested in the high-frequency permeability which depends on the amplitude of B. In many applications the materials are used in the demagnetized state ($B=0$, $H=0$), in very weak alternating magnetic fields, so that we are then interested in $(dB/dH)_{B=0, H=0} = \mu_i$, the initial permeability. When a larger field is applied, the induction first follows the virginal curve, and on decreasing the field a hysteresis loop occurs in which the remanence B or $4\pi I$ for $H=0$ and coercive force H for $B=0$ (or for $I=0$) are characteristic points. On the loop obtained after saturation we call these constants the retentivity $B_r = 4\pi I$ and the coercivity ${}_BH_c$ (or ${}_IH_c$) respectively. (For small values of H_c, ${}_BH_c = {}_IH_c = H_c$) (see Fig. 24). When a static biasing field is used in addition to an alternating field δH which causes an alternating induction δB the permeability for $\delta H \to 0$ is $\mu_{rev} = \delta B/\delta H$, the reversible permeability. This permeability is usually quite different from the relevant slope of the B-H curve, and is, if the biasing field is too small to saturate material, of the same order as μ_i. We shall concern ourselves here with the variation of μ_i with temperature and with frequency, and the variation of μ with the amplitude of H and with frequency.

In paramagnetic salts the magnetic field has to turn the isolated spins of the magnetic ions against the disordering influence of thermal agitation. In ferromagnetics, where below the Curie temperature in large domains all spins are parallel due to the exchange interaction, thermal motion obviously has no influence on the total domain magnetization. If there were no opposing forces an infinitely small field would turn all spin vectors in the field direction; *i.e.*, μ would be infinite. In fact there are several factors binding the domain magnetization vectors to certain directions, and energy has to be supplied to turn these vectors into the field direction. These factors are listed in the following sections.

B. Crystalline Anisotropy

In each crystal the spontaneous magnetization is bound more or less tightly to certain crystal directions. The energy E_c depends on the orientation of the magnetization with respect to the crystal axes, and for cubic crystals in first approximation is given by

$$E_c = K_1(\alpha_x^2\alpha_y^2 + \alpha_x^2\alpha_z^2 + \alpha_y^2\alpha_z^2) + K_2(\alpha_x^2\alpha_y^2\alpha_z^2) + \cdots \quad (25)$$

in which α_x, α_y and α_z are the cosines between the direction of magnetization and the three crystal axes x, y and z respectively. These constants K may be either positive or negative. For $K_1 > 0$, E_c is a minimum for, *e.g.*, $\alpha_x = 1$ and $\alpha_y = \alpha_z = 0$, or any 100 (cube-edge) direction is a direction of preferred magnetization. For $K < 0$, E_c is a minimum for $\alpha_x = \alpha_y = \alpha_z$, or any 111 (cube-diagonal) direction is a direction of preferred magnetization. (K_2 is neglected in both cases.)

A good measurement of K_1 and K_2 involves measurement on not too small single crystals which are more easily obtained pure for metals than for ferrites. For Fe: $K_1 = +5 \times 10^5$, $K_2 = +1.5 \times 10^5$; for Ni: $K_1 = -0.6 \times 10^5$ erg/cm³. Some values of K_1 for ferrites[12] are given in Table V. Until recently all ferrites were thought to have negative values of K_1, but recently Bozorth has shown it is positive for $CoFe_2O_4$. It may be noted that large values of K_1 occur in those ferrites which contain ions with a high g factor (for completely inverse ferrites equal to g_{eff}. (see Part II section F (24)) so this high K_1 is presumably due to spin-orbit coupling, the orbital momentum being strongly coupled to the lattice. In ferrites with magnetic ions with $L=0$ in the ground state, in which therefore practically no spin-orbit coupling can occur, the crystalline anisotropy is low in the case of $MnFe_2O_4$, but certainly higher for $MgFe_2O_4$ and for $Li_{0.5}Fe_{2.5}O_4$; this is due to unknown causes. The values of K_1 and K_2 of Fe and Ni are known to decrease steeply with temperature, so that they are zero well below the Curie temperature; it seems likely that this is a general phenomenon.

C. Strain Anisotropy

When a rod of a ferromagnetic material is magnetized lengthwise, it either decreases or increases in length.

[12] Most of these have not the correct stoichiometric ratio and may contain cation vacancies or ferrous ions.

The relative change in length $dl/l=\lambda$ is then negative or positive respectively. Conversely, when a tensile stress σ is applied, the direction of stress for $\lambda>0$ is a preferred direction of magnetization, but for $\lambda<0$ any direction perpendicular to the stress, when no other anisotropy is present. When the material is magnetized to saturation the magnetostriction is λ_s and the strain anisotropy $(3/2)\lambda_s\sigma$.

For a certain direction of magnetization the strain anisotropy energy is again a function of the angle between the magnetization and the crystal axis, and the behavior is usually adequately described by two terms containing the constants λ_{100} and λ_{111} respectively. Some values of λ_s and of λ_{111} and λ_{100}, obtained for the same single crystals as above, are given in Table V. It may be noted that all λ_s values and all λ values measured in the preferred direction are negative, except for Fe_3O_4. The magnetostriction also decreases with temperature, but more slowly than the crystalline anisotropy in most cases.

D. Shape Anisotropy

In a magnetized piece of a ferromagnetic "free poles" exist on the surface which create a demagnetizing field in the material. Only for an ellipsoid is this demagnetizing field homogeneous, and here the demagnetization energy is

$$E_d = \tfrac{1}{2}(N_xI_x^2 + N_yI_y^2 + N_zI_z^2)$$

in which N_x, N_y and N_z are demagnetization coefficients connected with the shape of the ellipsoid, and I_x, I_y and I_z are the projections of I on the three axes x, y, and z of the ellipsoid. The components of this demagnetizing field are N_xI_x etc. We have $N_x+N_y+N_z=4\pi$; for a comparatively very long dimension z, $N_z=0$; *i.e.*, we have

	N_x	N_y	N_z
for a long thin needle	2π	2π	0
for a large thin plate	0	0	4π
for a sphere	$(4/3)\pi$	$(4/3)\pi$	$(4/3)\pi$.

The values of $4\pi I_s$ at room temperature for completely dense ferrites are also given in Table V. An air pore in a spontaneously magnetized domain acts in the same way: free poles exist and a demagnetizing field is created. Since nonspherical pores will generally lie in random directions, the smallest demagnetization energy is caused by spherical pores. For this reason, the presence of crystals of a second nonmagnetic[13] phase like α-Fe_2O_3, which separates in very large thin plates (Fig. 7(d)) causes a larger shape anisotropy than spherical air-filled pores, in addition to increasing the internal strain and therefore the strain anisotropy. Recapitulating, we see that the crystal anisotropy is dependent on the chemical composition, the strain anisotropy both on chemical composition (λ) and manufacturing technique (σ), the shape anisotropy on chemical composition (I_s) and on manufacturing technique (N-values). The values of K, λ and I_s in the ferrites are such that the demagnetization energy is largest; *i.e.*, free poles are unlikely to occur.

[13] The extremely small saturation magnetization of $\alpha-Fe_2O_3$ of the order of $1/200\times$ that of the ferrites, may be disregarded here.

E. Weiss Domains and Bloch Walls

The reason that a spontaneously magnetized piece of ferromagnetic is not a permanent magnet is that by the formation of Weiss domains the number of free poles and therefore the demagnetization energy is decreased. For predominant crystal anistropy the angles between the domain orientations are 180° and (for $K>0$) 90° or (for $K<0$) 71 or 109° (= the angle between the cube diagonals). The domains are separated by boundaries, *Bloch walls*, in which adjacent spins (meaning, in a ferrimagnetic, the excess spin of one of the sublattices) are not exactly parallel, but make a very small angle, so that the orientation of the spins gradually changes from the first domain orientation to the second in such a way that no free poles are created in the wall.

For the creation of the wall, energy has to be supplied against the exchange energy for the formation of angles of adjacent spins, and against the crystal anisotropy because the spins in the wall are not in a preferred direction. The thickness of the wall is a simple function of these two quantities. This energy must be supplied by the decrease in demagnetization energy, and since this is connected with volume, but the wall energy with surface, there is a certain critical diameter below which no wall can be present. For the cubic ferrites, however, this is extremely small ($10^{-5}-10^{-6}$ cm).

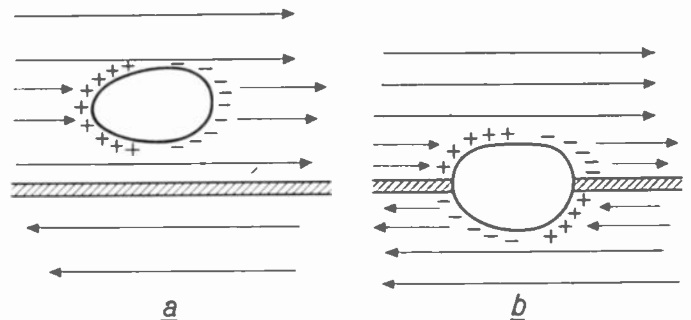


Fig. 15—Decrease of demagnetization energy when a domain wall intersects a pore or inclusion.

Under the influence of an external field the wall moves in such a way that the domain having a magnetization vector nearest to the field grows at the expense of its neighbor. It would seem that such a wall displacement might take place very easily since its energy does not change with its location, disregarding pores and strains. Due to the fact that the spins precess during wall movement, damping of this precession will cause energy dissipation. Sintered ferrites, moreover, contain a number of pores, and the Bloch walls will be fixed to these pores, thus diminishing the number of free poles on the pore (Fig. 15). Intricate domain patterns for various cases have been calculated by Néel,

and these have been observed by Williams in metals and by Galt in ferrites.

The other magnetization process is the rotation of the domain magnetization vectors towards the direction of the applied field in which energy must be supplied to overcome the crystal anisotropy and the strain anisotropy. For pure domain rotation and negative K_1 it can be shown that

$$\mu - 1 = 2\pi I_s^2 / |K_1| . \tag{26}$$

We shall now see how the factors discussed above influence the initial permeability and the hysteresis loop. From the foregoing it is clear that in order to have very high permeability, we expect we must have:

a) a very dense material

b) $\lambda = 0$, $K_1 = 0$. We shall disregard K_2 in what follows. We can make $\lambda = 0$ either by making solid solutions between ferrites with $\lambda < 0$ and Fe_3O_4 with $\lambda > 0$, or we can make λ small by taking a material in which the Curie temperature is not too far above room temperature. This can be achieved by using solid solutions with zinc ferrite, which, as we have seen, lowers the Curie temperature. As regards K_1 we can again make $K_1 = 0$ by making solid solutions between a ferrite with $K_1 < 0$ and Co ferrite which, as we saw, has a very large positive K_1. Another method to reduce K_1 is to take a ferrite with a Curie temperature not far above room temperature. Snoek found that materials with low Curie temperature, *e.g.*, $Cu_{0.4}Zn_{0.6}Fe_2O_4$, $\approx Mn_{0.5}Zn_{0.5}Fe_2O_4$ and $Ni_{0.3}Zn_{0.7}Fe_2O_4$, can have very high initial permeabilities of 1,500, $\approx$1,000 and 4,000 respectively, which he ascribed to the fact that K_1 and λ are very low, not far below the Curie temperature. He, moreover, reduced λ to practically zero by incorporating a few per cent of Fe_3O_4 into these solid solutions, and found that for the MnZn ferrite μ_i was substantially increased; *e.g.*, to 3,000. The Curie temperature effect cannot be the entire story, however, since nickel ferrite and nickel zinc ferrites with higher Curie temperatures do show a peak of μ_i just below the Curie temperature, but a very much lower one (Fig. 16). It seems probable that already at low temperatures the anisotropies of the solid solutions with Zn ferrite are lower than in the simple ferrites and that λ does not decrease with temperature so steeply as Snoek supposed. It is noteworthy that a decrease of the Curie temperature by the replacement of Fe^{3+} ions, *e.g.*, by Al^{3+} ions does not give a rise in μ_i, which even allowing for the influence of density and saturation magnetization, does not fit in the "Curie temperature effect" picture. As regards the influence of density on μ_i, it appears that this is very marked if K_1 is very small. Snoek [29] showed that the same composition $Ni_{0.3}Zn_{0.7}Fe_2O_4$, which showed the highest initial permeability, only had $\mu_i = 80$ when fired at a very low temperature of 1,000°C. In the latter case the shape anisotropy determines μ_i.

Wijn [11] showed, however, that a material with higher K_1, $Ni_{0.66}Zn_{0.34}Fe_2O_4$, shows no difference in

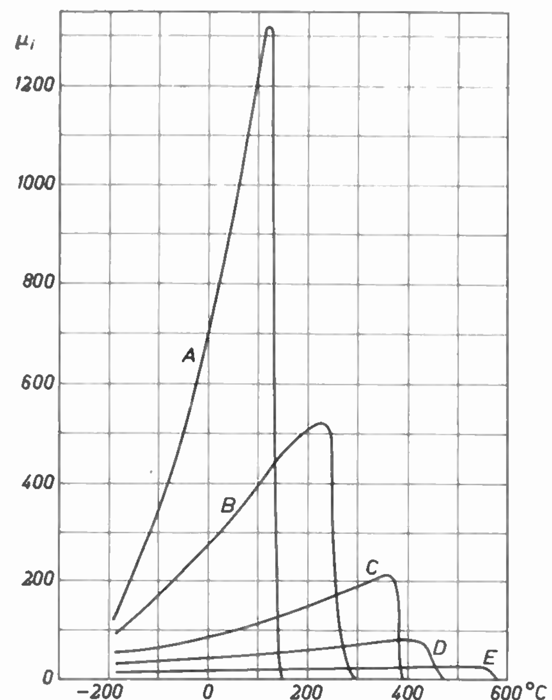


Fig. 16—Initial permeability vs temperature for a series of nickel-zinc ferrites with the composition $\approx Ni_{1-a}Zn_aFe_2O_4$. *A*: $a \approx 0.36$; *B*: $a \approx 0.5$; *C*: $a \approx 0.64$; *D*: $a \approx 0.2$; *E*: $a \approx 0$.

$\mu_0(\approx 90)$ when fired at 1,250°C or at 1,460°C (in the latter case some Fe^{2+} ions are formed and probably a small amount of cubic second phase, which acts mainly in the same way as air pores). The coercive force is however decreased four-fold by higher firing temperature, and maximum permeability increased five-fold (Fig. 17).

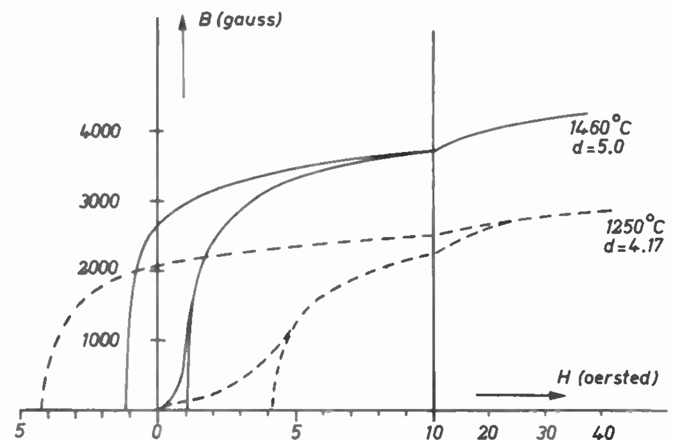


Fig. 17—Hysteresis loops of NiZn ferrite, composition like *C* in Fig. 16. Broken line curve: fired at 1,250°C in oxygen, porosity 23 per cent. Solid line curve: fired at 1,460°C in oxygen, porosity 12 per cent.

A very steep rise (or fall) of *B* vs *H* (near H_c and where μ_{max} occurs) must be due to irreversible wall displacements, (irreversible because reversal of the field gives a totally different slope of *B* vs *H*, and wall displacements, because rotations give a more gradual change of *B* vs *H*) (see also below). The decrease of H_c with higher firing temperatures means that wall displacements become much easier, which is understood from Fig. 15 if

we take H_c to be the field necessary to dislodge the walls from the pores. The constancy of μ_i is, therefore, proof that μ_i is not influenced by wall displacements and that μ_i has to be caused by rotations only. The coercive forces for this material and for other NiZn ferrite compositions, fired at these different temperatures, have been found by Wijn to agree well with an expression derived by Néel for H_c as a function of the porosity p. ($p=1-(d_{\text{apparent}}/d_{\text{x-ray}})$, in which d_{apparent} = weight/external volume of the sintered piece and $d_{\text{x-ray}}$ = weight /volume of a unit cell $=(8A/N)a^3$). In this expression I_s and K_1 appear and K_1 is derived from μ_i, using (26).

In the above picture, it was assumed that the steep slope of B vs H at field strengths of $\approx H_c$ is due to wall displacements. This has been shown by Wijn to be true, from measurements on rods under lengthwise tensile stress or pressure. Since the magnetostriction is negative, tensile stress will give preferred orientations perpendicular to the length of the rod, and pressure a preferred orientation in that direction.

It follows that tensile stress will favor the occurrence of rotational processes and pressure that of 180°-wall displacements. In practice the stresses and pressures are limited by the mechanical strength of the material, so that the spins will not all turn into these stress anisotropy directions. Nevertheless for fields $\geqq H_c$ tensile stress increased B, and pressure decreased B, which proves that the magnetization process at these field strengths is related to Bloch wall displacements.

F. Frequency Dependence of the Permeability

1. Frequency dependence of the amplitude permeability

Since ferrites are used at high frequencies it is interesting to see how its useful properties depend on frequency.

A sinusoidal field with amplitude $H_{\max}$ gives an induction with amplitude $B_{\max}$, which lags behind H in time. This time lag gives rise to losses. Wijn [30] measured $B_{\max}/H_{\max}$ curves for several fixed frequencies for many ferrites. The results for two NiZn ferrites with approximately equal composition, but sintered at different temperatures, are given as an example in Figs. 18(a) and (b). In Fig. 18(a) the magnetization curve; *i.e.*, the amplitude permeability defined by $\mu=B_{\max}/H_{\max}$, is independent of frequency up to ≈ 1.8 mc/sec. In Fig. 18(b) the magnetization curve changes considerably with frequency. From a comparison with the results mentioned in the foregoing section it is clear that in the latter case the irreversible domain wall displacements cannot follow the high-frequency field because they are damped. The theory of a possible damping mechanism is not given here. It suffices to say that a certain damping coefficient derived from domain wall velocities measured on single crystal specimens [31] agrees qualitatively with that obtained from Wijn's measurements. It may be deduced from the latter experiments that generally all ferrites with $\mu>400$ show a decrease of the amplitude permeability μ with fre-

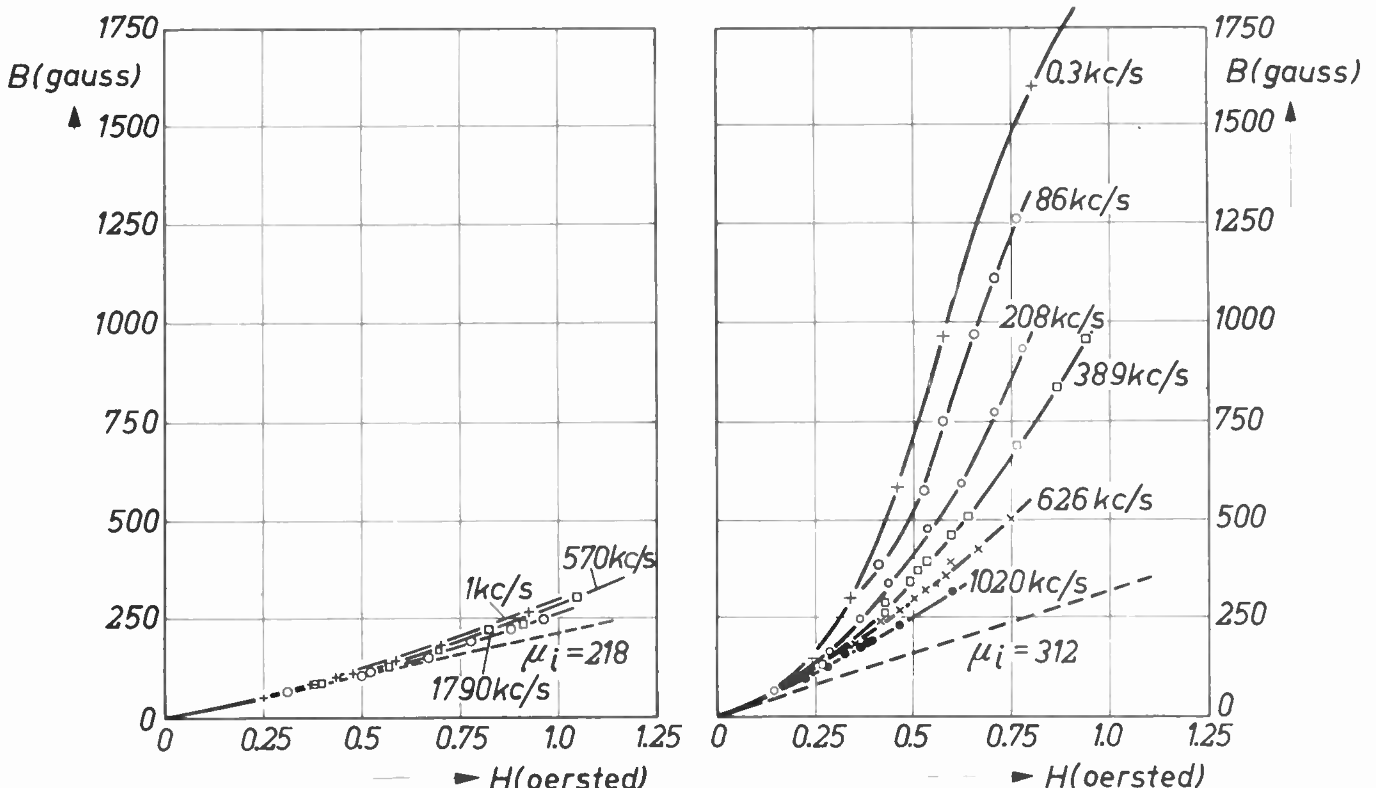


Fig. 18—Amplitude $B_{\max}$ as a function of field amplitude $H_{\max}$, for different frequencies, of two NiZn ferrites with compositions approximately as in B in Fig. 16. (a) Fired at 1,250°C in oxygen. (b) fired at 1,450°C in oxygen.

quency; those with $200 < \mu < 400$ show a marked decrease with frequency only when fired at a high temperature; those with $\mu < 200$ do not show such a decrease even when fired at $\approx 1{,}450°C$.

2. Natural Ferromagnetic Resonance

We saw in Fig. 18 that the initial permeability μ_i of the ferrites concerned was independent of frequency up to ≈ 1.8 mc/sec. We shall now see how the initial permeability changes when the frequency is further increased. Since this change is connected with by far the most important losses occurring in ferrites in small alternating fields, we shall take into account the phase angle δ of B with respect to H. For $(B_{max}/H_{max})\cos\delta = \mu'$ and $(B_{max}/H_{max})\sin\delta = \mu''$ we can write $\mu_i = \mu' - j\mu''$ and $\mu''/\mu' = \tan\delta$. Graphs giving μ' and μ'' as a function of the frequency f on a doubly logarithmic scale are given for a number of NiZn ferrites in Fig. 19 [32]. It is seen

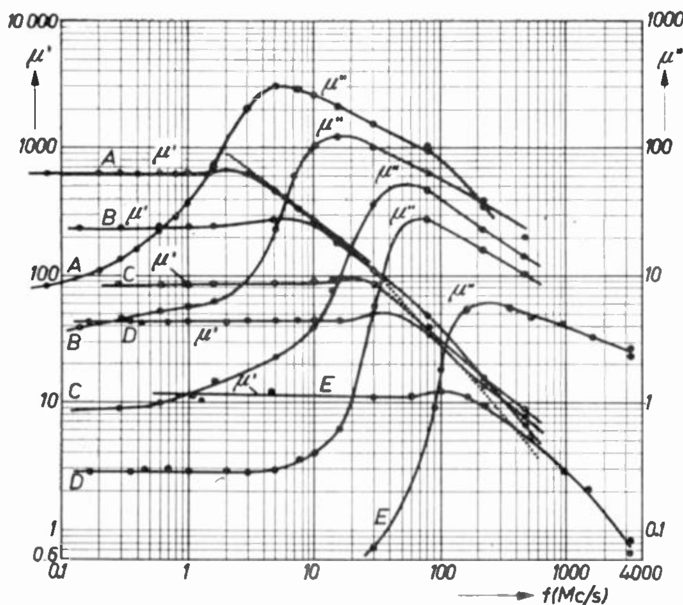


Fig. 19—Real and imaginary parts of the initial permeability for nickel-zinc ferrites with the same composition as those of Fig. 16, as a function of frequency f. The broken line connects the points where μ' has dropped to half its low-frequency value.

that for every material μ' first slightly increases with frequency and then starts to decrease at a frequency slightly below that at which μ'' has a maximum, and that for the material E, $\mu' - 1$ is shown to be negative at the highest frequencies.

Snoek, who first observed this phenomenon [33], ascribed this to a resonance mechanism, theoretically predicted already in 1935 [34], which is briefly explained here.

We have seen that the electron has a magnetic moment $e\hbar/2mc$ and an angular spin momentum $\hbar/2$. Thus the electron can be considered as a spinning top of which we know that it can perform precessional motions the frequency of which depends on the stiffness with which it is bound to its equilibrium position. Let the stiffness be only caused by an external static magnetic field H_z, then the application of an ac field perpendicular to H_z will cause resonance when its frequency is equal to the natural precession frequency in the static field, and it follows from the equation of motion that this resonance frequency f_r obeys

$$\omega_r = 2\pi f_r = \gamma H_z = -g\left(\frac{e}{2mc}\right)H_z. \quad (27)$$

In paramagnetics such resonance phenomena can be observed. In ferromagnetics, due to the exchange interaction, the spins of a complete Weiss domain precess in unison.

When there is no external field H_z there is still the stiffness of the anisotropy forces. Their action can be simulated for small deviations from the equilibrium position by that of a magnetic field which we call the crystalline anisotropy field H_A. Its magnitude is easily shown to be $H_A = 4|K_1|/3I_s$ for $K_1 < 0$, and it points into the equilibrium orientation of the magnetization. Eliminating K_1 by using (26) we obtain Snoek's equation

$$(\mu_i - 1)f_r = \frac{4}{3}\gamma I_s \quad (28)$$

i.e., for high low-frequency permeability μ_i one obtains a low f_r vice versa. In the above picture the resonance would occur at one sharp frequency f and have infinite amplitude. In reality this is not true, due to damping of the precession of the spins, which may be compared with the viscosity of a liquid, and which is due to interaction with the other spins and with the crystal lattice. Therefore the resonance amplitude is finite and the resonance peak occurs in a frequency range around f_r.

If we take f_r to be the frequency at which μ' is half its low-frequency value μ_i, $(\mu_i = \sqrt{\mu'^2 + \mu''^2})$ and μ'' at low frequencies is very small), then the practically straight dotted line through the points $\mu' = \frac{1}{2}\mu_i$ in the doubly-logarithmic plot shows that the relation (28) is very well obeyed, since I_s varies only by a factor <2 [11].

Nevertheless the slow fall-off of μ' with frequency is not that expected for a resonance in a fixed field H_A even with damping. This is due to the fact that the expression (27) is only valid for a sphere. For an elongated shape the demagnetizing fields are also active and the expression becomes for an ellipsoid with a long axis in the x direction [35]:

$$\omega_r = \gamma\sqrt{[H_A + I_s(N_y - N_x)][H_A + I_s(N_z - N_x)]}. \quad (29)$$

Normally the maximum stiffness is $H_A + 2\pi I_s$ since $N_x + N_y + N_z = 4\pi$, but Polder and Smit [36] have shown that for such an ellipsoid with a number of domain walls in the yz plane, an ac field parallel to the walls sets up free poles on the wall, with the result that the maximum stiffness becomes $H_A + 4\pi I_s$; *i.e.*, the highest resonance frequency is

$$\omega_r = \gamma(H_A + 4\pi I_s). \quad (30)$$

This explains the wide resonance frequency range. The resonance losses in the highest resonance frequency region, corresponding to (30) disappear when $4\pi I_s < \omega_r/\gamma$, if H_A can be neglected (small K_1), as was shown by varying $4\pi I_s$ by heating up the sample at a constant frequency (9,300 mc/sec) [37]. Since these losses are connected with the occurrence of Bloch walls, they vanish when the material is saturated. This results in the possibility of the application of ferrites as absorption modulators in resonance cavities.

The changes of μ with frequency that have been observed in ferrites can all be explained by the above resonance phenomena, so that it does not seem necessary to assume the occurrence of reversible wall displacements in very low fields, which also may give rise to a resonance phenomenon [38], as has been shown for other materials [39].

The technically interesting property $\tan\delta/\mu'$, which is independent of the introduction of an air gap, is shown in Fig. 20 for a number of ferrites.

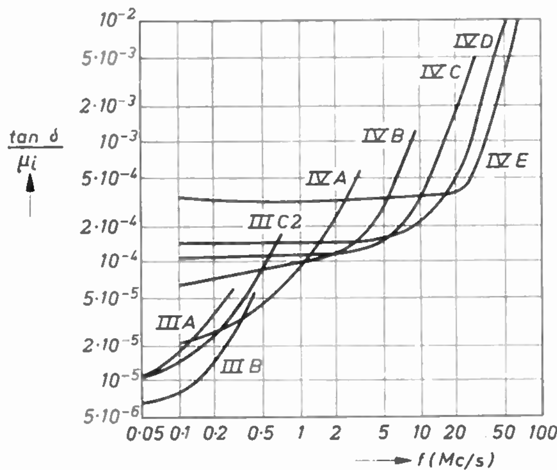


Fig. 20—The relative loss factor $\tan\delta/\mu_i$ for different grades of Ferroxcube. Ferroxcube III materials are MnZn ferrites; Ferroxcube IV *A–E* have compositions approximately equal to those given as *A–E* in Fig. 16.

G. The Influence of Resistivity

1. Eddy current losses

The ferrites thank their existence to their high resistivities which prevent the occurrence of eddy currents. It is of course well known that an alternating magnetic flux only penetrates into a conducting medium to a skin depth s;

$$s = \frac{1}{2\pi\sqrt{\mu' f/\rho}}$$

in cgs units, in which ρ is the resistivity in ohm cm.

Even for a complete flux penetration there are eddy current losses amounting to $W_e f = \pi^2 d^2 B^2 f^2 / n\rho$ in which d is the smallest dimension of the sample perpendicular to the flux and n depends on the shape of the sample ($n = 6$ for sheets, 16 for cylinders, 20 for spheres).

In most ferrite applications that have been developed, the eddy current losses may be neglected with respect to resonance losses discussed here; nevertheless conductivity gives rise to other losses we shall mention.

2. Dimensional resonance losses

We saw that ferrites containing Fe^{2+} ions show a low resistivity: ($\rho = 5.10^{-3}$ ohm cm for Fe_3O_4, $\rho = 10^2$ ohm cm for MnZn ferrite containing some Fe^{2+} ions).

Such materials with low ρ show an exceptional high dielectric constant ϵ' at low frequencies. At increasing frequency, ϵ' and the resistivity ρ decrease and again reach a constant level at very high frequencies [40]. This behavior of the conductivity $1/\rho$ and ϵ' can be described with exactly the same relaxation formulas, with the same relaxation time [40].

The assumption to which one seems to be forced from several investigations [41] is that the grains of the ferrite make poor contact with each other and that the frequency dependence of $1/\rho$ and ϵ' is caused by the dielectric breakdown of the air gaps.

The high dielectric constant ϵ' causes a new type of losses, as has been found by Brockman et al. [42], which, like eddy current losses, are dependent on the size of the sample. The wavelength inside the material is about equal to $\lambda_m = (c/f)(\mu\epsilon)^{-1/2}$, if the losses are small. When this wavelength is smaller than twice the relevant dimension of the core, standing waves are set up, resulting in large losses. Because of the high ϵ' (*e.g.*, 10,000) together with the high μ (*e.g.*, 1,000), one obtains losses already for a dimension of 2 cm at 2.5 mc/sec., as Brockman, Dowling and Steneck have shown.

As an example of the behavior of μ' and ϵ' with respect to μ_1' and ϵ_1', the real parts of μ and ϵ at very low frequency, are given as a function of frequency in Fig. 21. The μ'/f curve for the laminated sample is the normal ferromagnetic resonance curve. These losses make it necessary to use high resistivity ferrites for applications where large pieces are to be used at high frequencies, *e.g.*, in the $17\times10\times2$ cm blocks from which transformer cores for Brookhaven Cosmotron were built.[14]

3. Relaxation losses

Another type of losses connected with the resistivity already occurs at very low frequencies, so far below the ferromagnetic resonance frequency that the resonance phenomenon cannot be held responsible.

The measurement of these losses at different frequencies and temperatures has shown that these losses, which were known since the invention of high-permeability ferrites [29], are due to a relaxation phenomenon: for every frequency the losses have a maximum at a different temperature (Fig. 22) and the relation between the relaxation time $\tau = 1/2\pi f$ and temperature is given by $\tau = \tau_\infty e^{E_m/kT}$; *i.e.*, the plot of $\log\tau$ vs $1/T$ shows a straight line.

[14] Indeed, it was on samples of MnZn ferrite ($\rho = 10^2$ ohm cm) for this project that the dimensional losses were first found. The material finally used was a NiZn ferrite with $\rho > 10^5$ ohm cm.

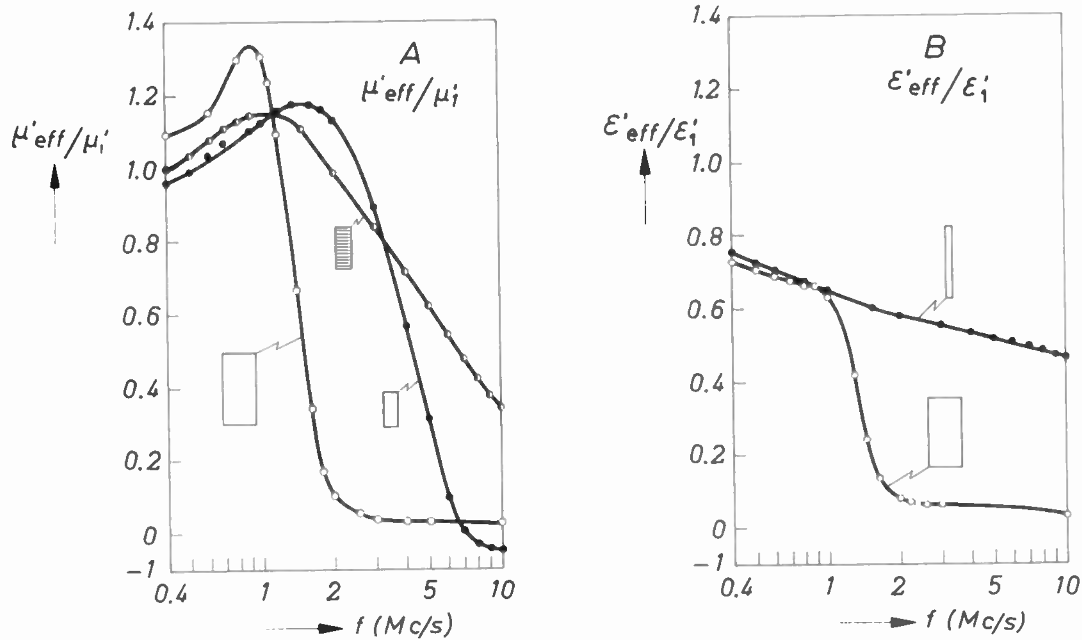


Fig. 21—(a) Initial permeabilities $\mu_{eff.}'$ measured on samples of different size plotted as $\mu_{eff.}'/\mu_i'$ (in which μ_i' is measured at 1,000 cycles/s) against frequency. (b) Similarly values of $\epsilon_{eff.}'/\epsilon_i'$ have been plotted against frequency. Relative sample cross sections are indicated in the figure; the hatched rectangle is a laminated core.

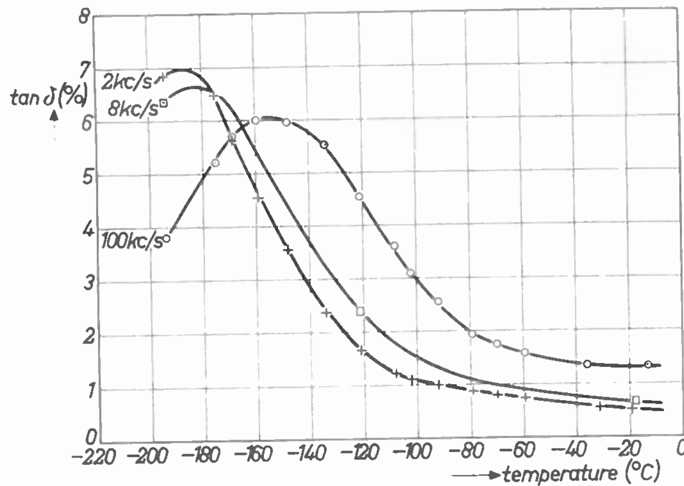


Fig. 22—Loss factor tan δ in per cent of a MnZn ferrite ($\approx Mn_{0.56}Zn_{0.38}Fe_{0.06}^{II}Fe_2O_4$) as a function of temperature, measured for various frequencies at a fieldstrength of 2.5 millioersteds.

For this MnZn ferrite the activation energy obtained from Fig. 22 [43] was $E_m = 0.11$ ev, a very small value, which may be connected with an electron diffusion process. Wijn determined E_m for a number of ferrites with widely different ρ, and also determined the activation energy E_r of the temperature dependence of resistivity: $\rho = \rho_\infty e^{E_r/kT}$. He found that the activation energies E_m and E_r are equal and range from 0.1 ev for fairly well-conducting ferrites to 0.4 ev for almost insulating ferrites.

These relaxation losses can be avoided by the prescriptions given by Snoek [29]: low ferrous ions content is obtained by firing at not too high temperature in oxygen, or by cooling of a porous ferrite in oxygen so that it can take up oxygen again.

It is also useful to use an excess of divalent oxide, and add some MnO in excess [44], because Mn^{3+} (or Mn^{4+}) and Fe^{2+} ions probably cannot coexist. The reason why the presence of Mn^{2+} and Mn ions with higher valency in the same lattice sites does not cause conductivity in Mn_3O_4 is not known. It has been suggested that Mn^{2+} and Mn^{4+} coexist, necessitating the transfer of two electrons simultaneously; also that Mn^{2+} and Mn^{3+} ions occur on different lattice sites, and that Mn^{3+} is semicovalently bound [12]. Whether the conductivity of $NiMn_2O_4$ [45] is due to the presence of Mn^{2+} and Mn^{3+} in B sites is not known.

H. Hysteresis Losses Distortion and Total Losses

In ferromagnetic metals it is the usual practice to separate the losses into eddy current, hysteresis, and residual losses; *e.g.*, by Legg's formula

$$\frac{R}{\mu f L} = ef + a\,B_{\max} + c \qquad (31)$$

in which e is loss coefficient for the eddy current losses, a for the hysteresis losses, and c for the residual losses. In ferrites the eddy current losses are negligible, the residual losses are composed of several types, and cannot be expressed by a simple constant in an equation for $R/\mu fL$. The hysteresis loss coefficient a can in principle be obtained from hysteresis measurements but it can be shown that such a simple constant cannot represent the hysteresis losses of ferrites, because higher order terms of H in a general relationship $B(H)$ for hysteresis loops are neglected in obtaining a, and it has been shown experimentally that this is not permissible when B exceeds 10 gauss. For several applications such as pupin coils and band-pass filters for telecommunica-

tion, the specifications contain a maximum value for a hysteresis loss factor, although the property that is really important is the distortion, defined as V_3/V_1 (=the voltage output of the third harmonic/idem of the fundamental), which is related to hysteresis loss factor.

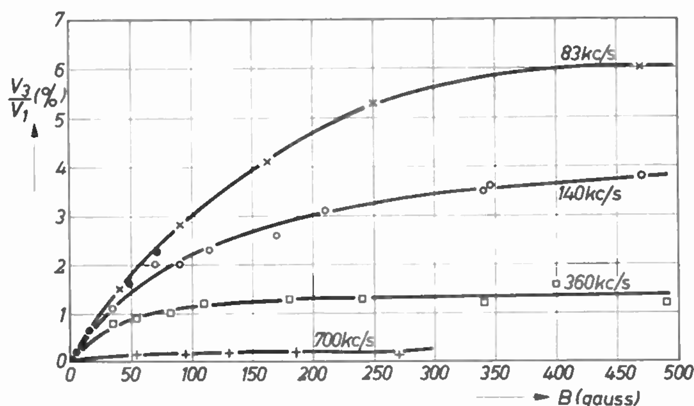


Fig. 23—Distortion V_3/V_1 in per cent as a function of the amplitude B_{max} measured at different frequencies. (● Distortion at 2 kc calculated from hysteresis resistance) for a $MnFe^{II}$ ferrite containing 43.5 per cent MnO and the balance $FeO+Fe_2O_3$.

As an example the distortion is given as a function of the amplitude B_{max} for different frequencies for a MnZn ferrite containing some Fe^{2+} ions in Fig. 23 [11]. The figure shows that the distortion strongly decreases with frequency and is practically zero at 700 kc. The shape of the hysteresis loop changes likewise with frequency; the typical loop with sharp tips has disappeared at 700 kc and the losses that remain are only due to a phase angle between B and H, as follows from the elliptical loop one finds at that frequency. In such ferrites where V_3/V_1 decreases with frequency, the hysteresis coefficient a as defined by (31) increases with frequency so that the relationship between a and V_3/V_1 is not valid here. The ferrites for which V_3/V_1 decreases with frequency are those for which, at higher amplitude H_{max}, a decrease of B_{max} with frequency occurs; *i.e.*, the distortion is also caused by irreversible domain-wall displacements.

Whereas in low-induction applications one is primarily interested in the quality factor Q and in the distortion, the losses in high induction applications are unpleasant because they cause power loss, temperature rise, and, if these become intolerable, increase in size of the core. One is then interested in the total energy dissipation only, and the losses may for higher inductions conveniently be determined calorimetrically.

Such total loss measurements have been made, and also at lower inductions by another method, but since it is not our aim to provide a list of properties of the ferrites, but only to show how they are influenced by the chemistry, we refer to the literature [11].

I. Square Hysteresis Loops

In the application of materials with a square hysteresis loop, properties are important which are completely different from those of interest for the applications of other ferrites.

In a coincident-current magnetic memory [46] a ferrite core receives one current pulse or two equal simultaneous pulses from two primary windings which bring it from the state I_0 (Fig. 24), along the dashed inner loop into the states $-\frac{1}{2}H_m$ and $-H_m$ respectively. After the pulse the material reverts to the states I_0 and $-I_0$ respectively, and the ratio of the voltage outputs in a secondary winding caused by the double and single pulse must be as large as possible; *i.e.*, the loop must be as rectangular as possible. The "squareness ratio" $R_s = I_{(-(1/2)H_m)}/I_{(H_m)}$ is often used as a quality factor, and since this ratio is dependent on H_m, the maximum value is $(R_s)_{max}$. For other applications $(I_0/I_{H_m})_{max}$ is important. Since it is desirable in many cases to use small current pulses, H_m, and therefore H_c, must be small. The first and most widely used square loop ferrite was a MnMg ferrite [47].

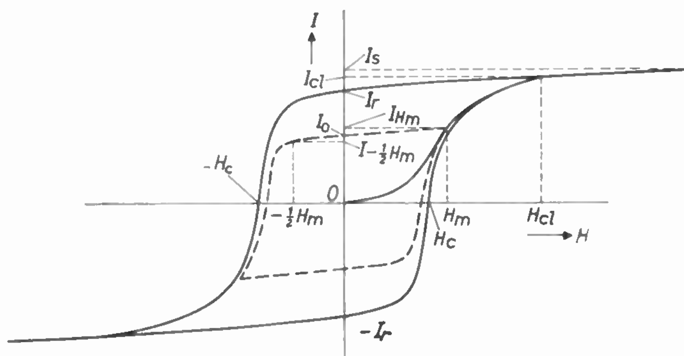


Fig. 24—Hysteresis loop with the symbols used in the text.

We shall now see which conditions must be fulfilled to obtain a square hysteresis loop [48], and we shall, therefore, first see what happens when a polycrystalline material with random crystallite orientation is magnetized to saturation and then allowed to revert to the remanent state (Fig. 25). The originally random orientation of the domain magnetizations (a) are first all turned into the direction of the saturating field (b), and then at the remanence point fan out into a sector of a sphere, the half-angle of which depends on the number of anisotropy directions in the crystals. If the crystalline anisotropy is predominant, the greatest angle a preferred direction of the magnetization can make with the field is 54.5° (d), *i.e.*, for $K_1<0$ the angle between a 111 axis and a field in a 100 direction.

If the strain anisotropy predominates, there are only two preferred directions at each point (*i.e.*, for $K_1<0$ parallel to the greatest compressive strain or the smallest tensile strain) so that for random strain orientation the half-angle of the sector will be 90° (c).

When air pores are present in the material, free poles will be present on their surface, causing demagnetizing fields opposed to the applied field. When a positive

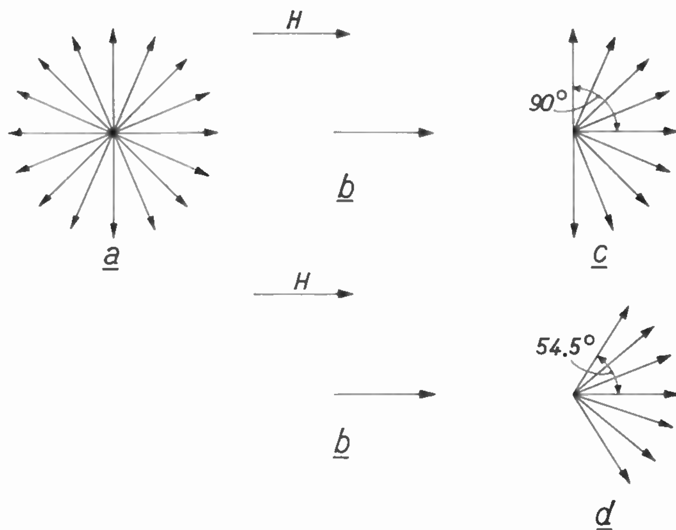


Fig. 25—(a) Random domain orientations in the demagnetized state. (b) In the saturated state all domain orientations are parallel. (c) Domain orientations in remanent state for uniaxial anisotropy. (d) For negative crystalline anisotropy.

applied field is reduced to zero the material will in the remanent state still feel a negative demagnetizing field so that the domain orientations will be more randomly distributed than in the case that the other anisotropies would be present only. Apart from the above "shearing" effect, specially nonspherical pores will cause local demagnetizing fields with random directions, which may be stronger than the "fields" due to the other anisotropies. These demagnetizing fields at any point can have only two preferred directions; *i.e.*, in remanence only one, so that the half-angle of the sector is again 90°.

It can be easily calculated that for the case of cubic crystalline anisotropy I_r/I_s = about 0.87, for the case of strain anisotropy and shape anisotropy (neglecting shearing) $I_r/I_s = 0.5$. The most favorable condition would be that in which there is only one preferred direction in the polycrystalline material; then I_r/I_s would be 1. This can in principle be brought about by using a ferrite with very small crystalline anisotropy and large magnetostriction strained in such a way that the magnetization vector due to the strain anisotropy lies in the field direction. This has been done by G. W. Rathenau and G. W. van Oosterhout [48] by melting a glass ring onto the outer cylindrical surface of a ring of NiZn ferrite with small K_1; *i.e.*, with high Zn content. Cooling caused tangential compression, and therefore preferred directions in the field direction in the wound core.

A family of hysteresis loops obtained on such a strained core is shown in Fig. 26.

Similar results have been obtained by enclosing a ferrite ring in a resin [49].

In unstrained ferrites, in order that I_r/I_s be as high as possible, the crystalline anisotropy must be pre-

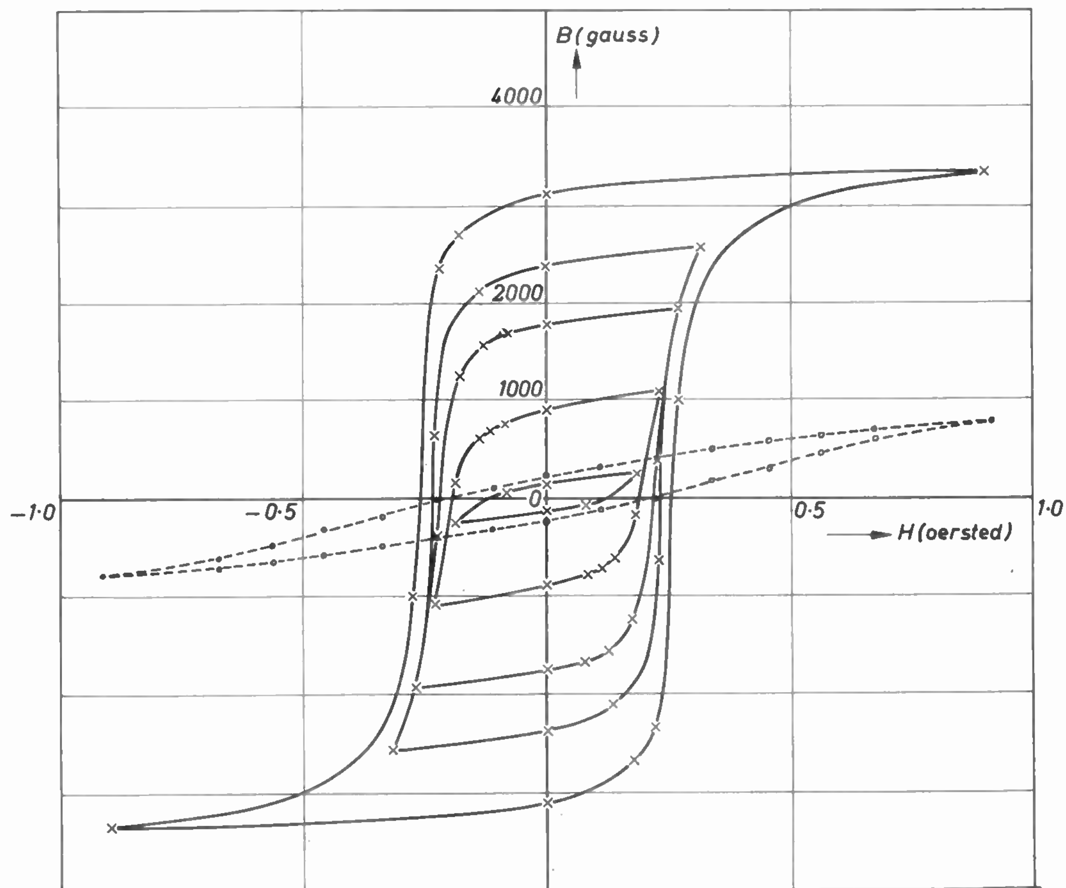


Fig. 26—Family of hysteresis loops for a ring of Ferroxcube IV *A* stressed by a ring of glass. Broken curve: hysteresis loop of Ferroxcube IV *A* without glass.

dominant; *i.e.*, it must be high, but even in the absence of other anisotropies the anisotropy field H_A should be high with respect to H_c in order to obtain a high $(R_s)_{max}$, because in small negative fields domain rotations must be prevented as much as possible. Since a small coercive force is desired, and H_c increases with K_1, one must choose a material with a moderate value of K_1. At the same time the other anisotropies must be reduced as much as possible, since they tend to reduce I_r/I_s.

Shape anisotropy can be suppressed by firing to very high densities (Fig. 7(c)), which moreover, lowers the coercive force.

Even in the (hypothetical) case that internal stresses are absent in the demagnetized state, there will still be stresses in the remanence point, because the change of domain orientations causes a change of the shape of the individual crystals. Only if (for negative K_1) $\lambda_{111}=0$ is there no change in strain when the domain magnetization changes from one preferred direction into another so that no extra stresses occur in the remanent point. This means that not λ_s, but the magnetostriction in the preferred direction must be zero which has been demonstrated for solid solutions of $NiFe_2O_4$ ($\lambda_s=-26.10^{-6}$) and Fe_3O_4 ($\lambda_s=+41$) [48]. For 56 per cent $NiFe_2O_4$, $\lambda_s=\approx 0$, but for 70 per cent $NiFe_2O_4$ ($\lambda_s=\approx -8.10^{-6}$) the hysteresis loop was most square. Since the magnetization process in low fields, of the order of H_c, is a change of the domain magnetization orientation from one preferred direction into another, the magnetostriction in such low fields is mainly determined by λ_{111}. This low-field magnetostriction was shown to be zero just for this composition.

The effect of magnetostriction is also notable in $Li_{0.5}Fe_{2.5}O_4$ which has $\lambda_s=-1.10^{-6}$ when pure. Firing at 1,150°C causes some Fe^{2+} ions to be present, and λ_{111} is slightly positive. This can be reduced to zero by starting from a solid solution of $Li_{0.5}Fe_{2.5}O_4$ and some $NiFe_2O_4$ with negative λ_{111}.

It has been possible to show that a large number of chemical compositions yield square-loop ferrites [48]. For a few of these R_s and I_0/I_{H_m} are plotted against H_m in Fig. 27.

About the background of properties under pulse conditions little has been published. It has been found that the product of the switching time and the "excess" field $(H-H_c)$ is a constant for a material and can be used as a quality factor of the material.

J. Ferrites at Microwave Frequencies

It can be seen from Fig. 19 that the ferrites cannot be used in the ordinary way at microwave frequencies ($\approx$1,000–24,000 mc), because of their very high losses. These losses can be explained almost entirely in terms of ferromagnetic resonance absorption in the effective fields due to the crystal anisotropy. We have seen that due to demagnetization on domain walls the highest effective field is $\approx 4\pi I_s$, and therefore these losses are decreased by the application of a dc magnetic field, and in the completely saturated state they should vanish.

When a magnetized ferrite is subjected to a hf field perpendicular to the dc field H_z, a circular precession of the magnetization vector will result, the sense of which is determined with the direction of the dc field. The precessing magnetization vector can be resolved into a fixed part and a rotating part (Fig. 28(a)) and the hf field into two circularly polarized fields. These rotating fields and magnetizations can have the same or opposite sense of rotation so that one can define two different susceptibilities [50], which can be shown to be $\chi^+=I_s/(H-\omega/\gamma)$ and $\chi^-=I_s/(H+\omega/\gamma)$. This results in different wavelengths of the circularly polarized fields in the ferrite, so that the plane of polarization of the traveling wave is rotated by the ferrite (Faraday rotation). This phenomenon has been found by Roberts [51] and can be applied [52] in unidirectional isolators, circulators, gyrators, etc.

Another class of devices is based on the facts that in a certain mode the fields in a waveguide are elliptically polarized and the sense of elliptical rotation is opposite for opposite directions of propagation, and also for points located symmetrically in a waveguide cross section.

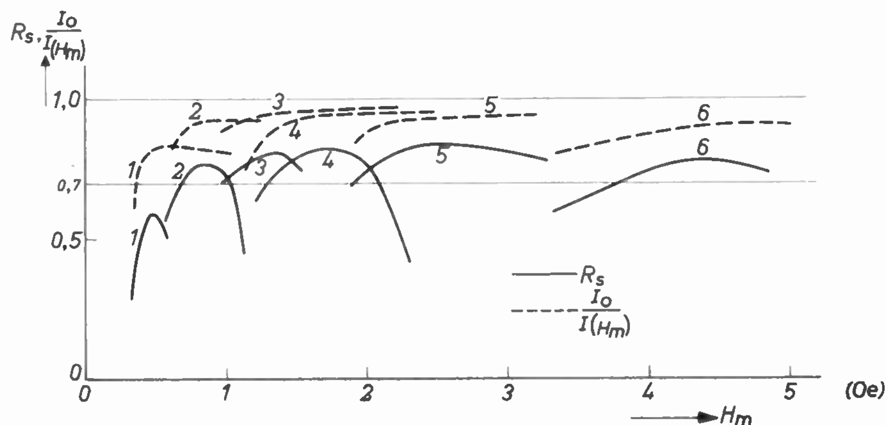


Fig. 27—R_s and $I_0/I_{(H_m)}$ against H_m for the following ferrites. (1) $Co_{0.02}Mn_{0.98}Fe_2O_4$. (2) $(CuO)_{0.1}\cdot(MnO_{1+\delta})_{1.1}\cdot Fe_2O_3$. (3) $(MgO)_{0.5}\cdot(MnO_{1+\delta})_{0.875}\cdot Fe_2O_3$. (4) $Mn_{0.1}Ni_{0.5}Mg_{0.4}Fe_2O_4$. (5) $Mg_{0.4}Ni_{0.6}Fe_2O_4$, (6) $Li_{0.46}Ni_{0.08}Fe_{2.46}O_4$.

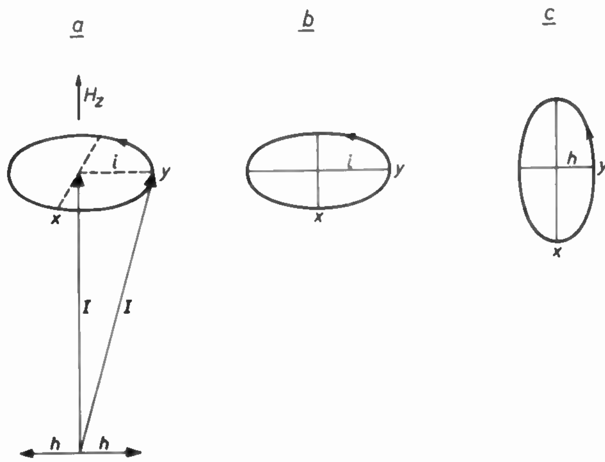


Fig. 28—(a) The precessing magnetization vector I can be resolved into a fixed part ($\approx I$) and a rotating part i. (b) The rotating part of the magnetization describes an ellipse because the stiffnesses in the x and y directions are different. (c) Elliptical polarization of h in a certain location in the waveguide. See text.

If a magnetized piece of ferrite with a certain shape, magnetized in the z direction, is introduced asymmetrically in a waveguide with length in the x direction, the resonance condition is, according to Kittel [35] and Snoek [33]:

$$\omega_r = \gamma [H_{\text{eff.}_z} H_{\text{eff.}_y}]^{1/2}$$
$$= \gamma [\{H_z + I_s(N_x - N_z)\}\{H_z + I_s(N_y - N_z)\}]^{1/2}.$$

Since the stiffness is different in the x and y directions due to the demagnetizing fields, the rotational part of I, i, will perform an elliptical rotation with axes inversely proportional to $(H_{\text{eff.}_z})^{1/2}$ and $(H_{\text{eff.}_y})^{1/2}$ (Fig. 28(b)). When the ellipticity is the inverse of that of the field at the location of the piece of ferrite (Fig. 28(c)) in the waveguide and the direction of rotation of i opposite to that of h, energy is not transferred during a cycle; *i.e.*, resonance is not excited. A traveling wave will therefore not be attenuated in one direction x, but in the opposite direction $-x$ the reversed sense of rotation of h causes resonance; *i.e.*, there is energy transfer to the ferrite, which results in an attenuation of the wave. Ratios of reverse to forward attenuation up to 50 have been obtained. Unidirectional attenuator devices based on this principle are described by Kales [53] and by Fox *et al.* [54].

At the dc field strengths H_z that are needed (up to $\approx$2,000 oersteds) the resonance losses such as occur in the nonmagnetized ferrite are apparently not completely absent, so that $H_A + 4\pi I_s$ of the ferrite has to be made smaller than ω/γ. This puts an upper limit to $4\pi I_s$, and for longer wavelengths the normal ferrites developed for ordinary applications are not suitable. The $4\pi I_s$ value is conveniently decreased by having a fair amount of $MgFe_2O_4$ ($4\pi I_s = 1{,}400$) in the solid solution [63] or by replacing part of the Fe^{3+} ions by Al^{3+} ions [44].

Very good results are obtained with Mg ferrite-aluminates. The $4\pi I_s$ value should not be too low because the energy absorption is proportional to I_s, so that for low I_s more ferrite is needed, which increases the forward attenuation.

V. History of the Development of Ferrites

The oldest ferrite is the mineral magnetite, Fe_3O_4. This is of magmatic origin and is therefore appreciably older than human life, so that it had to wait a long while for its mention in literature. Thales of Miletus (640–548 B.C.) "observed that the lodestone had the power to lift small pieces of iron and thus counteract their natural desire to fall downward." "He endowed the lodestone with a soul in the belief that nothing less than an inherent living force could account for the observed effect" [55]. We shall skip some 2,500 years and for the older literature on natural magnetite refer to [55].

Although there is a number of patents on the use of cores made from powered natural magnetite, we shall concern ourselves only with made-to-measure ferrites.

List [56] noted the formation of magnetic oxides consisting of MO (M = Mg, Mn, Zn, Ni or Cu) and Fe_2O_3, which were attracted by a magnet.

Hilpert [57] was the first to realize that such materials might be useful magnetic core materials, better than Fe_3O_4, because of their higher resistivities. In 1909 he took out a number of patents, but his ferrites had very low permeability and must have had enormous losses, considering his methods of preparation.

Forestier [58] made several ferrites by precipitation with alkali hydroxide from chloride solutions and heating the precipitate up to 900°C. His method of investigation was the measurement of the magnetization in a field of about 500 oersteds as a function of temperature. By this method he established the existence of a compound $MgFe_2O_4$, which, however, had a Curie temperature of 315°C. He then made materials with the composition MFe_2O_4 in which M = Cu, Ni, Mg, Sr, Ba, Pb, Ca, Cd, and determined the magnetizations and Curie temperatures, also for those which we now know to be nonferromagnetic, which seems to indicate that the materials had not fully reacted. Forestier was the first of a great number of pupils of Chaudron, who may be considered the father of a school of French chemists working on ferromagnetic spinels, all using the same method of investigation. Their work, which greatly increased our knowledge on the substitution possibilities, is extensively cited in [59]; not all of it can stand criticism founded on more recent knowledge.

Kato and Takei [60] published a number of papers in Japanese from 1933 to 1939 which have since become known. Because they used low firing temperatures, they were specially interested in $CuFe_2O_4$ which sinters at a low temperature, and made M-Zn ferrites, which, however, did not have initial permeabilities above 50. Van Arkel, Verwey, and van Bruggen [61] in 1936 established the phase relations in several systems MO—Fe_2O_3 (M = Mn, Cu, Ni, Mg) and Snoek in 1936 [62] confirmed and supplemented their results by measurements

of coercive force and saturation magnetization in these systems.

It was Snoek who in 1941 established the importance of the correct oxygen content, not possessed by Hilpert ferrites, for the quality of ferrites for high frequency use, even in those materials where the ferrous content is so small that eddy current losses are negligible. He also established the importance of extreme homogeneity, not possessed by the Japanese ferrites.

Dr. J. L. Snoek, whose work laid the foundations for an important new industry. Dr. Snoek died in 1950 in a road accident.

The requirement for the oxygen content limits the firing temperature, and homogeneity is best obtained at very high firing temperatures. Snoek showed how these two conditions could be fulfilled simultaneously. Thirdly, he found that solid solutions M-Zn ferrites showed the highest initial permeabilities, together with low losses. These losses are low up to a certain frequency, which frequency is about inversely proportional to the permeability, which in turn he showed to be connected with the Zn ferrite content and with the firing temperature. Finally he showed that for the lower frequency range ferrites; *i.e.*, those with higher Zn ferrite contents, the permeability was further increased by incorporating small amounts of Fe_3O_4 in the solid solution, because this could reduce the magnetostriction to practically zero. These very important results are laid down in a number of patents and in his book [29] which was the starting signal for a great many investigations both of a scientific and technical nature, and of a very rapid growth of technical applications. The great majority of present-day applications use ferrites which are based on Snoek's work, although research and development in many laboratories have resulted in a gradual improvement of several properties.

These present-day applications include

For radio: IF band-pass filters, antennas, elements for permeability tuning, transformers.

For television: line output transformers, deflection coils, linearity[15] and width control of image, aerial transformers, noise suppression filters.

[15] Biased by a permanent magnet of *e.g.*, Magnadur (Ferroxdure).

For telecommunication: filters for carrier telephony, hf chokes, equalizing cores, delay lines, broad band audio transformers, pulse transformers[15].

For miscellaneous applications: pulse transformers, ultrasonic vibrators, mechanical filters, recording heads for tape recorders, etc., frequency modulation, hf coupling, ignition coils, low-power transmission, magnetic amplifiers, and saturable reactors, various vital parts in proton accelerators and other elementary-particle accelerating machines, modulation of uhf in waveguides, gyrators, etc.

Two important additions to the arsenal of ferrite applications useful in industry originate from the United States [63, 47]: the ferrites used for unidirectional attenuators and gyrators in microwave technique, and the ferrites with square hysteresis loops, that can be used as switching cores for magnetic memories and numerous other applications, neither of which class of ferrites contains zinc ferrite as an essential component.

The present-day yearly output of ferrites may be estimated to be well over 6,000 tons, which represent a value of $15,000,000.

Bibliography

[1] Néel, L., *Annales de Physique*, Paris, Vol. 17 (1932), p. 64; and Vol. 5 (1936), p. 256.

[2] Clark, C. A., and Sucksmith, W., *Proceedings of the Royal Society* (London) A, vol. 225 (1954), pp. 147–159.

[3] Néel, L., *Annales de Physique*, Paris, Vol. 3 (1948), pp. 137–198.

[4] Shull, C. G., and Smart, J. S., *Physical Review*, Vol. 76 (1949), pp. 1256–1257.

[5] Kramers, H. A., *Physica*, Vol. 1 (1934), p. 182.

[6] Anderson, P. W., *Physical Review*, Vol. 79 (1950), pp. 350–356, 705–710.
Van Vleck, J. H., *Journal de Physique et le Radium*, Vol. 12 (1951), pp. 262–274.

[7] Foëx, G., and Graff, M., *Comptes Rendus de l'Academie des Sciences*, Paris, Vol. 209 (1939), p. 160.
Volger, J., *Nature*, Vol. 170 (1952), p. 1027.

[8] Jonker, G. H., and van Santen, J. H., *Physica*, Vol. 16 (1950), pp. 337–349.

[9] de Boer, J. H., and Verwey, E. J. W., *Proceedings of the Physical Society*, Vol. 49A (1937), pp. 59–71.

[10] Zener, C., *Physical Review*, Vol. 82 (1951), pp. 403–405.

[11] Smit, J., and Wijn, H. P. J., "Physical Properties of Ferrites." *Advances in Electronics and Electron Physics*, Vol. 6 (1954), pp. 69–126, (see there also for literature refs.).

[12] Goodenough, J. B., and Loeb, A. L., *Physical Review*, Vol. 98 (1955), pp. 391–408.

[13] Bragg, W. H., *Nature*, Vol. 95 (1915), p. 561; *Philosophical Magazine*, Vol. 30 (1915), pp. 305–315.
Nishikawa, S., *Proceedings of Tokyo mathematical-physical Society*, Vol. 8 (1915), pp. 199–209.

[14] Gorter, E. W., *Philips Research Reports*, Vol. 9 (1954), pp. 295–365, 403–443.

[15] Verwey, E. J. W., *Zeitschrift für Kristallographie*, Vol. 91A (1935), pp. 65–69.
Hägg, G., *Zeitschrift für physikalische Chemie*, Vol. 29B (1935), p. 95.

[16] Barth, T. F. W., and Posnjak, E., *Zeitschrift für Kristallographie*, Vol. 82 (1932), pp. 325–341.

[17] Verwey, E. J. W., and Heilmann, E. L., *Journal of Chemical Physics*, Vol. 15 (1947), pp. 174–180.

[18] Bertaut, F., *Journal de Physique et le Radium*, Vol. 12 (1951), pp. 252–255.

[19] Pauthenet, R., *Comptes Rendus de l'Academie des Sciences*, Paris, Vol. 230 (1950), pp. 1842–1843.
Néel, L., *Comptes Rendus de l'Academie des Sciences*, Paris, Vol. 230 (1950), pp. 190–192.

[20] Verwey, E. J. W., de Boer, F., and van Santen, J. H., *Journal of Chemical Physics*, Vol. 16 (1948), pp. 1091–1092.
de Boer, F., van Santen, J. H., and Verwey, E. J. W., *Journal of Chemical Physics*, Vol. 18, (1950), pp. 1032–1034.

[21] van Santen, J. H., and van Wieringen, J. S., *Recueil des travaux chimiques des Pays-Bas*, Vol. 71 (1952), pp. 420–430.
Romeijn, F. C., *Philips Research Reports*, Vol. 8 (1953), pp. 304–342.
[22] Yafet, Y., and Kittel, C., *Physical Review*, Vol. 87 (1952), pp. 290–294.
[23] Néel, L., *Comptes Rendus de l'Academie des Sciences*, Paris, Vol. 230 (1950), pp. 375–377.
[24] Corliss, L. M., and Hastings, J. M., N.O.L. Conference on Ferrimagnetism, Washington, D. C., October 11–12, 1954.
[25] Guillaud, C., and Creveaux, H., *Comptes Rendus de l'Academie des Sciences*, Paris, Vol. 230 (1950), p. 1459.
Guillaud, C., *Journal de Physique et le Radium*, Vol. 12 (1951), pp. 239–248.
[26] Maxwell, L. R., and Pickart, S. J., *Physical Review*, Vol. 92 (1953), pp. 1120–1126.
[26a] McGuire, T. R., *Physical Review*, Vol. 93 (1954), pp. 682–686.
Smart, J. S., *Physical Review*, Vol. 94 (1954), pp. 847–850.
[27] Maxwell, L. R., and Pickart, S. J., *Physical Review*, Vol. 96 (1954), pp. 1501–1505.
[28] Gorter, E. W., and Schulkes, J. A., *Physical Review*, Vol. 90 (1953), pp. 487–488.
[29] Snoek, J. L., *New Developments in Ferromagnetic Materials*. New York-Amsterdam, Elsevier Publishing Company, 1947.
[30] Went, J. J., and Wijn, H. P. J., *Physical Review*, Vol. 82 (1951), p. 269.
Wijn, H. P. J., and Went, J. J., *Physica*, Vol. 17 (1951), p. 976.
[31] Williams, H. J., and Shockley, W., *Physical Review*, Vol. 75 (1949), p. 178.
Galt, J. K., *Physical Review*, Vol. 85 (1952), p. 664.
Galt, J. K., Andrus, J., and Hopper, H. G., *Reviews of Modern Physics*, Vol. 25 (1953), p. 93.
[32] Wijn, H. P. J., Gevers, M., and van der Burgt, C. M., *Reviews of Modern Physics*, Vol. 25 (1953), pp. 91–92.
[33] Snoek, J. L., *Nature*, Vol. 160 (1947), p. 90.
Snoek, J. L., *Physica*, Vol. 14 (1948), p. 207.
[34] Landau, L., and Lifshitz, E., *Physikalische Zeitschrift der Sowjetunion*, Vol. 8 (1935), p. 153.
[35] Kittel, C., *Physical Review*, Vol. 73 (1948). p. 155.
[36] Polder, D., and Smit, J., *Reviews of Modern Physics*, Vol. 25 (1953), p. 89.
[37] Beljers, H. G., van der Lindt, W. J., and Went, J. J., *Journal of Applied Physics*, Vol. 22 (1951), p. 1506.
[38] Rado, G. T., Wright, R. W., and Emerson, W. H., *Physical Review*, Vol. 80 (1950), p. 273.
Rado, G. T., Wright, R. W., Emerson, W. H., and Terris, A., *Physical Review*, Vol. 88 (1952), p. 909.
Rado, G. T., *Reviews of Modern Physics*, Vol. 25 (1953), p. 81.
[39] Wijn, H. P. J., *Physica*, Vol. 19 (1953), p. 555.
[40] Koops, C. G., *Physical Review*, Vol. 83 (1951), p. 121.
[41] Kamiyoski, K., *Physical Review*, Vol. 84 (1951), p. 374.
Volger, J., *Semi-Conducting Materials*. New York, Butterworth Scientific Publications, 1951; p. 162.
Fairweather, A., and Frost, E. J., *Proceedings of the Institute of Electrical Engineers* (London), 1953.
[42] Brockman, F. G., Dowling, P. H., and Steneck, W. G., *Physical Review*, Vol. 75 (1949), p. 1440. Vol. 77 (1950), p. 85.
[43] Wijn, H. P. J., and van der Heide, H., *Reviews of Modern Physics*, Vol. 25 (1953), p. 98.
[44] van Uitert, L. G., Schafer, J. P., and Hogan, C. L., *Journal of Applied Physics*, Vol. 25 (1954), pp. 925–926.
[45] Kamyama, M., and Nara, Z., *Chemical Abstracts*, Vol. 47 (1953), p. 8505(i).
[46] Papian, W. N., PROCEEDINGS OF THE IRE, Vol. 40 (1952), pp. 475–478.
[47] Albers-Schoenberg, E., *Journal of Applied Physics*, Vol. 25 (1954), pp. 152–154.
[48] Wijn, H. P. J., Gorter, E. W., Esveldt, C. J., and Geldermans, P., *Philips Technical Review*, Vol. 16 (1954), pp. 49–58.
[49] Williams, H. J., Sherwood, R. C., Goertz, M., and Schnettler, F. J., *Communication Electronics*, Vol. 9 (1953), p. 531.
[50] Polder, D., *Philosophical Magazine*, Vol. 40 (1949), p. 99.
[51] Roberts, F. F., *Journal de Physique et le Radium*, Vol. 12 (1951), p. 305.
[52] Hogan, C. L., *Bell System Technical Journal*, Vol. 31 (1952), p. 1.
Reviews of Modern Physics, Vol. 25 (1953), p. 253.
[53] Kales, M. L., *Journal of Applied Physics*, Vol. 24 (1953), pp. 604–608.
[54] Fox, A. G., Miller, S. E., and Weiss, M. T., *Bell System Technical Journal*, Vol. 34 (1955), pp. 5–103.
[55] Still, A., *The Soul of Lodestone*, New York-Toronto, Murray Hill Books Inc., 1946.
[56] List, K., *Berichte der deutschen chemischen Gesellschaft*, Vol. 11 (1878), p. 1512.
[57] Hilpert, S., *Berichte der deutschen chemischen Gesellschaft*, Vol. 42 (1909), pp. 2248–2261.
[58] Forestier, H., *Annales Chimie Xe série Tome IX* (1928), pp. 353–401.
[59] Michel, A., Chaudron, G., and Bénard, J., *Journal de Physique et le Radium*, Vol. 12 (1951), pp. 189–201.
[60] Kato, Y., and Takei, T., *Journal of the Institute of Electrical Engineers*, Japan, Vol. 53 (1933), pp. 408–412.
Takei, T., *Journal of the Institute of Electrical Engineers*, Japan, Vol. 59 (1939), pp. 274–282.
[61] van Arkel, A. E., Verwey, E. J. W., and van Bruggen, M. G., *Recueil des travaux chimiques des Pays-Bas*, Vol. 55 (1936), pp. 331–339.
Verwey, E. J. W., van Arkel, A. E., and van Bruggen, M. G., *Recueil des travaux chimiques des Pays-Bas*, Vol. 55 (1936), pp. 340–347.
[62] Snoek, J. L., *Physica*, Vol. 3 (1936), pp. 463–483.
[63] Albers-Schoenberg, E., and Goldsmith, H. A., *Electrical Manufacturing*, Vol. 44 (1949), p. 86.

Correspondence

Noise Factor Measurement*

When a receiver is being designed for the lowest possible noise factor, the working out of the theoretical design is usually followed by an experimental program in which various adjustments are made concurrently with a series of noise factor measurements. When these measurements are being carried out, the importance of the temperature of the noise diode resistor is not always fully realized. While obtaining the best absolute accuracy in the noise factor measuring equipment is a laudable aim, it is the day-to-day and hour-to-hour repeatability of the noise factor readings which will ultimately limit the approach to the best design. In the usual case, the noise diode resistor will be warming up, because of its proximity to various components, as the tests proceed throughout the day. A very gratifying improvement in repeatability of readings is often achieved by the following method of correction for temperature.

The noise diode resistor is ventilated as well as possible, and a thermometer is mounted near the resistor. A correction of

$$\Delta I = (r - 1)\frac{\Delta T}{290}\cdot\frac{1}{20R} \qquad (1)$$

is subtracted from the noise diode current reading before it is inserted in the usual formula,

$$F = \frac{1}{r - 1}\cdot 20IR. \qquad (2)$$

Here, F = noise factor as a power ratio, ΔT = temperature rise of the resistor above 16.8°C., I = noise diode current, R = noise diode load resistance, and r = the ratio of amplified noise power with the noise diode on to its value due to receiver noise only (usually 2).

A convenient method of applying this correction is to fasten a scale of ΔI to the thermometer.

It is not easy to predict whether or not the temperature correction is worthwhile in a particular case. The maximum temperature rise may be measured, say ΔT_{max}. The lowest noise factor to be expected may be called F_{min}, as a power ratio. Then:

$$\Delta F_{DB} \doteq 4.343\frac{\Delta T_{max}}{290}\frac{1}{F_{min}} = \frac{\Delta T_{max}}{66.8F_{min}}, \qquad (3)$$

where ΔF_{DB} is the error in noise factor in decibels. As an example, if F_{min} is 1.20 (as a power ratio) and ΔT is 10.0°C., ΔF_{DB} is found to be 0.125 decibels. In general this error will be increasing from zero to its maximum value throughout the day, and can distort the experimental results much more than in the case of a random error.

The derivation of the above equations is straightforward, and the modifications for push-pull or other special input connections are obvious. It might be pointed out that the error has nothing to do with the temperature coefficient of resistance of the resistor, but rather is a consequence of two facts; first that noise factor must be defined at a given temperature (otherwise it will not be a constant of a receiver), and second that the thermal noise power from the noise diode resistor is proportional to its absolute temperature.

In the above equations standard temperature has been assumed to be 16.8°C., this value is implied when using the usual numerical coincidence $F = 20IR$, for the case $r = 2$.

A. C. HUDSON,
Microwave Section,
Div. of Radio and Electrical Engrg.,
National Res. Council.
Ottawa, Ontario, Canada.

* Received by the IRE, August 25, 1955.

Germanium Transistor Amplifiers Stable to 95°C.*

Reference to manufacturers specifications shows that current gain of some germanium transistors (the GE 2N43 for example) does not change greatly up to 100°C. The operating point, however, does shift radically, sending open-loop amplifiers into cut off at far lower temperatures. If the operating point can be stabilized, amplifier gain can be maintained over wide temperature ranges.

In the process of developing a series of temperature insensitive amplifiers, a rather simple means was found of stabilizing the operating point when the transistors are connected in triode amplifier-cathode follower pairs. A parallel RLC filter is put into a feedback loop about each pair. The percentage of feedback is high at dc due to the low DC resistance of L and stabilizes the operating point. The gain of the pair at the resonant frequency of the LC combination is then set by the value of the shunt resistance R which may be the equivalent shunt resistance of the inductor but is usually a separate component. For closed loop gains of ten or less, sufficient dc feedback is provided by the resistor alone and the L and C may be eliminated. The battery fixes the operating point at 6 volts, since the dc loop gain will hold the input dc voltage near zero. In effect, the dc gain of the amplifier is utilized to provide operating point stability rather than being thrown away through a coupling capacitor.

In a gain-of-ten configuration utilizing 2N43A transistors as shown in Fig. 1, closed loop gain has been consistently maintained within ±2 per cent from room temperature to 98.5°C. Using moderately high-gain silicon transistors (Texas Instrument #904) closed-loop gain has been maintained within the same limits to over 135°C. With low gain silicon transistors (Texas Instrument #903) closed-loop gain has been maintained ±5 per cent to 150°C. Surprisingly enough, silicon and germanium units show similar open-loop high temperature effects,

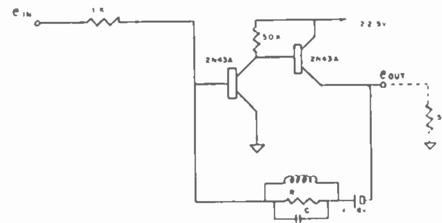


Fig. 1

drifting into cut off at 40° to 50°C in our configurations. Above 100°C, no amount of dc feedback will hold germanium units stable but silicon units are still easy to stabilize. Germanium units have been stored at 101°C for half-hour periods and immediately operated for another half hour at 98°C at 10 per cent of maximum rated power. No degradation of performance or permanent damage to the transistors has been observed although this treatment is in excess of ratings.

W. GREATBATCH
Dept. of Elec. Engrg.
Univ. of Buffalo and Cornell Aeronautical Lab., Inc.
Buffalo, N.Y.

W. HIRTREITER
Flight Res. Dept.
Cornell Aeronautical Lab., Inc.
Buffalo, N.Y.

* Received by the IRE, July 21, 1955. This work was performed as part of U. S. Navy contract NOa(s) 53-1107-C.

Power Flow*

Most of the discussions on terminology in an established engineering field lead nowhere, and ordinarily I would refrain from this type of debate. There are many cases of inaccurate use of terms which are harmless because everyone involved understands what is meant. Thus, one is seldom misled by the use of "megacycles" for "megacycles per second," although the term "hertz" has been, and is sometimes employed to make this inaccuracy unnecessary.

The recent exchange of letters on the terms "instantaneous frequency" and "instantaneous spectra" has been illuminating, I think, for here a somewhat different concept is involved, and careful thinking is necessary.

However, when there is a well-defined usage available, and where dimensions are involved, it would seem that it is best to adhere to the correct terminology. Therefore, I wish to point out this once the increasing substitution of the word "power" where "energy" is meant. This is something like the awkward question of "current" and

* Received by the IRE, July 11, 1955.

"charge flow," but in the case in point usage has been correct until recently. Unfortunately, when I looked for examples, I found that the Editor of the PROCEEDINGS OF THE IRE furnishes the majority of the examples. Hence, I wish to make it clear, that the quality of the work is not impugned, but I do protest the careless terminology.

The first case is where the term appears even in the title of the paper.[1] Here, in the second paragraph we find, "P_e is the Poynting vector, which we may interpret as the electromagnetic power flow per unit area. . . . " Further on, one finds

$$\nabla\cdot(P_e + P_k) + \frac{\partial}{\partial t}(W_e + W_k) = 0. \quad (11)$$

"As P_e and W_e are electromagnetic power flow and electromagnetic energy, so P_k and W_k are kinetic power flow and kinetic energy."

Why the superfluous "flow"? A search of the better books on electromagnetic theory reveals that Poynting's vector has always been defined as "energy flow." If "flow" is to be taken as a time rate, then the dimensions in (11) are incorrect.

Further examples are found in many papers. Heffner is in and out of this mistake all through his paper,[2] and it was in this paper that I first was really confused. We find, "The power flowing from the gap to the beam, is

$$P = \tfrac{1}{2}\,\mathrm{Re}\, Vi^* \quad (15)."$$

However in the same paragraph, one finds, "The fact that the waves carry power in opposite directions has been pointed out previously by Chu and Walker. It arises because of the opposite signs attached to the stored energy of the two waves."

This seems to be enough. All I am asking is that reasonable care be employed in writing on this subject so that one does not confuse, unnecessarily, an already difficult subject.

W. M. GOTTSCHALK
Res. Div., Raytheon Mfg. Co.,
Waltham 54, Mass.

Rebuttal[3]

I do not regard *flow* as having the dimensions $1/t$, but merely as a descriptive word, perhaps superfluous, but scarcely confusing. Schelkunoff[4] speaks of "longitudinal power flow" and just plain "power flow" in connection with a waveguide, so I feel that if I am wrong I am wrong in good company. I do not object either to current flowing in a circuit, though it may be that to the pure minded the current either stands still or else merely exists.

J. R. PIERCE
Bell Telephone Labs., Inc.
Murray Hill, N.J.

[1] W. H. Louisell and J. R. Pierce, "Power flow in electron beam devices," PROC. IRE, vol. 43, pp. 425–427; April, 1955.
[2] H. Heffner, "A coupled mode description of beam-type amplifiers," PROC. IRE, vol. 43, pp. 210–217; February, 1955.
[3] Received by the IRE, July 14, 1955.
[4] S. A. Schelkunoff, "Electromagnetic Waves," D. Van Nostrand Co., Inc., New York, N. Y., p. 319; 1943

Nonuniform Transmission Lines as Impedance-Matching Sections*

A great improvement upon the lines proposed by Scott[1] and Yang[2] can be obtained if the characteristic impedance of the matching section (Fig. 1) can be varied according to one of the following equations:

$$\frac{d}{dy}(\ln Z_0) = (1 - 0.889\cos 2\pi y + 0.0112\cos 4\pi y)\ln(Z_2/Z_1), \quad (1)$$

and

$$\frac{d}{dy}(\ln Z_0) = \left(\sin\pi y - \frac{1}{3}\sin 3\pi y + \frac{1}{54}\sin 5\pi y\right)\frac{\frac{\pi}{2}\ln(Z_2/Z_1)}{1 - \frac{1}{9} + \frac{1}{270}}, \quad (2)$$

where

$$y = \frac{x}{l}.$$

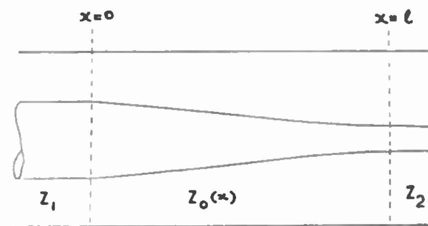


Fig. 1—Matching section between impedances Z_1 and Z_2.

The equation for the magnitude of the reflection-coefficient can be rewritten as

$$|\rho| = \int_0^1 \frac{1}{2}\frac{d}{dy}(\ln Z_0)e^{-j4\pi l/\lambda y}dy. \quad (3)$$

Observe that in (1) and (2), expressions for $|\rho|$ can be readily integrated for both cases. The solutions are,

$$|\rho| = \frac{1}{2}\ln(Z_2/Z_1)\left[\frac{1}{(4l/\lambda)} - \frac{3.556\,l/\lambda}{(4l/\lambda)^2 - 4} + \frac{0.0448\,l/\lambda}{(4l/\lambda)^2 - 16}\right]\sin 2\pi l/\lambda \quad (4)$$

$$|\rho| = \left[\frac{1}{(4l/\lambda)^2 - 1} - \frac{1}{(4l/\lambda)^2 - 9} + \frac{1}{10.8}\frac{1}{(4l/\lambda)^2 - 25}\right]\cdot\frac{\frac{1}{2}\ln(Z_2/Z_1)}{1 - \frac{1}{9} + \frac{1}{270}}\cos 2\pi l/\lambda. \quad (5)$$

The values of $|\rho|$, normalized with respect to $\frac{1}{2}\ln(Z_2/Z_1)$, are plotted against l/λ in Fig. 2. Curve (1) is for the line proposed in (1) and curve (2) for the line proposed in (2). For comparison, curve (3) has been drawn for the parabolic line proposed by Yang.

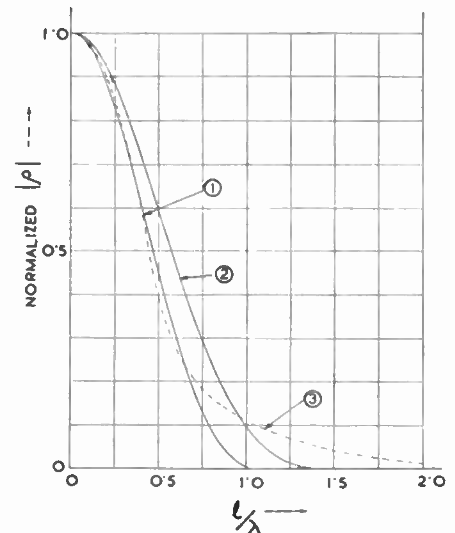


Fig. 2—Amplitude variation of the reflection-coefficient with l/λ.

It will be seen that the curves (1) and (2) show very little frequency-sensitivity and also give low reflection. For curve (1), the normalized value of $|\rho|$ is less than 0.006 for l/λ greater than one, and for curve (2) the normalized value of $|\rho|$ is less than 0.0005 for l/λ greater than 1.5.

TABLE I

l/λ	0.25	0.5	0.75	1.0
$\|\rho\|$ calculated from (5)	0.305	0.205	0.102	0.034
$\|\rho\|$ obtained from the Computer	0.297	0.205	0.104	0.035
l/λ	1.25	1.50	1.75	2.0
$\|\rho\|$ calculated from (5)	0.0057	0.00002	0.00000	0.00002
$\|\rho\|$ obtained from the Computer	0.0060	0.00015	0.00006	0.00001

It may be added that (3) is only approximate and has been derived on the assumption that $(1-\rho^2)\cong 1$. To verify the accuracy of this assumption, the nonlinear differential equation for ρ was solved on the Manchester University Digital Computer for $Z_2/Z_1 = 2$. The values thus obtained for the line assumed in (2) agreed closely with those obtained from (5), and are given in the above table:

Optimum nonuniform line-matching sections of any given length can be obtained by finding a suitable Fourier series for $(d/dy)(\ln Z_0)$. This will be discussed in a forthcoming paper.

J. WILLIS and N. K. SINHA
Faculty of Tech., University of Manchester
Manchester 1, Eng.

* Received by the IRE, September 15, 1955.
[1] H. J. Scott, "The hyperbolic transmission line as a matching section," PROC. IRE, vol. 41, pp. 1654–1657; November, 1953.
[2] R. F. H. Yang, "Parabolic transmission line," PROC. IRE, vol. 43, p. 1010; August, 1955.

Contributors

R. F. Blunt was born in Greeley, Kans., on August 24, 1920. He entered Rice Institute, Houston, Tex. in 1938. His education was interrupted by service in the Navy where he attended the Radio Materiel School.

R. F. BLUNT

At the conclusion of World War II, he returned to Rice Institute, from which he obtained his Ph.D. degree in 1949. The subject of his doctorate thesis pertained to low temperature physics.

After receiving his degree, Dr. Blunt joined the staff of the National Bureau of Standards Solid State Physics Section where he has recently been engaged in the investigation of intermetallic compounds.

Dr. Blunt is a member of the American Physical Society.

W. H. Brattain was born in Amoy, China, on February 10, 1902. He received a B.S. degree from Whitman College in 1924, an M.A. degree from the University of Oregon in 1926, and the Ph.D. degree from the University of Minnesota in 1928.

W. H. BRATTAIN

Dr. Brattain, with Dr. John Bardeen, is the co-inventor of the point contact transistor. Since joining Bell Telephone Laboratories in 1929, Dr. Brattain's research work has been concerned primarily with the study of semiconductors. He has also engaged in the study of frequency standards, thermionics, magnetometers and infra-red phenomena.

During World War II Dr. Brattain was associated with the Division of War Research, Columbia University and in 1952 he was a visiting lecturer at Harvard University.

Earlier this year Dr. Brattain received honorary Doctor of Science degrees from Whitman College, and from Union College. Portland University awarded him the same degree in 1952. In 1954 he received the John Scott Medal in recognition of his work on the transistor, and in 1952 he was awarded the Stuart Ballantine Medal of the Franklin Institute.

Dr. Brattain is a fellow of the American Physical Society and of the American Association for the Advancement of Science, and a member of the Franklin Institute, Phi Beta Kappa and Sigma Xi.

R. H. Bube was born on August 10, 1927, in Providence, R. I. He received the B.S. degree in physics from Brown University in 1946, and the M.A. and Ph.D. degrees in physics from Princeton University in 1948 and 1950, respectively.

R. H. BUBE

Since 1948, Dr. Bube has been with the RCA Laboratories Division where he has worked as a solid-state research physicist in the Physical and Chemical Research Laboratory.

He has been engaged primarily in research on luminescent materials and photoconductors.

Dr. Bube is a member of the American Physical Society and Sigma Xi.

G. Destriau was born August 1, 1903, in Bordeaux, France. He attended the Sorbonne and the École Centrale des Arts and Manufactures, both in Paris, and received the engineering degree from the latter institution. He was awarded the doctorate in 1936 by the Sorbonne; his work for his doctorate thesis was done in the Institut du Radium.

G. DESTRIAU

In 1939 he became Chargé de Cours at the Faculté des Sciences of the University of Bordeaux, and in 1942 Director of Studies of the École Centrale des Arts and Manufactures, Paris.

From 1944 to 1953, he was Chargé de Cours at the Faculté des Sciences of the University of Paris (Sorbonne) and professor of the Faculté des Sciences of the University of Poitiers. In 1953 he was made a professor at the Sorbonne.

Since 1952 he has also been a consultant to the Lamp Division of the Westinghouse Electric Corp., Bloomfield, N. J.

Although Professor Destriau is known primarily for his pioneering work in electroluminescence and related effects of electric fields on phosphors, he has also made several notable contributions in the fields of alpha particle scintillations, x-ray therapy, radiography, optics, and the properties of thin metal films.

Dr. Destriau is a member of the Société Française de Physique, the Société Française de Chemie-Physique, and of the Société Française des Electriciens.

H. P. R. Frederikse was born in The Hague, Netherlands, on July 13, 1920. He studied at the University of Leiden, Netherlands, where he received his Ph.D. degree in 1950.

H.P.R. FREDERIKSE

From 1950 to 1953 he was a member of the semiconductor research group at Purdue University, in Indiana, working on thermoelectricity in germanium. In 1953 Dr. Frederikse joined the Solid State Physics Section of the National Bureau of Standards where he is presently engaged in the investigation of intermetallic compounds.

Dr. Frederikse is a member of the American Physical Society and Sigma Xi.

G. F. J. Garlick was born on February 21, 1919 in Staffordshire, England. He received the B.Sc. degree with honors from the University of Birmingham in England in 1940. In 1943, he received the Ph.D. degree and in 1955 the D.Sc. degree from the same university.

G. F. J. GARLICK

He has directed luminescence research in the department of physics at Birmingham since 1944, and is now a lecturer there. His publications include several papers and a book on phosphors.

Dr. Garlick is a fellow of the British Institute of Physics.

E. W. Gorter received his science degree in 1939 from Leyden University, where his major subjects were physical and inorganic chemistry. After a short period of teaching at a secondary school at the Hague, he received a Ramsay Memorial Fellowship to work at the University of Manchester in England.

E. W. GORTER

After the invasion of Holland in 1940, he joined the Netherlands forces in England, and for the greater part of the war worked on chemical warfare problems with a Dutch research group in London.

When he returned to civilian life in 1946, he joined the staff of Philips Research Laboratories. He is presently working in the chemistry section under Dr. E. J. W. Verwey on ferromagnetic materials.

He received the doctor's degree cum laude from Leyden University in 1954 for his thesis "Saturation Magnetization and Crystal Chemistry of Ferrimagnetic Oxides." Dr. Gorter is co-inventor of the oxidic permanent-magnet material Magnadur (Ferroxdure).

F. Herman (S'45-A'52) was born in New York, N.Y., on March 21, 1926. He received the B.S. in E.E. degree in 1945, the M.S. in E.E. degree in 1949, and the Ph.D. degree in physics in 1953, all from Columbia University.

F. Herman

From 1945 to 1947, Dr. Herman served with the U. S. Navy as an electronic technician. From 1947 to 1949, he was an instructor in electrical engineering at Cooper Union in New York City. In 1949, he joined the RCA Laboratories, Princeton, N. J., where he has been ever since. While on the Technical Staff at the RCA Laboratories, Dr. Herman has devoted himself chiefly to theoretical problems in solid state physics. In the summer of 1954 he delivered an invited paper on his work on the electronic energy band structure of germanium at the International Conference on Semiconductors, Amsterdam.

Dr. Herman is a member of Tau Beta Pi, Sigma Xi, and the American Physical Society.

H. F. Ivey (M'46-SM'51) was born in Augusta, Ga., on June 16, 1921. He received the degrees of A.B. summa cum laude and M.S. from the University of Georgia in 1940 and 1941. In 1944, he was awarded the degree of Ph.D. in physics by M.I.T.

H. F. Ivey

From 1942 to 1945, he was a staff member of the M.I.T. Radiation Laboratory where he was concerned with problems of tenebrescence as applied to dark-trace cathode-ray tube screens for radar applications. From 1945 to 1946 he was a senior engineer in the research department of the National Union Radio Corp. He entered the employment of the Lamp Division of the Westinghouse Electric Corp. in Bloomfield, N. J., as a research engineer in 1946. Here he worked on problems of thermionic electron emission and electron space charge. In 1952 his activities were shifted to the field of luminescence. Since 1954, he has been in charge of the Phosphor Research Group of the Westinghouse Lamp Division, which is concerned with phosphors for fluorescent and other lamps, phosphors for television, and with electroluminescence.

He is a member of the American Physical Society (Divisions of Electron Physics and Solid-State Physics), the Electrochemical Society (Electronics Division), Phi Beta Kappa, Sigma Xi, Sigma Pi Sigma, and Pi Mu Epsilon.

E. T. Jaynes (SM'54) was born in Waterloo, Iowa, on July 5, 1922. He attended Cornell College and Iowa State University, receiving the B.A. degree in physics from the latter in 1942. He studied in the graduate school of the University of California in Berkeley and at Princeton University from which he received the M.A. degree in 1948 and the Ph.D. degree in theoretical physics in 1950.

E. T. Jaynes

From 1942 to 1946 he was engaged in microwave research and development as a project engineer at the Sperry Gyroscope Co., in Garden City, N. Y., and in the combined research group of the Naval Research Laboratory.

Since 1950, he has been on the faculty of Stanford University, and at present holds the titles of associate professor in the microwave laboratory and lecturer in physics.

F. Jona was born in Pistoia, Italy, on October 10, 1922. He received his undergraduate degree in 1945 and the Ph.D. degree in physics in 1949, both from the Eidgenössische Technische Hochschule in Zurich, Switzerland.

F. Jona

From 1945 to 1946, he was an instructor in the physics department of the University of Berne, Switzerland. He returned to the Eidgenössische Technische Hochschule as a research associate from 1946 to 1952.

He came to Pennsylvania State University as a research associate in 1952. Since 1954, he has been an assistant professor.

Solid state physics, especially elasticity and ferroelectricity, and crystal analysis, is Dr. Jona's field of research. He is a member of the American Physical Society, Schweizerische Physikalische Gesellschaft and Società Italiana di Fisica.

B. Kazan was born in New York, N. Y. on May 8, 1917. He received the B.S. degree in physics from the California Institute of Technology in 1938 and the M.A. degree in physics from Columbia University in 1940.

B. Kazan

In 1940 he joined the Signal Corps Engineering Laboratories and was engaged in early experimental work with radar equipment. From 1944 to 1950 he was chief of the Special Purpose Tube Section at the Evans Signal Laboratory, where he was responsible for development and application engineering of traveling wave tubes, klystrons, display tubes, and transistor devices. Since 1951, Mr. Kazan has been engaged in research on television tubes and display devices at the RCA Laboratories, Princeton, N. J.

Mr. Kazan is a member of the American Physical Society, Sigma Xi and Tau Beta Pi.

F. A. Kröger was born in Amsterdam, The Netherlands, on September 11, 1915. He studied physical chemistry at the University of Amsterdam, and received the Ph.D. degree in 1940. His doctorate thesis was on luminescence of solids containing manganese.

F. A. Kröger

In 1938, Dr. Kröger joined the staff of the Research Laboratories of the N. V. Philips Co., Eindhoven. He has since led research groups on luminescence and semiconductors.

Dr. Kröger is a member of the Dutch Chemical Society.

E. E. Loebner (M'52) was born in 1924 at Pilsen, Czechoslovakia. He graduated with honors in 1945 from the Higher Industrial Institute in Pilsen, receiving a degree in mechanical engineering. In 1947, he came to the United States as a B'nai B'rith Hillel scholar to continue his studies. He received the B.A. and Ph.D. degrees in physics from the University of Buffalo in 1950 and 1955 respectively.

Dr. Loebner was active in mechanical design and hydraulics both abroad and in the United States. Later, as an ONR research assistant, he did work on the thermoelectric power of carbons. In July, 1952, he joined Sylvania Electric Products Inc., and was engaged in research on applications of electroluminescence at the Radio and Television Division in Buffalo, N. Y. In 1954, Dr. Loebner transferred to the Waltham Laboratories, Waltham, Mass., later a part of Sylvania's Electronic Systems Division, where he continued his work on electroluminescent devices. Since October, 1955, he has been on the technical staff of the RCA Laboratories Division in Princeton, N. J.

E. E. Loebner

Dr. Loebner is a member of the American Physical Society, the American Association for the Advancement of Science, and Sigma Xi, and a fellow of the Physical Society.

J. L. Moll (A'51) was born on December 21, 1921, in Wauseon, Ohio. He received the B.Sc. degree in engineering physics in December, 1943 from the Ohio State University, and the Ph.D. degree in electrical engineering in 1952.

J. L. Moll

Dr. Moll worked at the Radio Corporation of America during the years 1944 and 1945. From 1945 to 1948 he was employed as a graduate assistant in the mathematics department at Ohio State University, and from 1948 to 1952 was an instructor in the electrical engineering department, as well as a research associate for the Ohio State University Research Foundation. Since September, 1952, Dr. Moll has been a member of the Bell Telephone Laboratories staff, where he has been working on transistor development.

Dr. Moll is a member of Sigma Xi.

G. A. Morton (A'39–SM'46–F'51) was born in New Hartford, N. Y., on March 24, 1903. He received the B.S. in E.E. degree in 1926, the M.S. degree in 1928, and the Ph.D. degree in physics in 1932, all from M.I.T.

G. A. Morton

From 1927 to 1933, he was a research associate and instructor at M.I.T. He became a research engineer for RCA in 1933, and since 1954 has been associate director of the Physical-Chemical Laboratory at RCA Laboratories. From 1946 to 1947, he was on leave from RCA doing special research work at Oak Ridge, Tenn.

Dr. Morton was a section member of National Defense Research, and a member of the Army Air Force Advisory Board during World War II. He received the IEE Overseas Premium Award, and certificates for war research from the Navy, Army, and Air Force.

He is a fellow of the American Physical Society, and a member of Sigma Xi, and the American Institute of Electrical Engineers. His research work has resulted in several inventions in television electronics, electron optics, infrared imaging, computers, and nucleonics.

T. S. Moss was born on January 28, 1921. He graduated first in his class from Cambridge University, receiving the B.A. degree in 1941. In 1944, he received the M.A. degree from Cambridge, and in 1951, the Ph.D. degree.

T. S. Moss

He is the author of "Photo-Conductivity in the Elements," published in 1952. At present, he is a Principal Scientific Officer at the Royal Aircraft Establishment in Farnborough, England.

Dr. Moss became an associate member of the Institute of Electrical Engineers in 1949.

F. H. Nicoll (A'39–SM'43) was born in Saskatchewan, Canada, on June 6, 1908. He received the B.Sc. degree in physics in 1929 and the M.Sc. degree in 1931 from Saskatchewan University. He held an 1851 Exhibition Scholarship to Cambridge University, England, for three years' research and received the Ph.D. degree from that university in 1934.

F. H. Nicoll

From 1934 to 1939 Dr. Nicoll was associated, as research physicist, with Electric and Musical Industries, Ltd. in London, England. He was employed by the RCA Victor Division of the Radio Corporation of America at Camden, N. J. as research engineer from 1939 to 1941. Since 1941 he has been engaged in research on cathode-ray tubes and electron optics at the RCA Laboratories Division in Princeton, N. J.

Dr. Nicoll is a member of the American Physical Society and of Sigma Xi.

G. L. Pearson was born in Salem, Ore., on March 31, 1905. He was graduated from Willamette University in 1926 with an A.B. degree and was awarded an M.A. degree in physics by Stanford University in 1929. He then became a member of the physical research department of Bell Telephone Laboratories, spending his early years there on studies of noise in resistors, vacuum tubes and carbon microphones.

G. L. Pearson

Except for a brief period during World War II, when he was engaged in military projects, Mr. Pearson has concentrated on semiconductor research since 1935. In this field he has been concerned with thermistors, transistors, silicon rectifiers, and photovoltaic devices including the solar battery.

Mr. Pearson holds a large number of patents on semiconductor devices, including thermistors and transistors. He is the author of numerous articles and papers in the same field. In 1948, the A.I.E.E. awarded him its annual prize for the best paper in research He is a fellow of the American Physical Society and a member of Sigma Xi.

R. Pepinsky was born in St. Paul, Minn., on January 17, 1912. He received the B.A. degree in physics in 1933, and the M.A. degree in 1934, both from the University of Minnesota.

R. Pepinsky

He was a university fellow at the University of Chicago in 1935, and a teaching assistant from 1936 to 1939. He received the Ph.D. degree in 1939.

For one year, he was a research physicist with the United States Rubber Co. in Providence, R. I. In 1941 he became an instructor, and in 1942, an assistant professor, at the Alabama Polytechnic Institute. He was on academic leave from 1942 to 1945 and was a member of the Radiation Laboratory at M.I.T. He returned to Alabama as a research professor from 1945 to 1949. Since 1949, he has been a research professor of physics at Pennsylvania State University. He has also been consultant in physics and part-time resident physicist at Brookhaven National Laboratory since 1952. During part of 1955 he was visiting professor at the National Autonomous University of Mexico, as a U. S. Department of State grantee; in 1954 he was a visiting lecturer on biochemistry in several European universities, under a Rockefeller Foundation grant.

Dr. Pepinsky's major fields of research include X-ray and neutron crystallography,

solid state physics and chemistry—particularly structural bases for physical properties of solids, radiation damage, biochemical structure analyses, structural studies of explosives, crystal engineering, complex ion chemistry, theory of the phase problem in diffraction analysis, and optical rotatory power. Dr. Pepinsky has also been concerned with X-ray tube and crystallographic instrumentation design, and electronic computer design. He is the inventor of X-RAC, R-PAC and S-FAC.

He is a fellow of the American Physical Society, Chemical Society of London, and American Mineralogical Society.

In addition, Dr. Pepinsky holds membership in the American Chemical Society, American Crystallographic Association, American Nuclear Society, American Association for the Advancement of Science, Association for Computing Machinery, and Physical Society of Japan. He is a member of the Solid State Advisory Panel of the Office of Naval Research.

A. Rose (A'36–M'40–SM'43–F'48) received the A.B. degree in 1931, and the Ph.D. degree in 1935, both from Cornell University in Ithaca, N. Y.

A. Rose

For twenty years, from the completion of his graduate studies until earlier this year, he was employed at RCA Laboratories where he was engaged in research on television camera tubes, electron optics, emission phenomena, quantum noise, and solid state.

Dr. Rose is currently at Laboratories RCA Ltd. in Zurich, Switzerland.

He is a fellow of the American Physical Society, and a member of Societe Suisse de Physique.

J. E. Rosenthal received her Sc.B. degree from the Sorbonne in Paris and her Sc.M. and Ph.D. degrees from New York University.

J. E. Rosenthal

She was National Research Fellow at Johns Hopkins University and received a fellowship at Columbia University from the American Association of University Women. Later, Dr. Rosenthal was an instructor in the graduate division of Brooklyn College, and joined the Signal Corps Engineering Laboratories as a physicist. In 1953, she became a project engineer for Allen B. Du Mont Laboratories, Inc.

Spectroscopy, particularly the spectra of polyatomic molecules, was Dr. Rosenthal's field of specialization before the war. Since 1942, her work has been in electronics.

She is a member of Sigma Xi and the American Mathematical Society, a fellow of the American Physical Society, and is a holder of several patents. Her married name is Bramley.

M. L. Schultz was born in Chicago, Ill., on May 24, 1912. He received the B.S. degree in chemistry from the University of Chicago in 1934, and the Ph.D. degree in physical chemistry from the University of Chicago in 1939.

From 1939 to 1943, he was instructor in chemistry at the Armour Institute of Technology and Illinois Institute of Technology, both in Chicago. He joined the SAM Laboratories of the Division of War Research at Columbia University in 1943. Since 1945, he has been associated with RCA Laboratories in Princeton, N. J., working on research on semiconductors and infrared photoconductors.

M. L. Schultz

Dr. Schultz is a member of the American Chemical Society, the American Association for the Advancement of Science, Phi Beta Kappa, and Sigma Xi.

G. Shirane was born in Nishinomiya City, Japan, on May 15, 1924. He received the B.E. degree in applied physics in 1947, and the D.Sc. in physics in 1954, from University of Tokyo.

G. Shirane

From 1948 to 1952, he was a research associate in the physics department of Tokyo Institute of Technology. He joined the physics department of Pennsylvania State University in 1952 as a research associate. In 1955, he became an assistant professor.

Dr. Shirane has engaged primarily in research on solid state physics, studies on ferroelectricity, and crystal analysis. He is a member of the American Physical Society and the Physical Society of Japan.

IRE News and Radio Notes

Calendar of Coming Events

Engineers Joint Council Nuclear Engineering and Science Congress, Cleveland, Ohio, Dec. 12–16.

PGEM-SIAM Operations Research Symposium, U. of Pennsylvania, Philadelphia, Pa., Dec. 14

URSI Fall Meeting, U. of Florida, Gainesville, Fla., Dec. 15–17

IRE-ASQC Second National Symposium on Reliability and Quality Control, Hotel Statler, Washington, D. C., Jan. 9–10

IRE National Simulation Symposium, Dallas, Tex., Jan. 19–21

IRE Symposium on Microwave Theory and Techniques, U. of Pennsylvania, Philadelphia, Pa., Feb. 2–3

IRE-AIEE-ACM Western Joint Computer Conferences, Fairmont Hotel, San Francisco, Calif., Feb. 8–10

IRE Annual Southwestern Conference, Oklahoma City, Okla., Feb. 9–11

Annual Banquet of Washington Section, Hotel Statler, Washington, D. C., Feb. 11

IRE-AIEE-U. of P. Conference on Transistor Circuits, U. of Pennsylvania, Philadelphia, Pa., Feb. 16–17

IRE-AIEE Scintillation Counter Symposium, Shoreham Hotel, Washington, D. C., Feb. 28–29

IRE National Convention and Radio Engineering Show, Waldorf-Astoria Hotel and Kingsbridge Armory, New York City, Mar. 19–22

PGIE-AIEE-ISA Conference on Magnetic Amplifiers, Syracuse, N. Y., Apr. 5–6

Seventh Regional Technical Conference and Trade Show, Hotel Utah, Salt Lake City, Utah, Apr. 11–13

Tenth Annual Spring Television Conference of the Cincinnati Section, Engineering Society of Cincinnati Building, Cincinnati, Ohio, Apr. 13–14

New England Radio Engineering Meeting, Sheraton Plaza Hotel, Boston, Mass., Apr. 23–24

PGCT-PIB Symposium on Nonlinear Network Theory, Engineering Society Building, New York City, Apr. 25–27

URSI Spring Meeting, National Bureau of Standards, Washington, D. C., Apr. 30–May 3

Operations Research to be Subject of December Symposium

The Philadelphia Section of the IRE, the Professional Group on Engineering Management, and the Delaware Valley Section of the Society for Industrial and Applied Mathematics will sponsor a one-day Operations Research Symposium at the University of Pennsylvania, Philadelphia, on December 14. P. M. Morse of the Massachusetts Institute of Technology will deliver the keynote address. Among topics to be discussed are: the scope of operations research in defense, commerce, and industry; statistical forecasting; scheduling and queuing problems; inventory and production control with emphasis on the use of computing machinery; and linear programming. Speakers scheduled to discuss these topics are Russell L. Ackoff, Case Institute of Technology; Wroe Alderson, Alderson and Sessions; Leroy A. Brothers, USAF; Leslie C. Edie, Port of New York Authority; James C. Hetrick, Continental Oil Company, and Alan O. Mann, S.K.F. Industries, Inc. Case histories will also be presented.

The registration fee at the symposium will be $3.50, but the advance registration fee will be only $2.50. Queries regarding advance registration and other matters may be addressed to R. V. D. Campbell, O. R. Symposium Registration, Burroughs Research Center, Paoli, Pennsylvania.

IRE Southwestern Conference Scheduled for February

The IRE Southwestern Regional Conference and Electronics Show for 1956 will be held in Oklahoma City, February 9–11, sponsored by the Oklahoma City Section.

The technical program for the conference will be directed by Clyde L. Farrar, Chairman of the School of Electrical Engineering, University of Oklahoma. Charles E. Harp, also of the Engineering School will lead the special committee appointed by the Section to handle all arrangements for the convention. Others on the committee are: George W. Holt, Southwestern Bell Telephone Company, *Vice-Chairman;* Nicholas Battenburg, Tinker Air Force Base, *Secretary;* R. R. Freeland, International Crystal Manufacturing Co., *Treasurer;* Ralph L. McClung, Civil Aeronautics Authority, *Registration;* A. J. Spooner, Two-Way Radio Communication Co., *Inspection Tours;* Clifford M. Easum, Radio Station KTOK, *Banquet;* Alfred B. Ashley, Civil Aeronautics Authority, *Housing;* Morris W. Thomas, KOMA-KWTV, *Exhibits;* Carl O. Hart, Tinker Air Force Base, *Publicity;* and Mrs. George W. Holt, *Ladies Activities.*

In developing the technical program, Professor Farrar will organize sessions along the lines of the Professional Groups. Since sessions will be simultaneous during a large part of the conference this plan will promote easier orientation and grouping of interests. Papers will be presented in such diversified fields as: Antennas and Propagation; Circuit Theory and Solid State Electronics; Electronic Devices and Component Parts; Computers, Information Theory, Automatic Controls; Aeronautical and Navigational Electronics; Management, Quality Control, and Production; Transmitters, Receivers and Audio; AM, FM, and Television; Communications and Microwaves; Ultrasonics, Medical and Industrial Electronics; Instrumentation, Telemetry and Nuclear Science.

A program of tours will round out the conference. One attraction will be an inspection, including mission briefing, of the mile-long aircraft maintenance line and supporting shops at Tinker Air Force Base. Tours of the CAA establishment and the Sylvania Plant are also contemplated. In addition a women's program, including entertainment and tours, will be offered.

Charles Harp (left), head of special committee, and Clyde L. Farrar, program chairman, discuss Southern Conference plans.

Vehicular Communications Group Holds Conference

Approximately 125 guests and members of the Professional Group on Vehicular Communications attended the Sixth National Meeting of the Group at the Multnomah Hotel, Portland, Oregon on September 26 and 27. Included on the program were four technical sessions at which fifteen papers were presented covering VHF and UHF propagation, new features of vehicular equipment design, and unique applications of mobile communication systems. The highlight of the meeting was the annual banquet at which Robert E. Lee of the Federal Communications Commission presented the principal address. Other events included a formal luncheon, an exhibit of vehicular equipment, and a tour of the Columbia River Gorge Highway to Bonneville Dam. The meeting was conducted jointly by the Vehicular Group and the Portland Section under the chairmanship of George C. Ellison of Portland.

G. C. Ellison

PIB Sponsors Nonlinear Circuit Analysis Symposium In April

The International Symposium on Nonlinear Circuit Analysis, II will be the sixth of a series sponsored by the Microwave Research Institute of the Polytechnic Institute of Brooklyn. It will be held on April 25, 26 and 27, 1956, at the Engineering Societies Building in New York City.

The program will consider basic methods and recent advances in the analysis and design of nonlinear networks and will emphasize the use of nonlinear network theory in the study of oscillators, switching and discontinuous systems, and nonlinear systems with random inputs. An exposition of the fundamental mathematical methods of analysis will be correlated with applications in such fields as automatic control where specific practical systems illustrate such nonlinear phenomena as subharmonic generation, parametric damping, jump resonance, and stabilized oscillation.

The cooperation of the PG on Circuit Theory, the Office of Naval Research, the Air Force Office of Scientific Research and the Signal Corps enable the symposium to eliminate admission charge or registration fees. Volume VI of the MRI Symposia Series, "Proceedings of the Symposium on Nonlinear Circuit Analysis, II" will be published by October, 1956 at five dollars per copy, clothbound. Orders for the "Proceedings," accompanied by check or money order made out to Treasurer, Symposium Committee, will be accepted in advance. Copies of the detailed program, hotel accommodation information and registration forms are available on request. All correspondence should be addressed to: Polytechnic Institute of Brooklyn, Microwave Research Institute, 55 Johnson Street, Brooklyn 1, N. Y.

TRANSACTIONS OF THE IRE PROFESSIONAL GROUPS

The following issues of TRANSACTIONS are available from The Institute of Radio Engineers, Inc., 1 East 79 Street, New York 21, N. Y., at the prices listed below:

Sponsoring Group	Publications	Group Members	IRE Members	Non-Members*
Aeronautical and Navigational Electronics	PGAE-5: A dynamic Aircraft Simulator for Study of Human Response Characteristics (6 pages)	$ .30	$ .45	$ .90
	PGAE-6: Ground-to-Air Cochannel Interference at 2900 MC (10 pages)	.30	.45	.90
	PGAE-8: June 1953 (23 pages)	.65	.95	1.95
	PGAE-9: September 1953 (27 pages)	.70	1.05	2.10
	Vol. ANE-1, No. 2, June 1954 (22 pages)	.95	1.40	2.85
	Vol. ANE-1, No. 3, September 1954 (27 pages)	1.00	1.50	3.00
	Vol. ANE-1, No. 4, December 1954 (27 pages)	1.00	1.50	3.00
	Vol. ANE-2, No. 1, March 1955 (41 pages)	1.40	2.10	4.20
	Vol. ANE-2, No. 2, June 1955 (49 pages)	1.55	2.30	4.65
	Vol. ANE-2, No. 3, September 1955 (27 pages)	.95	1.45	2.85
Antennas and Propagation	PGAP-4: IRE Western Convention, August 1952 (136 pages)	2.20	3.30	6.60
	Vol. AP-1, No. 1, July 1953 (30 pages)	1.20	1.80	3.60
	Vol. AP-1, No. 2, October 1953 (31 pages)	1.20	1.80	3.60
	Vol. AP-2, No. 1, January 1954 (39 pages)	1.35	2.00	4.05
	Vol. AP-2, No. 2, April 1954 (41 pages)	2.00	3.00	6.00
	Vol. AP-2, No. 3, July 1954 (36 pages)	1.50	2.25	4.50
	Vol. AP-3, No. 4, October 1954 (36 pages)	1.50	2.25	4.50
	Vol. AP-3, No. 1, January 1955 (43 pages)	1.60	2.40	4.80
	Vol. AP-3, No. 2, April 1955 (47 pages)	1.60	2.40	4.80
	Vol. AP-3, No. 3, July 1955 (66 pages)	2.05	3.10	6.15
Audio	PGA-7: Editorials, Technical Papers and News, May 1952 (47 pages)	.90	1.35	2.70
	PGA-10: November–December 1952 (27 pages)	.70	1.05	2.10
	Vol. AI-1, No. 1, January–February 1953 (24 pages)	.60	.90	1.80
	Vol. AU-1, No. 2, March–April 1953 (34 pages)	.80	1.20	2.40
	Vol. AU-1, No. 3, May–June 1953 (34 pages)	.80	1.20	2.40
	Vol. AU-1, No. 4, July–August (19 pages)	.70	1.05	2.10
	Vol. AU-1, No. 5, September–October 1953 (11 pages)	.50	.75	1.50
	Vol. AU-1, No. 6, November–December 1953 (27 pages)	.90	1.35	2.70
	Vol. AU-2, No. 1, January–February 1954 (38 pages)	1.20	1.80	3.60
	Vol. AU-2, No. 2, March–April 1954 (31 pages)	.95	1.40	2.85
	Vol. AU-2, No. 3, May–June 1954 (27 pages)	.95	1.40	2.85
	Vol. AU-2, No. 4, July–August 1954 (27 pages)	.95	1.40	2.85
	Vol. AU-2, No. 2, September–October 1954 (22 pages)	.95	1.40	2.85
	Vol. AU-2, No. 6, November–December 1954 (24 pages)	.80	1.20	2.40
	Vol. AU-3, No. 1, January–February 1955 (20 pages)	.60	.90	1.80
	Vol. AU-3, No. 2, March–April 1955 (32 pages)	.95	1.40	2.85
	Vol. AU-3, No. 3, May–June 1955 (30 pages)	.85	1.25	2.55
	Vol. AU-3, No. 4, July–August 1955 (46 pages)	1.15	1.75	3.45
Broadcast Transmission Systems	PGBTS-1: March 1955 (102 pages)	2.50	3.75	7.50
Broadcast and Television Receivers	PGBTR-1: Round-Table Discussion on UHF TV Receiver Considerations, 1952 IRE National Convention (12 pages)	.50	.75	1.50
	PGBTR-3: June 1953 (67 pages)	1.40	2.10	4.20
	PGBTR-5: January 1954 (96 pages)	1.80	2.70	5.40
	PGBTR-6: April 1954 (119 pages)	2.35	3.50	7.00
	PGBTR-7: July 1954 (58 pages)	1.15	1.70	3.45
	PGBTR-8: October 1954 (20 pages)	.90	1.35	2.70
	Vol. BTR-1, No. 1, January 1955—Papers presented at the Radio Fall Meeting, 1954 (68 pages)	1.25	1.85	3.75
	Vol. BTR-1, No. 2, April 1955 (40 pages)	.95	1.45	2.85
	Vol. BTR-1, No. 3, July 1955 (51 pages)	.95	1.45	2.85
Circuit Theory	PGCT-2: Papers presented at the Circuit Theory Sessions of the Western Electronic Show and Convention, San Francisco, Calif., August 19–21, 1953 (106 pages)	1.95	2.90	5.85
	Vol. CT-1, No. 1, March 1954 (80 pages)	1.30	1.95	3.90
	Vol. CT-1, No. 2, June 1954 (39 pages)	1.00	1.50	3.00
	Vol. CT-1, No. 3, September 1954 (73 pages)	1.00	1.50	3.00
	Vol. CT-1, No. 4, December 1954 (42 pages)	1.00	1.50	3.00
	Vol. CT-2, No. 1, March 1955 (106 pages)	2.70	4.05	8.10
	Vol. CT-2, No. 2, June 1955 (113 pages)	2.60	3.90	7.80

*** Public Libraries, Colleges and Subscription Agencies may purchase at IRE Member rate.**

(Continued on page 1982)

WRH

TRANSACTIONS OF THE IRE PROFESSIONAL GROUPS (cont'd)

Sponsoring Group	Publications	Group Members	IRE Members	Non-Members*
Communications Systems	Vol. CS-2, No. 1, January 1954 (83 pages)	$1.65	$2.50	$4.95
	Vol. CS-2, No. 2, July 1954 (132 pages)	2.25	3.35	6.75
	Vol. CS-2, No. 3, November 1954: IRE Symposium on Global Communications, June 23–25, 1954, Washington, D. C. and IRE-AIEE Symposium on Military Communications, April 28, 1954, New York, N. Y. (181 pages)	3.00	4.50	9.00
	Vol. CS-3, No. 1, March 1955: Papers presented at the Symposium on Marine Communications and Navigation, October 13–15, 1954, Boston, Mass. (72 pages)	1.00	1.50	3.00
Component Parts	PGCP-1: March 1954 (46 pages)	1.20	1.80	3.60
	PGCP-2: September 1954: Papers presented at the Component Parts Sessions at the 1954 Western Electronic Show and Convention, Los Angeles, Calif. (118 pages)	2.25	3.35	6.75
	PGCP-3: April 1955 (44 pages)	1.00	1.50	3.00
Electronic Computers	Vol. EC-2, No. 2, June 1953 (27 pages)	.90	1.35	2.70
	Vol. EC-2, No. 4, December 1953 (47 pages)	1.25	1.85	3.75
	Vol. EC-3, No. 1, March 1954 (39 pages)	1.10	1.65	3.30
	Vol. EC-3, No. 3, September 1954 (54 pages)	1.80	2.70	5.40
	Vol. EC-3, No. 4, December 1954 (46 pages)	1.10	1.65	3.30
	Vol. EC-4, No. 1, March 1955 (48 pages)	1.10	1.65	3.30
	Vol. EC-4, No. 2, June 1955 (36 pages)	.90	1.35	2.70
Electron Devices	PGED-4: December 1953 (62 pages)	1.30	1.95	3.90
	Vol. ED-1, No. 1, February 1954 (72 pages)	1.40	2.10	4.20
	Vol. ED-1, No. 2, April 1954 (75 pages)	1.40	2.10	4.20
	Vol. ED-1, No. 3, August 1954 (77 pages)	1.40	2.10	4.20
	Vol. ED-1, No. 4, December 1954: Papers presented at the 1954 Symposium on Fluctuation Phenomena in Microwave Sources, November 18–19, 1954, New York, N. Y. (280 pages)	3.20	4.80	9.60
	Vol. ED-2, No. 2, April 1955 (53 pages)	2.10	3.15	6.30
	Vol. ED-2, No. 3, July 1955 (27 days)	1.10	1.65	3.30
Engineering Management	PGEM-1: February 1954 (55 pages)	1.15	1.70	3.45
	PGEM-2: November 1954 (67 pages)	1.30	1.95	3.90
	PGEM-3: March 1955 (52 pages)	1.00	1.50	3.00
Industrial Electronics	PGIE-1: August 1953 (40 pages)	1.00	1.50	3.00
	PGIE-2: March 1955 (81 pages)	1.90	2.85	5.70
Information Theory	PGIT-3, March 1954 (159 pages)	2.60	3.90	7.80
	PGIT-4: September 1954 (234 pages)	3.35	5.00	10.00
	Vol. IT-1, No. 1, March 1955 (76 pages)	2.40	3.60	7.20
Instrumentation	PGI-2: Data Handling Systems Symposium: IRE Western Electronic Show and Convention, Long Beach, Calif., August 27–29, 1952 (111 pages)	1.65	2.45	4.95
	PGI-3: April 1954 (55 pages)	1.05	1.55	3.15
Medical Electronics	PGME-2: October 1955—Panel Discussion on Medical Electronics (39 pages)	.85	1.25	2.55
Microwave Theory and Techniques	Vol. MTT-1, No. 2, November 1953 (44 pages)	.90	1.35	2.70
	Vol. MTT-2, No. 2, July 1954 (67 pages)	1.25	1.85	3.75
	Vol. MTT-2, No. 3, September 1954: Papers presented at the Joint IRE Professional Group—URSI meeting, Washington, D. C., May 5, 1954 (54 pages)	1.10	1.65	3.30
	Vol. MTT-3, No. 1, January 1955 (47 pages)	1.50	2.25	4.50
	Vol. MTT-3, No. 3, April 1955 (44 pages)	1.40	2.10	4.20
	Vol. MTT-3, No. 4, July 1955 (54 pages)	1.60	2.40	4.80
Nuclear Science	Vol. NS-1, No. 1, September 1954 (42 pages)	.70	1.00	2.00
	Vol. NS-2, No. 1, June 1955 (15 pages)	.55	.85	1.65
Reliability and Quality Control	PGQC-2: March 1953 (51 pages)	1.30	1.95	3.90
	PGQC-3: February 1954 (39 pages)	1.15	1.70	3.45
	PGQC-4: December 1954 (56 pages)	1.20	1.80	3.60
	PGRQC-5: April 1955 (56 pages)	1.15	1.75	3.45
Telemetry and Remote Control	PGRTRC-1: August 1954 (16 pages)	.85	1.25	2.55
	PGRTRC-2: November 1954 (24 pages)	.95	1.40	2.85
	Vol. TRC-1, No. 1, February 1955 (24 pages)	.95	1.40	2.85
	Vol. TRC-1, No. 2, May 1955 (24 pages)	.95	1.40	2.85

* Public Libraries, Colleges and Subscription Agencies may purchase at IRE Member rate.

(Continued on page 1983)

M. J. Kelly and Sir Gordon Radley Receive First Columbus Communication Prize

Mervin J. Kelly, President of Bell Telephone Laboratories and an IRE Fellow, and Sir Gordon Radley, Director General of the British Post Office, were selected by the city of Genoa, Italy, to receive the first Christopher Columbus International Communication Prize. The award was conferred in Genoa on Columbus Day, October 12.

Dr. Kelly and Sir Gordon received the prize in recognition of "the planning, now being placed into practice, of the submarine telephone cable which will make it possible to establish 36 telephone circuits across the Atlantic between Scotland and Canada with extension to New York." The award also honored "the numerous scientists, research workers and engineers who have contributed in the planning, production and placing in operation of the intercontinental submarine telephone line."

M. J. Kelly

The Christopher Columbus Prize was instituted recently under the auspices of the City of Genoa as a memorial to the discoverer, who was a native of Genoa. Its aim will be to honor any outstanding development of the previous four years which aids communications among men. The award will be conferred yearly in one of four communications categories: maritime, air or land communications, and telecommunications. Awards, based on recommendations of the Italian Higher Institute of Telecommunications and the Superior Research Council, will be given to an individual, organization or group of persons. Presentation will take place each October in Genoa, at the International Meeting of Communications.

CORRECTION

In November, Proceedings of the IRE published a paper titled "The Cascade Backward-Wave Amplifier: A High-Gain Voltage-Tuned Filter for Microwaves," coauthored by M. R. Currie and J. R. Whinnery.

Dr. Currie's biography, which should have appeared in the Contributor's Section of the November issue, inadvertently was published in the October Proceedings, a month previous to the time the paper reached print.