PROGRESS IN SEMICONDUCTORS

VOLUME 2

General Editor
ALAN F. GIBSON, B.Sc., Ph.D.

American Editor
Prof. R. E. BURGESS
Vancouver, B.C.

European Editor
Prof. P. AIGRAIN
Paris

LONDON
HEYWOOD & COMPANY LTD.
1957

PREFACE

In this, our second volume of *Progress in Semiconductors* we have extended our coverage of the broad field of semiconductor research with eight further papers by acknowledged authorities in their own branches of the subject. Our general aims remain unchanged. The difficulties that arise from intensive specialization are well known and it is the purpose of this series of review articles to help in alleviating this problem.

The best treatment of a particular aspect of semiconductors clearly varies with the subject matter. Frequently, however, a broad physical picture is more valuable than a detailed, perhaps highly mathematical description. In our present issue Dr. Rose contributes a paper describing a general model from which the carrier lifetime in a very wide range of materials can be deduced. Thus, in this treatment, for example, the familiar equation originally deduced by Shockley and Read appears as a special case appropriate to a material like germanium.

The growing importance of semiconductors other than silicon and germanium is reflected in the papers by Dr. Cunnell and Dr. Saker and by Dr. Herman and his collaborators. Since the early work by Prof. Welker on compounds between Group III and Group V elements, interest in this family of materials has grown rapidly and it would appear that, by exploiting some of their unique properties, the already impressive array of semiconductor devices will be considerably extended. In their present paper Dr. Cunnell and Dr. Saker describe the properties of these materials and indicate some of their applications, a field to which the authors have made significant contributions.

The paper by Dr. Herman, Dr. Glicksman and Dr. Parmenter on 'Semiconductor Alloys' is of a more theoretical nature. The contributions of these workers and their colleagues at Princeton to the study of germanium–silicon alloys are widely appreciated and it would be difficult to find a group of more qualified authors.

It is a platitude that the growth of high-quality single crystal material is the foundation stone of semiconductor research. The surprising thing is that significant advances in the art of growing germanium are still being made. The recent interest in horizontal crystal growing encouraged us to ask Mr. Cressell and Dr. Powell, long standing exponents of this technique, to state their case and to describe some of their recent work. The low dislocation densities indicated by some of their photographs are convincing evidence of the efficacy and value of their method. On the other hand, the direct study of dislocations and other defects in crystals is obviously facilitated by controlled methods of introducing high densities of defects by, for example, radiation damage. The effects of nuclear radiation on semiconductors and insulators have been studied for many years but it must be admitted that there remain as many questions as answers. The paper by Dr. Crawford and Mr. Cleland describes both the progress and the problems in this field of research.

Dr. Dunlap's name has long been associated—in my mind, at least—with somewhat exotic impurities in germanium and silicon. His paper stresses these rather than the simpler and better understood Group III and Group V impurities. The properties of gold or manganese doped germanium, for example, are not only interesting in their own right but contribute significantly to the understanding of impurity effects in semiconductors generally. A review of progress at the present time is fully justified, though here again is a field in which further developments can be expected.

Our two remaining papers are concerned with electronic effects at high electric fields in semiconductors. Dr. Curie's paper describes and compares the various theories of electroluminescence. The potential application of electroluminescence to lighting and light amplification has engendered widespread interest in this effect in recent years. Mr. Gunn's paper is concerned with high field effects in germanium and silicon, where considerably more is known about the mechanism of carrier scattering, though even here there are some significant discrepancies between theory and experiment. The experimental study of deviations from Ohm's law in semiconductors, to which Mr. Gunn has made important contributions, is of great value in the whole field of transport phenomena.

Finally, may I be allowed to take this opportunity of offering my sincere congratulations to Prof. Bardeen, Dr. Brattain and Dr. Shockley on the recent recognition of their great contribution to semiconductor research. The award of a Nobel prize in any field of scientific endeavour marks the coming of age of that field. In the presence instance the work of these prize-winners converted a physicists playground into a major industry.

A. F. GIBSON.

CONTENTS

SEMICONDUCTOR ALLOYS

F. HERMAN, Ph.D.

M. GLICKSMAN, Ph.D.

and

R. H. PARMENTER, Ph.D.

David Sarnoff Research Center,
R.C.A. Laboratories, Princeton, N.J., U.S.A.

MS. received 25 *July* 1956

SEMICONDUCTOR ALLOYS

1. INTRODUCTION

The object of this paper is to review some of the recent work on crystalline semiconductors whose atomic structure is characterized by a substantial degree of disorder. Such substances are properly called semiconductor alloys. To be specific, the term *alloy* will be restricted to disordered crystals of the solid substitutional solution type. The ordered phases of alloy systems will be regarded as compounds (ordered stoichiometric crystals). Interest will be centered on the effect of disorder on various physical properties, notably the electronic energy level scheme, the lattice vibrational spectrum, and miscellaneous electrical, magnetic, thermal and optical characteristics.

During the past decade a great deal of effort has been devoted to the study of semiconducting materials[1]. Until relatively recently, most of this activity has been concerned with silicon and germanium[2, 3, 4]. Much of the current pioneering work is being done on semiconducting compounds [5, 6, 7]. An extensive search for new semiconducting materials, both crystalline and non-crystalline, is now under way in many laboratories. The search is motivated partly by the need for substances having optimum properties for particular device applications, partly by the hope of discovering materials with unusual physical properties, and partly by scientific curiosity.

As an out-growth of the current interest in new materials, a number of laboratories have initiated research programs on semiconducting alloy systems. While the objective of discovering substances with unique or desirable properties has not always been realized, these investigations have nevertheless provided a good deal of valuable information.

From a scientific standpoint, the most significant advance has been the work, both experimental and theoretical, on the germanium–silicon alloy system. Since the energy band structure of these alloys can be predicted tolerably well, it has proved possible to interpret, in satisfying detail, the observed variation with alloy composition of certain major electrical and optical properties.

In the present review, no attempt will be made to summarize the results of the numerous experimental investigations which have been carried out in recent years on diverse semiconducting alloy systems. However, adequate references will be given so that the interested reader may trace the original papers. The authors have adopted the view that it is better to indicate the present status of the theory of semiconducting alloys, and to illustrate the theory by typical examples, than to summarize vast amounts of experimental data, most of which, unfortunately, lack suitable theoretical interpretation.

It is the hope that the appearance of this review will not only call attention to interesting but perhaps little-known work, such as that published recently in the Russian literature (see Table 1 below), but will also stimulate further research activity on semiconducting alloy systems.

2. ATOMIC STRUCTURE[8]

There is a fundamental distinction between ordered and disordered crystals, just as there is a basic difference between crystalline and non-crystalline solids. In an ideal ordered crystal, the constituent atoms are arranged in space in the form of a regular array, with the proper type of atom occupying each lattice site. Such solids necessarily satisfy a stoichiometric requirement, and exhibit both long- and short-range order. Substitutional alloys (interstitial alloys will not be considered here) resemble ordered crystals in that the equilibrium positions of all the atomic nuclei form, very nearly, a regular lattice. In such alloys all the normal lattice sites are occupied, with the solute atoms replacing the solvent atoms at some of these sites. The disorder in alloys results from the essentially random arrangement of the solute atoms among the available lattice sites.

In crystalline substances, both ordered and disordered, there is long-range order in the sense that the positions in space occupied by atoms form a lattice. Of course, the identity of the atoms must be ignored when dealing with disordered crystals from this point of view. On the other hand, there is no such long-range order in non-crystalline (amorphous) materials. In the thin amorphous films[9] obtained by evaporation, however, there is usually some degree of short-range order present. (The same statements hold for liquid semiconductors[10].) In the case of amorphous germanium[11], each atom is found to have four nearest neighbors at 2·40 Å and twelve next-nearest neighbors at 3·95 Å, with an essentially random atomic distribution beyond second neighbors. These numerical values are virtually identical to those in single-crystal germanium. (In contrast, each atom in liquid germanium[12] has eight nearest neighbors at 2·70 Å, and there is a random arrangement beyond this distance.) The basic difference between disordered crystals and amorphous solids may be stated as follows: In the former, there is a correlation in the relative positions of different atoms—the identity of the atoms being ignored—for all orders of neighbors; in the latter, there is no correlation beyond a certain order of neighbor.

The distinction between disordered crystals and amorphous solids just stated has a number of important physical consequences. For example, the one-electron energy spectrum and the lattice vibrational spectrum of disordered crystals may be characterized by a Brillouin zone structure, though the reduced wave vector concept has a more restricted meaning than in ordered crystals. Since amorphous materials have no periodicity properties whatever, there is no associated reciprocal lattice, and hence a description of electronic and vibrational quantum states in terms of the zone scheme seems inappropriate. While some of the work to be discussed in subsequent sections has a bearing both on crystalline and non-crystalline materials, the major emphasis will be on the former.

For the purposes of classification, it is convenient to subdivide semiconductor alloys into the following categories.

(1) *Lightly doped ordered crystals.* This class includes nearly perfect crystals having impurity concentrations so small that the individual impurity atoms can be treated independently. If the solute atoms (in substitutional alloys) have a different valency than the solvent atoms they replace, there will usually be localized donor or acceptor states associated with the solute atoms. On the other hand, if the solute and solvent atoms have the same valency, as in

germanium–silicon alloys, localized electronic states having energy levels in the forbidden band are less likely to occur. Since the effect of the impurity atoms on the properties of the host crystal can be treated by perturbation methods, the theory of lightly doped ordered crystals has reached a high state of development. The same is true for the crystals in the next class.

(2) *Crystals with missing or extra host atoms.* If a crystal deviates from the ideal form due to the presence of vacancies and/or interstitials in limited numbers rather than the presence of foreign atoms, the crystal belongs to this category. There will usually be localized donor or acceptor states associated with the vacant lattice sites or the interstitial host atoms. It is often possible to produce vacancy-interstitial pairs by irradiating perfect crystals[13], or by synthesizing crystals under carefully controlled conditions[14].

(3) *Heavily doped ordered crystals.* Included in this category are crystals having impurity concentrations so great that the various impurity atoms can no longer be dealt with separately. For convenience, this class of materials will be restricted to the case of solute atoms having a valency different from that of the solvent atoms they replace. The new feature which arises here is the merging of impurity levels into bands, or, what amounts to the same thing, the interaction of neighboring impurity atoms. The theory of impurity bands is still in an early stage of development. This circumstance may be traced to the inherent mathematical difficulties, and also to the lack of clear-cut experimental results capable of guiding the theory.

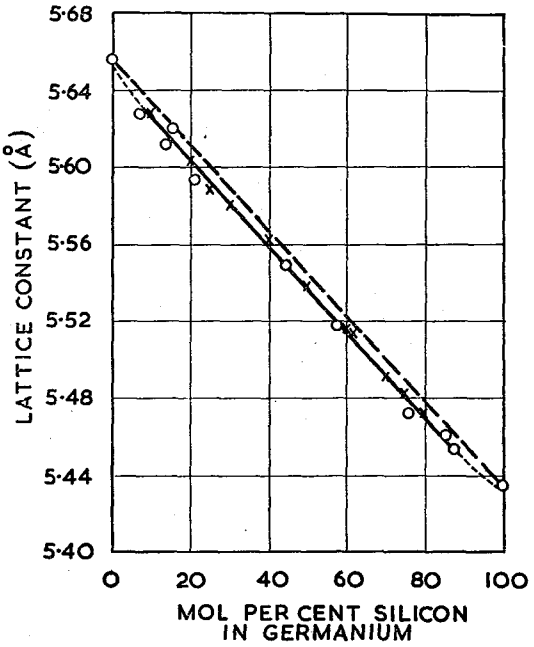
(4) *Mixed crystals.* This category differs from the previous one in two respects: Firstly, it contemplates a wider composition range; secondly, the corresponding solute and solvent atoms here must have the same valency. In most cases this stipulation will be equivalent to the condition that the presence of the solute atoms does not lead to the creation of localized donor or acceptor states. Only if this latter condition is fulfilled can there be semiconducting alloys over a wide composition range. If the impurity concentration in (3) is too great, the impurity bands may become so broad that there is no longer any forbidden band, and the crystal will behave more like a metal than a semiconductor. In the present case, a similar valency does not ensure a wide range of solubility; there are also factors such as ionic radius and electronegativity differences to consider. However, if a wide solubility range exists, it is likely that the system will be semiconducting over the whole of this range.

Perhaps the two simplest examples of mixed crystal systems are the binary systems Ge–Si[15] and Se–Te[16]. Stöhr and Klemm[15] showed that germanium and silicon were miscible in all proportions, and noted that the lattice constant varied monotonically from that of silicon (5·430 Å) to that of germanium (5·657 Å). This curve is shown in Figure 1, with some later data of Johnson and Christian[17] included. X-ray analysis[18] indicates that the germanium and silicon atoms are arranged at random at the sites of a diamond-type lattice over the entire composition range. There is no X-ray evidence[18] for the presence of interstitial germanium or silicon atoms. Analogous studies of the Se–Te systems have been carried out by Grison[16], who showed that selenium and tellurium are mutually soluble in any concentration.

Some work has recently been done on the Ge–Sn alloy system by Trumbore[18a]. Germanium single crystals were grown from melts containing relatively large

amounts of very pure tin. The distribution coefficient (atomic percent Sn in Solid/atomic percent Sn in melt) was found to be approximately 0·02 in the range 0–5 atomic percent Sn in the melt. In the range 5–6 atomic percent Sn in the melt, difficulty was encountered with polycrystalline growth, occlusion, and undoubtedly with constitutional supercooling.

According to X-ray lattice constant determinations, the expansion of the germanium lattice by the inclusion of tin is in agreement with the assumption



[By courtesy of the Editor, Physical Review.

Figure 1. *Variation of lattice constant of germanium–silicon alloys with mol per cent silicon.* X = *Values given by Stöhr and Klemm*[15] *corrected to the more modern values of lattice constant for pure germanium and silicon.* O = *Values of lattice constant measured and reported by Johnson and Christian*[17] *(Johnson and Christian*[17]*)*

that Vegard's law holds over the entire composition range, within experimental error. Since the lattice constant of pure germanium is 5·6576 Å, and that of pure tin (diamond structure) 6·489 Å, Vegard's law predicts a lattice constant of 5·6663 Å for a Ge–Sn alloy with 0·0105 atomic percent Sn, while the observed value is 5·6667 $\pm$0·0014 Å.

Electrical measurements[18a] indicated that at concentrations of tin as high as 10^{20} atoms cm^{-3}, it is truly neutral insofar as its effect on the electrical conductivity of germanium is concerned. Measurements of the free carrier lifetime suggested that at these concentration levels tin is definitely not an effective recombination center for electrons or holes in germanium.

In a mixed crystal of the ternary type, such as ZnS–ZnSe[19, 20, 21], the zinc atoms form one sub-lattice, while the sulphur and selenium atoms are arranged at random on another sub-lattice. In a quaternary alloy such as ZnS–CdSe[19], the zinc and cadmium atoms are randomly arranged on one sub-lattice, and the sulphur and selenium atoms are similarly dispersed on another. Listed in Table 1 are some mixed crystal systems involving three or more constituent atomic species which have been studied experimentally and which have come to our attention. No attempt has been made to compile a complete list of known mixed crystal systems; the contents of Table 1 should be regarded as representative, rather than as exhaustive, especially in the case of the Russian work.

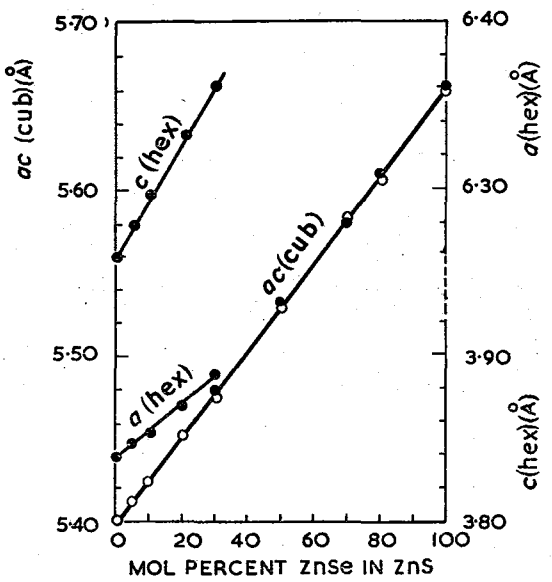


Figure 2. *Variation of lattice constant of ZnS–ZnSe mixed crystals with mol per cent ZnSe. The open circles represent samples prepared at* 850° *C; the heavy dots represent samples prepared at* 1100° *C.* (*Asano*[21])

One of the most interesting features of mixed crystal systems is the relationship between crystal structure and composition. In the case of isomorphous components (substances having the same crystalline form and similar chemical formulae), the same crystal structure will usually persist at all concentrations. However, if one or more of the ingredients can crystallize in more than one crystal form—this situation is described by the term 'polymorphism'—the structure of an intermediate alloy may depend upon the composition, as well as upon the method of preparation. For example, pure zinc sulphide can exist in a number of distinct crystal forms[22], the cubic (sphalerite) and hexagonal (wurtzite) modifications being the most common. Asano[21] has shown that mixtures of zinc sulphide and zinc selenide form cubic solid solutions in any proportion as long as they are crystallized at 850° C. The lattice constant varies linearly with

composition (see Figure 2). In contrast, for a crystallization temperature of 1100° C, the predominant crystal structure tends to be hexagonal up to about 30 mol per cent ZnSe; beyond this proportion the cubic form is the more common. Similar studies on other mixed crystal systems have been reported (see references to Table 1).

Table 1. Mixed Crystals.

Chemical formula	Reference	Chemical formula	Reference
Mg_2Ge–Mg_2Sn	*a*	Tl_2Se–Sb_2Se_3	*g*
HgS–HgSe, HgSe–HgTe	*b*	As_2Se_3–Sb_2Se_3	*h*
HgTe–HgS	*b*	Tl_2Se–$[As:Sb]_2Se_3$	*h*
InAs–InSb	*c, d*	Ga_2Se_3–ZnSe	*i*
InP–InAs, GaP–GaAs	*e*	Ga_2Te_3–ZnTe	*i*
InSb–GaSb	*d, f*	ZnS–ZnSe	*j*
InSb–AlSb, GaSb–AlSb	*f*	[Zn:Cd]–[S:Se]	*k*
GaAs–GaSb, GaAs–InAs	*d*		

a. G. Busch and U. Winkler. *Helv. Phys. Acta.* **26,** 578 (1953).
b. E. I. Nikol'skaya and A. R. Regel. *Z. Tek. Fiz.* **25,** 1347 and 1352 (1955) [In Russian].
c. C. Shih and E. A. Perretti. *J. Amer. Chem. Soc.* **75,** 608 (1953); *Trans. Amer. Soc. Metals.* **46,** 389 (1954); *Trans. Amer. Soc. Metals Preprint.* **48,** (45) (1955).
d. N. A. Goryunova and N. N. Fedorova. *Z. Tek. Fiz.* **25,** 1339 (1955) [In Russian].
e. O. G. Folberth. *Z. Naturf.* **10a,** 502 (1955).
f. W. Köster and B. Thoma. *Z. Metallkunde.* **46,** 293 (1955); J. S. Blakemore. *Can. J. Phys.* **35,** 91 (1957); J. C. Woolley, B. A. Smith, and D. G. Lees. *Proc. Phys. Soc.* **B69,** 1339 (1956).
g. B. T. Kolomiets and N. A. Goryunova. *Z. Tek. Fiz.* **25,** 984 (1955) [In Russian].
h. N. A. Goryunova and B. T. Kolomiets. *Z. Tek. Fiz.* **25,** 2069 (1955) [In Russian].
i. N. A. Goryunova, V. A. Kotovich and V. A. Frank-Kamenetskii. *Doklady Akad. Nauk. S.S.S.R.* **103,** 659 (1955); N. A. Goryunova, V. A. Grigor'eva, B. M. Konovalenko and S. M. Ryvkin. *Z. Tek. Fiz.* **25,** 1675 (1955) [In Russian].
j. J. A. Amick (private communication); H. W. Leverenz. *An Introduction to Luminescence of Solids.* (Wiley, New York, 1950); H. A. Klasens. *J. Electrochem. Soc.* **100,** 72 (1953).
k. H. W. Leverenz, E. J. Wood, R. E. Shrader and S. Lasof. Contribution to G. R. Fonda (Ed.). *Preparation and Characteristics of Luminescent Materials* (Wiley, New York, 1948).

(5) *Crystals with random stacking.* Many crystals are known[22] to consist of perfectly ordered two-dimensional layers stacked with respect to one another in a random fashion. Such crystals are ordered along directions parallel to the layers, and disordered along the direction perpendicular to the layers. Polymorphism is closely related to the phenomenon of random stacking, and the transition from one polymorphic modification to another is believed to take place through the rearrangement of the stacking of the various layers. An isolated stacking fault in an otherwise perfect crystal may be regarded as a twin boundary[22a]. It has recently been observed[22b] that the electroluminescence properties of certain phosphors can be correlated with the degree of structural disorder apparently arising from random stacking.

(6) *Crystals with randomly arranged host atoms.* In certain polyatomic crystals, some or all of the host atoms are arranged at random among the available lattice sites. If the randomly arranged atoms lying on a given sub-lattice have the same valency, we have mixed crystals of the type belonging to category (4). If such atoms have different valencies, we have a somewhat different situation. As an example, we will consider stoichiometric or nearly stoichiometric Cs_3Sb.

According to recent work by Jack and Wachtel[22c], the crystal structure of Cs_3Sb consists of two inter-penetrating diamond-type sub-lattices. Taken together, these two sub-lattices form a body-centered cubic lattice. X-ray measurements indicate that all the sites of one sub-lattice are occupied by caesium atoms, and that caesium and antimony, in equal or nearly equal proportions, are arranged at random at the sites of the other sub-lattice. It is not clear from the work of Jack and Wachtel what types of impurity states arise from the disordered arrangement of the caesium and antimony atoms. Whatever their nature, the presence of these impurity states distinguishes this case from that of mixed crystals belonging to category (4).

It should be emphasized that so far as atomic disorder is concerned, *i.e.*, disorder due to point imperfections, the concentrations of various types of imperfections are not independent of each other[14, 22d]. For example, during the growth of a crystal from the melt, the amount of a given impurity in the melt that will be dissolved in the crystal depends upon the types and concentrations of impurities already present in the crystal.

In summary, we observe that disorder can arise from the presence of foreign atoms ((1) and (3) above), from the presence of missing or extra host atoms (2), from the random arrangement of host atoms ((4), (5), and (6)), and from 'architectural' disorder (5). Of course, 'architectural' disorder can also occur through the presence of dislocations; as the dislocation density is increased, we go over gradually from a crystalline to a non-crystalline substance.

3. ONE-ELECTRON ENERGY LEVELS

A number of attempts to calculate the electronic energy level distribution in disordered crystals have been made in recent years. A good deal of useful qualitative information has already been obtained by the examination of idealized one-dimensional alloys. Some inroads into the more difficult problem of a realistic three-dimensional alloy have been made with the aid of perturbation techniques. In the present section we will survey some of the recent theoretical work on alloys, emphasizing those conceptual aspects which are particularly pertinent to semiconducting alloys.

As is well known, the problem of determining the energy levels of electrons in a perfect crystal is greatly simplified by the translational symmetry of the crystal. In fact, it is possible to obtain an eigensolution valid for the entire crystal volume merely by solving the appropriate wave equation within a single unit cell. Since this simplification can no longer be made in the case of a disordered crystal, it is much more difficult to study a disordered crystal than an ordered one.

In a perfect crystal, the electronic energy levels fall naturally into well-defined bands. The transition from an allowed to a forbidden band of energies is abrupt. Within a given band, each quantum state is characterized by a reduced wave vector $\mathbf{k}$. This follows from the fact that $\mathbf{k}$ is a constant of the motion, or, as it is often expressed, $\mathbf{k}$ is a 'good' quantum number. Much of the theoretical work to be described below has been concerned with determining the extent to which the concepts of allowed bands, forbidden bands, and quantum states characterized by $\mathbf{k}$ carry over to the case of disordered crystals.

James and Ginzbarg[23] and Landauer and Helland [24] approached the problem by constructing an idealized one-dimensional random alloy, and then determining the energy level distribution by numerical methods. It was found that the presence of disorder tends to smear out the transition regions between allowed and forbidden bands, so that the density of states decreases gradually, rather than abruptly, in the passage from allowed to forbidden bands. While these direct calculations provide a useful qualitative picture of the effect of disorder on the energy level distribution, they provide no general rules for the one-dimensional case, and, more seriously, they cannot be extended in an obvious manner to the practical case of three-dimensional alloys.

Various authors, including Saxon and Hutner[25], Luttinger[26], Kerner[27], Hart[28] and des Cloizeaux[29], have treated idealized one-dimensional alloys by analytical rather than numerical methods, in the hope of formulating general principles relating to the energy level distribution. Some progress has been made in the direction of answering such questions as the following: What is the probability that a given energy in a binary alloy AB is allowed (forbidden) if the corresponding energy in the pure constituents A and B is allowed (forbidden)?

While the above mentioned one-dimensional studies are instructive, they do not provide the type of quantitative information—such as the detailed form of the transition regions—that is essential to the analysis of three-dimensional alloys. In particular, there is reason to believe that the nature of the transition region depends upon the dimensionality of the crystal, so that the results for one-dimensional cases are not directly applicable to two- or three-dimensional cases. For information pertinent to three-dimensional problems, it is necessary to turn to the early work of Nordheim[30] and Muto[31], and to the more recent work of Parmenter[32, 33, 34], all of which are based on perturbation methods. Since some of this work employs Nordheim's virtual crystal model of an alloy, it is desirable to introduce the virtual crystal concept at this point.

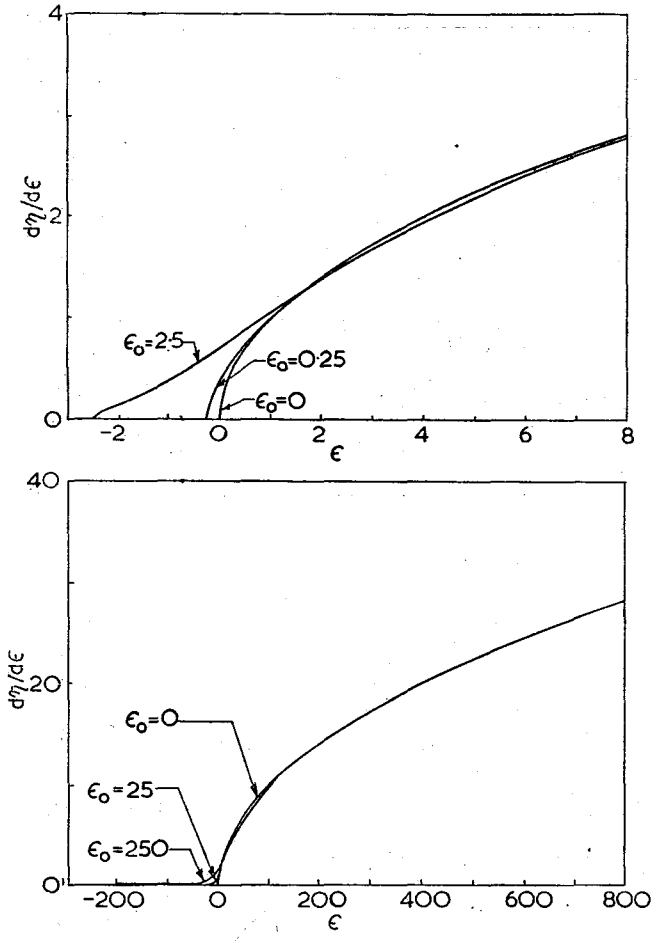
Consider an aggregate of atoms arranged at random on the sites of some periodic lattice. The potential acting on an electron moving through the disordered alloy formed by the random configuration of atoms can be decomposed into a periodic part and a non-periodic part, as follows:

$$V_{\text{alloy}}(\mathbf{r}) = V_{\text{periodic}}(\mathbf{r}) + V_{\text{non-periodic}}(\mathbf{r}).$$

The decomposition is carried out in such a manner that the periodic part is the average of the alloy potential over all possible random configurations of atoms in the alloy, consistent with the given composition of the alloy. $V_{\text{non-periodic}}(\mathbf{r})$ is then the deviation of $V_{\text{alloy}}(\mathbf{r})$ from its average, $V_{\text{periodic}}(\mathbf{r})$. In a given alloy, the non-periodic part will have a form which depends upon the specific arrangement of the atoms, while in an ensemble of alloys of given composition, the non-periodic part will have definite statistical attributes. The virtual crystal is defined by the periodic component, and the effect of disorder is represented by the non-periodic component.

Nordheim[30] used the virtual crystal concept in the course of a study of the residual resistance of alloys. He proceeded essentially by first order perturbation theory, regarding the virtual crystal potential $V_{\text{periodic}}(\mathbf{r})$ as the unperturbed potential, and the disorder potential $V_{\text{non-periodic}}(\mathbf{r})$ as the perturbing potential. Somewhat later, Muto[31] attempted to calculate the energy level distribution of an

alloy on the basis of the virtual crystal viewpoint. Muto reasoned that the average over the Bloch functions of the virtual crystal is zero, so that there is no shift in the energy levels produced by disorder in the first order of perturbation theory. He suggested that a smearing of the band edges would take place in the second order of perturbation theory.

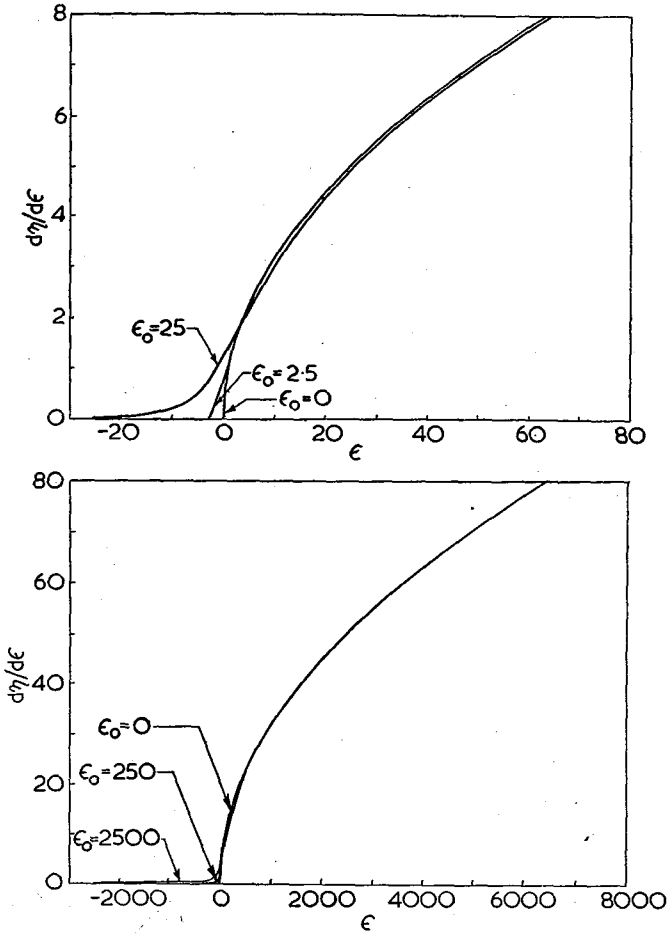


[*By courtesy of the Editor, Physical Review*

Figure 3a. *$d\eta/d\epsilon$ versus ϵ for various values of ϵ_0. (Parmenter*[32]).

Subsequent work by Parmenter[32, 34] extended the virtual crystal approach to the higher orders of conventional perturbation theory. Parmenter showed that the higher orders have only a minor effect on the energy level distribution near the middle of an unperturbed band. Near the top or bottom of an unperturbed band, however, the higher orders can have an important effect. In general, even orders of perturbation theory tend to raise or lower band edges according to whether the perturbing potential, $V_{\text{non-periodic}}(\mathbf{r})$, is predominantly repulsive or

attractive, respectively. Odd orders of perturbation theory, other than first order, always tend to move the band edge out into the forbidden band, irrespective of the nature of the perturbing potential. The net effect is that band edges are smeared out into the forbidden bands. For three-dimensional alloys, this smearing is always finite in extent, so that the forbidden bands need not disappear together.



[*By courtesy of the Editor, Physical Review*

Figure 3b. $d\eta/d\epsilon$ *versus* ϵ *for various values of* ϵ_0 (*Parmenter*[32]).

The contribution of second order perturbation theory to the smearing of a band edge can be seen from Figure 3, taken from Parmenter[32]. In this figure, ϵ is proportional to the energy and $d\eta/d\epsilon$ to the density of allowed states per unit energy. The parameter ϵ_0 determines the shape of the energy tail extending into the forbidden band; *e.g.* $\epsilon_0 = 0$ corresponds to a vanishing perturbation potential.

Under conditions that a perturbation approach is valid (see below), it is

possible to label the quantum states in the disordered alloy by band indices and reduced wave vectors **k**. This follows from two considerations:

(1) A conventional Brillouin zone scheme and conventional energy bands can be associated with the periodic virtual crystal potential.

(2) Perturbation theory provides a bridge between the case of strict periodicity and the case of disorder in the limit of vanishingly small disorder.

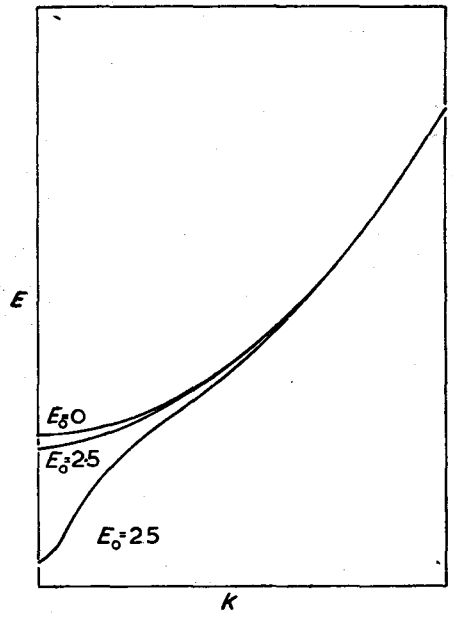


Figure 4. **E** *versus* **k** *for various values of* E_0. (*Parmenter*[32]).

Although **k** is not really a good quantum number in a disordered alloy, however small the disorder, it can be used to label the quantum states in a restricted sense. To this extent, it can be employed[32] in connection with the determination of such quantities as the density of allowed states per unit energy, the expectation value of velocity of an electron, and the electron effective mass (the proportionality constant relating an externally applied force and the resultant acceleration of the electron).

In Figure 4 are shown the $E(\mathbf{k})$ curves corresponding to some of the density-of-states curves of Figure 3. An examination of the low-energy tail suggests that rather large velocities and rather small effective masses may result from the alloying in this energy region. While these effects tend to increase the electron mobility, they are probably offset by the incoherent scattering resulting from the disorder, so that the net effect would be a decrease in mobility with alloying, rather than an increase. A further effect which may be predicted theoretically[34]

is that, upon alloying, anisotropic effective masses at the bottom or the top of an energy band will become more nearly isotropic.

The virtual crystal approach leads to wave functions which can be misleading unless properly interpreted. In the case of a germanium–silicon alloy, for example, it is expected on physical grounds that a state near the bottom of the valence band has a wave function which resembles a germanium 4s atomic orbital in the neighborhood of germanium nuclei, just as this wave function should resemble a silicon 3s atomic orbital near the silicon nuclei. If the virtual crystal concept is misinterpreted it might be thought that the nodal structure of the wave function is averaged in the same manner as the potentials, so that the state just considered would have a wave function with a nodal structure intermediate between germanium 4s and silicon 3s near all the crystal nuclei.

When properly interpreted, the virtual crystal approach requires that a given wave function have the appropriate nodal structure everywhere in space. A recent formulation by Parmenter[33], which, incidentally, is particularly suitable for actual numerical calculations, guarantees that the electronic wave functions behave properly in the vicinity of each crystal nucleus.

Muto's[31] utilization of non-degenerate perturbation theory, in conjunction with the virtual crystal model of a disordered alloy, has been criticized by James and Ginzbarg[23] on the ground that the perturbing potential mixes originally degenerate states together. While it is true that degenerate perturbation theory, rather than non-degenerate perturbation theory, must be employed when the calculation is restricted to a given order, Parmenter[32, 33, 34] has pointed out that the non-degenerate form is justified as long as one is prepared to carry the calculation to all orders of perturbation theory.

Under certain conditions, all of the states in the virtual crystal energy bands are modified only slightly by the presence of disorder. When this situation obtains, it is expected that all the states can be correctly represented by rapidly convergent perturbation-type expansions. If the presence of imperfections leads to a splitting-off of some states from the virtual crystal energy bands, the split-off states, and the remaining states as well, cannot be treated by perturbation methods, since the perturbation expansions diverge. When there is very substantial disorder present, the perturbing potential may be so large that a perturbation treatment fails. Such questions as these have received some attention[34], but deserve further study.

The electronic states which are split-off from the unperturbed energy bands by the presence of imperfections are of particular interest in semiconductors, since they form the donor and acceptor states. For sufficiently small impurity concentrations, the energy levels associated with these localized states will be sharp, and the associated density-of-states curve will be a delta function centered on the energy of these levels. As the impurity concentration increases, the impurity-level wave functions centered on neighboring impurity atoms will start to overlap, causing the originally sharp levels to spread into a narrow band of energies—the so-called impurity band.

A number of attempts have been made to study impurity bands theoretically, in particular, to determine the density-of-states curve. The one-dimensional model of an impurity band studied by James and Ginzbarg[23] has already been mentioned. The main objection to this model is that it is one-dimensional, so

that conclusions reached need not necessarily apply to three-dimensional problems. For example, there will always be at least one discrete state associated with an impurity atom in a one-dimensional problem, no matter how small the perturbing potential of the impurity is. On the other hand, small localized perturbing potentials do not necessarily give rise to discrete states in three dimensions.

A number of workers[35–38] have assumed that the impurity atoms form a superlattice in the host crystal. By this assumption, the problem is given spatial periodicity, so that a conventional energy-band solution may be obtained. Unfortunately, such an assumption removes one of the most important features of the problem—*viz.* the disorder. A three-dimensional model which attempts to take the disorder into account has been studied by Aigrain and Jancovici[39, 40]. These authors obtained the expectation value of energy with respect to approximate, mutually orthogonal wave functions which were so constructed as to incorporate the disorder of the problem. It was possible to define a pseudo-quantum number $\mathbf{k}$, having the significance that the density-of-state curve could be obtained from the energy $E(\mathbf{k})$ in exactly the same manner as for an ordinary energy band.

The width in energy of a band formed by a superlattice of impurity atoms will vanish in an exponential fashion as f, the mol fraction of impurity atoms, goes to zero. This results from the nature of the overlap of the impurity atom wave functions. On the other hand, the bandwidth associated with a disordered set of impurity atoms will vanish much less rapidly with vanishing f. This difference in behavior results from the fact that the disorder admits of the possibility that two impurity atoms may be close together even when f is small.

In summarizing this section, it may be said that considerable progress has been made, particularly in recent years, in obtaining a conceptual understanding of the nature of electronic energy levels in disordered alloys. Despite this, much work remains to be done. For example, the present status of this work in no way compares with the advanced state of development of the theory of energy bands of a perfect crystal[41]—especially with regard to numerical calculations on specific systems. Concerning the even more difficult problem of energy levels in liquids[42, 43] or amorphous solids[9], very little work has been done.

[*Note added in proof.*] Mott[43a] has recently raised an objection to the view expressed by some writers on impurity band conduction that, with only one type of impurity center present, an array of singly charged centers can be treated as a half-filled band, the gap between this band and the conduction band decreasing to zero as the concentration increases. The view to which Mott objects is based on the Bloch–Wilson theory of semiconductors. According to this theory, there will always be conduction in the band produced by an array, ordered or disordered, of singly charged centers, and the effective mass of the free carriers will decrease *continuously* as the impurity concentration is increased. If the only type of impurity present is an aggregate of divalent impurity atoms, the Bloch–Wilson theory predicts that, at absolute zero temperature, there is no conduction in the (filled) ground state impurity band until such high concentrations are reached that this band overlaps the next highest impurity band.

Mott argues that the Bloch–Wilson theory does not provide the correct

description of the wave functions of electronic states associated with sufficiently dilute concentrations of monovalent or divalent impurity atoms embedded in a semiconductor. Below a critical impurity concentration, Mott believes these wave functions to be highly localized at the impurity sites, rather than non-localized and extending over the entire crystal (Bloch-like wave functions). There would be no electrical conduction according to such a picture. When the impurity concentration exceeds a certain critical value, Mott suggests that the impurity-state wave functions suffer a drastic change in character, changing from ones having a localized nature to others having a non-localized nature. It is predicted that a *sudden* onset of conduction in the impurity band system accompanies this drastic change in the character of the impurity-state wave functions. Although this theoretical prediction is not borne out by the available experimental evidence, Mott feels that more refined measurements on single crystal samples will bear out his contention.

In considering Mott's work, it should be recalled that the Bloch–Wilson energy band theory is based on the one-electron approximation. What Mott is doing, in effect, is going beyond the conventional one-electron approximation, and introducing correlations between particles (electrons at impurity centers and their parent impurity atoms) not given by the Block–Wilson (or Hartree–Fock) scheme. The adequacy of Mott's qualitative approach cannot be judged in the absence of a simple, quantitative many-electron theory of electrons in crystals.

It has been emphasized by Conwell[43b], and also by Mott[43a], that impurity conduction in compensated samples can have a character quite different from that in uncompensated samples, *i.e.*, in samples having only one type of impurity present. Consider an N-type semiconductor having (per unit volume) N_D donor levels and a smaller number N_A acceptor levels. It will be assumed that the donor and acceptor impurity atoms have hydrogen-like impurity states. At the absolute zero of temperature, all the acceptor levels will be occupied, and negatively charged, since there are enough donor electrons present to fill the holes in the valence band created by the presence of the acceptor atoms. Under these conditions, there will be $N_D - N_A$ neutral donor atoms, and N_A positively charged donor atoms with an unfilled level. These vacant sites (positive holes) can always jump to adjacent, occupied donor sites, provided these adjacent sites are sufficiently close. Clearly, the jumping of the positive holes from donor site to donor site allows for P-type electrical conduction.

According to Mott[43a] and others[43c, 43d], an activation energy must be associated with the jumping of vacant sites, and the existence of such an activation energy will have a profound effect upon the temperature dependence of the associated electrical conduction. James[43d] and Anderson[43d] have gone even farther, and postulated the existence of networks of closely spaced sites, along which the conduction occurs in a threshold impurity concentration range. This follows from the fact that the probability of jumping decreases with increasing separation between sites, so that the actual, overall jump path will be determined by networks of the most closely spaced sites. James[43d] has shown that the formation of these networks can be treated to advantage by methods well known in high polymer theory, while Anderson[43d] has reported a successful treatment of this problem by a different method.

4. LATTICE VIBRATIONAL SPECTRUM

The problem of determining the vibrational spectra of disordered crystals can be approached theoretically from several points of view. First, the disordered crystal can be replaced by the corresponding virtual crystal, and the spectrum of the latter can be calculated by standard methods. Next, an attempt can be made to obtain rigorous solutions for highly idealized models, and then infer from the results the nature of the solutions for more realistic cases. Finally, in the limit of extremely dilute alloys (very small amount of disorder), the problem can be treated in a variety of ways by perturbation methods. Some of the work which has already been done along these lines will be reviewed briefly below.

It should be noted at the outset that the problem of deducing the vibrational spectrum of an alloy is in many respects analogous to that of finding the spectrum of one-electron energy levels in an alloy. In each case there is an eigenvalue equation to be solved, and in each case the complicating factor is the presence of disorder. Information of a qualitative nature obtained by solving either problem is often directly applicable to the other. For example, just as certain interruptions in the crystal periodicity give rise to localized electronic states, so are particular structural imperfections expected to result in localized normal modes of vibration.

According to the virtual crystal approximation, an alloy is replaced by an equivalent perfect crystal in which each lattice site is occupied by the appropriate type of virtual atom. The force constants which enter into the dynamical theory must be taken as suitably averaged quantities. As an illustration of the method[44] consider a germanium–silicon alloy having the composition Ge_fSi_{1-f}. The virtual crystal is a diamond-type crystal, so that the results of Smith's work on diamond[45] are pertinent to the mathematical development of the theory.

The lattice vibrational Hamiltonian for a germanium–silicon alloy has the following form[44]:

$$\mathcal{H} = -\hbar^2/2\ [f\sum_{Ge} M_{Ge}^{-1}\ \partial^2/\partial X^2_{Ge} + (1-f)\sum_{Si}\ M_{Si}^{-1}\ \partial^2/\partial X^2_{Si}] + f^2/2\ \sum_{Ge}\sum_{Ge} \partial^2/\partial X_{Ge}\ \partial X_{Ge} + 2f\,(1-f)/2\sum_{Ge}\sum_{Si}\partial^2/\partial X_{Ge}\,\partial X_{Si} + (1-f)^2/2\sum_{Si}\sum_{Si}\partial^2/\partial X_{Si}\ \partial X_{Si}. \quad \ldots (1)$$

where M_{Ge} and M_{Si} are the masses of germanium and silicon atoms, and X is the nuclear displacement from equilibrium. The summations are carried out over all lattice sites as indicated. In order to apply the virtual crystal approximation, it is necessary to average the various terms in the Hamiltonian (1) over all random configurations consistent with the composition Ge_fSi_{1-f}. The outcome is that an atom of suitably averaged mass $M(f)$ is placed at each lattice site; also, each pair of atoms; i, j, is coupled by a suitably averaged force constant $[\partial^2/\partial X_i \partial X_j]_f$. It follows from (1) that:

$$M^{-1}(f) = f\,M_{Ge}^{-1} + (1-f)\,M_{Si}^{-1} \quad \ldots (2)$$

$$[\partial^2/\partial X_i \partial X_j]_f = (f^2\,\partial^2/\partial X_{Ge}\partial X_{Ge} + 2f(1-f)\partial^2/\partial X_{Ge}\,\partial X_{Si} + (1-f)^2\,\partial^2/\partial X_{Si}\,\partial X_{Si}. \quad \ldots (3)$$

If we assume that the coupling between a Ge–Si pair is the arithmetic average of

the couplings between Ge–Ge and Si–Si pairs, (3) reduces to:

$$[\partial^2/\partial X_i \partial X_i]_f = f\,\partial^2/\partial X_{\mathrm{Ge}} \partial X_{\mathrm{Ge}} + (1-f)\,\partial^2/\partial X_{\mathrm{Si}} \partial X_{\mathrm{Si}}. \quad \ldots (4)$$

If attention is restricted to nearest-neighbour interactions, the coupling constants on the right-hand side of (4) can be determined from elastic constant data[46]. Carrying out the necessary calculations[44], it is found that the virtual crystal approximation predicts a substantially linear variation with composition of nearly all the normal modes of vibration, and hence of the lattice spectrum as a whole. Further analysis[44] suggests that the virtual crystal approximation is satisfactory for long wavelength acoustical modes, but not for other types of modes.

A rigorous solution for the vibrational spectrum of a one-dimensional chain has been given by Dyson[47]. The chain is composed of point masses with nearest neighbors coupled by elastic springs obeying Hooke's law. Considering only longitudinal vibrations, and regarding the masses and strengths of the springs as random variables, Dyson derives in closed form an expression for the normal mode distribution. The result depends, of course, on the probability laws that are taken to represent the distribution of the masses and spring constants. Dyson's study of a special case suggests that as these distributions become wider and wider, the high-frequency portion of the spectrum grows at the expense of the low-frequency portion, and the upper extremity of the spectrum changes from a form having a sharp cut-off to one having a long tail.

Unfortunately, multi-dimensional problems do not appear amenable to treatment by the elegant formalism developed by Dyson for the one-dimensional case. About the best that can be done at present on the theoretical side is to consider multidimensional cases in the limit of vanishingly small disorder.

One of the first studies along these lines was that of Stripp and Kirkwood[48]. They investigated the vibrational spectrum and partition function of very dilute (three-dimensional) alloys by means of a powerful perturbation technique. A classical, closely packed, monatomic crystal served as the model, and the nearest-neighbor, harmonic approximation was employed in formulating the problem. The only defects considered were small numbers of vacant lattice sites. It was found that the influence of these vacancies on the thermodynamic properties increases in relative importance with the absolute temperature. Another interesting result was that a pair of vacancies will attract each other with an average force whose potential diminishes asymptotically as $1/R^3$, where R is the distance of separation of the vacancies.

Montroll and Potts[49, 50] have also examined theoretically the effect of localized defects on the vibrations of a crystal. Their approach is somewhat more general than that of Stripp and Kirkwood, in that the localized defects can be impurity atoms and interstitial atoms as well as vacancies. Montroll and Potts[49] find: (1) that the vast majority of the normal modes are perturbed only slightly by the presence of defects, and (2) that various types of defects can cause localized normal modes whose effect dies out rapidly with distance from the defect. These localized modes have vibrational frequencies which are displaced out of the frequency bands of the perfect crystal. In this respect they resemble localized electronic impurity states in semiconductors.

Montroll and Potts[49] also develop a theory of the interaction of two defects

as a function of their distance of separation. The change in the zero-point energy of the crystal produced by the introduction of a defect pair was determined in the range of very low temperatures. The calculations show that like defects attract each other in a monatomic lattice. In their most recent paper[50], they also consider the interaction between localized defects and the boundaries of the crystal. They find that the defects will be attracted or repelled by the boundaries, depending upon the nature of the defects.

The lattice spectrum of a non-dilute alloy can be determined by experimental methods as well as by theoretical ones. In fact, certain experimental techniques seem ideally suited to the problem. The methods of temperature-diffuse X-ray scattering[51] and cold neutron scattering[52] have already proved their value in studies of metals and metallic alloys; if applied to semiconductor alloys they might yield valuable information. Investigations of the infra-red lattice absorption and of the Raman effect are also likely to provide important information bearing on the lattice spectrum.

5. OPTICAL PROPERTIES

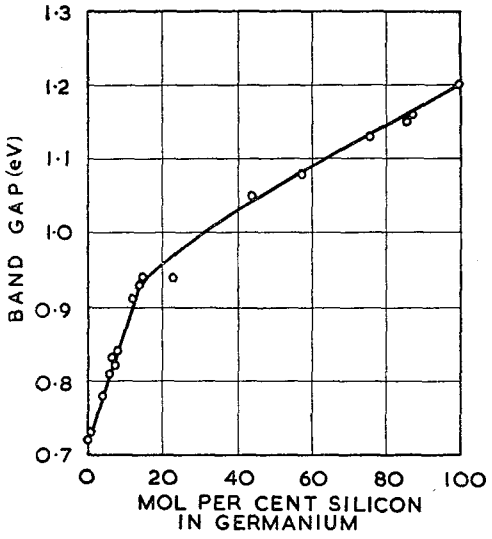
In general, the intrinsic and extrinsic optical properties of a semiconducting alloy will depend upon (1) the nature of the electronic quantum states; (2) the nature of the lattice vibrational spectrum; and (3) the coupling between the electrons and the lattice vibrations. As long as ordering effects do not occur, these three factors may be expected to vary in a continuous manner with composition in a given alloy system. Therefore, a continuous variation of the optical properties with composition is anticipated.

This expectation is generally borne out by the experimental results (see references in Table 1). For example, the threshold for fundamental absorption (due to electronic transitions across the forbidden band) is found to shift gradually with composition in a number of alloy systems, including the following: Ge–Si[17, 53, 54]; Se–Te[55]; InP–InAs[56]; and GaP–GaAs[56]. For the Ge–Si system, data are available over the entire composition range, while in the other three cases some ranges have not yet been examined. In some systems, notably Ge–Si and GaP–GaAs, the rate of change of the absorption edge with composition has markedly different values at different compositions. While a gradual variation in the rate of change does not call for any detailed explanation, a *discontinuity* of the type observed in the Ge–Si systems (see below) needs an interpretation.

Because of this unusual behavior, it is instructive to consider the Ge–Si alloys in some detail. In Figure 5 is shown the experimental data for the variation of the optical gap with composition, as reported by Johnson and Christian[17]. Note that the curve exhibits a marked change in slope in the neighborhood of 15 atomic per cent silicon. In order to explain this unexpected behavior, it is necessary to consider the energy band structure of the separate components, and of the intermediate alloys (see Figure 6).

For both germanium and silicon, the valence band edge (*i.e.* absolute maximum) occurs at the central point of the reduced zone. The conduction band edge lies along the [111] and the [100] axes of the reduced zone in germanium and silicon, respectively. Since a good deal is known about the band structure of

the two extremes[2, 3, 4], including the gross topology of the lowest conduction bands and the highest valence bands, it is a relatively simple matter[57] to predict what sort of change in the band structure with alloying would lead to the experimental results shown in Figure 5. For example, it may be inferred from the results of some theoretical work[58] that the valence band edge remains at the central zone point over the entire composition range, and, further, that the major changes occur in the conduction band structure. Of course, the spin-orbit splitting between certain pairs of bands will vary with composition, but this feature is of no real importance for the problem at hand.



[*By courtesy of the Editor, Physical Review*

Figure 5. *Variation of the energy gap in germanium–silicon alloys with mol per cent silicon, as determined by optical absorption measurements. (Johnson and Christian*[17])

As silicon atoms are substituted for germanium atoms, the three types of minima in the lowest conduction band of germanium (see Figure 6*c*) move away from the highest valence band state, but at different rates. The [000] minimum rises most rapidly, and the [100] minima least rapidly, the motion of the [111] minima being intermediate. For the moment, the role that phonons play in assisting the optical absorption will be ignored, and the intermediate alloys will be treated in terms of their virtual crystals. The threshold for absorption is then determined by the energy separation between the top of the valence band and the bottom of the conduction band, irrespective of the location of the latter in the reduced zone.

On the basis of these remarks, it is clear that the rapid change of the optical gap in the range 0 to 15 atomic per cent silicon can be attributed to the rapid motion of the [111] minima relative to the valence band edge. At about 15 per cent, the [111] minima rise above the [100] minima, and the relatively slow

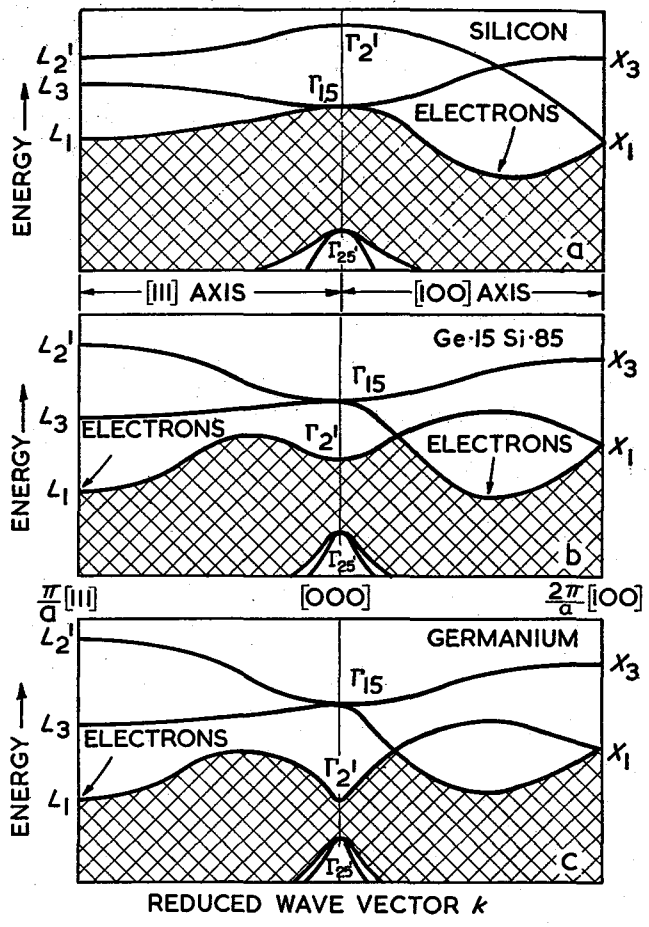


Figure 6. *Schematic diagrams of electronic energy bands of* (a) *pure silicon,* (b) *germanium–silicon alloy, composition* 85/15 *and* (c) *pure germanium. For reasons of simplicity, the spin-orbit splitting is not shown. In all three cases, the top of the valence band is at the central zone point. In* (a) *and* (c), *the bottom of the conduction band lies along the* [100] *and the* [111] *axes respectively. In* (c), *the* [000] *minimum and the* [100] *minima lie* 0·10 *and* 0·18 *eV above the* [111] *minima, respectively. In* (b), *the bottom of the conduction band occurs along both the* [100] *and the* [111] *axes. Symbols such as* X_1 *denote the symmetry classification of the electronic states, in the conventional group theoretical notation.* (*After Herman*[4, 57, 58])

change in the remaining composition range is due to the slower motion of the [100] minima. In going from pure germanium to pure silicon, the conduction band electrons first occupy states only near the [111] minima, then near both the [111] and the [100] minima, and finally only near the [100] minima. The conduction band edge switches its location in the reduced zone at some critical

composition, which, according to the experimental evidence[17], is approximately 15 atomic per cent silicon. While the [111] and the [100] minima are within a few kT of each other on an energy scale, states near both types of minima will be occupied.

As has already been mentioned, the valence and conduction band edges in Ge–Si alloys occur at different positions in the reduced zone over the entire composition range. Since optically induced electronic transitions involving a change in $\mathbf{k}$ require the assistance of phonons, a detailed study of the optical absorption as a function of composition must take into account not only the change in band structure with alloying, but also the corresponding changes in the lattice spectrum and in the electron–phonon coupling.

Detailed measurements of the fundamental optical absorption in Ge–Si alloys have been carried out by Braunstein[53] and by Moore[54]. Using the phenomenological theory of Macfarlane and Roberts[59] as a guide, Braunstein and Moore are able to determine the composition dependence of the forbidden bandwidth to a high degree of precision. Their results agree qualitatively with those reported previously by Johnson and Christian[17]. In addition, Braunstein and Moore are able to deduce the frequencies of the phonons which assist the nonvertical electronic transitions. The variation of these phonon frequencies with composition lends further credence to the theoretical interpretation[57] of the optical behavior. Finally, Moore[54] has examined the effect of alloying upon the absorption due to vertical ($\mathbf{k}$-conserving) electronic transitions, at the germanium-rich end of the range. He found that the motion of the [000] minimum is roughly twice as rapid as that of the [111] minima, a result which is consistent with the theory[57].

In summary, it might be observed that a study of the intrinsic optical properties can shed considerable light on the energy band structure of an alloy system. The more that is known in advance about the band structure, the more can be deduced from a careful examination of the optical properties. If the band structures of the two extreme cases are known in sufficient detail, it will usually be possible to predict the intermediate behavior with the aid of symmetry arguments and topological considerations.

Let us now turn to some of the extrinsic optical properties, particularly the absorption and emission associated with electronic transitions to or from localized states due to the presence of impurity atoms.

Consider first an impurity atom having a ground state energy level which lies close (within a few hundredths of an eV) to the valence or conduction band edge. Since the associated electronic orbit will extend over many hundreds or thousands of atomic cells, this impurity state will not be greatly influenced by the particular host atoms which happen to be its immediate neighbors. In such cases, there will be only one donor or acceptor ground state over the entire composition range, and the energy level of this state will vary essentially continuously with composition. Interesting effects may take place if there is a switchover in the band structure bordering the impurity level. For example, in the case of column V donors in the Ge–Si alloy system, there should be a change in the multiplet structure[60] and in the order of certain excited states as the critical composition (15 atomic per cent silicon) is traversed.

For purposes of comparison, consider next an impurity atom having a ground

state energy level which lies deep within the forbidden band, say at least a few tenths of an eV away from the nearest band edge. The electronic orbit will now be confined to relatively few atomic cells, so that the nature of the state will be strongly influenced by the particular host atoms which occupy positions adjacent to the impurity atom in question. Under such conditions, a multiplication in the number of ground states will occur as a sweep across the composition range is made.

This actually occurs in the emission spectrum of silver-activated zinc sulphoselenides[20]. As selenium is substituted for sulphur, the original emission bands shift to longer wavelengths, and at the same time new emission bands appear which are also shifted to longer wavelengths with increasing selenium content. The original emission bands are undoubtedly associated with a silver atom substituting for a zinc atom and having four sulphur atoms as nearest neighbors. On this interpretation, the new bands would arise from silver atoms having one or more selenium atoms as nearest neighbors. While the overall behavior is *continuous* in the sense that emission bands grow or decay very gradually with composition, it is certainly more complicated than the behavior in the previous example, where instead of a multiplicity of ground state impurity levels, there is only one.

The variation with composition of certain ionization energies of copper and gold impurities in germanium-rich germanium–silicon alloys has recently been studied by Schultz[61]. Now it is known[61a] that both copper and gold introduce three acceptor levels in the forbidden band of germanium. In the case of copper, there are levels at 0·04 eV and 0·32 eV from the valence band, and another at 0·26 eV from the conduction band. In the case of gold, the three acceptor levels lie at 0·15 eV from the valence band, and at 0·20 eV and 0·04 eV from the conduction band. Gold also introduces a donor level at 0·05 eV from the valence band. Schultz finds that the 0·05 eV gold donor level moves away from the valence band at a rate of 0·0028 eV per atomic percent silicon in the range 0 to 11 or 12 atomic percent silicon. At the latter composition, a rather abrupt increase in the rate sets in. The 0·15 gold acceptor level also exhibits a nonlinear composition dependence, as do certain copper and indium impurity levels. According to Schultz, the rapid change of ionization energy with alloy composition which has been found for gold, copper, and indium cannot be due simply to a change of dielectric constant with composition, but rather must involve a specific interaction between the impurity and its environment.

The present state of our knowledge of the extrinsic optical properties of mixed crystals may be summarized as follows: In the case of crystalline phosphor materials, there is a large amount of experimental data available[19]. Most of these data were accumulated in the course of investigations where practical objectives were paramount. As a result of this situation, there exists a vast body of information which is extremely difficult to interpret theoretically, partly for lack of suitable theories, and partly for lack of precise information concerning the composition and structure of the materials in question. If progress in scientific directions is to be made, it seems essential that careful measurements be carried out on highly purified single crystals. This approach is bearing fruit in the case of the Ge–Si alloys[61], and should prove equally valuable for more complicated situations.

6. TRANSPORT PROPERTIES

The phenomenological treatment of the transport of charge and heat in semiconducting alloys is formally identical to the approach in pure crystals. A relaxation time representing the scattering of the appropriate carriers is assumed, and the electric or heat current obtained from the solution of the Boltzmann transport equation. The disordered alloy character introduces effects which must be included in the Boltzmann equation. Perhaps the most important of these is the modification in the relaxation time brought about by the disorder. In those alloy systems which have been studied in detail, the experiments indicate a variation of the carrier relaxation time with the degree of alloying. Under certain conditions, the disorder scattering can become the dominant scattering mechanism.

The transport properties will also be affected by changes in the electronic and vibrational quantum states, whether they are due to a variation of the virtual crystal with composition in a given alloy system, or to the presence of disorder. Unfortunately, it is difficult to separate these two classes of quantum state changes by conventional experiments. The experimental data on the germanium–silicon alloy system can be successfully interpreted in terms of virtual crystal changes alone.

6.1. Virtual Crystal Effects

As noted in Section 3, an alloy can be represented by a virtual crystal, which by definition is strictly periodic. It follows from the definition that an energy band structure can be associated with the virtual crystal, and hence with the alloy itself. It must be borne in mind, however, that the presence of disorder tends to blur the form of the band structure. In spite of this blurring, the virtual crystal band structure remains a good first approximation to the electronic energy level distribution. The object of this section is to present experimental evidence, gathered from diverse studies of transport phenomena, supporting this statement.

Cyclotron resonance absorption measurements[62] have provided detailed information about the band structure of germanium and silicon, particularly the shape of the energy surfaces near the bottom of the conduction band and the top of the valence band. These experiments were carried out at microwave frequencies (in the range 10^{10} to 5×10^{10} c/s) and required samples with long scattering relaxation times τ ($\omega\tau \geqslant 1$) to show appreciable absorption at the resonant magnetic field value. Some crystals from the germanium–silicon alloy system have also been studied[63] by this technique. Surprisingly enough, resonance was observed with as much as 5 per cent silicon in germanium, whereas 10^{-5} per cent of column III or column V impurities usually results in too small a value of τ for resonance to be resolved. The relaxation times in the alloys were in the range 10^{-11} to 4×10^{-11} sec. Studies at $4{\cdot}8 \times 10^{10}$ c/s were limited to alloys of less than 5 per cent composition because of the decreasing relaxation time with increasing alloying.

Alloys of 0·9 and 5·4 per cent silicon in germanium exhibited the band structure of germanium. At the other end of the system, at least 1 per cent germanium in silicon does not change the band structure from that of pure silicon. The

effective masses in the alloys were within experimental error of the values found for the pure crystals. The theoretically predicted changes[34] in the effective masses in germanium lie within the experimental errors of both the germanium and the alloy measurements. Thus, the measurements do not disagree with the predictions of the theory.

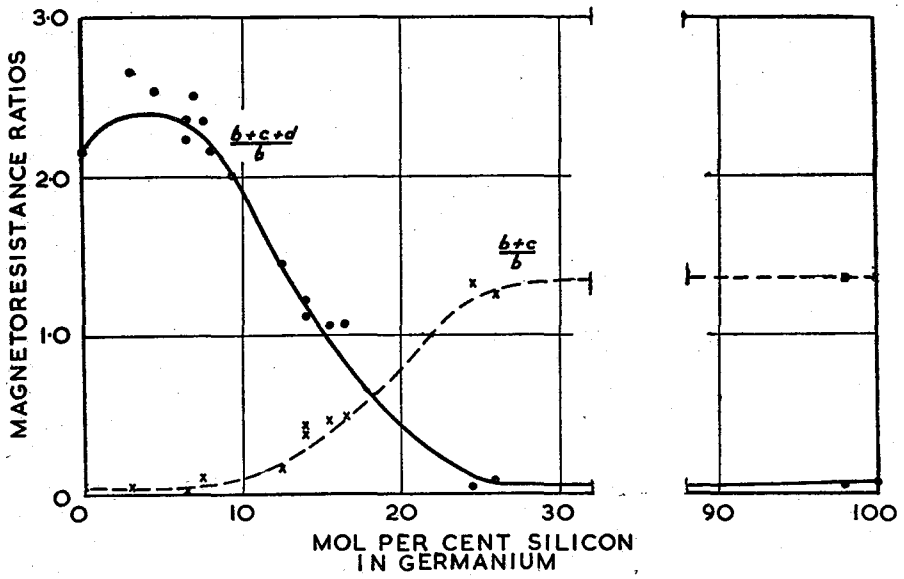


Figure 7. *Magnetoresistance of germanium–silicon alloys as a function of silicon content. For energy surfaces with* [111] *symmetry, the function* $b + c$ *should be zero, while* d *depends directly on the deviation from spherical symmetry. Thus* $b + c \ll b + c + d$ *for* [111] *surfaces, as is the case for germanium. For energy surfaces with* [100] *symmetry, the function* $b + c + d$ *should be zero, with* d *again depending directly on the deviation from spherical symmetry. Thus* $b + c + d \ll b + c$ *for* [100] *surfaces, as is the case for silicon. (Glicksman*[68]*).*

It may be concluded from the cyclotron resonance observations that alloys of silicon with small amounts of germanium have a virtual crystal like the pure silicon crystal; similarly, alloys of germanium with small amounts of silicon have a virtual crystal like the pure germanium crystal. Since these observations do not extend beyond 5 per cent alloy composition, they do not provide information about the behavior at intermediate compositions, in particular at 15 per cent silicon in germanium, where the most interesting effects should occur.

D.C. measurements may also give details of the band structure. Observations of the galvanomagnetic effects in germanium[64] and silicon[65] have been interpreted[65, 66, 67] to yield information on the electronic band structure. The magnetoresistance observed has a symmetry (with respect to the crystalline axes) which depends on the symmetry of the energy–momentum relations obeyed by the electrons. Fortunately, the electrons in germanium (with [111]–symmetry energy surfaces) have a magnetoresistance effect considerably different from that of the electrons in silicon (with [100]–symmetry energy surfaces). The combinations of observed magnetoresistance data[68] which bring out the

differences in symmetry are plotted in Figure 7. There is an obvious change in this symmetry in the region 10–20 per cent silicon in germanium from the [111]–symmetry for the alloys of less than 10 per cent silicon to the [100]–symmetry for alloys of greater than 20 per cent silicon in germanium. In the 10–20 per cent silicon range, the data have been interpreted[69] as due to a sum of contributions from electrons in [111]–oriented energy minima and from electrons in [100]–oriented energy minima. These results substantiate the band structure suggested by Herman[57] for the alloys to explain the optical absorption experiments[17].

A combination of Hall and magnetoresistance measurements can be used to calculate the asymmetry of the energy surfaces[65, 66, 67], provided the relaxation time introduces no additional asymmetry. If this latter assumption is correct, the asymmetry in the effective mass of the alloys' [111] surfaces does not vary outside experimental error for alloys of composition from pure germanium to about 16 per cent silicon in germanium[69]. The observations on alloys of 25 per cent silicon in germanium give an asymmetry[68] in the [100] surfaces which is the same as that found in silicon, within experimental error. The galvanomagnetic observations on the germanium–silicon alloy system show that the lowest conduction band changes, from that found in germanium to that found in silicon, at an alloy composition of about 15 per cent. The character of the bands in the alloys seems quite similar to that found in the pure material.

The information available on other systems is not as extensive as that for the germanium–silicon system. Hall, Hedden and Turner[70] have made use of their measurements of the variation of the conductivity of selenium–tellurium alloys with temperature, to calculate the energy gap as a function of composition. Their first observations yielded an energy gap *versus* composition curve containing both a maximum and a minimum; this result has yet to be verified. Nussbaum[71] has measured the Hall mobility of alloys of up to 13 per cent selenium in tellurium. There is no consistent variation of the data with alloy composition in that range. However, it was found that the Hall mobilities of holes in the alloys varies as $T^{-3/2}$. The complexity and corresponding lack of complete knowledge of the band structures of selenium and tellurium make the data difficult to interpret in terms of a consistent set of virtual crystals for the alloys.

The system HgSe–HgTe has been studied by Nikol'skaya and Regel[72]. Their measurements of the Hall mobility of electrons in polycrystalline samples of the alloys could be fitted by a smooth curve, with an indication of a maximum toward the middle of the alloy range. Again, lack of a frame of reference hampers an interpretation; knowing little about the band structure of the pure components, one can say nothing with confidence about the alloy behavior.

6.2. The Relaxation Time

In alloy systems, the relaxation time differs from that for a pure crystal because of the scattering of the carriers by the disorder present. This problem was considered by Nordheim[30], who calculated the scattering relaxation time τ for such random scattering in a metal alloy. It is instructive to compare this scattering with that caused by the thermal vibrations of the atoms in the lattice. In the alloy disorder case, Nordheim found that:

$$1/\tau = C_1 f(1-f)E^{1/2},$$

where E represents the energy and where C_1 is a temperature-independent factor which includes the strength of the scattering and the effective masses of the electrons.

For thermal vibrations, the relaxation time is given by the following expression:

$$1/\tau = C_2 E^{1/2},$$

where C_2 includes the strength of the scattering and is proportional to the temperature. Since the two relaxation times have the same energy dependence, the mobility, Hall effect, and magnetoresistance can be easily calculated when both must be included. The reciprocal of the carrier mobility in a semiconductor alloy will then be the sum of two terms, one representing the disorder scattering, and the other the thermal lattice scattering. The former will vary with the concentration as $f(1-f)$, and with the temperature as $T^{1/2}$, whereas the latter will be proportional to $T^{3/2}$.

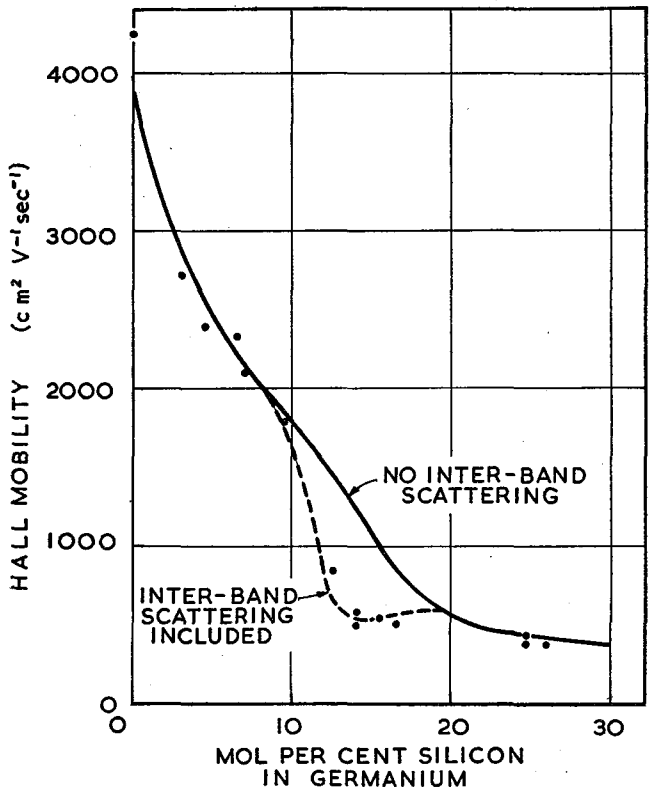
A somewhat different approach has been followed by Brooks[73] in considering the disorder scattering, with application to the germanium–silicon system. The potential function of the alloy is represented in terms of the band energies of the pure elements arranged at random. The scattering relaxation time has a form similar to the one quoted above, and the mobility has the same $T^{-1/2}$ temperature dependence as noted above. Brooks estimated the parameters in his expressions from measured data on germanium and silicon. He predicted that the disorder mobility would equal the thermal mobility at room temperature for an alloy of 8 per cent silicon in germanium.

Levitas[74] has measured the conductivity and Hall effect in the range 100–300° K for P-type germanium–silicon alloys. Unfortunately, most of the samples were polycrystalline, with varying impurity concentrations, but the main features observed are significant. Levitas noted that the mobility decreased with alloying, and varied less strongly with temperature in the alloys (of less than 60 per cent silicon) than in pure germanium or pure silicon. This is the behavior expected if the disorder scattering contributes to reducing the mobility, because the disorder mobility varies less strongly with the temperature than does the thermal scattering.

Some measurements by Glicksman[68] of the room temperature Hall mobility of electrons in germanium–silicon alloys are plotted as a function of alloy composition in Figure 8. The alloy samples studied were N-type single crystals. Observations of the mobility between 80 and 300° K for some of the samples (5–7 per cent silicon) could be interpreted as a sum of contributions from thermal scattering and disorder scattering. At room temperature the disorder contribution was about equal to the thermal vibration contribution for an alloy of 9 per cent silicon in germanium. This result is in good agreement with the values calculated by Brooks[73].

The solid curve in Figure 8 is calculated using the above information to determine the strength of the disorder scattering. The change from the high mobility [111] minima to the lower mobility [100] minima causes the calculated decrease in the 10–20 per cent region. However, the observed mobility in this two-band region is considerably lower than that represented by the solid curve. In order to fit the data, an additional scattering mechanism must be included which lowers

the mobility when the two bands are close together. Inter-band scattering between the [100] family of spheroids and the [111] groups may be the mechanism. This was suggested by Brooks and Paul[75] to explain the great decrease in mobility they observed in measurements of germanium and of germanium–silicon alloys under pressure. The application of pressure is believed to cause the [111] and the [100] minima to approach each other. The results of the calculations including inter-band scattering are shown in Figure 8 as a broken curve, which is a good



[*By courtesy of the Editor, Physical Review*

Figure 8. *Hall mobility of electrons in germanium–silicon alloys as a function of silicon content. The curves are calculated for the two cases labelled; the circles are the experimental points.* (*Glicksman*[69]).

fit to the data. The mobility in the alloys falls off with alloy composition just as is to be expected if due to disorder scattering, at least for the alloys where the conduction is almost entirely in one band.

Mobility measurements by Nussbaum[71] in the selenium-tellurium system showed no consistent variation with alloy composition; in particular, there was no strong decrease evident. It is possible that the failure to see disorder scattering is due to a thermal vibration scattering so large that it entirely masks the disorder scattering.

Folberth[56] has reported values of the Hall mobility in polycrystalline samples of InAs–InP alloys. The mobility drops rapidly from the InAs value of 20,000 $cm^2\ V^{-1}\ sec^{-1}$ to about 8,000 $cm^2\ V^{-1}\ sec^{-1}$ for 20 per cent InP, the latter value being about twice that observed in pure InP. Since the band structures of the individual components are as yet undetermined, the virtual crystal properties are unknown, and definite conclusions cannot be drawn from these data with regard to disorder scattering effects. Measurements on single crystals of higher purity in these systems should prove as enlightening as those made in the simpler alloy system of silicon and germanium.

Very little attention has been directed to the thermal properties of semiconductor alloys. In a review article[76], Joffe and Joffe report results obtained for Ge–Si alloys, Si–Sn alloys, and PbTe–PbSe alloys. In all these cases it is found that alloying reduces the thermal conductivity considerably. Values for the Ge–Si alloys at room temperature given in this article are 0·145 cal cm^{-1} sec^{-1} deg K^{-1} for pure germanium, and 0·008 cal cm^{-1} sec^{-1} deg K^{-1} for a high composition alloy (assumed to have about 50–60 percent silicon content). The authors do not quote the source for these measurements nor the condition of the alloy material. Observations made by Steele[77] on homogeneous polycrystalline alloys also show a rapid decrease of the thermal conductivity with alloying in this system. His measurements give a minimum thermal conductivity at room temperature of about 0·025 cal cm^{-1} sec^{-1} deg K^{-1}, for alloys containing 50–60 percent silicon. The alloys also show a temperature dependence of the thermal conductivity considerably weaker than that found in the pure components. In the case of the PbTe–PbSe alloys[76]. the thermal conductivity is independent of temperature over a considerable range below and above room temperature.

In order to explain these results, Joffe and Joffe assume that the transport of heat energy by the diffusion of phonons is the dominant mechanism governing the thermal conductivity. They then consider the factors determining the mean free paths of the phonons. Because phonons have rather short wavelengths, at least compared with the electrons or holes responsible for electrical conduction, their mean free path is much more sensitive to local deviations from the regular periodicity of the perfect crystal than is the mean free path of electrons or holes. In an alloy, the disorder provides for abundant phonon scattering; the phonon mean free path in a 50 per cent silicon–germanium alloy is of the same order as the atomic spacing. These considerations explain why the thermal conductivity is much more sensitive to alloying than is the electrical conductivity.

7. CONCLUDING REMARKS

Present efforts to interpret the properties of semiconducting alloys lean heavily on the results of the earlier investigations[8, 78] of metallic alloys, especially in the crystallographic domain. From a structural standpoint, semiconducting and metallic alloys have much in common, and the similarities have been used to advantage in recent semiconductor work. For example, it has been presumed that the crystal structure of an intermediate (semiconducting) alloy would be isomorphous with those of the pure constituents; and, further, that the lattice constant in an isomorphous series would vary uniformly with composition.

While other possibilities are conceivable, the above conjectures have invariably been borne out by the experimental investigations.

The simplest approach to the theory of disordered crystals or alloys is the virtual crystal approximation, which, paradoxically, ignores the disorder altogether. According to this approximation, an alloy can be represented by a suitable virtual crystal whose form depends upon the alloy composition. The nature and distribution of the electronic and vibrational quantum states of a given alloy are determined in large measure by the constitution of its virtual crystal counterpart. The experimental evidence gathered to date suggests that the virtual crystal approximation is quite adequate for studying the composition dependence of the energy band structure of alloy systems. Refined theoretical treatments, which specifically take into account the effect of disorder, predict additional effects, such as the 'tailing' of energy bands. Since these specific disorder effects have so far escaped experimental notice, they are apparently very small.

The germanium–silicon alloy system has been the main source of information concerning the variation of the energy band structure in alloys with composition. It has been possible to explore the band structure behavior as a function of composition in a rational manner, and to interpret the optical and galvanomagnetic data in considerable detail. The successful study of the germanium–silicon alloy system owes a great deal to the previous work on the pure constituents. It is probably a general rule that a detailed knowledge of the band structure of the separate components is a prerequisite to the successful exploration of the intermediate alloys.

The disorder in alloys has a number of important consequences, the most significant of these being the effect on the relaxation times of electrons, holes, and phonons. In general, the mean free paths of carriers will be reduced by the random deviations from strict periodicity. The reduction of the electrical and thermal conductivities with alloying has been observed in a number of alloy systems. In the case of the germanium–silicon alloys, the disorder contributes importantly to the scattering of electrons, reducing the electron mobility in the alloys to values well below those in the pure elements. A further reduction in the electron mobility takes place in the neighborhood of 15 per cent silicon, where the conduction band is defined by two groups of energy minima, the [111] and the [100] groups. This further reduction can be explained in terms of inter-band scattering, a mechanism which has often been employed in the interpretation of the conductivity of transition metals and their alloys[77]. As already mentioned, the presence of disorder modifies the electronic energy level distribution, but this effect is small and the ultimate effect on the transport properties is probably smaller still.

There are numerous problems which cannot be solved for the case of concentrated alloys, but which can be readily solved in the limit of infinitely dilute alloys with the aid of standard perturbation methods. For example, in the theory of lattice vibrations, it can be shown that localized structural defects usually perturb most of the normal modes only slightly, and in addition, often give rise to localized normal modes near the defects. These latter are analogous to the bound impurity state of an electron in a semiconductor.

Relatively few semiconductor alloy systems have been studied so far, and the behavior of only one of these, the germanium–silicon system, can be explained

in some detail. While many of the results obtained in these studies bear a close resemblance to the results for metallic alloy systems, some phenomena of common occurrence in metallic alloys have not yet been observed for semiconductor alloys. For example, order–disorder transitions have yet to be seen. The discovery of a semiconducting alloy exhibiting an order–disorder transition should prove very fruitful because, for semiconductors, details of the energy band structure and carrier relaxation time are more directly available than for metals.

While much more information is required before every aspect of alloy systems is understood, the recent studies of semiconductor alloy systems have yielded a good deal of information. When these studies are fully interpreted, they should prove useful in furthering a general understanding of the solid state.

ACKNOWLEDGEMENTS

The authors are indebted to many colleagues at the R.C.A. Laboratories whose penetrating and constructive criticism of this article was of great assistance in its preparation. In particular, thanks are due to Dr. L. S. Nergaard, Dr. J. O. Kessler, Dr. M. C. Steele, Mr. M. A. Lampert, and Dr. S. M. Christian for their contributions.

REFERENCES

1. E. Burstein and P. Egli. Contribution to *Advances in Electronics and Electron Physics*. Vol. VII, p. 1 (Academic Press, New York, 1955); L. Pincherle and J. M. Radcliffe. *Advances in Physics*, **5**, 271 (1956); G. Busch and U. Winkler. *Erg. Exakten Natur.* **29**, 145 (1956); H. Welker. *Erg. Exakten Natur.* **29**, 275 (1956); O. Madelung. Contribution to *Encyclopedia of Physics*. Vol. XX, p. 1 (Springer-Verlag, Berlin, 1957).
2. H. Brooks. Contribution to *Advances in Electronics and Electron Physics*. Vol. VII, p. 87 (Academic Press, New York, 1955).
3. H. Y. Fan. Contribution to *Solid State Physics*. Vol. I, p. 283 (Academic Press, New York, 1955).
4. F. Herman. *Proc. Inst. Radio Engrs., N.Y.*, **43**, 1703 (1955).
5. H. Welker. *Physica.* **20**, 893 (1954); H. P. R. Fredeirkse and R. F. Blunt. *Proc. Inst. Radio Engrs., N.Y.* **43**, 1828 (1955); see also *J. Electron. (Lond.)*, (Sept. 1955); H. Welker and H. Weiss. Contribution to *Solid State Physics*, Vol. III, p. 1 (Academic Press, New York, 1956).
6. F. A. Cunnell, this volume.
7. U. Winkler. *Helv. Phys. Acta.* **28**, 633 (1955); E. Mooser and W. B. Pearson. *Phys. Rev.* **101**, 1608 (1956); E. Mooser and W. B. Pearson, *Can. J. Phys.* **34**, 1369 (1956).
8. For a more detailed discussion of the atomic structure of alloys, see: A. H. Cottrell. *Theoretical Structural Metallurgy*, (Edward Arnold, London, 1948); also A. F. Wells, *Structural Inorganic Chemistry*, Chap. V, XXIV. (Oxford University Press, Oxford, 1950). For a recent review of the theory of alloys, particularly of metallic alloys, see: G. V. Raynor. *Contribution to Progress in Metal Physics.* Vol. 1 (Butterworths, London, 1949), p. 1.
9. I. D. Konozenko. *Uspekhi Fiz. Nauk.* **52**, 561 (1954). This is a survey article on the formation, structure and electrical properties of thin metallic and semiconducting layers. Recent theoretical developments are discussed briefly.
10. See, for example, G. M. Pound, G. Derge and G. Osuch. *J. Metals.* **7**, 481 (1955): [molten Cu–Fe sulphide matter].
11. O. Fürst, R. Glocker and H. Richter. *Z. Naturf.* **4a**, 540 (1949).
12. H. Hendus. *Z. Naturf.* **2a**, 505 (1947).

13. G. Kinchin and R. S. Pease. *Reports on Progress in Physics.* Vol. 18, p. 1 (The Physical Society, London, 1955); J. W. Glen. *Advances in Physics* **4,** 381 (1955); F. Seitz and J. S. Koehler. Contribution to *Solid State Physics,* Vol. II, p. 305 (Academic Press, New York, 1955).
14. F. A. Kröger, H. J. Vink and J. van den Boomgaard. *Z. Phys. Chem.* **203,** 1 (1954); F. A. Kröger and H. J. Vink. Contribution to *Solid State Physics,* Vol. III, p. 310 (Academic Press, New York, 1956).
15. H. Stöhr and W. Klemm. *Z. Anorg. Allgem. Chem.* **241,** 305 (1939).
16. E. Grison. *J. Chem. Phys.* **19,** 1109 (1951).
17. E. R. Johnson, and S. M. Christian. *Phys. Rev.* **95,** 560 (1954).
18. J. A. Amick (private communication).
18a. F. A. Trumbore, *J. Electrochem. Soc.* **103,** 597 (1956).
19. H. W. Leverenz. *An Introduction to Luminescence of Solids* (Wiley, New York, 1950).
20. H. A. Klasens. *J. Electrochem. Soc.* **100,** 72 (1953).
21. S. Asano. *Science of Light (Tokyo).* **4,** 16 (1955).
22. L. W. Strock. *Sylvoinia Tech.* **8,** 71 (1955); L. W. Strock and V. A. Brophy. *Amer. Mineral.* **40,** 94 (1955); F. G. Smith. *Amer. Mineral.* **40,** 658 (1955); H. Jagodzinski. *Acta Cryst.* **2,** 201, 208, 298 (1949) and **7,** 17 and 300 (1954); C. J. Schneer. *Acta Cryst.* **8,** 279 (1955); F. G. Smith and V. G. Hill. *Acta Cryst,* **9,** 821 (1956).
22a. R. W. Cahn, *Advances in Physics* **3,** 363 (1954); P. Haasen, *J. Metals* **209,** 31 (1957).
22b. M. A. Short, E. G. Steward, and T. B. Tomlinson. *Nature, Lond.* **177,** 240 (1956); G. Diemer. *Philips Research Reports* **10,** 194 (1955); A. H. McKeag and E. G. Steward. *J. Electrochem. Soc.* **104,** 41 (1957); P. Zalm. *Philips Research Reports* **11,** 353 and 417 (1956).
22c. K. H. Jack and M. M. Wachtel, private communication, to be published.
22d. H. Reiss, C. S. Fuller and F. J. Morin. *Bell. Syst. Tech. J.* **35,** 535 (1956); H. Reiss. *J. Chem. Phys.* **25,** 400 and 408 (1956); R. J. Hodgkinson. *J. Electron.* **1,** 612 (1956) and **2,** 201 (1956).
23. A. S. Ginzbarg, *Ph.D. Thesis* (Dept. of Physics, Purdue University, 1949); H. M. James and A. S. Ginzbarg. *J. Phys. Chem.* **57,** 840 (1953).
24. R. Landauer and J. C. Helland. *J. Chem. Phys.* **22,** 1655 (1954).
25. D. S. Saxon and R. A. Hutner, *Philips Res. Rep.* **4,** 81 (1949).
26. J. M. Luttinger. *Philips Res. Rep.* **6,** 303 (1951).
27. E. H. Kerner. *Phys. Rev.* **95,** 687 (1954); *Proc. Phys. Soc. Lond.* A**69,** 234 (1956).
28. E. W. Hart. *Phys. Rev.* **95,** 666 (1954); and private communication.
29. J. des Cloizeaux. *J. Phys. Radium.* **16,** 320 (1955).
30. L. Nordheim. *Ann. Phys., Lpz.* **9,** 607, 641 (1931).
31. T. Muto. *Sci. Pap. Inst. Phys. Chem. Res. (Tokyo).* **34,** 377 (1938).
32. R. H. Parmenter. *Phys. Rev.* **97,** 587 (1955).
33. R. H. Parmenter. *Phys. Rev.* **99,** 1759 (1955).
34. R. H. Parmenter. *Phys. Rev.* **104,** 22 (1956).
35. G. W. Castellan and F. Seitz. *Proceedings of the Reading Conference on Semiconducting Materials.* p. 8. (Academic Press, New York, 1951).
36. C. Erginsoy. *Phys. Rev.* **88,** 893 (1952).
37. W. Baltensperger. *Phil. Mag.* **44,** 1355 (1953).
38. F. Stern and R. M. Talley. *Phys. Rev.* **100,** 1638 (1955).
39. P. Aigrain and B. Jancovici (unpublished).
40. P. Aigrain. *Physica.* **20,** 978 (1954).
41. See, for example, J. R. Reitz. Contribution to *Solid State Physics.* Vol. I, p. 1 (Academic Press, New York, 1955).
42. S. Schubin. *Phys. Z. Sowjet.* **5,** 81 (1934).
43. A. I. Gubanow. *J. Exp. Theor. Phys.* **28,** 401 (1955).
43a. N. F. Mott. *Can. J. Phys.* **34,** 1356 (1956).
43b. E. M. Conwell. *Phys. Rev.* **103,** 51 (1956).
43c. D. Pines, E. Abrahams, and P. W. Anderson, unpublished; quoted in discussion of reference 43b.
43d. H. M. James, unpublished, and P. W. Anderson, unpublished; quoted in F. Herman. *J. Phys. Chem. Solids,* January, 1957 issue, in press.
44. F. Herman (to be published).
45. H. M. J. Smith. *Phil. Trans. Soc. (Lond.).* A**241,** 105 (1948).
46. Y. C. Hsieh. *J. Chem. Phys.* **22,** 306 (1954).
47. F. J. Dyson. *Phys. Rev.* **92,** 1331 (1953); see also R. Bellman. *Phys. Rev.* **101,** 19 (1956); [*Note added in proof:* The vibration spectrum of disordered crystal lattices has also been treated analytically by I. M. Lifshitz and his collaborators; see, for example, I. M. Lifshitz. Supplement to *Nuovo Cimento* **3,** 716 (1956); I. M. Lifshitz and G. I. Stepanova. *Soviet Physics J.E.T.P.* **3,** 656 (1956).]

48. K. F. Stripp and J. C. Kirkwood. *J. Chem. Phys.* **22,** 1579 (1954).
49. E. W. Montroll and R. B. Potts. *Phys. Rev.* **100,** 525 (1955).
50. E. W. Montroll and R. B. Potts. *Phys. Rev.* **102,** 72 (1956).
51. See, for example: R. E. Joynson, *Phys. Rev.* **94,** 851 (1954); also E. H. Jacobsen. *Phys. Rev.* **97,** 654 (1954).
52. B. N. Brockhouse and A. T. Stewart. *Phys. Rev.* **100,** 756 (1956); R. S. Carter, D. J. Hughes and H. Palevsky. *Bull. Amer. Phys. Soc.* **1,** 204 (1956).
53. R. Braunstein (private communication, to be published); see also R. Braunstein and S. M. Christian. *Bull. Amer. Phys. Soc.* **1,** 126 (1956).
54. A. R. Moore (private communication, to be published).
55. A. Nussbaum. *Phys. Rev.* **94,** 337 (1954).
56. O. G. Folberth. *Z. Naturf.* **10a,** 502 (1955).
57. F. Herman. *Phys. Rev.* **95,** 847 (1954).
58. F. Herman. *Physica.* **20,** 801 (1954).
59. G. G. Macfarlane and V. Roberts. *Phys. Rev.* **97,** 1714 (1955); **98,** 1865 (1955).
60. W. Kohn and J. M. Luttinger. *Phys. Rev.* **98,** 915 (1955).
61. M. L. Schultz (private communication).
61a. H. H. Woodbury and W. W. Tyler, *Phys. Rev.* 105, 84 (1957).
62. G. Dresselhaus, A. F. Kip and C. Kittel. *Phys. Rev.* **98,** 368 (1955); B. Lax, H. J. Zeiger, R. N. Dexter and E. S. Rosenblum. *Phys. Rev.* **93,** 1418 (1954); R. N. Dexter, H. J. Zeiger and B. Lax. *Phys. Rev.* **95,** 557 (1954).
63. G. Dresselhaus, A. F. Kip, H. Y. Ku, G. Wagoner and S. M. Christian. *Phys. Rev.* **100,** 1218 (1955).
64. G. L. Pearson and H. Suhl. *Phys. Rev.* **83,** 768 (1951).
65. G. L. Pearson and C. Herring. *Physica.* **20,** 975 (1954).
66. B. Abeles and S. Meiboom. *Phys. Rev.* **95,** 31 (1954).
67. M. Shibuya. *Phys. Rev.* **95,** 1385 (1954).
68. M. Glicksman. *Phys. Rev.* **100,** 1146 (1955); M. Glicksman and S. M. Christian *Phys. Rev.* **104,** 1278 (1956).
69. M. Glicksman. *Phys. Rev.* **102,** 1496 (1956).
70. H. H. Hall, D. T. Hedden and T. J. Turner. *Phys. Rev.* **89,** 340 (1953).
71. A. Nussbaum. *Phys. Rev.* **94,** 337 (1954); A. Nussbaum 'Electrical Properties of Pure Te and Te–Se Alloys,' *University of Pennsylvania, Contract AF-33-(616)-78, Tech. Rep. No. I* (Nov. 15, 1953, unpublished).
72. E. I. Nikol'skaya and A. R. Regel. *Z. Tek. Fiz.* **25,** 1347, 1352 (1955).
73. H. Brooks (private communication).
74. A. Levitas. *Phys. Rev.* **99,** 1810 (1955).
75. H. Brooks and W. Paul. *Bull. Amer. Phys. Soc.* **1,** 48 (1956).
76. A. V. Joffe and A. F. Joffe. *Izvestia Akad. Nauk S.S.S.R.* (Physics) **20, 65 (1956);** see also A. F. Joffe, Supplement to *Nuovo Cimento* **3,** 702 (1956); also A. F. Joffe, *Can. J. Phys.* **34,** 1342 (1956).
77. M. C. Steele (Private communication, to be published).
78. For a detailed discussion of the theory of electrical and thermal conductivity in metallic alloys, see H. Jones. Contribution to *Encyclopedia of Physics.* Vol. XIX, p. 227 (Springer-Verlag, Berlin, 1956).

PROPERTIES OF THE III-V COMPOUND SEMICONDUCTORS

F. A. CUNNELL, Ph.D.
and
E. W. SAKER, Ph.D., A.Inst.P.

Services Electronics Research Laboratory, Baldock, Herts., U.K.
MS. Received 29 *September*, 1956

PROPERTIES OF THE III–V COMPOUND SEMICONDUCTORS

1. INTRODUCTION

It is now eight years since semiconductor research was stimulated by the discovery of triode action in germanium, and the properties of germanium and silicon continue to be exploited in many new devices. Although the importance of these two materials remains as yet unrivalled in the semiconductor field, there is a rapidly growing interest in a class of semiconducting compounds formed from a combination of some Group III and Group V elements. This active interest originated principally in a publication by Welker[1] in 1952, in which he outlined their properties and indicated their more obvious potentialities. The present article will be concerned with a discussion of these materials although they constitute only a part of the whole range of compound semiconductors[2].

The nine III–V compounds we shall consider are formed by combination of elements from the two groups of three, aluminium, gallium and indium, and phosphorus, arsenic and antimony, as shown in the portion of the Periodic Table (Table 1). Of these the indium compounds have been the most extensively

Table 1. Part of the Periodic Table.

Period	Group III B	IV B	V B
3	Al	Si	P
4	Ga	Ge	As
	'A' sub-groups: fourth period transition metal		
5	In	Sn	Sb
	'A' sub-groups: fifth period transition metal		

studied (in particular InSb), and aluminium compounds the least. This fact is at least partly explained by the high melting points of the aluminium compounds, which makes their technology more difficult.

All these compounds crystallize in the zinc blende structure similar to the diamond structure characteristic of the Group IV elementary semiconductors. Although the binding is predominantly covalent, an essential difference exists between their properties and those of the Group IV elements. This originates in the difference in the electronegativity of the constituent atoms which introduces an ionic component into the binding forces. In particular, the introduction of an ionic component appears to lower the total energy of the III–V system relative to the corresponding Group IV element (*e.g.* GaAs and Ge, InSb and Sn) with the result that the bonding is strengthened and the melting point raised. The fact that the energy gap is increased relatively more than the

melting point leads to a series of semiconductors with lower melting points for a given energy gap than the Group IV semiconductors. This is important, since a large energy gap is necessary for high-temperature operation of rectifiers and transistors, and a low melting point simplifies the processing of the material.

In addition, the type of bond formed in the III–V compounds has an important effect on the mobility of charge carriers. Although some of the compounds have not been obtained in a sufficiently pure form for the lattice carrier mobility to be assessed, it is apparent from experimental data on InSb, InAs, InP, GaAs, GaSb that large electron mobilities are obtained, coupled with much lower hole mobilities. The large electron mobility is in qualitative agreement with theory which predicts that lattice mobility should be proportional to $m^{*-5/2}$ where the effective mass parameter m^* is small in the case of most of these compounds.

This group of materials covers a range of energy gaps from 0·17 eV for InSb to several eV in the case of AlP. In view of the requirements for rectifiers and transistors capable of relatively high temperature operation the compounds AlSb, GaAs, and InP are of potential importance. The latter two materials may offer an improvement on silicon in frequency behaviour owing to their greater electron mobility. Against these possible advantages one must set the disadvantage of increased difficulty in the preparation and purification of a compound semiconductor consisting of elements of markedly different volatility. The compounds InSb and InAs, although unsuitable as rectifier and transistor materials, possess extraordinarily large electron mobilities giving large Hall and magnetoresistance effects which have been exploited in a number of devices. It seems likely that another of their principal applications will be as long-wave photoresponders, since InSb is sensitive out to 7 μ, a longer wavelength limit than for any other known responder.

It is intended to illustrate briefly in the following paragraphs the present-day state of knowledge of these new materials, the difficulties associated with their preparation and purification and also to give an impression of their ultimate potential importance.

2. PREPARATION AND PURIFICATION OF III–V COMPOUNDS

In general, these compounds are prepared by direct combination of a stoichiometric mixture of the constituent elements at a suitable temperature either *in vacuo* or in an inert atmosphere. The antimonides are the most readily prepared since antimony has a comparatively low vapour pressure by comparison with arsenic and phosphorus. Even at 1080° C, the melting point of AlSb, the vapour pressure of antimony is only of the order of 10 mm of mercury. On the other hand, phosphides and arsenides must be prepared with care if dangerously high pressures are to be avoided at fairly low temperatures relative to their melting points. For example, although the vapour pressure of arsenic at 600° C is less than 1 atm, it rises rapidly above that temperature to approximately 8 atm at 700° C and 26 atm at 800° C. If the method of direct combination of the elements is to be used to prepare arsenides, it is desirable to allow sufficient time to elapse at a temperature below 650° C for appreciable combination to occur. For example, GaAs, (melting point 1250° C) can be prepared safely in

sealed evacuated silica tubes by raising the temperature to 1300° C in a continuous manner over a period of 48 hours. Similar considerations are necessary in the case of phosphides.

As an alternative to forming the compound from a stoichiometric mixture of the components it is possible to crystallize quite large single crystals from a 10–20 per cent solution of phosphorus, arsenic or antimony in aluminium, gallium or indium[3]. This method has the advantage that it is only necessary to heat the solution to a temperature slightly in excess of the solubility curve temperature corresponding to the particular composition. Figure 1 shows phase diagrams of the In–Sb, Ga–As, Ga–Sb, Al–As, Al–Sb, In–As systems. The crystals, which are formed by slow cooling, must be separated either chemically or mechanically from the excess Group III constituent. Single-crystal plates 12 mm long × 1·5 mm thick have been obtained in this way.

The possibility of solid solubility of either constituent in a binary compound semiconductor is an important consideration. PbS is one example of a material in which small changes in the composition, producing large modifications of the electrical properties, are possible. Although phase diagrams for most of the III–V systems are known, little detail is available in the immediate vicinity of the stoichiometric ratio. It is believed, however, that the solid solubility of the constituent elements in these materials may be small. For example, single crystals of InSb have been grown from non-stoichiometric melts of different composition (0·5 per cent either side of the 1:1 ratio) and no detectable change in electrical properties has been observed[4]. The limit of detection in these experiments was determined by the foreign impurity content.

The fact that solid solubility of the component elements does not appear to be favoured in materials of the III–V type is important from a practical point of view, particularly in subsequent refining of a compound in which one element is considerably more volatile than the other. Excess of the less volatile element is easily removed by the method of zone-refining[5] in which a molten region is traversed through the material. The excess is swept ahead of the freezing front and accumulated at one end of the ingot.

Most of the Group III and Group V elements are readily obtainable in a state corresponding to an impurity content of 10–100 p.p.m. Generally, it is worthwhile purifying the components further, prior to the preparation of the compound. This is particularly true in a case where foreign impurity atoms present in the elements are known to have unfavourable segregation coefficients in the compound. A variety of methods has been used to purify the elements. The zone-refining technique has been applied to antimony, aluminium and indium, but it is not suitable in the case of arsenic and phosphorus owing to their great volatility. A repeated sublimation process is favoured for the latter materials. Gallium has been purified by zone-refining of its chloride, followed by recovery of the metal by electrodeposition[6]. An electrolytic process has also been used for indium[7].

Zone-refining is the only method available for purification of the compounds. The results published for several of the III–V semiconductors indicate varying degrees of success. As might be expected, the antimonides are most amenable to this process. InSb, for instance, presents no practical difficulties, although the efficiency with which some of the most commonly occurring impurities are

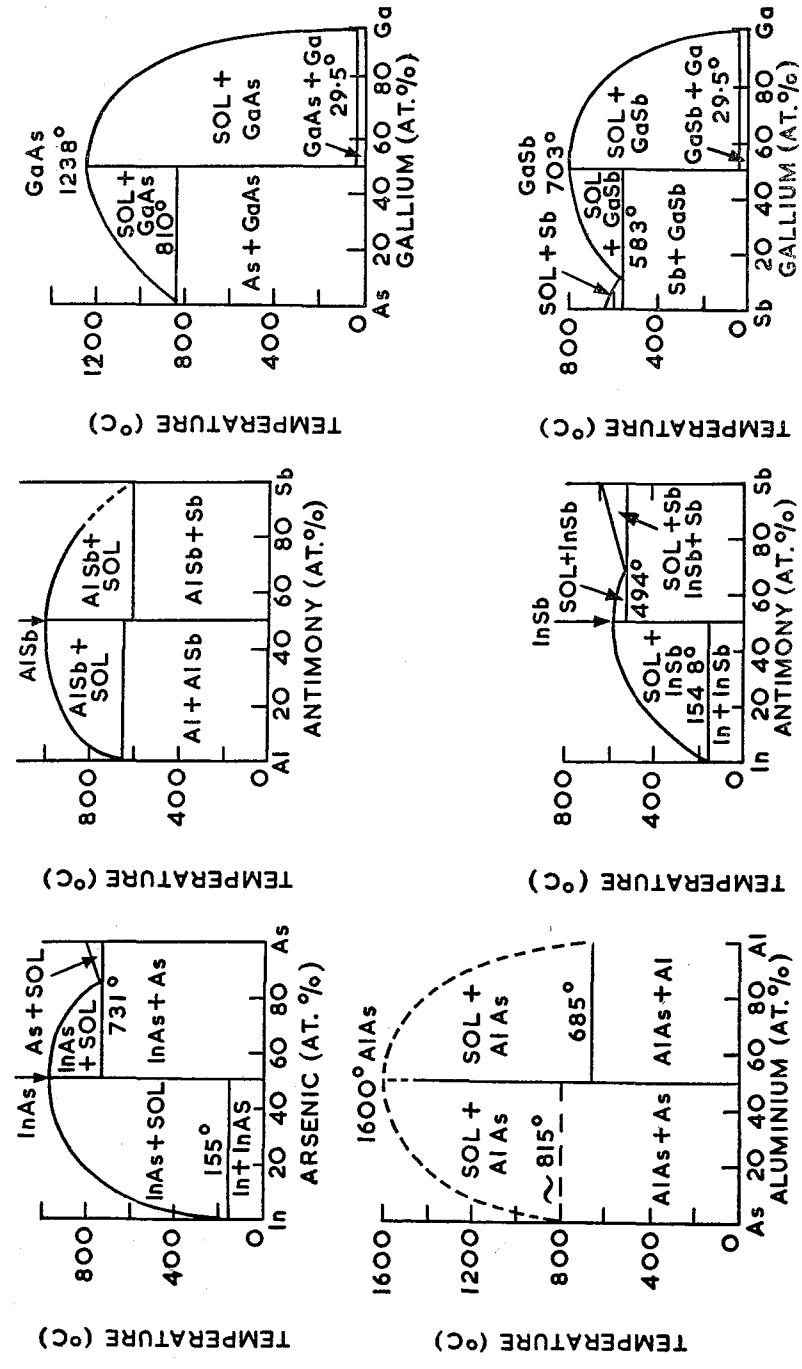


Figure 1. *Phase diagrams for the systems In–As, Al–Sb, Ga–As, Al–As, In–Sb, Ga–Sb (GaSb, AlAs, GaAs after Köster and Thoma; InSb, InAs after Liu and Peretti; AlSb after Guertler and Bergmann)*

swept to the ends of an ingot is relatively small, *e.g.* segregation coefficients for some N-type and P-type impurities have been reported to be approximately 0·6 and 1·2 respectively[8]. Consequently, to obtain a total impurity content of 10^{15} cm^{-3} atoms or less, a series of zone refining and cropping operations may be required.

Figure 2 shows a typical Hall constant variation along an InSb ingot after thirty passes. The P-type impurities accumulate at the first end to freeze,

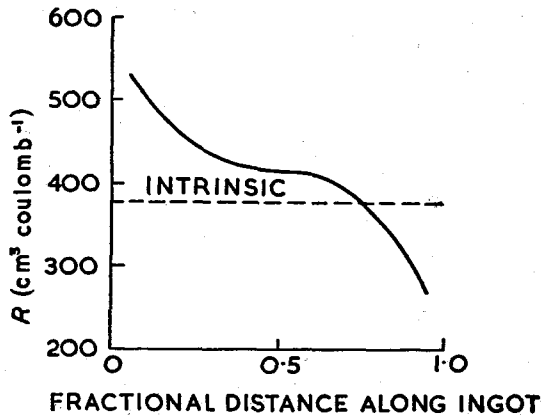


Figure 2. *Hall constant variation along a zone refined InSb bar.*

raising the Hall constant above the intrinsic value, and N-type impurities at the other end. Attempts to obtain very pure GaSb and AlSb by zone-refining have not been successful so far.

The process of refining the arsenides and phosphides is more involved. The method used for GaAs purification may be considered typical of the type of technique required for these materials. GaAs tends to dissociate *in vacuo* above about 850° C. It follows that arsenic will escape from the molten zone and condense on the colder portions of the system. The escape is quite rapid, with the result that an arsenic-deficient zone remains. Clearly the transport of the arsenic away from the zone is a continuous process which ultimately results in complete dissociation of the compound. In order to avoid dissociation, it is essential not only to use a closed system, but also to maintain the entire system at a temperature above the sublimation point of arsenic. A small quantity of arsenic is lost from the molten zone initially, but equilibrium is attained when the arsenic pressure within the tube corresponds to the vapour pressure of arsenic over GaAs at the temperature of the zone. Refinements of this method are possible[9].

Single crystals of all of the antimonides have been prepared by the Kyropoulus method, the technique widely used for silicon and germanium. An inert atmosphere rather than a vacuum is favoured, otherwise the loss of antimony vapour from the crucible may be appreciable. This method is not so easily applied to arsenides and phosphides because of the increased dissociation problem, although it has been successfully applied to GaAs and InAs using an entirely sealed

system[10]. As an alternative, large crystals can be obtained by careful directional freezing of a molten ingot within a sealed silica tube.

The Stockbarger or Bridgman method of growing single crystals is unsuitable for materials which expand appreciably on solidification. All materials possessing the zinc blende or diamond structure have this characteristic.

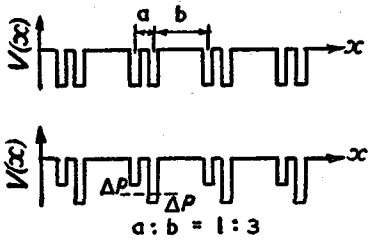


Figure 3. *Simplified potential contours along a* [111] *direction in a crystal of the zinc blende type. (After Seraphin.)*

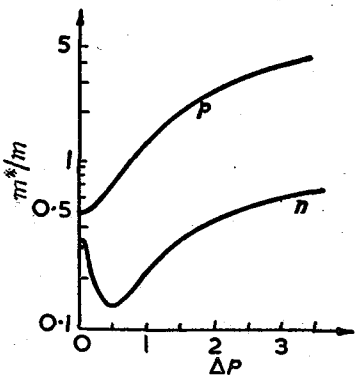


Figure 4. *Electron and hole effective mass variation as a function of the asymmetry of adjacent potential walls. (After Seraphin.)*

3. THE ENERGY BAND SCHEME OF THE III–V COMPOUNDS

No detailed calculations have so far been published for the energy band scheme of any of the III–V compounds. The first contribution to this aspect of the work was made by Seraphin[11]. He considered a simplified one-dimensional model of a diamond-type lattice and obtained a solution to the Schrödinger equation for the system, the implications of which were in agreement with experimental observations. The model he chose consisted of a linear chain of potential wells representing the potential distribution along a [111] direction. With this model, the case of the Group IV element can be transformed into the case of a III–V or a II–VI compound of the same structure by modifying the depths of the wells, as shown in Figure 3.

Seraphin expresses his conclusions in terms of the difference in depth of adjacent wells (defined as $2\Delta P$, where $\Delta P = 0$ for Group IV elements). The results can be summarized as follows. With increasing inequality in the depth of adjacent wells, *i.e.* increasing ionicity:

(1) The conduction band rises slightly.
(2) The valence band is depressed to lower energies implying a stronger binding force. As a result of both this and (1) the forbidden energy gap increases.
(3) The electron effective mass decreases to a minimum value and then increases (see Figure 4).
(4) The hole effective mass increases uniformily.

Since the lattice mobility is proportional to $m^{*-5/2}$, (3) and (4) can equally be expressed as a mobility variation.

These conclusions are in general agreement with fact. It appears that the III–V compounds possess a blend of covalent and ionic bonding, which leads to a minimum effective mass for electrons and consequently a large electron mobility which is at a maximum in InSb.

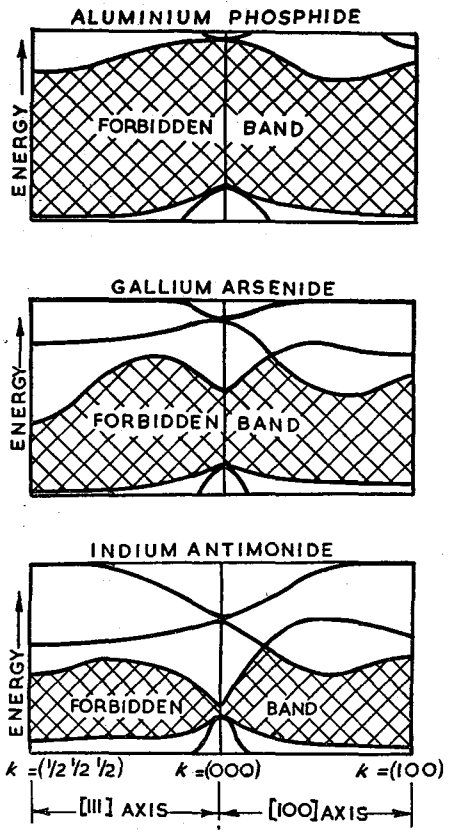


Figure 5. *Hypothetical energy band contours for AlP, GaAs, InSb. (After Herman.)*

A semi-empirical relationship between the degree of ionicity in the atomic binding and the carrier mobility has been established by Goodman[12].

More recently, Herman[13] has made some speculations as to the possible general features of the energy bands in several of the III–V compounds. These speculations were based on qualitative arguments and the existing knowledge of the band structures of germanium and silicon. Figure 5 shows the hypothetical band structure of AlP, GaAs, InSb. Herman believes the schemes for AlAs and GaP would be similar to each other, and intermediate in form between those shown for AlP and GaAs.

Very little experimental information has been obtained to confirm or deny these models suggested by Herman. InSb and InAs have been examined for cyclotron resonance effects, but only in InSb is evidence concerning the band

structure available. The results appear to indicate a spherical symmetry of the conduction band but a more involved valence band[14]. Two resonance lines, both anisotropic have been tentatively identified with holes. Some magneto-resistance data for N-type InSb[15] supports the belief that the conduction band is spherically symmetrical and low-temperature measurements on P-type material[16] have indicated a complicated valence band. It should be pointed out, however, that published results on N-type material to date have been concerned with a conduction band modified by the presence of impurity levels which themselves form a narrow energy band merging with the main set of energy levels. This topic will be dealt with further in the next section. In the following tables some of the properties of the III–V compounds are set down. Table 2 illustrates the trend in properties within a horizontal isoelectronic series, and Table 3 compares the III–V compounds with the corresponding Group IV semiconductors.

Table 2.

	Grey Sn	InSb	CdTe	AgI
ϵ_G (at 0° K)	0·08	0·23	1·8	~ 2·8
μ_n (at 300° K)	~3000	70,000	~600	~30

Table 3.

Material	Melting point (° C)	ϵ_G (at 300° K)	Lattice constant (Å)
Si	1420	1·08	2·34
AlP		3·0	2·36
Ge	958	0·66	2·44
GaAs	1240	1·35	2·44
Grey Sn	232	0·1	2·8
InSb	523	0·17	2·8

In Figure 6, energy gap is plotted against melting point illustrating the point that the III–V compounds possess, in general, a larger energy gap for a given melting point than the Group IV elements. InSb and InAs are exceptions to this rule.

4. GENERAL PROPERTIES OF III–V COMPOUNDS

Although all nine of the compounds mentioned previously have been prepared the greater part of the research in this field has been concentrated on InSb. It follows that an account of the general properties of the III–V compounds must of necessity contain a disproportionate amount of information concerning this one material. It should also be remembered that a detailed comparison of the properties of the compounds is made difficult by the fact that their state of purity is widely dissimilar in certain cases. For the sake of clarity it is intended to deal with each material in turn.

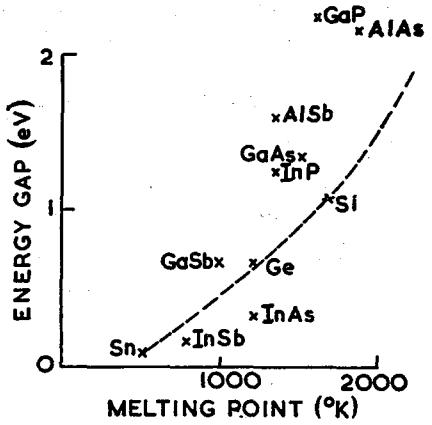


Figure 6. *Energy gap versus melting point for the III–V semiconductors and the Group IV elements.*

Table 4.

	Melting point	ϵ_G (300° K)	μ_n (at 300° K)	μ_p (at 300° K)
AlP		~ 3·0		
AlAs	> 1600	2·16		
AlSb	1080	1·60	~ 100	~ 200
GaP	1340	2·25		
GaAs	1240	1·35	~ 4600	~ 300
GaSb	720	0·67	~ 4000	~ 800
InP	1070	1·25	~ 3400	~ 650
InAs	936	0·33	~ 28,000	> 200
InSb	523	0·17	70,000	~ 800

The mobilities given in this table are only approximate since the proportionality constant which converts $R\sigma$ values to true mobility values, μ, varies from one material to another depending on the scattering process involved and is not always known.

Table 4 summarizes the more important parameters of the III–V compounds.

4.1. Indium Antimonide

The particular attention which has been shown to InSb over the last two years can be attributed to the comparative simplicity of its processing together with its extraordinarily high electron mobility. It has been mentioned in a previous section that a minimum in the effective electron mass, and therefore a maximum in electron mobility, is expected from Seraphin's calculation for the III–V member of an isoelectronic series. Of all the III–V compounds InSb displays this characteristic most strongly, and many of its more interesting properties are directly associated with its very low effective electron mass.

(1) *Carrier effective masses.* Estimates of the effective mass of both electrons

and holes have been obtained in a number of ways. Tanenbaum and Maita[17] deduced $(m_n m_p/m^2)^{1/2}$ from electrical measurements and a knowledge of the optical energy gap at room temperature. Using the expression for the density of charge carriers:

$$np = 4(2\pi mk/h^2)^3 T^3 (m_n m_p/m^2)^{3/2} \exp(-\epsilon_G/kT)$$

and the fact that the lattice mobility ratio $\mu_n/\mu_p = b$ is proportional to $(m_p/m_n)^{5/2}$ it is possible to deduce m_n and m_p separately. The values obtained in this way are $m_n = 0{\cdot}034\ m$ and $m_p = 0{\cdot}2\ m$.

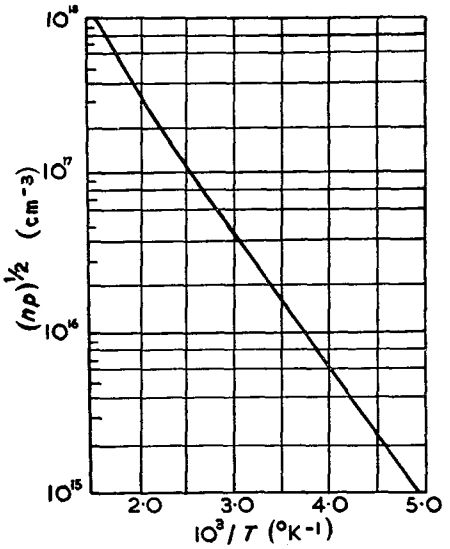


Figure 7. $(np)^{\frac{1}{2}}$ *versus* $10^3/T$ *for InSb.*
(After Hrostowski et al.)

Thermoelectric power measurements lead to similar values[18, 19, 20, 21, 22]. However, these measurements yield only average values of effective mass while the more recent cyclotron resonance experiments[14] are able to resolve effective masses corresponding to carriers in different, though overlapping, bands and show their dependence on crystallographic direction. The evidence so far, suggests an isotropic mass for electrons of $0{\cdot}013\ m$ and two masses for holes $m_p \sim 0{\cdot}18\ m$ and $m_p > 1{\cdot}2\ m$, both of which appear somewhat anisotropic. This latter figure for electrons is rather lower than the figures obtained by other methods but since the cyclotron resonance experiment was done at a temperature in the region of 2° K the discrepancy could possibly be accounted for in terms of a temperature dependence of the effective mass. Some thermoelectric measurements[21] above room temperature seem to be consistent with such a dependence, but a total variation of effective mass from $0{\cdot}05\ m$ near the melting point to $0{\cdot}013\ m$ at liquid helium temperatures appears surprisingly large. Chasmar and Stratton[22] suggest that m_n from cyclotron resonance experiments should be associated with the lowest impurity levels whereas at higher temperatures m_n is characteristic of the conduction band.

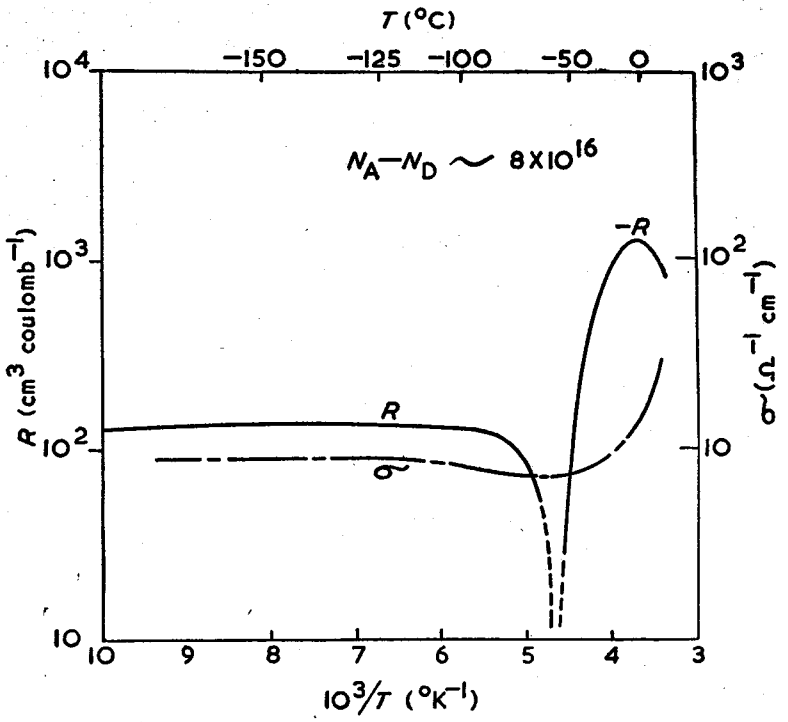


Figure 8. *Hall constant and conductivity versus* $10^3/T$ *for P-type InSb.*

(2) *The energy gap and degeneracy in InSb.* The energy gap in InSb at a temperature T can be expressed as

$$\epsilon_G = 0{\cdot}27 - (3 \times 10^{-4}T) \text{ eV}.$$

This small value implies a large intrinsic carrier concentration at room temperature. Figure 7 shows the variation of n_i with T. It follows that material of reasonable purity (*i.e.* 10^{15}–10^{16} cm^{-3} impurities) is in the intrinsic range at room temperature. Furthermore, the large value of b (approximately 85 at room temperature) can lead to N-type behaviour of the Hall constant in material where $N_A > N_D$ even at fairly low temperatures. Figure 8 illustrates this point for $N_A - N_D \sim 8 \times 10^{16}$cm^{-3}. Hall constant and conductivity measurements in P-type (*i.e.* $N_A > N_D$) material must be interpreted with considerable caution for that reason. In the intrinsic region, the Hall constant expression reduces to:

$$R = -\frac{3\pi}{8} \cdot \frac{1}{n_i e} \cdot \frac{b-1}{b+1}$$

and since $b \gg 1$ the effect of holes can almost be neglected.

The position of the Fermi level in intrinsic material can be expressed as

$$\varphi_i = \tfrac{1}{2}(\epsilon_c + \epsilon_v) + \frac{kT}{2} \ln \left(\frac{m_p}{m_n}\right)^{3/2}$$

Here ϵ_c and ϵ_v represent the energies of an electron at the bottom of the conduction band and the top of the valence band respectively, measured with respect to an electron at infinity. Since $m_p/m_n \sim 6$, the Fermi level lies substantially above the centre of the forbidden energy gap. At room temperature this displacement is 0·04 eV and the Fermi level lies only 2 kT below the bottom

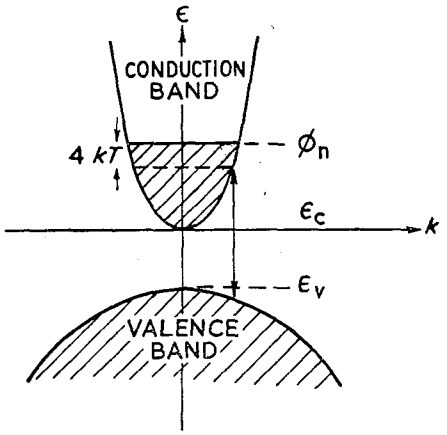


Figure 9. *A simplified energy versus wave vector diagram for degenerate N-type InSb.*

of the conduction band. It follows that intrinsic material even at room temperature is on the fringe of degeneracy and classical statistics are not strictly valid. Degeneracy has the effect of making n_i rise less rapidly and the mobility fall more rapidly with increasing temperature than for the case of non-degenerate semiconductors. Consequently the deviation of the conductivity from classical behaviour is quite pronounced above room temperature.

The movement of the optical absorption edge to shorter wavelengths with increasing electron concentration is just another aspect of degeneracy. This effect has been explained in terms of a small density of states (proportional to $m_n^{1/2}$) in the conduction band arising from the small electron effective mass[23, 24]. At electron concentrations above about 10^{17} cm^{-3} the variation of the position of the absorption edge with n becomes appreciable as the lower available levels in the conduction band become occupied. Figure 9 illustrates the energy level scheme in the case of degenerate N-type InSb showing the optical transition of an electron from the valence band to the lowest unoccupied level in the conduction band arbitrarily placed 4 kT below the Fermi level. The absorption edge measured in pure material corresponds to the energy interval $\epsilon_c - \epsilon_v$ which is in reasonable agreement with thermal activation energy data. A similar effect in P-type InSb is not found owing to the considerably larger effective mass for holes.

(3) *Dependence of Hall constant and magnetoresistance effect on magnetic field strength.* In the simple derivation of the Hall constant and magnetoresistance, weak fields are normally assumed. However, in these calculations the magnetic field parameter appears in association with the mobility and it is the magnitude

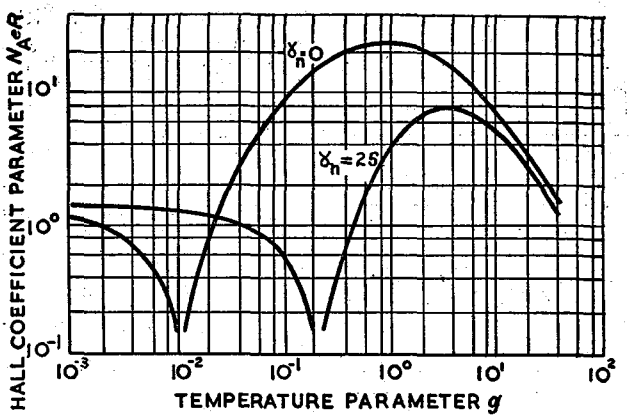


Figure 10. *Hall coefficient versus temperature for P-type InSb under strong and weak magnetic fields. (After Herman et al.)*

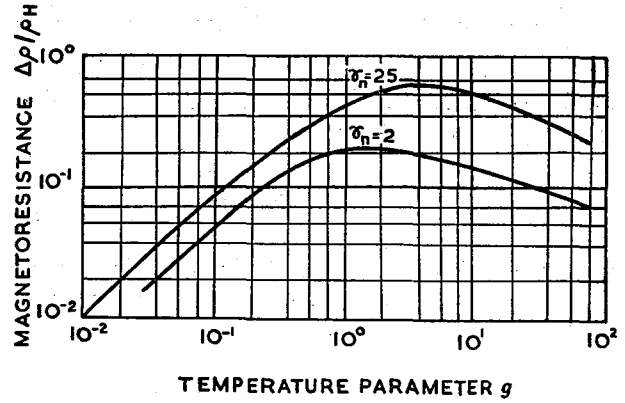


Figure 11. *Magnetoresistance versus temperature for P-type InSb under strong and weak magnetic fields. (After Herman et al.)*

of the product μH which is important. In the case of a material like InSb which possesses very large electron mobility, a divergence from the simple expressions takes place at much lower field strengths than for germanium or silicon. These divergences are most pronounced for material in which holes and electrons make comparable contributions to the conduction process. Figures 10 and 11 show curves of the Hall coefficient and magnetoresistance for P-type material of acceptor density N_A plotted against a temperature parameter:

$$g = bn/p$$

for a mobility ratio of $b = 85$ and different values of

$$\gamma_n = 1{\cdot}77\,\mu_n^2\,H^2$$

($\gamma_n = 25$ and $\gamma_n = 2$ correspond to $H \sim 5300$ oersteds and $H \sim 1500$ oersteds respectively for InSb at room temperature).

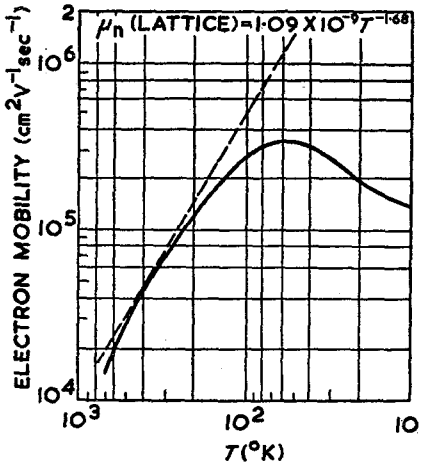


Figure 12. *Electron mobility versus temperature for N-type InSb. (After Hrostowski et al.)*

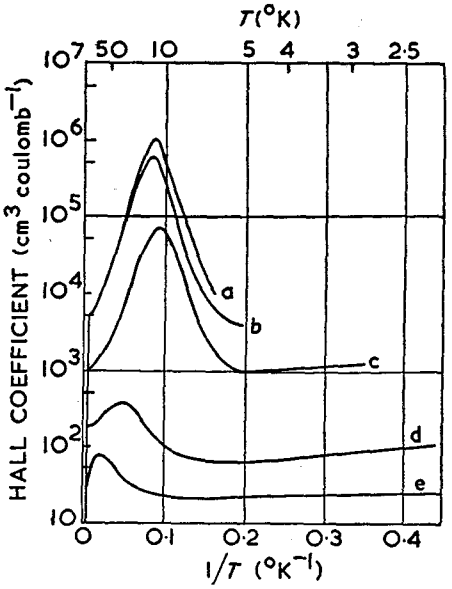


Figure 13. *Hall coefficient versus* $1/T$ *for P-type InSb of various impurity concentrations in the range 2·5° K to 50° K. Total impurity concentration increasing in the order c, a, b, d, e. (After Fritzsche and Lark-Horovitz.)*

These curves are calculated on the assumption of lattice scattering, spherical energy surfaces and classical statistics. Clearly, care must be taken in interpreting Hall constant measurements since the dependence on field is significant. Figure 10 shows that it is possible under certain conditions to obtain a reversal in sign of the Hall constant merely by an increase in H.

(4) *Electron lattice mobility.* As a consequence of the low electron effective mass, the ratio of impurity scattering to lattice scattering at room temperature in InSb is larger than in the case of germanium or silicon for a given impurity concentration.

Impurity scattering of electrons is therefore significant even at a concentration of 10^{15} cm^{-3} impurities. The observed mobilities have increased steadily with improvement in the purity of the material over the last two years and figures in the region of 75,000 cm^2 V^{-1} sec^{-1} for $R\sigma$ at 300° K are frequently quoted in the literature. Figure 12 shows an experimental curve of electron mobility as a function of T. At low temperatures, the mobility is limited by impurity scattering and at high temperatures by lattice scattering. The onset of degeneracy at the higher-temperature end of the curve causes the mobility to fall at a rate which increases with T. The experimental curve has been fitted to a combination of calculated lattice and impurity mobility curves obtained from a knowledge of the impurity content (approximately 10^{15}cm^{-3}) and the electron effective mass. The temperature dependence of the electron lattice mobility can be described by $\mu_L = 1{\cdot}09 \times 10^9\ T^{-1{\cdot}68}$ which leads to a value of 70,000 cm^2 V^{-1} sec^{-1} at

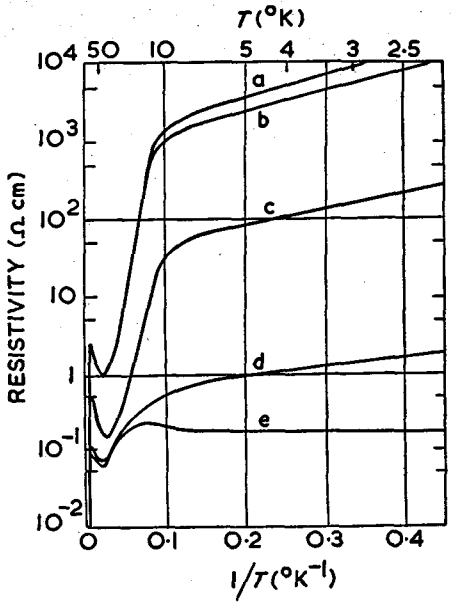


Figure 14. *Resistivity versus* $1/T$ *for P-type InSb of various impurity concentrations in the range* 2·5° *K to* 50° *K.* (*After Fritzsche and Lark-Horovitz.*)

300° K, *i.e.* approximately 83,000 $cm^2\ V^{-1}\ sec^{-1}$ for $R\sigma$. (The figures for $R\sigma$ and $8R\sigma/3\pi$ are often confused in the literature.)

(5) *Low-temperature measurements on InSb and impurity band conduction.* Hall constant and conductivity measurements have been made on InSb down to liquid helium temperatures[26, 27, 28]. For P-type material much the same behaviour has been found as in the case of germanium at low temperatures[29, 30]. Figure 13 shows typical curves for P-type material which illustrate the dependence of the magnitude and position of the Hall constant maximum on the impurity content. The corresponding conductivity curves are shown in Figure 14. These results appear consistent with Hung's model of conduction in an impurity band, although a detailed theoretical explanation of this behaviour has not yet been produced. For N-type material of relatively high purity ($\sim 10^{15} cm^{-3}$ impurities) no de-ionization of impurities is found even at a very low temperature. The implication is that, in material of that purity, the impurity levels overlap the bottom of the conduction band[31]. The extent of this overlap should be a function of impurity content becoming zero below a certain critical value. Hrostowski *et al.*[25] estimate an activation energy for donor centres of 10^{-3} eV. Ionization should be observable at about 2° K for material with $N_D \sim 10^{13}\ cm^{-3}$. The existence of an overlap raises the question as to whether measurements on N-type material up to the present time can be considered characteristic of a true InSb conduction band, and if not, to what extent the band is modified by the impurities. It remains for future work with improved material to answer this question.

Associated with the behaviour of the Hall constant in P-type InSb at low temperatures is the phenomenon of a negative magnetoresistance change[26, 27]. The ratio $\Delta\rho/\rho$ reverses sign below the Hall maximum and then remains negative down to the lowest temperature of the measurements, as shown in Figure 15. There is no satisfactory explanation of this behaviour.

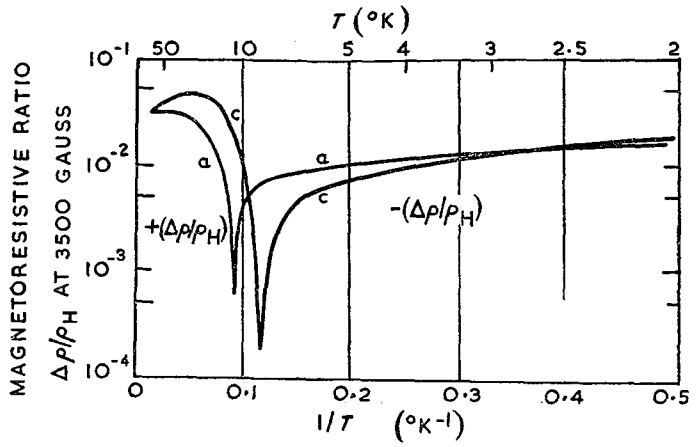


Figure 15. *Magnetoresistance versus* $1/T$ *for P-type InSb in the range* 2·5° *K* to 50° *K. (After Fritzsche and Lark-Horovitz.)*

(6) *Carrier lifetime in InSb.* It is well known that it is possible to disturb the equilibrium concentration of holes and electrons in a semiconductor by injecting minority carriers either from a point-contact or a P–N junction, or by illuminating the material with suitable radiation. The disturbance decays at a rate determined by the concentration of recombination centres (a function of purity and the state of perfection of the lattice) and the probability of direct recombination of holes and electrons. The average lifetime of the carriers is important in some semiconductor applications since together with the diffusion constant it determines the range over which a disturbance may be transmitted. In InSb, the lifetime appears in expressions for the photoconductive and photo-electromagnetic effects both of which may be used for the detection of radiation out to seven microns.

Rough estimates of the photon radiation recombination rate of electrons and holes in intrinsic InSb have been made from a knowledge of the absorption spectrum[32, 33]. A lifetime of the order of 1 microsec is obtained which is of the same order as experimental values derived from the photo-electromagnetic and photoconductive effects and the light 'spot' method. It seems likely, therefore, that in InSb the photon radiation recombination process may be partly responsible for limiting the carrier lifetime. This is in contrast to germanium for which the calculated lifetime in intrinsic material[34] is 0·75 sec, approximately three orders of magnitude larger than maximum observed values.

(7) *Infra red absorption in InSb.* The absorption edge of pure InSb has now been examined experimentally in some detail[35, 36]. It has been shown that the

results can be described by a theoretical expression derived on the assumption that indirect transitions are allowed in which the absorption of a photon is accompanied by the emission or absorption of a phonon. Since cyclotron resonance experiments have shown that the minimum of the electron energy surfaces is in all probability at the centre of the Brillouin zone, the absorption results imply that the maximum of the valence band is not at the centre of the zone. This is in accord with other experiments to which reference has been made previously. Furthermore, the optical measurements are consistent with the existence of two different types of holes in the valence band.

The investigation of the infra-red absorption of InSb has been extended out to 200 microns[37, 38]. A number of lattice absorption bands have been observed indicating some ionic bonding in the crystal. The free electron theory of absorption proposed by Drude and Krönig has been applied to InSb and it is found to be in reasonable agreement with experiment except in the regions of lattice absorption.

4.2. Gallium Arsenide

Little published information is available on this material. However, there is every indication that interest is growing particularly in its application to power rectification and solar energy conversion.

Its fairly high melting point and the tendency to decompose in vacuum above about 850° C makes the purification rather more involved than InSb purification. The technique used has been outlined in a previous section.

Estimates of the energy gap at room temperature have been made from measurement of the position of the absorption edge[39], photoconductive response[40] and recombination radiation[41]. A reasonable value for ϵ_G at a temperature T is given by the expression

$$\epsilon_G = 1{\cdot}52 - (4{\cdot}9 \times 10^{-4}\, T)\ \text{eV}$$

Making use of approximate effective masses of holes and electrons inferred from a series of thermoelectric power measurements on both P- and N-type specimens over a range of temperature[42], the intrinsic concentration of holes and electrons at 300° K can be calculated. For $m_n = 0{\cdot}06\ m$, $m_p = 0{\cdot}5\ m$ and $\epsilon_G = 1{\cdot}37$ eV at 300° K

$$n_i^2 \sim 6 \times 10^{12}$$

Clearly, for all practical purposes the minority carrier will be of completely negligible concentration.

Figures 16, 17, 18, 19, 20 show the temperature dependence of R, of σ, and of the product $R\sigma$ over a large temperature range for both N- and P-type samples. It can be seen from these results that the room temperature values of $R\sigma$ for electrons is higher than the accepted figure for silicon, and comparable with the figure for germanium[42, 43]. It is to be expected that further purification will lead to a higher value. In fact, Welker[44] has predicted a lattice mobility greater than 6800 $cm^2\ V^{-1}\ sec^{-1}$. The anomalous reversible rise in the Hall constant at high temperatures is not yet explained; nor is the unusual temperature dependence of the electron mobility above room temperature.

The highest room temperature hole mobility (*i.e.* $R\sigma$) reported[42] is 340

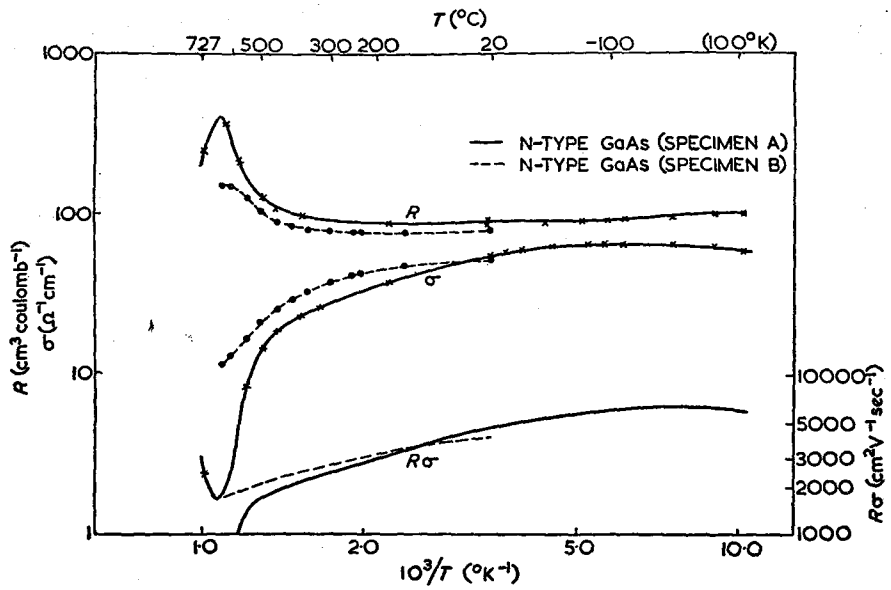


Figure 16. *Hall constant, R, conductivity, σ, and Rσ versus $10^3/T$ for N-type GaAs in the temperature range 90° K to 1000° K. (After Edmond et al.)*

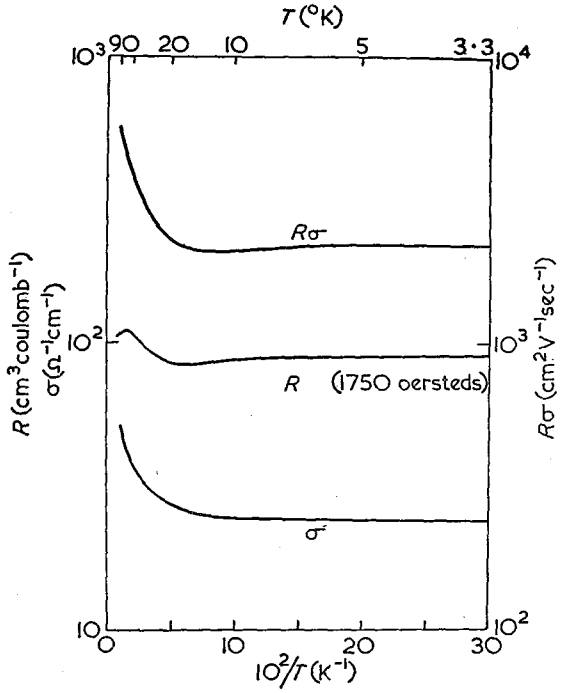


Figure 17. *Hall constant, R, conductivity, σ, and Rσ versus $10^2/T$ for N-type GaAs in the temperature range 3° K to 90° K. (After Edmond et al.)*

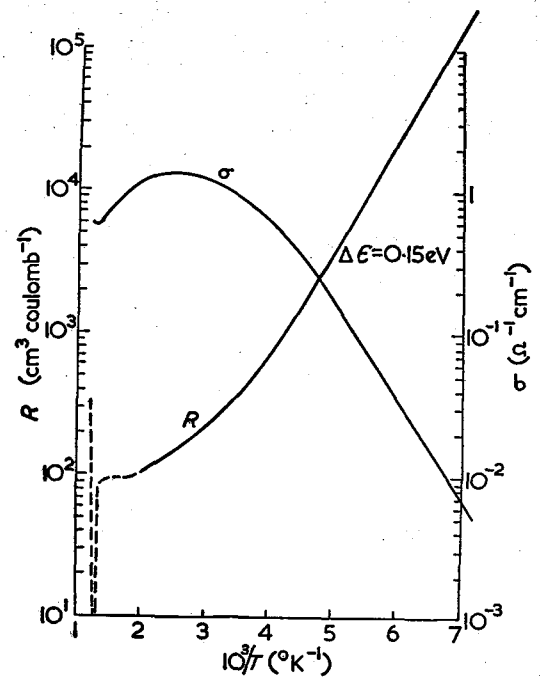


Figure 18. *Hall constant and conductivity versus* $10^3/T$ *for P-type GaAs in the temperature range* 140° *K to* 500° *K. (After Edmond et al.)*

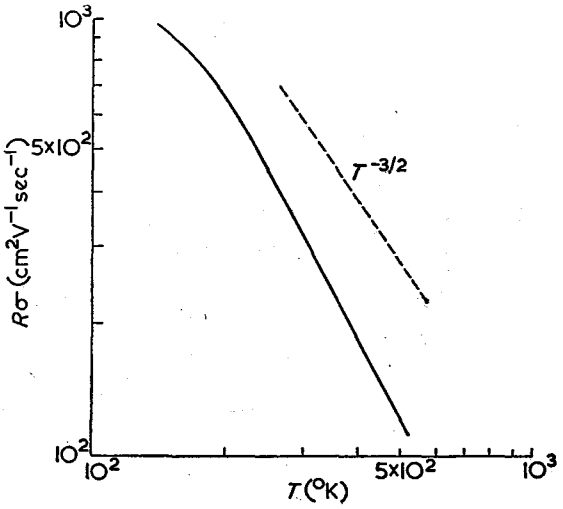


Figure 19. *Rσ versus T for P-type GaAs in the temperature range* 140° *K to* 500° *K. (After Edmond et al.)*

$cm^2\ V^{-1}\ sec^{-1}$. Consequently a lattice mobility ratio of between ten and twenty seems likely. Figure 20 refers to a P-type specimen of GaAs which exhibits the characteristic behaviour at low temperatures attributed to conduction in an impurity band. A similar though less obvious behaviour has been reported for N-type material[42]. Figure 17 illustrates the slight hump in the Hall constant curve attributed to this effect.

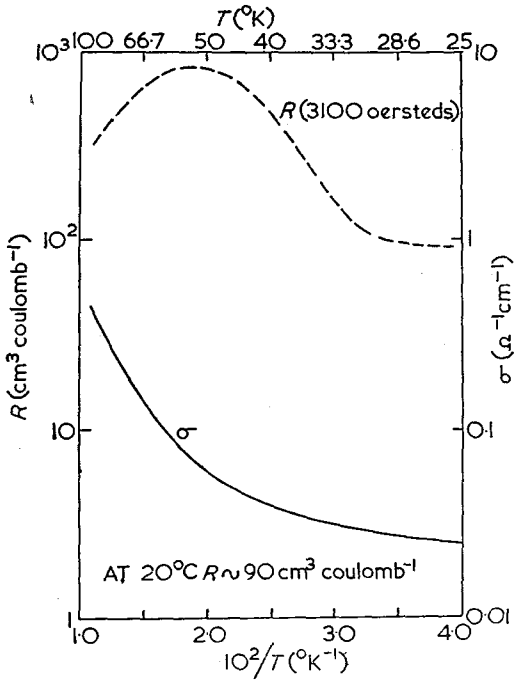


Figure 20. *Hall constant and conductivity versus* $10^2/T$ *for P-type GaAs in the temperature range* 25° *K to* 90° *K. (After Edmond et al.)*

The optical absorption coefficient of GaAs is shown as a function of wavelength in Figure 21. The absorption edge falls on the short-wave side of the silicon edge. Consequently P-N junctions formed in this material are potentially more efficient than silicon junctions in the conversion of solar energy to electrical energy[45, 46]. A GaAs solar battery[47] has been made at the Bell Telephone Laboratories which has given a performance comparable with its silicon counterpart, despite the fact that the design was not optimum. Similar work has been reported by Gremmelmaier[48].

Turnover voltages of between 10 V and 50 V are common for point-contacts on relatively impure GaAs. Diffused P–N junctions have been made in this material using cadmium as the diffusing acceptor impurity[47]. However, improved material is necessary before full advantage may be taken of its favourable energy gap and electron mobility for rectification applications.

4.3. Gallium Antimonide

Attempts to produce this material in a state of high purity by zone refining have so far met with only limited success. In general, a limiting carrier concentration is achieved ($\sim 10^{17}$ cm^{-3} holes) which has been explained in terms of a departure from the stoichiometric composition due to preferential antimony evaporation rather than by the presence of foreign impurity atoms[49]. N-type material is normally produced by doping with a Group VI element.

The energy gap has been estimated from thermal data as approximately 0·8 eV at absolute zero[50, 51]. Optical absorption measurements indicate a room temperature energy gap of approximately 0·7 eV and imply a temperature coefficient of about $3{\cdot}0 \times 10^{-4}$ eV deg^{-1}. A value for $(m_n m_p/m^2)^{1/2}$ of 0·28 is obtained from these results[50]. A rough estimate of the mobility ratio from data at the Hall effect inversion temperature has been made giving $b \sim 5$. Using this value to deduce the ratio of the two carrier masses (assuming lattice scattering for both types of carriers) the separate masses $m_n \sim 0{\cdot}2\,m$ and $m_p \sim 0{\cdot}39\,m$ have been calculated. These figures give $n_i \sim 2{\cdot}7 \times 10^{12}$ cm^{-3} at 300° K. Reported room temperature mobilities tend to cover a considerable range but the highest values so far are approaching 4000 cm^2 V^{-1} sec^{-1} for electrons, and 850 cm^2 V^{-1} sec^{-1} for holes. The temperature dependence of mobility at high temperatures obeys a $T^{-3/2}$ law for holes. There is a stronger dependence for electrons[49]. Some P–N junctions have been grown from the melt by doping P-type material with tellurium. From the optical and electrical properties of such a junction it has been deduced that the lifetime of minority carriers in this material is probably very short[49].

4.4. Aluminium Antimonide

The intrinsic energy gap of this material is approximately 1·6 eV. For any realizable impurity content the intrinsic carrier concentration should therefore be negligible. It follows that the properties of the material should be only weakly temperature-dependent up to higher temperatures when compared with silicon, provided that the impurity activation energy is small against kT at room temperature. It is principally for this reason that it has been considered as a potential rectifier material. In addition, the energy gap lies within the range most suited to solar energy conversion[45, 46].

From the practical and economic point of view, the relatively low melting point, the comparative simplicity of growing single crystals by the conventional Kyropoulus method and the small cost of the component elements are all in its favour. It has, however, the disadvantage that it is attacked by moisture in the air, although its resistance to this action is said to decrease as the purity of the material is improved.

It is difficult to assess the future importance of AlSb at this stage since little detailed information has been published. Blunt *et al.*[52] report no improvement in the material with zone refining. This they attribute to the predominance of lattice defects over foreign impurities as a source of carriers. A similar situation occurs with GaSb and may be due to preferential evaporation of antimony.

Rectification at AlSb–metal contacts has been studied by several workers[53, 54, 55], although not with material of very high purity. Resistivity and Hall constant data

indicate comparable electron and hole mobilities in the region of 100 $cm^2\ V^{-1}\ sec^{-1}$ at room temperature[52]. An effective mass of electrons of $(1{\cdot}8 \pm 0{\cdot}8)\ m$ has been estimated from thermoelectric power measurements[54].

The absorption spectrum of both P-type and N-type material shows a characteristic absorption edge corresponding to 1·6 eV, approximately. This is illustrated in Figure 21. For P-type and N-type material, additional absorption

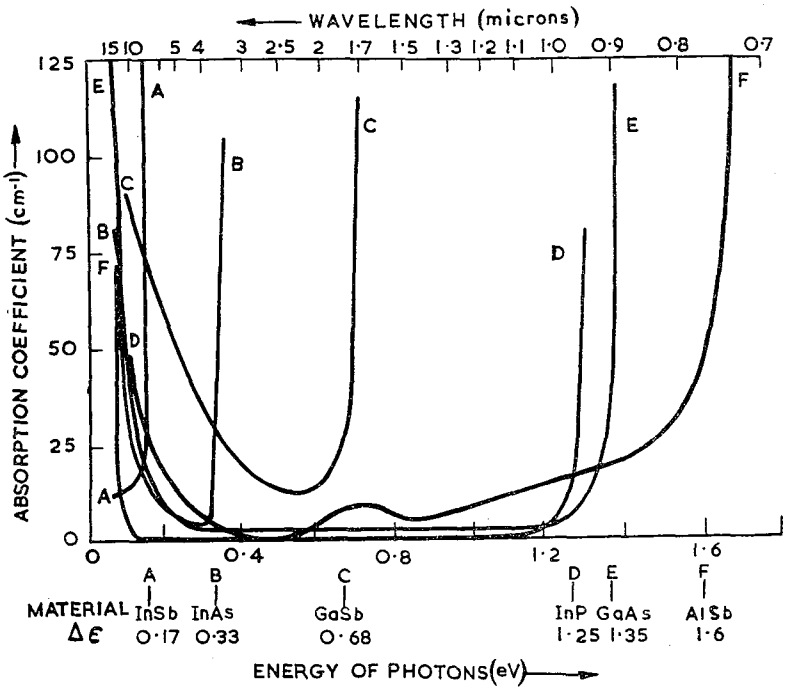


Figure 21. *Optical absorption versus wavelength for the III–V compounds. (After Welker.)*

at 1·6 microns and 4 microns respectively has been found[52]. These absorption bands can presumably be attributed to the presence of impurity levels in the forbidden energy gap. The change in the character of the absorption in converting P-type to N-type material can be associated with a redistribution of electrons over the impurity levels.

4.5. Indium Arsenide

Although InAs is a very interesting substance, possessing an electron mobility which is only exceeded by that of indium antimonide, much less work has been done on InAs than on the latter substance. To some extent this is due to difficulties in purification, since InAs does not zone refine as easily as InSb.

The preparation of the material by direct combination presents no difficulties apart from those discussed earlier. The phase diagram has been investigated by Liu and Peretti[56], who find a line phase corresponding to InAs with lattice constant of 6·058 Å and melting point of 942° C.

The electrical properties of InAs have been investigated by Folberth, Madelung and Weiss[57]. The maximum value of $R\sigma$ measured at room temperature is about 28,000 $cm^2\ V^{-1}\ sec^{-1}$, although, by extrapolation from high temperatures, values of about 35,000 $cm^2\ V^{-1}\ sec^{-1}$ would be expected. The hole mobility at room temperature is of the order of 200 $cm^2\ V^{-1}\ sec^{-1}$, giving a mobility ratio of greater than to 120:1. This is even higher than in InSb.

The energy gap obtained from resistance *versus* temperature measurements is 0·47 eV, whereas from optical measurements on pure material at room temperature a value of 0·33 eV is obtained[58, 59]. Hence the temperature coefficient is obtained as $\alpha = -4{\cdot}5 \times 10^{-4}$ eV deg^{-1}. The position of the absorption edge is a function of the impurity content, as in indium antimonide.

Few values for the effective mass in InAs have been reported. From cyclotron resonance experiments at low temperatures Kittel[60] obtains a tentative value between 0·02 m and 0·03 m for m_n.

Chasmar and Stratton[22] from measurements of thermo-power and Hall effect find that m_n increases from 0·04 m to 0·055 m in the temperature range 80° C to 180° C, thereafter remaining constant at 0·055 m up to 400° C. Folberth *et al.*[57] give a value of 0·1 for $(m_n m_p/m^2)^{1/2}$ implying $m_p = 0{\cdot}3\ m$ if m_n is taken as 0·03 m at room temperature.

Talley and Enright have found[61] that P–N junctions cut from a polycrystalline ingot show photo-voltaic effects, with maximum response at 3·5 microns and time constant of 7 microsec. Photoconductivity and the PEM effect at room temperature have also been observed[62].

4.6. Indium Phosphide

This compound was reported by Welker in an early publication[63] to show rectification and possible transistor action, but very little work has been reported on the material to date. Folberth and Weiss[43] give electron and hole mobilities $R\sigma$ at room temperature as 4000 $cm^2\ V^{-1}\ sec^{-1}$ and 60 $cm^2\ V^{-1}\ sec^{-1}$ respectively, although higher hole mobilities have been reported elsewhere[64]. Measurements of electron mobility at higher temperatures show a $T^{-3/2}$ dependence, and by extrapolation of these results a room temperature mobility of nearly 6000 $cm^2\ V^{-1}\ sec^{-1}$ would be expected. Folberth and Weiss give the following relation for the intrinsic number of carriers as a function of temperature:

$$n_i^2 = 7 \times 10^{31} T^3 \exp(-\epsilon_{GO}/kT)$$

where ϵ_{GO}, the thermal activation energy, has the value 1·34 eV.

The optical properties of InP have been measured by Oswald[65, 66]. The refractive index on the long-wave side of the absorption edge is approximately 3·1, and the optical activation energy is given by the relation:

$$\epsilon_G = 1{\cdot}41 - (4{\cdot}6 \times 10^{-4}\ T)\ \text{eV}$$

Indium phosphide shows remarkably high transparency in the infra-red beyond 1 micron extending as far as a lattice absorption commencing at 15 microns.

By utilizing the above optical value for the energy gap together with the expression given previously for n_i, Folberth and Weiss give a tentative value for $(m_n m_p/m^2)^{1/2}$ of 0·23.

Although InP shows rectifier action, no experiments have been reported

confirming Welker's initial claim of transistor action. Jenny[67] shows that an effect similar to transistor action can be produced by resistive coupling between emitter and collector probes. This does not mean that InP will not show transistor action when sufficiently pure, but indicates that a more stringent criterion, *i.e.* power gain between emitter and collector, must be applied.

4.7. Aluminium Phosphide, Aluminium Arsenide and Gallium Phosphide

The intrinsic energy gaps of these three materials have been reported to be 3·1, 2·16, and 2·39 eV, respectively[68]. Since melting points are correspondingly high the preparation of crystals is difficult and this probably accounts for the scarcity of information on these compounds.

There have been several publications concerning the properties of GaP, which has the lowest melting point of the three (1340° C). Absorption measurements have given the temperature variation of energy gap as $2{\cdot}4-(5{\cdot}5 \times 10^{-4} T)$ eV, and values of the refractive index in the visible and infra-red regions have been found to be 3·37 and 2·9 respectively[66, 69]. The phenomenon of electroluminescence in single crystals of GaP has also been investigated[68].

Mixed crystals of the form $Ga_xAs_yP_{x-y}$ can be grown with a continuous variation of energy gap between that of GaAs and GaP[70]. Similar work has been described for the system $In_xAs_yP_{x-y}$[70, 71].

4.8. Applications of the III–V Compounds

As mentioned in sub-section 2, impetus was given to the early investigations of the III–V compounds by the hope that these materials, possessing large energy gaps and high mobilities together with comparatively low melting points would be useful in the transistor field. Although several of the compounds show good rectification (*e.g.* AlSb, GaAs, InP), there are at the moment no clear-cut reports of transistor action (*i.e.* evidence of power gain) in any of the compounds. This is due at least partly to the fact that the materials as prepared at the moment are comparatively impure, and a great deal of further effort will have to be given to their purification before their future potentialities became clear.

However, it is evident that certain of the compounds will be useful in other spheres. InSb in particular promises to be an extremely useful material, largely on account of two features: the small energy gap, and the high electron mobility. Coupled with these attributes is the fact that InSb is easy to prepare and purify, and is one of the cheaper of the compounds.

The small optical energy gap in pure InSb gives rise to a photoresponse which extends to beyond 7 microns. This means that the material will respond to radiation from objects only a few degrees above the temperature of the cell. Current noise in the material appears to be very low, so that the sensitivity is limited by Johnson noise, and for a cell of area 2 mm^2 and 1 c/s bandwidth an equivalent noise input of 10^{-9} W can be obtained.

The photoresponse may be utilized in several ways. Avery *et al.*[72] have investigated the bulk photoconductivity, Mitchell *et al.*[73] have produced a P–N junction photovoltaic cell, and several workers[74, 75, 76] have investigated

the photo-electromagnetic effect. The latter effect has a theoretical sensitivity comparable with the other effects, and is easier to work with under practical conditions. Some uses of the infra-red sensitivity of InSb are in infra-red spectrometry, pyrometry of low-temperature objects, gas analysis, *etc.* The short response time of less than 1 microsec makes amplification far easier than in the case of conventional bolometers having a response time of the order of 0·1 sec.

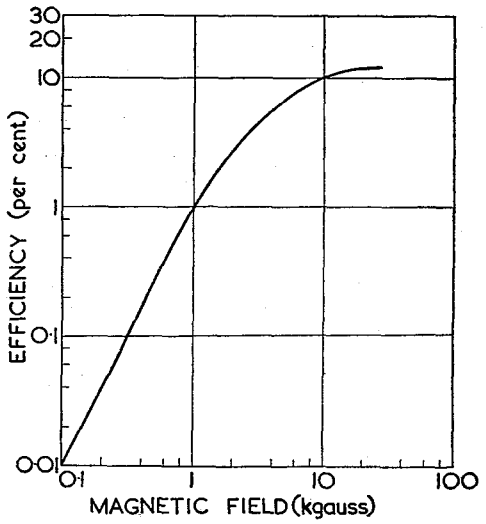


Figure 22. *Efficiency of typical InSb Hall generator as a function of magnetic field strength.*

The high electron mobility in InSb gives rise to very large Hall and magneto-resistance effects. If the Hall voltage created in a rectangular specimen by the application of a magnetic field is fed to a matched load, the efficiency of this Hall generator may be calculated in terms of the power supplied to the specimen, and it is easily shown[77] that in small fields the efficiency is proportional to $(\mu H)^2$ where μ is the carrier mobility in the material. The Hall constant R is eliminated during the calculation and does not apppear. Hence InSb, having an electron mobility of 70,000 $cm^2\ V^{-1}\ sec^{-1}$ has an efficiency approximately 300 times that of germanium (electron mobility 3,900 $cm^2\ V^{-1}\ sec^{-1}$). Since the mobility ratio is very high in InSb, the presence of the holes may be ignored if the material is intrinsic. In fields of more than about 1 kgauss, the transfer impedance between the input and output of the Hall generator must be taken into account and the above expression no longer holds, but nevertheless a high mobility is essential for good efficiency. It has been shown theoretically[78, 79] that for infinitely high fields and mobilities, roughly 18 per cent of the power supplied to a Hall specimen could be transferred to a suitable load. In practice, a transfer of about 12 per cent can be achieved. Figure 22 shows efficiency of a typical Hall specimen as a function of applied magnetic field.

The first practical application of the Hall effect in InSb was to the measurement of magnetic field[77]. Since then, many other applications have been

envisaged[80, 81]. They may be conveniently divided into devices which utilize (1) the sensitivity of the Hall effect in InSb to low magnetic fields, (2) the multiplying property, (3) the non-reciprocal property or (4) the switching property of the Hall effect. In the first category come magnetometers for high or low fields, a simple electrical compass very suitable for automatic course

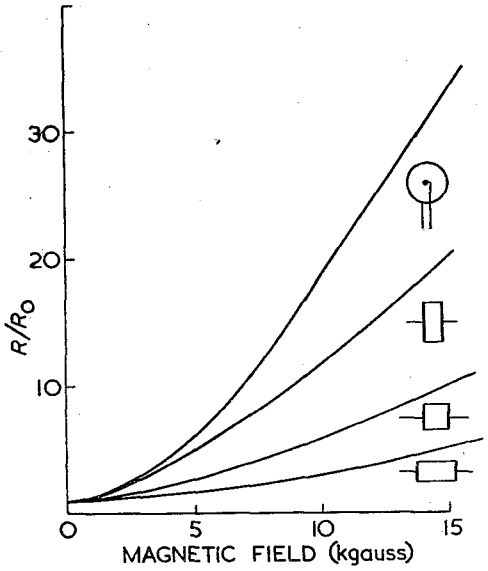


Figure 23. *Magnetoresistance as a function of magnetic field strength for specimens of various geometries.*

control, and ammeters in which the magnetic field created by the current to be measured is used to give a Hall voltage, either directly or (for smaller currents) by utilizing a nearly closed core. Also in this group comes the Hall amplifier[82], in which the flux produced in a nearly closed core by a signal winding gives rise to a Hall output power of greater magnitude than the signal power. Other possible devices utilizing the sensitivity to low magnetic fields are magnetic tape pick-up heads with response down to zero frequency, units for detecting cracks in ferromagnetic materials, and susceptibility bridges. The second group of applications utilizes the fact that the Hall voltage is proportional to the product of the current through the specimen and the magnetic field acting. If one signal provides the current through the specimen, and the other signal energizes a coil providing the magnetic field, the Hall output is proportional to the product of the two quantities. This property is used in the construction of wattmeters, modulators for transforming small D.C. signals into A.C. to facilitate amplification, and in multipliers for calculating machines. In the third group, utilizing the non-reciprocal property of the Hall effect, is the gyrator, a passive device which can be arranged to pass A.C. signals in one direction only and which operates down to zero frequency. Devices utilizing the switching property of the Hall effect depend on the fact that application of a fairly large (several kgauss)

magnetic field can cause appreciable amounts of power (typically 0·1 W) to be switched to a load connected to the Hall terminals. Such a switch is noiseless and has no moving parts. Examples being considered are for automatic train control and for counting of ferromagnetic objects.

The magnetoresistive effect, which is also a function of the carrier mobility, is much larger in InSb than in any other known substance. The increase in resistance obtained in a magnetic field is, however, a function of the geometry of the specimen. The deflection of the current carriers by the magnetic field is normally opposed by the Hall field. This can be short-circuited either by connecting together the Hall electrodes or, more simply, by using specimens in which the current electrodes themselves are used to short the Hall field. Thus, for large magnetoresistance changes, specimens in which the length is short compared with the width are used. The limiting case giving the greatest magnetoresistive effect is the Corbino disc, in which the current flows radially, and in which no potential gradient can be built up in a direction at right angles to a radius. However, nearly as great effects can be obtained in rectangular specimens, with greater simplicity in design. Figure 23 shows magnetoresistive effects in various shapes of specimens.

The magnetoresistance effect can equally well be used for many of the applications outlined above[80, 83]. Points to note are that sensitivity is low in small fields, the effect is not normally dependent on the sense of the magnetic field, and is not inherently balanced so that a bridge arrangement must be used for null output. However, in fields greater than about 1 kgauss the magnetoresistive effect can give greater output than the Hall effect. A particular use of the magnetoresistive effect is likely to be in the form of transducers, in which the specimen is mounted in a region of steep magnetic gradient[86]. Microphones, strain gauges, thickness gauges, *etc.* may be constructed. The sensitivity should be very high. Other uses of the magnetoresistive effect are in control circuits. If a magnetoresistance specimen is connected in series with an iron-cored coil of which the magnetic field acts on the specimen, the effect is to produce a non-linear resistance which increases with increasing current. A magnetoresistance amplifier which works on lines somewhat similar to the Hall amplifier can be constructed[84].

Uses for other of the III–V compounds are not at the moment clearly defined. Indium arsenide, having the second highest mobility of the group, shows Hall and magnetoresistance effects which are only exceeded by those of InSb. It has a larger energy gap and consequently a higher resistance and could therefore be made less subject to temperature effects than InSb.

The use of certain of the III–V compounds as solar energy converters materials has already been mentioned. Although little more than very preliminary work has been done on this subject, it is clear that InP, GaAs and AlSb are capable of providing greater efficiency than silicon[85]. The final choice between these materials will probably be decided by technological considerations.

ACKNOWLEDGEMENT

Our thanks are due to Dr. O. Simpson for his interest and criticism, and also to the Admiralty for permission to publish.

REFERENCES

1. H. Welker. *Z. Naturf.* **7a,** 744 (1952).
2. E. W. Saker and F. A. Cunnell. *Research.* **7,** 114 (1954).
3. G. Wolff, P. H. Keck and J. D. Broder. *Phys. Rev.* **94,** 753 (1954).
4. F. A. Cunnell and H. B. Clarke (Unpublished work at S.E.R.L., 1955).
5. W. G. Pfann. *J. Metals.* **4,** 747 (1952).
6. J. L. Richards. *Nature, Lond.* **177,** 182 (1956).
7. T. C. Harman. *J. Electrochem. Soc.* **103,** 128 (1956).
8. S. W. Kurnick, A. E. Goldberg, G. M. Mitchell and R. N. Zitter. *Electrochem. Soc. Abstracts.* Abs. No. 98. Spring Meeting (May 1955).
9. J. van den Boomgaard, F. A. Kröger and H. J. Vink. *J. Electronics.* **1,** 212 (1955).
10. R. Gremmelmaier. *Z. Naturf.* **11a,** 511 (1956).
11. B. Seraphin. *Z. Naturf.* **91,** 450 (1954).
12. C. H. L. Goodman. *J. Electronics.* **1,** 115 (1955).
13. F. Herman. *J. Electronics.* **1,** 103 (1955).
14. G. Dresselhaus, A. F. Kip, C. Kittel and G. Wagoner. *Phys. Rev.* **98,** 556 (1955).
15. G. L. Pearson and M. Tanenbaum. *Phys. Rev.* **90,** 153 (1953).
16. J. Hatton, B. V. Rollin. *Proc. Phys. Soc., Lond.* A.**67,** 385 (1954).
17. M. Tanenbaum and J. P. Maita. *Phys. Rev.* **91,** 1009 (1953).
18. R. Barrie and J. T. Edmond. *J. Electronics.* **1,** 161 (1955).
19. H. P. R. Frederikse and E. V. Mielczarek. *Phys. Rev.* **99,** 1889 (1955).
20. J. Tauc and M. Matyas. *Czech. J. Phys.* **5,** 369 (1955).
21. H. Weiss. *Z. Naturf.* **11a,** 131 (1956).
22. R. P. Chasmar and R. Stratton. *Phys. Rev.* **102,** 1686 (1956).
23. E. Burstein. *Phys. Rev.* **94,** 1431 (1954).
24. T. S. Moss. *Proc. Phys. Soc., Lond.* B. **67,** 848 (1954).
25. H. J. Hrostowski, F. J. Morin, T. H. Geballe and G. H. Wheatley. *Phys. Rev.* **100,** 1672 (1955).
26. H. Fritzsche and K. Lark-Horovitz. *Phys. Rev.* **99,** 400 (1955).
27. B. V. Rollin and A. D. Petford. *J. Electronics.* **1,** 171 (1955).
28. R. F. Broom and A. C. Rose-Innes (*Proc. Phys. Soc., Lond.* B. **69,** 1269 (1956).
29. C. S. Hung and J. R. Gliessman. *Phys. Rev.* **96,** 1226 (1954).
30. H. Fritzsche. *Phys. Rev.* **99,** 406 (1955).
31. R. Talley and F. Stern. *J. Electronics.* **1,** 186 (1955).
32. I. M. Mackintosh and J. W. Allen. *Proc. Phys. Soc., Lond.* B. **68,** 985 (1955).
33. D. G. Avery and D. P. Jenkins. *J. Electronics.* **1,** 145 (1955).
34. W. Van Roosbroeck and W. Shockley. *Phys. Rev.* **94,** 1558 (1954).
35. V. Roberts and J. E. Quarrington. *J. Electronics.* **1,** 152 (1955).
36. E. Blount, J. Callaway, M. Cohen, W. Dumke and J. Phillips. *Phys. Rev.* **101,** 563 (1956).
37. H. Yoshinaga and R. A. Oetjen. *Phys. Rev.* **101,** 526 (1956).
38. W. G. Spitzer and H. Y. Fan. *Phys. Rev.* **99,** 1893 (1955).
39. F. Oswald and R. Schade. *Z. Naturf.* **9a,** 611 (1954).
40. R. Barrie, F. A. Cunnell, J. T. Edmond and I. M. Ross. *Physica.* **XX,** 1087 (1954).
41. R. Braunstein. *Phys. Rev.* **99,** 1892 (1955).
42. J. T. Edmond, R. F. Broom and F. A. Cunnell. *Proceedings of the Physical Society Conference on Semiconductors at Ashorne Hill,* 1956 (to be published).
43. O. G. Folberth and H. Weiss. *Z. Naturf.* **10a,** 615 (1955).
44. H. Welker. *J. Electronics.* **1,** 181 (1955).
45. M. B. Prince. *J. Appl. Phys.* **26,** 534 (1955).
46. E. S. Rittner, *Phys. Rev.* **96,** 1708 (1954).
47. D. A. Jenny, J. J. Loferski and P. Rappaport. *Phys. Rev.* **101,** 1208 (1956).
48. R. Gremmelmaier. *Z. Naturf.* **10a,** 501 (1955).
49. D. P. Detwiler. *Phys. Rev.* **97,** 1575 (1955).
50. H. N. Leifer and W. C. Dunlap, Jr. *Phys. Rev.* **95,** 51 (1954).
51. R. F. Blunt, W. R. Hosler and H. P. R. Frederikse. *Phys. Rev.* **96,** 576 (1954).
52. R. F. Blunt, H. P. R. Frederikse, H. J. Becker and W. R. Hosler. *Phys. Rev.* **96,** 578 (1954).
53. R. K. Willardson, A. C. Beer and A. E. Middleton. *J. Electrochem. Soc.* **101, 354** (1954).
54. W. Sasaki, N. Sakamoto and M. Kuno. *Phys. Soc. Japan.* **9,** 650 (1954).
55. E. Justi and G. Lautz. *Abl. Braunschweig. Wiss. Ges.* **5,** 36 (1953).
56. T. S. Liu and E. A. Peretti. *Trans. Amer. Soc. Met.* **45,** 677 (1952).
57. O. G. Folberth, O. Madelung and H. Weiss. *Z. Naturf.* **9a,** 954 (1954).
58. F. Oswald and R. Schade. *Z. Naturf.* **9a,** 611 (1954).

59. H. J. Hrostowski and M. Tanenbaum. *Physica.* **20,** 1065 (1954).
60. Reported by Talley and Stern, ref. 31.
61. R. M. Talley and D. P. Enright. *Phys. Rev.* **95,** 1092 (1954).
62. C. Hilsum (to be published).
63. H. Welker. *Z. Naturf.* **8a,** 248 (1953).
64. H. P. R. Frederikse and R. F. Blunt. *Proc. Inst.Radio Engrs., N.Y.* **43,** 1828 (1955).
65. F. Oswald. *Z. Naturf.* **9a,** 181 (1954).
66. F. Oswald. *Z. Naturf.* **10a,** 927 (1955).
67. D. A. Jenny. *Z. Naturf.* **10a,** 1032 (1955).
68. G. A. Wolff, R. A. Hebert and J. D. Broder. *Phys. Rev.* **100,** 1144 (1955).
69. O. G. Folberth and F. Oswald. *Z. Naturf.* **9a,** 1050 (1954).
70. O. G. Folberth. *Z. Naturf.* **10a,** 502 (1955).
71. H. Weiss. *Z. Naturf.* **11a,** 430 (1956).
72. D. G. Avery, D. W. Goodwin, W. D. Lawson and T. S. Moss. *Proc. Phys. Soc., Lond.* B. **67,** 761 (1954).
73. G. R. Mitchell, A. E. Goldberg and S. W. Kurnick. *Phys. Rev.* **97,** 239 (1955).
74. S. W. Kurnick, A. J. Strauss and R. N. Zitter. *Phys. Rev.* **94,** 1791 (1954).
75. C. Hilsum, D. J. Oliver and G. Rickayzen. *J. Electronics.* **1,** 134 (1955).
76. S. W. Kurnick and R. N. Zitter. *J. Appl. Phys.* **27,** 278 (1956).
77. E. W. Saker, F. A. Cunnell and J. T. Edmond. *Brit. J. Appl. Phys.* **6,** 217 (1955).
78. W. P. Mason, W. H. Hewitt and R. F. Wick. *J. Appl. Phys.* **24,** 166 (1953).
79. R. F. Wick. *J. Appl. Phys.* **25,** 741 (1954).
80. I. M. Ross and E. W. Saker. *J. Electronics.* **1,** 223 (1955).
81. W. Hartel. *Siemens Z.* **28,** 376 (1954).
82. I. M. Ross and N. A. C. Thompson. *Nature, Lond.* **175,** 518 (1955).
83. R. H. Willardson and A. C. Beer. *Elect. Mfg.* (Jan. 1956).
84. H. J. Thuy. *Arch. Elekt. Ubers. Tragung.* **8,** 269 (1954).
85. J. J. Loferski. *J. Appl. Phys.* **27,** 777 (1956).
86. I. M. Ross and E. W. Saker. *Nature, Lond.* **178,** 1196 (1956).

RADIATION EFFECTS IN SEMICONDUCTORS

J. H. CRAWFORD, Jr., Ph.D.
and
J. W. CLELAND, M.S.

Solid State Division, Oak Ridge National Laboratory,
Oak Ridge, Tenn., U.S.A.
MS. received August 1, 1956

RADIATION EFFECTS IN SEMICONDUCTORS

1. INTRODUCTION

The sensitivity of the electrical properties of semiconductors to lattice imperfections is too well known to need emphasis. Acceptor and donor impurity atoms, whether foreign to the host system or arising from non-stoichiometry, account for the extrinsic semiconducting properties. In fact, such impurities are almost entirely responsible for the electrical behavior, at ordinary temperatures in materials with large forbidden-gap width. Lattice defects may also profoundly affect the electrical behavior under certain conditions of defect concentration and temperature. This fact has long been known for photoconductors (alkali halides containing F-centers, *etc.*), and suspected for many semiconductors. Fast particle bombardment of the diamond lattice semiconductors germanium and silicon, however, provided the first conclusive evidence that point defects, *viz.* interstitial atoms and vacant lattice sites, produce acceptor and donor states. Consequently, fast particle bombardment has been extensively employed as a tool for investigating the defect sensitivity of semiconductors. Conversely, because of their great sensitivity to lattice defects, semiconductors have aided considerably in the study of the nature of radiation damage. More recently, other techniques of defect introduction have been explored. Quenching experiments[1] on germanium and silicon from high temperature have been performed to freeze-in defects, and plastic deformation[2] has been used to increase the concentration of edge dislocations which in turn, increases the total acceptor concentration of germanium.

We wish to summarize in this review the results of fast-particle bombardment experiments on a number of semiconducting systems. Historically, this field of study opened in 1947 when Johnson, Lark-Horovitz and co-workers at both Oak Ridge and Purdue University[3, 4] showed that exposure of germanium and silicon specimens to both reactor irradiations and cyclotron particles affected their conductivity and Hall coefficient too greatly to be attributed to impurities introduced by nuclear reactions. It was therefore concluded that the marked change in both concentration and type of charge carrier was caused by localized energy states associated with lattice defects resulting from nuclear bombardment. This view was confirmed by the observation that such effects could be almost completely removed by appropriate vacuum annealing. Since this early work, considerable effort has been devoted to studies of radiation effects in a variety of semiconductors of the diamond-type lattice using a number of radiation sources. In addition to studies of the energy level structure of radiation-produced defects which determine the effect on charge carrier concentration and type, effects on mobility, minority carrier lifetime, minority carrier trapping processes and their associated photoconductivity, annealing studies, optical properties and magnetic behavior have all been investigated.

It is convenient to include first a brief discussion of the mechanism of defect production by ionizing radiation. We shall then proceed to an examination of radiation effects in various materials. Since the primary effort has been placed on

germanium and silicon these will be treated in separate sections and the remaining materials will be covered in a single section. Although semiconductors in addition to those considered here have been bombarded, the results are too fragmentary to warrant inclusion in this review.

2. THE NATURE OF RADIATION PRODUCED DEFECTS

2.1. Production of Defects by Energetic Radiations

The theory of defect production by high-energy particles has received considerable attention since the prediction of E. P. Wigner[5] in 1943 that extensive radiation effects might be expected in reactor materials. The essentials of the energetics of the radiation damage process were first established by Seitz[6] and since refined by several investigators[7, 8, 9, 10].

When an incident energetic particle collides with a lattice atom while traversing a solid material, the lattice atom may receive sufficient recoil energy to be displaced to an interstitial site, thus creating an interstitial–vacancy pair. Moreover, if the energy of the primary recoil is large enough, subsequent collisions may produce secondary and tertiary displaced atoms. Consequently, a large number of lattice defects can result from the collision of a single energetic particle with a lattice atom.

The initial interaction may take place by one of three processes: (1) elastic collision with the nucleus; (2) coulomb interaction with the nucleus (Rutherford's scattering); and (3) elastic collision with the atom. The first process requires an uncharged particle, *i.e.* a fast neutron; the second, a charged particle with sufficient energy to penetrate the screening effect of the electron cloud of the stationary atom; the third process results when the energy is not large enough to penetrate the screening electrons and amounts to closed-shell interaction. It should be remembered that charged particles are important even in neutron bombardment, since recoiling atoms usually become ionized.

In addition to energy losses resulting from elastic collisions, an incident charged particle may also lose an appreciable fraction of its energy by electrostatic interactions which cause electronic excitation and ionization of the lattice atoms. In fact, above a certain energy ϵ_c, corresponding to the situation in the free atom case where the velocity of the moving particle is equal to that of the electrons in the outermost orbit, most of the energy of the moving ion is lost through such electronic interactions. For energies less than ϵ_c, the orbital electrons of the fixed atom can relax without experiencing excitation. In a solid, ϵ_c takes on a somewhat different significance because of the band structure. Seitz[6] has presented arguments which indicate that in an insulator or semiconductor:

$$\epsilon_c = (M_1/m)\, I_t/8 \qquad \ldots (1)$$

where M_1 is the mass of the moving atom, m is the mass of an electron and I_t is the value of the fundamental adsorption edge in energy units. In a metal, the valence electrons are essentially free, and ionization processes occur for nearly all energies of the moving ion. However, energy losses by electronic processes are expected to be negligible when the velocity of the incident charged particle

falls below that of the electrons at the top of the Fermi distribution in the metal. For fast neutrons, electronic processes are negligible for most solids ($Z_2 > 10$, where Z_2 is the nuclear charge of the stationary atom) since the average energy transferred by pile spectrum neutrons to a lattice atom of such a mass is seldom above ϵ_c. In the case of charged particles in the MeV range (protons, deuterons, α-particles, electrons, *etc.*), this situation does not hold for the incident particle. When calculating the number of displaced atoms in this case, it must be remembered that much, if not most, of the energy of the primary particle is lost in electronic interactions.

Snyder and Neufeld[8] have shown that for hard-sphere or elastic collisions with relatively heavy atoms ($Z_2 > 10$), the total number of displaced atoms (including secondaries and tertiaries) per incident neutron of energy ϵ is given approximately by:

$$A = \overline{T}/2\alpha \qquad \ldots (2)$$

where α is the energy required to displace an atom (assumed to be a definite threshold) and $\overline{T}$ is the average energy transferred to the recoiling atom by the particle during the collision, given by the relation:

$$\overline{T} = 2\epsilon M_1 M_2/(M_1 + M_2)^2 \qquad \ldots (3)$$

Here ϵ is the energy of the incident particle, M_1 its mass and M_2 the mass of the lattice atom. For a 1 MeV neutron, $\overline{T}$ for germanium is 27 keV. In order to determine the total number of atoms displaced by pile spectrum neutrons it is necessary to integrate equation (2) over the neutron energy spectrum, using the appropriate neutron scattering cross-section for the material in question.

In the case of energetic particles of charge Z_1, most of the energy range of interest falls within the Rutherford scattering range. Using the Bohr theory for the unshielded nucleus it can be shown[7] that in the non-relativistic range the cross-section for producing a primary displacement is;

$$\sigma_d = \eta/\epsilon \qquad \ldots (4)$$

where ϵ is the energy of the incident particle and

$$\eta = 4\pi a_0^2 \frac{M_1}{M_2} \cdot \frac{Z_1^2 Z_2^2 \epsilon_R^2}{\alpha} \qquad \ldots (5)$$

where a_0 is the Bohr radius and ϵ_R is the Rydberg energy. The mean energy of the primary recoil is approximately:

$$\overline{T} = \alpha \ln \left\{ \frac{4M_1 M_2 \epsilon}{\alpha(M_1 + M_2)^2} \right\} \qquad \ldots (6)$$

Using this value, the approximate number of displaced atoms per particle may be obtained from equation (2) or, alternatively, from an approximate relation due to Seitz[7], which is valid over a greater range of energy. For a thick target, *i.e.* when the incident charged particle spends an appreciable portion of its range in the solid, the total number of displaced atoms is obtained by integrating over its path. Here it is important to employ the rate of energy loss in electronic

interactions as well. Approximate methods for this calculation have been developed by Seitz and Koehler[7].

The situation for electron bombardment is more complex since the threshold energy required of an electron for displacing an atom normally lies in the relativistic range[7, 10]. However, the energy transferred to the recoiling atom is considerably smaller so that multiplication of defects is small except for very large incident electron energies.

The foregoing discussion considers the problem primarily from a two body picture. A more statistical picture has also been evolved that employs the concept of a 'thermal spike' *i.e.* a small region of the crystal, heated to high temperatures by the incident particle and rapidly quenched ($> 1000°$ C for $\sim 10^{-11}$ sec)[11]. More recently Brinkman[9] has suggested a similar 'displacement spike' resulting from the high rate of energy loss of an incident particle toward the end of its range and Seitz[7] has suggested that both processes might occur concurrently with the displacement processes. These have been invoked in explaining the disordering of super-lattice alloys (Cu_3Au, CuNi, *etc.*) whose rate of disordering under bombardment is too great to be understood by displacement processes alone. It has also been suggested that the spike picture can lead to additional defects such as regions of misorientation and small dislocation loops[7, 9]. It should be pointed out, however, that the thermal spike concept is far from satisfactory, and recently Brooks[12] has examined this mechanism critically in the light of recent refinements in the theory.

Kinchin and Pease[10] have indicated an alternative approach for understanding the disordering process in that, although a recoil atom may not have sufficient energy to displace a second atom, it may replace it with a much smaller energy expenditure. Hence an atom knocked interstitially may disorder other atoms by successive replacements without increasing the defect density. This mechanism also permits displaced atoms to travel appreciable distances before the recoil energy is completely degraded.

Although on the basis of the above discussion the results obtained on a given solid are expected to be qualitatively the same for any incident particle of sufficient energy and dose, there are some subtle differences which may have an important bearing on processes occurring during, and subsequent to, bombardment. The most important of these is the distribution of defects. An energetic massive particle may transmit sufficient energy to a primary particle so that the number of secondary and tertiary displacements may be several hundred. Consideration of the range of the primary particle itself[12] indicates that this large number of defects must be confined in a small region of the crystal, encompassing perhaps 10^4 to 10^6 atoms. At the other extreme, bombardment with electrons whose energy is near the threshold for atomic displacements (such that $T < 4\alpha$) means only one displaced atom will result per collision, hence the defects will be randomly distributed throughout the effective electron range. Unfortunately, the range of electrons, of energy < 1 MeV, is quite short, which makes investigations on bulk properties of specimens difficult. It is possible to overcome this experimental limitation and still obtain randomly distributed defects by employing the Compton electrons, and photo-electrons which result from γ-ray absorption in solids. Such experiments will be discussed below.

2.2. Displacement Energy

One of the most important experiments from the standpoint of radiation damage theory is the determination of the displacement energy, thereby justifying the assumption that lattice atoms sit in a simple potential well of depth α. M. M. Mills pointed out that this parameter could be determined by electron bombardment and such measurements have been made on several types of materials. E. E. Klontz[13] made such measurements on N-type germanium, using the resistivity as an index of lattice damage. The specimens were irradiated

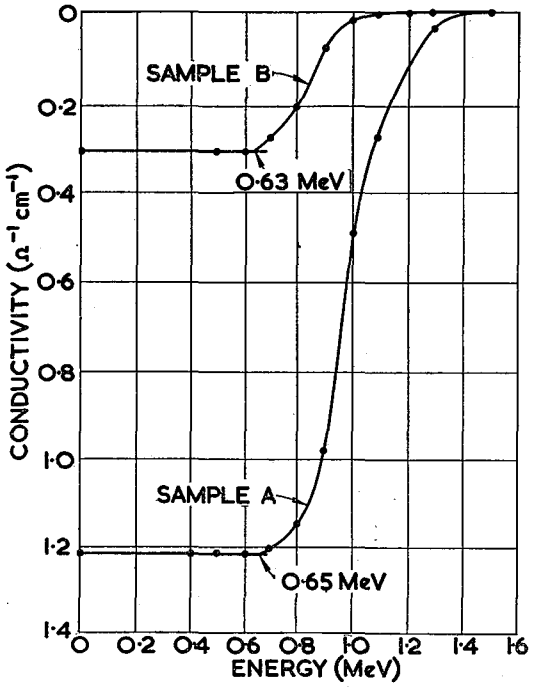


Figure 1. *Conductivity of two germanium samples during electron irradiation as a function of the energy of the electrons (Klontz[13]).*

at liquid nitrogen temperature and the electron energy was decreased until no resistivity change could be detected for a prescribed exposure. The results are shown in Figure 1, from which it is evident that the threshold energy is $\sim 0{\cdot}63$ MeV. Because of the rapid falling off of displacement cross-section near the threshold, the actual value may be somewhat smaller. This value of the threshold energy leads to an α of 31 eV. This value is quite near the value of 25 eV which is usually assumed for most materials[7]. Recently Loferski and Rappaport[14] have re-examined the problem of determining the displacement energy in both germanium and silicon. As an index of damage they employed the minority carrier lifetime which, because bombardment introduces recombination centers, is much more sensitive to lattice damage than concentration change. The short circuit current of photo-diodes was used to reflect the change

in lifetime. Room temperature bombardments on both silicon and germanium indicated a threshold energy of the electrons required to produce an effect that corresponded to a displacement energy of 12·9 eV for both materials. This value is less than that of Klontz by more than a factor of two, and is therefore too large a discrepancy to be attributed to experimental error.

There are several possible explanations for the discrepancy. The lifetime is a considerably more sensitive index of lattice damage than Klontz's resistivity measurements, which were made on specimens with an electron concentration of $\sim 10^{14}$ cm^{-3}. Kohn[15] has calculated that the energy required to place a germanium atom interstitially is less than half of Klontz's value, and he has suggested that either the first interstitial position is unstable or that closely coupled interstitial–vacancy pairs might not lead to acceptor states. The latter view is certainly not unreasonable and yet such a closely coupled pair might still be an effective recombination center for excess electrons and holes. A second explanation questions the suitability of assuming a discrete value for α. The energy required to displace an atom is unquestionably dependent on direction, hence the displacement energy should be treated as a distribution. Sampson *et al.*[16] have suggested a probability function extending from zero at $\epsilon = \epsilon_1$ to unity at $\epsilon = \epsilon_2$ for a displacing particle of energy ϵ, hence the displacement cross-section would vary from zero at ϵ_1 to a small but finite value at ϵ_2. If such is the case, a small concentration of acceptors may escape notice while the change in lifetime may yet be detected.

2.3. The Rate of Defect Formation During Bombardment

The theory as outlined above has been applied to calculate the number of displaced atoms in Ge for various particles. The results are compared with experiment in Table 1. The first three columns list the particle, its energy and the temperature of exposure. Column 4 lists the particle range or sample thickness (whichever is the larger) of the specimens employed. Column 5 lists the rate of conduction electron removal per incident particle. The thickness was used to compute the experimental number of displacements per incident particle (the number displaced per cm of path) which is given in Column 5. The numbers of displaced atoms per incident particle as calculated are listed in the last column. Most of these calculations except for neutrons are due to Seitz and Koehler[7] and have been suitably adjusted for specimen thickness.

Before comparing the experimental and calculated values of Table 1, it should be mentioned that because of scatter in the data the experimental values are perhaps not reliable to better than a factor of 1·5 and evidence will be presented below that two electrons are trapped per Frenkel defect. On this basis, the experimental values of atoms displaced at low temperature (77° K and 90° K) agree with the theoretical ones to within a factor of 2. In view of theoretical approximations and experimental uncertainty, this constitutes what appears to be excellent agreement and the wide range of particles and energies so correlated may be taken as evidence of the general applicability of the displacement theory. The larger deviation toward lower values at higher temperatures is probably due to thermal annealing of the defects during irradiation as will be indicated later.

Table 1.

Particle	Energy (MeV)	Exposure temperature (°K)	Sample thickness (or range) (cm)	$-dn/d(nvt)$ (cm^{-1})	$dN/d(nvt)$ (cm^{-1})
Neutrons	Reactor spectrum	300		3·2(*a*)	10(*f*)
α-particles	5·3	300	$1·9 \times 10^{-3}$	4100(*b*) 5000(*c*)	6500 6500
Deuterons	9·6	205	$1·5 \times 10^{-2}$	1100(*c*)	2100
Electrons	1·5	77	$4·5 \times 10^{-2}$	1·3(*d*)	1·1
		300	$4·5 \times 10^{-2}$	0·6(*d*)	1·2
	3·0	300	$1·6 \times 10^{-1}$	1·4(*e*)	2·3
	4·5	90	?	8(*c*)	3·0
		205	?	3·7(*c*)	3·0

a. Crawford and Lark-Horovitz. *Phys. Rev.* **78,** 815 (1950).
b. Brattain and Pearson. *Phys. Rev.* **81,** 846 (1950).
c. Fan and Horovitz. *Report on Conference on Defects in Crystalline Solids,* p. 232. (Physical Society, London, 1955).
d. Klontz. Doctoral Thesis (Purdue University, 1952, unpublished).
e. Brown. Fletcher and Wright. *Phys. Rev.* **92,** 591 (1953).
f. Calculated assuming the average energy of the fast flux is ~ 0·08 MeV.

3. DEFECTS IN DIAMOND LATTICE SEMICONDUCTORS

3.1. General Considerations

In the previous section the mechanism by which lattice defects are produced by bombarding particles was outlined and certain quantitative tests of the theory were considered. In order to understand the influence which these defects exert on the electrical behavior of semiconductors, it is necessary to examine in some detail the consequences they have for electronic structure. Perhaps the most important effect is that on carrier type and concentration. Hence we shall be directly concerned with localized energy states which arise as a result of defects in the diamond lattice. Such localized states also have an important effect on optical absorption, minority carrier lifetime and minority carrier trapping phenomena. Somewhat more indirectly, since they distort the charge distribution in the crystal, they also alter the mobility, particularly in the impurity scattering range.

The tacit conclusion of the preceding section was that bombardment introduces Frenkel defects, and it is these with which we shall be concerned while discussing a possible energy level model. However, there is much evidence that a simple description in terms of equal numbers of interstitials and vacancies alone may be a gross over-simplification except for certain limited cases. Because of the short range of the primary recoils and the high damage density along the path of a charged particle, the lattice disorder resulting from massive high energy particles is not randomly distributed, but rather is better described by regions of high local disorder distributed throughout a more nearly perfect matrix. As we shall point out later there are indications of a much more complex spectrum of energy levels resulting from such a non-random defect distribution than can be expected for simple randomly distributed Frenkel defects. In addition,

presumably because of the appreciable mobility of some of the defects (most probably the interstitials), extensive alteration of the defect energy-level structure occurs during and subsequent to exposure at room temperature and extensive rearrangement is indicated even at very low temperature (—150° C). These activated rearrangements of disorder are expected to produce more complex defect distributions with a concomitant complexity in their level structure. With these qualifications in mind, we proceed to a discussion of current models of energy level structures.

3.2. Localized Energy Levels of Defects

One of the puzzling features of early studies of radiation effects was the markedly different behavior of germanium and silicon. N-type germanium was converted to P-type and P-type specimens increased in conductivity σ with exposure whereas both N- and P-type silicon tended towards intrinsic. The behavior of germanium could be explained qualitatively on the basis of acceptor introduction alone, while in silicon it was necessary to postulate the introduction of both deep-lying acceptors and donors to account for the removal of carriers of both signs. In view of the close chemical and electronic similarity of these materials, this apparent difference in defect energy level structure was not to be expected. James and Lehman[17], however, pointed out that donors or hole traps in germanium might escape detection if they lay near the top of the valence band. Further investigation of low-resistivity P-type germanium indeed demonstrated their presence[18]. Moreover, a closer examination of σ *v.* bombardment curves of initially N-type germanium revealed that the initial linear slope extrapolated to $\sigma = 0$ at an exposure only about one-half that required to cause conversion to P-type[19].

This type of behavior led James and Lark-Horovitz[20] to propose a model of multiple ionization of both interstitials and vacancies to account for the observed effects of bombardment. The interstitial atom was treated as an isolated system in a dielectric medium and the ionization potential of each of its valence electrons (considered from a hydrogenic model) was estimated from the relation:

$$\epsilon_i = \frac{i^2 \epsilon_H}{(\kappa_i')^2} \qquad \ldots (7)$$

where i refers to the ionization potential in question and is therefore the effective nuclear charge, ϵ_H is the ionization energy of the hydrogen atom and κ_i' is the effective dielectric constant to which the electron in question is subject. κ' is expected to decrease with increased ionization taking a value somewhere between that of the material and unity. For example in germanium, $\kappa = 16$ leading to a first ionization potential of $\sim 0{\cdot}05$ eV. This, in effect, is a localized energy level that would serve as a donor state lying $0{\cdot}05$ eV below the conduction band. The second electron, since its orbit is now much restricted because of the higher effective nuclear charge, is subject to a much smaller κ' and its ionization potential might be expected to place a localized energy level near or below the middle of the forbidden band. The remaining valence electrons would probably be too tightly bound to contribute additional levels.

Such an atomistic approach cannot be used so readily for vacancies; hence James and Lark-Horovitz considered the levels split out of the energy bands

when a core ion was removed. They concluded that as many as two electrons might be introduced into a vacancy before the electrostatic repulsion becomes too great and that each such event would also introduce a localized energy level in the forbidden band. The first vacancy-produced level would lie at approximately 0·05 eV above the filled band and would therefore serve as an acceptor or electron trap. The second vacancy-produced level would presumably

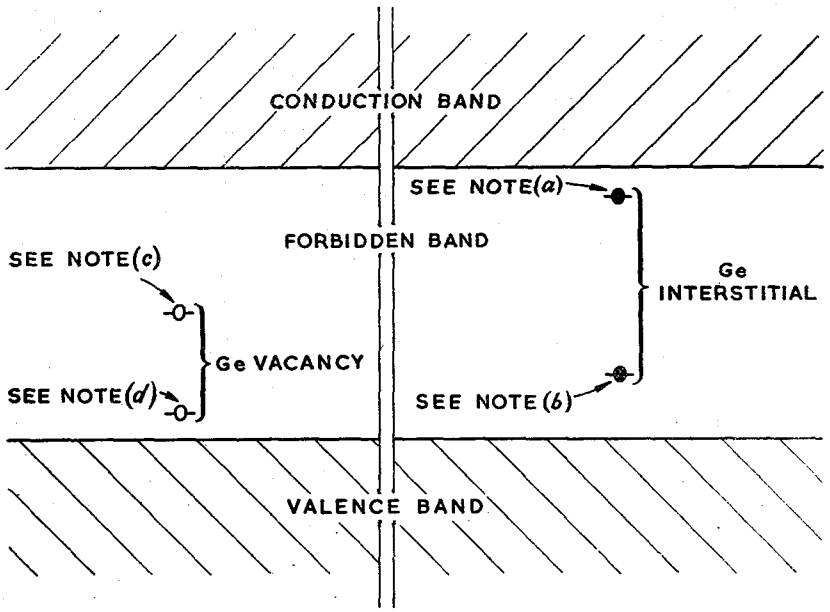


Figure 2. *Level diagram for vacancies and interstitials in germanium according to James and Lark-Horovitz.* (*a*) *If occupied, the interstitial atom is neutral;* (*b*) *If not occupied, interstitial atom is doubly charged;* (*c*) *If occupied, net charge of two electrons near vacancy;* (*d*) *If not occupied, zero net charge near vacancy.*

require a higher ionization energy. The net result of their analysis is shown in Figure 2 in which the levels are arbitrarily indicated for both interstitials and vacancies. It should perhaps be emphasized that when both types of levels are present concurrently, as would be the case for Frenkel defects, the electrons from the interstitial atoms would redistribute themselves over available levels to produce the configuration of lowest energy. Therefore, whether a level is occupied or vacant depends on its energy position rather than on the defect with which it was originally identified. The important contribution of this model is that it indicates the most important parameters for energy level positioning.

Although this model is both qualitative and somewhat speculative it has been of considerable assistance in determining the rather complex energy level structures of irradiated semiconductors. Perhaps its best justification is that it appears reasonably consistent with the gross features of many experiments and is the best description currently available.

Another defect energy level model applicable to the diamond lattice has

recently been proposed by Champion[21] in an attempt to explain the electronic behavior of diamond. He uses a quite different approach employing localized, resonating orbitals and postulating a series of partial covalent-type bonds to describe the electronic properties of the various defects. As might be expected from the different approach, his model leads to strikingly different conclusions. Thus the interstitial acts as an electron trap or acceptor while the vacancy may either trap electrons or act as a donor. There has been no attempt as yet to apply this model to germanium and silicon nor to obtain experimental confirmation. Some of the conclusions, however, do not seem applicable to the Group IV semiconductors. For example, it hardly appears likely that in materials of high dielectric constant an interstitial would prefer to add electrons and thereby become negatively charged. Although such a process may be possible for diamond, it is indeed questionable whether Ge^- or Si^- could be stable unless the complex resonating orbitals which are postulated are exceedingly stable. Future work will perhaps clarify this point as well as the general applicability of Champion's model to the Group IV semiconductors.

It should be mentioned here that Dienes and Kleinman[22] have suggested that the early specific heat results obtained by Keesom, Pearlman and Lark-Horovitz[23] might be explained by a rearrangement of bonds into an unsaturated structure which is not unlike Champion's model for clusters of defects. Keesom *et al.* found that in one silicon specimen the Debye temperature apparently decreased as a result of fast neutron bombardment. This behavior was explained[22] as resulting from a bond rearrangement into a configuration analogous to a large condensed ring structure which would produce a softer phase containing ~ 5000 atoms imbedded in the otherwise undeformed matrix. Although such an unsaturated structure may be stable in the diamond system, the well-known instability of double bonds in silicones makes it questionable whether such a situation could exist in crystalline silicon. Moreover recent specific heat results obtained on a silicon single crystal by Keesom *et al.*[24] appear to throw doubt on the earlier work and additional experiments are necessary before the decrease in Debye temperature mentioned above can be considered conclusive.

4. EXPERIMENTAL INDICATIONS OF DEFECT STATES

4.1. Germanium

Considerable effort has been spent on the determination of the type and position of defect states in germanium produced by a wide variety of energetic particles. The results of fast neutron bombardment have been reported by Cleland *et al.*[25, 26] for both N-type and P-type germanium. The data were analyzed in terms of the James–Lark-Horovitz model[20], *i.e.* a four-level model with two levels occupied, under the assumption that the usual chemical donors or acceptors were completely ionized, the position of the levels being treated as parameters to be determined from experiment. In order to avoid complications arising from annealing of the defects during exposure, the initial behavior was used wherever possible. For N-type germanium the situation was easily analyzed, since only one level was found to lie above the middle of the forbidden band. In this case the data were consistent with the model and indicated two vacant

states (electron traps or acceptors) present in equal concentration identifiable with the interstitial: a shallow one (0·20 eV below the conduction band), and a deep one lying below the middle of the conduction band. The deep state, of course, is the dominant acceptor level in bombarded P-type specimens.

Since the model predicts three states (one vacant and two occupied) below the middle of the forbidden gap, the situation in P-type germanium is much more difficult to analyze. However, it can be readily shown that if the lowest-lying occupied state is several kT below the higher one, the limiting value of the Fermi level φ^* will lie midway between the vacant state and the occupied state. φ^* is that position approached by φ at saturation. If the initial value of φ lies below φ^*, a decrease in hole concentration p will result from bombardment while, if σ initially lies above this point, exposure will cause an increase in p. Consequently, φ^* should be readily determined from the variation of initial slope of the σ *v.* bombardment curve with initial hole concentration p_0. Such results for exposure at 195° K are shown in Figure 3. The rate of change in p with defect concentration N was obtained from the conductivity variation, assuming the mobility was unaffected. It was further assumed that 1·6 Frenkel defects were introduced per incident fast neutron as indicated by the rate of electron removal in N-type specimens[19]. The zero value of $dp/dN_{N=0}$ at this temperature leads to $p^* = (1{\cdot}5 \pm 0{\cdot}2) \times 10^{15}$ cm^{-3} which, on using the best value for effective hole mass[27] ($m_h \simeq 0{\cdot}3\ m_0$), yields $\varphi^* = 0{\cdot}123 \pm 0{\cdot}003$ eV. At pile ambient temperatures ($\sim 50°$ C) a similar series of measurements yields $p^* = (6{\cdot}5 \pm 1{\cdot}5) \times 10^{16}$ cm^{-3} and, correspondingly, $\varphi^* = 0{\cdot}120 \pm 0{\cdot}006$ eV. Hence, within experimental error, φ^* is temperature independent as expected. It should be mentioned here that a study by Forster and Fan[28] of deuteron-bombarded Ge at a much higher rate of defect introduction shows some interesting differences. They find that at 195° K $\varphi^* = 0{\cdot}089$ eV, a value considerably smaller than the neutron value. Moreover, specimens with p_0 only slightly greater than the value of p^* obtained with fast neutrons show first a σ decrease followed, with further deuteron bombardment, by an increase. This behavior suggests that one type of lattice defect is being preferentially removed by annealing during exposure or that a non-equilibrium distribution of carriers exists. This point will be considered in greater detail below.

If the deepest occupied state is neglected (assumed to lie several kT below the next highest state), determination of the position of the acceptor and hole trap (occupied) from the initial rate of hole removal curve reduces to a one-parameter problem once φ^* is known. The data of Figure 3 have been fitted with $\epsilon_1 = 0{\cdot}180$ eV and $\epsilon_2 = 0{\cdot}066$ eV, where ϵ_1 refers to the vacant state and ϵ_2 to the occupied state measured from the top of the valence band. The second occupied state presumably lies much nearer the valence band. It should be pointed out that the correlation is sensitive to the rate of production of defects per incident fast neutron (taken to be 1·6) and is, therefore, sensitive to the neutron energy spectrum. Also, any variation in mobility would cause the slope values to be decreased somewhat. Consequently, these positions can only be considered approximate.

Fan and co-workers[29] have analyzed the results of electron and deuteron bombardment in a more general fashion. They have determined the positions of defect states by an examination of the variation of φ with bombardment:

φ is expected to show a maximum rate of change when it passes through a level of discrete states. The results for 4·5 MeV electrons are shown in Figure 4. They find that the shallow vacant level lies 0·20 eV below the conduction band in excellent agreement with the fast neutron results. In contrast, however, the deep acceptor lies 0·24 eV above the valence band rather than at 0·18 eV as the neutron results indicate. This value of 0·24 eV has been substantiated by the γ-ray bombardments of Cleland *et al.*[30] As was mentioned earlier, since damage during γ-irradiation results from energetic Compton electrons and

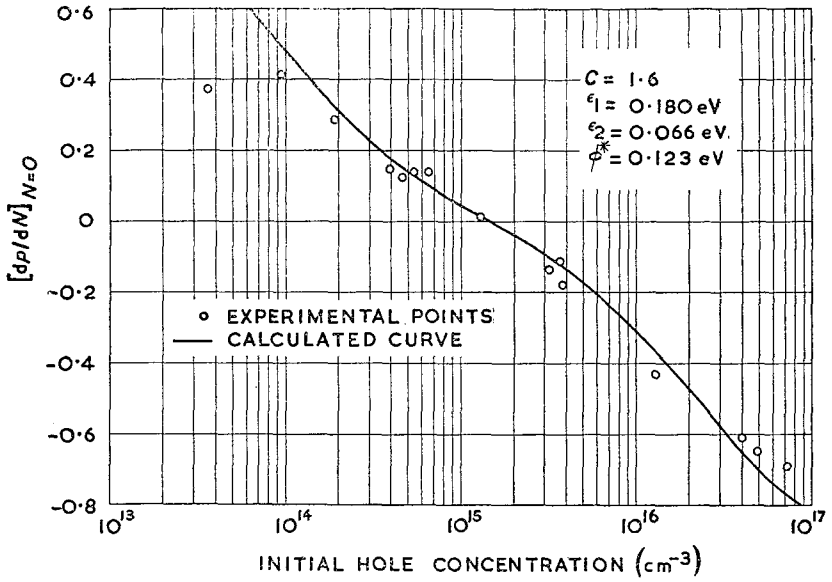


Figure 3. *The initial rate of change of hole concentration per Frenkel defect v. the initial hole concentration during bombardment at 195° K. The points are experimental and the solid curve is calculated assuming that C, the number of defects introduced per incident neutron, is 1·6.* (*Cleland et al.*[26])

photo-electrons, these experiments correspond to internal electron bombardment. The γ-ray experiments also indicate a shallow state 0·20 eV below the conduction band.

The results of deuteron bombardment are generally quite different from those obtained with electrons and fast neutrons. For 9·6 MeV deuteron bombardment at 200° K, Fan *et al.*[29] find evidence for a distribution of states less than 0·10 eV from the conduction band and a similar set of levels less than 0·10 eV from the top of the valence band. Moreover, they find that the distribution of levels varies with exposure.

Since they do not suffer from an inhomogeneous distribution of defects and are therefore less subject to smearing out of localized levels and preferential annealing during exposure, the electron and γ-ray results are the most reliable as far as Frenkel defect energy levels are concerned. Unfortunately, because of the much smaller damage rate with these radiations, no information relative

to the occupied states near the valence band has yet been obtained. An experiment similar to the initial slope study in neutron bombarded P-type germanium (Figure 3) utilizing electrons or γ-rays would be quite valuable.

In summary, we might say that, although there is some disparity in these results, they appear to confirm qualitatively the more important features of the James–Lark-Horovitz model of Frenkel defect states.

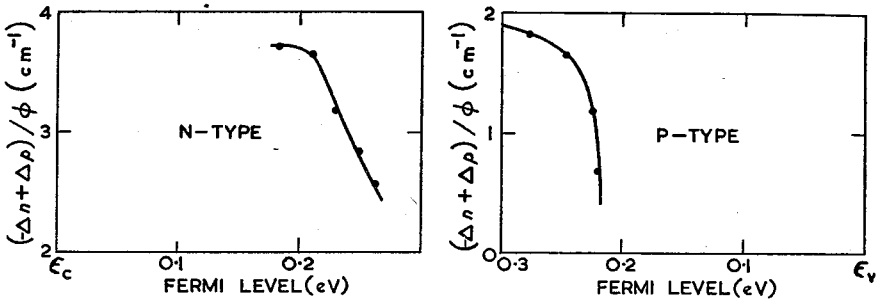


Figure 4. *Number of conduction electrons removed or holes added by a 4·5 MeV electron per unit length, as a function of the Fermi level. Irradiation at 200° K. (Calculated by taking the total change of concentration (n cm^{-3}) at a time of irradiation t divided by the total flux cm^{-2} incident on the sample up to the time t.) (Fan and Lark-Horovitz.*[29])

There is another method of defect state spectrometry, currently being considered by Schweinler, Cleland and Crawford[31], which is very attractive. This method, should it prove feasible, has the advantage of great generality and may be applied to specimens doped with multi-level impurities, *i.e.* gold, manganese, nickel, *etc.*, as well as bombarded specimens. The nuclear properties of Ge are such that nuclear doping, as a result of radio-active decay, can depress φ across the forbidden energy gap and localized states can be detected from the variation of φ during decay time. The predominant decay products of Ge, resulting from thermal neutron capture, are ^{71}Ga and ^{75}As, both of which are stable nuclides which ultimately result from (n, γ) processes on ^{70}Ge and ^{74}Ge. The thermal neutron capture cross-sections are such that ^{71}Ga and ^{75}As are produced to the extent of 3 per cent and 1 per cent of the integrated thermal flux $(nvt)_{th}$, respectively[32]. The ^{75}Ge activity decays to As with an 82 min half-life, while ^{71}Ge produces Ga with a half-life of 11·4 days. Hence, by appropriate choice of initial electron concentration and thermal neutron exposure, an N-type specimen containing energy levels of interest can be converted to P-type by compensation of impurities in a period convenient for observation, provided that the decay product sits substitutionally in the lattice. Investigation shows that the decaying radiation is too soft to eject the decay product from its position by recoil and, although the spectrum of γ-rays emitted on thermal neutron capture by ^{70}Ge and ^{74}Ge is not known, experiments using nearly pure thermal neutrons do not indicate that germanium atoms are displaced during neutron capture. Consequently, the decay products are expected tc behave as normal substitutional impurities.

Preliminary experiments have revealed an inherent difficulty with this method. Since the resolution in energy is $\sim 4\,kT$, it is desirable to follow φ during decay at as low a temperature as practicable. However, the bombardment-produced

states (and also the states of multi-level impurities) are efficient minority carrier traps (see sub-section 6.4) and the decaying radiation is sufficient to saturate these traps at the temperatures necessary for acceptable resolution. Consequently, the charge carrier concentrations are not in thermal equilibrium at liquid nitrogen temperature, which renders the method ineffective. However, experiments aimed at eliminating this difficulty are now underway.

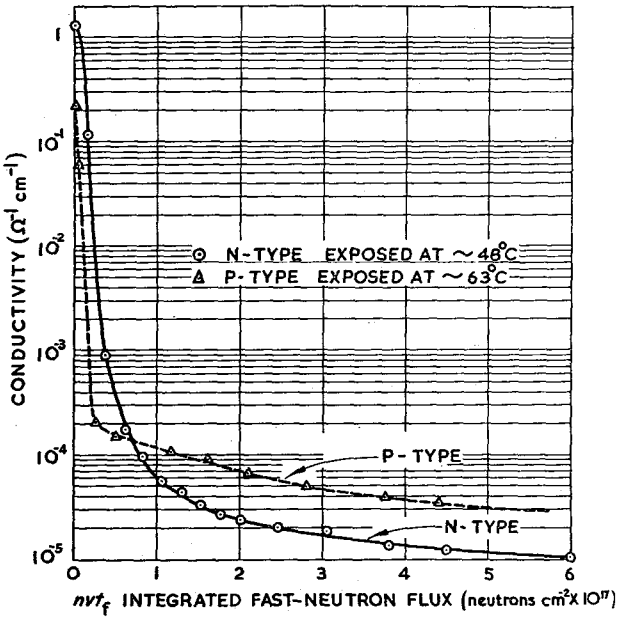


Figure 5. *Conductivity of N- and P-type silicon during fast-neutron irradiation. (Cleland et al. unpublished.)*

Parenthetically, two interesting consequences of nuclear doping might be mentioned. The half-life of ^{71}Ge has been checked by the variation of carrier concentration during decay[33]. The results are in reasonable agreement with the conventional determination. Nuclear doping by careful incremental exposures to thermal neutrons has been used to render germanium intrinsic to temperatures below dry ice temperature[34]. The latter application promises to be valuable for studies of the intrinsic behavior over a wide temperature range. Additional uses for this technique appear to be promising. Since acceptors and donors are introduced in a fixed ratio (1 As to 3 Ga), the known compensation which results can be used to study impurity scattering phenomena and to test the Brooks–Herring scattering theory[35]. It may also find valuable application in studies of impurity-banding phenomena. Investigations of these applications are currently being planned.

4.2. Silicon

Although the initial radiation effect studies were made on silicon at the same time as on germanium, considerably less extensive work has been done on this

material. This results from the facts that high-purity, single-crystal specimens have only recently become generally available, and that ohmic contacts to silicon are more difficult to prepare. In addition, the positioning of the defect states in the forbidden band of silicon are such that it is more difficult to locate them. Curves of σ *v.* nvt_f for N- and P-type silicon for fast neutron exposure are shown in Figure 5. After an initial rapid decrease (the rates of carrier removal are about the same for each type and are nearly twice that for N-type germanium), σ approaches the intrinsic value. The difference between the saturation values for the two specimens is due to the different irradiation temperatures. As might be expected, the conductivity reached after prolonged exposure is somewhat lower than the intrinsic value because of a bombardment-induced decrease in mobility. The results of deuteron bombardment obtained by Forster[38] and Longo[36] are qualitatively the same as with neutrons. Measurements of Hall coefficient on heavily irradiated P-type silicon by Longo and Kleitman[37] indicate a negative value as expected for intrinsic silicon.

These results suggest that the two occupied and two vacant states predicted by the James–Lark-Horovitz model are symmetrically located above and below the middle of the forbidden band respectively. The larger energy gap and smaller κ of silicon make a more symmetrical positioning quite reasonable. The problem of the actual positioning of levels, however, remains, and electrical measurements have not been very successful in revealing the solution. In contrast with the behavior of neutron-bombarded germanium which indicated two vacant states, the log σ *v.* *nvt* curve rather than the σ *v.* *nvt* curve for both N- and P-type Si bombarded near room temperature is approximately linear initially which suggests a distribution of levels such that a linear depression of φ toward the middle of the gap occurs with exposure. This behavior has also been observed by Forster[28].

Electron bombardment, as in the case of germanium, again gives more clear-cut results than massive particles. Hill and McKay[38] have found that the rate of carrier removal per 4·5 MeV electron shows a sharp break when φ reaches a position 0·36 eV below the conduction band and 0·37 eV above the valence band in high-purity N-type and P-type silicon respectively. Furthermore, impure specimens, with carrier concentrations nearly in the degenerate range, also give indications of localized states close to both the conduction and valence bands. Shallow states such as these would also explain the almost linear initial portion of the log σ *v.* exposure curves mentioned above for neutrons and deuterons. Since the shallow states are indicated by electron bombardment experiments, they are presumably real and not a result of inhomogeneities as suspected for germanium (sub-section 4.1).

The defect energy level structure indicated by these results is summarized schematically in Figure 6. It should be emphasized that this diagram is at best tentative. The four sets of levels, two shallow and two deep, are symmetrically located with respect to the center of the gap. The shallow states which lie near the band edges have been arbitrarily placed at 0·05 eV from the conduction band and valence band respectively. These states may be interpreted in the light of the model of James and Lark-Horovitz[20] as follows: levels A and B are attributed to the first and second ionization potentials of the interstitial, while levels C and D are taken to be the energies required to place one and two electrons

respectively into a vacant site. This interpretation is supported to some extent by optical absorption measurements on irradiated silicon (see sub-section 6.2). According to the model in question, the states associated with each type of defect are expected to lie further from the band edges than in the case of germanium because of the smaller dielectric constant of silicon ($\kappa = 11$). Brooks[12] has pointed out that this contention is substantiated by the position of levels

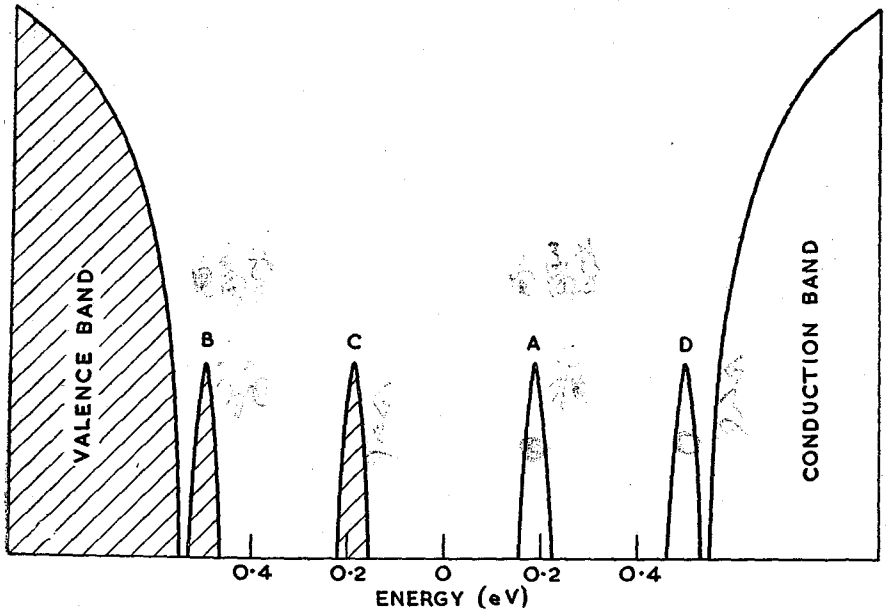


Figure 6. *Energy level diagram for defects in silicon. Levels A and B are attributed to the first and second ionization potentials of the interstitial while levels C and D are taken to be the energies required to place one and two electrons respectively into a vacant site and are therefore indicated as being occupied.*

associated with gold impurity in the two materials. In germanium[39], gold produces an occupied state at 0·05 eV above the valence band, whereas in silicon[40], this state moves up to 0·35 eV. It is noted that this is very close to the displacement suffered by the deepest lying occupied defect-state in going from germanium to silicon. Consequently, the state of affairs in silicon seems to correspond with the prediction of the model.

4.3. Indium Antimonide

Although, because of their binary nature, the Group III–V compounds of interest crystallize in the zinc blende lattice, this would reduce directly to the diamond lattice if the constituents were the same. Therefore, these materials may be conveniently treated in this section with only a few qualifications. Because of its availability in high-purity single-crystal form, the most attention given to semiconductor compounds has been concentrated on InSb[41]. Its behavior under fast neutron bombardment is almost the inverse of germanium in that P-type specimens are converted to N-type, and N-type specimens

approach saturation at a low value of σ, after prolonged exposure to neutrons at room temperature. The effect on σ (measured at 77° K) for both N- and P-type specimens is indicated in Figure 7 and the temperature dependence of the Hall coefficient after successive exposures is shown in Figure 8. Because of the large thermal neutron cross-section of indium which transmutes to tin, a donor impurity when substituted for indium in the lattice, it is necessary to

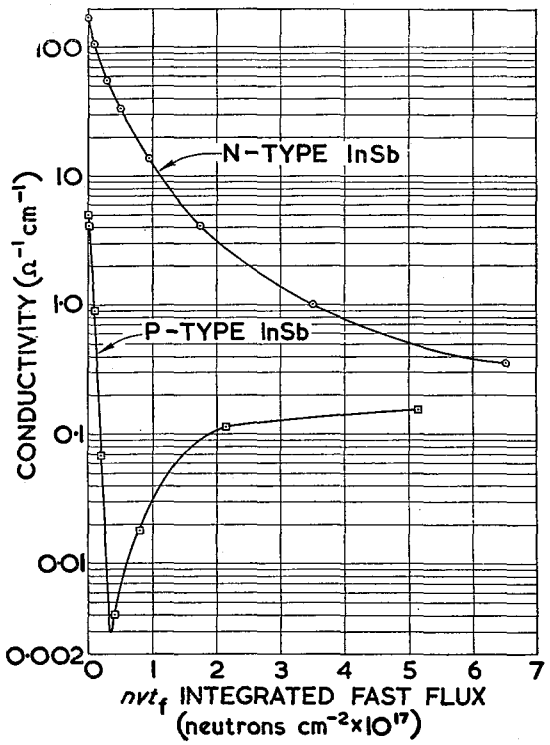


Figure 7. *Conductivity of InSb at — 196° C. after successive reactor irradiations.* (*Cleland and Crawford*[41].)

exclude this type of nuclear doping when studying displacement effects. This was done by shielding the specimens with layers of indium and cadmium foil which stop only the thermal neutrons.

Provided it can be assumed that ratios of states are not altered during exposure by preferential annealing of one type of defect, the temperature dependence of limiting electron concentration n^* and hence φ^* should be obtainable from a curve lying between curves P_2 and N_2 of Figure 8. Consequently, if this saturation condition is determined by two states above the middle of the gap, one occupied and one vacant, φ^* should, as in the case of germanium, be essentially temperature independent and lie half-way between the two levels. The value of φ^*, calculated from Figure 8 using an electron effective mass of 0·015 m_0 as indicated by cyclotron resonance experiments[42], is found to be

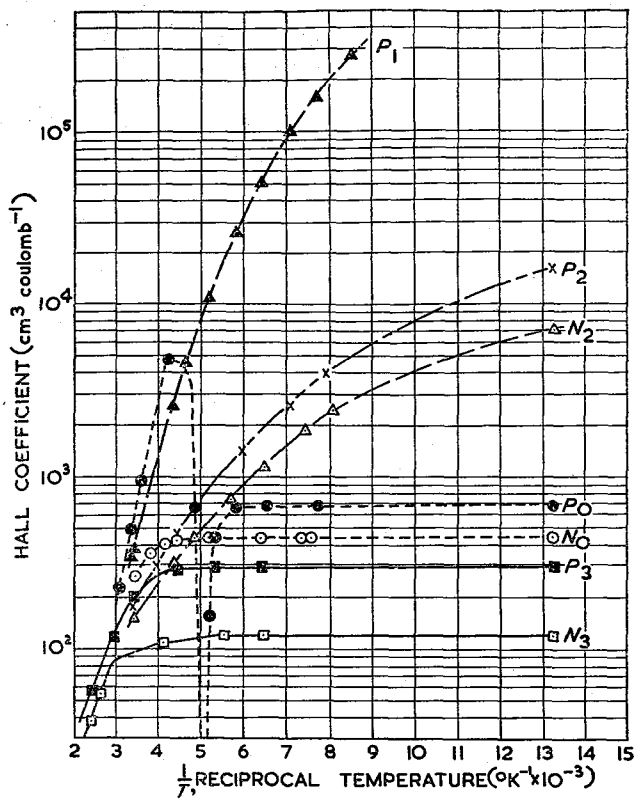


Figure 8. *Hall coefficient of InSb as a function of temperature after various reactor irradiations. N_0, original N-type InSb; N_2, after $6{\cdot}5 \times 10^{17}$ neutrons cm^{-2}; N_3, 40 hr at 350° C in vacuum; P_0, original P-type InSb; P_1, after 4×10^{16} neutrons cm^{-2}; P_2 after 5×10^{17} neutrons cm^{-2}; P_3, after 40 hr at 350° C in vacuum.* (*Cleland and Crawford*[41].)

0·015 eV below the conduction band and is temperature independent within the errors of interpolation.

Examination of the bombardment behavior of InSb suggests three groups of fast neutron-induced defect-states:

(1) A shallow group of hole traps or occupied states, which lie below the middle of the gap, as indicated by the strong curvature of the σ *v.* nvt_f curve and the appearance of the log R *v.* $1/T$ curve for P-type specimens before conversion.

(2) A second group of occupied states which, since they act as donors in converted material, must lie above the center of the gap.

(3) A group of shallow electron traps (vacant states expected to trap electrons when $n > n^*$).

Because of the binary nature of the compound, there are four possible simple defects available to explain these levels. Cleland and Crawford[41] have speculated

about the nature of these states. It was concluded that, because of the large dielectric constant ($\kappa = 16$) and the small effective mass for electrons, the first ionization potential of both interstitial indium and antimony is nearly the same and quite small ($\sim 0{\cdot}05$ eV). However, the two types of vacancies may behave quite differently. The indium vacancy has five electrons extending into it from the four antimony neighbors which form three half-filled and one completed Sp^3 orbitals. The antimony vacancy, on the other hand, has only three electrons associated with it. If a vacancy can be considered to some extent analogous to an atomic system, then by Hund's rule one would expect less energy to be required to place an electron into the antimony vacancy than into the indium

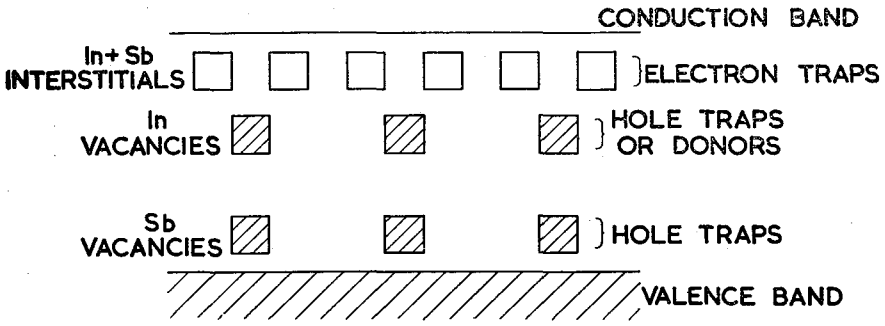


Figure 9. *A possible defect energy level scheme for irradiated InSb.* (*Cleland and Crawford*[41]).

vacancy. Since the gap is quite narrow ($\sim 0{\cdot}2$ eV at room temperature), it is reasonable to assume that the second ionization state of each type of defect lies outside the forbidden band and need not be considered further. After redistribution of electrons to states of lowest energy, the energy-level scheme of Figure 9 results, consisting of a set of shallow hole traps arising from the antimony vacancy, a set of hole traps or donor states from the indium vacancy, and finally a set of vacant states, twice as numerous as the occupied levels, just below the conduction band which arises from the indium and antimony interstitials.

There is an additional possible type of lattice damage in these binary compounds which does not exist for elemental semiconductors. Except for the relatively large covalent binding energy, InSb and the other III–V compounds are somewhat analogous to an ordered superlattice alloy. Should extensive atomic interchange result from bombardment by the displacement mechanism of Kinchin and Pease[10], as is the case for Cu_3Au[43], extensive electronic effects would result. In a disordered InSb system the basis for the Sp^3 covalent orbitals would be largely destroyed, and the open diamond structure would be expected to collapse to a more close-packed metal-type lattice. Such a collapse would undoubtedly be accompanied by a transition to metallic conduction. Attempts to observe superlattice disordering in InSb by fast neutron bombardment have been made using both X-ray diffraction techniques after neutron exposures up to 10^{21} n cm^{-2} and electrical measurements after neutron exposures up to 10^{18} n cm^{-2} near room temperature, but with negative results. Irradiations up to 5×10^{17} n cm^{-2} carried out at temperatures as low as $-150°$ C to reduce

the mobility of defects were also unsuccessful. Hence, it appears that the disordered structure is too unstable to be frozen in at these temperatures. Perhaps the increase in available free volume in a highly disordered region allows the lattice to re-order without the annihilation of all of the point defects.

In addition to fast neutron studies, results of electron bombardment have been reported[44]. These experiments were carried out at 90° K and are complicated by minority carrier trapping effects. Because of this and the small range of property changes investigated, they cannot be directly compared to the fast neutron results.

4.4. Gallium Antimonide

The behavior of GaSb on irradiation[45] is quite similar to that of germanium. In view of their close electronic similarity (comparable band gap, *etc.*), this is not surprising. A detailed study of the defect level structure is hampered by the fact that high-resistivity specimens of this material have not yet been obtained. Fast neutron bombardment was found to convert N-type specimens to P-type, and the conductivity of P-type specimens approaches a saturation value. There is evidence of a hole trap 0·14 eV above the valence band and an acceptor $\sim 0{\cdot}2$ eV from the valence band. Prolonged exposure near room temperature and at −125° C gave no electrical evidence of the superlattice disordering discussed in the preceding section.

5. ANNEALING OF BOMBARDMENT PRODUCED DEFECTS

5.1. Early Indications

Early studies[3, 4, 18] of neutron and deuteron bombarded Ge and Si demonstrated that essentially all of the bombardment-introduced lattice disorder could be removed by appropriate heat treatment. For example, the original conductivity of Ge (except for the small permanent change resulting from nuclear doping) could be restored by annealing at 450° C for 24 hr. Similarly, irradiated Si was annealed but at a somewhat higher temperature. Since the defects were expected to be predominantly of the Frenkel type, a naïve view suggests that simple interstitial–vacancy annihilation is responsible for restoration of the crystal perfection.

The first annealing study made by Brattain and Pearson[46] on Ge bombarded with polonium α-particles indicated a first- rather than a second-order relaxation process. They found that the curvature of the σ *v.* exposure curve for room temperature was consistent with a first order build-up process and that the relaxation subsequent to exposure was described by the same rate constant provided it was assumed that a certain fraction of the acceptors did not anneal. In addition, Cleland *et al.*[18] found that the variation of σ in P-type germanium, with fast neutron exposure near room temperature, could be described by a first order build-up and saturation process. It was also pointed out, however, that a similar curvature was observed if the exposure were carried out at 195° K, and subsequent studies have shown that the dependence of change in carrier concentration on φ accounts for most of the curvature. More recent studies indicate that the annealing processes are indeed complex.

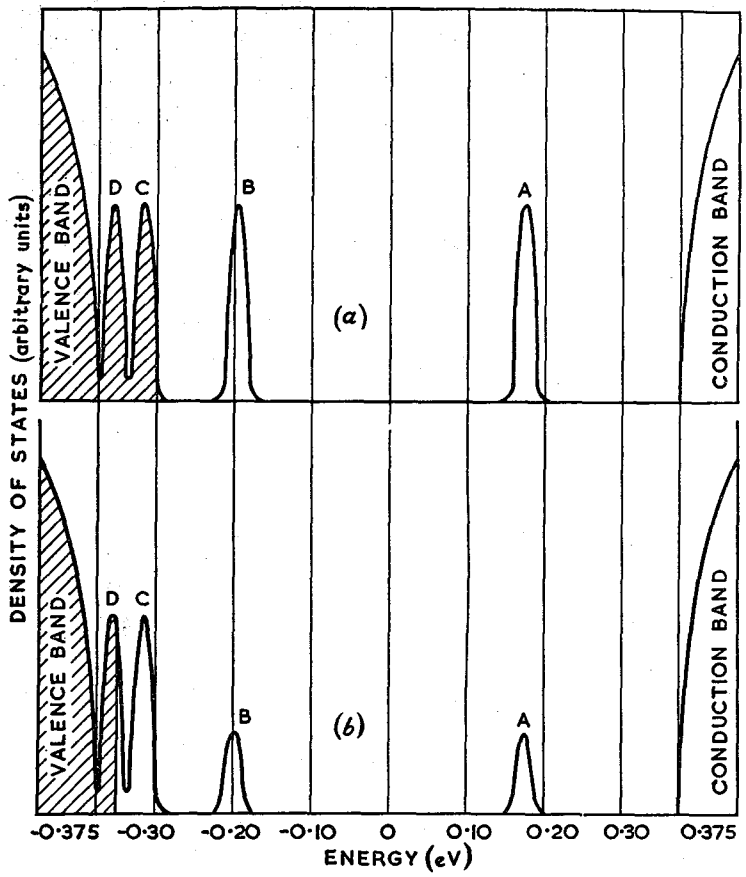


Figure 10. *A diagram of the distribution of energy levels produced in germanium by bombardment.* (*a*) *indicates the distribution expected for equal concentrations of interstitial and vacancies.* (*b*) *shows the effect expected to result from partial interstitial removal.* (*Cleland et al.*[26])

5.2. Preferential Annealing of Interstitial in Germanium

It was mentioned in sub-section 4.1 that for germanium the Frenkel defect energy level model breaks down after appreciable exposure, presumably as a result of alteration of the relative proportions of the various defect-states caused by preferential annealing of one type of defect. This is demonstrated[26] by the fact that the temperature dependence of hole concentration after appreciable exposure is considerably less at low temperature than suggested by the energy-level model suggesting the appearance of more shallow acceptors as irradiation proceeds. Moreover, subsequent ageing at room temperature and annealing at low temperature cause an increase in p and a further decrease in the slope of the log R *v.* $1/T$ curve at low temperature. This suggests that the interstitials, which are evidently the more mobile component, are removed by some process,

e.g. clustering or migration to dislocations taking their electrons with them, thereby emptying the states associated with the vacancies so that these are now available as acceptors, with a much smaller ionization energy than the ones they replace. The situation both before and after interstitial annealing is shown schematically in Figure 10. Irradiated N-type specimens give similar indications[25], since a light anneal near 100° C produces a further decrease in electron concentration, suggesting that the vacant state 0·2 eV below the band edge is replaced by a much deeper one which is more effective in trapping electrons. Similar effects have been observed with deuteron and α-particle bombardment. The behavior of deuteron-bombarded P-type germanium at 200° K[28], which was mentioned in sub-section 4.1, is particularly striking and leads to a smaller apparent value of φ^*.

A similar preferential annealing has been indicated in GaSb[45]. In silicon, however, although its annealing has not been extensively investigated, there is no definite evidence of such a process. Instead, annealing of a P-type specimen near 200° C appears to cause a continuous decrease in ionization energy from the intrinsic value induced by prolonged neutron exposure[37].

5.3. Low Temperature Relaxation after Neutron Bombardment

Defect relaxation during and subsequent to fast neutron exposure has been observed at temperatures as low as −160° C[25, 26]. Although this is masked to a considerable degree by minority carrier trapping effects (see sub-section 6.4), thermal cycling in the presence of radiation has revealed that irreversible changes occur which cannot be attributed to changes in the non-equilibrium distribution of charge in traps but rather to the disappearance of the traps themselves. These low temperature relaxation effects raise the question of the identity of those defects remaining at room temperature. Presumably, if sufficient defect mobility exists at these temperatures for the annihilation of simple defects, these would not be expected to be stable at or above room temperature. Since the room temperature disorder during the initial stages of fast neutron bombardment is adequately described by the Frenkel defect model, this state of affairs does not seem likely. In the case of massive-particle bombardment (fast neutrons and deuterons), the question may be resolved in the following way. Because of the highly localized nature of the damage production, it has been suggested that, at the lower temperatures where atomic motion is restricted, the lattice disorder around the point of a primary collision is so extensive that individual defects can no longer be distinguished, and the structure of the region is best described as approximating to that of a liquid. As the temperature is increased the disordered region may rearrange to configurations of successively higher thermal stability leading to a large local concentration of point defects well below room temperature. Such an explanation, however, will not suffice for electron bombardment for which low temperature relaxation is also observed.

5.4. Annealing of Electron Damage in Germanium

Klontz[13] has observed extensive annealing at 0° C and room temperature of the damage introduced into N-type germanium by Van de Graaff electron

bombardment at 77° K. The recovery of resistivity followed a first order process with a relaxation time of 2,300 sec at 0° C. Further annealing at 70° C caused a marked decrease in the slope of the log R *v.* $1/T$ curve. The first-order mechanism suggests that the process at 0° C involves recombination of close pairs of interstitials and vacancies. Such a process is not unreasonable since, because the energy transferred during a collision is not much greater than the threshold energy, an appreciable concentration of relatively closely spaced interstitials and vacancies should be produced.

By far the most comprehensive study of annealing of radiation damage has been performed by Brown *et al.*[47] on N-type germanium exposed to 3 MeV electrons at room temperature. Isothermal anneals were carried out in the range

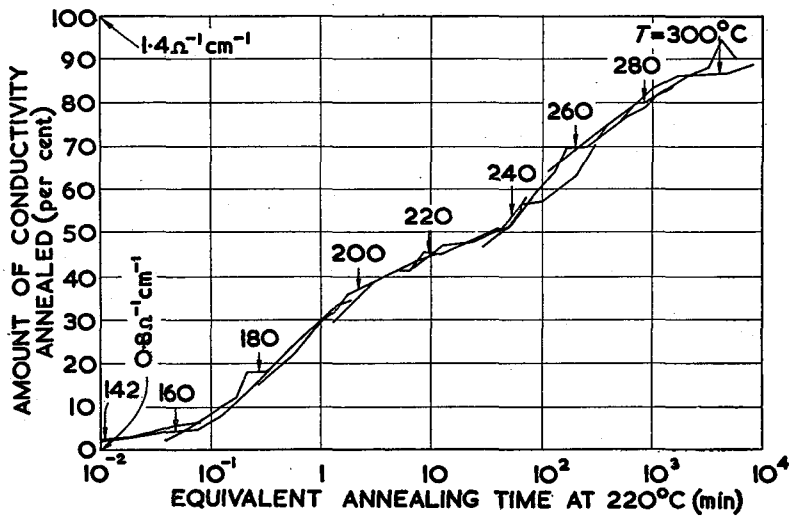


Figure 11. *Composite annealing curve obtained by adjusting the time scales of annealing curves at different temperatures, for a group of samples having the same initial conductivity and given the same bombardment, for superposition with respect to the curve at 220° C. The arrows above the curves indicate the ten minute annealing points for the original curves.* (*Brown, Fletcher and Wright*[27])

from room temperature to 340° C. The progress of annealing was followed by means of σ and in every case a monotonic decrease in $\triangle\sigma.(= \sigma_0 - \sigma)$ induced by bombardment was observed. On plotting the fraction of $\triangle\sigma$ annealed *v.* log time, it was found that, although the character of these curves varied from one end of the temperature range to the other, they could be joined to give a continuous composite curve by adjusting the time scale appropriately. Since the abscissa is logarithmic, the adjustment involves only a horizontal shift of the curves for matching. The composite curve in which the time scales for all annealing runs are adjusted to that at 220° C is shown in Figure 11. The fact that superposition of these curves is possible suggests that time is involved only in the ratio t/τ, where τ, the temperature-dependent relaxation time, is inversely proportional to the adjustment factor. The activation energy for the

annealing was obtained from a semi-logarithmic plot of τ_{220}/τ_T against reciprocal temperature. This curve is a reasonably good straight line with a slope corresponding to 1·8 eV at high temperature and 1·6 eV at the low end of the range.

In a second theoretical paper[48], these authors have attempted to account for the observed annealing behavior in terms of known processes. They assume that only one component of the Frenkel defect is mobile, *viz.* the vacancy, and they reason that as long as the vacancy is within the stress field of the interstitial there will be an activation energy gradient toward the interstitial causing a high probability of recombination after the vacancy makes one jump. Outside the critical radius circumscribing the strained region surrounding the interstitial the vacancy is expected to move in an undistorted lattice with the characteristic activation energy for vacancy motion. These two regions are treated separately. The first of these corresponds to the recombination of close pairs by a simple first-order process, characterized by the activation energy of the first jump toward the vacancy. The second annealing range refers to the capture of vacancies after random-walk migration by distant vacancy sinks such as other interstitials, dislocations and small-angle boundaries. The best fit to the data was obtained using only the close pair recombination, and recombination during liberation. It is not surprising that the other sinks do not enter since in good single crystals the dislocation density is quite small ($\sim 10^4$ lines cm^{-2}).

The combination of first and second order processes accounts adequately for the very long extension in time of the observed annealing behavior. Moreover, the only process, considered by theory, which is thermally activated, is the migration of the mobile defect. This is also in agreement with observation and the small variation of the experimental activation energy with temperature is adequately accounted for by the barrier lowering in the vicinity of the interstitial for the first order process.

Although these authors arbitrarily assumed that the vacancy is the more mobile component, work cited earlier in this section suggests that the interstitial does the moving. Consequently, perhaps their arguments should be applied to the interstitial rather than the vacancy. The alternative description, however, introduces one difficulty which they have pointed out[48]. Whereas, because of the lattice strain, the easy direction of motion of the vacancy is toward the interstitial, such may not be the case for an interstitial moving in the stress field of a vacancy. Here the easy direction of motion would appear to be tangential to the vacancy rather than toward it. Such a state of affairs would tend to impede the first order process, a behavior not experimentally observed. It would appear to be worthwhile to re-examine the annealing behavior in the light of this complication. In this connection, 'reverse' annealing, resulting from the preferential removal of one type of defect discussed in connection with neutron bombardment studies, was not observed in these experiments, whereas recent studies by Cleland and Crawford[30] on γ-ray bombarded germanium unmistakably indicate that such a process does occur on annealing near 100° C. Since randomly distributed defects are expected in each case, the reason for the difference in annealing behavior is not clearly understood. Perhaps the 'reverse' process in the electron bombardment experiments occurred to a large extent during exposure because of heating of the specimens. If this is the case, the concentration of interstitials and vacancies would no longer be equal.

6. EFFECTS ON VARIOUS PROPERTIES

6.1. Mobility

The rate of carrier removal during exposure is usually obtained from the initial slope of the σ *vs.* exposure curve under the assumption that the mobility remains constant. This assumption is at best an approximation which is valid only in the lattice scattering range and becomes progressively worse the greater the impurity scattering contribution to the mobility. In the range where impurity scattering alone limits the mobility, the change in σ caused by the decrease in mobility may be comparable to or even greater than the variation produced by the change in carrier concentration. Consequently, it is best where possible to use the Hall coefficient measurements to determine changes in carrier density.

According to the Frenkel defect energy level model of James and Lark-Horovitz[20], the redistribution of electrons from interstitial atoms over the available localized states leads to charged centers or sites even before the charge concentrations are altered. Hence, depending on the distribution of levels in the energy gap, two doubly charged sites as in the case of germanium, or two singly charged sites as may be the case for silicon, are expected to result from the presence of the defects alone. For N-type germanium under conditions where both vacant states are occupied, the interstitial is no longer charged and only the doubly charged vacancy is available for scattering carriers. In P-type germanium, on the other hand, the interstitials and vacancies may be either singly or doubly ionized depending on φ. Because of the different relative arrangement of states in silicon, doubly ionized defect states are not expected except at very low temperatures, and only singly charged and neutral defects are normally expected.

In view of the fact that the scattering probability of a charged scattering center is approximately proportional to the square of the charge z, the behavior of the impurity mobility μ_I might also be expected to reflect the nature of the defect energy level structure. There are, however, two complicating factors which must be taken into account. The first of these arises because the carrier concentration in irradiated specimens is not equal to the concentration of scatterers. Hence, the simple scattering theory of Conwell and Weisskopf[49] is no longer applicable. Instead, the Brooks–Herring theory[35] which accounts for the extent of screening of the scattering charge by free carriers must be used. As long as the defects are randomly distributed as in the case of electron bombardment this presents no problem, but for the localized defect distributions expected for massive particles serious complications may arise. The second difficulty is associated with the extensive preferential annealing behavior discussed in the previous section. Consequently, the defect model cannot be applied over an extended range of exposure to massive particles, though some success with electron or γ-ray bombardment might be expected.

The most extensive study of the effect of radiation on carrier mobility has been performed on germanium bombarded with fast neutrons at room temperature[25, 26]. Values of μ as a function of temperature were obtained from measurements of ρ and R after successive exposures of both N- and P-type specimens. The most surprising aspect of the study is the markedly different effect on the temperature dependence of μ for electrons as compared to that for holes. The

results are summarized in Figures 12 and 13 in which log μ is plotted against log T after successive exposures for N-type and P-type germanium respectively. The temperature dependence for electrons has the proper sign, but exhibits a greater than normal magnitude in the low-temperature, impurity scattering range. Since the electron concentration varies strongly with temperature after appreciable exposure, the enhanced temperature dependence (as great as T^6 before conversion in a specimen with $n_0 = 7 \times 10^{17}$ cm^{-3}) is consistent with a

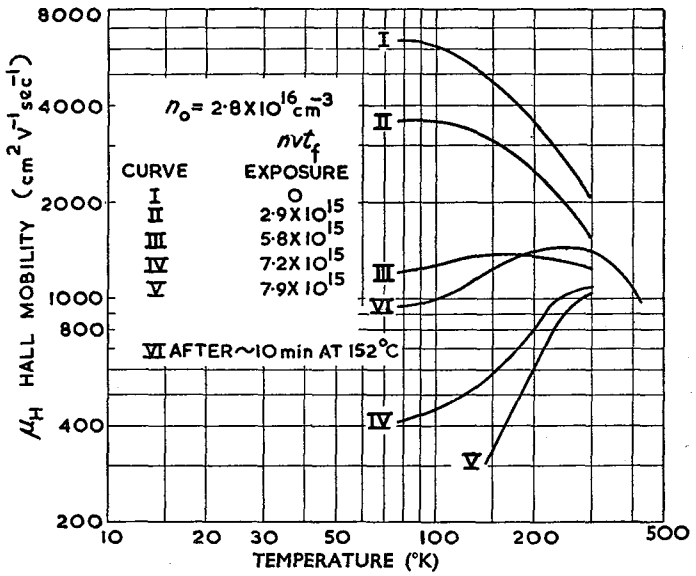


Figure 12. *Hall mobility v. temperature for N-type germanium after the indicated exposures, for a specimen with* $n_0 = 7 \times 10^{17}$ *cm*$^{-3}$. *(Cleland et al.*[25])

rapid decrease in screening and a corresponding increase in charged center scattering probability according to the Brooks–Herring theory[35].

In contrast, the hole mobility is decreased by a factor which varies by not more than a factor of two over the entire temperature range. Moreover, there is no evidence of the maximum which moves to higher temperature as expected for an increase in the charged impurity concentration. Even in P-type specimens variation in screening is expected to play a role, since the rate of increase in p is less than the rate of defect introduction. However, there is no evidence that such is the case.

The reason for this peculiar disparity in behavior is not clear. However, there is one possible explanation based on the localized, non-random distribution of fast neutron produced defects which seems to account for the difference. As mentioned above, the primary recoil from a fast neutron collision, because of its restricted range, loses its energy by producing secondary displacements within a relatively small volume. If it is assumed that the mobility in this region is strongly decreased, a crystal containing many such regions could be treated

to a first approximation as having a smaller effective cross-section for conduction. The decreasing mobility would then approach the value characteristic of the highly disordered region, and saturation would occur as the damaged regions begin to overlap. This description seems to account for the apparent mobility decrease over the entire range in P-type germanium. The different

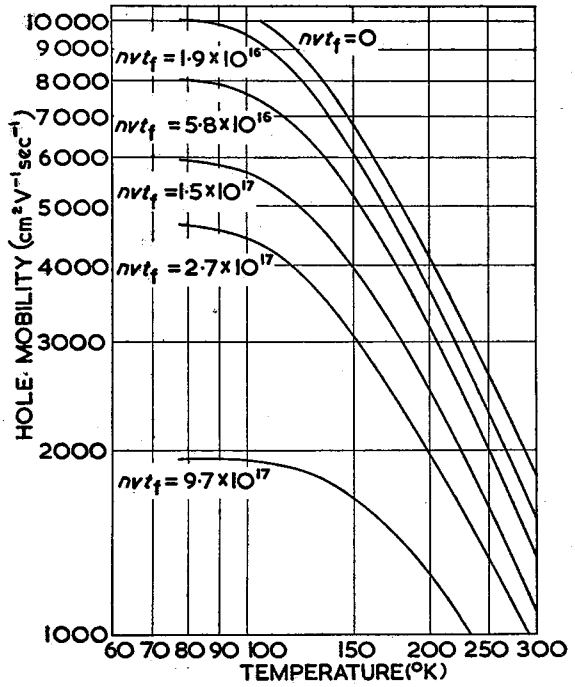


Figure 13. *Hall mobility v. temperature after successive exposures for P-type germanium. The exposures are indicated on the curves.* (*Cleland et al.*[26])

behavior of N-type material may then be understood as follows. Although the damaged regions themselves do not contribute to impurity scattering, they trap electrons from the undistorted material and thereby reduce the screening of the charged chemical impurities already present causing an increase in the scattering probability. In P-type specimens, on the other hand, since the disordered regions increase the hole concentration, screening is increased rather than decreased with the result that the impurity contribution to the mobility decreases slightly instead of increasing.

It should be pointed out that the explanation offered here is considerably over-simplified. The size of the damaged regions will vary widely depending on both the neutron energy spectrum and the amount of energy transferred to the recoiling atom during a primary collision. Moreover, what will be the effect of annealing within the damaged regions is not understood, though such an effect would be expected to level the defect distribution to some extent (Curve VI, Fig. 12). Perhaps a better description of the defect distribution would be in

terms of randomly dispersed damaged regions of widely varying size superposed on a small but uniform distribution of individual defects. However, this naïve explanation does appear to have some merit. Treating the apparent rate of mobility decrease at 300° C for P-type germanium according to this picture leads to a damaged region whose average size is $\sim 10^5$ atoms, a value which is not inconsistent with the estimate of Brooks[12].

There are, no doubt, alternative explanations of the different effects of neutron bombardment on electron and hole mobilities. For example, the non-random defect distribution is reflected by a localized perturbation of the floor of the conduction band and a corresponding fluctuation of the valence band edge. The magnitude of these fluctuations will depend on the position of the Fermi level in the undamaged region to that in the region of high defect density. Since in the absence of preferential annealing φ in the damaged region is expected to approach φ^*, the perturbations in band edge and hence extraneous scattering are considerably greater in initially N-type specimens than in P-type germanium.

Whatever the proper explanation of the effect on mobility, it is recognized that the above pictures are inconsistent with the energy-level analysis of P-type material given previously. If the non-randomness of the defect distribution is so pronounced, it is questionable whether the analysis in terms of thermal equilibrium of carriers with the localized states of uniformly distributed defects can be more than a rough approximation. Such being the case, the excellent agreement of the defect-level positions obtained in neutron-bombarded specimens using the Frenkel defect model with those obtained for electron and γ-ray bombardment is indeed surprising.

In contrast to the neutron results, the mobility in electron[29], γ-ray[30], and deuteron[28] bombarded specimens of germanium appears to be consistent with the charged impurity scattering theory. This suggests that even in the case of deuteron exposure a more uniform damage distribution results. Additional work is needed before the results can be correlated quantitatively with theory and the questions raised in connection with neutron bombardment can be answered conclusively.

6.2. Optical Absorption

The effect of fast neutron bombardment on the infra-red absorption spectrum of both silicon and germanium has been investigated with the most extensive studies being made on silicon. The first studies on silicon were made by Fan and Becker[50] on relatively impure polycrystals. They observed a shift of the absorption edge to longer wavelengths and a discrete absorption which sharpened and moved to shorter wavelengths with decreasing temperature. In the specimens they investigated, the neutron exposure was sufficient to reduce the conductivity to near the intrinsic value, so that the free carrier absorption is almost entirely eliminated thus making visible the 9 μ band which had been previously observed only in much purer specimens.

Recent work by Longo and Kleitman[51] on high-purity single-crystal specimens confirms and extends the earlier work of Fan and Becker. The absorption spectra they obtained is shown in Figure 14. The band at 1·75 μ increases monotonically in intensity with exposure and additional absorption between

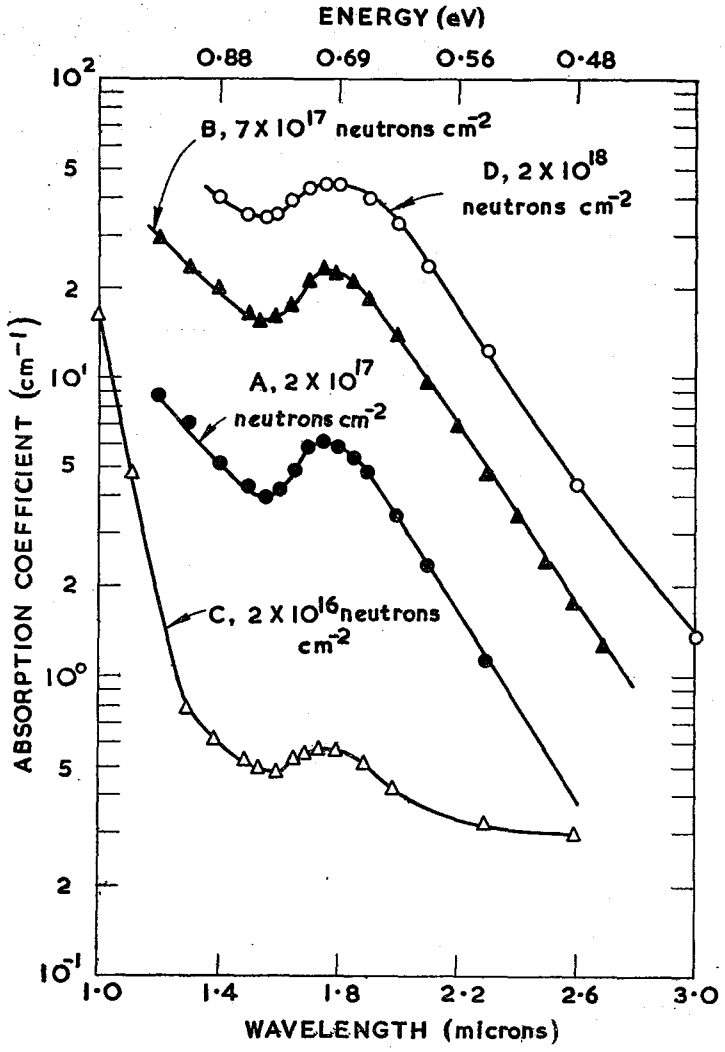


Figure 14. *Optical absorption v. wavelength of neutron-bombarded silicon at room temperature. (Longo and Kleitman[37]).*

1 μ and 1·5 μ also increases with exposure. The temperature variation of the 1·75 μ (0·71 eV) band, which is quite similar to that of the F-band in colored alkali halides, as was pointed out by Fan[50], indicates that it is associated with an optical transition between localized states and hence must represent photon absorption by an electron localized at a defect exciting it to a state 0·71 eV above the ground state. In view of the fact that the energy separation between the defect-states proposed on the basis of electrical evidence is approximately this value it is tempting to identify the optical transition with excitation of electrons from low-lying interstitial and vacancy levels to the upper ones.

However, as was pointed out by James and Lark-Horovitz[20], the energy separation between the states of the interstitial is the difference between the first and second ionization energies and in the case of the vacancies, this separation represents the difference in energy required to place one and two electrons in the vacancy. Therefore, these states do not exist simultaneously, but their existence depends on the state of ionization of the defect. In consequence, the upper pair of levels of the diagram (Figure 6) bears no relationship to the excited states of the electrons bound to the defects. Since there is negligible photoconductivity associated with the 1·75 μ transition (see sub-section 6.3) it is assumed that the transition involves a singly ionized interstitial atom whose lowest energy electron lies in the level B of Figure 6. Excitation of an electron bound to a vacancy in level C on the other hand would place it in a state very near the conduction band and perhaps lead to enhanced conductivity.

The additional absorption beyond the absorption edge (1 to 1·5 μ) also increases monotonically with exposure and also increases nearly linearly with decreasing wavelength. At low temperature (77° K), the magnitude of the absorption coefficient in this spectral range is decreased. Although available information is not sufficient to permit a complete interpretation, the following points are significant:

(1) Absorption in this range is accompanied by appreciable photoconductivity at both low temperature (105° K) and at room temperature (see sub-section 6.8), suggesting that electrons or holes or perhaps both are produced either by direct excitation to the respective bands or by thermal ionization subsequent to excitation to shallow states.

(2) The decrease in absorption with temperature indicates either that a portion of the absorption results in transitions to states near the band edges followed by thermal ionization or that release to the conduction band is assisted by a phonon interaction leading to an indirect transition similar to the case for band-to-band transitions in which the crystal momentum K is not conserved.

(3) One interpretation of the energy dependence of absorption is that the localized states involved in the transition are distributed in energy with respect to the band edges. These points will be discussed in greater detail in the next section.

In one case Longo and Kleitman[51] examined the absorption after successive annealings of neutron-irradiated Si in the range of 150–250° C. Anneals at 150° C caused a small but perceptible decrease in the bombardment-induced absorption over the entire spectral range. Additional annealing near 200° C causes a much larger decrease in the absorption from the absorption edge to and including the 1·75 μ peak but beyond this point an absorption tail is developed. The tail increases with additional annealing and, after the 1·75 μ band is no longer visible, the tail becomes quite flat extending out to the longest wavelength used (150 μ). Although the nature of this continuous absorption produced by the annealing is far from completely understood, it can be taken as evidence of the complicated thermal relaxation processes that occur in neutron-bombarded silicon on annealing in the intermediate temperature range (200–250° C), *i.e.* complex clusters of defects of an unspecified nature which develop during annealing may be responsible for the flat absorption tail.

In contrast to silicon, the absorption spectra of irradiated germanium has little definite structure. Rather, only additional absorption beyond the absorption edge has been observed. Lipson, Burstein and Smith[52] have compared this to absorption in deformed specimens and have shown that the shift of absorption edge toward longer wavelengths, which occurs in plastically deformed specimens, is not observed in irradiated specimens (Figure 15). Speculations as to the nature of the fast neutron-induced absorption will be deferred until the next section.

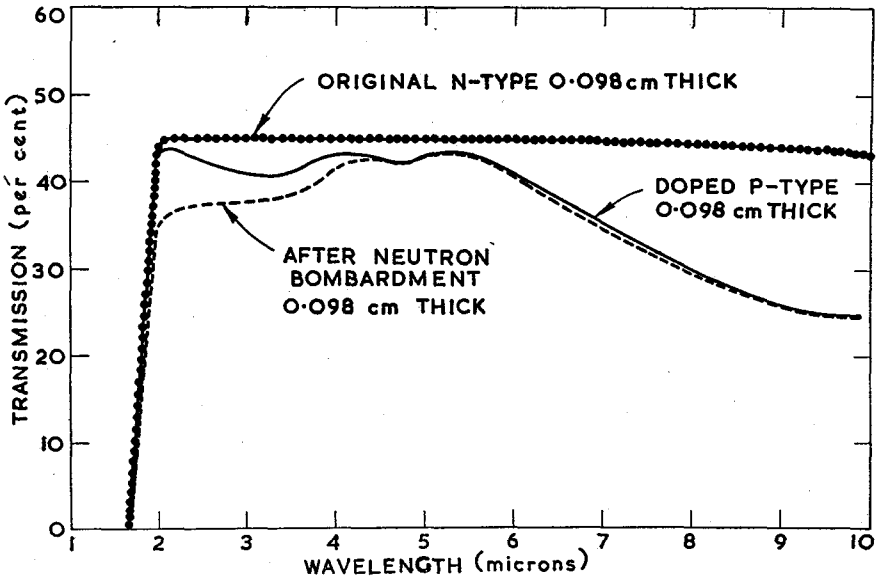


Figure 15. *Transmission of germanium specimens converted from N- to P-type by neutron irradiation compared with an un-irradiated P-type specimen of the same final charge carrier concentration.* (*Lipson et al.*[37])

It might be pointed out in this connection that the bombardment-induced absorption in the diamond lattice elements appear to become simpler, with less structural detail, as the mass of the element increases. For example, studies by Ditchburn *et al.*[53] of bombarded diamond show that the additional absorption resulting from reactor exposure has considerable structural detail. Moreover, the detail is considerably sharpened after high-energy electron bombardment, a fact which is expected to result from the more uniform distribution of defects. In this last connection, it appears that electron bombardment should be employed in future studies of silicon and germanium to bring out greater detail in the defect absorption.

6.3. Spectral Dependence of Photoconductivity

Since photoconduction in semiconductors may arise from two distinct processes: (1) excess majority and minority carriers after direct band-to-band

transitions; and (2) carriers resulting from transitions involving localized states in the forbidden band, we shall treat these separately. In this section only the latter type, *i.e.* that arising from photon energies less than the band gap, will be considered.

Studies of the spectral dependence of photoconductivity in silicon exposed to both fast neutrons and 9·6 MeV deuterons have been made by Longo and Kleitman[51]. Typical results at 169° K on an initially P-type specimen ($p_0 = 1{\cdot}6 \times 10^{15}$ cm^{-3}), after receiving 7×10^{17} n cm^{-2} are shown in Figure 16.

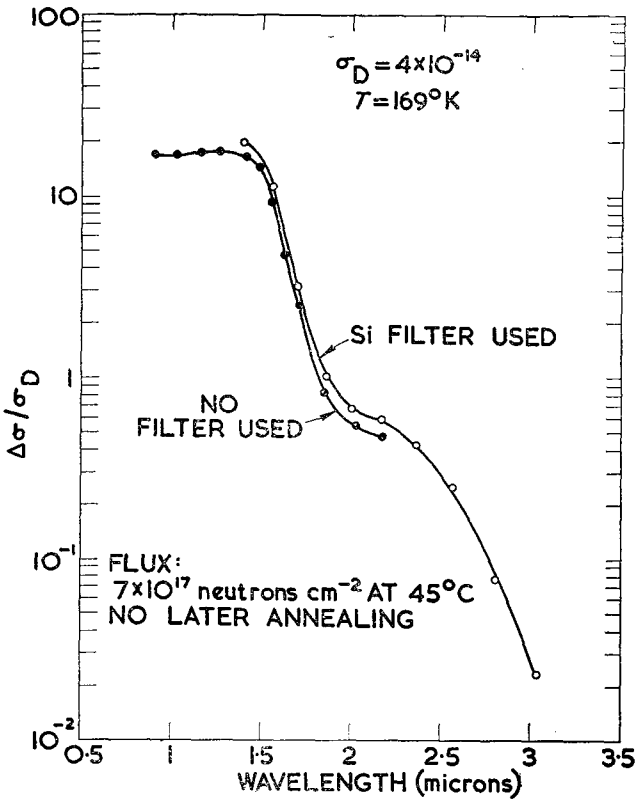


Figure 16. *Photoconductivity v. wavelength of neutron-irradiated silicon at 169° K. (Longo and Kleitman[37].)*

From 1 to 1·5 μ $\triangle\sigma/\sigma_0$, the relative photoconductivity, is high and flat. In the region of the 1·75 μ absorption band, *i.e.* from 1·5 to 2 μ, the photoconductivity drops sharply and exhibits a shoulder beyond 2 μ from which it falls rapidly to a negligible value beyond 3 μ. The behavior of a deuteron-bombarded specimen was quite similar. Annealing in the range from 200 to 250° C decreases the photoconductivity over the range from 1 to 2 μ but produces a tail extending past 4 μ, an effect which is similar to that for the optical absorption. The implications of the spectral response of photoconductivity have already been indicated

in the previous section. The high flat response between 1 and 1·5 μ is presumably associated with either excitation from a low-lying state to one near the conduction band, direct excitation from the low-lying state into the conduction band or from the valence band into a localized state. Measurements of the Hall coefficient on specimens exposed to 1·36 μ light indicate that both electrons and holes are involved in the excess conductivity. The reason for the response above 2 μ in irradiated specimens before and after annealing is not understood.

The spectral response of photoconductivity has also been studied in irradiated germanium, but here measurements have been confined primarily to electron-bombarded specimens. Fan *et al.*[54] first reported the spectral behavior of specimens exposed to 4·5 MeV electrons at room temperature. They found that, for exposures not large enough to produce conversion to P-type but great enough to empty all of the chemical donors, photoconductivity existed far beyond the fundamental absorption edge out to photon energies of 0·24 eV and hence was presumably due to direct transition from a state $\sim$ 0·24 eV below the band edge into the conduction band. The relaxation time for the process at 83° K was about 0·5 sec. In N-type specimens, bombarded with 6 MeV electrons at 90° K, Stoeckman[55] *et al.* found evidence for two photoconductive absorption bands, one at 0·55 eV and one at 0·35 eV. After conversion to P-type the 0·35 band disappeared and illumination in the 0·55 eV band decreased the conductivity. Once again the relaxation time was of the order of 1 sec. On warming to room temperature, the spectral response was altered yielding a behavior similar to that reported by Fan *et al.*[54] This change on warming is another indication that extensive relaxation effects occur well below room temperature even in electron-bombarded specimens.

These results, particularly the latter, are consistent with the Frenkel defect energy level of James and Lark-Horovitz. The two levels at 0·35 and 0·55 eV may tentatively be identified with levels A and B of Figure 4 as occupied states of the interstitial atom.*

After conversion to P-type, the shallow state of the interstitial should no longer be occupied and, indeed, it is observed that photoconductivity at this wavelength disappears. Moreover, as Brooks[12] has pointed out, illumination in the 0·55 eV band quenches photoconductivity resulting from minority carrier trapping; thus it appears that the second interstitial state is an electron trap in P-type germanium.

The results of Stoeckman *et al.* have recently been confirmed for deuteron bombardment at 90° K by Kleitman and Longo[56] who, in addition have observed a negative photoconductivity peak at 0·35 eV shortly after conversion to P-type.

6.4. Minority Carrier Phenomena

Because of the extensive lattice perturbations produced in the diamond lattice semiconductors by fast particle bombardment, it is perhaps not surprising

* The reason for the larger ionization energy of the shallow interstitial state as indicated by photoconductivity is not completely understood. However, since the minimum of the conduction band in Ge does not occur at $\mathbf{k} = 0$, a $\mathbf{k} = 0$ transition from the localized state to the band would require a greater energy than that to the minimum of the band.

that marked effects on the behavior of excess carriers result. These effects are twofold:

(1) Irradiation causes a marked reduction in minority carrier lifetime, *i.e.* the relaxation time τ for recombination of an excess electron–hole pair; and

(2) bombardment-produced metastable minority carrier traps are introduced.

Although these two effects should not be considered entirely independent, they will be treated separately for convenience.

The drastic reduction of τ in Ge was first observed[57] while attempting to use charged particle bombardment to produce barrier layer configurations suitable for transistors. The impairment of τ was so great that transistor action was inhibited. The consequences of reduction in τ have also been observed in irradiated junction transistors[58], wherein the decrease in bulk lifetime causes the amplification factor to fall to a small value after fast neutron exposures of $\sim 10^{12}$ n cm^{-2}. Florida *et al.*[59] have used this effect to improve the switching characteristics of germanium point-contact transistors by reducing the hole storage of the base material. They also report τ measurements on irradiated bulk specimens. A decrease in τ from 1·6 to 0·6 microsec was observed in a 2·03 Ω cm specimen after 10^{14} n cm^{-2} exposure in the Harwell reactor. A decrease in τ from 50 to 6 microsec in 6 Ω cm material was also reported. The use of lifetime as an index of damage in both germanium and silicon by Loferski and Rappaport[14] was discussed in sub-section 2.2.

The nature of the recombination centers in irradiated germanium and silicon has not been thoroughly investigated. Consequently we may only speculate as to their identity. Because of their multi-level structure, the isolated defects themselves would be expected to act as recombination centers at or near room temperature. This conclusion seems reasonable since it is well known that under appropriate conditions chemical impurities with a multiple-level structure act as efficient recombination centers. In the case of fast neutron or massive charged-particle bombardment, there is an additional possibility which bears consideration, namely, the localized nature of damage distribution may lead to sufficient lattice distortion in these regions that the usually applicable selection rules for band transition are to a large extent relaxed and, therefore, the rate of recombination of excess electrons and holes is markedly enhanced. This latter hypothesis may be subjected to test by studying the differences of recombination behavior in specimens exposed to both electrons and massive fast particles.

Photoconductivity associated with trapping of minority carriers, which has been studied extensively in un-irradiated specimens, has also been observed in irradiated germanium. Shulman *et al.*[60] have investigated the kinetics of minority carrier trapping in N-type Ge exposed to high energy electrons at room temperature. Using the technique of Haynes and Hornbeck[61], they found that in N-type specimens bombarded at room temperature, the hole trap concentration was comparable to the concentration of electrons removed during bombardment. The capture cross-section of the traps was estimated to be 2×10^{15} cm^{-2} and decay curves were simple exponentials corresponding to a trap depth of 0·25 eV.

In specimens bombarded at 78° K both electron traps and hole traps were

observed[62]. The concentration of hole traps introduced by the bombardment was considerably less than the total number of defects, accounting for only 3 per cent of the conductivity change. The hole traps anneal out at 140° K. In P-type specimens no electron traps are observed unless the energy of the bombarding electrons is 2·5 MeV or greater. For such energies, however, two classes of electron traps with different capture cross-sections are produced. Temperature dependence of the decay curves indicates an electron trap 0·2 eV below the conduction band in agreement with equilibrium carrier concentration measurements (sub-section 4.1). The electron traps anneal at 185° K. Recently Klontz *et al.*[63, 55] have confirmed this work with 4·5 and 6 MeV electron bombardment of germanium. They find time constants for hole traps in N-type germanium of ~ 1 sec at 90° K. The photoconductivity decreases when the specimens are converted to P-type indicating that either the effectiveness of electron traps or their concentration is considerably less than those observed for hole traps in N-type germanium.

Studies of germanium exposed to fast neutrons[25, 26] at low temperature (120° K) also reveal that minority carrier traps are introduced into both N- and P-type germanium. These traps anneal out well below room temperature. Moreover, as would be expected, deuteron bombardment also appears to introduce minority carrier traps[28].

The minority carrier trapping processes produced by bombarding germanium raise some interesting problems which have yet to find solution. For example, it is not known why these traps anneal at low temperature, particularly in the case of electron bombardment which is expected to lead to a random distribution of defects. Moreover, the threshold energy (2·5 MeV) of bombarding electrons required for the production of electron traps in P-type specimens is indeed curious and suggests that closely spaced defects arising from secondary displacements are necessary for the formation of a metastable electron trap. In any event, these processes are not well correlated by the energy level model applied to defects produced at room temperature and may be taken as further evidence that the current model is over-simplified. In order to account for these and other inconsistencies in behavior, it may prove necessary to postulate the existence of energy levels associated with complexes of defects as well as the localized states of isolated interstitials and vacancies. Additional experimental work is necessary before these questions can be answered.

6.5. Magnetic Susceptibility

Since unpaired electrons in solids give rise to paramagnetism, the magnetic susceptibility is another property which is of value in detecting the distribution of charge in lattice defects. For example, Jensen[64] first showed that alkali halides containing F-centers possessed an excess paramagnetism of the magnitude expected from the concentration of unpaired electrons in anion-vacancies. This work has been extended recently by Scott *et al.*[65] and by Rauch and Heer[66]. According to the James–Lark-Horovitz model[20], electrons trapped in N-type germanium should also exert a paramagnetic contribution.

Attempts to observe excess paramagnetism in bombarded N-type germanium have been made by Stevens[67] *et al.* The situation is complicated to some extent

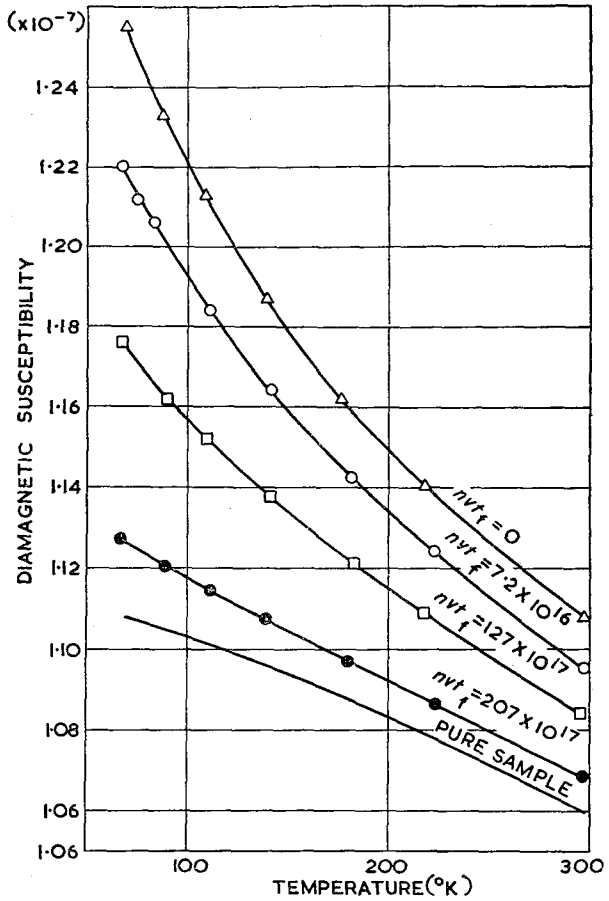


Figure 17. *The diamagnetic susceptibility v. temperature of a fast-neutron-bombarded N-type germanium specimen. The exposures are indicated on the curves. The curve of a pure specimen is shown for comparison. (Stevens et al.*[67])

by the free carrier contribution which in the classical range is given by:

$$\chi_c = \frac{\beta^2 n}{3\rho kT}\left(3 - \left\langle \frac{\overline{m}}{m^*} \right\rangle^2\right)$$

where β is the Bohr magneton, ρ is the density, and the quantity in angular brackets is the reciprocal of the effective mass squared, averaged appropriately for motion of the electron in the magnetic field. Because of the small effective mass of electrons in germanium $\left(\left\langle \frac{\overline{m}}{m^*} \right\rangle^2 = 51\right)$, the free carrier contribution is diamagnetic. Consequently, as the carriers are removed during bombardment the susceptibility becomes more positive. The behavior of the total

susceptibility after successive fast-neutron exposure is shown by Figure 17 in which the diamagnetic susceptibility is plotted against temperature. It is observed that as the electrons are removed, the curve approaches that for a high-purity un-irradiated specimen. At the point of conversion all of the electrons in the specimen ($7{\cdot}5 \times 10^{17}$ cm^{-3}) have been removed into localized states, yet there is no evidence of the expected paramagnetic contribution. In Figure 18, the curve for a high purity specimen (Curve 0) is compared to the curves just before (Curve 1) and just after (Curve 2) conversion. The curve calculated for $7{\cdot}5 \times 10^{17}$ trapped electrons per cm^3 is also shown (Curve 0′). It is evident

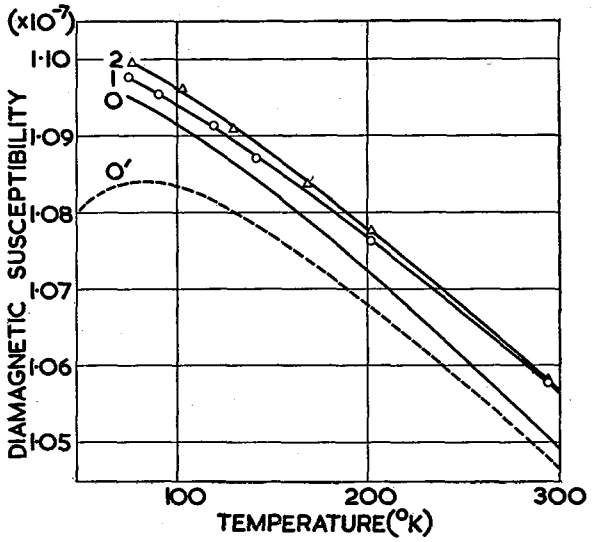


Figure 18. *The diamagnetic susceptibility v. temperature of a fast neutron bombarded N-type germanium specimen after irradiation into the intrinsic region. Curve 1 is for the specimen with approximately 5 × 10^{15} electrons per cm^3 remaining in the conduction band at 300° K. Curve 2 is after the specimen has become P-type and has less than 10^{15} carriers per cm^3 at 300° K. Curve 0 is for a pure specimen, displaced downward by 0·7 per cent for clarity. Curve 0′ is curve 0 to which has been added the expected paramagnetic contribution of 7·5 × 10^{17} unpaired spins per cm^3. (Stevens et al.*[67])

that, except for a smaller slope at high temperature, the susceptibility of the irradiated specimen approaches that of high-purity germanium with no indication of paramagnetic centers. The absence of unpaired-spin contribution was unexpected. However, it may result from the highly localized distribution of neutron-produced defects in that the defect-states may be so dense that banding of states can occur. If this is the case, classical statistics would no longer be valid to describe the paramagnetism and a Pauli type description leading to a temperature independent paramagnetic contribution would be more realistic. Measurements on γ-ray or electron-bombarded specimens would aid in clearing up this question.

Preliminary measurements of the susceptibility of irradiated silicon have

been made by Stevens[68]. Although the results are not yet complete, they indicate that trapped electrons in N-type silicon bombarded with fast neutrons contribute excess paramagnetism, while trapped holes in irradiated P-type specimens are magnetically inactive.

ACKNOWLEDGEMENT

It is a pleasure to acknowledge the assistance of many members of the Solid State Division of the Oak Ridge National Laboratory in the preparation of this paper. We particularly wish to thank D. K. Holmes, H. C. Schweinler and R. H. Silsbee for stimulating discussion and critical comment, and J. C. Pigg and D. K. Stevens for permission to use unpublished data. Finally we wish to acknowledge the efforts of K. Lark-Horovitz who, in his capacity of Consultant to the Oak Ridge National Laboratory, has worked in close co-operation with the authors for a number of years. His persistent efforts and stimulation gave much of the impetus which has led to the present development of this field of study.

REFERENCES

1. R. A. Logan. *Phys. Rev.* **91,** 757 (1953); S. Mayburg and L. Rotondi, *Phys. Rev.* **91,** 1015 (1953); S. Mayburg. *Phys. Rev.* **95,** 38 (1954); R. L. Hopkins and E. N. Clarke. *Phys. Rev.* **100,** 1786 (1955).
2. C. J. Gallagher and A. G. Tweet. *Phys. Rev.* **96,** 834A (1955); C. J. Gallagher. *Phys. Rev.* **88,** 721 (1952); W. C. Dunlap, Jr. *Phys. Rev.* **91,** 1282 (1953); Pearson, Read and Morin. *Phys. Rev.* **94,** 750A (1954); Pearson, Read and Feldmann. *Phys. Rev.* **100,** 1251A (1955).
3. For a review of the early work on silicon and germanium see K. Lark-Horovitz, Contribution to *Semiconducting Materials* p. 47 (Academic Press, New York, 1951); Davis, Johnson, Lark-Horovitz and Siegel. *Phys. Rev.* **74,** 1255 (1948); W. E. Johnson and K. Lark-Horovitz. *Phys. Rev.* **76,** 442 (1949).
4. Lark-Horovitz, Bleuler, Davis and Tendam. *Phys. Rev.* **73,** 1256 (1948).
5. E. P. Wigner. *J. Appl. Phys.* **17,** 857 (1946).
6. F. Seitz. *Disc. Faraday Soc.* **5,** 271 (1949).
7. F. Seitz and J. S. Koehler. *Solid State Physics,* vol. 2 (Academic Press Inc., New York, 1956).
8. W. S. Snyder and J. Neufeld. *Phys. Rev.* **97,** 1636 (1955).
9. J. A. Brinkman. *J. Appl. Phys.* **25,** 961 (1954).
10. G. H. Kinchin and R. S. Pease. *J. Nucl. Energy* **1,** 200 (1955); *Reports on Progress in Physics.* vol. **18,** pp. 1–51 (Physical Society, London, 1955).
11. H. Brooks. *Knolls Applied Physics Laboratory declassified report* 360 (1950).
12. H. Brooks. *Ann. Rev. Nucl. Sci.* (to be published).
13. E. E. Klontz and K. Lark-Horovitz. *Phys. Rev.* **86,** 643 (1952); see also E. E. Klontz. *Thesis* (Purdue University, 1952, unpublished).
14. P. Rappaport and J. J. Loferski. *Phys. Rev.* **100,** 1261A (1955); J. J. Loferski and P. Rappaport. *Phys. Rev.* **98,** 1861 (1955).
15. W. Kohn. *Phys. Rev.* **94,** 1409A (1954).
16. Sampson, Hurwitz and Clancy. *Phys. Rev.* **99,** 1657A (1955).
17. H. M. James and G. W. Lehman. *Semi-Conducting Materials,* p. 74 (Academic Press, New York, 1951).
18. Cleland, Crawford, Lark-Horovitz, Pigg and Young. *Phys. Rev.* **84,** 861 (1951).
19. Cleland, Crawford, Lark-Horovitz, Pigg and Young. *Phys. Rev.* **83,** 312 (1951).
20. H. M. James and K. Lark-Horovitz. *Z. Phys. Chem.* **198,** 107 (1951).
21. F. C. Champion. *Proc. Roy. Soc.,* A **234,** 541 (1956).
22. G. J. Dienes and D. A. Kleinman. *Phys. Rev.* **91,** 238A (1953).
23. Keesom, Lark-Horovitz and Pearlman. *Science,* **116,** 630 (1952).
24. Keesom *et al.* (unpublished results); *Purdue Progress Reports to Signal Corps P.R.F.* 1046 (1955) (unpublished).

25. Cleland, Crawford and Pigg. *Phys. Rev.* **98,** 1742 (1955).
26. Cleland, Crawford and Pigg. *Phys. Rev.* **99,** 1170 (1955).
27. Dresselhaus, Kip and Kittel. *Phys. Rev.* **92,** 827 (1953); **95,** 568 (1954); **98,** 368 (1955); Lax, Zeiger, Dexter and Rosenblum. *Phys. Rev.* **93,** 1418 (1954); Dexter, Zeiger and Lax, *Phys. Rev.* **95,** 557 (1954); Lax, Zeiger and Dexter. *Physica,* **20,** 818 (1954).
28. Forster, Fan and Lark-Horovitz. *Phys. Rev.* **86,** 643A (1952); see also J. H. Forster. *Thesis* (Purdue University, 1953) (unpublished).
29. H. Y. Fan and K. Lark-Horovitz. *Report of the Bristol Conference on Defects in Crystalline Solids,* p. 232 (Physical Society, London, 1955); H. Y. Fan *et al. Phys. Rev.* **95,** 1087 (1954).
30. Cleland, Crawford and Holmes. *Phys. Rev.* **102,** 722 (1956).
31. Schweinler, Cleland and Crawford (unpublished data).
32. Cleland, Lark-Horovitz and Pigg. *Phys. Rev.* **78,** 814 (1950).
33. Cleland, Crawford, Pigg and Schweinler. *Oak Ridge National Laboratory, ORNL-1852, Solid State Division Semiannual Progress Report,* p. 17 (1955).
34. J. W. Cleland and J. H. Crawford, Jr. (unpublished data).
35. H. Brooks. *Phys. Rev.* **83,** 879 (1951); P. P. Debye and E. M. Conwell. *Phys. Rev.* **93,** 693 (1954).
36. T. A. Longo and D. Kleitman. *Purdue University Progress Reports to Signal Corps P.R.F.* 1046 (Oct. 1955) (unpublished).
37. T. A. Longo and D. Kleitman. *Purdue University Progress Reports to Signal Corps P.R.F.* 1046 (Apr. 1955) (unpublished).
38. D. Hill and J. McKay. *Purdue University Progress Reports to Signal Corps P.R.F.* 1046 (Oct. 1955) (unpublished).
39. W. C. Dunlap, Jr. *Phys. Rev.* **100,** 1629 (1955).
40. E. A. Taft and F. H. Horn. *Phys. Rev.* **93,** 64 (1954).
41. J. W. Cleland and J. H. Crawford, Jr. *Phys. Rev.* **93,** 894 (1954); **95,** 1177 (1954).
42. Dresselhaus, Kip, Kittel and Wagoner. *Phys. Rev.* **98,** 556 (1955).
43. S. Siegel. *Phys. Rev.* **75,** 1823 (1949).
44. Pepper, Klontz, Lark-Horovitz and McKay. *Phys. Rev.* **94,** 1410A (1954).
45. J. W. Cleland and J. H. Crawford, Jr. *Phys. Rev.* **100,** 1614 (1955).
46. W. H. Brattain and G. L. Pearson. *Phys. Rev.* **80,** 846 (1950).
47. Brown, Fletcher and Wright. *Phys. Rev.* **92,** 591 (1953).
48. R. C. Fletcher and W. L. Brown. *Phys. Rev.* **92,** 585 (1953).
49. E. Conwell and V. F. Weisskopf. *Phys. Rev.* **77,** 388 (1950).
50. H. Y. Fan and M. Becker. *Semi-Conducting Materials* p. 132 (Academic Press, New York, 1951).
51. T. A. Longo and D. Kleitman. *Phys. Rev.* **100,** 1260A (1955); see also Ref. 37.
52. Lipson, Burstein and Smith. *Phys. Rev.* **99,** 444 (1955).
53. Ditchburn, Mitchell, Paige, Custers, Dyer and Clark. *Report of the Bristol Conference on Defects in Crystalline Solids,* p. 92 (Physical Society, London, 1955).
54. Fan, Kaiser, Klontz and Lark-Horovitz. *Phys. Rev.* **95,** 1087 (1954).
55. Stoeckman, Klontz, Fan and Lark-Horovitz. *Phys. Rev.* **98,** 1535 (1954).
56. D. Kleitman and T. A. Longo. *Phys. Rev.* **100,** 1261A (1955).
57. K. Lark-Horovitz (private communication).
58. J. C. Pigg. *Oak Ridge National Library, ORNL-1506, Solid State Division Semi-annual Progress Report,* p. 46.
59. Florida, Holt and Stephen. *Nature* **173,** 397 (1954).
60. Shulman, Brown and Fletcher. *Phys. Rev.* **96,** 833 (1954); R. G. Shulman. *Phys. Rev.* **102,** 1451 (1956).
61. J. R. Haynes and J. A. Hornbeck. *Phys. Rev.* **94,** 1438A (1954).
62. Brown, Fletcher and Wright. *Phys. Rev.* **96,** 834 (1954).
63. Klontz, Pepper and Lark-Horovitz. *Phys. Rev.* **98,** 1535 (1955).
64. P. Jensen. *Ann. Phys., Lpz.* **34,** 161 (1939).
65. Scott, Hrostowski and Bupp. *Phys. Rev.* **79,** 346 (1950).
66. C. V. Heer and C. Rauch. *Phys. Rev.* **90,** 530 (1953).
67. Stevens, Cleland, Crawford and Schweinler. *Phys. Rev.* **100,** 1084 (1955).
68. D. K. Stevens. Unpublished data.

LIFETIMES OF FREE ELECTRONS AND HOLES IN SOLIDS

A. ROSE, Ph.D.

Laboratories RCA Ltd., Zurich, Switzerland

MS. received 30 *July* 1956

LIFETIMES OF FREE ELECTRONS AND HOLES IN SOLIDS

1. INTRODUCTION

The lifetimes of free electrons and holes in solids are of great importance in the broad fields of luminescence, photoconductivity and semiconductor devices. At the same time, analysis of the factors that control lifetimes quickly leads to a complex array of algebraic relations or physical concepts, or both. Most of the complexity is inescapable. The still unplumbed variety of behaviour, particularly of luminescent materials and photoconductors, bears direct witness to a maze of inner parameters in these materials. Some of the complexity may, however, be relieved by a happy choice of approximations or concepts. This is one of the aims of the present paper. It must be stated as an aim since the judgement of whether or not a particular concept is simple is largely a matter of opinion.

There exist, in the appropriate literature, numerous analyses of the lifetime or recombination properties of special models. The models have been chosen to fit the behaviour of semiconductors[1, 2], of phosphors[3], of photoconductors[4] and of phosphors and photoconductors combined[5]. On the one hand, these models have brought with them separate systems of concepts, terminology and operating conditions; on the other hand, the physical processes governing lifetimes are the same in all three fields. In fact, it is becoming increasingly common to examine new materials for all three properties: electrical control of conductivity, luminescence and photoconductivity.

It is well worthwhile trying to develop a method of analysis that is not peculiar to any one of the fields and which also allows one to change the operating conditions smoothly between wide limits. That is the second aim of the present paper.

The writer has no illusions that one can set up a tidy universal model, ready to illuminate all problems. Quite the contrary. Most real problems involve the simultaneous influence of several different types of recombination centres. The shifting competition between these centres as the temperature or the rate of generation of free pairs changes is not always easy either to describe or comprehend. At the same time, a useful insight into the problem can be attempted by emphasizing the physical concepts.

2. SOME BACKGROUND INFORMATION

To compute the lifetime of a free carrier, one must know the distribution of electrons and holes among the discrete states lying in the forbidden zone. For a solid in thermal equilibrium, the distribution is characterized by a single parameter, the *Fermi level.* If, now, additional free carriers *of one sign only* are introduced, the new distribution can, to a good approximation, be described by a new value for the Fermi level, called the *steady-state* or *quasi Fermi level.*

Such an analysis has been carried out for carriers optically excited out of deep-lying discrete states[6] and also for carriers injected at one electrode giving rise to space-charge-limited current flow[7].

Another type of problem that can be treated readily is that of an insulator in which additional free carriers of both signs are generated—the new free carrier densities being large compared with the thermal equilibrium densities. Here the use of two steady-state Fermi levels, one for free electrons and one for free holes, reproduces most of the observed behaviour[5, 8]. But even here the treatment remains simple only for one type or class of discrete states in the forbidden zone. A class is defined by its capture cross-sections for free electrons and holes; the states may still be distributed over a range of energies. When more than one class of states is introduced, a pair of steady-state Fermi levels must be computed for each class, taking into account possible redistributions of electrons and holes between the classes. Much of the behaviour of phosphors and photoconductors is traceable to the presence of several classes of states[3, 4, 5, 9, 18].

In contrast with the many analyses of lifetimes for phosphors and photoconductors, there are only a few discussions especially centred on semiconductors. Three papers[1, 2, 10] all treat the problem of lifetime in the presence of a single level of discrete states. The solutions are involved algebraic expressions that do not lend themselves either to visualizing the physical processes or to including states distributed in energy. The extensive and careful analysis of Haynes and Hornbeck[11] illustrates the complexity of the problem when more than one class of discrete states is present.

What is missing is a framework within which the shifting distribution of electrons and holes in the various levels of discrete states can be visualized. This lack is especially evident in the transition from semiconductors, in which the thermal concentrations of carriers play a large role, to insulators, in which the optically (or electronically) excited carriers dominate the distribution.

Quick reference is made here to an earlier paper[4] in which the following items were discussed: the enormous range of lifetimes encountered in solids, extending from 10^{-13} sec to, at least, several seconds; the large range of capture cross-sections extending from about 10^{-13} cm^2, a value derivable from capture by an attractive coulomb field, to values less than 10^{-22} cm^2, which are probably associated with capture in the face of a repulsive coulomb field (capture of an electron by an already negatively charged centre); the existence generally of two independent lifetimes, one for electrons and one for holes; and the logical collapse of these two lifetimes into a single common lifetime when the added carrier densities are greater than the densities of discrete states.

A final introductory remark should be made. In a field as broad as the study of carrier lifetimes in solids, understanding is enhanced by a variety of approaches. The present analysis is supplementary to other analyses already published. Its emphasis is on the physics rather than the mathematics of recombination.

3. SURVEY OF METHOD OF ANALYSIS

Consider a free electron wandering in the conduction band. In an energy level diagram, the electron looks down and sees a number of energy levels (empty

states from the point of view of an electron) into which it can jump. These include the very shallow-lying states, called 'shallow traps' or 'ionized donors'; the deeper-lying states, called variously by the names 'traps', 'recombination centres', 'activators', 'killers', and 'ground states'; and finally the free holes in the valence band. From the point of view of the electron about to jump, the only significant information is the number of these states and their capture cross-sections for free electrons. It is necessary, then, as a first step to know the distribution of empty places into which the electron can jump. This implies knowing the distribution of both electrons and holes in all the states of the energy level diagram.

After the electron has jumped into one of the empty states, it has two possible paths. It may jump back up into the conduction band, by thermal excitation, or it may be captured by a free hole. It is significant to know which of these events will take place first.

If the electron jumps back into the conduction band, one calls the state from which it had jumped, a 'trap'. That is, the life of the free electron has been interrupted or suspended but not terminated. Furthermore, if the jumping back and forth between the conduction band and the trap is more frequent than the capture of the trapped electron by a free hole, the fractional occupation of these states by electrons is determined by the density of electrons in the conduction band. The electrons in these states are said to be in thermal equilibrium with the conduction band. In particular, one can draw a Fermi function pinned down at one point by the fractional occupation in the conduction band. The Fermi function extending below the conduction band described the occupation of those states that are in thermal equilibrium with the conduction band.

If the electron is captured by a free hole before it is thermally re-excited into the conduction band, the life of the free electron has in fact been terminated. A new optical excitation, electronic excitation or injection is required to replace the lost electron. Furthermore, the states in which capture by a free hole is much more probable than thermal excitation into the conduction band, are not in thermal equilibrium with the conduction band. The occupancy of these states is determined either by the kinetic processes of first capturing a free electron and then a free hole or by thermal exchange with the filled band. For those states in which the kinetic processes are dominant, one can easily verify that the ratio between the time a state is occupied by an electron and the time it is occupied by a hole is ns_n/ps_p; where n and p are the free electron and hole densities; and s_n and s_p are the capture cross-sections of the state for free electrons and holes, respectively. Finally, the occupancy of these states is not dependent on their position in the energy scale.

The second step is then to separate the states in the forbidden zone into those whose occupancy is determined by thermal exchange with one of the band edges, and those whose occupancy is determined by the kinetics of recombination. It is significant to make this separation only if the transition between the two types of occupancy is sharp. For most materials and most temperatures of interest, the transition can be considered sufficiently sharp. For example, at room temperature the rate of thermal excitation varies by a factor of almost 100 for a 0·1 V shift in energy. Since the kinetic processes are not a function of energy, this means that the ratio of thermal to kinetic processes also varies at the

rate of a factor 100 per 0·1 V. Thus, if one locates the energy at which thermal and kinetic processes are equal, one can be confident that, in moving 0·1 V towards the nearest band edge, thermal processes are dominant and similarly that a movement of 0·1 V in the opposite direction will ensure the dominance of kinetic processes. At lower temperatures, the transition is, of course, still sharper.

The argument can be summarized at this point by reference to Figure 1. Here, it is assumed that the added densities of carriers are large compared with their thermal equilibrium values, not because the added densities are large but because the thermal densities are negligibly small. The states in the forbidden zone have been separated into those for which thermal exchange with free

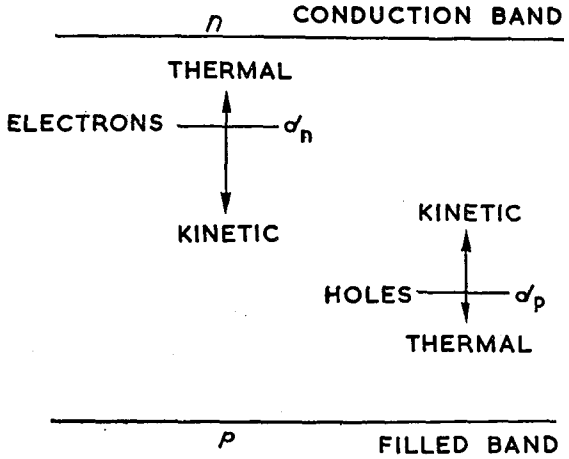


Figure 1. *Demarcation levels for excited carrier densities large compared with thermal densities.*

carriers is dominant and those for which the kinetics of capture are dominant. This separation is carried out once for electrons and once for holes. The demarcations between thermal and kinetic processes are labelled d_n and d_p for electrons and holes respectively. The occupancy of states between d_n and the conduction band is determined by thermal exchange of electrons with the conduction band, and is measured by the Fermi function adjusted to fit the known occupancy of the conduction band. Similarly, the occupancy of states between d_p and the filled band is determined by thermal exchange of holes with the filled band, and is measured by the Fermi function adjusted to fit the known occupancy of the filled band by holes. The occupancy of states between d_n and d_p is uniform or independent of energy and is measured, as already mentioned, by the ratio ns_n/ps_p.

The lifetime of a free electron can be visualized easily. It is approximately the time required for an electron to be captured by any one of the empty states lying below d_n. The actual lifetime is somewhat shorter since a small fraction of the electrons falling into the states lying above d_n are captured by holes. The correction can readily be introduced by computing the 'kinetic' traffic of holes

into these states. Similarly, the lifetime of a hole is approximately the time required for a hole to be captured by any one of the electron-occupied states above d_p.

4. DETERMINATION OF THE DEMARCATION LEVELS d_n, d_p

We seek here the energy level d_n at which a trapped electron has an equal chance of being thermally excited into the conduction band and of being captured by a free hole. Quantitatively this is expressed by:

$$\nu^* \exp(-\epsilon/kT) = pvs_p \qquad \ldots (1)$$

where the left-hand side gives the rate of thermal excitation and the right-hand

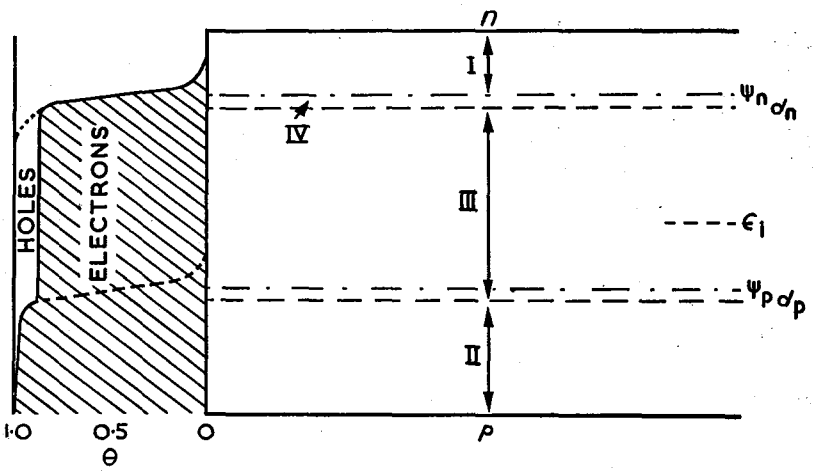


Figure 2. *Distribution of electrons and holes for excited carrier densities large compared with thermal densities.*

side the rate of capture by free holes. ν^* is the attempt-to-escape frequency multiplied by a transition probability; ϵ is the energy difference between the trapping state and the conduction band; p is the free hole density; v the thermal velocity; and s_p the capture cross-section of an electron in the trapping state for a free hole. In the earlier paper[4], equation (1) was solved for ϵ with the result that the level in question, d_n, to a first approximation was located at the steady-state Fermi level for electrons. Similarly, the level d_p for holes was located approximately at the steady-state Fermi level for free holes. Both levels d_n and d_p are actually shifted a small amount, $kT \ln(N_n/N_p)$, from their respective steady-state Fermi levels. N_n refers to the density of states occupied by electrons and lying between d_n and d_p, and N_p refers to the density of such states occupied by holes.†

The same results, obtained above mathematically, can be deduced by inspection of Figure 2 which shows the distribution function for electrons and

† In reference 8, the states $N_g = N_n + N_p$ were called ground states because the lifetimes of free electrons and holes were terminated by capture into these states.

holes in the states in the forbidden zone and in the free states. Starting with known values for the free electron and hole densities, n and p, the corresponding steady-state Fermi levels, ψ_n and ψ_p, are located. Next the Fermi functions are sketched in at these two levels. It is to be noted that the Fermi function shows a very sharp transition at the Fermi level between almost total and almost zero occupation values.

A determination is now made of the fraction of states lying between ψ_n and ψ_p which are occupied by electrons. The fraction occupied by holes is, of course, 1 minus the fraction occupied by electrons. The fractional occupancy is drawn in, as in Figure 2, as a line of constant occupancy, since the occupancy of these states is controlled by kinetic processes and is independent of energy. Finally, the intersections of the line of constant occupancy with the two Fermi functions defines the two demarcation levels d_n and d_p. The complete occupancy curve is shown in Figure 2 as a solid line extending from the conduction band to the filled band, including two parts of a Fermi function at its ends and a straight-line portion in the middle. From the point of view of free electron and hole densities, the forbidden gap has been narrowed by an amount equal to the straight-line portion of the curve. This is a result of the excitation that generated the free carrier densities in excess of their thermal equilibrium values.

The use of the states lying between ψ_n and ψ_p in order to determine the demarcation levels d_n and d_p is an approximation, but an approximation that becomes increasingly good as d_n and d_p approach ψ_n and ψ_p respectively. Since the Fermi function rises so steeply near the Fermi level, the intersections with the straight line portion of the curve are not likely to depart far from the Fermi levels. The facts that d_n and d_p are shifted by the same amounts from ψ_n and ψ_p and that the shift is a logarithmic function of the ratio N_n/N_p are clear from Figure 2 and the Fermi function.

5. LIFETIMES—EXCITED DENSITIES PREDOMINANT

When the densities of electrons and holes, excited by light or other means, are large compared with their densities at thermal equilibrium, the two steady-state Fermi levels are well separated, as shown in Figure 2. This is the situation common in phosphors and in many photoconducting insulators. There is a large region (marked III in Figure 2) of states whose occupancy is determined by kinetic processes. These states dominate the lifetimes of electrons and holes since most of the recombination traffic passes through them.

Consider, for example, a free electron looking down into the various empty states into which it might be captured. Capture into any of the states below d_n essentially terminates the life of the electron. There are many of these states lying between d_n and d_p. Below d_p, however, the fraction of empty states into which an electron can be captured falls off exponentially. For that reason, the states in region II of Figure 2 can usually be neglected. Only if the density of these states becomes large enough to offset the exponentially decreasing Fermi function or if their capture cross-sections for electrons is unusually large, need they be considered.

In region I (states lying above ψ_n), there are many states into which an electron

can jump since, substantially, all these states are empty. An electron jumping into these states, however, will be thermally re-excited into the conduction band long before it can capture a free hole. A measure of the contribution of these states to lifetime can better be obtained by looking at the traffic of free holes into them. Any free hole which is caught in these states (see Figure 1) will remain there until neutralized by a free electron. That is, the rate of 'permanent' capture of free electrons into the states of region I must equal the rate of capture of holes into the same states. But the rate of capture of holes in region I is small compared with that in region III, since the fraction of electron-occupied states into which a hole can be captured falls off exponentially above ψ_n. This is the same argument used for neglecting the contribution of the states in region II to lifetime.

The states in region IV present a more complex problem. Even though electrons captured into these states are likely to be thermally re-excited before capturing a hole, the fraction of electrons 'permanently' captured is significant. Again, one uses the traffic of free holes into these states as a guide. The states in region IV, because they lie below the Fermi level, are more than half occupied by electrons and therefore have substantially the same occupancy as the states in region III. The rate of capturing holes, by states in region IV, will accordingly be approximately the same as if the states were in region III. Since the recombination electron traffic into these states must equal the hole traffic, the states in region IV can be as effective in determining lifetimes as those in region III. On the other hand, the width of region IV is likely to be small compared with region III in insulators in which the excited densities of carriers are large compared with their thermal densities. For the purposes of many discussions, region IV may be included with region III.

In summary, the states in regions III and IV, that is, those lying between ψ_n and d_p, are the most important for determining the lifetimes of electrons and holes. It remains to determine the fractional occupancy of these states by electrons and holes.

5.1. Excited Carrier Densities less than the Densities of Recombination Centres

Let the densities of electrons in the conduction band plus those in region I be small compared with the density of holes (N_p) in region III. Similarly, let the densities of holes in the filled band plus those in region II be small compared with the density of electrons (N_n) in region III. This is a common situation for photoconducting insulators and for optically excited phosphors. The above conditions ensure that the densities of holes and electrons in regions III and IV remain substantially the same after excitation as they were before excitation. (By the condition of charge neutrality, the difference between the excited electron and hole densities outside III and IV must be reflected by the opposite change inside III and IV.) The density of holes (N_p) in III and IV is then equal to the density of states lying between φ, the thermal equilibrium value of the Fermi level, and ψ_n; the density of electrons (N_n) in III and IV is equal to the density of states lying between ψ and D_p. These are the electron and hole densities that existed in regions III and IV before excitation.

The electron and hole lifetimes can now be written in the usual form:

$$\tau_n = 1/vs_nN_p \quad \ldots (2)$$

$$\tau_p = 1/vs_pN_n \quad \ldots (3)$$

These lifetimes are independent and, in general, widely different since the capture cross-sections s_n and s_p are, in general, widely different. By way of example, it is known[12] that the electron lifetime in sensitized crystals of cadmium sulphide is about 10^{-3} sec. The lifetime of a free hole in these crystals may be estimated, from the poor efficiency of the photo-voltaic effect, to be less than 10^{-9} sec. It is only in some relatively pure and insensitive crystals of CdS that the electron and hole lifetimes have been observed[13, 14] to approach equality at about 10^{-7} sec.

One can readily see in terms of Figure 2 that as the excitation is increased the levels ψ_n and d_p, embracing the states through which recombination takes place, pull apart to include more states. If the states are all of one kind, the effect is to shorten the lifetimes of carriers by increasing N_n and N_p. This is a comparatively slow effect and readily accounts for the frequent observations that photocurrents increase as the 0·7 to 0·9 power of the light intensity[12, 15, 16]. Such odd powers over large ranges of light intensity are difficult to account for without a distribution of states in energy.

The increased separation of ψ_n and d_p at higher excitations means also that higher energy jumps are possible from the conduction band into one of the N_p states, or between the valence band and one of the N_n states. One sees here, in simple fashion, a mechanism that can account for the frequent observation that the luminescence of phosphors shifts towards the blue at higher excitations[17].

An increase in temperature, on the other hand, tends to bring the levels ψ_n and d_p closer together. The distance of ψ_n from the conduction band is defined by:

$$\psi_n = kT \ln (N_c/n) \quad \ldots (4)$$

where N_c is the density of states in the bottom of the conduction band. Again, if there is only one class of states and if the temperature does not affect the capture cross-sections, an increase in temperature has the opposite effect to an increase in rate of excitation. There are many instances in photoconductivity and luminescence where the opposing actions of temperature and excitation are evident. Experimentally, this means that an effect that occurs at a given temperature will occur at a higher temperature if the excitation is increased. Good quantitative evidence for this correlation has been reported by Bube[18] in measurements of supra-linearity in photoconductors. Other evidence is found in the known facts that the colour changes of luminescence occur at higher excitations for higher temperatures and that the temperature 'break point' for phosphors increases with increasing excitation. The temperature changes slowly with carrier density by the relation:

$$\triangle T/T = kT\psi_n \ln (n_2/n_1) \quad \ldots (5)$$

obtained from equation (4). $\triangle T$ is the temperature change needed to compensate for an increase of carrier density from n_1 to n_2. In this way the Fermi level ψ_n is maintained constant. The relative shifts of the temperature break points, for excitation densities in the ratios 1:3:40, found by Nail, Pearlman and Urbach and shown in Figure 106 of *Luminescence of Solids*[17], appear to fit equation (4).

5.2. More than One Class of States

The behaviour of a single class of states in the forbidden zone, even though distributed in energy, is comparatively simple. One need only bear in mind that more of these states are embraced by ψ_n and d_p as the intensity of excitation is increased. In this way, the lifetimes of electrons and holes are continuously shortened. Most real solids, however, have two or more significant classes of states in the forbidden zone. These states arise variously from chemical impurities, non-stoichiometry, and departures from perfect crystallinity. The arguments outlined above for locating the demarcation levels d_n and d_p for one class of states must be repeated for each class of states. But, for each class of states, the recombination traffic of electrons into those states must be identically equal to the recombination traffic of holes. This is the condition for steady state. When only one class of states was considered, the equality of traffic was easily satisfied by adjusting the densities of free electrons and holes to fit the already-given capture cross-sections and occupancies of the ground states N_n/N_g and N_p/N_g (see equations (2) and (3)). These occupancies were the same after excitation as before excitation. When, now, a second class of ground states is introduced, it would be highly fortuitous that its capture cross-sections and occupancies would fit the densities of electrons and holes already determined by the first class of states. It is almost certain that there will be an unbalance of electron and hole recombination traffic into the second class of states. This unbalance will continue until the occupancies of the second class are sufficiently altered to cancel the unbalance†. But if the numbers of electrons and holes in the second class of states is appreciably altered there must be some place for these electrons and holes to go to. They cannot go into the free states or into regions I and II, since the densities of electrons and holes in these states have been assumed to be small in comparison with the ground state densities. The only remaining possibility is that the excess electrons or holes from the second class of states be transferred to the first class of states. In brief, the introduction of a second class of states compels a redistribution of electrons and holes between it and the first, or already existing, class of states.

It is this redistribution of electrons and holes between different classes of states that lies at the bottom of some of the most complex, varied and dimly decipherable behaviour of phosphors and photoconductors[16, 17]. The redistribution, by changing the values of N_n and N_p, see equations (2) and (3), may obviously either increase or decrease the lifetimes of the free carriers. Further, the recombination traffic may be shifted from one class of states to another. Some of these effects have been described in earlier papers[4, 18]. One can readily imagine how a second class of states can sensitize or desensitize a photoconductor, activate or poison a phosphor. These are some of the manifold results that come from deliberately incorporating impurities in a solid by physical or chemical means.

5.3. Electronic Doping

There is a completely parallel set of results for electronically incorporating impurities. The meaning of this statement is easily seen from Figure 2. Suppose

† In the steady-state, the ratio $s_p N_n / s_n N_p$ must be the same for each class of states.

that in region III there is one class of states while in regions I or II there is a second class of states. The lifetimes of the free carriers will be determined substantially only by the states in region III, that is, by the first class of states. But the demarcations d_n and d_p, setting the boundaries of region III, are electronically adjustable. At high excitations, these levels pull apart so that states that were originally in regions I or II are now included in region III. The same adjustments can be made by variations in the temperature of the solid. Since the states in region III play the dominant role for the lifetimes of free carriers, *it is the introduction of a second class of states into region III rather than just into the solids that is significant.* From this point of view, it is all the same whether new states are introduced into region III by physical diffusion or by electronic envelopment by the levels d_n and d_p. One might describe the last process as *electronic doping.*

What is conceptually significant is that the long catalogue of effects which are known to arise from the physical and chemical addition of impurities should also be producible electronically by varying the excitation intensity or temperature. The electronic doping is, of course, reversible. The supra-linear curves found in photoconductors[18] and phosphors[17] are the electronic equivalent of sensitization and activation. The sublinear curves are the electronic equivalent of desensitization and deactivation. Infra-red quenching is the electronic equivalent of poisoning. And the colour shifts, found in luminescence, as the excitation intensity is varied, are the electronic equivalent of multiple activation.

5.4. Interlaced Densities

In the foregoing discussion, the densities of free and shallow-trapped electrons and holes were taken to be small enough to have little effect on the occupancy of the states in region III. Most photoconducting insulators and phosphors fit these conditions for moderate intensities of optical excitation. Also, it is generally true that the density of one sign of carrier is much larger than the density of the other sign of carrier. This means that the difference between the densities of carriers is substantially equal to the larger density. This difference must appear (with opposite sign) in the states of region III. As the intensity of excitation is increased, it will happen, for example, that the free plus the shallow-trapped electrons exceed the number of holes normally present in region III. Holes will then be forced to appear in region III equal in number to the free plus the shallow-trapped electrons. Since the shallow-trapped electrons are in thermal equilibrium with the free electrons, their numbers will be proportional to the numbers of free electrons. Consequently, the number of holes in region III will also be proportional to the number of free electrons. The result is immediately recognizable as bi-molecular recombination.

In this intermediate range of excitations, the density of majority carriers (electrons) increases only as the square root of the excitation. The lifetime of the majority carrier is becoming shorter. Meantime, the minority carrier density (holes), is increasing linearly, or slightly faster than linearly, with excitation. This pattern continues until the numbers of free electrons and holes, and therefore, their lifetimes, approach equality. The densities of free carriers will, at that point, be greater than the densities of states in region III or greater than

about 10^{15} cm^{-3}. It is difficult to optically excite photoconductors or phosphors into this range. The excitation density needed is likely to be greater than 10^{22} sec^{-1} cm^{-3}. Excitation by electron beams, however, can achieve this and even higher rates.

5.5. Excited Carrier Densities Larger than Recombination Centre Densities

If one can excite an insulator to such high levels that the free carrier densities are greater than the densities of states in region III, then the free carrier densities and their lifetimes must be substantially equal. This follows logically since there are insufficient available states into which the electrons or holes resulting from any appreciable difference in free carrier densities can be located. It is likely that one needs the high excitation densities of electron beams to achieve this condition.

It was emphasized earlier that, at low excitations, the various classes of states in region III had to exchange electrons and holes in order to satisfy the boundary conditions—*i.e.* equal rates of electron and hole flow into each class. When the free carrier density is larger than the density of states in region III, this exchange of electrons and holes between classes of states no longer takes place. Rather the occupancy of each class of states is adjusted, independently, by exchange of electrons or holes with the free carriers. A large adjustment of occupancy in a class of states can be made without altering the substantial equality of free carrier densities. The result is then quite simple. Owing to the equality in free carrier densities, the occupancy of each class of states is given by (see equation (2) and (3) with equal lifetimes):

$$N_n/N_p = s_n/s_p \qquad \ldots (6)$$

One finds now that the introduction of any type of additional states into region III can only have the effect of reducing the lifetime of both signs of carriers. (It is to be understood that the number of additional states is still less than the number of free carriers.) By way of contrast, at low excitations the introduction of states into region III could increase or decrease the lifetime of one sign of carrier while having the opposite effect on the other sign of carrier. The sensitivity of photoconductors like CdS, for example, may, by the addition of states in region III, be made 10^5 times more sensitive at low excitations[19]. That is, the electron lifetime has been increased by a factor of 10^5. In the present high excitation range, the sensitivity (electron lifetime) must recede at least to where it was before the introduction of the additional states. For this reason, the current *versus* light curves of some very sensitive photoconductors appear to be saturated towards high light intensities.

Finally, one must, at these high densities of carriers, begin to be concerned with direct recombination between free electrons and free holes. The capture cross-section of a free hole for a free electron is quite small[4, 20], in the neighborhood of 10^{-20} cm^2. The small cross-section and low density of free holes normally allows one to neglect their contribution to electron lifetime. When, however, the density of free holes approaches 10^{18} cm^{-3}, the lifetime of an electron looking down at the free holes alone would be

$$(vs_n p)^{-1} = 10^{-5} \text{ sec}$$

If competing recombination processes are no faster than this, direct recombination of free electrons and holes will predominate. As pointed out by Redfield[21], this is more likely to occur in materials with a small forbidden gap ($< 0{\cdot}3$ V) where there is less room for states in the forbidden zone to contribute to recombination.

6. LIFETIMES—THERMAL CARRIER DENSITIES PREDOMINANT

Up to this point, the major properties of phosphors and photoconductors could be discussed by reference to Figure 2. Materials with forbidden gaps greater than 1 eV can be described by this figure, since in these materials the thermal density of carriers can be negligibly small. At lower temperatures, proportionally

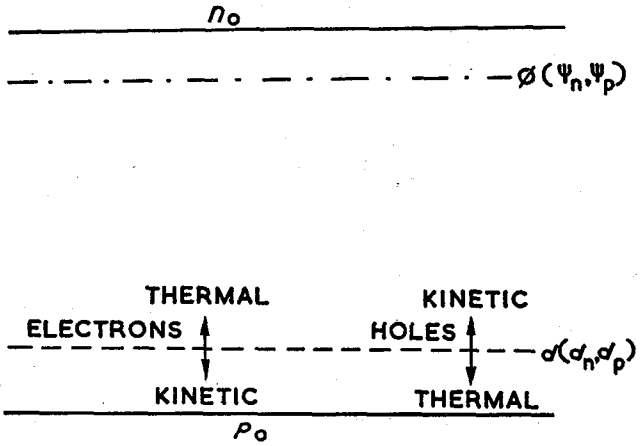


Figure 3. *Demarcation level for thermal densities of carriers large compared with excited densities.*

smaller gap widths would be included. The major feature of Figure 2 is the division of the states in the forbidden zone into a group near the middle of the zone, through which most of the recombination takes place and whose occupancy is determined by kinetic processes, and two groups near the band edges, which are in thermal equilibrium with the free carrier densities and do not contribute significantly to recombination.

If we now postulate that the thermal equilibrium densities of carriers are large compared with the optically excited densities, a new division of states is necessary. By postulate, thermal processes predominate throughout the forbidden zone. Consequently, region III, in which kinetic processes had been predominant, disappears. The boundaries d_n and d_p of region III collapse into a single level d ($= d_n, d_p$), see Figure 3. Also, the two steady-state Fermi levels collapse into the single thermal-equilibrium Fermi level φ of Figure 3. The location of d relative to φ is determined by the same argument given in equation (1). To a first approximation, the level d is the same distance from the filled band as the Fermi level is from the conduction band. The correction term is $kT \ln (s_n/s_p)$. (See Appendix I.)

The level d is such that an electron located there is equally likely to be thermally excited into the conduction band or to be captured by a free hole. Similarly, a hole at the level d is equally likely to be thermally excited into the valence band or to be captured by a free electron. One can still use the level d to make the kind of separations already made in Figure 1. That is, the states above d are dominated by thermal exchanges with the conduction band. Electrons falling into these states are more likely to be thermally re-excited into the conduction band. Electrons falling into states below d are more likely to be captured by a free hole. Similarly, from the point of view of free holes, the states below d are dominated by thermal exchange with the filled band. Holes, captured into these states, are more likely to be thermally re-excited into the filled band. Holes captured into states above d are more likely to capture a free electron.

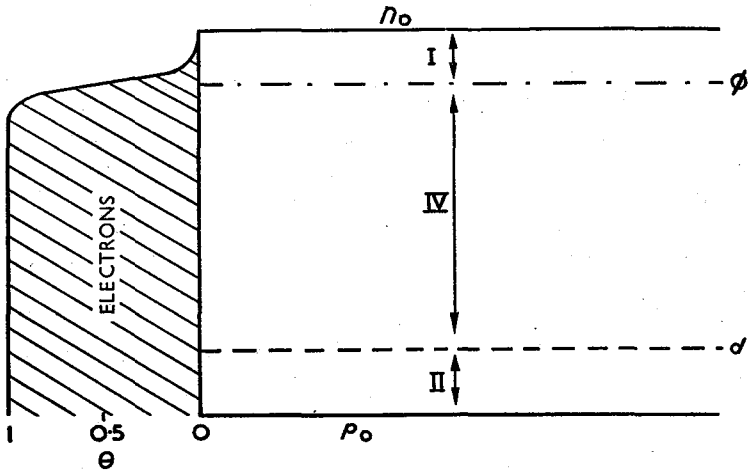


Figure 4. *Distribution of electrons and holes for thermal densities of carriers large compared with excited densities.*

Three groups of states in the forbidden zone, using the same designations as in Figure 2, are shown in Figure 4. Region III which was predominant in Figure 2 has collapsed into a line at the level d. Region IV which was of secondary importance in Figure 2 is now of major importance in Figure 4. Regions I and II remain, and retain the same significance as in Figure 2.

One uses the same procedure as before to examine the lifetime of an electron and a hole. A free electron looks down into the various empty states into which it might be captured. The states in region I are substantially empty so that, other factors being equal, an electron has an equal chance of falling into any of the states. If the electron falls into a state lying near the conduction band, it is more quickly thermally excited into the conduction band, and has less chance of capturing a free hole than if it falls into a state near the Fermi level. Since the rate of thermal excitation increases exponentially above the Fermi level, one can neglect the contribution to lifetimes by states more than kT higher than the Fermi level relative to states at the Fermi level.

The number of empty states in region IV decreases exponentially as one

departs downward from the Fermi level. At the same time, the residence time of an electron in these states before being thermally re-excited increases exponentially. The combination of these two opposing factors makes the states at various energy levels equally effective for recombination traffic. The same result will be even more evident when one considers the recombination traffic of free holes. A strong caution needs to be inserted here. The arguments leading to equal effectiveness of states in region IV for recombination do not necessarily mean that they make an equal contribution to electron lifetime. In transistor materials, where the electron and hole lifetimes are equal, equal contribution to recombination means equal contribution to electron lifetime. In many sensitive, high dark current photoconductors, where the electron lifetime is longer than the hole lifetime, the states in region IV contribute equally to electron recombination but not necessarily, as will be discussed below, to electron lifetime.

The effectiveness of the empty states in region IV for electron recombination was found to be constant because the probability of capture by a free hole increased as fast as the density of empty states decreased. At the level *d*, the probability of capture by a hole is already near unity. This has two consequences. It means that the contribution to recombination of all the empty states in region IV would be the same if the density of empty states in region IV were everywhere equal to the density of empty states at *d*, and if an electron captured in these states were certain to be captured by a free hole. In brief, region IV could be replaced by a region III in which the density of holes at each energy level was that at the level *d*.

The second consequence is that the effectiveness of the states below *d* for electron recombination decreases exponentially. Already an electron captured near *d* is almost certain to be captured by a hole. This factor cannot increase below *d*. The number of empty states below *d*, however, does decrease exponentially so that their contribution to recombination also decreases exponentially.

In summary, the various levels in region IV were found to be equally effective for electron recombination while the levels outside region IV make an exponentially decreasing contribution to recombination. Region IV, then, plays the same role as did region III in the discussion of insulators. The arguments leading to this conclusion have been more involved than they need be. In treating electron recombination, one is trying to separate out a small capture process from a large thermal exchange process. The same conclusions follow more quickly in treating the recombination of the minority carrier.

Consider a hole looking up into the various electron-occupied states. The states in region II may be neglected by arguments, already given, that a hole captured in these states is more likely to be thermally re-excited into the filled band. The states in region IV are substantially filled with electrons and, since these states lie above *d*, a hole captured in them is almost certain to recombine with a free electron from the conduction band. Thus, the recombination traffic of holes is uniformly distributed throughout the states in region IV, a result already obtained for the recombination traffic of electrons.

The number of electron occupied states into which a hole may jump above φ decreases exponentially and so, therefore, does their contribution to recombination. Again, the conclusion is that the states in region IV carry the major recombination traffic.

6.1. Minority Carrier Lifetime—Single Level—Small Perturbation

Once having established the importance of region IV for recombination, the analysis of the *minority* carrier lifetime proceeds quickly. The analysis of the *majority* carrier lifetime is, in general, more complex and is treated below in the section on high dark current photoconductors.

We review, as is done by Shockley and Read[1], the behaviour of a single level of recombination states as the Fermi level is swept through the forbidden zone. The review is carried out for state densities large compared with thermal carrier densities. The slight modification needed to fit the opposite case of small state densities is then indicated.

The review of minority carrier lifetime is at the same time the review of majority carrier lifetime *for transistor materials*. The simplifying feature of transistor materials is that electron and hole lifetimes are equal and one need only examine the lifetime of one sign of carrier—the more convenient one—in this case the minority carrier. Equality of lifetimes is ensured, for logical reasons, by having the added carrier densities larger than the density of states in region IV. While this was difficult to achieve in photoconducting insulators and phosphors because the state densities were likely to be greater than 10^{15} cm^{-3}, it is comparatively easy to achieve in transistor materials where, by careful preparation, the state densities can be brought below 10^{12} cm^{-3}. The higher densities of the usual donor and acceptor levels are neglected here because they lie close to the band edges in regions I and II. At sufficiently high densities or at sufficiently low temperatures, they can, of course, contribute significantly to recombination.

Figure 5 shows six successive stages in the movement of the Fermi level (φ) across the forbidden zone from strongly N-type to strongly P-type material. The geometric centre of the forbidden zone is denoted by ϵ_1. The demarcation level (d), except for a small correction term $kT \ln (s_n/s_p)$, is the mirror image of the Fermi level reflected at ϵ_1. The single level of recombination states (N_t) is located above the middle of the forbidden zone.

The argument followed below can be briefly summarized. As the Fermi level φ_f sweeps through the forbidden zone from conduction band to filled band, the demarcation level d sweeps through the forbidden zone at the same rate but in the opposite direction. States embraced by these two levels are counted in full, states outside the two levels make an exponentially decreasing contribution to recombination.

In Figure 5(a) the Fermi level is near the conduction band; the demarcation level is near the filled band. The states N_t are embraced by the levels φ and d and these states are almost fully occupied by electrons. The lifetime of a hole, the minority carrier, is therefore given by:

$$\tau_{p0} = 1/v s_p N_t \qquad \ldots (7)$$

The same result would be obtained for any other location of the N_t states between φ and d. This means that as φ moves away from the conduction band towards N_t the lifetime remains constant and equal to τ_{p0}.

In Figure 5(b), the material is still N-type but the Fermi level now lies below the N_t states but above ϵ_1. By previous arguments, the efficacy for recombination

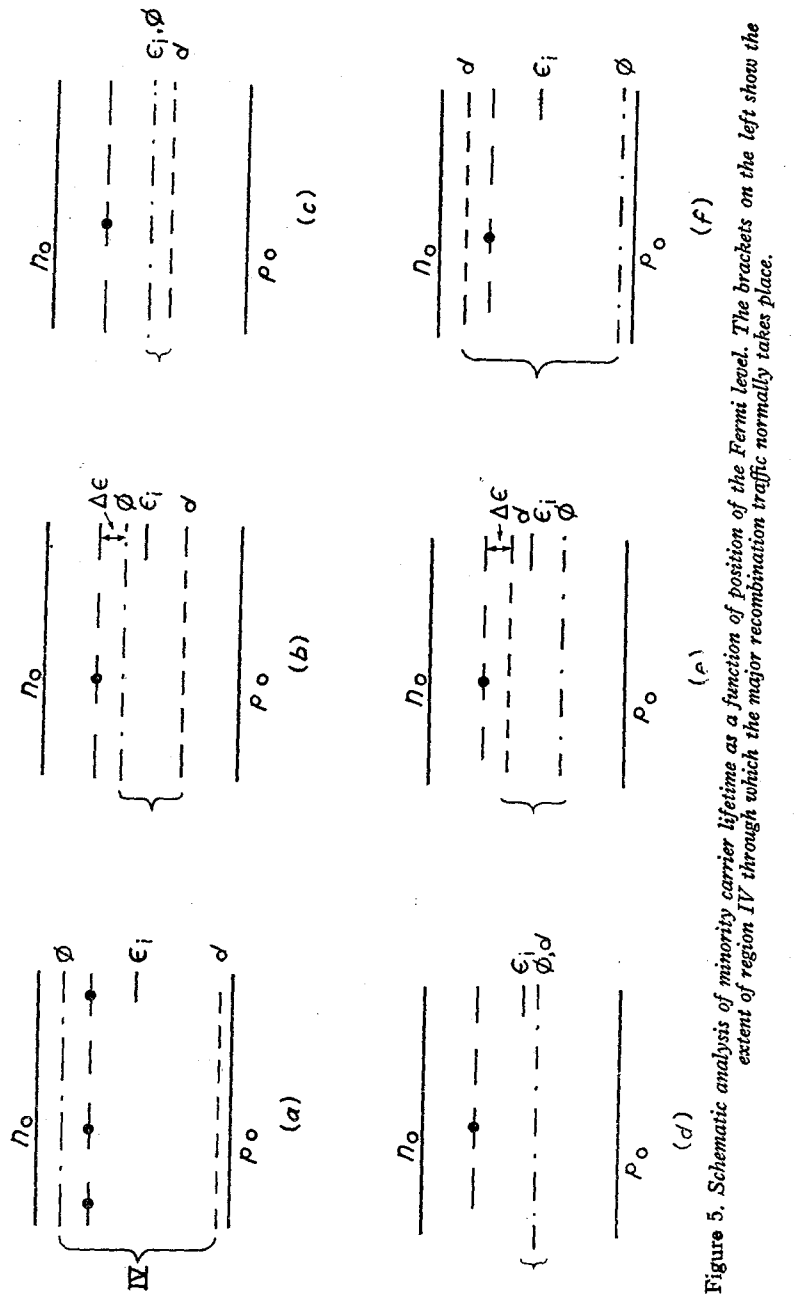


Figure 5. *Schematic analysis of minority carrier lifetime as a function of position of the Fermi level. The brackets on the left show the extent of region IV through which the major recombination traffic normally takes place.*

of states lying outside the range $\varphi - d$, decreases as $\exp(-\triangle\epsilon/kT)$ where $\triangle\epsilon$ is the distance to φ or d whichever is nearer. In this instance, it is clear why the contribution of N_t to recombination decreases. The thermal equilibrium occupancy of N_t by electrons decreases as $\exp(-\triangle\epsilon/kT)$. Thus a free hole sees fewer electron-occupied states into which it can be captured. As the Fermi level moves from N_t to ϵ_i the lifetime increases exponentially and is given by:

$$\tau_{p0} \exp(+\triangle\epsilon/kT)$$

In Figure 5(*c*) the Fermi level coincides with ϵ_i and the material is intrinsic. Note that φ and d do not coincide. The energy interval $\varphi - d$ is the small correction term $kT \ln (s_n/s_p)$. The maximum lifetime for N-type material is achieved here, but not the maximum lifetime for P-type. As the Fermi level moves through and below E_i the material becomes P-type. The minority carrier is now a free electron. There is a discontinuity in minority carrier lifetime since the capture cross-section s_n is larger than s_p. The lifetime of a free electron increases as the Fermi level continues to move from ϵ_i to a distance $\frac{1}{2}kT \ln (s_n/s_p)$ below ϵ_i. This is a result of the increasing distance between the N_t states and the Fermi level.

In Figure 5(*d*), the Fermi level and the demarcation level coincide at the energy level $\frac{1}{2}kT \ln (s_n/s_p)$ below ϵ_i. The maximum lifetime of a free electron is achieved here because this represents the maximum departure of N_t from φ and d.

In Figure 5(*e*), φ and d have moved past each other, φ moving toward the filled band and d toward the conduction band. The N_t states still lie outside the interval $\varphi - d$. The efficacy of the N_t states for recombination is reduced by the factor $\exp(-\triangle\epsilon/kT)$ compared with their efficacy when embraced by the levels φ and d. As φ approaches the filled band, its approximate mirror image d approaches the N_t states. The interval $\triangle\epsilon$ becomes smaller and the lifetime decreases at the rate

$$\tau_{n0} \exp(+\triangle\epsilon/kT)$$

τ_{n0} is used here in anticipation of Figure 5(*f*).

In Figure 5(*f*), the Fermi level has moved so close to the filled band that its mirror image d now lies above N_t. Since N_t is again embraced by the levels $\varphi - d$, these states have attained their maximum contribution to recombination. The N_t states are substantially empty and the time required for them to capture a free electron is:

$$\tau_{n0} = 1/vs_nN_t \qquad \ldots (8)$$

This lifetime remains constant as long as N_t is embraced by φ_f and d.

The variation of minority carrier lifetime with Fermi level is summarized in Figure 6. This is the same type of curve as shown by Shockley and Read for large state densities. The various parts of the curve are identified with the series of Figures 5(*a*)–5(*f*) by the letters (*a*–*f*).

The treatment for low state densities—sufficiently low to ensure the equality of electron and hole lifetimes—parallels that for large state densities except near the intrinsic range. Here, there can be no discontinuity since electron and hole lifetimes are equal. The approximate behaviour is shown by the dotted line.

Figure 5 shows visually a number of the conclusions reported by Shockley and

Read. When $\tau_{n0} = \tau_{p0}$, the capture cross-sections s_n and s_p are equal, according to equations (7) and (8), and the correction term $kT \ln (s_n/s_p)$ vanishes. Under these conditions, complete symmetry is obtained in Figures 5 and 6. The states N_t have the same effect on lifetime on either side of the intrinsic level ϵ_i. Asymmetry is introduced when the two capture cross-sections are not equal. The ratio of capture cross-sections is likely to be two or three powers of 10, reflecting the difference between capture by a neutral state and capture by a charged state. The correction term is then only about a 0·1 V.

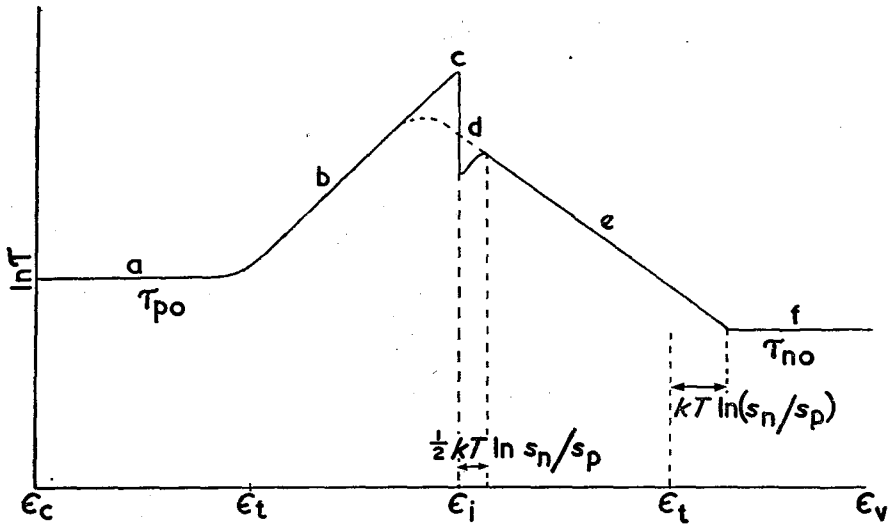


Figure 6. *Minority carrier lifetime plotted as a function of position of the Fermi level. The single level of recombination states is located at* ϵ_t. *The level* (ϵ_t) *is the mirror image of* ϵ_t *reflected at* ϵ_i. *The solid line shows the behaviour for recombination state densities large compared with thermal carrier densities. The dotted line is the estimated correction for the case of small densities of recombination states—that is, equal lifetimes of electrons and holes.*

6.2. Single Level—Large Perturbation

Figure 5 also helps in visualizing what happens to the lifetime when the added carrier densities become large, *i.e.* larger than the thermal densities. If one starts with the states N_t embraced by the levels φ and d, then these states are either substantially filled by electrons or substantially empty, depending upon whether the material is N-type or P-type. These occupancies are controlled by thermal equilibrium processes. The corresponding lifetimes τ_{n0} and τ_{p0} are given by equations (7) and (8). At sufficiently high densities of added carriers the states N_t are embraced by the two levels d_n and d_p. This means, as discussed earlier, that the occupancy of the N_t states is determined by kinetic processes. In particular, the ratio of electron-occupied states N_n' to hole-occupied states N_p' is:

$$N_n'/N_p' = s_n/s_p \qquad \ldots (9)$$

The lifetime is $\tau_\infty = 1/vN_n's_p = 1/vN_p's_n$

$$= (v\frac{s_n}{s_n + s_p}N_t s_p)^{-1} \qquad \ldots (10)$$

If the capture cross-sections are nearly equal, then $\tau_\infty = 2\tau_{n0}$. Other types of behaviour are easily deduced from equations (7), (8) and (10).

If one started at low excitations with the N_t states well outside the interval $\varphi - d$, the lifetime would be quite large because these states are only partially effective. At high excitations, the N_t states would be embraced by the levels d_n, d_p and would be completely effective. Under these conditions τ_∞ would be much smaller than the starting lifetime.

6.3. Multiple Levels

What is perhaps more significant is that Figure 5 allows a quick estimate to be made of the effect on lifetime of adding other levels of states, either of the same type or different types. Suppose, for example, that in Figure 5(*a*) the N_t states, instead of being only at one energy level, were distributed uniformly between φ and d. Qualitatively, the same dependence of lifetime on position of the Fermi level would be observed as already obtained for a single level. That is, the lifetime would increase as the Fermi level approached the middle of the forbidden zone, starting from either N- or P-type material. The quantitative difference is that the rise of lifetime toward intrinsic material, instead of being exponential, would be a slow rise varying approximately as the reciprocal of the interval $\varphi - d$.

This is a significant result for the rate of generation of free pairs in not too-well purified materials. The rate of thermal generation is given by the minority carrier density divided by the minority (= free pair) lifetime. As the Fermi level approaches the middle of the forbidden zone, the minority carrier density must increase exponentially. If, at the same time, the lifetime increases only slowly, as it would for a distribution of recombination states, the thermal generation of pairs would increase almost exponentially. This behaviour can readily account for the frequent observations that the reverse currents in P–N junctions are generated predominantly in the intrinsic section rather than at the N and P terminals.

It is not necessary that the φ states be distributed continuously in energy in order to void the exponential pattern of Figure 6. States scattered at a few levels throughout the interval $\varphi - d$ in Figure 5(*a*) are sufficient. The new levels may have the same or different cross-sections from the N_t states. The effect will always be to attenuate the exponential dependence of lifetime on position of the Fermi level.

It must be noted here that the addition of other classes of states in the interval $\varphi - d$ (region IV) is quite different from their addition in the interval $d_p - d_n$ (region III). In the latter case, as already discussed, there was an exchange of electrons or holes between the added and the previously existing states, which resulted in a great variety of possible behaviour. The new states added in region IV, on the other hand, do not interact with the existing states. Their occupancy is determined by thermal exchange with the conduction band.

The Shockley–Read, as well as the Hall, analysis of a single level of states has pointed out that the location of the level could be established in transistors by observing the temperature-dependence of lifetime. This is true for a level lying, for example, above the Fermi level. Also this present analysis for a fixed Fermi level, shows that the efficacy of these states for recombination will increase

exponentially with increasing temperature at a rate given by the energy interval $\triangle\epsilon$ between these states and the Fermi level. If there is more than one level of states present, the simple exponential dependence is voided, and a slower, more complex dependence on temperature will occur. Further, if the capture cross-section varies with temperature, the lifetime may either increase or decrease with increasing temperature.

6.4. High-Dark-Current Photoconductors

There are many materials in which, like transistors, the thermally generated carrier densities are high and in which significant photocurrents are observable. These include conducting samples of cadmium sulphide, cadmium selenide, zinc sulphide, zinc oxide, grey selenium, lead sulphide, thallium sulphide, intermetallic compounds, and silicon and germanium, even when the latter two materials are not of transistor quality. Unlike transistors, these materials, in general, do not have a sufficiently low density of states in the forbidden zone to ensure the equality of electron and hole lifetimes. It has been stated already that when the added carrier densities exceed the density of states in the forbidden zone, particularly in regions III and IV, the equality of lifetimes is ensured for logical reasons. When the added carrier densities are less than the state density, the electron and hole lifetimes may or may not be equal and, in general, are not.

In the case of transistors, since the two lifetimes were equal, one could choose to analyze the more convenient lifetime, that of the minority carrier. Even when the state density is large and the two lifetimes are not equal, one can still analyze the minority carrier lifetime using Figures 4 and 5, as was carried out in the preceding section. The lifetime of the majority carrier, however, must be separately analyzed. The analysis does not lend itself to easy visualization. For that reason an algebraic analysis, similar to that of Shockley and Read, is used (see appendix II) to obtain the ratio of electron to hole lifetimes for a single level of states. The following result applies to an N-type semiconductor in which the level of recombination states lies above the middle of the forbidden zone. In this way, terms involving the minority carrier density can be neglected. The ratio is:

$$\tau_n/\tau_p = \frac{(n_0 + n_1)s_n + N_n s_p}{(n_0 + n_1)s_n + N_p s_n} \qquad \ldots (11)$$

Here, n_0 is the thermal equilibrium density of free electrons, n_1, the density of free electrons (as in Shockley–Read), if the Fermi level were at the level of recombination states; N_n and N_p, the number of recombination states occupied in thermal equilibrium by electrons and holes, respectively; and s_n and s_p, the capture cross-sections of these states for electrons and holes. All three terms in the denominator are multiplied by s_n; the separation into two groups has been made for reasons of visual symmetry.

Shockley and Read make the statement that the two lifetimes will be equal if either n_0 or n_1 is large compared with the density of recombination states $N_n + N_p$. The statement is valid if the two capture cross-sections are equal or nearly so. But the very nature of a sensitized photoconductor, *i.e.* one made sensitive by the addition of recombination states, is that these states have a very small cross-section ($\sim 10^{-20}$ cm^2) for one sign of carrier[4, 5, 12, 16, 22, 23]. Further, in so far as this small cross-section is achieved by having doubly

charged states, the capture cross-section for the other sign of carrier is likely to be large ($\sim 10^{-13}$ cm^2). The ratio of capture cross-sections can then be as high as 10^7. In equation (11), if $s_p \simeq 10^7\, s_n$, the free electron density n_0 could be many powers of 10 greater than the density of recombination states and still allow the ratio τ_n/τ_p to be greater than unity. The more exact criterion for equality of lifetimes is that $n_0 s_n$ or $n_1 s_n$ be large compared with $N_n s_p$.

The equality of lifetimes is attained by large n_0 and by large n_1 in opposite ways. Large n_1 means that the level of recombination states approaches close to the conduction band. For $n_1 = n_0$, the states are at the Fermi level, and for $n_1 > n_0$ the states lie above the Fermi level, in region I. As the states depart from the Fermi level in the direction of the conduction band, their contribution to recombination decreases exponentially. The minority carrier lifetime *increases* until it equals the majority carrier lifetime. It is as if the effective number of recombination states decreases until they are less than the added carrier densities. Under these conditions, the two lifetimes must be equal. Large n_1, then, brings about equality by increasing the minority carrier lifetime.

Large n_0, on the other hand, has little effect on the minority carrier lifetime, but a first order effect on the majority carrier lifetime. The latter decreases as n_0^{-1} until it becomes equal to the minority carrier lifetime.

If either n_0 or n_1 is large enough to ensure equality of lifetimes, the problem becomes that already treated in the preceding section. When n_0 and n_1 are small, so that $N_n s_p$ and $N_p s_n$ are the dominant terms, the ratio of lifetimes is:

$$\tau_n/\tau_p = N_n s_p / N_p s_n \qquad \ldots (12)$$

Here one sees immediately that the lifetime of the free electron is given by the time to be captured by one of the empty recombination states, and the lifetime of the hole by the time to be captured by one of the electron-occupied recombination states. The occupation ratio N_n/N_p is the thermal equilibrium value. The ratio $\tau_n \tau_p$ can be either greater or less than unity.

The intermediate case is that $N_n s_p$ is the dominant term in the numerator and either $n_0 s_n$ or $n_1 s_n$ is the dominant term in the denominator. The term $n_1 s_n$ is of less interest. Its contribution to recombination when $n_1 > n_0$ is likely to be small compared with states lying below the Fermi level. To a good approximation one can then write:

$$\tau_n/\tau_p = N_n s_p / n_0 s_n \qquad \ldots (13)$$

Since $N_n s_p$ was taken to be the largest term in the numerator and accordingly larger than $n_0 s_n$, the electron lifetime is larger than the hole lifetime. Also, the added electrons will exceed the added holes. The difference will appear as holes in the recombination states and will be substantially equal to the number of added electrons. In a sense then, the lifetime of a free electron is the lifetime of an added hole in the recombination states. It is the time taken for one of the free electrons to recombine with this hole. That time is, as in equation (13), $(v s_n n_0)^{-1}$.

One has here a clue to account for an optimum in photosensitivity, as a photoconductor is made more N-type. The photoconductor can be made more sensitive by making it more conducting if, by so doing, doubly negatively charged states are introduced into region IV. Once these states are introduced, further increases in conductivity will decrease the electron lifetime by the relation $(v s_n n_0)^{-1}$.

6.5. Multiple Levels and Multiple Classes

A curious and complex problem arises now if the recombination states are distributed in energy. Consider, for example, one level of states near the Fermi level, such that $N_{p1} > n_0$. The level by itself would give an electron lifetime, from equation (12), of $(vs_n N_{p1})^{-1}$. Let a second level of states, equal in number to the first level, be located well below the Fermi level so that $N_{p2} < n_0$. Then, using equation (13), this level by itself would give an electron lifetime of $(vs_n n_0)^{-1}$. The lifetime for the second level alone is larger than that for the first level alone. The problem is how to combine their contributions to lifetime when both levels are present.

Ordinarily, one expects that the shorter of two lifetime mechanisms predominates. Actually, the reverse is true here. The flow of holes into both levels is substantially equal. The 'extra' holes appearing in each level give rise to 'extra' free electrons numerically equal to the 'extra' holes. The problem then is to compute the steady-state increase in density of holes at each level. This requires a knowledge of the lifetime of the hole at each level. In the states near the Fermi level, the lifetime of a hole is $(vs_n N_p)^{-1}$ and is shorter than the lifetime $(vs_n n_0)^{-1}$ of a hole in the states located well below the Fermi level. Accordingly, there will be a larger density of 'extra' holes in the second group of states and these will make the larger contribution to the density of 'extra' free electrons.

One has here two lifetimes that remain operationally separate. This means that when the exciting radiation is cut off, the 'extra' free electrons decay in two steps. A small fraction decays rapidly into the states near the Fermi level; the rest decays slowly into the lower level of states.

Another feature worth noting is that even though the same recombination traffic passes through both levels of states and even though they make equal contributions to the lifetime of free holes, they do not make equal contributions to the lifetime of free electrons.

Finally, one must recognize that only half the exciting radiation, the half that contributes holes to the lower group of states, is supporting the bulk of the 'extra' free electrons. As discussed elsewhere[4], this increases the noisiness of the photo-current.

The recombination properties of states at different energy levels can be reproduced by states at the same energy level but of different classes. The concepts are somewhat easier to visualize here. Consider two classes of states at the same energy level located below the Fermi level. Let the two classes have the same number of states and also the same capture cross-sections for holes. This ensures that the recombination traffic of holes will be evenly split between the two classes. Let the capture cross-sections for electrons be widely different. The free electron lifetime, referred to each class separately, will be $(vs_n n_0)^{-1}$ with s_n taking on different values for each class. Since the rates at which free holes pour into each class are equal, the rates at which free electrons pour into each class must also be equal. These rates are

$$\triangle n_I \, (vs_{nI} n_0) = \triangle n_{II} \, (vs_{nII} n_0) \qquad \ldots \ (14)$$

where $\triangle n_I$ and $\triangle n_{II}$ are the extra free electrons associated with approximately

the same numbers of extra trapped holes in each class. From equation (14)

$$\left(\frac{\triangle n_{\mathrm{I}}}{\triangle n_{\mathrm{II}}}\right)_{\mathrm{free}} = \frac{s_{\mathrm{nII}}}{s_{\mathrm{nI}}} = \left(\frac{\triangle p_{\mathrm{I}}}{\triangle p_{\mathrm{II}}}\right)_{\mathrm{trapped}} \quad \ldots (15)$$

If $s_{\mathrm{nII}} \gg s_{\mathrm{nI}}$, then $\triangle n_{\mathrm{I}} \gg \triangle n_{\mathrm{II}}$. One sees here again that almost all of the extra free electrons are created by only half the absorbed light. The fraction one-half could, by another choice of parameters, be a much smaller fraction. In that case a small fraction of the absorbed light is responsible for most of the extra free electrons, and the photocurrents become proportionately more noisy. One sees here also that while each class of states contributes equally to the recombination traffic they make unequal contributions to the lifetimes of the extra free electrons.

7. TRANSITION FROM SEMICONDUCTORS TO INSULATORS

The lifetimes of electrons and holes have been discussed thus far for two broad classes of materials: relative insulators, in which the added carrier densities were large compared with the thermal densities and in which the major recombination traffic is carried by states (region III, Figure 2) whose occupancy is determined by kinetic processes; and relative semiconductors, in which the added carrier densities were small compared with the thermal densities and in which the major recombination traffic is carried by states (region IV, Figure 4) whose occupancy is substantially that of thermal equilibrium. What is missing is the transition from semiconductors to insulators. It is useful to outline the transition in terms of the shift from thermal to kinetic control of the occupancy of states. This is done below in Figure 7 using the concepts of the demarcation levels already developed.

The lifetime of the minority carrier can always be estimated with good accuracy as the time it takes a free hole to be captured by any of the electron-occupied states in region III and IV. If the state density is small compared with the added carrier density, the minority carrier lifetime is also the majority carrier lifetime. In general, this will not be true and one must separately compute the lifetime of the majority carrier.

If the lifetimes are not equal and if, at the same time, the thermal density of electrons n_0 is large compared with the density of states in regions III and IV then, by arguments given in the previous section, the electron lifetime is $(vs_{\mathrm{n}}n_0)^{-1}$. If the thermal density is not large compared with the density of states, then there will be some level (below the electron Fermi level) for which the thermal density n_0 is large compared with the density of empty states N_{p}. The electron lifetime is still $(vs_{\mathrm{n}}n_0)^{-1}$ but is referred to the states in regions III and IV for which $n_0 > N_{\mathrm{p}}$. The significance, as discussed in the previous section, is that only part of the exciting radiation, the part that contributes holes to these lower states, supports the bulk of the added free electron density.

As the added carrier densities increase, in Figure 7, region III, the region of kinetic processes, increases. The first significant increase comes (Figure 7(*b*)) when the minority carrier density exceeds its thermal value. The arrows on the right show the direction of movement of the levels with increasing excitation.

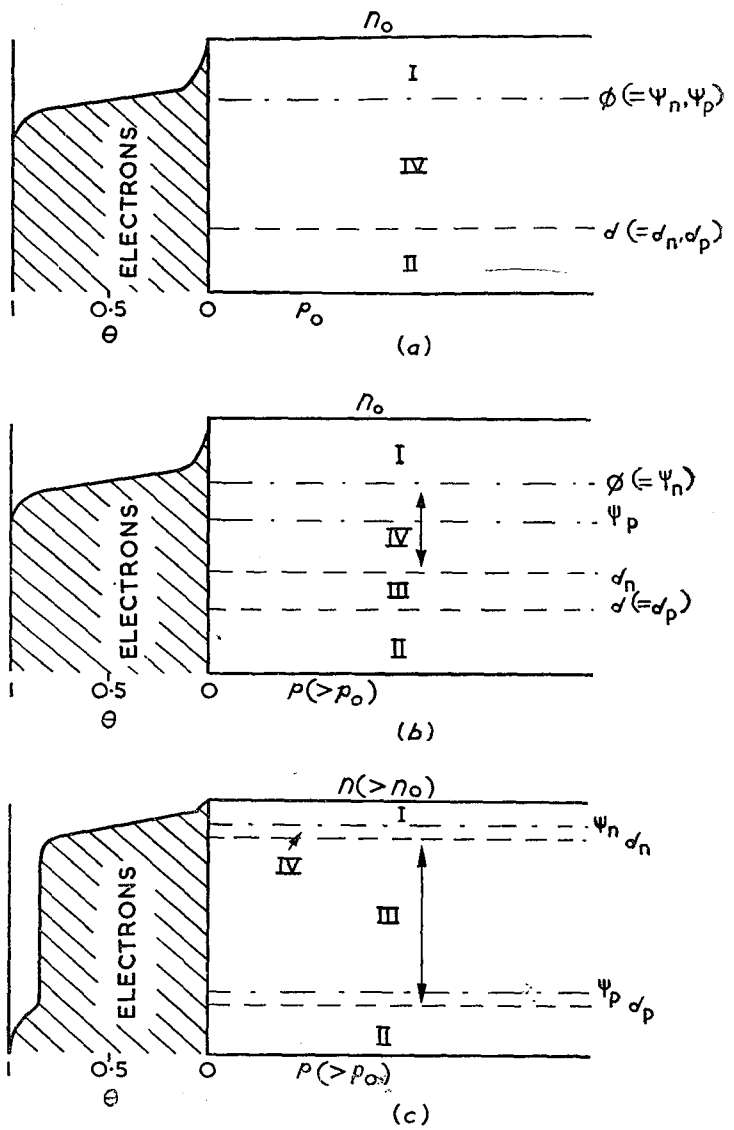


Figure 7. *Transition of distribution function for electrons and holes from semiconductors to insulators*

The occupancy of region III is constant and controlled by kinetic processes as opposed to thermal exchange processes. At still higher excitations (Figure 7(*c*)) the majority carrier density also exceeds its thermal value. By this time region III plays the dominant role in recombination and lifetime. The problem becomes that discussed at the beginning of this paper.

APPENDIX I—DEMARCATION LEVEL *d*

The condition for equality of kinetic and thermal processes for a hole-occupied level in N-type material is (like equation (1); Section 4).

$$v^*_p \exp(-\triangle\epsilon/kT) = n_0 v s_n \qquad \ldots \text{(A1)}$$

where $\triangle\epsilon$ is the energy distance from the filled band.
From equation (A1):

$$\triangle\epsilon = kT\ln\left(\frac{v^*_p}{n_0 v s_n}\right)$$

$$= kT\ln\left(\frac{v^*_p}{n_0 v s_p}\cdot\frac{s_p}{s_n}\right) \qquad \ldots \text{(A2)}$$

From detailed balance, $v^*_p/vs_p = N_v = N_c$

Equation (A2) then becomes:

$$\triangle\epsilon = kT\ln\left(\frac{N_c}{n_0}\right) + kT\ln\left(\frac{s_p}{s_n}\right)$$

$$= \varphi_n + kT\ln\left(\frac{s_p}{s_n}\right) \qquad \ldots \text{(A3)}$$

Equation (A3) indicates that the demarcation level d is the same distance from the filled band as the Fermi level φ is from the conduction band plus the correction term $kT\ln(s_p/s_n)$.

If the same argument is carried out for an electron-occupied level the result is:

$$\triangle\epsilon = \varphi_p - kT\ln(s_p/s_n) \qquad \ldots \text{(A4)}$$

where $\triangle\epsilon$ is measured from the conduction band and φ_p from the filled band. In a material in thermal equilibrium φ_n and φ_p are the same level measured from the conduction and filled bands respectively. Equation (A4) then, defines the same level d as does equation (A3).

APPENDIX II—GENERAL FORMULAE FOR SINGLE LEVEL OF RECOMBINATION STATES

In thermal equilibrium one can write:

$$n_0 N_p v s_n - N_n v^*_n \exp(-\epsilon_{tn}/kT) = 0 \qquad \ldots \text{(A5)}$$

and

$$p_0 N_n v s_p - N_p v_p^* \exp(-\epsilon_{tp}/kT) = 0 \qquad \ldots \text{(A6)}$$

where the first term in each expression describes the capture of electrons or holes into the recombination states, and the second term describes the thermal generation of electrons or holes from these states. ϵ_{tn} and ϵ_{tp} are the energy separations of the recombination states from the conduction and filled bands respectively. If thermal equilibrium is now disturbed by optical excitation, n_0, p_0, N_n and N_p become $n_0 + \delta n$, $p_0 + \delta p_0$, $N_n + \delta N_n$, $N_p + \delta N_p$. These values are inserted in equations (A5) and (A6) and the two equations are set

equal to satisfy the condition of steady state. Further side conditions are:
(1) by charge neutrality,

$$\delta n - \delta p = \delta N_p$$
$$\delta N_p = -\delta N_n$$

(2) by detailed balance,

$$v^*_n/vs_n = N_c$$
$$v^*_p/vs_p = N_v$$

(3) by definition,

$$n_1 = N_c \exp(-\epsilon_{tn}/kT)$$
$$p_1 = N_v \exp(-\epsilon_{tp}/kT)$$
$$\tau_n/\tau_p = \delta n/\delta p$$

The result is:

$$\tau_n/\tau_p = \frac{(N_0 + n_1)s_n + (N_n + p_0 + p_1)s_p}{(n_0 + n_1 + N_p)s_n + (p_0 + p_1)s_p} \quad \ldots \text{(A7)}$$

For N-type material and for the recombination states located above the middle of the forbidden zone, one may frequently neglect the terms in p_0 and p_1 compared with terms in n_0 and n_1.
The approximate result is then:

$$\tau_n/\tau_p = \frac{(n_0 + n_1)s_n + N_n s_p}{(n_0 + n_1)s_n + N_p s_n} \quad \ldots \text{(A8)}$$

ACKNOWLEDGEMENT

The writer is indebted to M. A. Lampert for some clarifying discussions on the significance of the demarcation levels.

REFERENCES

1. W. Shockley and W. T. Read. *Phys. Rev.* **87,** 835 (1952).
2. R. N. Hall. *Phys. Rev.* **87,** 387 (1952).
3. M. Schön. Techn.-Wiss. Abhandlungen, Osram-Ges. **6,** 49 (1953).
4. A. Rose. *Atlantic City Conference on Photoconductivity* (Wiley, New York, 1956).
5. I. Broser and R. Broser-Warminsky. *Ann. Phys., Lpz.* **16,** 361 (1955).
6. A. Rose. *R.C.A. Rev.* **12,** 362 (1951).
7. A. Rose. *Phys. Rev.* **97,** 1538 (1955).
8. A. Rose. *Phys. Rev.* **97,** 322 (1955).
9. C. A. Duboc. *J. Appl. Phys. Suppl.* **4,** 107 (1955).
10. W. H. Isay. *Ann. Phys., Lpz.* **13,** 327 (1953).
11. J. R. Haynes and J. A. Hornbeck. *Phys. Rev.* **94,** 1438 (1954).
12. R. W. Smith. *R.C.A. Rev.* **12,** 350 (1951).
13. R. W. Smith. *Phys. Rev.* **93,** 347 (1954); **98,** 1169 (1955).
14. H. S. Somers, Jr., R. E. Berry and I. Sochard. *Phys. Rev.* **101,** 987 (1956).
15. S. V. Forgue, R. R. Goodrich and A. D. Cope. *R.C.A. Rev.* **12,** 335 (1951).
16. R. H. Bube. *Proc. Inst. Radio Engrs., N.Y.* **43,** 1836 (1955).
17. H. W. Leverenz. *An Introduction to Luminescence of Solids* (Wiley, New York, 1950).
18. R. H. Bube. *Atlantic City Conference on Photoconductivity* (Wiley, New York, 1956).
19. R. H. Bube and S. M. Thomsen. *J. Chem. Phys.* **23,** 15 (1955).
20. W. Van Roosebroeck and W. Shockley. *Phys. Rev.* **94,** 1558 (1954)
21. D. Redfield. *Phys. Rev.* **100,** 1094 (1955).
22. F. Stöckmann. *Z. Phys.* **143,** 348 (1955).
23. R. Newman and W. W. Tyler. *Phys. Rev.* **96,** 882 (1954).

THE PRODUCTION OF HIGH-QUALITY GERMANIUM SINGLE CRYSTALS

I. G. CRESSELL, B.Sc., A.Inst.P., A.M.I.E.E.
and
J. A. POWELL, M.A., D.Phil.

Baddow Research Laboratories
Marconi's Wireless Telegraph Company Ltd., Chelmsford, U.K.
MS. received 27 August, 1956

THE PRODUCTION OF HIGH-QUALITY GERMANIUM SINGLE CRYSTALS

1. INTRODUCTION

Since the original production by Teal and Little[1] of single-crystal germanium for semiconductor device purposes, increasing stress has been placed upon improvements in crystal structure, the distribution of deliberately added impurities and the elimination of deleterious trace elements. These developments have been accelerated by the need for larger uniform single crystals, not only for transistor production, but also for power rectifiers and other devices. Furthermore, research on the higher calibre material resulting from the improved techniques is leading to a better understanding of fundamental semiconductor properties (see Pearson and Brattain[2]). Modern trends demand the rigid control of semiconductor parameters for a range of resistivities far wider than was originally envisaged, and have also necessitated a continuous reduction in crystal defects, an increasing homogeneity and an improved minority carrier lifetime.

The purpose of the present paper is to survey the available methods of crystal growth and to discuss in particular the horizontal zone melt technique as applied to the production of germanium single crystals, ranging in resistivity from the intrinsic resistivity to that normally used for transistors. As an introduction to the problems of single crystal growth a brief summary is given of the types of imperfection which may enter the crystal during growth and of the effects of these imperfections upon the electrical properties. The bibliography is chosen to be representative rather than exhaustive.

2. CRYSTAL IMPERFECTIONS

A synthesis of imperfections in single crystals has been given by Seitz[3]. This summary will be confined to single-crystal germanium (*i.e.* free from wide-angle grain boundaries), which may contain imperfections of the atomic type, *viz.*:

(1) Foreign atoms.

(2) Dislocations.

(3) Vacant lattice sites and interstitial atoms.

Such imperfections can give rise to localized states, with energy levels within the energy gap of the semiconductor. Their presence may be inferred from the electrical behaviour, or, in the case of certain types of dislocation, detected by etch techniques. It should be emphasized that interaction between imperfections of the same type and of different types is to be expected in a real crystal, and that only in special cases can the dominant effect of a single type of imperfection be established experimentally. The influence of the free surface will not be considered, except in so far as it affects the measurement of bulk properties.

2.1. Foreign Atoms

The foreign atoms in the crystal may be conveniently sub-divided into two main groups:

(1) Substitutional impurities from Groups III or V of the periodic table; these are deliberately added to the germanium to give rise to acceptor and donor levels respectively in the energy gap. Available substitutional impurities are boron, aluminium, gallium and indium from Group III, and phosphorus, arsenic and antimony from Group V. Such elements have been shown to possess a differential solubility between the solid and liquid phases of germanium, and their distribution coefficients have been comprehensively reviewed by Burton[4]. A correlation exists between the distribution coefficient and the tetrahedral covalent radius of the impurity atom, and the wide range of values permits the choice of an additive suited to the particular crystal growing process. The ionization energies appropriate to these substitutional impurities in germanium are similarly listed by Burton[4], and the theory of their electrical behaviour in germanium has been recently reviewed by Harvey Brooks[5].

(2) Metallic impurities from other groups of the periodic table; these may be broadly classified as trace elements and will normally fall into the category of contaminants. Most of these solute elements either have large radii or do not normally form covalent bonds; they tend to have small distribution coefficients, small solubilities and large diffusion coefficients. At comparatively large concentrations, they may exhibit the phenomenon of retrograde solid solubility, which leads to complex behaviour on heat treatment. These properties of copper, nickel, lithium, silver, cobalt, zinc and gold in germanium are summarized by Burton[4].

The most extensively investigated contaminant is copper[6, 7, 8, 9, 10, 11], the presence of which, in germanium, has been shown to give rise to two energy levels in the energy gap and to catalyze the electron–hole recombination process. Burton, Hull, Morin and Severiens[12] have studied the influence of copper and nickel on the lifetime of minority carriers in germanium, using the Hall–Shockley–Read model of electron–hole recombination[13]. They point out that the minority carrier lifetime in 'pure' germanium could be accounted for by residual traces of these elements at concentrations of the order of 10^{-10} to 10^{-11} atom fraction.

The presence of iron[14], cobalt[15], manganese[16] and platinum[17] in germanium have each been studied, and multiple ionization energy levels in the energy gap have been introduced to explain the observed electrical behaviour. As many of the above impurities give rise to energy levels near the centre of the forbidden energy gap, they act as traps or recombination centres and are instrumental in reducing the minority carrier lifetime.

Such contaminants may enter the germanium during the final crystal growth operation, following the zone purification processes, or during any subsequent heat treatment of the material. Their elimination is necessary in the production of high-quality material.

Gaseous elements, *e.g.* nitrogen, oxygen and hydrogen, dissolved in germanium may also be of importance in the preparation of high-quality material since their presence has been indicated by the vacuum fusion gas analysis of

germanium reported by Hassion, Thurmond and Trumbore[18]. The presence of 10^{15} to 10^{16} cm^{-3} interstitial oxygen atoms is also suggested by the infra-red absorption measurements on germanium by Kaiser, Keck and Lange[19]. Although no direct information of the electrical effects of these dissolved gaseous elements in germanium is available, experiments described later, in Section 5.2, suggest that a reduction in the minority carrier lifetime of germanium may be attributed to dissolved hydrogen in the lattice.

2.2. Dislocations

The importance of dislocations in a crystal lattice has long been realized and for comprehensive reviews of the subject the articles by Read[20] and Cottrell[21] should be consulted. Dislocations in germanium are believed to originate from one or more of the following causes:

(1) As a result of slip, occurring to effect stress relief in the lattice, with the simultaneous formation of edge dislocations. This may occur either during the growth process or in subsequent heat treatment, since germanium is plastic[22] above about 500° C.
(2) As a result of the condensation of lattice vacancies during the growth process[23].
(3) When solute atoms are misfits in the germanium lattice and are in a sufficiently large concentration (fractional atomic per cent) to cause excessive lattice strain[14, 15, 16].
(4) As a defect inherent in crystal growth.

When a germanium crystal surface is subjected to etching treatments a microscopic investigation reveals several types of surface structure, some of which are characteristic of edge-type dislocations. The etch techniques introduced by Vogel[24] have provided a powerful tool for the estimation of edge dislocation densities. He found, from X-ray and microscopic studies, that a correlation between the observed etch pit spacings and the dislocation spacings existed at a low-angle boundary. He concluded from this that the observed etch pits occur at the points of emergence of edge dislocations. This type of low-angle boundary is commonly called lineage and is introduced during crystal growth. It has been shown by Vogel, Read and Lovell[25] that a single lineage line acts as a centre for the recombination of minority carriers and the problem has been examined critically by McKelvey and Longini[26].

Experiments on the plastic deformation of germanium single crystals by Gallagher[22], Greiner[27] and Treuting[28] provide evidence that slip occurs on a macroscopic scale in a $<110>$ direction, that the slip plane is the {111}, and that there is an incubation period prior to any substantial plastic flow, following the application of a tensional stress. The traces of active slip planes are revealed by the arrays of etch pits associated with edge dislocations, as first observed by Gallagher[29] and, by introducing controlled numbers of edge dislocations by bending, it was shown that plastic deformation introduced acceptor levels[30] in the energy gap. A theory of dislocations is given by Read[31], using Shockley's model[32] of the dangling unpaired bond, in which such bonds are identified with the dislocation acceptor density.

It is believed that a single dislocation line gives rise to about 4×10^6 cm^{-1} potential acceptor centres, reduces the mobility of charge carriers through the

agency of scattering and also acts as a recombination centre for minority charge carriers, as shown by Kurtz, Kulin and Averbach[33], who find a correlation between minority carrier lifetime and dislocation densities in the range 10^4 to 10^7 cm^{-2}. At the time of writing the authors have found no such correlation for dislocation densities less than about 10^3 cm^{-2}. Rosi[34], in recent work, observes that data on minority carrier lifetime as a function of edge dislocation density exhibit a maximum at about 2×10^3 cm^{-2} and that such behaviour may be interpreted in terms of interactions between dislocations and various other recombination centres in the crystal.

The etch pit technique has been refined and extended by Ellis[35] to include slow etch procedures, by which means the structure of individual etch pits may be resolved. He identifies three types of pit:

(1) Large pits characteristic of primary edge dislocations similar to those observed by Vogel[24].
(2) Small pits with closed terraces.
(3) Pits with spiral terraces, indicative of a dislocation with a screw component, normally obtained in dendritic specimens only.

Class (1), the primary edge dislocations, have already been discussed and class (3) will not normally be observed under the conditions of near equilibrium crystal growth. The class (2) etch pit is most easily revealed on {111} surfaces of low primary edge dislocation density. Owing to their depth and character, a type of composite dislocation is suggested for the small pits, but no evidence exists of their electrical effects in the germanium crystal.

2.3. Vacant Lattice Sites and Interstitial Atoms

A temperature dependent concentration of lattice vacancies and interstitial atoms may be generated as equilibrium imperfections in a crystal and under certain conditions may be free to migrate. When germanium is heated to a high temperature and quenched rapidly, acceptor levels are introduced into the energy gap and the minority carrier lifetime suffers degradation. Several workers have endeavoured to identify these effects in germanium with the creation and 'freezing in' of lattice vacancies or interstitial atoms or both, but it is increasingly recognized that a major cause of these thermally induced acceptors may arise from the rapid solid phase diffusion of chemical impurities, notably copper[9], and the dislocations associated with plastic deformation[30] during the heat treatment.

The effects of copper in the lattice have been studied by Logan[36], who shows than when copper-doped germanium is cooled from a high temperature, copper precipitation occurs at a rate dependent upon the dislocation density. In conjunction with Schwartz[11], Logan also describes experiments on the restoration of resistivity and lifetime of heat-treated germanium. In these, copper is intentionally diffused into germanium and then 'gettered out', using the favourable distribution coefficient ratio of copper in germanium and copper in the ternary system, which germanium and copper form with other metals such as gold, indium and lead.

With this increased understanding of the behaviour of copper in germanium, procedures have been devized for minimizing the effects of both copper and

other chemical contaminants, and of plastic deformation during heat treatment[37, 38, 39] and interest has centred on the origin and behaviour of the remanent thermally produced acceptor centres. As Logan[39] in his more recent work points out, however, the observed characteristics of these centres do not permit an unambiguous identification or detailed understanding of their behaviour, and interpretation of data is difficult, probably owing to the interactions between the same type or other types of defects in the crystal lattice.

Alternative approaches designed to improve the understanding of vacancy-interstitial behaviour have been made by studies of the electrical properties of nucleonic-bombarded germanium with subsequent annealing[40], and of the mechanism of self-diffusion[41] but, at the time of writing, no consistent theory is available.

3. CRYSTAL GROWTH FROM THE MELT

Crystal growth from the melt has been recently discussed by Chalmers[42] and Ellis[35], and reviewed by Martius[43], and it is believed that the screw dislocation mechanism invoked by Frank[44] for the growth of single crystals from the vapour phase is not applicable. Chalmers proposes other possible mechanisms which can account for the experimental findings of growth from the melt and concludes that there is no evidence to suggest that dislocations are a prerequisite.

In the preparation of single-crystal germanium for device or research purposes the following conditions must be satisfied:

(1) *A stable orientation.* Growth must be promoted on a stable predetermined orientation in order to eliminate the possibilities of twinning, lineage and polycrystallinity, since crystals 30 cm long or greater may be grown. The growth planes commonly used for germanium are the {100} and {111} and the choice is primarily governed by the degree of perfection that may be achieved in growth, with the actual device or research requirement a secondary consideration.

In terms of crystal perfection the {100} growth plane is believed to be disadvantageous and lineage is commonly observed. Such lineage is evident along lines of intersection of {110} planes with the etched {100} cross-sectional surfaces of a grown crystal. Typical patterns for surfaces examined in zone-levelled crystals are illustrated in Figure 1. Arrays of dislocations lying in the two {110} planes which are orthogonal to the {100} cross-section tend to remain stable, whilst the others are more irregular and unstable. These experimental findings are in general agreement with those reported by Ellis[35], and Pfann and Lovell[45].

The {111} growth plane is highly stable and lineage is much more uncommon. Lineage, when present, consists of arrays of dislocations which may be revealed by a suitable etchant, and appears in the lines of intersection of {110} planes and the {111} growth plane as shown in Figure 2. As discussed in Section 5.3 proper control of the solid–liquid interface during crystal growth coupled with the stability of temperature gradients and growth rate eliminates completely the incidence of twinning and lineage on the {111} growth plane.

(2) *A controlled level of dislocation density.* In addition to the effect that edge dislocations have on the minority carrier lifetime, it is believed that they also have a considerable influence on the reverse breakdown voltage of an alloy junction device and upon the wetting properties of a metal when melted in

contact with the germanium, *e.g.* indium on germanium during the junction alloying process. The control of dislocation densities is of primary importance therefore in material for device fabrication, and there is a general trend to include an etch pit count in the germanium specification, in addition to the normal parameters.

Two major types of primary edge dislocation can be distinguished on the etched {111} plane of a grown crystal. The most common are those associated with the comparatively widely spaced etch pits lying in arrays on the lines of intersection of {111} planes with the {111} etched surface, which appear to be the traces of operative slip planes. In view of the low plastic temperature of germanium[29] it is essential that the grown crystal remains unstrained after growth, particularly whilst it is moving through the axial temperature gradient of the crystal growing furnace. The stabilization of radial temperature gradients, and their reduction to a low value throughout that region of the crystal above the plastic temperature, to give near parallel isotherms will fulfil this condition. In the absence of lineage and with a reduction in slip, a random distribution of primary edge dislocations may be revealed on an etched {111} surface. The density of these dislocations appears to be correlated with the angular separation between the solid–liquid interface and the principal growth plane, and tends to zero when they are coincident.

To obtain dislocation densities lower than 10^3 cm^{-2} therefore, the condition of near plane parallel isothermals in the grown crystal, normal to the growth direction, must be satisfied. Methods of achieving this will be discussed in Section 4.

(3) *A predetermined impurity distribution.* The distribution of solute in crystals grown from the melt has been investigated analytically and experimentally by Burton, Prim, Slichter, Kolb and Struthers[46] and is shown to be critically dependent upon stirring of the melt. Their paper is devoted primarily to mechanically stirred melts controlled at a temperature just above the melting point, using the vertical puller. However, much of their work is applicable to the horizontal zone melt, where, in contrast, stirring is by convection currents in the molten zone, where resistance heated furnaces are used, and by induced eddy currents, where directly coupled radio-frequency heating is used. Unless a high degree of mixing is maintained in the melt, solute striation effects may be introduced into the crystal. Such striations may be detected either by auto-radiographic techniques[46] or, more simply, by selective plating[47]. As shown by Haynes and Pearson[46], they cause irregularities in the resistivity and in the minority carrier lifetime.

Burton *et al.*[46] have discussed the inter-relation between growth rate and distribution coefficient for a range of stirring conditions and the formation of an enriched diffusion-dominated solute layer immediately ahead of the growing solid–liquid interface is now well understood. Solute elements such as antimony and gallium are highly sensitive to growth rate variations.

Thus, the requirements to preserve the axial and radial temperature gradients for stability and planarity of the liquid–solid interface, throughout growth, are further emphasized by the need to eliminate inhomogeneous solute distribution arising from growth rate fluctuations.

There are two different techniques which have been applied with varying

degrees of success to the growing of single crystals from the melt, *viz.* the vertical pulling method and the horizontal zone melt method.

3.1. The Vertical Pulling Method

The earlier favoured technique for the growth of germanium single crystals was developed from an apparatus originated by Czochralski[48] for determining the speed of crystallization of different metals. Teal with Little[49], and with Sparks and Buehler[50] are chiefly responsible for adapting and engineering this method to a stage where the modern vertical puller is a very precise machine capable of high orders of control over the crystal growth. Other similar apparatus is described by Lehovec *et al.*[51] and by Roth and Taylor[52].

Briefly, the method consists of dipping a small orientated single crystal of germanium into a graphite crucible containing molten germanium, then adjusting the heat input and heat losses from the crucible and along the crystal until the desired size of crystal is obtained at the desired pull rate. A photograph of a vertical puller is given in Figure 3. Heating of the charge may be achieved by either radio-frequency or resistance, and a servo-mechanism normally controls the heat input to maintain the melt temperature to within 1° C of a preset value, which may be programmed to give desired properties to the growing crystal. The chuck holding the seed crystal is attached to a rod which is mechanically linked to a variable-speed electric motor for withdrawing the crystal at a preset or programmed rate. A second motor is normally used to rotate the seed crystal and impart a degree of mechanical stirring to the molten germanium to improve the homogeneity of impurity distribution in it.

If subjected to gross strain during growth, twinning, polycrystallinity and other uncontrolled defects will result in the crystal. Since germanium expands by about 5 per cent on freezing, it is most important that all constraints imposed upon the crystal by a containing vessel are avoided. It is probably the absence of any form of container for the grown crystal that has led to the vertical puller becoming the preferred method of growing germanium crystals. Other advantages of this technique are:

(1) The conditions of thermal symmetry in the growing crystal are such that the isothermals will be approximately parallel, although not necessarily planar.

(2) The steep temperature gradient maintained in the crystal near the solid–liquid interface will tend to minimize the back diffusion of impurities into the crystal from the melt, the diffusion of any contaminants which may have condensed on the grown crystal, and the effects of heat treatment. It will also add stability to the solid–liquid interface where the ratio of heat flux down the crystal to the rate of release of the latent heat of fusion of germanium is required to be large.

(3) Continuous stirring by the rotating crystal maintains a more nearly uniform distribution of impurity in the melt and, at sufficiently large linear velocities at the interface, will partially overcome the normal tendencies for non-planarity of the interface.

(4) The near circular symmetry of the crystal cross-section is more suitable for some devices (*e.g.* large area power rectifiers).

(5) The vertical puller possesses a flexibility which renders it adaptable to several specialized techniques. In particular there are the *sequential doping* and

rate growing methods of producing several P–N junctions in a single crystal growing operation as described by Teal, Sparks and Buehler[53], Hall[54] and Bridgers[55].

For the production of large quantities of single crystal germanium of specific characteristics, however, the vertical puller possesses several inherent disadvantages, not the least of which originates from the segregation phenomenon, which occurs at the solid–liquid interface during the crystal growing operation. As the solute elements with which one is chiefly concerned have distribution coefficients less than unity, the germanium melt is continually enriched with solute as the crystal grows. Consequently, the crystal will exhibit a solute concentration (and hence a specific conductivity) continuously increasing along its length from the end first pulled, according to the relationship

$$C = kC_0(1 - x)^{(k-2)}$$

C_0 is the initial concentration in the melt, C is the concentration of solute in the solid at any point x, where x is the fraction of the original volume which has solidified, and k is the effective distribution coefficient (see Pfann[56]). Since the distribution coefficient of an impurity is a function of pull rate[46], attempts have been made to counter this variation in solute concentration by programming the crystal growth rate and melt temperature accordingly. These systematic changes in growth parameters, however, cause a departure from those conditions leading to uniformly high quality material, and the minority carrier lifetime suffers degradation. Other workers, notably Nelson[57], have endeavoured to overcome this difficulty by continuously adding doped germanium to the melt in such a manner that the solute concentration is maintained constant. Such a system requires much elaboration, particularly on the molten germanium level control, and adds considerably to the complexity of the apparatus.

In the experimental results described later, it is shown that critical control over the shape of the interface and the temperature gradient is necessary to produce uniformity in the crystal. Without still further elaboration in furnace control these conditions are not readily achieved in the vertical puller over long lengths of crystal (see Billig[58]).

For these reasons, there is now a trend towards the horizontal zone-melt method for the production of large quantities of germanium of predetermined characteristics, since the vertical puller, though extremely adaptable to specialized techniques, does not readily lend itself to complete automation.

3.2. The Horizontal Zone Melt Method

The progressive solidification from one end of a molten metal contained in a horizontal crucible is a well-known metallurgical technique. This technique has been applied to germanium processing largely by Pfann[56, 59] and his co-workers, who have developed the zone-melting technique for the purification of germanium, and the zone-levelling technique for the preparation of single crystals with a uniform resistivity along their length.

In the zone-melt process a horizontal, carbon-coated silica crucible is loaded with an orientated single crystal seed, a germanium slice containing a measured amount of impurity, and an ingot of purified germanium. A molten germanium

zone, produced by either radio-frequency or resistance heating, is then established in contact with the seed crystal and is caused to traverse the length of the charge at a constant rate. The whole ingot is then a single crystal having the orientation of the seed, with a solute concentration distributed according to the relation (see Pfann[56, 59])

$$C = kC_0 \exp - (kx/l)$$

where C is the solute concentration in the solid, C_0, the initial solute concentration in the molten zone, k the distribution coefficient, l the molten zone length and x is the distance moved by the solidifying interface. The above formula is valid provided the following assumptions are satisfied: (1) there is complete mixing in the melt; (2) k is constant; (3) diffusion in the solid is negligible; and (4) the molten zone remains constant in volume. The furnace temperature is controlled to within $\pm$ 1° C of a preset value, and a stabilized traverse rate is employed.

The simplicity of the apparatus and the consequent ease with which automation may be applied, together with the advantages of zone-levelling, render this method of growing germanium single crystals with uniform characteristics, most attractive for both production and research purposes. Whilst the theories of zone-melting and subsequent zone-levelling processes have been fully explored, little has been published on the experimental conditions necessary to obtain satisfactory single crystal growth by this method. In their very recent work Bennett and Sawyer[60] have achieved considerable progress towards the control of dislocation density and homogeneous solute distribution for the resistivity range 2–8 Ωcm. Their methods will be discussed later.

The advantages of this technique over the vertical puller may be summarized thus:

(1) The human element may be eliminated after the initial setting-up, thus adding to the degree of reproducibility which may be achieved in the grown crystals.
(2) The comparative ease with which crystals of uniform predetermined cross-section may be obtained.
(3) The support, by the crucible, of the grown crystal allows the production of single crystals up to several kg in weight.
(4) The relative ease with which predetermined concentrations of impurity solute may be introduced into the crystal and the large proportion of uniform material obtained.
(5) The greater independence of growth parameters which permits an extremely precise control to be established.

That this method has not been universally adopted is probably the result of the following disadvantages associated with the use of a crucible:

(1) As the thermal conductivity and emissivity of the refractory crucible differ widely from those of the melt, radial thermal symmetry about the crystal growth axis may be difficult to achieve.
(2) The constraint which may be imposed upon the crystal during solidification places restrictions on the crucible geometry.
(3) Owing to the large area of contact between the melt and the crucible, there is an increased risk of contamination.

(4) Greater care is needed to ensure adequate stirring of the melt by radio-frequency eddy currents or thermal convection currents to secure homogeneity of impurity distribution in the solid.

In the following sections, methods will be discussed for overcoming the above difficulties and results will be presented to show that the horizontal zone-melt technique is capable of producing high quality germanium single crystals with near uniform resistivity and a low level of dislocation density in large quantities.

4. THE SINGLE CRYSTAL ZONE-MELT PROCESS

The principles of zone-levelling have been discussed in detail by Pfann[56, 59] and are briefly summarized in the previous section. Experimentally, the assumptions in Pfann's analysis are most nearly satisfied by imposing steep axial temperature gradients on the ingot at the two molten–solid interfaces forming the boundaries of the molten zone. This ensures a stable zone-length and a high degree of thermal stirring; it also minimizes the back diffusion of impurities from the melt into the crystal and avoids exposing the grown crystal to elevated temperatures for extended periods of time. As originally developed, the molten zone was produced by the radiation from a graphite susceptor heated by radio-frequency induction, and movement of the zone was achieved by sliding the silica crucible containing the ingot, over silica guide rods. Examination of zone-levelled crystals produced in this manner, using axial temperature gradients of the order of 150° C cm^{-1}, revealed that the solid–molten interface of the growing crystal was concave and that lineage and a high density of edge dislocations were common, as shown in Figure 2.

To understand why this should be so, it is necessary to consider the heat flow pattern in the ingot and its environment, and the effect these have both on the growth process at the solid–liquid interface and subsequently on the grown crystal due to thermal stress. Consider first the distribution of heat in the idealized case of a bar of circular cross-section, occupying the same location as a growing crystal in the horizontal zone melt furnace, with steep axial temperature gradients. In the steady state, heat loss from the bar will be radially symmetrical, and a radial temperature gradient will exist which will be dependent upon the thermal properties of the bar and the temperature differences between the radiating surface and the surroundings. Under these conditions the isothermal surfaces in the bar will be concave towards the centre of the furnace, and *thermal coring* will be present, as shown schematically in Figure 4. The presence of such isothermal surfaces in a germanium crystal will cause the outermost parts of its cross-section to be in a state of tension and its inner core to be in a state of compression. The converse will be true if the isothermal surfaces are convex towards the centre of the furnace. As germanium is plastic above 500° C, there will exist a critical radius in the isothermal surfaces for which strain can no longer be sustained and slip along {111} planes will occur to effect stress relief in the crystal. This problem has recently been discussed in detail by Bennett and Sawyer[60], Rosi[34], and Billig[58] to explain the star-like patterns of etch pits observed on the cross-sections of crystals grown by the Czochralski technique. Similar patterns, as shown in Figure 5, have been observed by the authors in

crystals grown by the horizontal zone melt technique in which the isothermal surfaces were convex towards the furnace.

Generally, however, there is a considerable departure from the idealized case described above for the horizontal zone-melt, since the crystal is normally supported by a silica crucible having a surface emissivity widely divergent from that

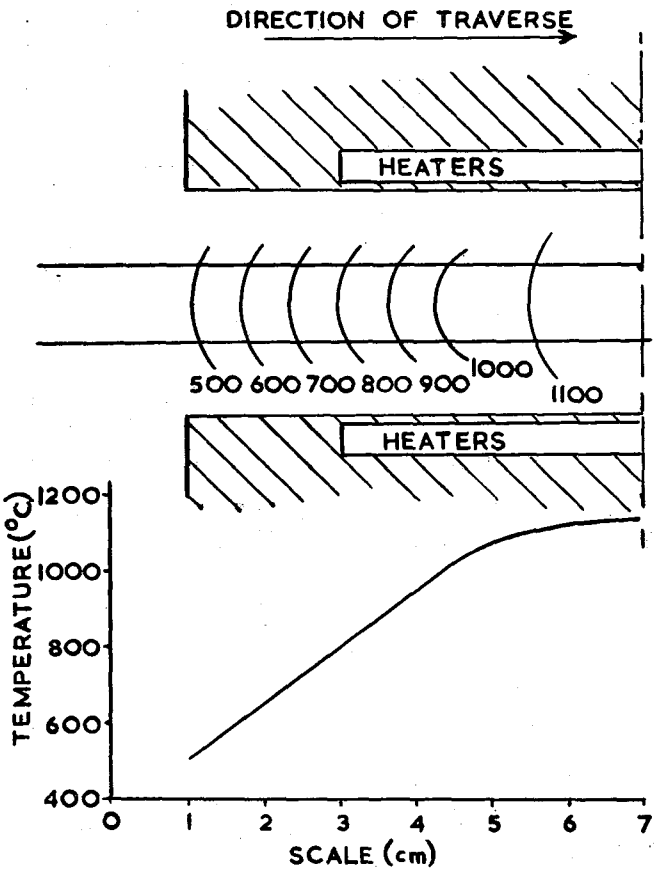


Figure 4. *Schematic diagram showing the symmetrical distribution of isothermals in a bar, in a crystal growing furnace.*

of the germanium free surface. This may lead to an angular displacement from the vertical plane of up to 10° of arc in the isothermal surfaces at the molten–solid interface, and to non-parallel isothermals down the crystal as shown schematically in Figure 6. Using similar arguments to those above, thermal stresses may be set up in the crystal and, again, slip may occur to effect stress relief. Under these asymmetrical conditions, the star pattern of etch pits is not normally observed and, in general, it is replaced by preferential slip due to resolved shear stress on particular {111} planes, the indices of which will depend upon the distribution of {111} planes with respect to the crucible, as shown in

Figure 7. A further asymmetry in the system is caused by convection currents in the ambient gas, but these may be suppressed by the correct adjustment of gas flow through the furnace tube. In addition to the edge dislocations associated with slip, which are believed to result from excessive thermal stresses in the solidified crystals, other dislocations, which are believed to be more directly related with crystal growth, are observed, *viz.* those associated with lineage and those which appear to be axial in the <111> growth direction, as discussed in Section 5.3. The elimination of all such dislocations demands a solid–liquid

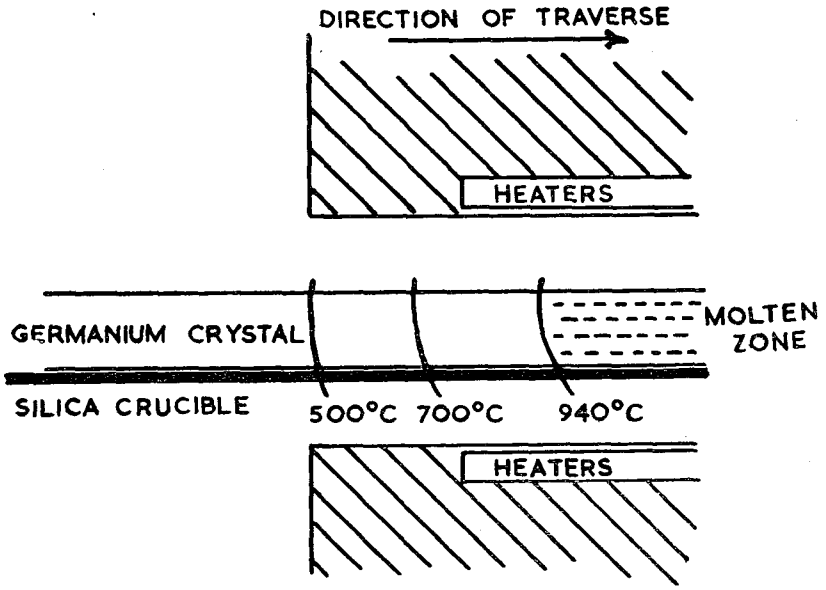


Figure 6. *Schematic diagram of horizontal zone melt, showing asymmetrical, non-parallel isothermals in a growing germanium crystal, with a steep axial temperature gradient.*

interface which is normal to the <111> growth direction, and isothermal surfaces in the grown crystal which are near parallel. To approach this condition in any crystal-growing furnace, it is necessary to suppress the radiation from the surfaces of the growing crystal and from the crucible, and thus reduce the radial temperature gradient to zero. In practice this may only be achieved if comparatively low axial temperature gradients are employed.

The radio-frequency induction after-heater technique of Bennett and Sawyer[60] reduces the heat loss by radiation from the crystal by imposing a boundary condition at the walls of the enveloping furnace tube, which is maintained at a constant temperature of about 800° C. In consequence of the small temperature difference between the crystal in its crucible and the furnace wall, the radial component of heat loss is largely suppressed and the asymmetry of radiation due to the presence of the crucible is reduced. By virtue of the one-dimensional heat flow down the crystal, therefore, a substantially planar interface normal to the direction of growth should be obtained. In addition to these desirable

features, the presence of fore- and after-heaters permits zone heating by direct radio-frequency induction without the usual intermediate susceptor, thus giving the advantages of eddy-current stirring of the melt and consequent homogenization of solute distribution. This method differs significantly from the earlier zone melt furnaces, and a compromise has been made in that the importance of a low temperature gradient at the growth interface has been recognized, whilst a steep temperature gradient is still maintained at the other interface to assist in stabilizing the zone length. Slow cooling of the whole crystal after completion of growth prevents the introduction of further strain and, as Bennett and Sawyer show, notable improvements in the uniformity of solute distribution are achieved together with a reduction in dislocation density, particularly at slow growth rates using their 12 in. after-heater. However, because the grown crystal enters the after-heater maintained at about 800° C and remains within it for several hours, some degradation of lifetime occurs and is presumed to arise from contamination of the crystal during this period, which can be demonstrated by the increase in minority carrier lifetime away from the seed end.

An alternative approach which avoids the degradation of minority carrier lifetime characteristic of the after-heater technique and yet retains the control of dislocation density is the use of controlled temperature gradient muffles in a resistance-heated furnace. These muffles appear to offer certain operational advantages. Furnaces of this type have been used by the authors over the past year for the preparation of material in the resistivity range 1 Ω cm to intrinsic, with minority carrier lifetimes of several msec. The method differs essentially from that of Bennett and Sawyer in that a much steeper but controlled axial temperature gradient is permitted to exist in the grown crystal; by virtue of the muffle, which has a similar axial temperature gradient, the radial temperature gradients in the crystal and crucible at elevated temperatures are reduced to approach zero, and near vertical planar solid-liquid interfaces are obtained. Owing to the high emissivity of the silica crucible, it was found desirable to maintain the bottom of the furnace slightly hotter than the top to minimize the effects of any small temperature mismatch between the crucible and the muffle. To preserve good zone length stability the temperature gradient at the melting solid–liquid interface is maintained as steep as possible, this also assisting in homogenization of the solute in the molten zone. To add further stability to the whole system of crystal growth, the silica crucible and ingot remain stationary throughout the process, the crystal-growing furnace traversing them, with a precision-controlled movement.

4.1. The Crystal Growing Apparatus

The principles outlined above have been embodied in the design of furnaces for growing crystals varying from 2 to 8 cm^2 in cross-section. A diagram of the furnace used for the 2 cm^2 cross-section crystals is given in Figure 8, and its construction will now be described. It consists basically of fourteen 6 in. longitudinal, helically wound elements, each consisting of a 3 ft. length of 0·025 in. diameter platinum–rhodium (10 per cent) wire. These are spaced around the circumference of a 2½ in. diameter circle, coaxial with the furnace tube, with eight distributed over the lower half and six over the upper half. The elements are surrounded by a fused alumina tube and the whole is enclosed in a

thermally insulated housing. Water-cooled, metal jackets are attached to each end of this lagged housing, coaxial with the furnace tube, into which are screwed stainless steel muffles, their penetrations into the furnace being adjustable. The

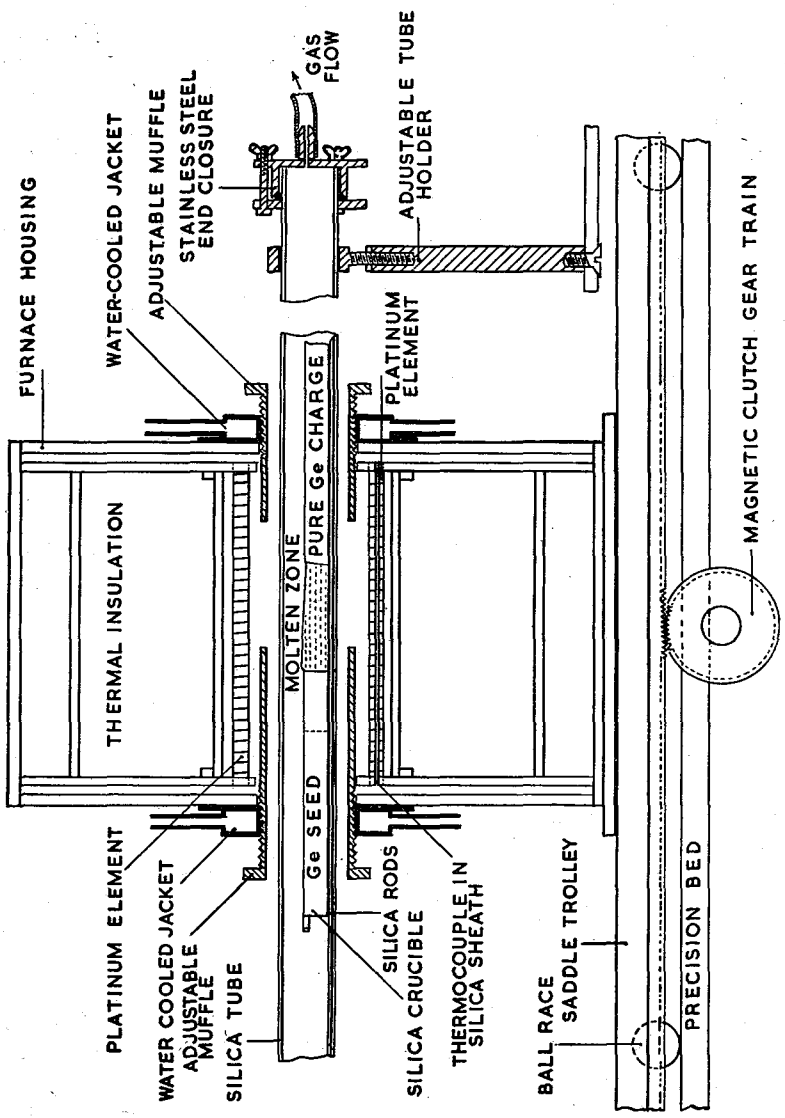


Figure 8. *Diagram of the horizontal zone melt crystal growing furnace.*

elements are connected in series through a control resistance and supplied with a ± 1 per cent servo-stabilized voltage, the total average power input being about 1 kW. Temperature control is achieved by means of a platinum/platinum–rhodium (13 per cent) thermocouple enclosed in a silica sheath and mounted at the bottom mid-point of the furnace, on the same pitch circle as the elements.

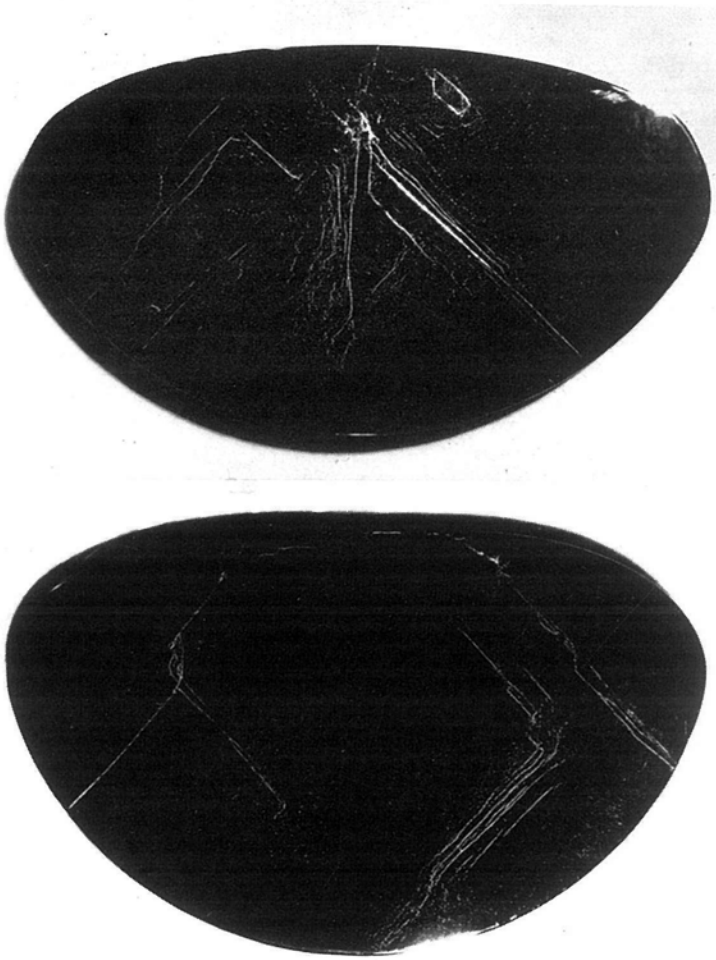


Figure 1. *Typical lineages on the etched* {100} *cross-sectional surfaces taken from the two ends of a zone-levelled germanium crystal, normal to the direction of growth.*

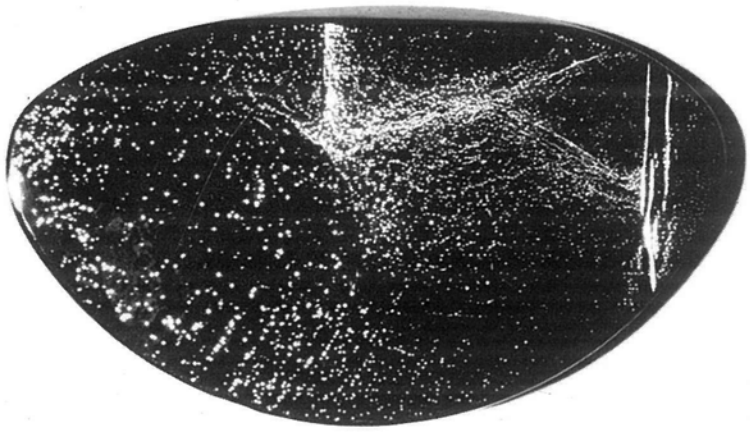


Figure 2. *Lineage on the etched* {111} *cross-sectional surface of a germanium crystal, normal to the direction of growth, obtained with a concave interface.*

[*Facing p. 152*

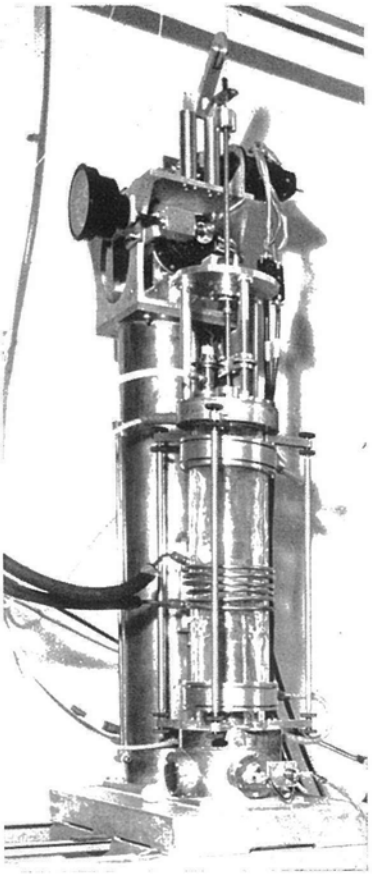


Figure 3. *A Czochralski apparatus for pulling single crystals of germanium, constructed by the Royal Radar Establishment, Ministry of Supply, Malvern, U.K.*

Figure 5. *Symmetrical slip, revealed by etch pits on the* {111} *cross-sectional surface of a germanium crystal obtained with a convex molten–solid interface.*

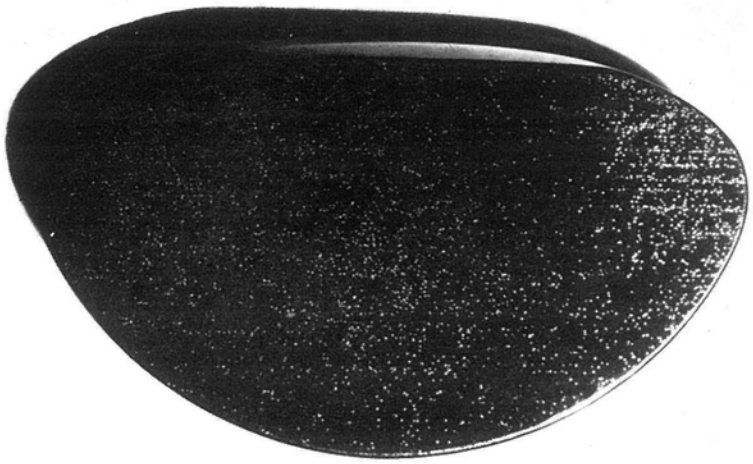


Figure 7. *Preferential slip, revealed by etch pits on the* {111} *cross-sectional surface of a germanium crystal, obtained with a molten-solid interface tilted with respect to the* {111} *growth plane*

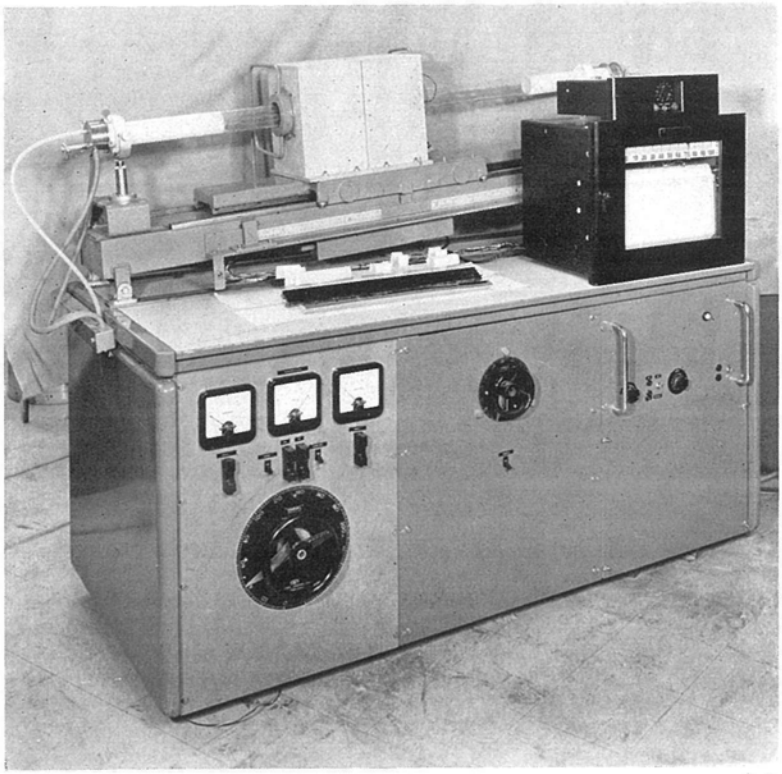


Figure 9. *The horizontal zone melt furnace for growing germanium single crystals 8* cm^2 *in cross-section.*

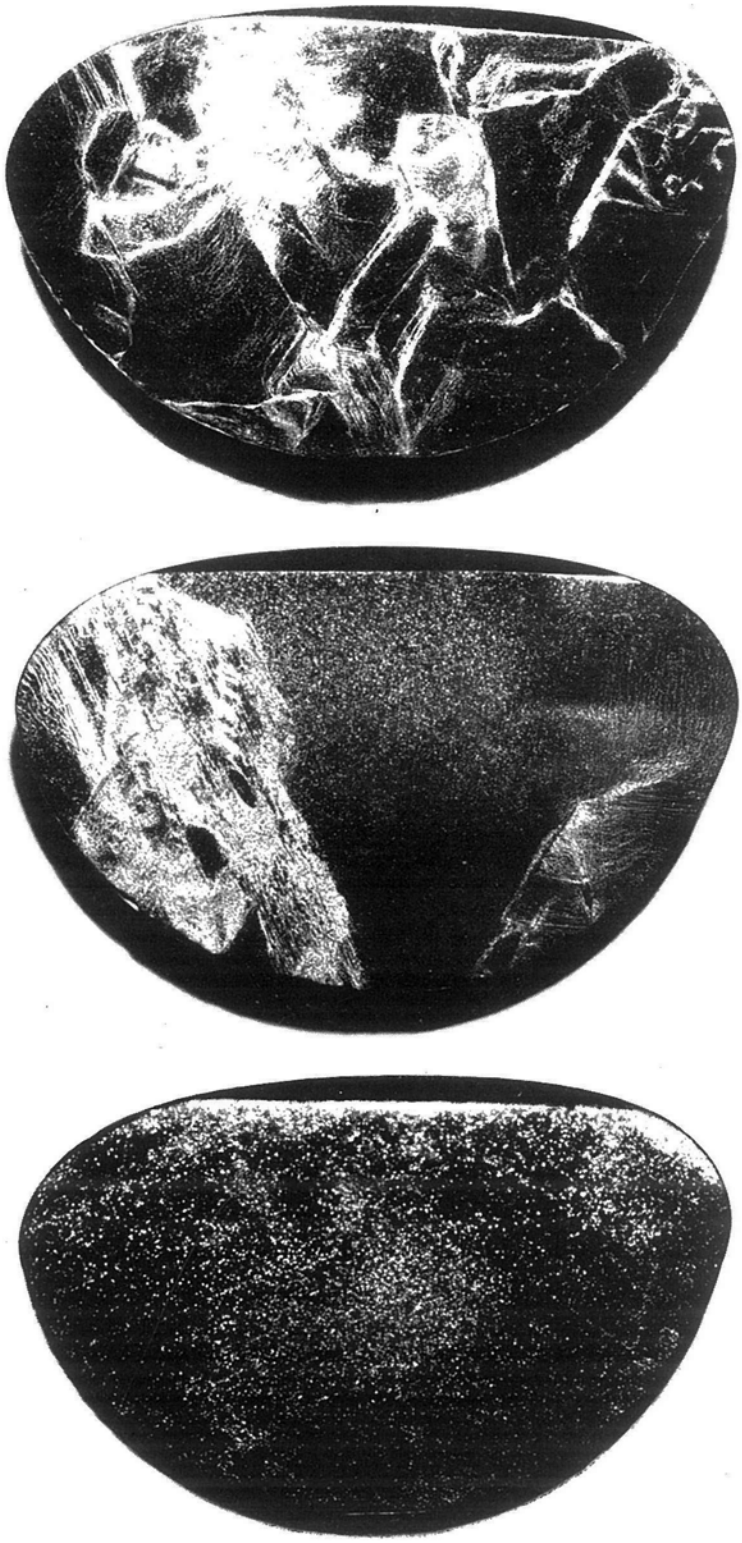


Figure 17. *Etched* {111} *cross-sectional surfaces taken at* 5*cm intervals in the direction of growth, illustrating the out-growth of deliberately introduced gross lineage, using a convex liquid–solid interface.*

Figure 18. *Fine (low angle) lineage on the etched* {111} *cross-sectional surface of a germanium crystal obtained with a slightly concave interface.*

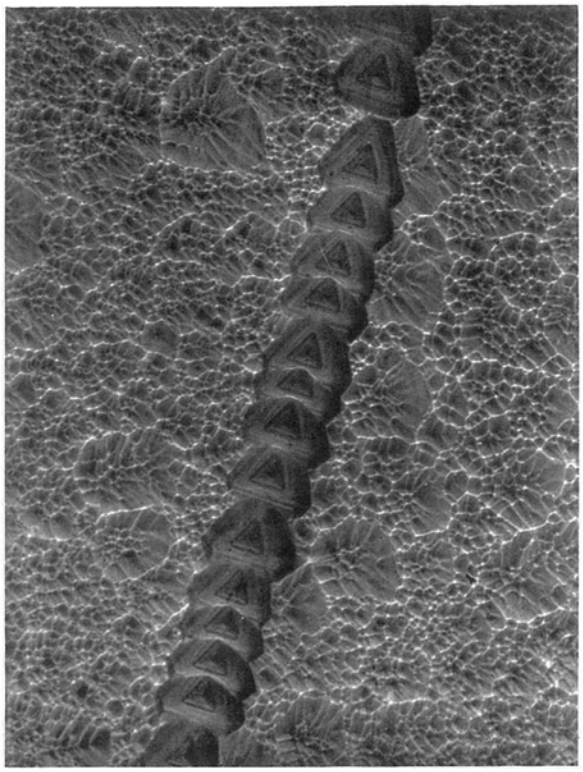
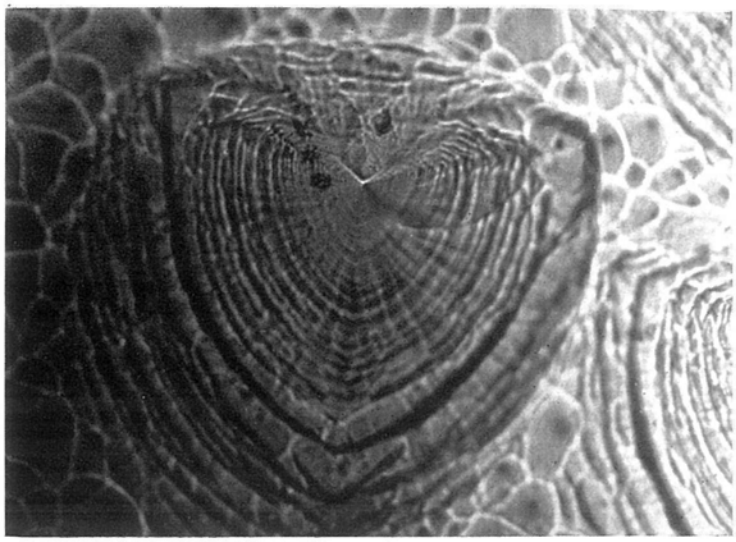


Figure 19. *Etch pits at a low-angle lineage, revealed by slow etch techniques on the* {111} *cross-sectional surface of a germanium crystal.* (*Magnification* × 100).

Figure 20. *Etched* {111} *cross-sectional surface of a germanium crystal showing increased density of etch pits at the side.*



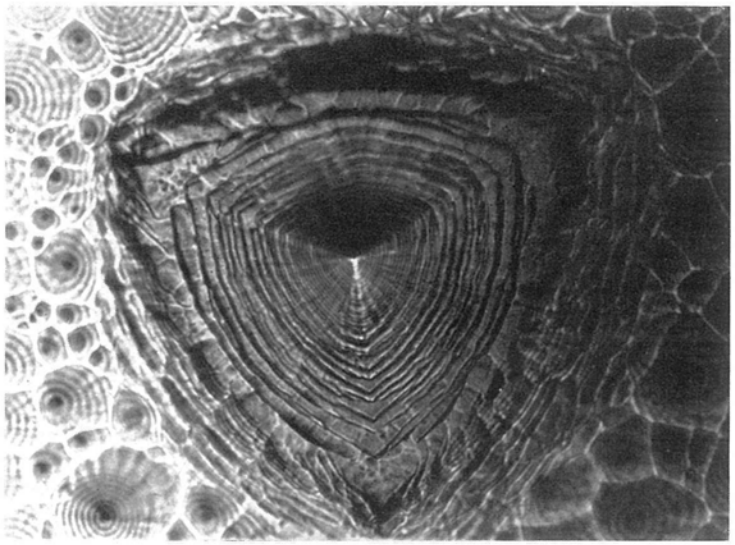


Figure 21. *The fine structure of etch pits on a* {111} *cross-sectional surface, revealed by slow etch techniques* (*Magnification* × 600).

Figure 22. *Typical etch pit distributions on the etched* {111} *cross-sectional surfaces of germanium crystals obtained as the angular separation of the liquid–solid interface and the* {111} *growth plane is reduced from about* 5° *to approach zero.*

Facing p. 153]

It is used in conjunction with a Honeywell Brown proportional control/recorder and the control resistance and gives $\pm 1°$ C long-term temperature stability. The furnace is mounted on a saddle-trolley running on ball-races, on a very rigid, precision-made bed, which is a standard item in the laboratory. Forward and reverse drives to the saddle-trolley are applied through magnetic clutches and gear trains from a servo-controlled velodyne unit mounted in the precision bed. The traverse may be preset to any speed in the range 1–26 cm hr^{-1} or programmed by means of a suitable cam and drive.

The furnace tube of transparent fused silica is 5 ft long and is mounted in a position coaxial with the furnace, by supports from the bed, provision being made for accurate alignment. The end closures through which a controlled gas flow passes, are of stainless steel and the tube is rendered air-tight by means of trapped neoprene compression seals. The silica crucibles are semi-circular in cross-section, 50 cm long and are carefully ground to be uniform; they are thoroughly cleaned, carbon-coated, and then prefired at 1200° C in a nitrogen–hydrogen atmosphere immediately prior to use. By means of jigs, the 10 cm <111> orientated germanium seed crystal, the germanium slice, containing a measured amount of impurity, and a zone purified ingot, of uniform cross-section, are then loaded into the crucible after rigorous cleansing treatments. The crucible, mounted on silica rods, is carefully adjusted in the furnace tube and remains in this position until completion of the crystal growing process. After thorough gas flushing of the system, the furnace is moved into position and the first molten zone in contact with the seed crystal is established. Thereafter, the furnace traverses the ingot at a preset rate using a controlled gas flow of 5 ft^3hr^{-1}, until it completely clears the grown crystal, which is then removed for examination, measurement of resistivity and minority carrier lifetime. A photograph of the horizontal zone-melt furnace, used for growing germanium crystals 8 cm^2 in cross-section, is given in Figure 9.

5. THE GERMANIUM CRYSTAL QUALITY AND SOME EXPERIMENTAL RESULTS

Extensive study of the crystals grown in the horizontal zone-melt process and of the experimental conditions for homogeneity of impurity distribution, high minority carrier lifetime and low-level dislocation density have led to an improved quality of germanium crystal. These three criteria of crystal quality are now discussed in terms of the crystal growth parameters and some of the experimental results presented.

5.1. Resistivity Variations

In Pfann's expression for the impurity concentration in a zone-levelled crystal, quoted in Section 3.2, it is clear that solutes having low segregation coefficient must be employed if a reasonably constant longitudinal resistivity is to be achieved. Hence, it is common practice to use antimony ($k = 0{\cdot}003$) and indium ($k = 0{\cdot}001$) in the preparation of N-type and P-type crystals respectively. In order to fully utilize the advantages offered by the zone-levelling technique and to approach the predicted impurity distribution, the objectives must lie in seeking the experimental conditions which lead to a general validity of those

assumptions made in Pfann's analysis. The conditions of constant zone length and thermal stirring of the molten zone have already been discussed in Section 4. Other causes of longitudinal resistivity variation may be attributed to variations in the cross-section and impurity content of the germanium charge, and to spurious chemical contamination from the crystal growing apparatus.

By the correct adjustment of furnace tilt to counter the effects of volume change upon solidification in the zone-purification and zone-levelling processes, the cross-section of a germanium ingot may be held constant to within $\pm$ 3 per cent over its entire useful length.

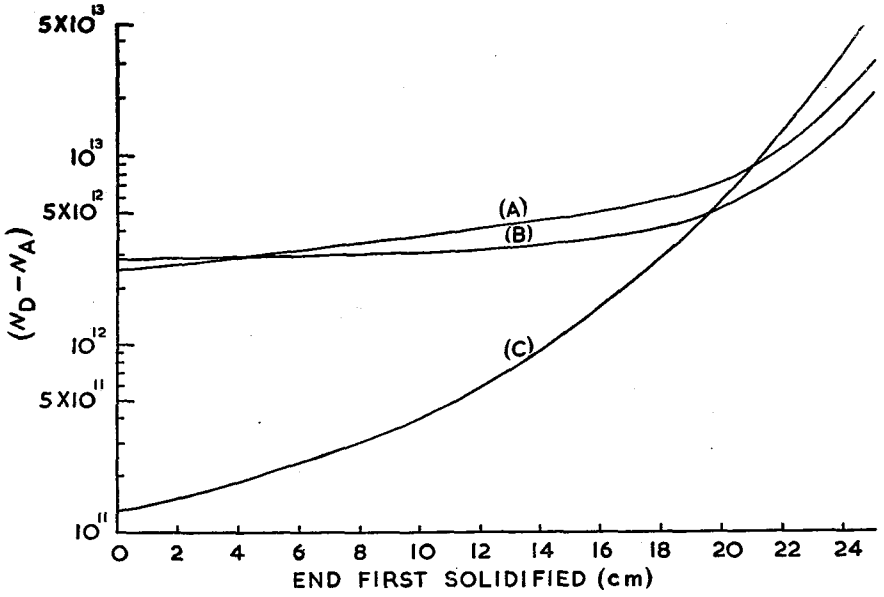


Figure 10. *Typical* $|N_D - N_A|$ *distributions in zone-purified single crystals of germanium. Curve A: after nine zone passes in* N_2 *ambient. Curve B: after nine zone passes in* N_2:H_2 (10 *per cent*) *ambient. Curve C: after fifteen zone passes in* N_2 *ambient.*

Where the preparation of high resistivity, long minority carrier lifetime material is envisaged, it is desirable to have a knowledge of the limiting residual concentration $|N_D - N_A|$ of impurities after zone purification. Also, for the purposes of assessment and more particularly to eliminate unpredictable behaviour resulting from impurity precipitation at grain boundaries, the germanium charge used should itself be a single crystal. The zone purification processes used in this laboratory yield purified ingots, 50 per cent of which are self-seeded single crystals of arbitrary orientation. To produce zone-levelled crystals of high resistivity, therefore, it is necessary only to select the germanium charge for processing, this selection leading to an enhanced reproducibility in the subsequently grown seeded single crystals.

Some loss of purity is inevitable by transference of material from the zone purification to the zone-levelling process as this involves cutting and cleaning operations on the seed crystal, the germanium charge and the germanium slice

containing the solute, and also the preparation of a suitable crucible and furnace tube.

It is also important, therefore, to have an estimate of the degree of contamination resulting from this practice, in order that the correct amount of impurity may be added to the molten zone to yield the resistivity required.

The specific resistivity as a function of temperature along the uncut lengths of self-seeded zone-purified crystals has been measured, followed by similar measurements on the same materials after one zone pass in the single-crystal process. These measurements yield a figure for the net $|N_D - N_A|$ associated with process contamination and, in the techniques used in this laboratory, it is of order 10^{13} cm^{-3}. Curves A and B in Figure 10 typify the distributions of the

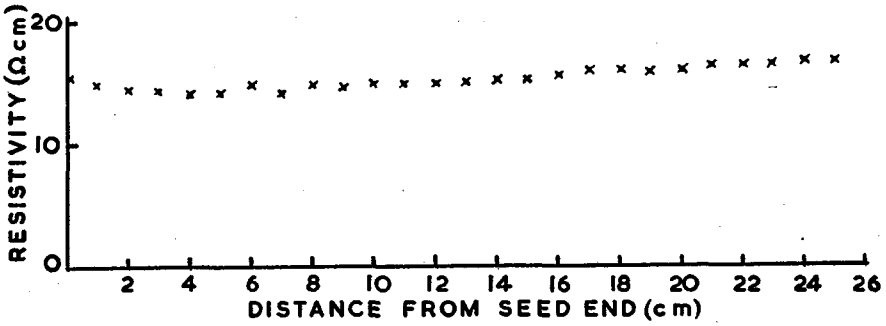


Figure 11. *Resistivity plot of a typical zone-levelled single crystal at* 20° *C.*

residual contaminant level $|N_D - N_A|$ of impurities after nine zone passes in a nitrogen and nitrogen–hydrogen (10 per cent) gas ambient respectively, whilst curve C shows the net density of donors $|N_D - N_A|$ for a triple zone-melted, intrinsic single crystal in a nitrogen ambient, for germanium which had been previously subjected to a multiple zone purification of fifteen zone passes.

Resistivity measurements by the calibrated four-probe technique similar to that described by Valdes[61] have been made at 1 cm intervals along the 26 cm lengths of a series of antimony-doped zone-levelled crystals grown at 6 cm hr^{-1}, in the resistivity range 1–40 Ω cm. The average deviation of $|N_D - N_A|$ from the mean level in each crystal was found to be $\pm$ 9·5 per cent. The first 85 per cent of zone-levelled crystal length possesses a longitudinal homogeneity within $\pm$ 5 per cent of the mean value, but the impurity concentration in the last 15 per cent shows a decrease in $|N_D - N_A|$ owing to the slight increase in zone length at the normal freeze end, resulting from the asymmetrical loading of the crystal furnace. A typical resistivity plot against distance from the seed is given in Figure 11 for an antimony doped crystal.

The cross-sectional variations in zone-levelled crystals have been found to be more serious in character, and a typical resistivity contour map for a crystal grown at 6 cm hr^{-1} is shown in Figure 12. The principal features of the distribution are a general gradient in impurity content from top to bottom of the cross-section, with regions of lower impurity content at the sides, and a localized region of higher impurity content at the bottom. These trends have been explored over

the range of growth rates 3–12 cm hr^{-1}. They show increased homogeneity at the slowest growth rate, whilst at 12 cm hr^{-1} the localized region of increased impurity content extends further over the cross-section. The results of such tests determine the choice of growth rate to be employed in the crystal growing process. Where high minority carrier lifetimes are the dominating requirement,

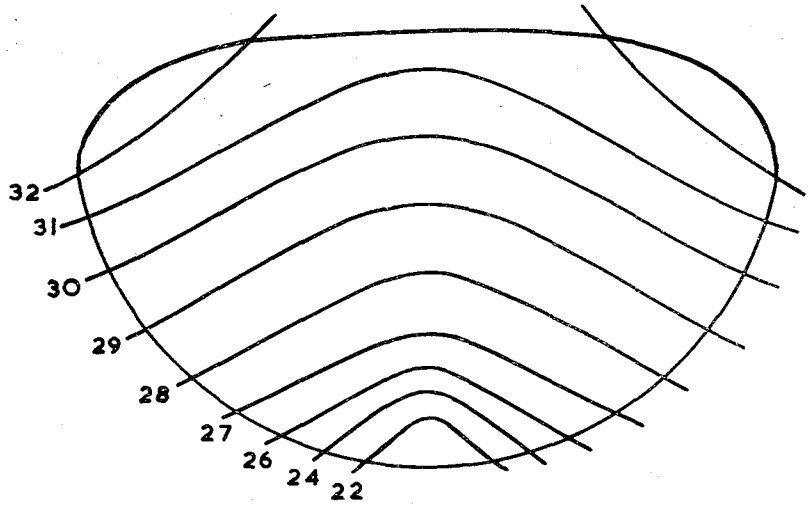


Figure 12. *Resistivity contour map of doped germanium single crystal cross-section.*

the faster pull rates should be employed, but if greater homogeneity in cross-sectional impurity distribution is required, the slower growth rates are necessary.

The general pattern shown in Figure 12 is held at all resistivities, and the inhomogeneity in impurity distribution is attributed to the inadequacies of thermal stirring in the melt.

5.2. The Lifetime of Minority Carriers

Whilst the minority carrier lifetimes of germanium recorded here are more than adequate for most device applications, their measurement is useful in the quality assessment of single crystals, since, as pointed out in Section 2, the values are critically dependent upon the crystalline perfection and the concentration of recombination centres, which may result from chemical contamination of the crystal. The minority carrier lifetime is most conveniently determined by the photoconductive decay method described by Hornbeck and Haynes[62]. For both production and research purposes it is desirable to determine the variation of this parameter along the length of the crystal and to make an absolute determination of its value for typical specimen slices from the crystal.

The measurement of minority carrier lifetime at 1 cm intervals along its length is performed before the crystal is cut. After de-greasing and suitable etching treatments to produce a low surface recombination velocity, graphitized pressure contacts are made at each end of the crystal and a small constant current is passed through it. For germanium crystals of 2 cm^2 cross-section with specific resistivities greater than about 5 Ω cm. the photoconductive decay may

be observed directly by using a fast light chopper and monitoring the potential across the specimen by means of a wide-band amplifier and cathode-ray display. For material of lower resistivity, where it is difficult to obtain an adequate photoconductive signal, a slow light chopper and narrow-band amplifier may be employed to measure the relative photoconductivity and hence, for zone-levelled crystals, to obtain a measure of the uniformity of the minority carrier lifetime at different points along the crystal.

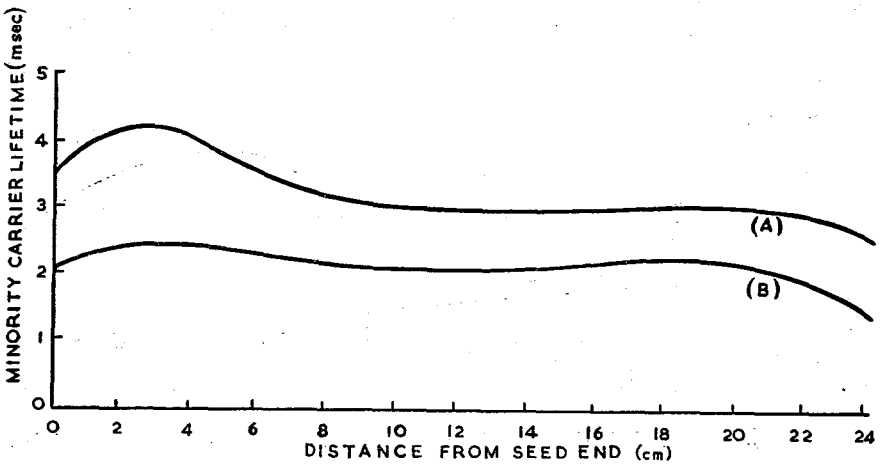


Figure 13. *Plots of minority carrier lifetime (uncorrected for surface recombination) at 20° C versus distance from seed crystal, for intrinsic material, grown at 12 cm hr^{-1}. Curve A: crystal grown in N_2 ambient from N_2:H_2 (10 per cent) ambient zone purification. Curve B: crystal grown in N_2 ambient from N_2 ambient zone purification.*

Measurements are made with the crystal enclosed in a quartz tube through which ambient gases may be circulated. For germanium with a low minority carrier lifetime it is usually sufficient to carry out measurements in dry oxygen, but for minority carrier lifetimes in excess of about 1 msec, where the effects of the surface become more critical, the crystal may be subjected to a standard Brattain and Bardeen[63] gas ambient cycling procedure, which will yield greater uniformity in the surface recombination velocity. If an electronic ozone generator is employed, permitting the measurements of lifetime during the ozone cycle, it is sufficient to pursue the cycling only to that point at which the ratio of lifetimes, measured in ozone and oxygen, becomes constant.

After checks for the homogeneity of minority carrier lifetime along the crystal, thick and thin longitudinal adjacent slices at representative regions in the crystal may be cut, lapped and etched, in order to make a precision determination of the bulk minority carrier lifetime, with its attendant surface recombination correction. For lifetimes in excess of about 1 msec, considerable care is necessary in evaluating the surface recombination. It has been found desirable to prepare and cycle, in the gas ambient, the two samples simultaneously in the same chamber and to observe the photoconductive decay during the cycling. For a thickness ratio of 6:1 between the two samples, consistent results may be obtained after about six gas ambient cycles, in common with the findings of Stevenson and Keyes[64].

Typical plots of minority carrier lifetime against linear distance from the seed are given in Figure 13 for single crystals of intrinsic resistivity and dislocation counts of the order of 10^3 cm^2, uncorrected for surface recombination. It may be observed that the minority carrier lifetime is substantially uniform with distance along the crystal, with a slight degradation in the region adjacent to the seed crystal; this is believed to be associated with chemical contamination of the seed which may occur during the extended period of heating necessary for the initial establishment of the molten zone prior to crystal growth.

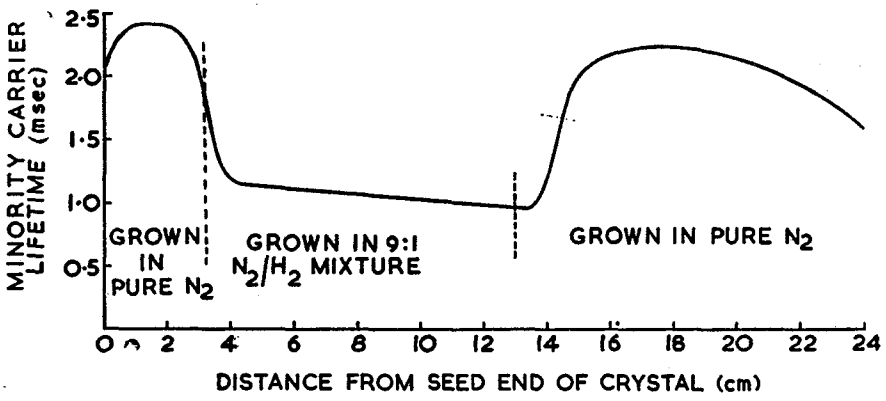


Figure 14. *Variation of minority carrier lifetime in a germanium crystal (uncorrected for surface recombination), in which the gas ambient was cycled during growth.*

The curves of Figure 13 illustrate the general features of the minority carrier lifetime in a grown crystal, and its near uniformity under steady growth conditions permits the study of the influence on minority carrier lifetime of gas ambients during crystal growth and of growth rates, provided the germanium charge is of intrinsic resistivity. Figure 14 shows the effects on minority carrier lifetime of ambient gas cycling during crystal growth, from a nitrogen ambient to a nitrogen–hydrogen (10 per cent) ambient and back to nitrogen, as measured on a crystal grown at 12 cm hr^{-1}. The data indicate that whilst both types of ambient gas will produce material of high minority carrier lifetime, there is an approximately 2:1 ratio favouring the use of a nitrogen ambient.

The reduction of minority carrier lifetime when hydrogen is introduced into the crystal growing process suggests either that a chemical contaminant is associated with the hydrogen or that hydrogen in the lattice, possibly interacting with other lattice imperfections, produces recombination centres which have so far not been identified.

In Table 1, comparative results of the minority carrier lifetime, uncorrected for surface recombination, are given for crystals in which the growth rates were varied. It may be seen that for both ambient gases, an improvement in minority carrier lifetime results from an increase in growth rate, in common with the

Table 1. Variations of minority carrier lifetime with growth rate. In both cases the zone purification gas ambient is N_2:H_2. In case A the single crystal gas ambient is N_2:H_2 and in case B it is N_2.

Growth rate (cm hr^{-1})	Minority carrier lifetime (msec)	
	A	B
6	0·71	1·40
12	0·99	1·44
16	—	1·48

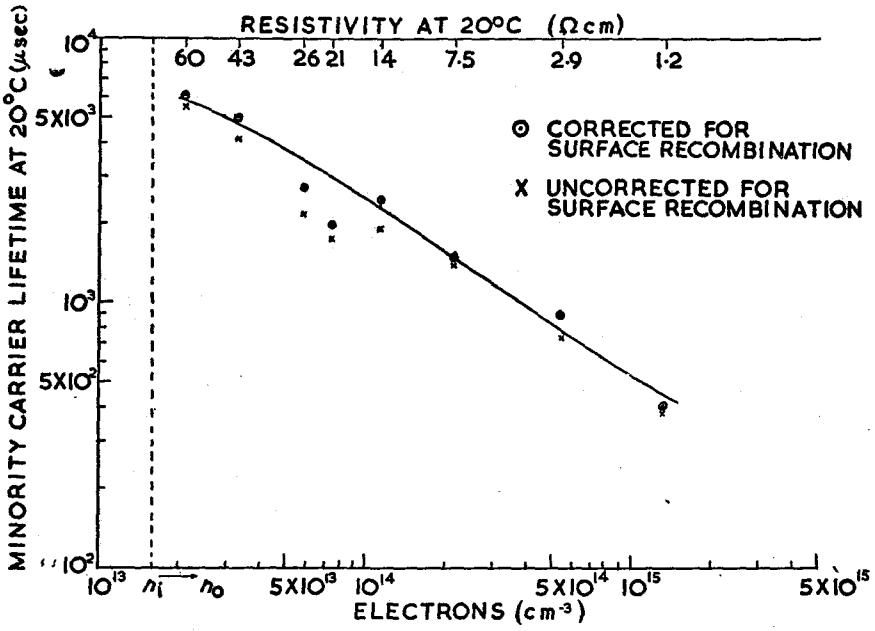


Figure 15. *Minority carrier lifetime as a function of resistivity of antimony doped germanium single crystals (experimental).*

findings of Bennett and Sawyer[60], but, as may be expected, this is achieved at the expense of greater inhomogeneity in the distribution of impurity content in doped crystals.

Precise determinations of the minority carrier lifetime for a series of zone-levelled N-type germanium crystals of differing antimony concentration grown under similar conditions, using a nitrogen ambient, are presented graphically in Figure 15. All these crystals are free of lineage, have a dislocation count of the order of 10^3 cm^2, and were grown at 6 cm hr^{-1}, this being the preferred compromise between minority carrier lifetime and homogeneity of impurity distribution. All crystals were grown from material which was zone purified by six zone passes in a nitrogen–hydrogen (10 per cent) ambient, followed by three zone passes in nitrogen. The antimony used in the zone-levelling was also zone-purified.

5.3. Crystalline Perfection

Experience has shown that a greater degree of crystalline perfection is attainable in germanium if crystal growth takes place on the {111} crystal plane, and in the many <111> orientated crystals produced under properly controlled conditions in this laboratory polycrystallinity and twinning are virtually nonexistent. From considerations of symmetry and for convenience in the analysis of crystalline defects from etch pit distributions, it is desirable to use a standard <111> orientation such that the vertical, longitudinal section of the crystal is a {110} plane. Under these conditions the lines of etch pits associated with lineage and slip when present, will be in the directions indicated in Figures 16*a* and *b* respectively.

The importance of lineage free crystals and of controlled dislocation densities has already been emphasized in Section 2.2. Studies of these imperfections and, in particular, their elimination have been made, and a correlation has been found between their occurrence in the grown crystal and the shape of the molten–solid interface during growth. The general character of the interface may be checked by the techniques of sequential doping to produce a P–N junction under the dynamic conditions of crystal growth. The interface can then be delineated by a selective etching technique[65] of the subsequently sectioned crystal.

Lineage, the term used to define a low-angle boundary in a single crystal, may exist in a variety of forms, the most commonly encountered being:

(1) Gross lineage, bordering on polycrystallinity.

(2) Fine lineage, which may constitute a single tilt boundary or a series of near parallel tilt boundaries, in which the angle of tilt is of the order of seconds of arc.

Gross lineage is evident in some self-seeded crystals of arbitrary orientation obtained from the zone purification processes, and it may also occur during single-crystal growth in the zone-levelling process in the event of faulty seeding or of silica–germanium adhesion occasioned by a poor carbon-coating technique. Under these latter conditions, gross strains are introduced into the crystal upon cooling owing to the differential thermal contraction, and the dense, complex arrays of dislocations introduced as a result may be revealed by the etch pit technique as shown in Figure 17*a*.

An illustration of fine lineage on the etched {111} cross-sectional face of a crystal grown by the zone melt technique is given in Figure 18. The characteristic etch pattern of a lineage consists of an array of regularly spaced etch pits, the actual etch pit spacing increasing as the tilt angle decreases, until, for angles of a few seconds of arc, the pits tend to lose their regular character and may be masked if any etch pits due to slip are present. These etch pits lie in the line of intersection of the {110} planes and the {111} growth plane and after subjecting a lineaged crystal to the slow etch techniques of Ellis[35], the characteristic triangular nature of the etch pits is revealed along the low-angle boundary as shown in Figure 19.

It is found that a positive control over the incidence of all lineage in germanium crystals grown in the {111} direction is contingent upon the maintenance throughout growth of a molten–solid interface which remains slightly convex. The convexity of the interface is amenable to control by the modification of the radial

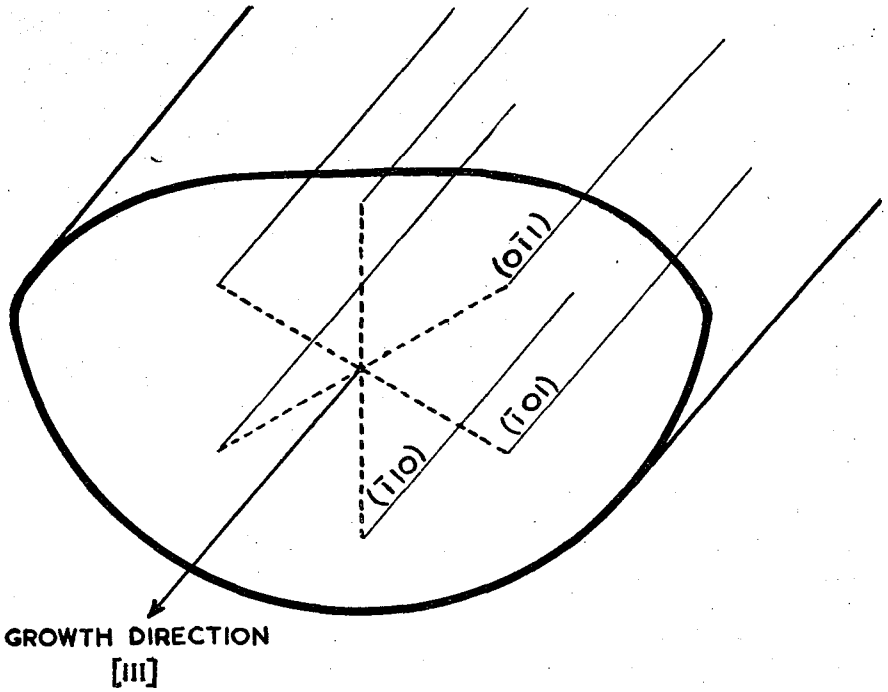


Figure 16*a*. *Standard orientation adopted. Dashed lines show intersections of* {110} *planes, normal to* (111) *growth plane, where lineage is most commonly observed.*

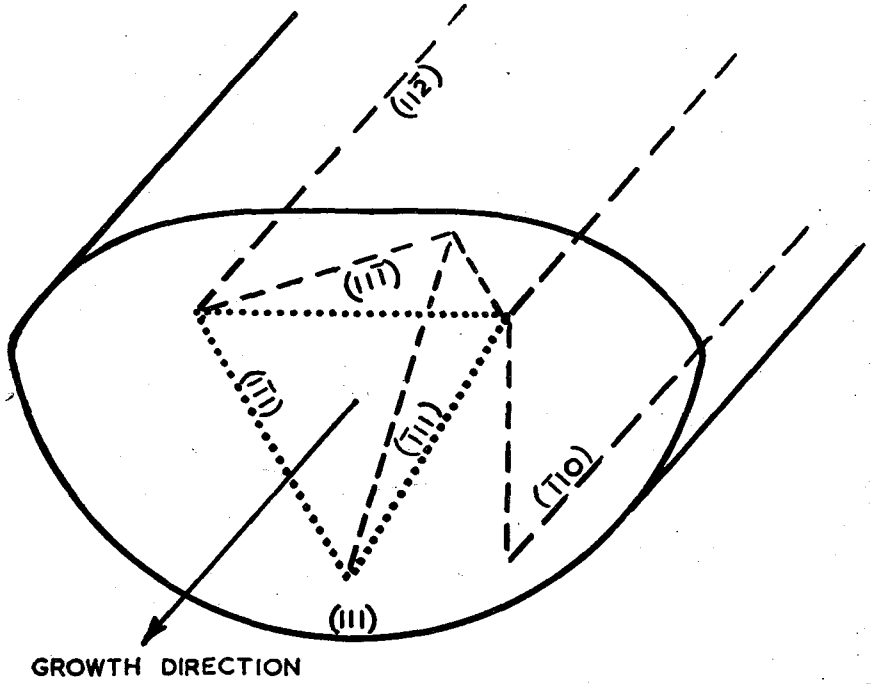


Figure 16*b*. *Standard orientation adopted. Dotted lines show directions on* (111) *growth plane of etch pits associated with traces of slip planes.*

temperature gradient in the crystal by adjustment either of the muffle penetration or furnace temperature. As a check on the sufficiency of this condition, the series of etched {111} cross-sectional surfaces in Figure 17*a*, *b* and *c*, taken at 5 cm intervals in the direction of growth, show the progressive reduction of lineage in a single crystal, where gross lineage was deliberately introduced from the seed.

However, the maintenance of a convex interface alone, throughout crystal growth, is not a sufficient condition to reduce dislocation densities to a low level, as evidenced by Figure 17*c*.

The experimental evidence resulting from the analyses of many cross-sectional {111} slices taken from crystals grown under slightly differing conditions strongly suggests that slip results from a non-planar interface and from non-parallel isothermal surfaces near the interface of the growing crystal, both of which may cause non-uniform thermal contraction. From studies of the etch pit distributions observed, the corrective action required to reduce dislocation densities may be deduced.

Case (1). If the molten–solid interface is convex or concave with radii of curvature smaller than some threshold value, slip may occur in that part of the crystal above the plastic temperature in a <110> direction to effect stress relief, and typical etch pit distributions are as shown in Figures 5 and 17*c*. By appropriate adjustment of furnace temperature or muffle penetration to cause the radial temperature gradient to approach zero, dislocation densities may be reduced as shown in Figure 20, where the radii of curvature have been raised above the threshold value except at the sides of the cross-section.

Case (2). Near planar interfaces are not necessarily parallel to the {111}, growth plane and crystals have been studied where the interface makes an angle of about $\pm$ 5° with respect to the growth plane, by adjusting the ratio of top-to-bottom heat in the furnace. Under these conditions a small amount of slip may still be observed, probably owing to the presence in the grown crystal of non-parallel isothermal surfaces, but the general pattern is now one of a random dislocation density which has been found to bear some correlation with the angular separation between the interface and the {111} growth plane. Detailed investigations on material of low-level dislocation densities (of the order of 10^2 cm^{-2}), at approximately 0·1 cm intervals in the direction of growth, reveal that in such random patterns, some dislocations lie in <110> directions and are identified with the Shockley-Read model of a dislocation resulting from slip, whilst others remain axially stable in a <111> direction. Microscopic investigations of these types of etch pits under slow etch conditions, examples of which are given in Figure 21, lend confirmation to these observations.

Case (3). By adjustment of the ratio of top-to-bottom heating in the furnace or by reducing the axial temperature gradient in the crystal, the angular separation between the interface and the growth plane may be reduced to approach zero.

Typical etch pit distributions for cross-sectional slices taken from crystals in which the angular separation is reduced from about 5° towards zero are shown in Figure 22*a*, *b* and *c*. The interface associated with the Figure 22*c* distribution appears to be macroscopically coincident with the {111} growth plane, and the etch pit count is less than 10 cm^{-2}.

Under those conditions, however, where dislocation counts are less than 10^2 cm^{-2}, furnace temperature and muffle penetration adjustments are critical and are only practicable when low axial temperature gradients of order 30–40 deg. C cm^{-1} are used. It has also been found that zone-levelling under these conditions increases the cross-sectional inhomogeneity of impurity distribution owing to the lack of thermal stirring, and may lead to growth instability.

As a compromise, to yield crystals of the homogeneity reported the authors have found it desirable to control dislocation levels at approximately 10^3 cm^{-2}. This permits a tolerable degree of interface convexity to give protection against the incidence of lineage and allows the use of an axial temperature gradient of the order of 40 deg. C cm^{-1} at the interface to increase thermal stirring. A typical etched cross-section of all doped crystals is that shown in Figure 22*a*; no correlation between impurity concentration and dislocation density for antimony or indium solutes being observed in the resistivity range 1 Ω cm to intrinsic.

6. CONCLUSIONS

It is evident from the recent literature that crystal growth of semiconducting materials is still a dynamic subject and that much remains to be explored. The more widespread preparation of single-crystal germanium of low dislocation density is giving new impetus to studies of heat treatment effects, diffusion of impurities, recombination mechanisms, and surface and mechanical properties, and it is hoped, will lead to further clarification of semiconductor behaviour.

ACKNOWLEDGEMENTS

The authors are indebted to Dr. E. Eastwood, Chief of Research at the Baddow Research Laboratories, Marconi's Wireless Telegraph Co. Ltd., for permission to publish, and to their colleagues; in particular, to Mr. R. Adlington for much painstaking photography.

REFERENCES

1. G. K. Teal and J. B. Little. *Phys. Rev.* **78,** 647 (1950).
2. G. L. Pearson and W. H. Brattain. *Proc. Inst. Radio Engrs., N.Y.* **43,** 1794 (1955).
3. F. Seitz. *Imperfections in Nearly Perfect Crystals.* (Wiley, New York, 1952).
4. J. A. Burton. *Physica* **XX,** 845 (1954).
5. Harvey Brooks. *Electronics and Electron Physics.* Vol. VII. (Academic Press, New York, 1955).
6. C. S. Fuller, and J. D. Struthers. *Phys. Rev.* **87,** 526 (1952).
7. C. D. Thurmond and J. D. Struthers. *J. Phys. Chem.* **57,** 831 (1953).
8. F. Van der Maesen and J. A. Brenkman. *Physica,* **XX,** 1005 (1954).
9. C. S. Fuller, J. D. Struthers, J. A. Ditzenberger and K. N. Wolfstirn. *Phys. Rev.* **93,** 1182 (1954).
10. R. A. Logan and M. Schwartz. *Phys. Rev.* **96,** 46 (1954).
11. R. A. Logan and M. Schwartz. *J. Appl. Phys.* **26,** 1287 (1955).
12. J. A. Burton, G. W. Hull, F. J. Morin and J. C. Severiens. *J. Phys. Chem.* **57,** 853 (1953).
13. W. Shockley and W. T. Read. *Phys. Rev.* **87,** 835 (1952).
14. R. Newman and W. W. Tyler. *Phys. Rev.* **96,** 882 and 874 (1954).
15. W. W. Tyler, R. Newman and H. H. Woodbury. *Phys. Rev.* **97,** 669 (1955).
16. H. H. Woodbury and W. W. Tyler. *Phys. Rev.* **100,** 659 (1955).
17. W. C. Dunlap, Jr. *Phys. Rev.* **96,** 40 (1954).
18. F. X. Hassion, D. C. Thurmond and F. A. Trumbore. *J. Phys. Chem.* **59,** 1076 (1955).

19. W. Kaiser, P. H. Keck and C. F. Lange. *Phys. Rev.* **101,** 1264 (1956).
20. W. T. Read, Jnr. *Dislocations in Crystals.* (McGraw-Hill, New York, 1953).
21. A. H. Cottrell. Contribution to *Progress in Metal Physics.* Vol. 4, p. 205. (Pergamon Press, London, 1953).
22. C. J. Gallagher. *Phys. Rev.* **88,** 721 (1952).
23. E. Teghtsoonian and B. Chalmers. *Can. J. Phys.* **29,** 370 (1951).
24. F. L. Vogel, W. G. Pfann, H. E. Corey and E. E. Thomas. *Phys. Rev.* **90,** 489 (1953); F. L. Vogel. *Acta Metal.* **3,** 245 (1955).
25. F. L. Vogel, W. T. Read and L. C. Lovell. *Phys. Rev.* **94,** 1791 (1954).
26. J. P. McKelvey and R. L. Longini. *Phys. Rev.* **99,** 1227 (1955).
27. E. S. Greiner. *J. Metals.* **7,** 203 (1955).
28. R. G. Treuting. *J. Metals.* **7,** 1027 (1955).
29. C. J. Gallagher. *Phys. Rev.* **92,** 846 (1953).
30. G. L. Pearson, W. T. Read, Jr. and F. J. Morin. *Phys. Rev.* **93,** 666 (1954).
31. W. T. Read, Jnr. *Phil. Mag.* **45,** 775 (1954); **45,** 1119 (1954); **46,** 111 (1955).
32. W. Shockley. *Phys. Rev.* **91,** 228 (1953).
33. A. D. Kurtz, S. A. Kulin and B. L. Averbach. *Phys. Rev.* **101,** 1285 (1956).
34. F. D. Rosi. *A.I.M.E. Symposium on Semiconductors* (Feb. 1956).
35. S. G. Ellis. *J. Appl. Phys.* **26,** 1140 (1955).
36. R. A. Logan. *Phys. Rev.* **100,** 615 (1955).
37. S. Mayburg. *Phys. Rev.* **95,** 38 (1954).
38. R. L. Hopkins and E. N. Clarke. *Phys. Rev.* **100,** 1786 (1955).
39. R. A. Logan. *Phys. Rev.* **101,** 1455 (1956).
40. W. L. Brown and R. C. Fletcher. *Phys. Rev.* **92,** 585 and 591 (1953).
41. H. Letaw, Jnr., W. Portnoy and L. Slifkin. *Phys. Rev.* **102,** 636 (1956).
42. B. Chalmers. *Impurities and Imperfections.* A seminar held during the 36th National Metal Congress, Chicago 1955. (American Society for Metals, 1955).
43. U. M. Martius. Contribution to *Progress in Metal Physics.* Vol. 5, 279 (Pergamon Press, London, 1954).
44. F. C. Frank. 'Advances in Physics'. *Phil. Mag. Supplement* 1. p. 91, (1952).
45. W. G. Pfann and L. C. Lovell. *Acta Metal.* **3,** 512 (1955).
46. J. A. Burton, R. C. Prim, W. P. Slichter, E. D. Kolb and J. D. Struthers. *J. Chem. Phys.* **21,** 1987 (1953).
47. P. R. Camp. *J. Appl. Phys.* **25,** 459 (1954).
48. J. Czochralski. *Z. Phys. Chem.* **92,** 219 (1917).
49. G. K. Teal and J. B. Little. *Phys. Rev.* **78,** 647 (1950).
50. G. K. Teal, M. Sparks and E. Buehler. *Proc. Inst. Radio Engrs., N.Y.* **40,** 906 (1952).
51. K. Lehovec, J. Soled, R. Koch, A. MacDonald and C. Stearns. *Rev. Sci. Instrum.* **24,** 652 (1953).
52. L. Roth and W. E. Taylor. *Proc. Inst. Radio Engrs., N.Y.* **40,** 1338 (1952).
53. G. K. Teal, M. Sparks and E. Buehler. *Phys. Rev.* **81,** 637 (1951).
54. R. N. Hall. *Phys. Rev.* **88,** 139 (1952).
55. H. E. Bridgers. *J. Appl. Phys.* **26,** 1188 (1955); **27,** 746 (1956).
56. W. G. Pfann. *Trans. Amer. Inst. Min. Metal. Eng.* **194,** 747 (1952).
57. H. Nelson. *Transistors* I. p. 66 (R.C.A. Laboratories, Princeton, N.J., 1956).
58. E. Billig. *Proc. Roy. Soc.* **235,** 37 (1956).
59. W. G. Pfann and K. M. Olsen. *Phys. Rev.* **89,** 322 (1953).
60. D. C. Bennett and B. Sawyer. *Bell. Syst. Tech. J.* **32,** 637 (1956).
61. L. B. Valdes. *Proc. Inst. Radio Engrs., N.Y.* **42,** 420 (1954).
62. J. R. Haynes and J. A. Hornbeck. *Phys. Rev.* **90,** 152 (1953).
63. W. H. Brattain and J. Bardeen. *Bell. Syst. Tech. J.* **32,** 1 (1953).
64. D. T. Stevenson and R. J. Keyes. *Physica,* **XX,** 1041 (1954).
65. *Brit. Pat.* 726913 (Filed 26 Nov. 1952).

IMPURITIES IN GERMANIUM

W. CRAWFORD DUNLAP Jr., Ph.D.

Bendix Aviation Corp. Research Laboratories, Detroit, Michigan, U.S.A.,
formerly of General Electric Corp., Syracuse, N.Y., U.S.A.

MS. received 8 *November*, 1956

IMPURITIES IN GERMANIUM

1. INTRODUCTION

Substitutional impurities are those impurities which replace lattice atoms in a crystal. They should be distinguished from interstitials, which occupy sites not normally occupied by lattice atoms. A given sample may contain impurities in both types of site. A list of typical examples of the two types in germanium is given in Table 1.

Table 1. Nature of Typical Impurities in Germanium

Substitutional impurities		Interstitial impurities	
Donors	Acceptors	Donors	Acceptors
Nitrogen	Boron	Lithium	Copper
Phosphorus	Aluminum		Iron
Arsenic	Gallium		Nickel
Antimony	Indium		Cobalt
Bismuth	Thallium		Manganese
	Zinc		Silver
	Gold		

The best-known impurities in germanium are the substitutional impurities of the third and fifth groups of the Periodic Table; *viz.* the donors phosphorus, arsenic, antimony and bismuth; and the acceptors boron, aluminium, gallium, indium and thallium.

The donor and acceptor action of these elements is easily understood. Since the impurity element occupies a substitutional site it tends to form four covalent bonds with the surrounding germanium neighbors. However, in the case of a donor element, there are five electrons available; in the case of an acceptor only three. Thus arsenic has an extra electron that can, with very little effort, move into the lattice structure (conduction band) as a free carrier. Likewise, aluminum can easily take up a lattice electron to form a missing mobile electron bond (positive hole).

In this paper we shall go into some detail regarding the effects of the various impurity elements in germanium, and those properties of germanium characterized by the doping element. A paper summarizing much of the available information up to June, 1954 has been presented by Burton[1]. The present paper goes into somewhat greater detail than that of Burton in certain respects, and also presents some more recent results. Other papers surveying parts of the impurity problems in germanium are those by Burstein and Egli[2], and Burstein, Picus and Sclar[3].

We shall discuss particularly other elements, besides those of the third and fifth groups, which have been shown to have electrical activity in germanium. We shall also discuss some elements that appear to be neutral.

Properties of importance for the present discussion are summarized in the following paragraphs.

1.1. Segregation Coefficient

This parameter specifies the tendency of an element to enter the lattice during the growth of a single crystal. We shall not discuss the techniques of crystal growing here, since this subject has been dealt with thoroughly elsewhere.

The segregation coefficient k may be defined as:

$$k = C_S/C_L$$

where C_S is the equilibrium concentration of the impurity element in the solid crystal and C_L is the concentration of the element in the liquid, both per unit mass. Since the definition above refers to equilibrium conditions, the values of k quoted apply only for small growth rates. 'Rate-growing' is discussed below.

A small segregation coefficient is an indication that an element prefers the liquid state; the growing crystal face acts as a sieve tending to force the impurities into the sprout end of the crystal. This effect is a powerful aid in the removal of those elements with a small k. If the segregation coefficient is greater than 1, as in the case of the boron, the impurity concentrates during the first part of the freezing. Elements with melting points higher than germanium, and not forming *eutectics* with germanium, tend to have $k > 1$.

1.2. Solubility

Solubility of an impurity is the maximum concentration of the element in the germanium–impurity solid solution when the system is in equilibrium at constant temperature. The relation between solubility and segregation coefficient is discussed below. Solubility is usually dependent upon temperature, and often possesses a strong maximum at 940° C, somewhat below the melting point. The solubility of various impurities ranges from about 10^{14} cm^{-3} for iron and nickel to over 10^{21} cm^{-3} for boron.

A useful illustration in discussing solubility as a function of temperature is the *phase diagram*, such as that shown in Figure 1. In this case the impurity is copper, but most of the important impurity elements show the same type of behavior. Included are both the *solidus* which represents the concentration of the impurity in the solid impurity–germanium solution at equilibrium at the temperature indicated, and the *liquidus* which gives the concentration of the impurity in the liquid solution. At points of inflection or discontinuity, compounds or eutectics are indicated. The eutectic composition has a lower melting point than nearby compositions, and liquid of eutectic composition freezes with no change in concentration.

1.3. Diffusion Coefficient

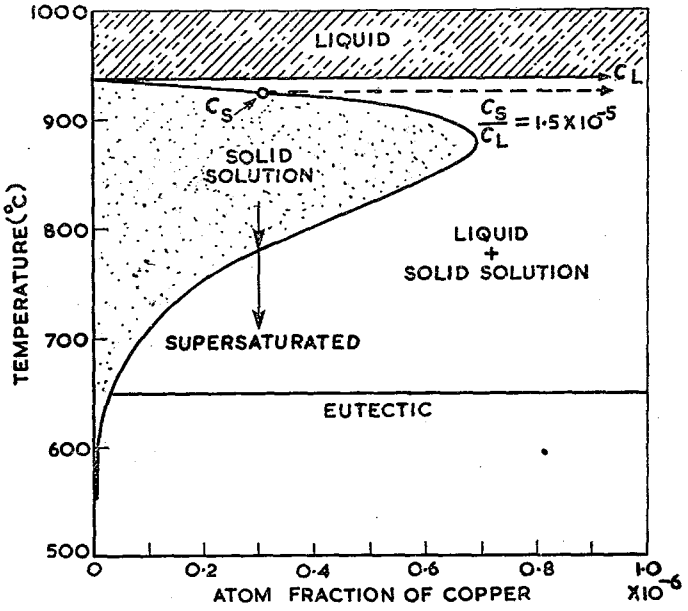
When a concentration gradient of the impurity exists, there is a mass flow in the direction of decreasing composition according to the relation:

$$D \cdot \frac{\partial^2 C}{\partial x^2} = \frac{\partial C}{\partial t} \qquad \ldots (1)$$

The diffusion coefficient, which is presumed independent of concentration at the levels usual in work with germanium, represents the ability of impurity

atoms to move about the lattice under the influence of thermal agitation. Although the theory of diffusion is not completely settled, present evidence indicates that for substitutional atoms at least, diffusion proceeds in two stages:
(1) The formation of a vacancy near the impurity atom or ion.
(2) The movement of the impurity into the vacancy.

In the present paper, factors affecting the number of vacancies will not be discussed. Evidence to be reviewed below indicates that both substitutional and interstitial atoms are important in germanium, diffusion measurements serving to distinguish between the two.



[By courtesy of Physica.

Figure 1. *Typical phase diagram for impurity–germanium system (Burton[1]).*

Interstitials probably diffuse through the open tunnels in the germanium lattice, and have to be present as small positive ions in order to make use of these tunnels.

Although many important properties depend upon atomic or ionic radii, reliable values of ionic radii applicable to germanium solutions are not available. Substitutional ions in germanium have a tetrahedral bonding structure, and for many elements it is difficult to evaluate the radii for this configuration. Also, it is probable that in many cases the element may change its ionic condition with temperature. This is probably true for many elements which act as acceptors but diffuse rapidly, *e.g.* copper and nickel.

Related to diffusion coefficient is ionic mobility, which is given by the Einstein relation $\mu = Dq/\mathbf{k}T$. The ionic mobility can be measured by direct methods[4] which give means of evaluating both the sign and the magnitude

of the charge on the diffusing entity. This is not possible with diffusion measurements.

Both diffusion and mobility are related to activation energy W according to the equation:

$$D = D_0 \exp(-W/\mathbf{k}T) \quad \ldots (2)$$

where D_0 is a constant of the material [5]. For substitutional impurities W tends to have a larger value than for interstitials.

1.4. Electrical Activity: Ionization Energy

The electrical effectiveness of an impurity element is governed by the density of impurity atoms in electrically active sites. There is little information available about this property, and it is generally assumed that each dissolved impurity atom is entirely effective in producing carriers. Preliminary measurements by Pearson, Struthers and Theuerer[6] indicated that there was, in the case of antimony, a close correspondence between the electrical activity and the density of impurity atoms as determined from radioactive tracer measurements. Such studies have not been repeated, and little further work along this line has been done.

Another significant parameter is the energy level of the impurity. The ionization energy, ϵ_i is usually obtained from the relation:

$$\epsilon_i = \frac{\Delta \log R}{\Delta\{1/(\mathbf{k}T)\}} = 0{\cdot}198 \times 10^{-4} \frac{\Delta(\log_{10} R)}{\Delta(1/T)} \quad \ldots (3)$$

where R is the Hall coefficient.

This relation is based upon a simple exponential, with no factor 2 in the denominator of the exponential, as is often given. Although this factor 2 should be present for a system in which the number of electrons or holes available is exactly the same as the number of states, such a situation is seldom realized. Exact determinations of ϵ_i require a precise knowledge of the state of compensation of donors and acceptors in the crystal. Further discussion on this topic is given in sub-section 5.7.

Ionization energy represents the strength of binding between carriers and the impurity center. It may be as small as 0·01 eV, or as large as half the band gap. The theory of impurity centers in germanium is discussed in sub-section 5.3.

The ionization energy may be dependent upon the Fermi level in the material. This is the case for multiple levels. Thus it may be necessary to specify the state of ionization as well as the ionization energy.

For a simple donor element, the ionization process is described by the reaction-type formula:

$$D^0 = D^+ + q^-$$

and for an acceptor:

$$A^0 = A^- + q^+$$

For a multiple acceptor, the formula is:

$$A^0 = A^- + q^+ \text{ (single acceptor)}$$
$$A^- = A^{2-} + q^+ \text{ (double acceptor)}$$
$$A^{2-} = A^{3-} + q^+ \text{ (triple acceptor)}$$

The ionization energy is strictly defined only if the impurity density is low enough to ensure little interaction between impurities. In germanium at levels above about 10^{15} cm^{-3}, enough of the impurity atoms are in such close proximity to each other that the observed ionization energy begins to drop. This leads to (1) a broadening of the impurity level compared to the sharp value which is characteristic of very dilute solutions; and (2) a narrowing of the gap between the impurity levels and the conduction (or valence) band. At impurity densities of 10^{18} cm^{-3} or more, the gap may be reduced to zero, and the germanium may be degenerate or 'semi-metallic'. The atomic theory of over-lapping impurity wave functions and the formation of the impurity band[8] is in a very preliminary stage at the present time. As with certain other phenomena discussed, the specific action of various elements in producing semi-metallic behavior is poorly understood, so we shall give little attention to this topic.

1.5. Impurity Scattering

Every impurity center has a tendency to scatter carriers, since it is an imperfection of the crystal. The effect is especially noticeable if the centers are charged, and present to a density of 10^{17} cm^{-3} or more[9]. Neutral centers may also scatter carriers, since they perturb the bands of the germanium in the vicinity of the center and thus create a 'deformation potential' that acts as a scattering potential. The theory of neutral scattering[10] is in a more uncertain state than that of ionized impurity scattering, the effects of which are accounted for to a satisfactory approximation merely by taking account of the coulomb forces between center and carrier.

1.6. Transport Properties

Except as noted above, impurities have little effect upon the purely electrical properties, once the density of carriers and the mean free path have been given. That is, the transport properties, *e.g.* magneto-resistance, Hall effect, conductivity and many others, do not depend in any detectable way upon the impurity used for doping, except to the extent that the nature of the impurity affects the mean free path and its dependence upon energy.

1.7. Photoconductivity, Optical Absorption, Recombination and Trapping

Among the most important properties where specific effects of doping materials are to be found are those involving optical properties and photoconductivity.

Photoconductivity is characterized particularly by *wavelength response* and *sensitivity*. The wavelength response is partly given by the threshold, on the long wavelength side of which the quanta in the incident radiation are of too low an energy to excite photocarriers. Thus the threshold should be related to ionization energy, and indeed the two techniques are often used to check the location of energy levels of impurities.

Sensitivity is often expressed in terms of a parameter Y, called quantum yield, which may be defined as the ratio of the number of carriers passing through the photoconductor to the number of photons incident. For 'primary' photoconductors such as undoped germanium, the yield is about 1. For 'secondary'

photoconductors, the presence of certain impurity centers, or 'traps' may greatly increase the yield, to values of 10,000 or more.

Related to the traps are the recombination centers, which are partly responsible for the limitation of intrinsic photoconductivity. Recombination states are centers that make possible the rapid return of an excess electron from the conduction band into the valence band by allowing it to drop first into the recombination state, where it may recombine with the hole more easily than it could directly.

Traps are distinguished from recombination centers in that traps can accept and deliver carriers to and from only one band. Recombination states are in contact with carriers from either band. As discussed below, a given impurity such as copper can switch from one type of behavior to the other if the cross-sections for capture specifying its ability to capture carriers vary differently with temperature for the two bands.

The 'capture cross-section' is the effective area of the center for a particular process of capture. It depends upon the energy of the carrier approaching the center, upon its sign, and other factors. The rate of capture R_C is given by:

$$R_C = \sigma N_A v$$

where σ is the cross-section, N_A the density of capturing centers, v the velocity of the particles being captured.

The phenomenological theory of photoconductivity and recombination is outside the scope of this paper.

1.8. Influence of Impurities upon Intrinsic Properties

A proper subject for discussion is the specific effects that various impurities may have in changing what are normally considered lattice properties. Our knowledge of many of these effects is limited and hence there will not be much to discuss in many cases. Some effects of this kind which are known to occur, are:

(1) High doping levels influence lattice scattering of carriers.
(2) Impurities may affect the self-diffusion coefficient.
(3) The thermal conductivity and specific heat may be affected by high doping levels.
(4) Impurities may affect the intrinsic optical absorption.

1.9. 'Interference' Effects of One Impurity upon Another

An interesting group of phenomena is caused by the interference of one impurity with the effects of another. Among these effects are:

(1) The solubility of one element may be influenced by the presence of another. Thus high doping levels of boron influence the solubility of a rapidly diffusing element such as lithium.

(2) The diffusion coefficient of an impurity element into germanium may be affected by high doping with another. Thus these effects probably occur through the intermediate agency of vacancies, which control the diffusion rate, and upon their density as a function of doping level. This reduces the phenomenon essentially to that of (1) above with vacancy density regarded as an acceptor

impurity. We shall not attempt discussion of structural imperfections in this paper.

(3) In compensated samples, 'ion-pairing' may occur. This is an attempt at 'bonding' or compound formation between impurity atoms.

2. METHODS FOR ANALYSIS OF IMPURITIES IN GERMANIUM

It has generally been assumed that when a certain amount of impurity is added to a melt and a crystal is grown, the electrical conductivity and Hall coefficient give the impurity density as well as the carrier density, since the normal carriers are usually ionized at room temperature. For quantitative work, this assumption is generally not valid, particularly for elements which are very insoluble, diffuse rapidly or have deep states. Thus, means of direct analysis in the extremely low concentration range of 1 part in 10^6 to 1 part in 10^9 are required.

2.1. Radioactive Tracers

One of the most satisfactory techniques for partially solving the problem is the use of radioactive tracers. Indium, arsenic, copper, silver, gold, iron, nickel, tin, and many other elements have radioisotopes with a half-life suitable for impurity studies.

The use of radioisotopes helps in establishing the uniformity of the distribution of impurity as well as its average density. Autoradiographic techniques and counting procedures are both of importance. Unfortunately, unless great care is taken, the tracer techniques may be insufficiently accurate and precise.

Tracer studies have played a considerable role in studies of solubility, diffusion, effects of growth conditions upon crystal properties, influence of bending upon electrical properties, segregation coefficient, and electrical activity. For cases where the tracer technique is inapplicable, other techniques have been developed.

2.2. Neutron Activation Analysis

The method of neutron activation analysis[12] is particularly suitable for both qualitative and quantative analysis of samples containing unknown impurities. It is derived from the tracer method, and depends primarily on the conversion of the impurities present into radioisotopes by a neutron source. Unfortunately, the germanium as well as the impurities becomes highly radioactive, and a radiochemical separation is usually necessary. The separated fractions are analyzed for radioactivity by counting methods. Both quantitative and qualitative analysis is possible. The sensitivity is low for certain light elements, *e.g.* lithium and boron, which have very small capture cross-sections for neutrons but in favorable cases, *e.g.* arsenic and antimony, it is extremely high. As little as 10^{-9} to 10^{-11}g of total impurity in a sample can be detected, provided a source of high neutron flux is employed.

2.3. Mass Spectrometry

The mass spectrometer[13] is useful for certain problems such as the study of gases trapped or occluded in the crystal during growth; fusion of the crystal in a chamber adjacent to the spectrometer permits the evaluation of very small amounts of such impurities.

2.4. Chemical Analysis

Chemical analysis[14] is useful for detecting certain impurities, although its range generally does not exceed that of the optical spectrograph. Enrichment of the impurity, for example by volatilization of the germanium with a halide, may precede a photometric or other evaluation of the impurity content. Chemical methods may be useful in the range 1 per cent to 1 p.p.m. depending on the particular impurity and the techniques employed.

2.5. Optical Spectrography

The optical spectrograph[15] is still used for some problems, although the impurity density in most germanium samples of interest now is well below the range of the spectrograph. The spectrograph is particularly useful in analyzing doping agents for impurities that might interfere with the results. Evaluation of germanium by enriching grown crystals using successive pulling techniques is helpful in the application of the spectrograph for germanium work.

3. THE SEGREGATION AND SOLUBILITY OF IMPURITIES

The segregation of impurities in germanium has been discussed by Pfann[16], Burton *et al.*,[17] Hall[18] and others. We shall give here only a short review of the most important conclusions.

Table 2

Element	C_S/C_L	Element	C_S/C_L	Element	C_S/C_L
B	20	As	0·04	Cu	$1{\cdot}5 \times 10^{-5}$
Al	0·10	Sb	0·003	Ag	10^{-4}
Ga	0·10	Bi	4×10^{-5}	Au	3×10^{-5}
In	0·001	Zn	0·029	Ni	5×10^{-6}
Tl	4×10^{-5}	Li	$>0{\cdot}01$	Co	10^{-6}
P	0·12	Sn	0·02		

Table 2 shows results on the segregation coefficients of various impurities in germanium. Although segregation coefficients probably depend somewhat upon ionic radius, attempts to correlate these radii with values of k have not been very successful.

Thermodynamically, the difficulty of placing an impurity from the liquid into the solid crystal is expressed in terms of a free energy difference. The thermodynamical relations determining the variation of segregation coefficient and solubility with temperature have been discussed by Thurmond and Struthers[19], by Hodgkinson[20] and by Hall[21]. Using only the simplest ideas, Hall has shown that if there is an activation energy W for moving an impurity atom from liquid to solid, then there can be found, with few, if any, further assumptions, a simple correlation between segregation coefficient and solubility.

The basic assumption leads to the following general expression for the segregation coefficient:

$$k = C \exp(-W/\mathbf{k}T) \qquad \ldots (4)$$

where C depends upon the relative number of possible sites in the liquid and solid phases and is probably of the order of unity.

For many impurities, k is not large enough, for most of the crystal growth, to lower the melting point. Thus the segregation coefficient, expressed in terms of the melting point T_m, is:

$$k_m = C \exp(-W/kT_m) \quad \ldots (5)$$

On eliminating W from these two equations, one finds:

$$k = C^{1-(T_m/T)} \cdot k_m^{T_m/T} \quad \ldots (6)$$

The quantity C, which is of the order of unity, is given by $C = \exp(S/R)$, where S is the entropy of fusion of solute and R the gas constant. Equation (6) gives us a formula for finding k even if there is an appreciable lowering of the melting point.

This simple theory accounts for the two main properties of many impurities in equilibrium with germanium: (1) For small changes below the melting point, the solubility first increases as the temperature decreases; and (2) as the lowering of the melting point increases, the segregation coefficient begins to change significantly with temperature. This is because the temperature becomes too low to excite the impurity atom across the energy jump W needed on the average to transfer it from liquid to solid. Below the maximum, the solubility drops off with further lowering of temperature. This is the retrograde solubility range. Most impurities that have been studied have been shown to be retrograde in germanium.

Previously, Thurmond and Struthers had developed the thermodynamic relations leading to similar expressions, on a more elegant basis, and had shown that any impurity obeying the following conditions will have retrograde solubility:

(1) The impurity must form a solution having positive heat of solution and ideal entropy of solution.
(2) The liquid must be an ideal liquid in the thermodynamic sense.
(3) The segregation coefficient should be less than 0·1.

Figure 2 shows the relation between solubility, segregation coefficient and temperature, for various segregation coefficients, according to the simple theory of Hall. Each curve is equivalent to the solidus of a phase diagram. It is seen that the smaller the segregation coefficient, the lower the solubility at any temperature and also the higher the temperature at which the impurity goes into the retrograde range. For $k = 10^{-6}$, $T_{max} = 850°$ C, while for $k = 10^{-1}$, $T_{max} = 600°$ C. Also shown are some data for copper[36], one of the most intensively studied impurities. The curve fits the pattern shown if the segregation coefficient is 5×10^{-6}. The experimental value has been given as 2×10^{-5}.

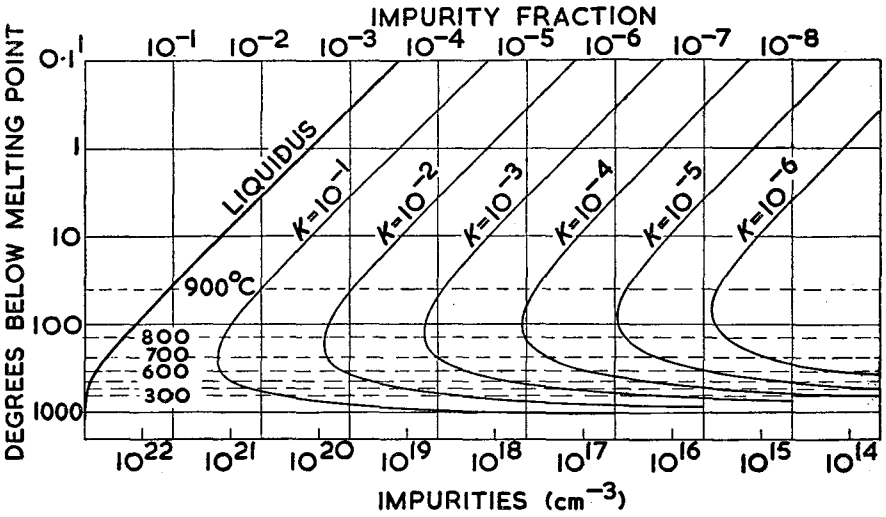
According to the theory, the liquidus is the same for all solutes having a small k value ($<0{\cdot}1$) and which form ideal liquids. The liquidus curve is given from the relation:

$$T_m - T = \frac{RTT_m \ln(1-x)}{H} \quad \ldots (7)$$

where x is the mol fraction of impurity in the liquid, T_m the melting point, H the heat of fusion of germanium, about 8,100 cal/gm-atom. Each solidus is then obtained from equation (7) by multiplying by k as given in equation (6).

Rate-growing is a specific effect of rate of growth upon k[22]. In the case of donors, the value of k may vary up to a factor of 3 or more, depending upon the rate of growth of the crystal. There is considerable difference in k for various orientations of the growth axis, the largest k occurring for the [111] direction. Acceptors show little or no dependence upon rate of growth. A useful method of growing P–N junctions is made possible by this difference.

The exact mechanism of rate-growing is not well understood. Presumably there is an absorption of impurity upon the growing crystal face in excess of the equilibrium concentration [18]. Diffusion makes possible the establishment



[*By courtesy of the General Electric Co.*

Figure 2. *Theoretical solidus curves for impurities with various segregation coefficients (Hall*[18]*).*

of equilibrium within the crystal. The difference between the diffusion coefficients of donors and acceptors is not great enough to account for the observed effects.

Longini and Greene[23] have suggested that rate-growing may depend upon the existence of a potential drop of the order of 0·2 eV at the interface between solid and liquid. The direction of the gradient would be such as to help acceptor ions into the interior. On the other hand, the movement of donors would be impeded, and in certain ranges of rate-growth their motion would be too slow to allow equilibrium to be established. Rate-growing effects occur at rates of about 2 in. hr^{-1} and higher, and varies according to the donor element considered.

4. IMPURITY DIFFUSION AND HEAT TREATMENT

Diffusion coefficients for germanium have been determined by the P–N junction method developed and first applied to germanium by the author[24, 25, 26, 27, 28], and by the method of radioactive tracers[29] which has long been a standard

technique for diffusion studies, particularly in metals. Special techniques that can be utilized include capacity measurements upon the dielectric layer formed by diffusion of impurities which have deep states[30]. Capacity versus voltage measurements on P–N junctions have also been used to measure diffusion[31].

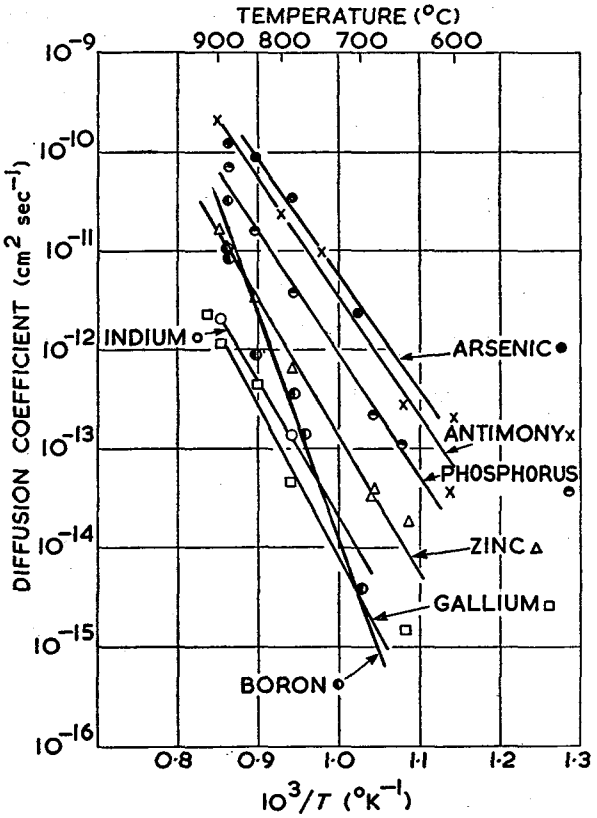


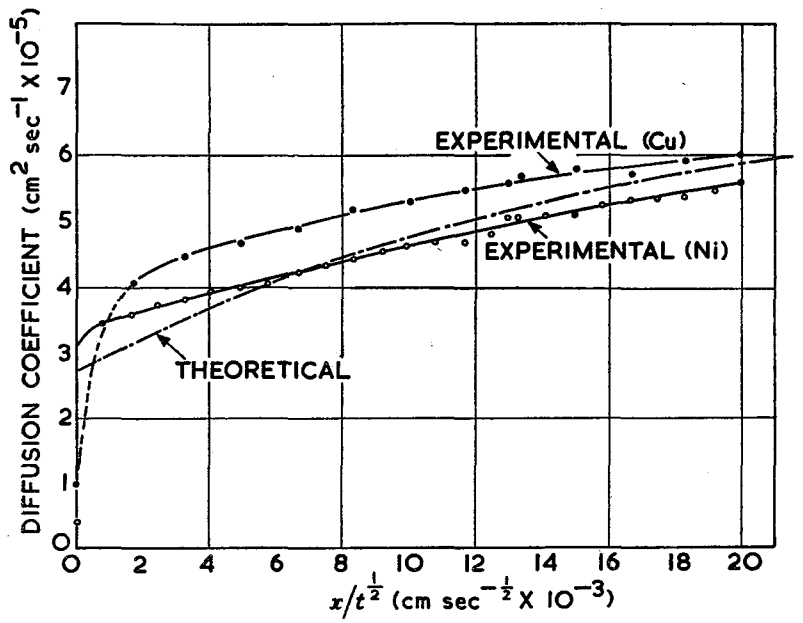
Figure 3. *Diffusion coefficients for various impurities in germanium.*

Diffusion coefficients for some important impurities in germanium are shown in Figure 3. The donors are seen to diffuse more rapidly than the acceptors, for the Group III–V elements. On the other hand, gold[51], copper, and other metals may diffuse much more rapidly than any of the donors. Results obtained by radioactive tracer measurements are in good agreement with those from P–N junctions[27].

Attempts to correlate the diffusion coefficient of substitutional impurities with the fundamental properties of the impurity itself and with those of the germanium lattice have not progressed beyond the rudimentary stage. It appears that substitutionals diffuse by the vacancy mechanism. Thus the activation energy should equal the energy required to form a vacancy, plus half the energy required for movement of the impurity into the vacancy[32].

Attempts to correlate diffusion coefficients with ionic radii have not been very successful. There is a correlation in the sense that the donors, which diffuse the more rapidly, are probably present as positive ions and the acceptors as negative ions. Positive ions are in general smaller than negative ions.

Longini and Greene[23] in the paper already cited have suggested that donors diffuse more rapidly than acceptors because they tend to promote the formation of acceptor-like vacancies when present in high concentration. The free energy



[*By courtesy of the Electrochemical Society.*

Figure 4. *Diffusion of copper into germanium.* (*van der Maesen and Brenkman*[37]).

term involved is the difference between the two energy levels involved. Acceptors, on the other hand, inhibit the formation of such acceptor vacancies. Recent diffusion studies indicate that the diffusion coefficient does, in reality, depend upon the base impurity in just this way[33].

4.1. Copper

The impurities that diffuse rapidly have been objects of much study. Probably the most widely studied of these are copper, whose rapid diffusion was discovered by Fuller *et al.* [34, 35, 36, 37, 38], lithium and nickel[37]. Figure 4 shows the penetration of copper into germanium as a function of distance[37]. The diffusion coefficient at 500° C is found to be about 10^{-5} cm^2 sec^{-1}; at 800° C it is about 5×10^{-5} cm^2 sec^{-1}. The activation energy, according to the results of Fuller *et al.*, is very small. The high rate of growth and the small activation energy are characteristic of interstitial diffusion, where, presumably, small ions move through open tunnels in the germanium structure.

One difficulty with such results is that they do not fit the theoretical diffusion expression well. Results on both nickel and copper have been analyzed by van der Maesen and Brenkman[37] who suggest that the diffusion is a two-particle process, involving both interstitial and substitutional copper: if the concentrations are N_s and N_i, the diffusion coefficients D_s and D_i, then the diffusion equations become:

$$\frac{\partial N_s}{\partial t} = D_s \frac{\partial^2 N_s}{\partial x^2} + k_1 N_i - k_2 N_s \qquad \ldots (8)$$

$$\frac{\partial N_i}{\partial t} = D_i \frac{\partial^2 N_i}{\partial x^2} - k_1 N_i + k_2 N_s \qquad \ldots (9)$$

where k_1 and k_2 are the reaction constants for the equations:

$$Cu_i \underset{k_2}{\overset{k_1}{\rightleftharpoons}} Cu_s$$

These authors suggest that copper precipitates as Cu_3Ge, although this is not an essential feature of the theory.

These equations have been solved in several limiting cases:
(1) when k_1 and k_2 are large; and (2) when they are small. For large values of k:

$$N_i = \frac{k_2}{k_1 + k_2} N_0 \operatorname{erfc} \tau,$$

$$N_s = \frac{k_1}{k_1 + k_2} N_0 \operatorname{erfc} \tau \qquad \ldots (10)$$

where $$\tau = \frac{x}{2\sqrt{(Dt)}}, D = \frac{k_2 D_i + k_1 D_s}{k_1 + k_2}$$

and N_0 = total concentration at $x = 0$.
On the other hand, for small values of k, the approximate solution is:

$$N_s = \frac{k_1}{k_1 + k_2} N_0 \operatorname{erfc} \tau_s + \left(\frac{k_1 k_2 N_0 t}{k_1 + k_2} \cdot \frac{1}{1 - \gamma^2}\right) \times$$

$$\left\{\frac{2}{\sqrt{\pi}} \gamma^2 \tau_s \exp(-\tau_s^2) + (1 + 2\tau_i^2)(\operatorname{erf}\tau_s - \operatorname{erf}\tau_i) + \frac{2}{\sqrt{\pi}} \exp(-\tau_s^2)\right\} \qquad \ldots (11)$$

where $$\tau_s = \frac{x}{2\sqrt{(D_s \tau)}}, \quad \tau_i = \frac{x}{2\sqrt{(D_i \tau)}}, \quad \gamma = \sqrt{\frac{D_s}{D_i}}$$

The diffusion coefficient D measured with respect to D_i is then, in terms of τ_i:

$$\frac{D}{D_i} = \frac{1}{3} \cdot \frac{4\tau_i^3 \operatorname{erfc}\tau_i + (2/\sqrt{\pi})(1 - 2\tau_i^2)\exp(-\tau_i^2)}{2i\operatorname{erfc}\tau_i} \qquad \ldots (12)$$

where i denotes the integrated error function erfc.

If a reasonable value of D_i is assumed, say 8×10^{-5} cm^2 sec^{-1}, then the complete curve can be plotted in terms of $x/t^{1/2}$. Figure 4 also shows such a

theoretical curve compared with the results for copper and nickel. Although this theory probably does not tell the whole story, it seems to be in the right direction.

Recent evidence indicates that diffusion of interstitial elements may be highly structure-sensitive. Tweet and Gallagher[40] have recently shown that apparent values for the diffusion coefficient of copper, determined by the use of radioactive ^{64}Cu and from electrical measurements, may vary from 2×10^{-8} to 4×10^{-5} cm^2 sec^{-1} depending upon whether the best available germanium was used, or whether the sample was full of small-angle grain boundaries. Since the value obtained for the imperfect germanium is in the range of the results of Fuller *et al.*, their results may be open to question.

It appears that the theory of van der Maesen and Brenkman may be correlated with the structure sensitivity of copper diffusion if structural factors determine the reaction constants k_1 and k_2. In particular, the conversion of interstitial copper into substitutional copper may be critically determined by dislocations and vacancies[41, 42].

Copper diffusion and precipitation account for the well-known heat treatment effect in germanium[34, 43, 44]. There is no evidence that this conclusion needs revising. Tracer studies with ^{64}Cu and many electrical studies have demonstrated a close correspondence between the properties of copper-diffused material and those of heat-treated material that might have acquired traces of copper from exposure to distilled or other water. Surface treatments with various cyanides now serve to eliminate these effects for practical purposes, although there appear to remain certain residual heat treatment effects that may be properly ascribed to the thermal formation of vacancies[45, 46, 47].

Related to both the diffusion and heat treatment effect of interstitials are annealing and precipitation. When a diffused specimen is heated for long times at low temperatures, precipitation may take place, since for many retrograde impurities the solubility drops with lowered temperature. It appears that this process is also highly structure-sensitive, and that dislocations and other imperfections furnish nuclei for the formation of the metal phase. Whereas germanium of good structure may require several hundred hours to reconvert to N-type at 450° C, if the quenching and annealing are done so as to minimize strains, use of imperfect starting material may allow reconversion in only a few hours.

4.2. Lithium

In 1950 the author, in collaboration with Jacobi[48], noted the occurrence of precipitation and solution phenomena in germanium, in the temperature range — 20° C to 200° C. This range is too low to be accounted for in terms of copper diffusion. It was first reported by Fuller and Ditzenberger[39] that lithium produces such effects. Table 3 shows the diffusion coefficient of lithium in germanium as a function of temperature. The solubility data are probably not very accurate but they are the best available at the present time. The rapid diffusion of lithium indicates that the metal diffuses as an interstitial element, and probably as a positive ion. Lithium is a donor element with an ionization energy of about 0·01eV. This value, in agreement with the results of the hydrogen model, indicates that lithium attains the noble gas configuration of its electrons by releasing its one valence electron.

Table 3. Diffusion Coefficient of Lithium in Germanium

Temperature (°C)	D (cm^2sec^{-1})
400	4×10^{-7}
500	1×10^{-6}
600	3×10^{-6}
700	6×10^{-6}

5. THE ELECTRICAL EFFECTS OF IMPURITIES IN GERMANIUM

5.1. The Hydrogen Model for Impurity Action

Figure 5 shows the hydrogen model for an impurity in germanium. The electron is shown in its elliptical orbit swinging about the impurity ion. The orbit includes many lattice atoms, and has a semi-major axis of about 8 Å, according to the simple Bethe theory. For germanium, the effective mass correction increases this value to about 45 Å.

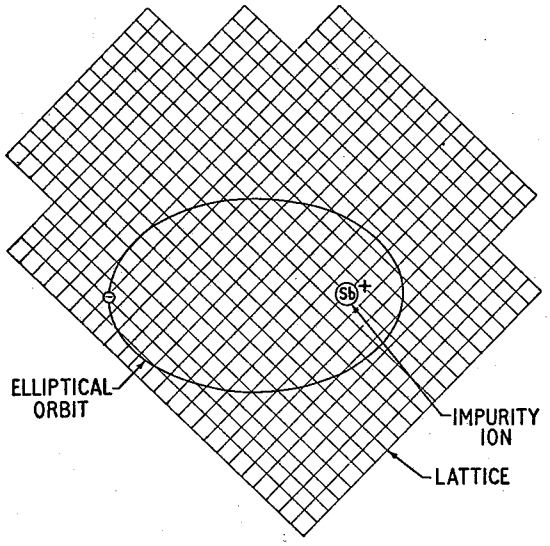


Figure 5. *Model of hydrogen-like impurity center in germanium.*

According to the Bethe model[49], the ionization energy ϵ_i for such a donor electron is:

$$\epsilon_i = \frac{13{\cdot}6}{\kappa^2} \cdot \frac{m^*}{m} \qquad \ldots (13)$$

The major axis of the orbit is:

$$r = 0{\cdot}53 \frac{m}{m^*} \kappa \qquad \ldots (14)$$

where κ is the dielectric constant, m^* is the effective mass, and m the mass of free electron. Using a value of 0·2 m for the effective mass, a value consistent with several experimental results, one obtains values of the order 0·01 eV, agreeing well with the experimental results[50, 51]. The Semi-major axis of the orbit is then about 45 Å.

5.2. Experimental Results—Group III–Group V

Table 4 shows the results of Geballe and Morin[51] for the ionization energies of the Group III–V impurities in germanium, determined from the slopes of the Hall and resistivity curves. Although all these results are in the neighborhood of 0·01 eV, there are small but, according to the above authors, real differences among them. The reason for the differences is to be sought in the differing electronic structure of the various impurities in the germanium lattice. A more profound analysis is needed for such effects than that which is provided by the simple Bethe hydrogen model.

Table 4. Ionization Energies of Group III and Group V Impurities in Germanium

P-type element	Ionization energy (eV)	N-type element	Ionization energy (eV)
B	0·0104	P	0·0120
Al	0·0102	As	0·0127
Ga	0·0108	Sb	0·0096
In	0·0112	Li	0·0093

Among the Group III–V elements not covered in the table are thallium, bismuth and nitrogen. Thallium has been studied by the author, and found to be similar to the other acceptor elements in behavior, although it has a smaller solubility and segregation coefficient than the others. The ionization energy appears to be normal, although no precise measurements have been attempted. Bismuth is probably similar.

Nitrogen has been used as a constituent of furnace gases (forming gas in particular being 90 per cent nitrogen and 10 per cent hydrogen). No evidence has been observed that nitrogen has any donor action when used in this form. Recently, however, Billig[52] has demonstrated that the molecular binding of N_2 inhibits the donor action. He states that if nitrogen is added to the furnace in the form of ammonia, NH_3, then strongly N-type germanium is formed. Again the impurity center appears to have the normal hydrogen-like behavior.

5.3. Theory of Wannier Orbitals for Impurities in Germanium

A beginning has been made on the development of a theory of the impurity centers in germanium and silicon by Luttinger and co-workers[53, 54, 55], by Lampert[56] and by Kittel and Mitchell[57, 58], who have studied various aspects of the problem. All have worked with the Wannier orbitals, solutions to a form of the Schrödinger equation that are particularly well suited to the problem.

The form of the time-independent Schrödinger equation used for calculating the allowed energy levels, E, for an electron on states of an impurity is:

$$\left\{-\frac{\hbar^2\nabla^2}{2m}+V(r)+U(r)\right\}\psi(r)=E\psi(r) \qquad \ldots (15)$$

where $V(r)$ is the lattice potential, $U(r)$ is the additional potential supplied by the impurity atom. We assume that $U(r)=-q^2/\kappa r$.

In the Wannier method, the above equation is written in terms of effective masses, the wave functions being expressed in the form:

$$\psi(r)=F\psi(k) \qquad \ldots (16)$$

where $\psi(k)$ is a Bloch wave function at the minimum energy of the energy band structure (in the many-valley models now in use for silicon and germanium, the conduction band structure consists of several ellipsoids oriented along the $\langle 100\rangle$ axes in the case of silicon, and along the $\langle 111\rangle$ axes in the case of germanium).

The F's obey the effective mass equations:

$$\left\{-\frac{\hbar^2}{2m_1}\cdot\frac{\partial^2}{\partial x_1^2}-\frac{\hbar^2}{2m_2}\left(\frac{\partial^2}{\partial x_1^2}+\frac{\partial^2}{\partial y_1^2}\right)-\frac{e^2}{\kappa r}-\epsilon_0\right\}F_1(r)=0 \qquad \ldots (17)$$

where m_1 and m_2 are the principal effective masses of the ellipsoids of revolution. For silicon $m_1=0{\cdot}98\ m$ and $m_2=0{\cdot}19\ m$. Since the calculations have been studied in detail mostly for silicon, we shall describe only general features of the results.

Although the Wannier theory as developed by Kohn and Luttinger is more exact than the Bethe model, it is inaccurate in describing the situation close to the donor ion. Difficulties arise firstly from the breakdown of the whole effective mass picture in this region, and secondly because of the use of the dielectric constant to give the potential near the donor ion. Thus, the inner orbits of the electron are not as accurately described as the outer ones. Similar circumstances also prevent the use of the theory for deep states such as those of gold and copper in germanium, where the binding is considerably stronger than for the usual impurities.

From the direct application of the theory, using the values $m_1=0{\cdot}98\ m$ and $m_2=0{\cdot}19\ m$ for the two effective electron masses in silicon, an ionization energy of 0·029 eV was found for phosphorus in silicon, compared to the experimental value of 0·044 eV. Presumably errors of the same percentage value would be found for germanium.

Although some experimental confirmation of the picture has been obtained from optical absorption studies of boron in silicon, little similar evidence is available as yet for germanium.

A corresponding theory has been worked out for the acceptor states in both silicon and germanium by Kohn and Schechter[55] and by Kittel and Mitchell[58]. The former, in particular, have found results in close agreement with experiment, for boron a value of $\epsilon_1=0{\cdot}0093$eV being found, compared with the experimental value of 0·010 eV. One result of these calculations is the coincidental character of the agreement between the values for the ionization energy of the acceptors relative to that of the donors.

5.4. General Methods for Studying Multiple Levels

For studying the direct action of a single level impurity as compared to multiple levels, the methods are relatively simple. Highly purified germanium is doped during the growth of a single crystal with a small amount of spectrographically pure impurity. The change of conductivity before and after doping

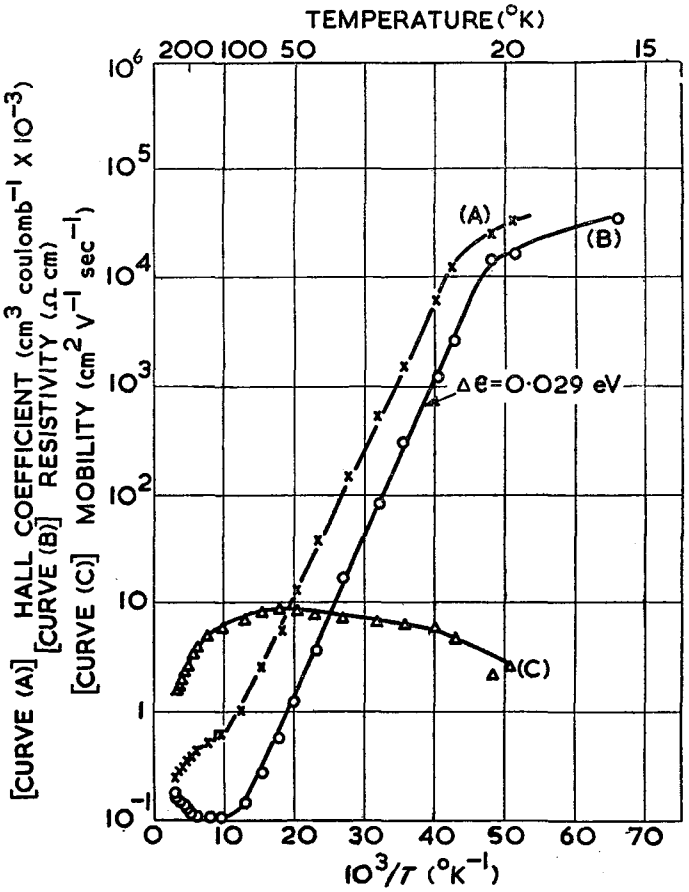


Figure 6. *Conduction properties of zinc-doped germanium.*

gives an indication of the electrical activity, type, and segregation coefficient. The temperature dependence of the Hall coefficient and resistivity give information about the energy levels. Solubility limits are indicated as segregation proceeds by the formation of lineage (numerous small angle grain boundaries usually appear). Microscopic examination usually shows inclusions of precipitated impurity.

The greatest attention must be given to the purity of the additions used, particularly if the segregation coefficient is small, as it is for so many of the interesting elements. The presence, for example, of a small amount of arsenic is more effective by a factor of about 10,000 than the same amount of an element

with $k = 10^{-5}$, because the amount in the crystal varies as the ratio of the two segregation coefficients.

For the study of multiple levels, more sophisticated methods must be used. In brief, the entire forbidden zone is 'scanned' by counter-doping with the proper Group III–V element, in order to bring out the various levels characteristic of the impurity. Since it cannot be assumed that all the levels of one element are either donor or acceptor, counter-doping in both the P- and N-direction may be needed. The principle of counter-doping may be understood in terms of Figure 9 which shows how the addition of compensating impurities tends to lock in the Fermi level close to the various impurity energy levels, to within terms of the order of kT. It is the value of the Fermi level, of course, which is obtained from the slope of the Hall curves.

It can be seen that it is the relative amount of the counter-doping agent with respect to that of the unknown in the crystal which is important. If the solubility is small, care in counter-doping is needed to obtain the desired 'resolving power'. Additional important information is obtained if equal amounts of counter-doping agent are required to shift from one level to the next. This indicates equality in the numbers of the various levels, and strongly indicates that they are various ionization states of the element, rather than coincidental levels arising, for example, from atoms located in various types of site.

Lack of equality in the apparent level density, however, may depend upon experimental uncertainties, and upon impurities in the additions that vary from run to run. It is for such reasons that many crystals often have to be made before definite assignments can be made as to levels and their relationships.

Resistivity profiles along the crystal in the neighborhood of the doping point often provide information about the diffusion coefficient. A sharp transition indicates a slowly diffusing substitutional impurity; a slow transition, with effects observed in the supposedly undoped portion, indicates a rapidly diffusing interstitial impurity.

Analysis of samples by Hall effect and resistivity may require facilities for measurements from liquid helium temperature to above room temperature. Ionization energies below 0·02 eV require liquid helium, those between 0·02 eV and 0·07 eV require liquid hydrogen, and those above 0·07 eV require liquid nitrogen. Although information about each class of impurity can be obtained from the shape of the Hall curve entering the saturation range in the next higher temperature range, this information is usually unsatisfactory and inaccurate.

Checking of impurity content and evaluation of electrical activity by use of radioactive tracers is often helpful. Radioisotopes are especially useful in determining whether apparently inactive elements are insoluble, or merely neutral in their action.

Diffusion may also be a useful technique both for counter-doping and for incorporation of the impurity element under study. It is possible, although the technique has not been worked out in detail, that a wide range of the forbidden zone could be covered in a single sample by diffusing into samples having initially wide variations of the counter-doping element, or by diffusing the counter-doping element into a crystal containing wide variations of the unknown. Point-contact spreading resistance measurements could serve to indicate the

Fermi level. Diffusion of the unknown, if it diffuses rapidly, is especially useful since the crystal then acts as a sieve to screen out slowly diffusing Group III–V impurities, although this is little help in preventing incorporation of other widely occurring interstitials such as copper.

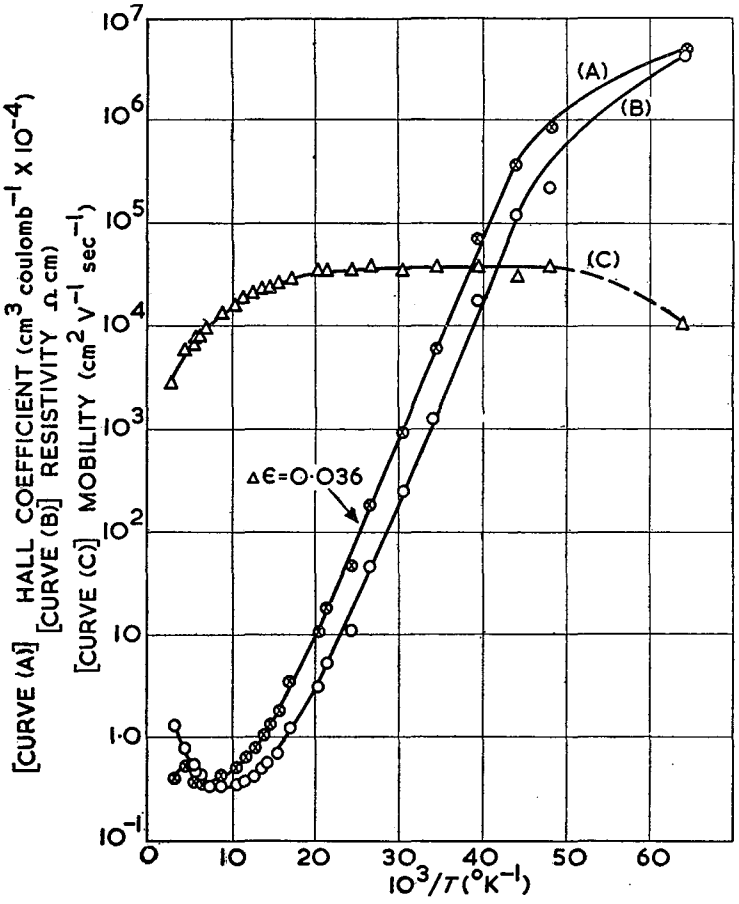


Figure 7. *Conduction properties of copper-doped germanium.*

5.5. Group II Elements: Zinc, Cadmium and Several Transition Elements: Double Acceptors

A striking development in the study of doping elements in germanium was the discovery that several elements that have electrical activity in germanium produce states unlike those of the Group III–V elements.

The first deep state found in germanium was that of zinc[59], which was found to produce an acceptor level 0·031 eV above the valence band. Shortly afterwards, copper[60, 61, 62] and platinum[60] were found to produce similar states. Figures 6, 7 and 8 show the conduction properties of zinc-, copper-, and platinum-doped germanium. The ionization energies were determined from the simple relation of equation (3).

The theory of the zinc-type levels is more unsatisfactory than that of the Group III–V impurities. Since the ionization energy is about 4 times that of the usual donors and acceptors, and since ionization energy should increase as the square of the charge on the center, it is logical to think of the zinc being present as a doubly charged negative ion, with two more or less incompletely

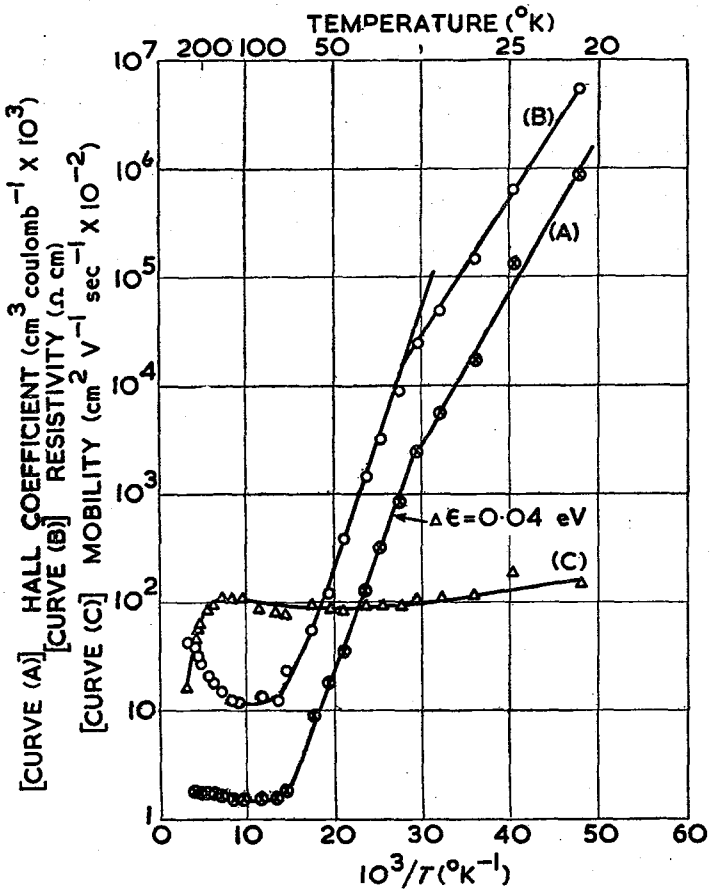


Figure 8. *Conduction properties of platinum-doped germanium* (40 Ωcm, *P-type sample*).

screened holes moving about this nucleus. Since the first helium ionization energy is about 24 eV, and if the dielectric constant can be applied as for the hydrogen model (a very dubious assumption), then one obtains by analogy the value 0·02 eV for ϵ_1.

The model is confirmed by recent work of Tyler and Woodbury[63] showing that zinc has a second acceptor level at about 0·1 eV. Since the second ionization potential of helium is 54 eV, the agreement is not extremely good, but is probably as good as can be expected from the dubious assumptions involved.

These results indicate that zinc does attain the four-bond tetrahedral structure characteristic of the germanium lattice and that it is substitutional. The diffusion

coefficient, being comparable to that of indium, also confirms this supposition. Similar results apply to cadmium.

The properties of transition metals have been among the most intensively studied[64, 65, 66, 67, 68, 69]. Burton *et al.* were the first to show that nickel produced deep levels in germanium and Tyler and his co-workers have studied iron, nickel, cobalt and manganese. All these transition metals are characterized by $4s^2$ electronic structures—that is, there are two valence electrons in the outer shell and a variable number of electrons in the incomplete 3d shell. Presumably these inner electrons play little or no part in the effects of the element in germanium. No theory, however, as to the action of these elements in germanium is available, as far as quantitative results are concerned.

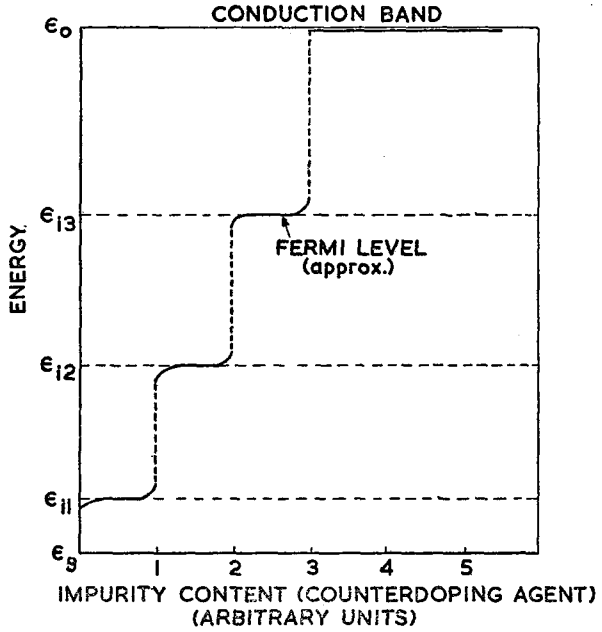


Figure 9. *Fermi level versus doping level for sample containing deep states.*

Manganese in particular produces an interesting level, almost in the middle of the band. Figure 10 shows the resistivity of manganese-doped germanium. Values of the order of 10^{14} Ωcm are attainable close to liquid air temperatures.

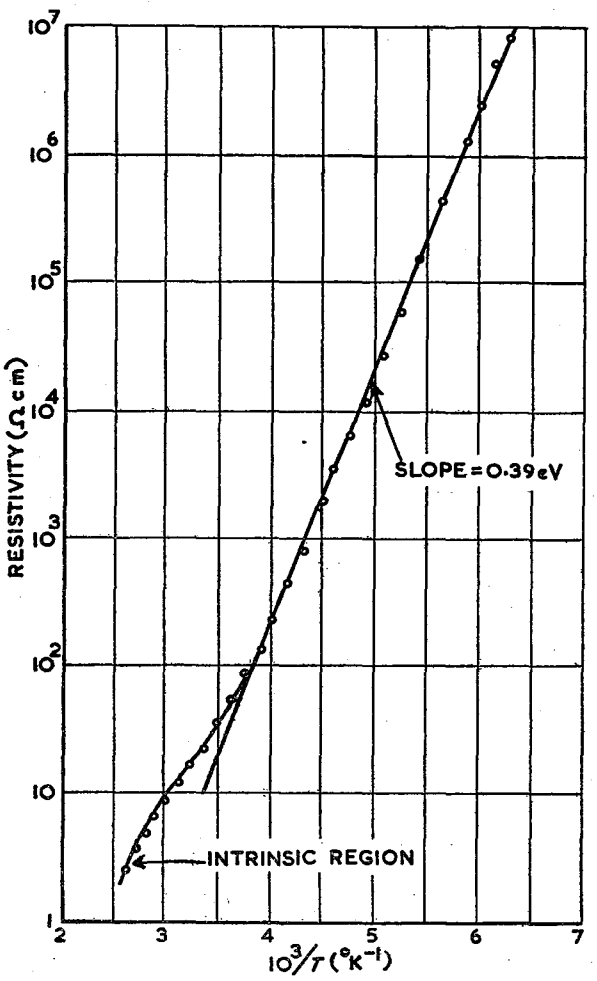
Table 5 shows the ionization energies of the acceptor levels obtained by Tyler *et al.* for these elements. The numbers are measured from the valence band for the lower level and from the conduction band for the upper level.

Although no quantitative studies to compare the level densities were made, it appears that the states are associated and represent successive ionization of the metal atom to give singly and doubly charged ions. Apparently, therefore, these elements also form the tetrahedral bonding arrangement in the germanium lattice.

All the transition elements mentioned above appear to diffuse rapidly in

Table 5. Ionization Energies of Transition Elements

Element	Lower level (eV)	Upper level (eV)
Fe	0·35	0·27
Co	0·25	0·31
Ni	0·22	0·30
Mn	0·16	0·37



[*By courtesy of the Physical Review.*

Figure 10. *Resistivity of manganese-doped germanium. (N-type sample). (Tyler, Newman and Woodbury*[69])

germanium, although quantitative data are lacking for all but nickel. It appears that the model of van der Maesen and Brenkman[37] will account for all the observed results if it is assumed that both substitutional and interstitial ions

are present. The electrical properties are determined by the density of ions in the substitutional sites, the diffusion by the interstitials. The exact nature of these properties for the various elements remains to be worked out.

Beside their direct electrical effects, the transition elements are particularly effective as recombination states, as discussed by Burton *et al.* They exhibit pronounced photoconductivity and trapping effects.

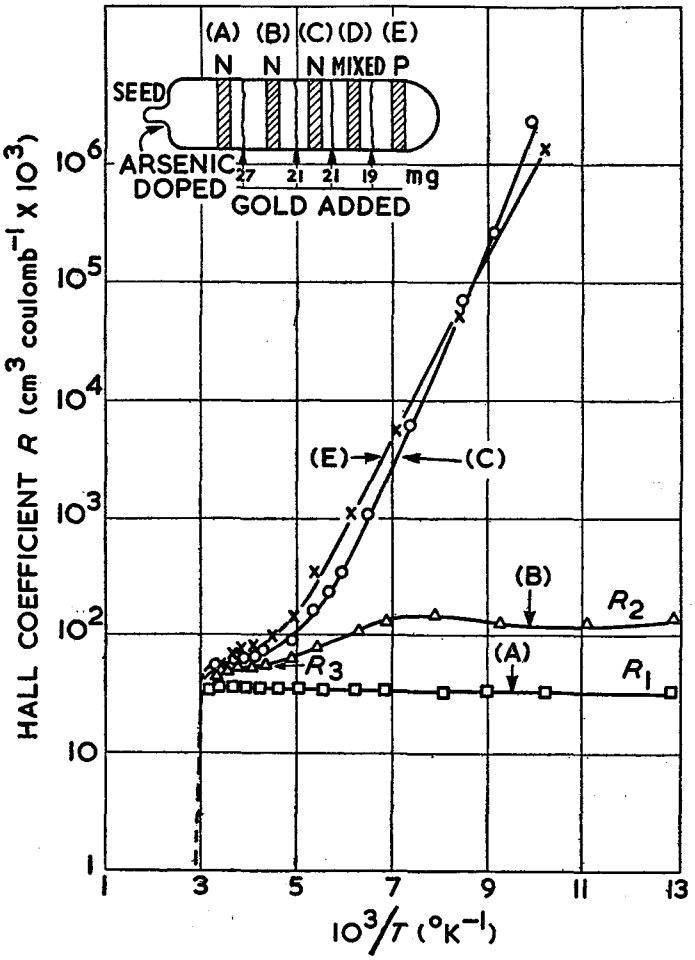


Figure 11. *Evidence for double acceptor levels in gold-doped germanium.*

5.6. Group I Elements: Triple Acceptors

The first element to be shown to have multiple acceptor levels in germanium was gold; it has also been the most thoroughly studied of the multiple acceptors[70, 71, 72, 73, 74, 75, 76]. Hence, we shall be able to give a more extensive discussion of gold than of some of the other elements.

The experimental evidence for multiple levels is indicated in Figure 11. The successive plots of Hall coefficient (or carrier density) apply to samples containing successively larger quantities of pure gold (purity 99·999 per cent). The base crystal contained arsenic.

Addition of the gold increases the Hall coefficient, and decreases the carrier density. This is an indication of the acceptor character. The presence of a new slope indicates that some of the electrons lost by the addition of gold are re-excited to the conduction band. This is primary evidence for an acceptor level in the upper half of the forbidden band. Calculation of the change in electron concentration shows that, within experimental error, the number of electrons re-excited is just half the total removed by the gold. This is good evidence

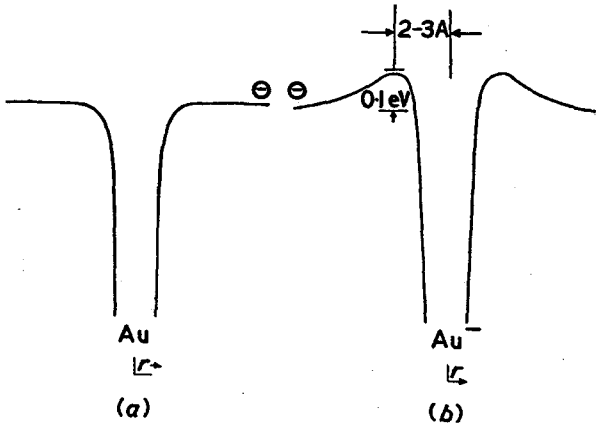


Figure 12. *Potential energy diagram for electrons and gold centers (a) electron plus neutral gold atom, (b) electron plus singly charged gold ion.*

that there are just as many acceptor states in the lower half of the band as in the upper half. The curve indicates that the second acceptor state is 0·2 eV below the conduction band.

In the above experiments, of course, the germanium remains N-type. When sufficient gold has been added to remove all the arsenic carriers, it can be seen (Curve E) that the gold becomes P-type. This enables us to obtain direct evidence of the acceptor state in the lower half of the band. The slope indicates a value of 0·15 eV.

Because of the apparent equality between the number of upper and lower states, the author suggested that the upper level is merely a second ionization state, in which an electron is added to the filled lower state. The gold ion would thus be Au^{2-} when the upper state is filled, Au^{-} when the lower state is filled.

Recent results[69] have also shown the presence of a third acceptor level at 0·05 eV below the conduction band. This is probably another ionization of the gold to form Au^{3-}.

It is remarkable that the Au^{-} center and the Au^{2-} are apparently capable of capturing an additional electron in spite of the coulomb repulsion between the electron and the ion. This repulsion leads to an energy barrier to be surmounted by the electron and to a rather small probability, or cross-section, for

capture of the electron. Diagrams illustrating the probable potential energy of an electron close to the Au^0 and the Au^- center are shown in Figure 12.

The negative charge on the Au^-, Au^{2-}, and Au^{3-} centers naturally tends to give them a large cross-section for capturing holes. The Au^0 and Au^- centers having the ability to capture either electrons or holes, fill the requirements for a recombination state, as discussed in Section 7.

It seems probable that the gold in the Au^{3-} state has the complete tetrahedral bonding structure proper to a substitutional impurity in germanium.

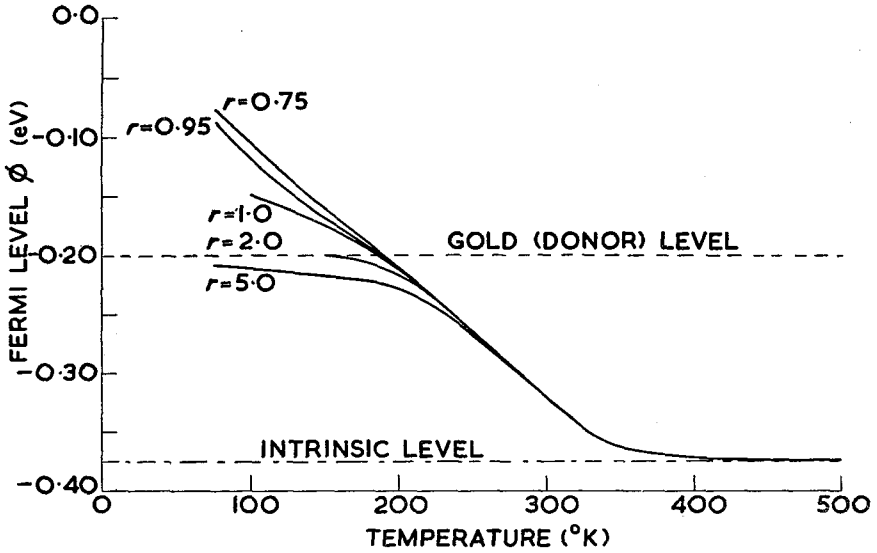


Figure 13. *Fermi level in N-type gold-doped germanium, as function of temperature.* $\Delta\epsilon_G = 0{\cdot}75eV$; $\Delta\epsilon_1 = 0{\cdot}20eV$; $N_e = 10^{14}\ cm^{-3}$.

It is clear from the above that both N- and P-type high- or low-resistivity germanium can be produced by doping with gold. The type of conductivity and the temperature dependence of the conduction properties relate to the relative proportion of gold atoms and other impurities, *e.g.* the usual donors and acceptors. In particular, we must now specify the more exact methods of obtaining the ionization energy values from the properties of compensated samples.

A convenient parameter to specify degree of compensation is the *trapping ratio r*. It is defined, for an N-type semiconductor, as N_d/N_e, where N_d is the number of deep (effectively donor) levels in the upper half of the band, and N_e is the number of electrons present over and above that needed to fill the valence band and the low-lying acceptor levels (in the present theory, the ordinary donor levels are lumped statistically with the conduction band). Then if $r > 1$, the number of electrons is insufficient to fill all the deep-lying levels, and the material is high-resistivity N-type; if $r < 1$, there is an excess of electrons above that number needed to fill the deep levels, and the material will have low-resistance for all temperatures above the range of deionization of the ordinary shallow levels.

Analysis of the conduction properties of such a two-level system has been made by the author. The first step is calculation of the Fermi level as a function of temperature, for various values of r. This is done graphically. The Fermi level results for N-type germanium containing 10^{14} cm^{-3} gold centers is shown in Figure 13. From the Fermi level the carrier density is easily calculated, also for various r. Results of these calculations, together with the experimental

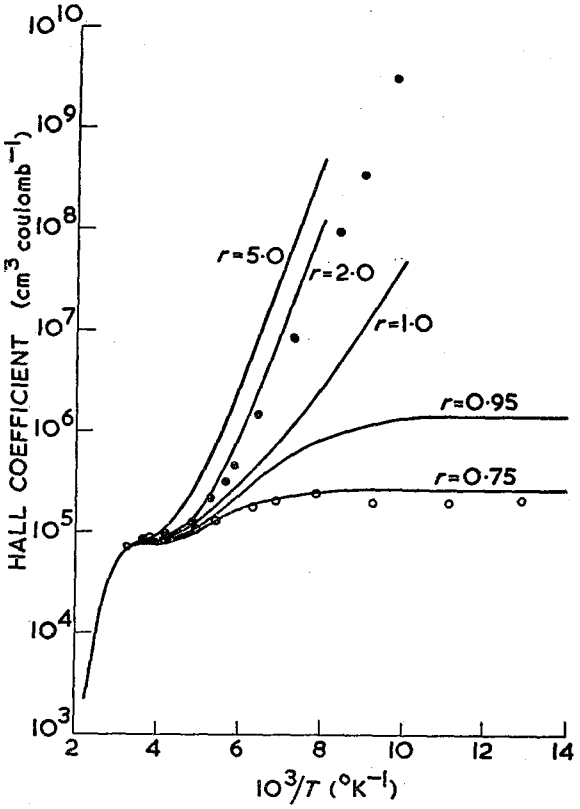


Figure 14. *Calculated Hall curves for N-type germanium of different trapping ratios.* $\Delta\epsilon_0 = 0{\cdot}75eV$; $\Delta\epsilon_1 = 0{\cdot}20eV$; $N_e = 10^{14}\ cm^{-3}$.

results for two samples, are shown in Figure 14. All the results observed to date appear to be consistent with this model. Extension to the case of the four known ionization energies for gold is easily made.

Figure 14 shows how the slope of the Hall curve can vary appreciably with r. Unless the degree of compensation is known well, considerable errors are possible.

In some ways, analysis of data for $r < 1$ is more suitable for determination of ionization energy than those just outlined for $r > 1$. An illustration is given in Figure 15. The method is based upon the fundamental property that the Fermi level is that energy at which the probability of a state being filled is one-half. Consequently, if the density $n_{1/2}$ is found at which just half the deep states

have been ionized, and if the Fermi level ϕ is found in terms of $n_{1/2}$, this Fermi level equals the ionization energy. This is done from the equation relating Fermi level to carrier density:

$$n_{1/2} = \left(\frac{2\pi m^* \mathbf{k}T}{h^2}\right)^{3/2} \exp\left(\phi/\mathbf{k}T\right) \qquad \ldots (18)$$

The effective mass to be used in this equation is 0·20 m. This is the 'density of states' mass.

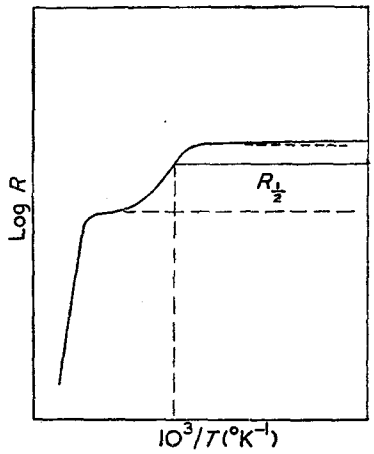
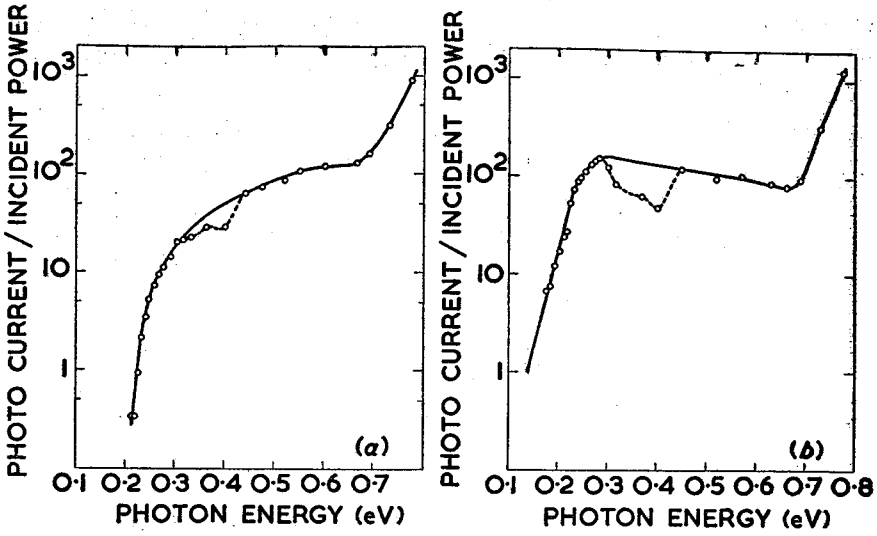


Figure 15. *Determination of ionization energy when* $r < 1$.

Application of the method requires that the ionization be complete well before the intrinsic conduction sets in. For ionization energies between 0·08 and 0·25 eV the conditions seem to be satisfactory if the doping level is not too high (saturation carrier density less than 10^{16} cm^{-3}). Values of the ionization energy for gold determined in this way from Curve B of Figure 11 agree well with those determined from the slopes of the Hall curve for high-resistivity samples.

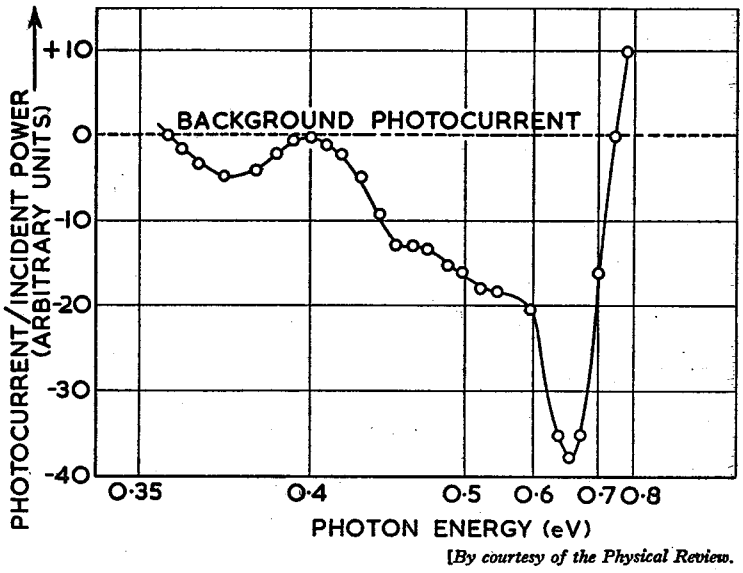
Photoconductivity measurements also permit an independent check on the thermal evaluation of ionization energy. Figure 16 shows some of Newman's results[76] on both P-type and N-type samples of gold-doped germanium. Although there is some uncertainty as to the proper wavelength to be chosen as the threshold for photoconductivity, the assumption of a value of 1 per cent of the plateau value gives results in good agreement with those of thermal measurement.

There is a characteristic difference in the shapes of the photoconductivity versus wavelength curves, depending on the type of conductivity. The curve cuts off more sharply for P-type than for N-type material. Although the detailed theory of the difference has not been worked out, it appears that the differing electric charge on the center in the two cases produces a different cross-section for the photo-ionization process and this leads to differing photo-currents close to the long wavelength limit.



[*By courtesy of the General Electric Co.*

Figure 16. *Photoconductivity versus wavelength for (a) N- and (b) P-type gold doped germanium, both at* 77° K. (*Newman*[76])



[*By courtesy of the Physical Review.*

Figure 17. *Quench effect in gold-doped germanium.* (*Newman*[76])

Results such as those of Figure 16 show that impurity photoconductivity of gold-doped germanium extends to 6 and 8 microns respectively for N- and P-type material. P-type gold-doped germanium has therefore considerable promise as an infra-red photoconductor. Previously the best photoconductors, such as PbTe and PbSe, have been restricted in their range to about 6 micron;

of course, the necessity for cooling germanium to 78° K is a hindrance to application. For a discussion of the performance of germanium photoconductors, see reference [3].

Also of interest is the 'quench effect' observed by Newman with N-type gold-doped germanium (Figure 17). This effect is a reduction of current through the sample below the dark value when light of a definite band in the infra-red

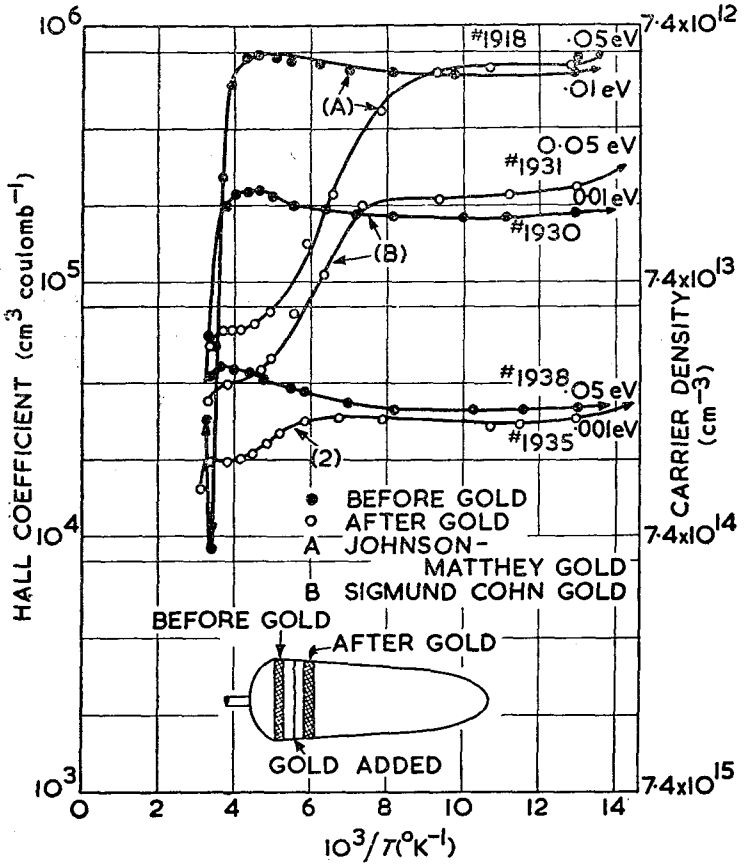


Figure 18. *Evidence for amphoteric action in gold-doped P-type germanium. In all cases the gold added was* 100*mg per* 100*g germanium.*

(0·45–0·64 eV) is incident upon the sample. When the light is turned on the sample there is an initial increase of current, followed by a decrease below the dark value. The effect is similar to the quench effect well known in the case of cadmium sulfide. As in the case of other photoconductivity studies, occurrence of the effect in germanium containing states at known energies and with more easily determinable properties than those in cadmium sulfide, gives hope that the quench effect, which is not well understood, will soon be clarified in both cases.

5.7. Amphoteric Impurity Action in Gold-Doped Germanium

Gold-doped germanium has proved of great interest in that another new type of impurity action occurs in it. This is amphoteric impurity action, which may be defined as the ability of an impurity to act as either a donor or an acceptor. There appear to be, at present, two possible types of amphoteric action.

(1) When the impurity is present in a single lattice position and under various conditions of the Fermi level, the center can either give off or accept electrons. This is the type of action observed in gold-doped germanium.

(2) When the impurity may take up either of several positions in the lattice, and may donate or accept electrons according to the type of site. It is possible,

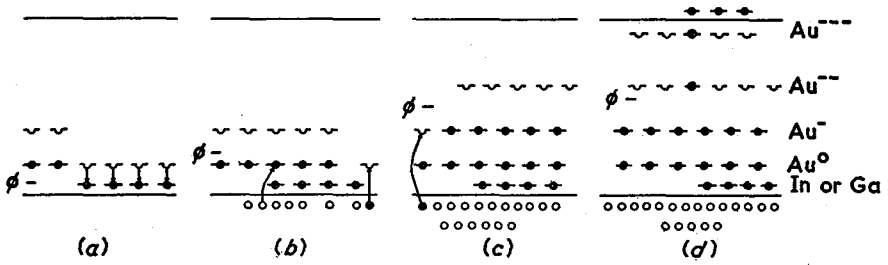


Figure 19. *Various stages of ionization in indium-gold doped germanium.* (*a*) *Gold added after Gallium*; $T = 0°$ K, (*b*) *First excitation region*, $T = 20$–$77°$ K, (*c*) *Second excitation region*, $T = 77$–$298°$ K; (*d*) *Intrinsic range.*

although it has not been definitely established, that copper in germanium acts in this manner, since ionic mobility measurements at high temperature show it to act here as a positive ion.

Experimental evidence for the amphoteric action of gold in germanium was found by the study of P-type gold-doped germanium containing indium or gallium, at low temperatures.

Results are indicated in Figure 18. There it is seen that the low temperature properties of gallium-doped germanium are changed in an interesting way by the addition of small amounts of gold. The carrier density at saturation remains the same to within a few per cent. but the slope, or ionization energy, has changed from the value of 0·01 eV, characteristic of gallium, to about 0·05 eV. This is precisely the type of behavior expected if the gold furnished electrons from states at a certain level, filling up the gallium acceptors and leaving acceptor-like levels to which the electrons could be re-excited.

A critical test of the theory comes with the addition of gallium in quantities just slightly exceeding the gold content. In this case, the gold should have insufficient electrons to fill up the gallium states, and the ionization energy should remain at 0·01 eV. This is the behavior observed.

This correspondence was determined using the gold content measured in terms of the lower acceptor level, shown being ionized between 78° K and room temperature. Since the correspondence is nearly perfect, we have evidence that the low donor states are numerically just equal to the lower gold acceptor levels. Again we have evidence that the donor state is linked with the acceptor states already discussed, and is simply the ionization state formed by removing

an electron from a neutral atom. The ionization energy is defined in this case as the energy required to add an electron to a positively charged Au^+ ion.

It is possible that the donor state could arise from the second type amphoteric action mentioned, *viz.* gold atoms located at different parts of the lattice. The correspondence just mentioned would then be accidental. This does not seem probable, but the possibility cannot be ruled out.

The complete energy diagram of gold in germanium is shown in Figure 19. There we show electrons dropped from the gold donor levels to fill up acceptors, leaving empty donor levels. At higher temperatures they can be re-excited. As soon as they are again filled, at still higher temperatures, the next stage of ionization, that of the lower acceptor level, can take place. It is significant that electrons cannot be excited from the donor level to the aceeptor levels since all the lower levels must be filled before the next higher acceptor level exists.

Thus gold in germanium appears to assume any of five ionization states, depending upon the Fermi level. It is presumed that the donor action results from the loss of the single 6s valence electron.

5.8. Other Deep States in Germanium

It appears that in the work already cited we have covered most of the types of states known to exist in germanium. Other doping elements produce states of the same general types as those already discussed. Among these other elements are silver, cadmium, mercury, and vanadium[77].

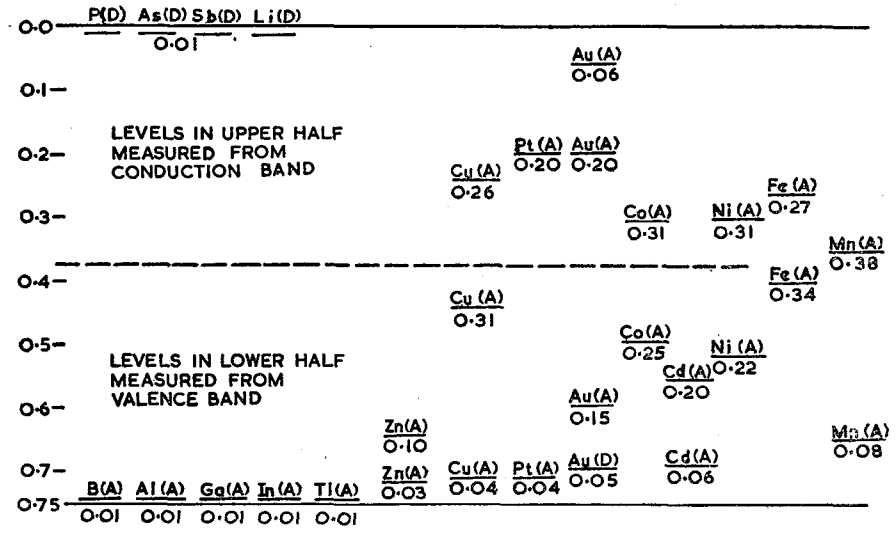


Figure 20. *Energy levels for impurities in germanium.*

Figure 20 illustrates the energy levels of some of the impurities that have been investigated. In most of these, the multiple levels are presumed to be of the multiple ionization type already established for gold. However, in many cases, careful counting of states has not been done to establish this. A few specific points of interest are those which follow.

(1) Copper is like gold in that it has three acceptor states[78], the uppermost being about 0·26 eV below the conduction band. It is possible that this level is involved in some of the properties of heat-treated and deformed germanium that have been ascribed to dislocation acceptors.

(2) Iodine may introduce donor levels that are 0·2 eV below the conduction band[82]. Evidence for this is indicated by studies on vapor-deposited germanium, although the evidence is not completely positive, and has not been confirmed. If iodine behaves like gold, for example, it should produce three donor levels, spaced at various intervals through the forbidden band.

5.9. Applications of Deep States in Germanium Studies

Ability to dope germanium with impurities having deep levels has opened a new field for germanium studies that were difficult if not impossible before the techniques outlined previously became available. A few of these are outlined below.

(1) *Dielectric constant measurements*[79]. The use of gold-doped germanium has made possible direct measurements of dielectric constant on bulk samples, using conventional bridge techniques at low frequencies. Previously such measurements were made either upon the barrier layers of P–N junctions or upon bulk samples at microwave frequencies. Because of the high resistivity of gold-doped germanium at 78° K and below (10^8 Ωcm is attainable), dielectric constant measurements were easily made on thin wafers coated with electrodes of gallium–aluminum paste. A value of $15{\cdot}8 \pm 0{\cdot}2$ was obtained, agreeing well with older values and having a somewhat smaller uncertainty.

(2) *Mobility measurements on photocarriers.* Interesting results, similar to those obtained a number of years ago by Engelhard[80] on cuprous oxide, have been obtained on gold-, cobalt-, and nickel-doped germanium[66, 69]. In these experiments, Hall measurements were made on samples illuminated by light in the fundamental absorption region. The results give information on the occurrence of trapping, and hence on the sign of the charge on the impurity centers. Thus the method serves as a valuable technique in studying unknown impurities in germanium.

Figures 21 and 22 show results for Hall mobility of N- and P-type samples of gold-doped germanium as a function of light intensity. For the N-type crystals, increased light intensity has little or no effect upon Hall mobility. This is a sign that photo-holes are removed rapidly, leaving only photo-electrons to carry the current. This in turn suggests that the centers are negatively charged, and is additional evidence that the multiple ionization model leading to formation of Au^{2-} centers in N-type gold-doped germanium is correct.

For the P-type material, on the other hand, light causes a reversal at very low intensity, again indicating that holes are trapped even in the P-type material, again presumably because the Au^{-} centers present have much larger capture cross-sections for holes than for electrons.

(3) *Measurements of surfaces and grain boundaries.* The use of deep-state doping permits the removal of part of the conductivity from parts of the sample, so that other parts may be studied. For example, gold and manganese doping has been used to freeze out the bulk conductivity at 78° K so that surface[82] and grain boundary conduction[83] could be studied. The results of studies on grain

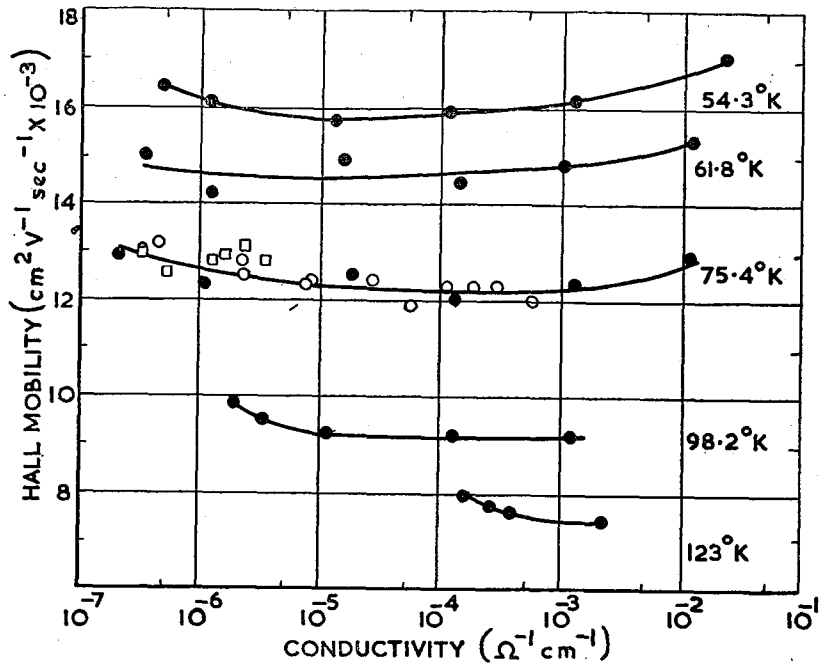


Figure 21. *Hall mobilities for photocarriers in N-type gold-doped germanium. (Tyler and Newman[81])*

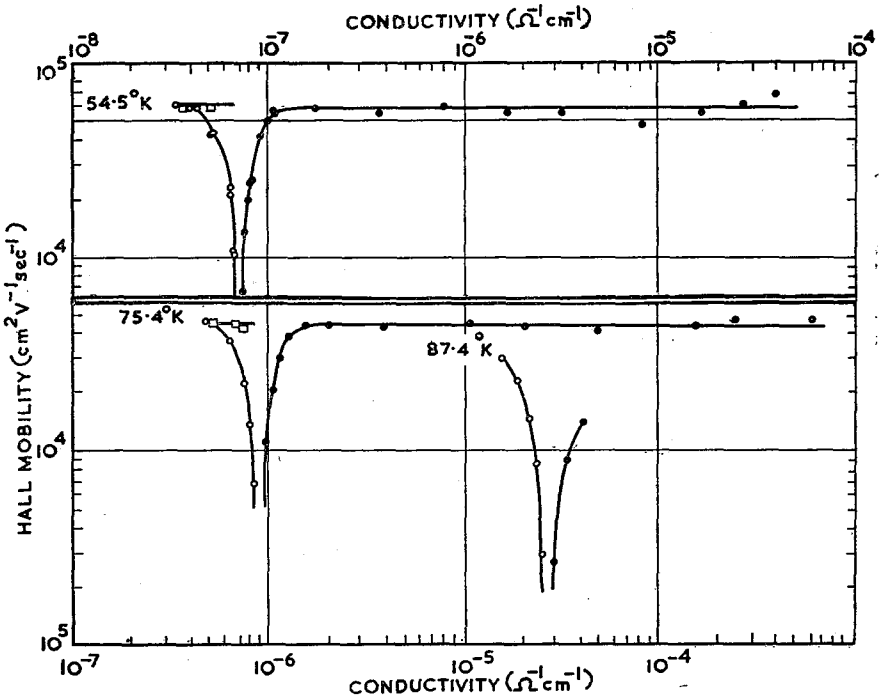
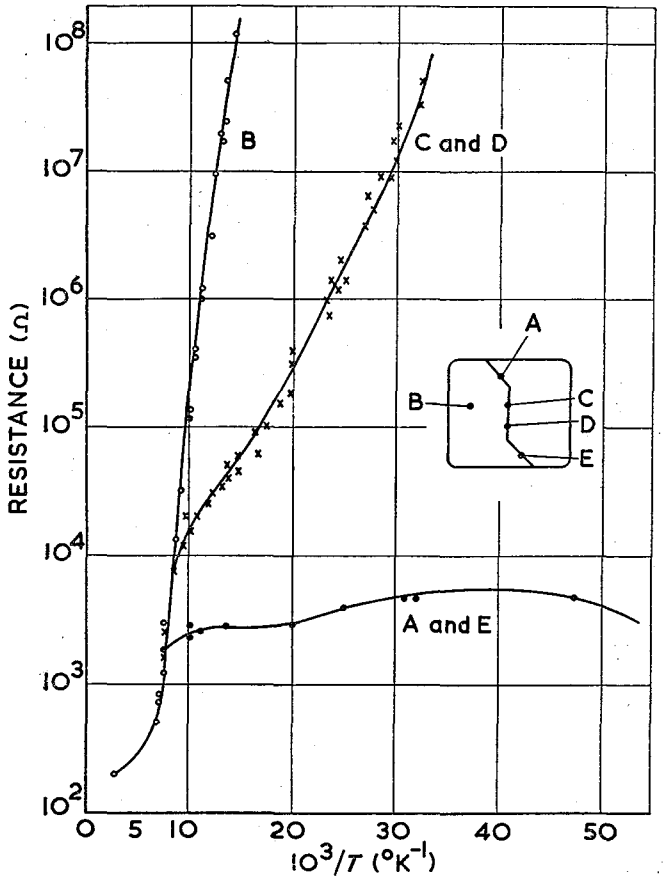


Figure 22. *Hall mobilities for photocarriers in P-type gold-doped germanium. (Tyler and Newman[81]).*

[Figs. 21 and 22 by courtesy of the Physical Review.

boundaries obtained by Tweet with gold-doped bulk crystal are shown in Figure 23. Curves are shown for the bulk conductivity as well as for the grain boundary. The ionization energy for the grain boundaries was about 0·05 eV. This is sufficiently smaller than the 0·15 eV value for the bulk conduction.



[*By courtesy of the Physical Review.*

Figure 23. *Conduction properties of gold-doped germanium containing grain boundaries.* (*Tweet*[83])

(4) *Impurity scattering from doubly charged impurity centers.* It has been suggested[71] that another confirmation of the multiple ionization model could be sought in the increased impurity scattering to be expected, assuming the latter to increase as Z^2. Although in these preliminary experiments, no notable success was met in measuring the scattering, due to doubly charged centers, more recently Tyler, Newman and Woodbury[69, 81], using photo-excitation, were able to demonstrate the effect. The photo-excitation leaves all other factors affecting mobility unchanged except for the sign of the charge on the centers, so that greater sensitivity is obtained than with the comparison of samples doped so as to contain ions of either single or double charge.

5.10. Neutral Impurities in Germanium: Carbon, Silicon and Tin

There appears to be a class of impurities that have few if any direct electrical effects upon germanium. In this class are carbon, silicon and tin, the Group IV elements.

(1) *Carbon.* Carbon has long been used as a crucible material for both zone-refining and single-crystal pulling of germanium. Chemical analysis has revealed in germanium, made in contact with carbon, traces of carbon up to a few hundredths of 1 per cent or more. In spite of this, the material may be nearly intrinsic in character and display good properties in all other respects. Thus it appears that carbon has little or no effect upon germanium.

(2) *Silicon.* Silicon–germanium alloys containing a few tenths of 1 per cent of silicon in germanium have been made[84]. Within this range of silicon content, there is very little change of germanium properties. Of course, at higher silicon contents, the lifetime and mobility drop and the band gap begins to increase significantly.

(3) *Tin.* Studies of the effects of tin upon germanium have been carried out by Trumbore[85]. This work involved careful evaluation of tin content by spectrochemical methods, and by use of radioactive tracers. Amounts of tin up to 10^{19} atoms cm^{-3} in germanium were found, although occlusions of tin in small particle forms were also found at this level of concentration. Little or no effect of tin upon lifetime or resistivity could be detected. It does appear that tin may affect heat treatment properties, presumably by promoting formation of acceptor-like vacancies. The segregation coefficient for tin in germanium was found to be 0·029.

Other neutral impurities in germanium include the noble gases, and perhaps also hydrogen and oxygen; the latter has definitely been found in germanium. More study is needed to decide the question of electrical activity[86].

For most of the other elements that have been added to germanium, electrical activity has been found. The exceptions are those elements with such small solubility that no effects were observed.

6. PHOTOCONDUCTIVITY AND ABSORPTION[3]

Photoconductive measurements have been referred to already as giving good methods of determining energy levels. Other aspects of interest are:

(1) The photoconductors made possible by doping germanium with various impurities may have interesting and important applications. We may mention the use of zinc-, gold-, and manganese-doped germanium as possibly important photoconductors. Zinc- and gold-doped P-type materials have been shown to be sensitive to 30 microns and to 10 microns respectively, to be fast in their action and to be reasonably low in noise content[87]. A thorough discussion on photoconductors of zinc-, gold- and indium-doped germanium, is given by Burstein, Picus and Sclar[3].

(2) Doped germanium shows a high degree of intrinsic photoconductivity because of the high resistivity attainable (10^8 Ωcm at 78° K for gold-doped germanium, over 10^{14} Ω cm for manganese-doped germanium).

Usually, associated with the photoconductivity is great sensitivity and transient effects arising from trapping by the impurity levels. These same phenomena

occur in other semiconducting photoconductors such as cadmium sulfide, but our knowledge of the properties of such material is incomplete because of lack of control of crystal growth, purity, *etc.* Thus, it is to be hoped that the study of deep states will shortly clarify many aspects of semiconducting photoconductors.

The multiple acceptor level picture is the key to the theory of photoconductivity in germanium. The negative charge on filled acceptor levels makes for very large differences between the cross-sections for capture of holes and of electrons. If the trapping time for one carrier is long enough, and for most of the germanium studied it seems to be always the holes that are trapped, the electrons remaining free for a long enough time to allow for the passage of considerable currents and for a large quantum yield.

Absorption studies are also of interest. Most of the interesting impurities, however, have such small solubilities that these measurements are difficult. The elements that are sufficiently soluble to be strongly absorbing show relatively little of interest and then, usually, at such long wavelengths that measurements are difficult.

7. RECOMBINATION AND TRAPPING

Recombination and trapping studies offer some of the most interesting and important applications of the deep states in germanium. Since these properties often require the presence of only 10^{12} or 10^{13} atoms, cm^{-3} they are especially useful for studying insoluble impurities such as nickel and silver.

Equation (20) gives the rate of recombination for a single center of density N_t, located at the energy ϵ_t below the conduction band, according to the theory of Hall[88] and Shockley and Read[89]:

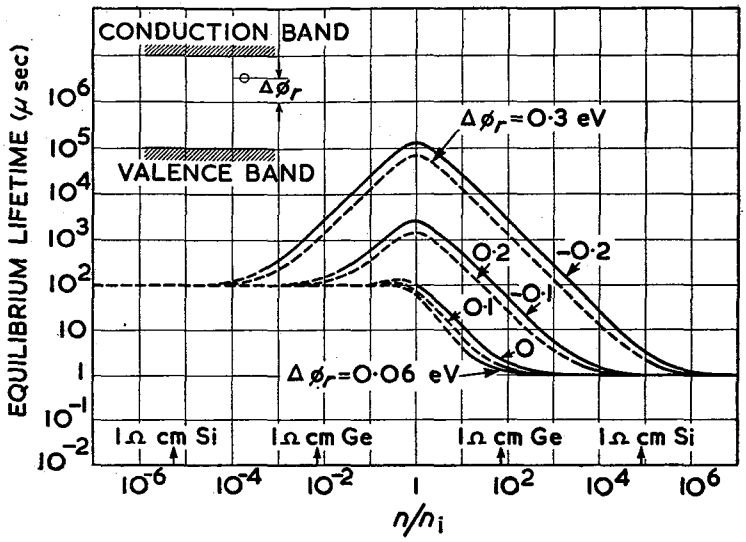
$$\tau_n = \frac{(1/C_p)(n_0 + n_1) + (1/C_n)[p_0 + p_1 + N_t(1 + (p_0/p_1)\}^{-1}]}{n_0 + p_0 + N_t\{1 + (p_0/p_1)\}^{-1}\{1 + (p_1/p_0)\}^{-1}} \quad \ldots (20)$$

Where C_p and C_n are the product of trap density N_t and probability per unit time of an individual trap capturing a hole or electron respectively, averaged over the valence or conduction band; n_0 and p_0 are the equilibrium electron and hole densities, and n_1 and p_1 are the electron and hole densities appropriate to a Fermi level located at the trap level. A similar relation applies to hole lifetime.

This theory is based upon simple statistical analysis, and the physical parameters of the problem include the capture cross-sections of the state for capture of holes and electrons.

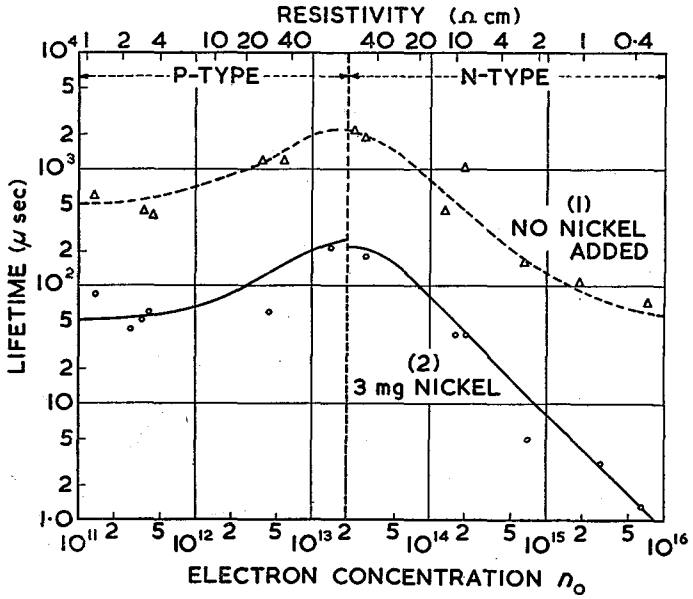
Figure 24 shows the predictions of this theory for the lifetime of germanium as a function of carrier density for several trap depths.

Figure 25 shows results of an attempt by Burton *et al.* to check the results of the theory. The points are the lifetimes measured for various carrier densities in a nickel-doped crystal, the curve the result of the Hall–Shockley–Read theory. The results are in sufficiently good agreement for the general ideas of the theory to appear to be well-confirmed. Better agreement with lifetime data is obtained if the recombination levels are assumed to lie at two discrete energy levels rather than one.



[*By courtesy of the General Electric Co.*

Figure 24. *Theoretical results of Hall–Shockley–Read theory for lifetime as function of carrier density, for states at various energies.* (*Hall*)



[*By courtesy of the Journal of Physical Chemistry.*

Figure 25. *Theoretical dependence and experimental results for lifetime in nickel-doped germanium.* (*Burton, Hull, Morin and Severiens*[64])

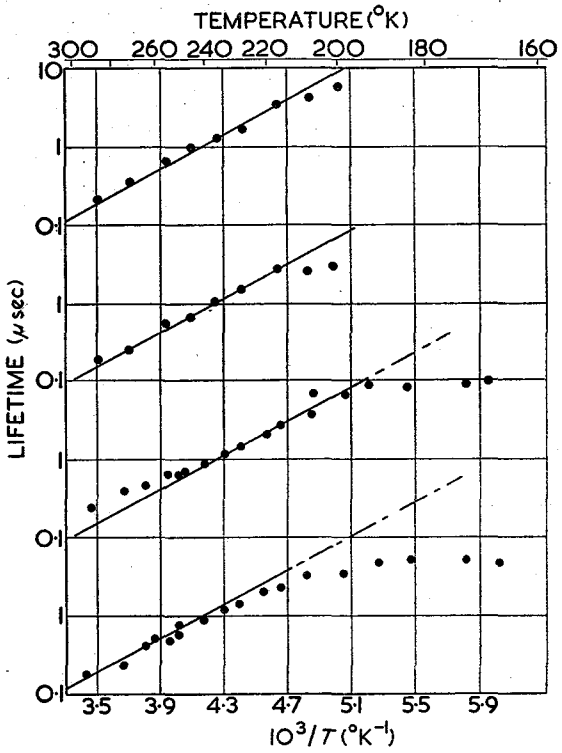


Figure 26. *Lifetime versus temperature for nickel-doped germanium.* (*Battey and Baum*[90])

[*By courtesy of the Physica Review.*

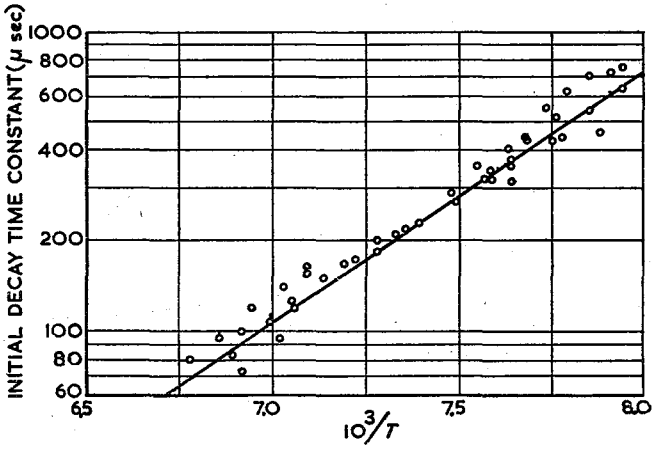


Figure 27. *Trapping in copper-doped germanium.* (*Shulman and Wyluda*[92])

[*By courtesy of the Physical Review.*

Hall[88] and Battey and Baum[90] have also studied lifetime as a function of temperature, and the last two authors have shown that the usual increase of lifetime with temperature is reversed if the Fermi level in the sample is located close to the trapping level (Figure 26). These results also agree with the Hall–Shockley–Read theory.

Trapping effects have been observed in germanium. Studies by Fan[91] and his co-workers on photoconductivity decay have shown trapping at — 78° C, due to unknown centers and, recently, Shulman and Wyluda[92] have made a study of the decay of photoconductivity in copper-doped germanium. Similar studies have also been carried out by Johnson[93] on gold-doped germanium, and by Tyler and Woodbury[94] on nickel-, manganese- and iron-doped germanium. The results of Shulman on copper are shown in Figure 27. The capture cross-sections were found to vary with temperature in such a way that the copper centers act as recombination centers at high temperatures and as trapping centers at low temperatures.

The principle objectives for trapping studies are the cross-sections for capture of carriers by the variously charged centers. Although data are as yet fragmentary, the results of Tyler and Woodbury indicate that the cross-sections of positively charged centers for electrons and singly negatively charged centers for holes is about 10^{-13} cm^2. This is, of course, enormous, and in fact much larger than the usually accepted value of the atomic area. The discrepancy arises from the coulomb attraction of the center for the carriers. This attraction creates a potential well many times the size of the ion itself.

According to the above authors, the cross-section of neutral impurity centers for electrons is about 10^{-16} cm^2, a normal value. The negatively charged centers have extremely small cross-sections for the capture of electrons, values of $\sim 10^{-22}$ cm^2 being typical. The capture cross-section of doubly charged centers for electrons is so small that it has not yet been measured, presumably, however, values in the neighborhood of 10^{-28} cm^2 might be expected.

An interesting trapping effect has been discovered by Tyler in iron-doped germanium[95]. The injection of carriers into this material at low temperatures leads to a thyratron-like effect, since the trapping removes the carriers of only one sign. As the time in the traps is long, the current continues to remain at high values even after the voltage to the injecting contact has been removed.

8. IMPURITY INTERACTION EFFECTS

Interaction effects between impurities have been observed in solubility and diffusion studies in germanium. The theory of interaction effects[96, 97, 98] is analogous to the mass action law as applied to the chemical equilibria in water solutions; for example, the ionization of a weak acid such as acetic acid is greatly inhibited by the presence of a stronger acid such as hydrochloric acid.

In this analogy, donors act as bases, since they furnish free negative carriers, acceptors as acids and the intrinsic conductivity plays the same role as the dissociation of water into hydroxyl and hydrogen ions.

In order that one impurity may affect the solubility of another, one impurity must be rapidly diffusing, such as lithium or copper, or else one must be prepared to wait long times for equilibrium. Also temperatures must not be too

high or the extrinsic conductivity will be unable to affect the concentrations to any great extent. For this reason, little effort has gone into studies of this type with germanium, and most of the work to date has been done with lithium and boron in silicon.

The nature of the effects is illustrated in the following scheme. If the mobile impurity is lithium, the immobile one boron, and if a lithium alloy is present to give a supply of lithium, then there will be the following set of reactions:

$$\begin{array}{rcl} \text{Li (alloy)} \rightleftharpoons \text{Li (germanium)} & \rightleftharpoons & \text{Li}^+ + q^- \\ & & + \\ \text{B (germanium)} & \rightleftharpoons & \text{B}^- + q^+ \\ & & \Downarrow \\ & & q^+q^- \end{array}$$

Boron contributes holes that combine with electrons from lithium and, if the intrinsic conductivity is low, shifts the equilibrium to the right, thus the solubility of lithium should be increased by the addition of boron, and this is the effect actually observed.

Although the use of an immobile acceptor and a mobile donor is the most practical case, the reverse situation, that of a mobile acceptor and an immobile donor is also possible. Similar results would be found for mobile donors and acceptors used together.

A second effect that may be of interest, and which also shows up in the study of the effects of one impurity on another, is that of ion pairing. This should lead to an increase in the solubility of the mobile donor (lithium) as a function of acceptor content, but at temperatures so high that the intrinsic conduction is great enough to make the previously discussed effect of little importance.

It has been suggested, for example, that the known effects of oxygen and water vapor on the electrical properties of germanium may arise from the reaction of the impurity atoms with dissolved oxygen to form electrically inactive oxide molecules. The whole problem of ion pairing and molecule formation in germanium is one on which much work remains to be done.

9. SUMMARY AND CONCLUSIONS

The study of specific impurity effects in germanium is one that has proved to be extraordinarily fruitful in new facts and interesting correlations that should provide the basis for much important theoretical work.

Although many impurities have been studied in germanium, it is not at all certain that the kinds of effects have yet been exhausted. Until many more impurities have been investigated, it may be presumed that still further effects involving conductivity, type, lifetime and trapping effects may be uncovered.

In addition, many of the effects discussed in the present paper are incompletely understood. Only partly understood are the electronic structure of the deep levels produced by gold, copper and other such elements, impurity banding, impurity scattering, photoconductivity, and many others. Recently acquired knowledge about the energy band structure of germanium should help provide the accurate wave functions that are needed for more exact calculations on these important problems.

Only those effects in which major, easily detectable, specific effects of the nature of the impurity center have been discussed in this paper. For many others, the effects are too small to be observed with materials or techniques now available, or else have not been studied at all. There is little doubt that all the major properties, including the simple transport phenomena, will show some dependence upon the type of impurity used, and further major developments are to be expected in this field.

REFERENCES

1. J. A. Burton. *Physica.* **20,** 845 (1954).
2. E. Burstein and P. H. Egli. *Advances in Electronics.* (Academic Press, New York, 1955.)
3. E. Burstein, G. Picus and N. Sclar. *Photoconductivity Conference.* (Wiley, New York, 1956.)
4. C. S. Fuller and J. C. Severiens. *Phys. Rev.* **96,** 21 (1954).
5. A. S. Nowick. *J. Appl. Phys.* **22,** 1182 (1951).
6. G. L. Pearson, J. D. Struthers and H. C. Theuerer. *Phys. Rev.* **75,** 344 (1949); **77,** 809 (1950).
7. W. Shockley. *Electrons and Holes in Semiconductors.* Chap. 11. (Van Nostrand, New York, 1950).
8. C. S. Hung and J. R. Gliessman. *Phys. Rev.* **96,** 1226 (1954).
9. E. M. Conwell and V. F. Weisskopf. *Phys. Rev.* **77,** 388 (1950).
10. C. Erginsoy. *Phys. Rev.* **79,** 1013 (1950).
11. See for example, E. S. Rittner. *Photoconductivity Conference.* (Wiley, New York, 1956).
12. G. H. Morrison and J. F. Cosgrove. *Anal. Ch.* **28,** 320 (1956).
13. N. B. Hannay and A. J. Ahearn. *Anal. Ch.* **26,** 1056 (1954).
14. C. L. Luke and M. E. Campbell. *Anal. Ch.* **25,** 1588 (1953); **27,** 1150 (1955).
15. P. H. Keck, A. L. MacDonald and J. W. Mellichamp. *Anal. Ch.* **28,** 995 (1956).
16. W. G. Pfann. *J. Metals,* **4,** 861 (1952).
17. J. A. Burton, E. D. Kolb, W. P. Slichter and J. D. Struthers. *J. Chem. Phys.* **21,** 1991 (1953).
18. R. N. Hall. *J. Phys. Chem.* **57,** 836 (1953).
19. C. D. Thurmond and J. D. Struthers. *J. Phys. Chem.* **57,** 831 (1953).
20. R. J. Hodgkinson. *Physica.* **20,** 1001 (1954).
21. R. N. Hall (unpublished).
22. R. N. Hall. *Phys. Rev.* **88,** 139 (1952).
23. R. L. Longini and R. F. Greene. *Phys. Rev.* **102,** 992 (1956).
24. W. C. Dunlap, Jr. *Phys. Rev.* **86,** 615 (1952).
25. R. N. Hall and W. C. Dunlap, Jr. *Phys. Rev.* **80,** 467 (1950).
26. W. C. Dunlap, Jr. *Phys. Rev.* **94,** 1531 (1954).
27. W. C. Dunlap, Jr. and D. E. Brown. *Phys. Rev.* **86,** 417 (1952).
28. C. S. Fuller. *Phys. Rev.* **86,** 136 (1952).
29. R. L. Hoffman. *Atom Movements.* (Society for Metals, 1952).
30. W. C. Dunlap, Jr., H. V. Bohm and H. P. Mahon, Jr. *Phys. Rev.* **96,** 822 (1954).
31. K. B. McAfee, W. Shockley and M. Sparks. *Phys. Rev.* **86,** 137 (1952).
32. N. F. Mott and R. W. Gurney. *Electronic Processes in Ionic Crystals.* (Clarendon Press, Oxford, 1939).
33. J. Bardeen (private communication).
34. C. S. Fuller, H. C. Theuerer and W. van Roosbroeck. *Phys. Rev.* **85,** 678 (1952).
35. C. S. Fuller, J. D. Struthers, J. A. Ditzenberger and K. B. Wolfstirn. *Phys. Rev.* **93,** 1182 (1954).
36. C. S. Fuller and J. D. Struthers. *Phys. Rev.* **87,** 526 (1952).
37. F. van der Maesen and J. A. Brenkman. *J. Electrochem. Soc.* **102,** 229 (1955).
38. J. C. Severiens and C. S. Fuller. *Phys. Rev.* **92,** 1322 (1953).
39. C. S. Fuller and J. A. Ditzenberger. *Phys. Rev.* **91,** 193 (1953).
40. A. G. Tweet and C. J. Gallagher. *Phys. Rev.* **103,** 828 (1956).
41. F. C. Frank and D. Turnbull. *Phys. Rev.* **104,** 617 (1956).
42. W. C. Dash. *Phys. Rev.* (in the press).
43. H. C. Theuerer and J. H. Scaff. *J. Metals.* **3,** 59 (1951).
44. W. P. Slichter and E. D. Kolb. *Phys. Rev.* **87,** 527 (1952).
45. S. Mayburg. *Phys. Rev.* **95,** 38 (1954).

46. R. A. Logan. *Phys. Rev.* **91,** 757 (1953).
47. H. Letaw, *Phys. Rev.* (to be published).
48. G. T. Jacobi and W. C. Dunlap, Jr. *Phys. Rev.* **78,** 349 (1950).
49. H. A. Bethe, *Radiation Laboratory Report* 43–12. (Department of Commerce Washington, D.C. Dec. 1942).
50. P. P. Debye and E. M. Conwell. *Phys. Rev.* **93,** 693 (1954).
51. T. H. Geballe and F. J. Morin. *Phys. Rev.* **95,** 1085 (1954).
52. E. Billig (private communication).
53. W. Kohn and J. M. Luttinger. *Phys. Rev.* **96,** 529 (1954).
54. W. Kohn and J. M. Luttinger. *Phys. Rev.* **98,** 915 (1955).
55. W. Kohn and D. Schechter. *Phys. Rev.* **99,** 1903 (1955).
56. M. Lampert. *Phys. Rev.* **97,** 352 (1955).
57. C. Kittel and A. H. Mitchell. *Phys. Rev.* **96,** 1488 (1954).
58. C. Kittel and A. H. Mitchell (to be published).
59. W. C. Dunlap, Jr. *Phys. Rev.* **85,** 945 (1952).
60. F. J. Morin and J. P. Maita. *Phys. Rev.* **90,** 337 (1953).
61. W. C. Dunlap, Jr. *Phys. Rev.* **96,** 40 (1954).
62. J. F. Battey and R. M. Baum. *Phys. Rev.* **94,** 1393 (1954).
63. H. H. Woodbury and W. W. Tyler. *Phys. Rev.* **100,** 1259 (1955).
64. J. A. Burton, G. W. Hull, F. J. Morin and J. C. Severiens. *Phys. Chem.* **57,** 853 (1953).
65. F. van der Maesen, P. Penning and A. van Wieringen. *Philips Res. Rep.* **8,** 241 (1953).
66. W. W. Tyler, R. Newman and H. H. Woodbury. *Phys. Rev.* **98,** 461 (1955).
67. W. W. Tyler and H. H. Woodbury. *Phys. Rev.* **96,** 874 (1954).
68. R. Newman and W. W. Tyler. *Phys. Rev.* **96,** 882 (1954).
69. H. H. Woodbury and W. W. Tyler. *Phys. Rev.* **100,** 659 (1956).
70. W. C. Dunlap, Jr. *Phys. Rev.* **91,** 1282 (1953).
71. W. C. Dunlap, Jr. *Phys. Rev.* **97,** 614 (1955).
72. W. C. Dunlap, Jr. *Bull. Amer. Phys. Soc.* **30,** No. 2, Paper F2 (1955).
73. W. C. Dunlap, Jr. *Phys. Rev.* **100,** 1629 (1955).
74. G. A. Morton, E. E. Hahn and M. L. Schultz. *Photoconductivity Conference* (New York, Wiley, 1956).
75. W. W. Tyler. *Phys. Rev.* (in the press).
76. R. Newman. *Phys. Rev.* **94,** 278 (1954).
77. H. H. Woodbury and W. W. Tyler. *Bull. Amer. Phys. Soc.* Series II, **1,** No. 3, Paper L9 (1956).
78. H. H. Woodbury and W. W. Tyler. *Phys. Rev.* (to be published).
79. W. C. Dunlap, Jr. and R. L. Watters. *Phys. Rev.* **92,** 1396 (1953).
80. E. Engelhard. *Ann. Phys., Lpz.* **17,** 539 (1933).
81. W. W. Tyler and R. Newman. *Phys. Rev.* **98,** 961 (1955).
82. W. C. Dunlap, Jr., J. C. Marinace and R. P. Ruth. *Bull. Amer. Phys. Soc.* Series II, **1,** No. 1, Paper N8, (1956).
83. A. G. Tweet. *Phys. Rev.* **99,** 1182 (1955).
84. E. R. Johnson and S. M. Christian. *Phys. Rev.* **95,** 560 (1954).
85. F. A. Trumbore (private communication).
86. W. Kaiser, P. H. Keck and C. F. Lange. *Phys. Rev.* **101,** 1264 (1956).
87. E. Burstein (private communication) (zinc and gold).
88. R. N. Hall. *Phys. Rev.* **87,** 387 (1952).
89. W. Shockley and W. T. Read, Jnr. *Phys. Rev.* **87,** 835 (1952).
90. J. F. Battey and R. M. Baum. *Phys. Rev.* **100,** 1634 (1955).
91. H. Y. Fan, D. Navon and H. Gebbie. *Physica.* **20,** 855 (1954).
92. R. G. Shulman and B. J. Wayluda. *Phys. Rev.* **102,** 1455 (1956).
93. L. F. Johnson. *Bull. Amer. Phys. Soc.* Series II, **1,** No. 3, Paper L8 (1956).
94. W. W. Tyler and H. H. Woodbury. *Bull. Amer. Phys. Soc.* Series II, **1,** No. 3, Paper L7 (1956).
95. W. W. Tyler. *Phys. Rev.* **96,** 226 (1954).
96. H. Reiss. *J. Chem. Phys.* **21,** 1209 (1953).
97. H. Reiss and C. S. Fuller. *Phys. Rev.* **97,** 559 (1955).
98. H. Reiss, C. S. Fuller and F. J. Morin. *Bell Syst. Tech. J.* **35,** 535 (1956).

HIGH ELECTRIC FIELD EFFECTS IN SEMICONDUCTORS

J. B. GUNN, B.A.

University of British Columbia, Vancouver, Canada
formerly of
R.R.E., Malvern, U.K.

MS. received 16 *October*, 1956

HIGH ELECTRIC FIELD EFFECTS IN SEMICONDUCTORS

1. INTRODUCTION

When an electric field is applied to a homogeneous semiconductor specimen which is in thermal equilibrium, this equilibrium is lost, and in the most general case the interactions between the field, the carriers and the scattering mechanisms become very complex. In the ordinary theory of carrier mobility a drastic simplification is made by assuming that, while the carriers may have a directed drift velocity superimposed upon their random thermal motions, their total energy remains unchanged by the application of the field. The carriers may then be said to be in equilibrium with the scattering mechanisms, and their 'temperature' equal to that of the lattice. The theory shows that, under these conditions, the drift velocity, and hence the current density must be proportional to the field. The effects discussed in this paper will be those which result from the fact that this assumption is strictly justified only for vanishingly small values of the field, and that for finite values the electron temperature exceeds that of the lattice.

Because the scattering is, in general, dependent on the carrier temperature, it follows that a change in mobility should be observed when the carrier temperature has been raised by the field. This prediction serves to introduce the first class of high-field effect, in which a departure from Ohm's law results from changes in carrier mobility while the carrier density remains constant. A second type of departure occurs when the field becomes so high that additional carrier pairs are generated by the ionization of lattice bonds. Subsidiary effects, *e.g.* the emission of light, may result from the fact that at the very highest fields the energy distribution of the carriers extends to very high values. The rate at which the carriers gain energy from the field, and hence their excess temperature, is proportional to their mobility, other things being equal. For this reason, experimental work has in the past been limited to silicon and germanium, which can be prepared in highly purified single-crystal form and have the necessary high mobilities.

2. THEORY

2.1. Mobility changes

The changes in mobility which occur in a high electric field depend on the attainment of a changed equilibrium distribution of energies among the charge carriers. The new distribution depends, in turn, upon the rate at which the energy of the carriers is dissipated by collisions with the scattering mechanism, so a study of the collision process must be the starting point for a theory of the mobility changes. The first such theory applicable to semiconductors appears to have been given in 1934 by Landau and Kompanejez[1]. More recent work includes that by Guth and Mayerhöfer[2] on metals, and by Seitz[3], Shockley[4],

Conwell[5,6], Yamashita and Watanabe[7], and Shibuya[8] on non-polar semiconductors. The treatment we shall give is an elementary version of that given by Shockley and Conwell.

When an electron is scattered by a lattice vibration, both the energy ϵ and the crystal momentum P must be conserved. For an electron with mass m and velocity v we have $P_n = mv$ and $\epsilon_n = P_n v/2$ while for a phonon of wavelength λ and phase velocity c, $P_\phi = h/\lambda$ and $\epsilon_\phi = cP_\phi$. Because of the conservation rules, a collision which changes P_n by δP_n results in the emission or absorption of a phonon with $P_\phi = \delta P_n$, and hence in a change of electron energy $\delta\epsilon_n = \epsilon_\phi = cP_\phi = c\delta P_n$. The biggest possible value of δP_n is $-2P_n$; hence the largest possible energy change is $\delta\epsilon_n = -2cP_n = -4c\epsilon_n/v$. In silicon and germanium at room temperatures, the thermal velocity v is about 20 times the value of c for the acoustic modes which are important in scattering, so the electron can at most lose only a small fraction of its energy in a single collision. The *average* energy change in a collision is, however, even smaller. It is[4]:

$$\langle\delta\epsilon_n\rangle = \frac{2mc^2}{kT}(2kT - \epsilon_n) \qquad \ldots (1)$$

where T is the temperature of the phonon distribution. This expression, which is valid only when ϵ_n is not too large compared with $2kT$, results from consideration of the relative probabilities of phonon absorption and phonon emission in a collision.

In order to simplify the mathematics, and to make the underlying physical principles clearer, we shall now introduce the assumption that all the electrons have the same random speed v. This procedure leads to results which are qualitatively correct and differ only by small factors from those obtained by averaging over the correct distribution. Under these conditions the mobility is given[9, 18] by:

$$\mu = \frac{q\tau}{m}\left\{1 + \frac{1}{3}\frac{\mathrm{d}(\ln\tau)}{\mathrm{d}(\ln v)}\right\} \qquad \ldots (2)$$

where τ is the mean free time between collisions. In general, both ionized impurity atoms and phonons will contribute to the scattering process, with characteristic mean free times τ_I and τ_L. We have[9]:

$$\tau_I = Bv^3 \qquad \ldots (3)$$

$$\tau_L = l/v \qquad \ldots (4)$$

$$\frac{1}{\tau} = \frac{1}{\tau_I} + \frac{1}{\tau_L} \qquad \ldots (5)$$

where B is a constant depending on the impurity density, and l is the constant mean free path for lattice scattering.

On substituting equations (3), (4) and (5) in equation (2), we obtain:

$$\mu = \frac{2q}{3m}\tau_L\frac{\{3(\tau_L/\tau_I)+1\}}{\{(\tau_L/\tau_I)+1\}^2} \qquad \ldots (6)$$

Denoting by the subscript 0 quantities relating to thermal equilibrium conditions, we have, from equation (1):

$$v_0 = (4kT/m)^{1/2} \qquad \ldots (7)$$

We introduce for convenience the parameters $x = v/v_0$, $a = \tau_{L0}/\tau_{I0}$, in terms of which:

$$\frac{\tau_L}{\tau_I} = ax^{-4} \qquad \ldots (8)$$

Hence:

$$\mu = \frac{\mu_{L0}}{x}\,\frac{(3a\,x^{-4}+1)}{(a\,x^{-4}+1)^2} \qquad \ldots (9)$$

where:

$$\mu_{L0} = \frac{2q\tau_{L0}}{3m} \qquad \ldots (10)$$

is the low-field mobility for pure lattice scattering.

In calculating the energy losses, we have only to consider the electron–phonon interactions, since the collisions with impurity ions are highly elastic.

The rate of energy loss per unit time by collision is thus:

$$\begin{aligned}\left(\frac{d\epsilon_n}{dt}\right)_{coll.} &= \langle\delta\epsilon_n\rangle/\tau_L \\ &= \langle\delta\epsilon_n\rangle x/\tau_{L0} = \langle\delta\epsilon_n\rangle\frac{2qx}{3m\mu_{L0}} \\ &= \frac{2mc^2}{kT}(2kT-\epsilon_n).\,\frac{2qx}{3m\mu_{L0}} \\ &= \frac{2mc^2}{\frac{1}{4}mv_0^{\,2}}(\tfrac{1}{2}mv_0^{\,2}-\tfrac{1}{2}mv^2)\,\frac{2qx}{3m\mu_{L0}} \\ &= \frac{8c^2q}{3\mu_{L0}}\,x\,(1-x^2) \qquad \ldots (11)\end{aligned}$$

where use has been made of equations (1), (4), (7) and (10). The rate of energy input from the field is simply:

$$\begin{aligned}\left(\frac{d\epsilon_n}{dt}\right)_{field} &= qv_dE = q\mu E^2 \\ &= q\mu_{L0}E^2\cdot\frac{(3ax^{-4}+1)}{x(ax^{-4}+1)^2} \qquad \ldots (12)\end{aligned}$$

In the steady state we have, evidently:

$$\left(\frac{d\epsilon_n}{dt}\right)_{coll.} + \left(\frac{d\epsilon_n}{dt}\right)_{field} = 0 \qquad \ldots (13)$$

and so the relation between x and E is:

$$\frac{3\mu_{L0}^{\,2}}{8c^2}E^2 = x^2(x^2-1)\frac{(ax^{-4}+1)^2}{(3ax^{-4}+1)} \qquad \ldots (14)$$

which, together with equation (9), gives an implicit relation between E and μ which is valid over a wide range of fields for any ratio of impurity to lattice scattering.

In order to clarify this relation between E and μ, it is convenient to expand the equations for certain special cases. First let E be small, so that $x^2 - 1 \ll 1$. Then equation (14) becomes:

$$x^2 = 1 + \frac{3\,(3a+1)\,\mu_{L0}{}^2}{8\,(a+1)^2\,c^2}\,E^2 \qquad \ldots\ (15)$$

and hence, from equation (9):

$$\mu = \mu_0 \left\{1 + \frac{3(9a+1)(a-1)}{16(a+1)^3}\frac{\mu_{L0}{}^2}{c^2}E^2\right\}$$

$$= \mu_0\,(1 + \beta E^2), \qquad \ldots\ (16)$$

where

$$\mu_0 = \mu_{L0}\,\frac{3a+1}{(a+1)^2}$$

is the actual mobility in low fields, and

$$\beta = \frac{3(9a+1)(a-1)}{16(a+1)^3}\cdot\frac{\mu_{L0}{}^2}{c^2}$$

is a function only of the temperature and composition of the specimen. Thus it appears that for $a < 1$ (lattice scattering dominant), β is negative and the mobility should depend quadratically on E, decreasing as E increases. Conversely, for $a > 1$ (impurity scattering) the mobility should at first increase with E. For $a = 1$ there should be no first-order dependence of μ on E^2. With pure lattice scattering ($a = 0$) we have:

$$\beta = -\frac{3\mu_{L0}{}^2}{16c^2}$$

which is to be compared with the more accurate value:

$$\beta = -\frac{3\pi\mu_{L0}{}^2}{64c^2}$$

obtained[4] by considering a Maxwellian, rather than a single-valued, velocity distribution. For impurity scattering, the largest positive value of β occurs when $a = 3{\cdot}6$; $(9a+1)(a-1)/(a+1)^3$ then has its maximum value of 0·89. For very large a, β tends to $27\mu_{L0}{}^2/16ac^2$. Thus, while the value of $|\beta|$ is always decreased by the simultaneous presence of impurity and lattice scattering, under certain conditions the positive mobility deviations may be nearly as large as the negative deviations characteristic of pure lattice scattering.

Consider now the case when the field E is very large, so that $(3a)^{1/2} \ll x^2 \gg 1$. Then equation (14) becomes:

$$\frac{3\mu_{L0}{}^2}{8c^2}\,E^2 = x^4 \qquad \ldots\ (17)$$

whence, and from equation (9), the mobility is:

$$\mu = \left(\frac{8}{3}\right)^{1/4}\left(\frac{c\mu_{L0}}{E}\right)^{1/2} \qquad \ldots\ (18)$$

so, under these conditions, the drift velocity μE varies as $E^{1/2}$ independently of the amount of impurity scattering at low fields. Physically, this arises because the electron temperature is now so high that impurity scattering is unimportant. The numerical constant $(8/3)^{1/4}$ again differs, by the omission of a factor $(4/\pi)^{1/4}$, from that given by the Maxwellian approximation[4], and is about 5·5 times larger than that quoted in the similar equation of Landau and Kompanejez[1].

Finally, when $a \ll 1$, equations (9) and (14) may be combined, giving μ as an explicit function of E:

$$\mu = \mu_{L0} \cdot \frac{2^{1/2}}{[1 + \{1 + \frac{3}{2}(\mu_{L0}^2/c^2)E^2\}^{1/2}]^{1/2}} \quad \ldots (19)$$

Equations (18) and (19) are not a complete description of the mobility variation in extremely high fields because, in calculating the rate of energy loss, equation (11), scattering by the optical modes of the lattice vibrations was deliberately excluded. The optical modes which have a small enough value of P_ϕ to be effective in electron scattering are those with the longest wavelengths. According to the measurements of Collins and Fan[10], which are in close agreement with the theoretical calculations of Hsieh[11], these modes have frequencies ν_0 of about 1·82 and 1·03 $\times$ 10^{13} $\sec^{-1}$ in silicon and germanium respectively. The characteristic temperatures $\theta_0 = h\nu_0/k$ are thus about 825 and 495° K, so that at and below room temperature the optical phonons are not appreciably excited*, and few electrons have sufficient energy to be scattered by phonon emission. The optical modes do not, therefore, contribute to the normal scattering process. In addition, the scattering probability for an electron with energy just equal to that of the optical modes may be very small. In the simplest theory, the tetrahedral symmetry of the diamond lattice permits no first-order change in the energy of the band edges for distortions of the type occurring in the optical modes i.e., for displacements of one sub-lattice with respect to the other[4]. There may, however, be a second or higher order change, and there may be a first-order change if the polarizability of the ions is taken into account[12], or if the energy bands do not have spherical symmetry[4]. The polarizability would appear to play a similar part in the infra-red absorption of silicon and germanium[10].

Because of the uncertainty of the theoretical picture, we shall somewhat arbitrarily assume the following relation between v and τ_0, the mean free time for scattering by an optical mode:

$$\left.\begin{aligned} \frac{1}{\tau_0} &= 0, \; v \leqslant v_0 \\ \frac{1}{\tau_0} &= A(v - v_0)^b, \; v > v_0 \end{aligned}\right\} \quad \ldots (20)$$

where $v_0 = (2h\nu_0/m)^{1/2}$ is the velocity of an electron with energy equal to that of the optical modes, and A, b are arbitrary constants. The temperature, and

* The parameters θ_0 should not be confused with the Debye temperatures θ_D which are about 670 and 370° K. θ_0 and θ_D are equal only for the simple ν^2 lattice vibration spectrum of Debye; for the diamond lattice, which has a very peaky vibration spectrum, the two need not be closely related.

hence the excitation of the optical modes, is supposedly so low ($\ll \theta_0$) that the absorption of an optical phonon is improbable. We shall also neglect acoustical scattering; this will be discussed later. It follows from the last two assumptions that an electron is scattered only when $v > v_0$, and that its initial velocity contains no component due to thermal motion. The velocity is thus always parallel to E, and the number $f(v)\mathrm{d}v$ of electrons having velocities between v and $v + \mathrm{d}v$ may be calculated as follows. In the notation of equation (13) we have, in the steady state:

$$\left(\frac{\mathrm{d}f}{\mathrm{d}t}\right)_{\text{coll}} + \left(\frac{\mathrm{d}f}{\mathrm{d}t}\right)_{\text{field}} = 0 \qquad \ldots (21)$$

where

$$\left.\begin{aligned}\left(\frac{\mathrm{d}f}{\mathrm{d}t}\right)_{\text{coll}} &= -\frac{f(v)}{\tau_0},\ v > v_0\\ \left(\frac{\mathrm{d}f}{\mathrm{d}t}\right)_{\text{coll}} &= +\frac{f(v+v_1)}{\tau_0},\ v \leqslant v_0\end{aligned}\right\} \qquad \ldots (22)$$

$$\begin{aligned}\left(\frac{\mathrm{d}f}{\mathrm{d}t}\right)_{\text{field}} &= -\frac{\mathrm{d}v}{\mathrm{d}t}\cdot\frac{\mathrm{d}f}{\mathrm{d}v}\\ &= -\frac{qE}{m}\cdot\frac{\mathrm{d}f}{\mathrm{d}v} \qquad \ldots (23)\end{aligned}$$

Substituting equations (20), (22) and (23) in equation (21) and integrating, we have, for $v > v_0$:

$$f_1 = B\exp\{-g(v - v_0)^{b+1}\} \qquad \ldots (24)$$

where $g = \dfrac{mA}{(b+1)qE}$ has been written for brevity.

For $v \leqslant v_0$:

$$f_2 = \frac{BAm}{qE}\int (v - v_0 + v_1)^b \exp\{-g(v - v_0 + v_1)^{b+1}\}\mathrm{d}v \qquad \ldots (25)$$

Here B is a constant. v_1 given by:

$$(v + v_1)^2 = v^2 + v_0^2 \qquad \ldots (26)$$

is the loss of velocity, and v the velocity after scattering of an electron which emits a phonon with $hv_0 = \frac{1}{2}mv_0^2$. Since the integrand in equation (25) is small except near $v = 0$, we may put $v_1 = v_0$ without serious error: the equation then becomes:

$$f_2 = C - B\exp(-gv^{b+1}) \qquad \ldots (27)$$

with C another constant. The constants B and C are determined by the relations $f_1(v_0) = f_2(v_0)$, and $\int_0^\infty f(v)dv = N$, where N is the total number of electrons. If

$$\frac{mAv_0^{b+1}}{(b+1)qE} \gg 1$$

as is inevitably the case, these give:

$$B = C = \frac{N}{v_0}.$$

The average drift velocity in the direction of E associated with the distribution function f is simply:

$$v_d = \frac{\int_0^\infty v f(v)\,dv}{\int_0^\infty f(v)\,dv} = \frac{1}{N}\int_0^\infty v f(v)\,dv \qquad \ldots (28)$$

Since the second term in equation (27) differs from zero only near $v = 0$, the contribution to v_d from f_2 is approximately:

$$\frac{1}{N}\int_0^{v_0} v f_2(v)dv = \frac{1}{v_0}\cdot \tfrac{1}{2}v_0^2 = \tfrac{1}{2}v_0 \qquad \ldots (29)$$

The corresponding contribution from f_1 is:

$$\frac{1}{N}\int_{v_0}^\infty v f_1(v) = \frac{1}{v_0}\int_{v_0}^\infty v \exp\left\{-g(v - v_0)^{b+1}\right\} dv \qquad \ldots (30)$$

For $b \neq 0$, the integral may be transformed by the substitution $u = (v - v_0)^{b+1}$ into the form:

$$\frac{g^{-1/(b+1)}v_0}{b+1}\int_0^\infty u^{\{1/(b+1)\}-1}\exp(-u)du + \frac{g^{-2/(b+1)}}{b+1}\int_0^\infty u^{\{2/(b+1)\}-1}\exp(-u)du$$
$$= \frac{g^{-1/(b+1)}v_0}{b+1}\Gamma\left(\frac{1}{b+1}\right) + \frac{g^{-2/(b+1)}}{b+1}\Gamma\left(\frac{2}{b+1}\right) \qquad \ldots (31)$$

Hence the drift velocity is:

$$v_d = \tfrac{1}{2}v_0 + \frac{\Gamma\{1/(b+1)\}}{b+1}\cdot\left\{\frac{(b+1)qE}{mA}\right\}^{1/(b+1)}$$
$$+ \frac{\Gamma\{2/(b+1)\}}{(b+1)v_0}\cdot\left\{\frac{(b+1)qE}{mA}\right\}^{2/(b+1)} \qquad \ldots (32)$$

When v_d is not much larger than $1/2v_0$, the third term may be neglected.

Physically, equation (32) represents the average velocity of an electron which is uniformly accelerated without scattering from a low velocity up to v_0, after which it becomes liable to be scattered by an optical phonon, with a velocity-dependent probability $1/\tau_0$ of losing the energy $\frac{1}{2}mv_0^2$. It is clear that, if the electron were scattered immediately on reaching v_0, v_d would be equal to $\frac{1}{2}v_0$ and independent of the field. The second and third terms on the R.H.S. thus represent the deviation from constant drift velocity resulting from the finite probability of optical scattering. The assumption that a separation can be made of the ranges of field strength over which the optical and acoustical scattering processes are effective is open to considerable objection; in practice, there must be a smooth transition from one process to the other, and the range over which both are effective may be so large as to obscure the $E^{1/2}$ behaviour predicted by equation (18).

2.2. Avalanche Ionization

Even though the probability $1/\tau_0$ for scattering by the optical modes may be large for $v \gg v_0$, there must be a few electrons having energies much greater than hv_0. In particular, a few electrons must possess the energy ϵ_0 required to ionize the neutral atoms of the lattice. For equal masses of holes and electrons, and spherically symmetrical bands, this energy is:

$$\epsilon_0 = \frac{3}{2}\epsilon_G \qquad \ldots (33)$$

where ϵ_G is the energy gap. The factor 3/2 arises because the energy and momentum of the single incident particle must be shared among three after an ionizing collision. It does not follow, of course, that an electron having $\epsilon = \epsilon_0$ will make ionizing collisions. The cross-section for this energy is zero, but should increase rapidly with ϵ above this point. Thus the 'effective' threshold energy may be rather greater than ϵ_0.

The ionization process is described by the ionization coefficient α. This is defined as the number of ionizing collisions per cm path length made by a single particle. The calculation of α is somewhat troublesome, and can be completed only numerically, as has been done in a classic paper by Wolff [13]. No semi-quantitative elementary theory, of the type we have used in discussing the mobility variations, seems possible, so we shall quote only the results of the correct theory. Wolff shows that, provided consideration is restricted to the range $hv_0 \ll \epsilon \leqslant \epsilon_0$, and the scattering process is assumed to be described by a constant mean free path l_0, the electron distribution function n_0 is nearly isotropic, and varies with energy approximately as:

$$n_0(\epsilon)\mathrm{d}\epsilon = \text{const.} \times \exp\left(-\frac{hv_0}{q^2E^2l_0^2}\epsilon\right)\mathrm{d}\epsilon \qquad \ldots (34)$$

For $\epsilon > \epsilon_0$, a different distribution function n_0' must be used which takes account of ionizing collisions, and the two distributions are matched at ϵ_0. Because $n_0(\epsilon_0)$ is such a rapidly varying function of E, its variation effectively controls $\int_{\epsilon_0}^{\infty} n_0' \, \mathrm{d}\epsilon$, to which α is very roughly proportional.

Thus we have, approximately:

$$\alpha = \text{const.} \times \exp\left(-\frac{hv_0\,\epsilon_0}{q^2l_0^2} \cdot \frac{1}{E^2}\right) \qquad \ldots (35)$$

Curves of accurate solutions of the problem have been given by Wolff [13] and Miller[14]. Even these may represent an over-simplification, as there is evidence to suggest that, at least in germanium, there is a difference between the value of α for holes and electrons, presumably as a consequence of the complex band structure.

The important feature of this type of ionization, which is absent from such phenomena as the internal field emission of Zener[15], is that the process is cumulative. For example, an ionizing collision by an electron at a certain point produces, in addition to the secondary electron, a hole which is accelerated by the field in the opposite direction to the electrons. The hole is capable of ionizing

and the electrons produced by it may produce further ionizing collisions at the original point. Thus, at a certain field strength, a form of positive feedback may take place, and the ionization process may become self-sustaining[16].

2.3. Avalanche Breakdown in a P-N Junction

Avalanche ionization is of considerable technological importance, as it is this process which eventually limits the magnitude of the reverse bias voltage which can be applied to a P–N junction. A theory of the avalanche breakdown has been given by McKay[16], and, in a more advanced form, by Miller[14], who has taken account of a possible difference between the values of α for holes and electrons.

The differential equation describing the growth of current density in a one-dimensional system is simply:

$$-\frac{\mathrm{d}j_\mathrm{p}}{\mathrm{d}x} = \frac{\mathrm{d}j_\mathrm{n}}{\mathrm{d}x} = \alpha_\mathrm{n} j_\mathrm{n} + \alpha_\mathrm{p} j_\mathrm{p} \qquad \ldots (36)$$

where $j_\mathrm{p}, j_\mathrm{n}$ are the hole and electron current densities, and α_p, α_n the ionization coefficients, respectively. If, as we shall assume, $\alpha_\mathrm{p} = \alpha_\mathrm{n} = \alpha(E)$, this becomes:

$$\frac{\mathrm{d}j_\mathrm{n}}{\mathrm{d}x} = J\,\alpha(E) \qquad \ldots (37)$$

where $J = j_\mathrm{p} + j_\mathrm{n}$ is the total current. Although the field E will be large, the drift velocity of the carriers may, as we have seen, be quite limited, and in principle the space charge of the carriers should be taken into account in calculating E. In practice, however, this space charge can be neglected by comparison with that due to ionized impurities in the P–N junction, and E can be taken as known.

Whatever the assumptions, equation (37) integrates at once to give:

$$j_\mathrm{n} = J \int \alpha\,\mathrm{d}x + \text{const.} \qquad \ldots (38)$$

Further progress can be made only when particular boundary conditions are assumed, although this is not always made clear in the literature. We shall assume that an electron current $j_{\mathrm{n}0}$ (> 0) enters the ionization region at $x = 0$, and that some field distribution $E(x)$ extends from $x = 0$ to $x = X$, outside which range $E = 0$. We shall also assume that at most a negligible hole current enters at $x = X$, *i.e.*, that $j_\mathrm{n}(X) = J$. Under these conditions it is readily seen that equation (38) becomes:

$$1 - \frac{j_{\mathrm{n}0}}{J} \equiv 1 - M^{-1} = \int_0^X \alpha\,\mathrm{d}x \qquad \ldots (39)$$

Since E is assumed known, $\alpha(x)$ can in principle be evaluated, and so the multiplication factor M can be calculated. A similar equation holds for the multiplication of a hole current entering at $x = X$, provided that $j_{\mathrm{n}0}$ can be neglected. A sufficient condition for the validity of both equations is $M \gg 1$.

The simplest practical case is that of a 'step' P–N junction, in which one side is much less pure than the other. Virtually all the reverse bias voltage

appears across the space-charge region in the pure side, and the field is a linear function of distance[9]. For a P–N junction, we may write:

$$\left.\begin{aligned} E &= -E(X)\frac{x}{X},\ 0 < x < X \\ E &= 0 \qquad\qquad \text{otherwise} \end{aligned}\right\} \qquad \ldots (40)$$

where

$$E(X) = 2V/X$$
$$X = (V\kappa/2\pi qN_I)^{1/2}$$

with V = effective bias voltage, N_I = net impurity density (donors minus acceptors), κ = dielectric constant.

On changing the variable of integration by means of equation (40), equation (39) reads:

$$1 - M^{-1} = \frac{X}{E(X)}\int_0^{-E(X)} \alpha\, dE$$
$$= \frac{\kappa}{4\pi qN_I}\int_0^{E(X)} \alpha dE \qquad \ldots (41)$$

This equation may be used to obtain $\alpha(E)$ from experimental data in two ways[16].

(1) If $E(X)$ is varied while N_I is held constant:

$$\alpha = \frac{4\pi qN_I}{\kappa}\frac{d(M^{-1})}{dE(X)} \qquad \ldots (42)$$

(2) If M is made infinite (complete breakdown), and $E(X)$ is measured as a function of N_I:

$$\alpha = \frac{4\pi q}{\kappa}\Big/\frac{dE(X)}{dN_I} \qquad \ldots (43)$$

When $\alpha_n \neq \alpha_p$, a very much more involved method must be employed, and explicit expressions for the α's can be obtained only if M can be expressed[14] as a simple function of V.

2.4. **Avalanche Injection**

In the last section, the space charge of the carriers resulting from the avalanche ionization process was deliberately neglected. While this approximation is frequently justified in normal P–N junctions, it ceases to be valid when the current density exceeds a certain critical value $J_0 = |\, qN_I v_{ds}\,|$, where $v_{ds} \sim \frac{1}{2}v_0$ is the approximately constant drift velocity of carriers in a high electric field[19]. It is clear that if $J > J_0$, the density of carriers must be greater than N_I, and gross space charge may be present. The recognition of this result leads to the consideration of the properties of an avalanche in which the potential distribution is determined mainly by the space charge of the carriers. The condition for this, the intrinsic avalanche, is just that $J \gg J_0$.

The simplest case for which $J \gg J_0$ evidently occurs when $J_0 = N_I = 0$; that is, in intrinsic material. The following equations can then be written for

the hole and electron densities p and n in terms of the currents and the drift velocities v_p and v_n of the carriers:

$$j_p = qpv_p \qquad \ldots (44)$$

$$j_n = -qnv_n \qquad \ldots (45)$$

The rate of generation of carriers is expressed by:

$$\nabla.j_p = -\nabla.j_n = \alpha_p j_p + \alpha_n j_n \qquad \ldots (46)$$

while Poisson's equation is:

$$\nabla.E = \frac{4\pi q}{\kappa}(p - n) \qquad \ldots (47)$$

If, for simplicity, we put $\alpha_p = \alpha_n$, $v_p = -v_n = v_{ds} = \text{const.}$, and restrict consideration to the single dimension x, we obtain, by differentiating equation (47) a second time and substituting from equations (44), (45) and (46)

$$\frac{d^2E}{dx^2} = \frac{8\pi J}{\kappa v_{ds}} . \alpha(E) \qquad \ldots (48)$$

which can, in principle, be solved when $\alpha(E)$ is known.

It is immediately evident from this equation that the space charge $(\kappa/4\pi)(dE/dx)$ cannot vanish everywhere unless $\alpha = 0$. Unless the boundary conditions restrict the avalanche to a very small range of x, it is clear that very large values of E and α will be reached as J increases.

Suppose that near $x = 0$, J, E, and v_{ds} are all negative (*i.e.*, in the $-x$ direction). Electrons then travel in the $+x$ direction, and the electron current grows as x increases from $x = 0$, at which point we may set $j_n = 0$. The magnitude of the electric field will then decrease with increasing x, reaching zero at some point $x = x_0$. Beyond $x = x_0$, E will become positive; this is physically ridiculous, and indicates that equation (48) cannot apply in this region. (In fact, the space charge and field both remain nearly zero, and the current is carried mostly by diffusion. We shall return to this point later.) The interesting part of the solution of equation (48) is the behaviour of the avalanche voltage V_a, measured between $x = 0$ and $x = x_0$, as a function of J.

Equation (48) can be integrated once by the substitution $4\pi\rho/\kappa = dE/dx$, giving:

$$\rho \frac{d\rho}{dE} = \frac{\kappa J}{2\pi v_{ds}} \alpha(E)$$

or

$$\rho^2 = \frac{\kappa J}{\pi v_{ds}} \int \alpha \, dE + \text{const.} \qquad \ldots (49)$$

Now at $x = x_0$, j_p and j_n must be comparable, since the flow is by diffusion for $x > x_0$. For simplicity, we neglect the difference of the diffusion constants so that the boundary condition at $x = x_0$ becomes $j_p = j_n$, $\rho = 0$. In conjunction with the definition $E(x_0) = 0$, this permits equation (49) to be written:

$$\rho^2 = \frac{\kappa J}{\pi v_{ds}} \int_0^E \alpha \, dE' \qquad \ldots (50)$$

We note that when $x = 0$, $\rho = -J/v_{ds}$ so that the field E_0 at $x = 0$ is given by:

$$\int_0^{E_0} \alpha \, dE' = \frac{\pi J}{\kappa v_{ds}} \qquad \ldots (51)$$

Thus E_0 increases as J increases, but only slowly if α is a rapidly increasing function of E.

It will be seen that further progress depends on the assumption of an explicit form for $\alpha(E)$, so that, in the equation:

$$\frac{dE}{dx} = 4\left\{\frac{\pi}{\kappa}\frac{J}{v_{ds}}\int_0^E \alpha \, dE'\right\}^{1/2} \qquad \ldots (52)$$

derived from equation (50), the right-hand side can be evaluated as a function of E. The simplest function which is even remotely realistic is:

$$\alpha = \alpha_1 \exp(\lambda E) \qquad \ldots (53)$$

While this is not a very good fit to the experimental data, it is a sufficiently good approximation to enable the general behaviour of the solution of equation (50) to be elicited in an analytical form. Both α_1 and λ are negative quantities, if, as here, E is negative.

With this assumption, equation (52) becomes:

$$\frac{dE}{\{\exp(\lambda E) - 1\}^{1/2}} = 4\left(\frac{\pi J \alpha_1}{\kappa v_{ds} \lambda}\right)^{1/2} dx$$

or

$$\tan^{-1}\{\exp(\lambda E) - 1\} = 2\left(\frac{\pi J \alpha_1 \lambda}{\kappa v_{ds}}\right)^{1/2} x + \text{const.} \qquad \ldots (54)$$

Now when $x = 0$, $E = E_0$; thus, from equation (51):

$$\exp(\lambda E_0) - 1 = \frac{\pi J \lambda}{\kappa v_{ds} \alpha_1} = \gamma \qquad \ldots (55)$$

while when $E = 0$, $x = x_0$. The constant of integration can thus be put in the form:

$$\tan^{-1}\{\exp(\lambda E) - 1\} = 2\gamma^{1/2}\alpha_1 (x - x_0) \qquad \ldots (56)$$

where

$$x_0 = -\frac{1}{2\gamma^{1/2}\alpha_1}\tan^{-1}\gamma \qquad \ldots (57)$$

The voltage V_a across the avalanche is thus:

$$-\lambda V_a = \lambda \int_0^{x_0} E \, dx$$

$$= \int_0^{x_0} \ln[1 + \tan\{2\gamma^{1/2}\alpha_1 (x - x_0)\}] \, dx \qquad \ldots (58)$$

With the substitution $u = \tan\{2\gamma^{1/2}\alpha_1 (x - x_0)\}$, and $\tan(2\gamma^{1/2}\alpha_1 x_0)$ expressed by means of equation (57), this becomes:

$$\lambda V_a = \frac{1}{2\gamma^{1/2}\alpha_1}\int_0^{\gamma} \ln\frac{(1+u)}{1+u^2} \, du \qquad \ldots (59)$$

For $\gamma \gg 1$, this is approximately:

$$V_a = \frac{1}{2\gamma^{1/2}\alpha_1\lambda}\int_0^\infty \frac{\ln(1+u)}{1+u^2}\,du$$

$$= \frac{1}{2\gamma^{1/2}\alpha_1\lambda}\left(\frac{\pi}{4}\ln 2 + G\right)$$

$$= 0{\cdot}414\left(\frac{\kappa v_{ds}}{\lambda^3\alpha_1}\right)^{1/2} J^{-1/2} \qquad \ldots (60)$$

(G is the Catalan constant 0·91597).

Thus we have the interesting result that, while the maximum field in the intrinsic avalanche increases with increasing current density, both the thickness and the voltage decrease. Such an avalanche is thus capable of exhibiting a negative resistance*. In addition, so far as the bulk of the intrinsic region ($x > x_0$) is concerned, the avalanche behaves as a copious source of electrons which are injected into the intrinsic material, despite the non-injecting nature of the p^+ electrode. The name *avalanche injection* has therefore been given to this process. Since a finite density of impurities will not affect the result provided $J \gg J_0$, avalanche injection must always occur when the density of majority carrier current flowing towards a non-injecting electrode exceeds the critical value J_0, even though the boundary condition $j_p = j_n$ at $x = x_0$ may not always be satisfied. It should be noted that the one-dimensional theory gives no indication of the transverse stability of the avalanche, and still leaves open the possibility that the stable configuration may be one in which all the current passes through a small region whose transverse dimensions are comparable with x_0.

2.5. Avalanche at a Current Constriction

For intrinsic material, a reversal of current flow in the model discussed in the last section leads only to the injection of holes by the normal process, without the intervention of any high-field phenomena. If the intrinsic region were changed to N-type, this would still apply, and the condition $J > J_0$ would imply nothing unusual. A change to P-type on the other hand, would prevent an increase of carrier density by injection, and large values of J would imply a field sufficient to produce avalanche multiplication throughout the P-region. This is not a very realistic situation, but very similar conditions must apply when a current is forced to flow through a constriction[20]. The field need then be high only in the small volume for which $J \ll J_0$.

In order to simplify the analysis, we shall assume a simplified form for the dependence of J on E. For low fields ($J < J_0$), Ohm's law will be assumed to hold, with the conductivity equal to the low-field value σ_0. Above the field E_1, given by $\sigma_0 E_1 = J_0$, the drift velocity is supposed to be independent of E. At some larger field E_2, avalanche multiplication will become important, and

* Similar results have been obtained by Leblond[28], starting from rather different equations.

it will be assumed that J then increases without limit. The $J(E)$ relation thus has the form

$$J = \sigma_0 E \qquad (E \leqslant E_1) \qquad \ldots (61)$$

$$J = J_0 \qquad (E_1 < E < E_2) \qquad \ldots (62)$$

$$J \to \infty \qquad (E \to E_2) \qquad \ldots (63)$$

as shown in Figure 1*a*.

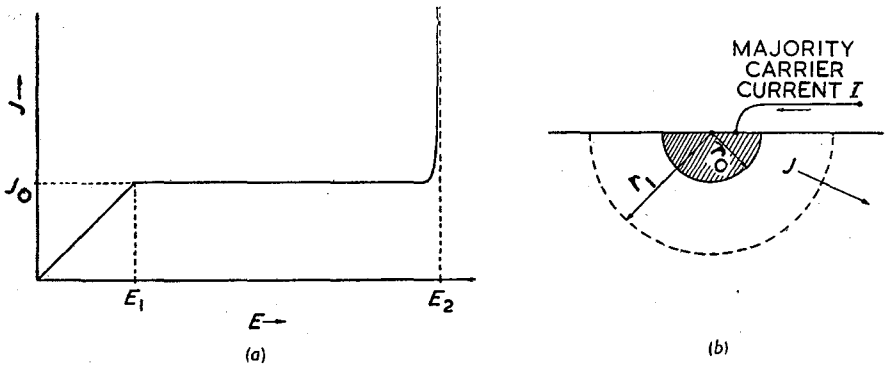


Figure 1(*a*) *Approximate form of* $J(E)$. (*b*) *A simple type of current constriction.*

The simplest geometrical configuration for a current constriction is one in which the current flow is purely radial. Such a flow pattern exists in the structure shown in Figure 1*b*. Here a current I of majority carriers enters a small hemispherical island (radius r_0) of heavily doped material, flows radially outward across its junction with a uniform semi-infinite region of purer material of the same conductivity type, and leaves at infinity. The current density at a radius r is evidently:

$$J = I/(2\pi r^2) \qquad \ldots (64)$$

If I is greater than a certain critical current $I_0 = 2\pi r_0^2 J_0$, a critical radius r_1, can be defined, for which $J(r_1) = J_0$, or:

$$r_1 = \left(\frac{I}{2\pi\sigma_0 E_1}\right)^{1/2} \qquad \ldots (65)$$

For $r \geqslant r_1$, Ohm's law is valid, E is given by equation (61), and the potential difference $V_1 - V_\infty$ between $r = r_1$ and $r = \infty$ is simply:

$$V_1 - V_\infty = \frac{I}{2\pi r_1 \sigma_0} = \left(\frac{I E_1}{2\pi\sigma_0}\right)^{1/2} \qquad \ldots (66)$$

For $r_0 < r < r_1$, on the other hand, equation (63) applies, and hence E is constant and equal to E_2. Thus the potential difference $V_0 - V_1$ between $r = r_0$ and $r = r_1$ is:

$$V_0 - V_1 = E_2(r_1 - r_0)$$

$$= \frac{E_2}{E_1}\left(\frac{IE_1}{2\pi\sigma_0}\right)^{1/2} E_2 r_0 \qquad \ldots (67)$$

For $r \leqslant r_0$, E may be neglected, since the heavily doped island may be made of arbitrarily high conductivity. The overall voltage V, measured between a point in the island and one at infinity, is thus:

$$V = V_0 - V_\infty = \left(\frac{E_2}{E_1} + 1\right)\left(\frac{IE_1}{2\pi\sigma_0}\right)^{1/2} - E_2 r_0 \qquad \ldots (68)$$

If I is less than I_0, then r_1 is less than r_0, Ohm's law applies everywhere, and the resistance of the structure is just the spreading resistance:

$$R_s = \frac{1}{2\pi r_0 \sigma_0} \qquad \ldots (69)$$

Thus the effect of the constriction is to make the resistance of the structure non-linear, and, for currents somewhat greater than I_0, the resistance will be much higher than the spreading resistance calculated according to Ohm's law. Although the theory has been given for the case of an L–H junction (*i.e.* one between lightly and heavily doped regions), the results will remain true when applied to other types of step junction (*e.g.* P–N, or metal–semiconductor) provided the current exceeds the critical value I_0. For, under these conditions, there must be avalanche ionization inside the r_1 hemisphere irrespective of the nature of the junction, and so any rectifying properties which it may possess at low currents will be ineffective. The corrections to the theory required to take account of deviations from spherical symmetry are generally small provided I is somewhat greater than I_0. It therefore follows from the theory that all small contacts to a semiconductor must have a high resistance to the outward flow of majority carriers, provided only that their drift velocity saturates and that the current exceeds I_0. For a given semiconductor, this resistance should be the same for all contacts of the same size.

The least plausible of the assumptions involved in the foregoing analysis is perhaps that of a constant field E_2 within the r_1 hemisphere. The validity of the assumption may be examined by using $E = E_2$ as a trial solution in the following system of equations:

$$\nabla \cdot \boldsymbol{j}_p = \alpha J \qquad \ldots (70)$$

$$\nabla \cdot \boldsymbol{j}_n = -\alpha J \qquad \ldots (71)$$

$$\nabla \cdot \boldsymbol{E} = \frac{4\pi\rho}{\kappa} \qquad \ldots (72)$$

$$J = I/2\pi r^2 \qquad \ldots (73)$$

$$\rho = \frac{1}{v_d}(j_p - j_n - J_0) \qquad \ldots (74)$$

which apply to P-type material for $r_0 < r < r_1$. On subtracting equation (71) from equation (70), substituting for J from equation (73), and expressing $\nabla \cdot$ in terms of the single co-ordinate r, we obtain:

$$\frac{d}{dr}\left\{r^2(j_p - j_n)\right\} = \frac{\alpha I}{\pi} \qquad \ldots (75)$$

Equations (72) and (74) may be combined to give:

$$r^2(j_p - j_n - J_0) = \frac{\kappa v_d}{4\pi}\frac{d}{dr}(r^2E) \qquad \ldots (76)$$

whence, from equation (75):

$$\frac{\alpha I}{\pi} = \frac{d}{dr}\left\{\frac{\kappa v_d}{4\pi}\frac{d}{dr}(r^2E) + r^2J_0\right\} \qquad \ldots (77)$$

If E is constant and equal to E_2, then v_d must be constant also, and equation (77) becomes:

$$\alpha I = \frac{\kappa v_d E_2}{2}\left(1 + \frac{4\pi r J_0}{\kappa v_d E_2}\right) \qquad \ldots (78)$$

Clearly the trial solution $E = \text{const.}$ will be a good approximation provided that the variation of α with r described by equation (78) is not so large as to imply a significant variation of E with r. Provided $\alpha(E)$ is a rapidly increasing function, quite large proportional changes in α will be permissible as the corresponding field changes will still be small.

The variation of α with r arises only from the second term on the R.H.S. of equation (78) and the largest possible variation of this is evidently $4\pi\, r_1\, J_0/\kappa\, v_d\, E_2$. If this quantity is less than unity, α changes by less than factor of two in the range $r_0 < r < r_1$, and the variation of E will be negligible. The condition can be expressed in the form

$$\frac{2(2\pi I J_0)^{1/2}}{\kappa v_d E_2} < 1 \qquad \ldots (79)$$

For germanium $\kappa\, v_d\, E_2$ is at least 0·4 A cm^{-1}, and the approximation is frequently justified.

Having established that the integration leading to equation (68) is approximately correct, we see from equation (78) that the variation of α with I should not lead to serious error in expressing V as a function of I with $E_2 = \text{constant}$. This is again because of the rapid variation of α with E.

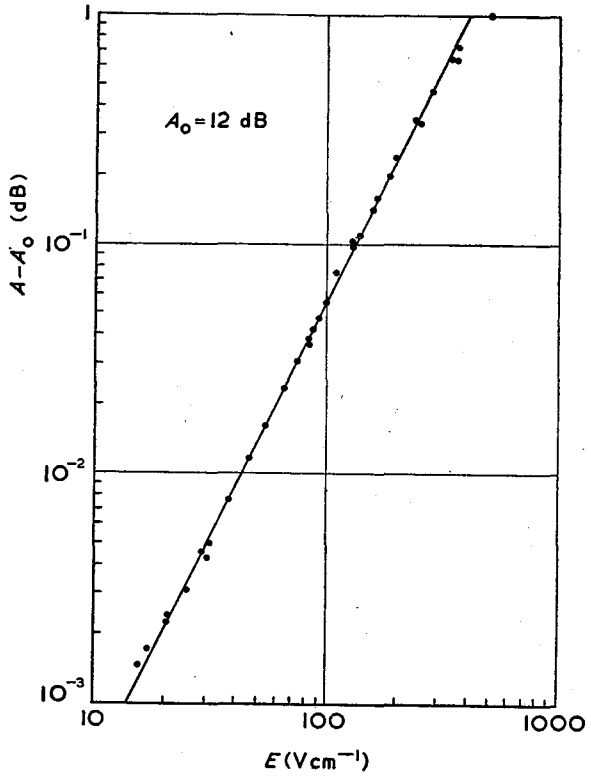
It is interesting to note that the foregoing analysis applies only when the majority carrier current is directed *outwards* from $r = 0$. If it were directed inwards, the two terms on the R.H.S. of equation (78) would have opposite signs, and, if their magnitude were comparable, the solution $E = \text{const.}$ would not be even approximately true. This result is associated with the fact that under these conditions the minority carriers produced by the avalanche near $r = r_0$ are injected into the semiconductor. They increase its conductivity and permit a current in excess of J_0 to be carried without the presence of space charge. This is an instance of the injecting avalanche discussed in the last section, and permits large values of I to exist without the correspondingly high values of V given by equation (68). In conjunction with the previous result for the high-resistance direction, this conclusion establishes as a theorem that any small contact to a semiconductor must be a rectifier and an emitter of minority carriers[19].

3. EXPERIMENTAL RESULTS

3.1. Small Quadratic Changes in Mobility

Measurements of the small quadratic changes in mobility predicted by equation (16) have so far been limited to N-type germanium. Two basic methods have been employed.

Arthur, Gibson and Granville[21] have measured the attenuation A of 8 mm electromagnetic radiation in passing through a germanium specimen, and its variation as a function of an externally applied pulsed electric field. The field



[*By courtesy of the editor of the Journal of Electronics.*

Figure 2. *The variation of attenuation with applied electric field (Arthur, Gibson and Granville[21]).*

was applied parallel to the electric vector of the wave; the absorption was thus a function of the *slope mobility* $\mu' = \mathrm{d}v_\mathrm{d}/\mathrm{d}E$, and not of the ordinary mobility $\mu = v_\mathrm{d}/E$. In the range of validity of equation (16), we have:

$$\mu' = \mu_0 (1 + 3\beta E^2) \qquad \ldots (80)$$

Under the particular experimental conditions used, A measured in decibels was nearly proportional to μ', so that, if A_0 is the value when $E = 0$, the relative

change is approximately:

$$\frac{A - A_0}{A_0} = 3\beta E^2 \qquad \ldots (81)$$

Hence β can be inferred from measurements of $A(E)$ and A_0. This method has the advantage that extremely small changes of μ' can be detected, and their functional dependence on E can be established with some certainty. The measurement of their absolute value, and of A_0, is, however, somewhat more difficult. Nevertheless, the values of β so obtained are probably reliable to better than a factor of 2. A major disadvantage of the method is that it is extremely difficult to vary the temperature of the specimen, which is placed across a copper waveguide.

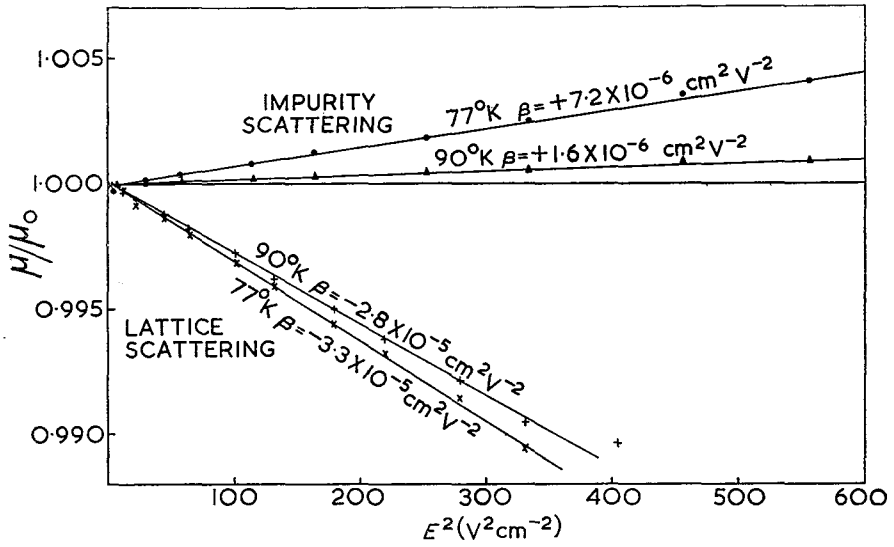


Figure 3. *Quadratic mobility changes at* 77° *and* 90° *K.*

A set of results obtained for a specimen of 4 Ω cm N-type germanium at 30 0° K is shown in Figure 2. These are quite typical of the method, and the value of β obtained from them, $-1{\cdot}5 \times 10^{-7}$ cm^2 V^{-2}, is near the mean value of the results of several determinations. This is to be compared with the theoretical value, which, since impurity scattering may be neglected, is $\beta = -(3\pi/64)(\mu_{L0}^2/c^2)$ for a Maxwellian distribution of velocities. If we adopt the values $\mu_{L0} = 3{\cdot}9 \times 10^3$ cm^2 V^{-1} sec^{-1} and $c = 5{\cdot}4 \times 10^5$ cm sec^{-1}*, (see references [24, 25]), the theoretical value of β is $-7{\cdot}7 \times 10^{-6}$ cm^2 V^{-2}. The observed magnitude of the mobility changes is thus some 50 times smaller than predicted.

A second method of measurement, used by the author, involves the direct measurement of resistance by means of a bridge circuit. In published work[22] a bridge excited by pulses less than 1 microsec long and using a cathode-ray

* For the velocity of longitudinal waves. Shear waves, for which $c = 3{\cdot}55 \times 10^5$ cm sec^{-1}, may be important for inter-valley scattering.

tube as the null detector was employed[23]. The low sensitivity of this apparatus necessitates the application of high voltages, and thus relatively high fields, to the specimen. Thus the truly quadratic region cannot be explored. Furthermore, at room temperature, measurements of resistance appear to be complicated by the presence of surface effects, which may obscure small changes of mobility. Taken together, these two facts are probably sufficient to account for the discrepancy in the functional dependence of μ on E, between the results of the R.F.[21] and pulse[22] methods. A modification of the bridge method has been evolved[26], in which the pulse length is increased to about 1 msec, the bridge arms are replaced by conventional wire-wound resistance boxes, and the detector consists of an amplifier and oscilloscope. These changes bring about a considerable decrease in the minimum electric field and an increase in the sensitivity. The unwanted surface effects can also be decreased considerably by taking care to keep the surface uncontaminated and by keeping the temperature below 90° K.

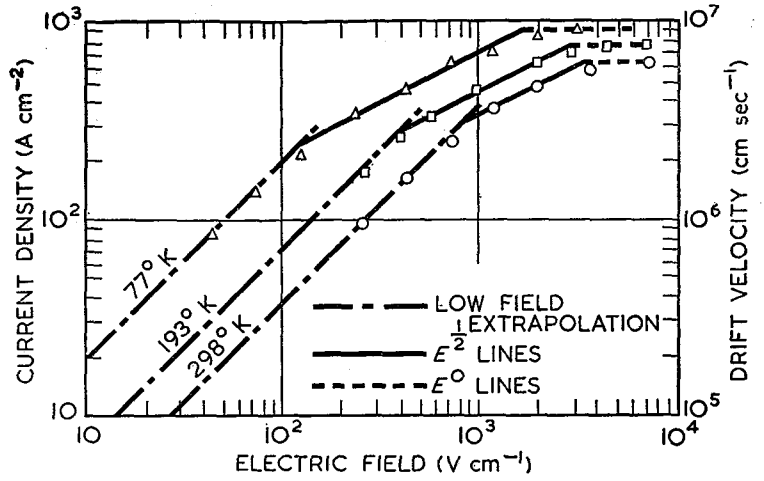
With these precautions, the measured variation of mobility with field is found to have the typical form shown by Figure 3. Here μ/μ_0 is plotted against E^2, the points lying close to a straight line, as predicted by equation (16). The slope of the line is evidently equal to β. For the points marked 'lattice scattering', the specimen had a net impurity content of about 4×10^{14} donors cm^{-3}. Assuming that there is no compensation of impurities the value of a at 77° K is thus of the order of 10^{-1}, and impurity scattering is negligible. If μ_{L0} is extrapolated from its value of $3{\cdot}9 \times 10^3$ cm^2 V^{-1} sec^{-1} at 300° K by using the measured ratios of the resistances at various temperatures, the values of β calculated as above are about $+ 3{\cdot}8 \times 10^{-4}$ and $- 2{\cdot}8 \times 10^{-4}$ cm^2 V^{-2} at 77° K and 90° K respectively. These are to be compared with the corresponding experimental values of $- 3{\cdot}3 \times 10^{-5}$ and $- 2{\cdot}8 \times 10^{-5}$ cm^2 V^{-2}.

In the case of the points marked 'impurity scattering' the net donor density was about $1{\cdot}7 \times 10^{16}$ cm^{-3}, and impurity scattering was important in determining β. The values of a may be calculated from the ratio of the resistance of the sample to that of the preceding specimen as a function of temperature. They are $a = 7{\cdot}7$ at 77° K, and $a = 6{\cdot}6$ at 90° K. Thus β is calculated to be $2{\cdot}8 \times 10^{-4}$ at 77° K and $2{\cdot}1 \times 10^{-4}$ cm^2 V^{-2} at 90° K. The corresponding measured values are $7{\cdot}2 \times 10^{-6}$ and $1{\cdot}6 \times 10^{-6}$ cm^2 V^{-2}. In this case the calculated values are quite sensitive to errors in the values of μ_{L0} and μ_I assumed.

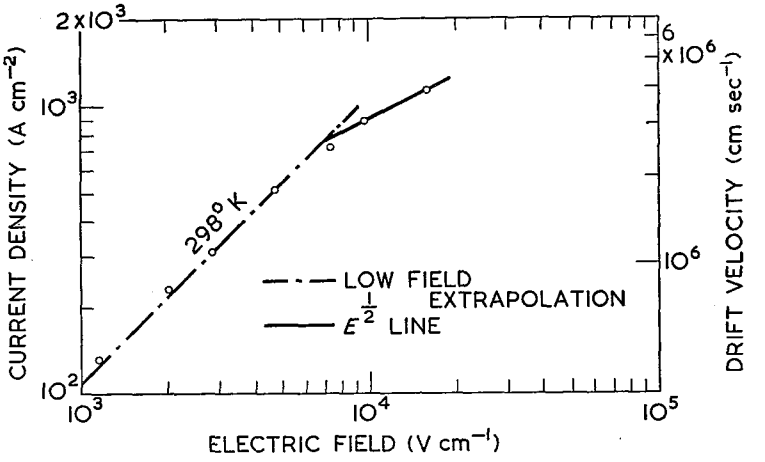
It is clear from the foregoing experimental data that, while theory and experiment are in agreement about the existence, sign, and form of the mobility variations, in all cases the experimental magnitudes are very much less than those calculated. It is difficult to see what the origin of this discrepancy may be, although it may be that equation (1) is in error, and that, because of the complex band structure of germanium, the average energy loss of an electron in collision with a phonon is greater than has been supposed[8, 17]. A factor of 10 on the R.H.S. of equation (1) would remove the greater part of these discrepancies, together with those noted in the next section.

3.2. Major Variations of Mobility

The large changes of mobility which are expected to occur at fields outside the quadratic range have been observed by Ryder[27] in N- and P-type silicon

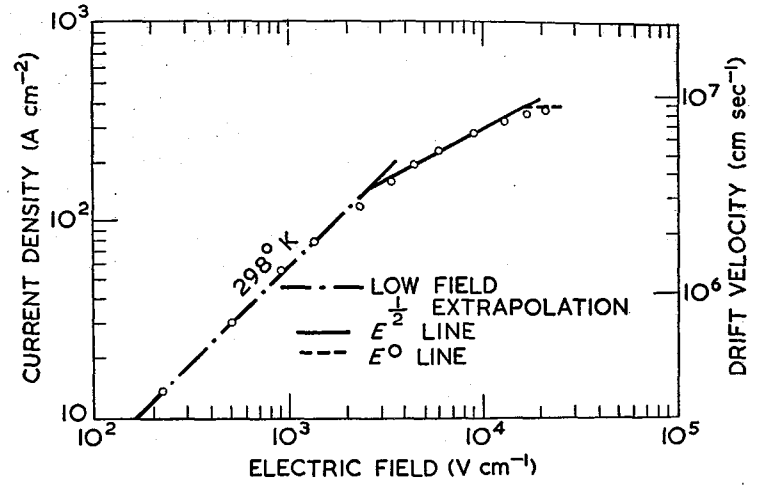


(Figure 4*a*). *Germanium: electrons.*

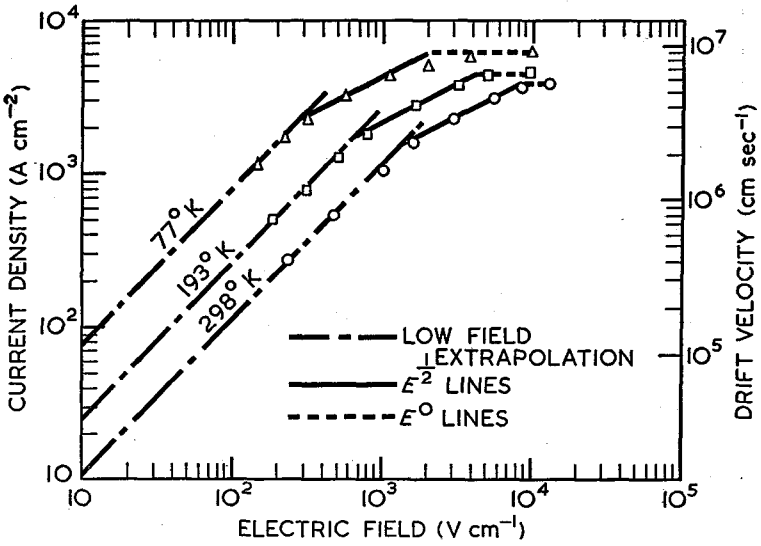


(Figure 4*b*). *Germanium: holes.*

and germanium, and by Gunn[22] in N-type germanium. Both authors measured the resistance of a rectangular section of the semiconductor under pulsed field conditions, and inferred the drift velocity from the conduction current. For values of electric field below about 10^4 V cm^{-1}, relatively long pulses, of the order of 10^{-7} sec, may be used, but at higher values the pulse length must be reduced to about 2×10^{-9} sec in order to avoid catastrophic thermal effects. Some data obtained[27] for the lower field range are shown in Figure 4. The general form of the curves is the same for all the cases studied. There is at low fields a region of approximately constant mobility, in which the accuracy is insufficient to show the quadratic deviation, followed by a range in which the



(Figure 4*c*). *Silicon: electrons.*



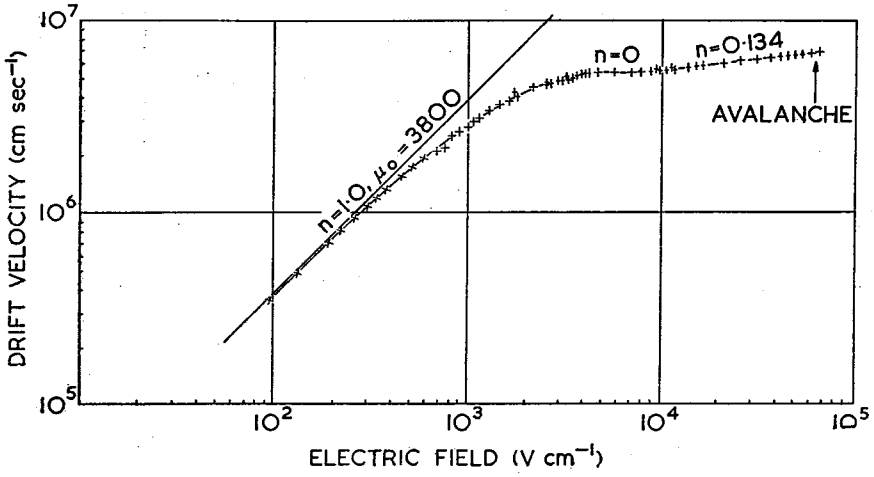
(Figure 4*d*). *Silicon: holes.*

Figure 4. *The variation of drift velocity with electric field (Ryder*[27]*).*
[*By courtesy of the editor of the Physical Review.*

ıobility decreases progressively. It is possible to separate the upper range into vo regions, in which the drift velocity varies first as $E^{1/2}$ and then indepen-ently of E, as predicted theoretically. It should be observed, however, that the ımber of experimental points in each range is hardly sufficient to suggest such relation in the absence of a theory to be fitted. Further doubt is thrown on ıe existence of a square-root region by the data[22] of Figure 5, in which the ɛperimental points are sufficiently dense for it to be apparent that, at any rate

for electrons in germanium, the slope of the ln v_d — ln E curve merely passes smoothly from 1 to 0, without ever remaining constant at the value $\frac{1}{2}$. The existence of a saturation range is, however, fully confirmed, although at the highest fields the velocity begins to increase once again.

By fitting a tangent of slope 1/2 to the experimental points it is at least possible to obtain an estimate of the degree of agreement between the theoretical and observed behaviour. A measure of the magnitude of the mobility



[*By courtesy of the editor of the Journal of Electronics.*

Figure 5. ***The variation of electron drift velocity with electric field for Germanium*** (*Gunn*[22]).

deviation is given by the field E_c at which the $E^{1/2}$ tangent intersects an extrapolation of the low-field range. According to equation (19), this field should be:

$$E_c = 1{\cdot}51\ c/\mu_0 \qquad \ldots\ (82)$$

which can be readily calculated from known data. In every case it is found that at room temperature the measured values of field are about 3 times greater than the theoretical. This points once again to the conclusion that the carriers are able to lose energy to the lattice vibrations about ten times more rapidly than would be expected theoretically.

In the range of 'saturated' drift velocity, on the other hand, the agreement between theory and experiment is much better. The theoretical and experimental values of $\frac{1}{2}\ v_0$ at room temperature are compared in Table 1, on the assumption that the carrier mass is equal to the free electron mass. In view of the small

Table 1.

Carrier:		Ge, electron	Ge, hole	Si, electron
$\frac{1}{2}v_0$ (cm sec^{-1})	theory	$6{\cdot}1 \times 10^6$	$6{\cdot}1 \times 10^6$	$8{\cdot}2 \times 10^6$
	experiment[27]	$6{\cdot}2 \times 10^6$	$6{\cdot}0 \times 10^6$	$8{\cdot}5 \times 10^6$
	experiment[22]	$5{\cdot}2 \times 10^6$		

and anisotropic effective masses of the carriers, we must suspect that the agreement is to some extent fortuitous.

The rise of v_d at the highest fields shown in Figure 5 can be ascribed to the existence of a finite scattering probability for $v > v_0$. According to equation (32), a plot of log $(v_d - \frac{1}{2} v_0)$ against log E should give a straight line of slope $1/(b+1)$. Figure 6 shows such a plot for $\frac{1}{2} v_0 = 5{\cdot}22 \times 10^6$ cm sec^{-1}. The slope of the line which best fits the experimental points is about 0·83. Thus A is about $3{\cdot}3 \times 10^{12}$ c.g.s. units, and b is about 0·2, compared with the value of 1 suggested by Shockley[4] on theoretical grounds. The fit of the experimental

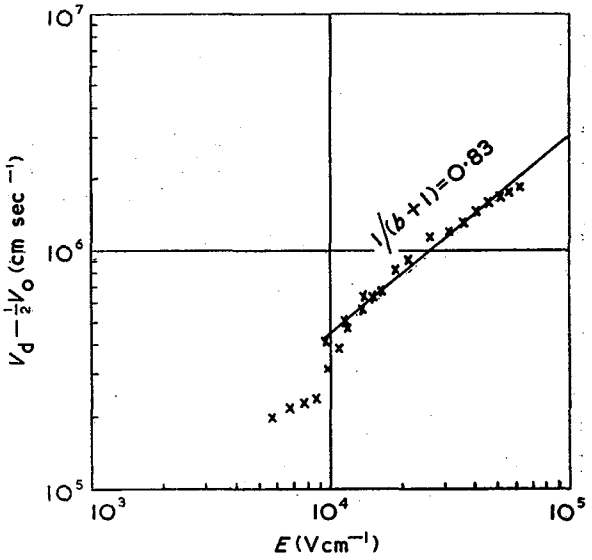


Figure 6. *The variation of* $v^d - \frac{1}{2}v_0$ *with E. Electrons in germanium (data of Figure 5)* $\frac{1}{2}v_0 = 5{\cdot}22 \times 10^6$ *cm sec*$^{-1}$.

points to the straight line is poor, but this is to be expected in view of the fact that the dependent variable is the small difference of large (observed) quantities. Nevertheless the accuracy is sufficient to show that the slope is not $\frac{1}{2}$, which would correspond to $b = 1$. In view of the general uncertainty about the probability of scattering by optical phonons, the disagreement is not serious.

3.3. Avalanche Ionization and P–N Junction Breakdown

The simplest structure in which avalanche ionization should be observable is a P–N junction, since, in this the high fields necessary are not accompanied by excessively high current densities. The problems of excessive power dissipation are therefore considerably reduced. Such junctions generally have a reverse current which, while remaining saturated at low voltages, rises more and more rapidly as a certain limiting breakdown voltage is approached. This rise has been ascribed by McKay and McAfee to avalanche multiplication of the current of thermally excited minority carriers constituting the saturation current[29]. They have shown that small changes in minority carrier current, produced

either by illumination of the junction or by α-particle bombardment, are multiplied in passing through the junction (Figure 7). At a given voltage, the amount of multiplication is just equal to that required to explain the normal reverse current by multiplication of the saturation component. These authors state that the Zener or internal field emission effect[15], which was previously

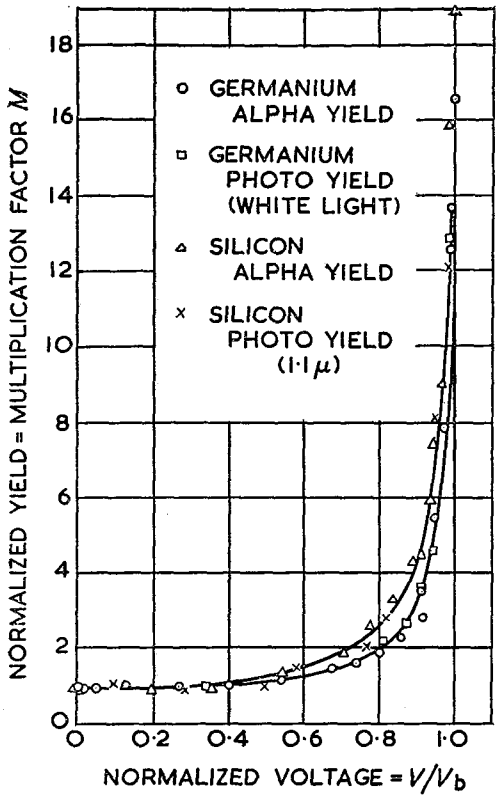


Figure 7. *Multiplication curves for grown silicon and germanium junctions (McKay and McAfee[29]).*

[*By courtesy of the editor of the Physical Review.*

believed to be the cause of junction breakdown, could not account for the observed multiplication. This statement requires modification in view of the fact that the passage of minority carriers through a junction may increase the space charge density ρ. For the step junction considered in Sub-section 2.3 a hole current j_p increases the density on the N-type side from $\rho = qN_I$ to $\rho = qN_I/ + j_p/v_{ds}$. Near the step, the field in the junction has its maximum value $E(X) = (8\pi\rho V/\kappa)^{1/2}$, and it is clear that this must be increased by the passage of the hole current. Since the rate of generation by the Zener effect should be a very steeply rising function of E, the total rate of generation ('Zener current') J_z should depend mainly on E_{max}. Thus it follows that, to a good approximation, J_z is a function only of the variable ρV. Hence the partial derivatives are connected by the relation:

$$\rho \frac{\partial J_z}{\partial \rho} = V \frac{\partial J_z}{\partial V} \qquad \ldots (83)$$

and the multiplication M is:

$$M - 1 \equiv \frac{\partial J_z}{\partial j_p} = \frac{V}{\rho}\frac{d\rho}{dj_p}\frac{\partial J_z}{\partial V} \quad \ldots (84)$$

For small values of j_p this reduces to:

$$M - 1 = \frac{V}{qN_I v_{ds}} \cdot \frac{\partial J_z}{\partial V}$$

$$= \frac{J_z}{J_0} \cdot \frac{\partial \ln J_z}{\partial \ln V} \quad \ldots (85$$

In the germanium junction studied by McKay and McAfee, the orders of magnitude were $J_z \sim 10^{-2}$ A cm^{-2}, $J_0 \sim 10^4$ A cm^{-2}, d ln J_z/d ln $V \sim 10^3$, so that $M - 1$ should be about 10^{-3}. Thus the Zener effect fails to account for the observed magnitudes of M (typically ~ 10), and it is on this ground that it must be rejected.

The alternative theory, based on avalanche ionization, gives a consistent explanation both of the growth of reverse current and of the multiplication effects. If the ionization coefficient α is the same for both holes and electrons, its variation with E may be deduced either from multiplication experiments by the aid of equation (42), or from measurement of the variation of junction breakdown voltage with resistivity using equation (43). In the latter method the maximum field $E(X)$ is subject to considerable uncertainty, and the former is therefore to be preferred. In interpreting his results on silicon, McKay[16] assumed the coefficients to be equal, basing his conclusions on the results of multiplication experiments on symmetrical junctions in both silicon and germanium in which M was found to be the same for holes and electrons. Unfortunately there is, as we shall see, reliable evidence that in germanium the values are different. Figure 8 shows the variation of α with E calculated by McKay from multiplication experiments, and in Figure 9 the experimental curve is compared with the theory of Wolff[13]. Considering the experimental difficulties and the doubts about the equality of ionization rates, the agreement is satisfactory. The theoretical curve is based on a mean free path l_0 of 200 Å and an energy threshhold ϵ_0 of 2·3 eV.

In germanium alloyed step junctions it has been found empirically by Miller[14] that the multiplication is well expressed by the relation:

$$M = \{1 - (V/V_b)^m\}^{-1} \quad \ldots (86)$$

where V_b is the breakdown voltage.

The value of m is found to be very different for junctions made on N-type ($m \sim 3$) and P-type ($m \sim 6$) material. The only possible interpretation for this result is that the ionization coefficients are different for holes and electrons. This complicates very considerably the mathematical analysis, but, thanks to the empirical relation (86), Miller has been able to calculate the values separately. The results are shown in Figure 10. The curve drawn through the experimental points for holes has been calculated from Wolff's theory using the values $l_0 = 130$ Å and $\epsilon_0 = 1{\cdot}5$ eV. That for electrons is drawn to fit the experimental points, and has roughly the same shape, but the absolute magnitudes are

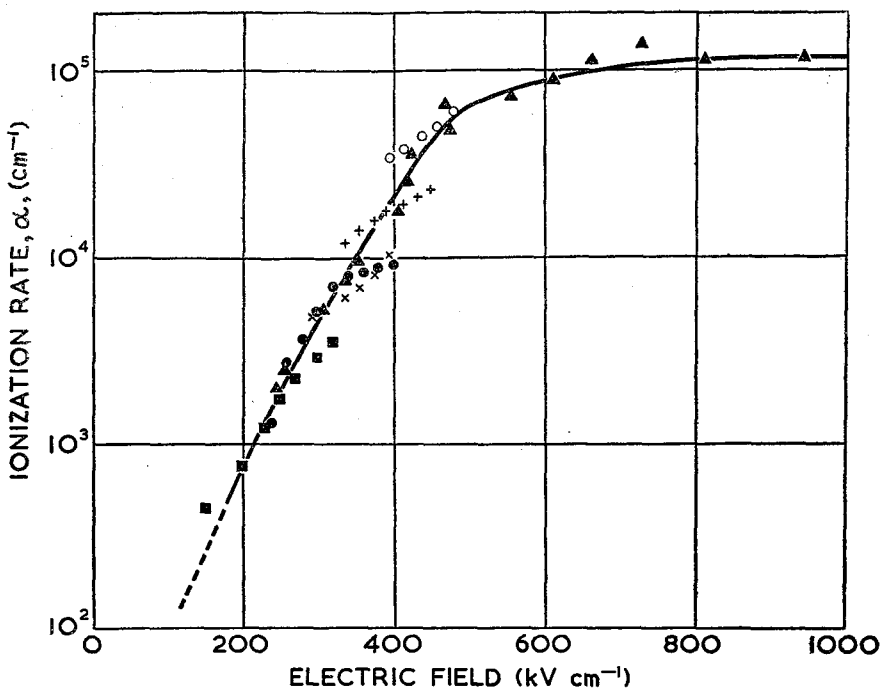


Figure 8. *The ionization coefficient α measured as a function of field in silicon (McKay[16]).* ▲ *step-junction breakdown data.* O + *linear-gradient multiplication data.* X *step-junction multiplication data.*

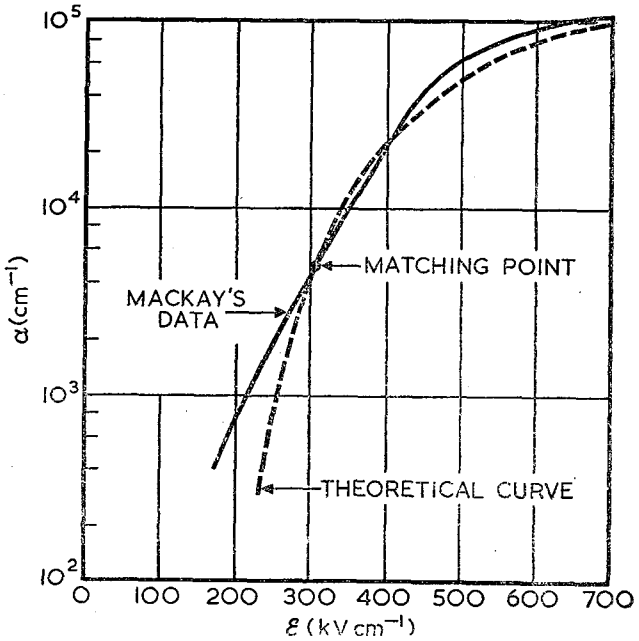


Figure 9. *Comparison of theoretical and experimental values of ionization coefficient in silicon (Wolff[13]).*

[Figures 8 and 9 by courtesy of the editor of the Physical Review.

smaller by a factor of about 2. The difference, while quite sufficient to make a substantial change in the shape of the multiplication curves, does not seem to make a difference between N- and P-type in the variation of V_b with the impurity density N_I, as shown in Figure 11. The slope of the log-log plot, which is $-0{\cdot}725$, is further evidence against the Zener effect, since for this the slope should be -1.

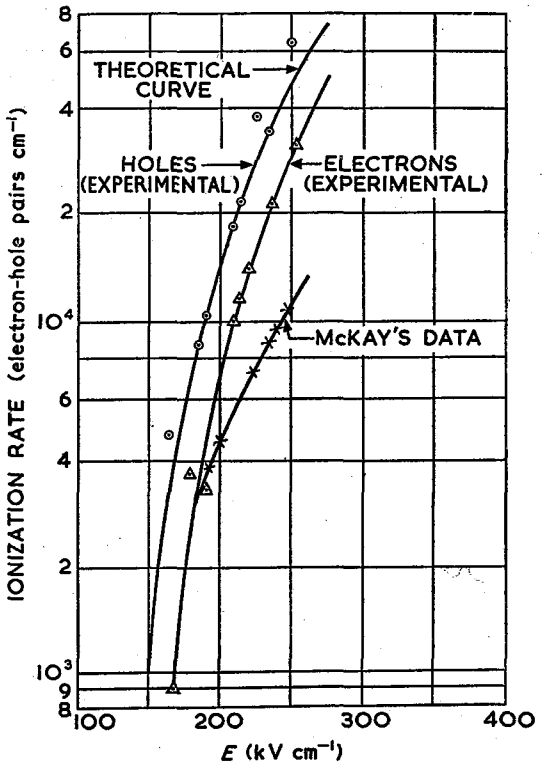


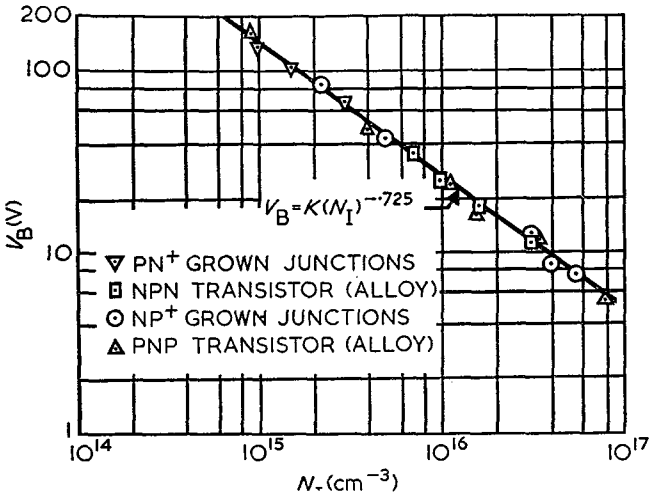
Figure 10. *Experimental values of the ionization rates for holes and electrons in germanium, and the theoretical curve (Miller*[14]*).*

[By courtesy of the editor of the Physical Review.

It is interesting to note that for both silicon and germanium the value deduced for the mean free path for optical phonon scattering is of the order of 200 Å, and that the energy threshold is about 0·4 eV greater than the theoretical minimum of $\frac{3}{2}\epsilon_G$. For germanium, the mean free path may be compared with the value obtained by extrapolation from equation (20), using the constants $b = 0{\cdot}2$, $A = 3{\cdot}3 \times 10^{12}$ c.g.s. units derived from Figure 6. For a velocity of $7{\cdot}3 \times 10^7$ cm sec^{-1}, corresponding to the threshold energy of 1·5 eV, the mean free path $l_0 = v\tau_0$ so obtained is about 60 Å. The agreement with the value of 130 Å obtained by Miller is satisfactory, considering the experimental difficulties and the approximate nature of the theory.

A further 'high field effect' allied to avalanche ionization is the ionization of neutral donor or acceptor atoms[30] which occurs in germanium at a field of about 4·5 V cm^{-1}. Such neutral atoms are, of course, present in significant numbers only at temperatures of the order of 30° K and below. Their ionization energy

is of the order of 0·01 eV, so that the field required is very much less than that required to ionize a germanium atom in the normal avalanche process. The increase of conductivity resulting from this process was at first[27, 6] interpreted as a mobility change resulting from the strong impurity scattering, but this is no longer believed to be the explanation[31].



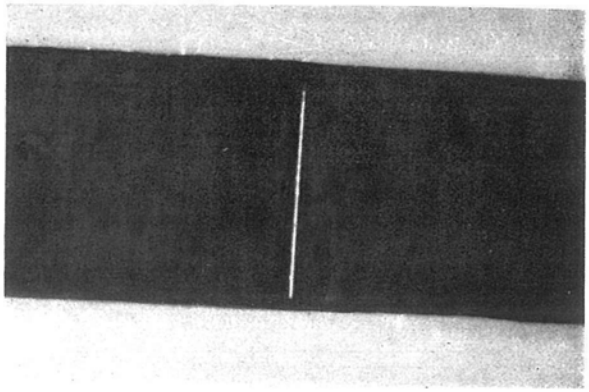
[*By courtesy of the editor of the Physical Review.*

Figure 11. *The breakdown voltage of germanium step junctions as a function of the net impurity density on the high resistivity side (Miller[14]).*

3.4. Light Emission from an Avalanche

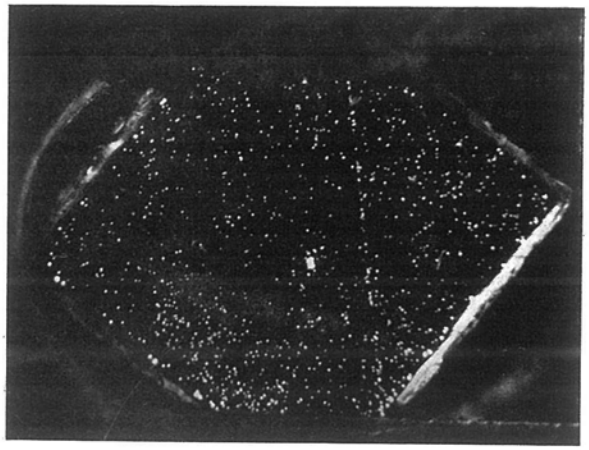
A particularly beautiful effect has been observed in silicon P–N junctions by Newman[32], and by Chynoweth and McKay[33]. This consists of the emission of visible light from the region in which avalanche multiplication is taking place. As the avalanche current is increased to large values, spots of yellowish light appear on the line in which the junction intercepts the surface, and their number increases as the current rises until a practically continuous line of light is formed (Figure 12). A spot can be initiated at a given point by slight local surface damage. The spectrum of the emitted light is shown in Figure 13, from which it will be seen that, by comparison with the normal recombination spectrum[34] shown in the same figure, the avalanche emission is notable for its content of photons having energies over twice the gap width.

In a grown junction only the light emitted at the surface can be seen, but in a specimen made by diffusion, in which the junction lies parallel with the surface and only a few microns below it, emission from the interior can be detected. The light suffers absorption by the thin overlying layer of silicon, and its colour is thus changed to a reddish tinge, in accordance with the known absorption spectrum of silicon. The light appears as a very large number of extremely small spots, which appear and disappear in a reproducible order as the avalanche current is varied (Figure 14). By comparison of the spectra of the emission from the surface and the interior with the absorption spectrum, it is



[*By courtesy of the editor of the Physical Review.*

Figure 12. *A photograph of the light emission from a grown silicon junction along the line in which the junction intersects the surface (Chynoweth and McKay*[33]).



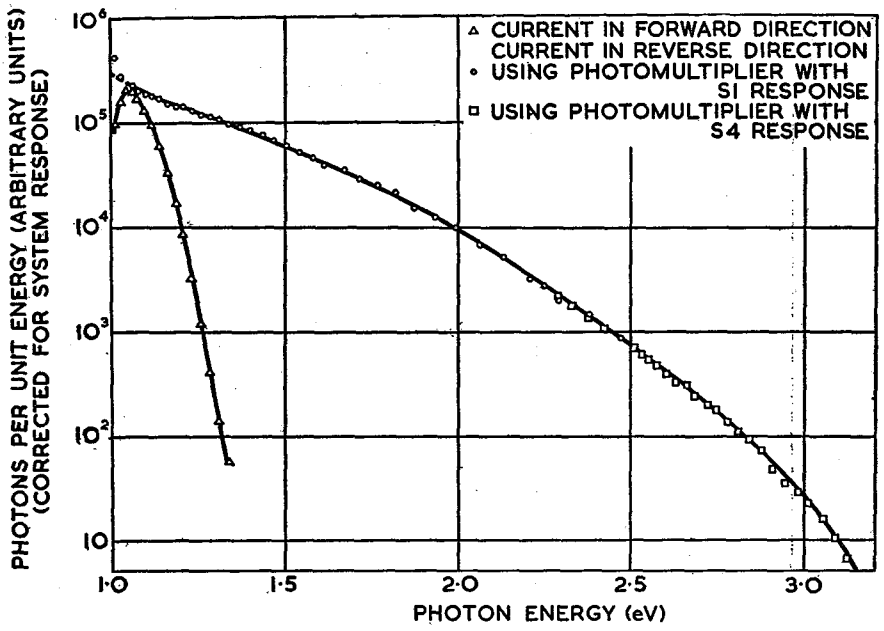
[*By courtesy of the editor of the Physical Review.*

Figure 14. *A photograph of the light emission from a silicon diffused junction (Chynoweth and McKay*[33]).

[*Facing p. 240*

possible to deduce that the emission comes from a depth which corresponds closely to the estimated position of the junction.

The origin of the light appears to lie partly in the radiative recombination of the energetic holes and electrons which exist in the avalanche. Theoretically, a photon with an energy of $\epsilon_G + 2\epsilon_0 \sim 5{\cdot}7$ eV could be emitted by the recombination of a hole and an electron each of which possessed the threshold energy for ionization. Such an event is highly improbable, of course, but it becomes more likely if the energies are below the threshold values. The low efficiency



[*By courtesy of the editor of the Physical Review.*

Figure 13. *The spectrum of the emission from an avalanche in silicon compared with that of the normal recombination radiation* (*Chynoweth and McKay*[33]).

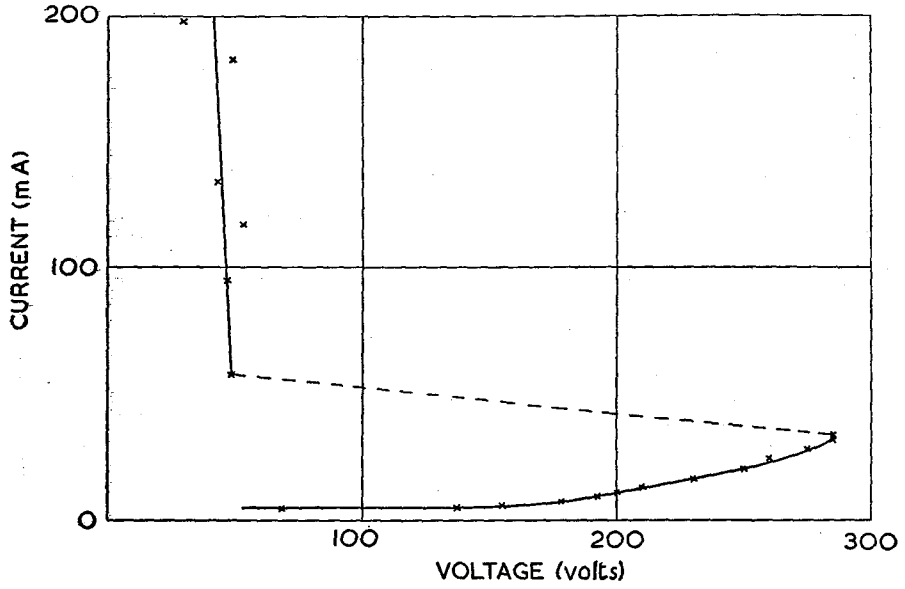
($\sim 7 \times 10^{-9}$ photons per carrier in the avalanche) and the shape of the spectrum above $1{\cdot}1\ \mu$ are in qualitative accord with this hypothesis. However, if this were the only source of radiation, the spectrum would be expected to have a sharp cut-off on the low-energy side at an energy corresponding to $\epsilon_G = 1{\cdot}1$eV. Although the experimental data in this region are meagre, they do seem to indicate that such a cut-off does not exist. An additional mechanism is thus required, and it has been suggested that transitions of electrons or holes within the same band, with the emission of photons with energies up to $\epsilon_0 = 2{\cdot}3$ eV, could be such a second mechanism. It has not been explained how momentum conservation is maintained. Evidently both mechanisms are necessary to explain the complete spectrum.

The existence of discrete spots of light suggests strongly that complete breakdown occurs only in a few isolated places. The current density in these areas is probably much greater than the critical value J_0, and, if so, the avalanche

should approach the intrinsic type, discussed in sub-section 2·4, in which the potential distribution is determined mainly by the space charge associated with the current. If this is true also for small amounts of avalanche multiplication, as is suggested by noise measurements[16], the ionization coefficients calculated from multiplication experiments are open to some doubt.

3.5. Avalanche Injection

The phenomenon of avalanche injection has so far been described only in germanium[19] although unpublished data suggest that it occurs also in silicon. In order to observe avalanche injection it is necessary to ensure that the current



[*By courtesy of the editor, Physical Society.*

Figure 15. *Voltage–current characteristic of a P-I-N avalanche injection diode* (*Gunn*[19]).

passed before the avalanche begins is held down to a manageable value. Since the carrier velocity in these circumstances is necessarily equal to v_{d_s}, this condition can be met by a reduction either of the carrier density or of the area of the current flow. The necessary reduction of carrier density can be achieved in a layer of highly pure germanium from which the intrinsic carriers have been extracted by P^+ and N^+ electrodes biased, respectively, negatively and positively. A region of high field, in which there is in general a space charge, will be formed as in a normal P–N junction. As the bias voltage is raised, this region will spread outwards from one of the electrodes, and at some voltage dependent on the impurity content N_I and thickness X of the intrinsic layer, it will extend right across from electrode to electrode. If the layer is not too thick, this will occur before avalanche multiplication begins at some field E_a. The necessary condition is:

$$X < \frac{\kappa E_a}{qN_I} \qquad \ldots (87)$$

Provided this is satisfied, the onset of avalanche ionization at some higher voltage will lead to a distortion of the field, and so to the production of two intrinsic avalanches, one at each electrode. These inject carriers into the central part of the intrinsic layer, which consequently develops a low resistance. The voltage drop across the structure of a whole thus becomes nearly equal to that across two intrinsic avalanches in series, and a negative resistance characteristic should be developed, as discussed in sub-section 2.4.

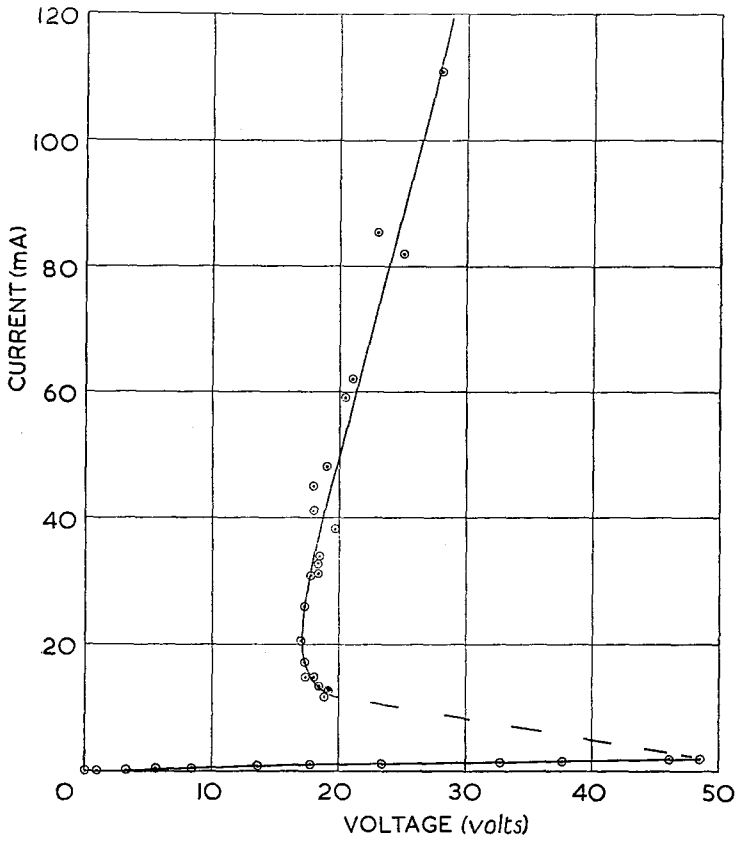
The current–voltage characteristic of a typical structure of this type is shown in Figure 15. While this exhibits roughly the features to be expected, it differs from the theoretical shape in two respects. Firstly, no negative resistance can be observed, but only a sudden decrease of voltage, in about 3×10^{-9} sec, as the current is increased. This results from the purely experimental difficulty of preventing the occurrence of relaxation oscillations. Secondly, the relationship $V \propto I^{-1/2}$ predicted by equation (60) is not satisfied, and at large currents the voltage remains practically constant. This is presumably a result of the failure of the assumption $\alpha = \alpha_1 \exp(\lambda E)$ (equation (53)) at very high fields of the order of 5×10^5 V cm^{-1} and above. By analogy with silicon, one might expect α to become nearly constant in this range, and if this happens the voltage V_a across an intrinsic avalanche should pass through a minimum with increasing J. As a transversely extended intrinsic avalanche is likely in any case to be unstable, on account of its negative resistance properties, this means that the avalanche will contract in area, so increasing J and decreasing V_a, until the minimum voltage V_s is reached. Changes in the current will then be accommodated by changes in area, with V_a remaining equal to V_s. Such a mechanism possesses obvious and striking analogies with the 'normal glow' in gas discharge theory.

The alternative approach to the observation of avalanche injection, that of decreasing the current-carrying area, leads logically to a structure like that shown in Figure 1. If biased so that minority carriers flow outwards, the resistance, at first equal to the spreading R_s, will rise with increasing voltage until the avalanche begins. When this happens there will be an immediate transition to an intrinsic avalanche and a decrease in voltage to a low value, as discussed in sub-section 2.5. This voltage should be equal to the sum of V_s for a single avalanche and the (normally negligible) drop across the spreading resistance which is greatly decreased by the conductivity modulation resulting from the injected carriers. Figure 16 shows the current–voltage characteristic of a structure of this type, which is atypical to the extent of possessing an abnormally high series resistance. The voltage is normally almost constant at about 13–17 V for large currents. Again, the negative resistance region cannot be explored over most of its range, although there is now definite evidence of its existence. With this structure it is possible to show conclusively that no injection occurs for currents below the break in the characteristic, and a large amount above it[19]. This, together with the rapid switching time quoted above, is strong evidence in favour of the theory of avalanche injection against other possible mechanisms.

According to the theory, the sustaining voltages of the characteristics shown in Figures 15 and 16 should be $2V_s$ and V_s respectively, and this is indeed approximately true if V_s is taken as 15 V. Unfortunately, the data for α in germanium do not extend far enough for a calculation of V_s to be made from them.

3.6. The Resistance of a Current Constriction

The theory outlined in sub-section 2.5 was first applied to the reverse resistance of a rectifying point-contact between a germanium surface and a metal wire, in an endeavour to overcome some of the difficulties experienced with the



[*By courtesy of the editor, Physical Society.*

Figure 16. *Voltage-current characteristic of an avalanche injection diode with radial flow (Gunn*[19]*).*

orthodox theories involving a P–N junction or potential barrier at the surface. The major experimental facts to be explained are:

(1) The resistance is relatively insensitive to surface conditions by comparison with area contacts between metal and germanium;
(2) The reverse resistance decreases with decreasing germanium resistivity;
(3) The current usually increases at first rapidly with increasing bias V, becomes roughly saturated at about $V = 5$–10 V, and then rises more slowly, approximately as V^2;
(4) When used as a transistor collector, the current multiplication properties are more like those of a filamentary than a junction transistor.

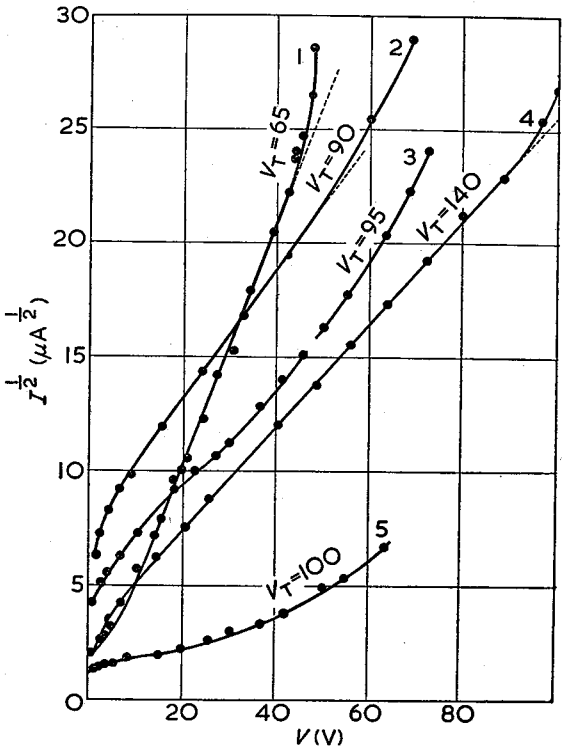


Figure 17. *Voltage–current characteristics of commercial point-contact diodes (V_T = turnover voltage). Curve 5 for a very heavily formed type.*

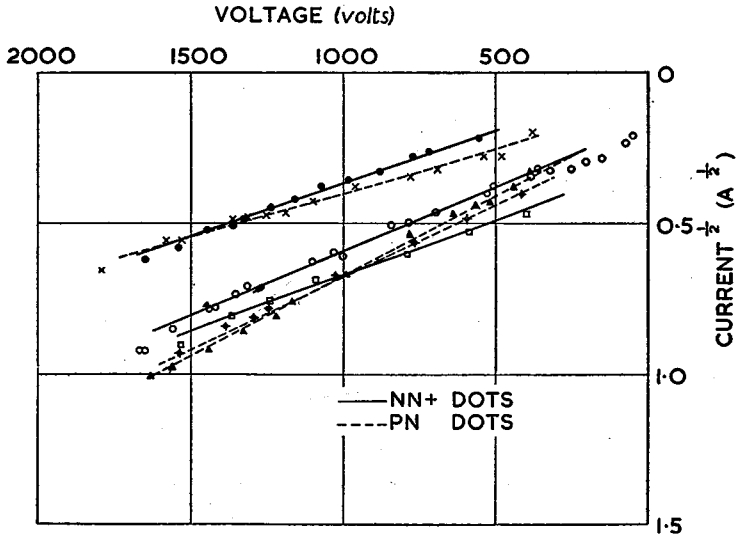


Figure 18. *Voltage–current characteristics of small alloy junctions (Paige[35]).*

Most of the available theories predict results which are basically rather different from these, and the differences are accounted for by the introduction of suitable *ad hoc* assumptions, many of them of a very special nature. While it does not give a complete and exact description of point–contact properties, the current-constriction theory seems to account quite naturally for the major features without the introduction of assumptions which cannot be verified[20]. It does appear, however, that surface effects such as channel formation may play an important, though secondary, role. For current flow in the forward direction the theory cannot account for the low resistance which is frequently observed, even with the assumption of avalanche injection. This last mechanism may, however, be important in those cases where the contact, though rectifying, has a relatively high forward resistance.

The verification of the theory by direct experiments on point–contacts is difficult because of their unreproducibility and the uncertainty in the contact area. In general, however, the current–voltage characteristic conforms quite well to equation (68) over a substantial voltage, in that a plot of $I^{1/2}$ against V usually gives a straight line with a negative intercept (Figure 17). The existence of fields at least as high as 3×10^4 V cm^{-1} near the contact has also been demonstrated directly[26]. In order to study the consequences of the purely geometrical features of the point-contact structure, Paige[35] has made measurements on alloy junctions of extremely small diameter having a known geometry and junction type (P–N or N–N$^+$). With these he had been able to show that, for currents greater than I_0; the resistance is indeed independent of junction type and obeys equation (68) quite accurately with values of E_1 and E_2 which are in agreement with those obtained by direct measurement[22]. Some typical results are shown in Figure 18.

4. CONCLUSION

The present position with respect to the study of high field effects in semiconductors may therefore be summarized as follows. The major effects observed—the non-linear relation between carrier velocity and field, the avalanche processes and the properties of small area contacts—have been explained at least qualitatively. Some of the new devices that can be developed using these effects have been indicated and no doubt the continuing development of these devices will stimulate further study of high electric field phenomena.

ACKNOWLEDGEMENT

The author is indebted to Dr. J. M. Radcliffe for helpful theoretical discussions.

REFERENCES

1. L. Landau and A. Kompanejez. *Phys. Z. Sowjet.* **6,** 163 (1934).
2. E. Guth and J. Mayerhöfer. *Phys. Rev.* **57,** 908 (1940).
3. F. Seitz. *Phys. Rev.* **76,** 1376 (1949).
4. W. Shockley. *Bell Syst. Tech. J.* **30,** 990 (1951).
5. E. M. Conwell. *Phys. Rev.* **88,** 1379 (1952).
6. E. M. Conwell. *Phys. Rev.* **90,** 769 (1953).
7. J. Yamashita and M. Watanabe. *Prog. Theor. Phys.* **12,** 443 (1954).
8. M. Shibuya. *Phys. Rev.* **99,** 1189 (1955).

9. W. Shockley. *Electrons and Holes in Semiconductors.* (Van Nostrand, New York, 1950).
10. R. J. Collins and H. Y. Fan. *Phys. Rev.* **93**, 674 (1954).
11. Y. S. Hsieh. *J. Chem. Phys.* **22**, 306 (1954).
12. F. Seitz. *Phys. Rev.* **73**, 549 (1948).
13. P. A. Wolff. *Phys. Rev.* **95**, 1415 (1954).
14. S. L. Miller. *Phys. Rev.* **99**, 1234 (1955).
15. C. Zener. *Proc. Roy. Soc.*, **145**, 523 (1934).
16. K. G. McKay. *Phys. Rev.* **94**, 877 (1954).
17. C. Herring. *Bell Syst. Tech. J.* **34**, 237 (1955).
18. J. Bardeen and W. Shockley. *Phys. Rev.* **80**, 69 (1950).
19. J. B. Gunn. *Proc. Phys. Soc., Lond.* B, **69**, 781 (1956).
20. J. B. Gunn. *Proc. Phys. Soc., Lond.* B, **65**, 908 (1952).
21. J. B. Arthur, A. F. Gibson and J. W. Granville. *J. Electronics* **2**, 145 (1956).
22. J. B. Gunn. *J. Electronics.* **2**, 87, (1956).
23. J. B. Gunn. *J. Sci. Instrum.* **33**, 364 (1956).
24. M. B. Prince. *Phys. Rev.* **92**, 681 (1953).
25. H. J. McSkimin. *J. Appl. Phys.* **24**, 988 (1953).
26. J. B. Gunn (unpublished work).
27. E. J. Ryder. *Phys. Rev.* **90**, 766 (1953).
28. A. Leblond. *C.R. Acad. Sci., Paris.* **242**, 1856 (1956).
29. K. G. McKay and K. B. McAfee. *Phys. Rev.* **91**, 1079 (1953).
30. E. J. Ryder, I. M. Ross and D. A. Kleinman. *Phys. Rev.* **95**, 1342 (1954).
31. E. M. Conwell. *Phys. Rev.* **94**, 1068 (1954).
32. R. Newman. *Phys. Rev.* **100**, 700 (1955).
33. A. G. Chynoweth and K. G. McKay. *Phys. Rev.* **102**, 369 (1956).
34. J. R. Haynes and H. B. Briggs. *Phys. Rev.* **86**, 647 (1952).
35. E. G. S. Paige. *J. Electronics* **2**, 378 (1957).

THEORIES OF ELECTROLUMINESCENCE

D. CURIE, D.Sc.

Maître de Recherches, Laboratoire de Luminescence Faculté des Sciences, Paris, France.

MS. received 30 *July*, 1956

THEORIES OF ELECTROLUMINESCENCE

1. INTRODUCTION

On the application of an electric field to luminescent solid crystals, different kinds of luminous effects have been observed. We shall deal with the following types:

(1) 'Pure' or 'intrinsic' electroluminescence (Destriau effect). Excitation of luminescence of phosphors by the sole action of a field, as obtained with electroluminescent cells in which a suitable phosphor powder is embedded in an insulator and submitted to an A.C. field (Figure 1). This effect may also be observed in a phosphor placed in an electric field without any contact with the electrodes and thus no direct current passing through the crystal.

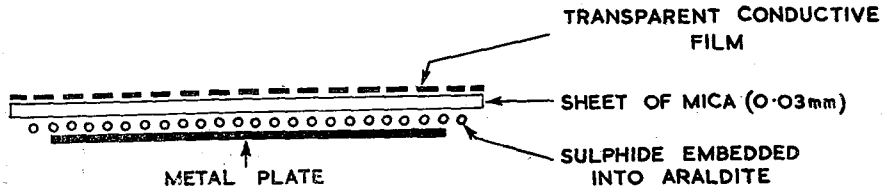


Figure 1. *An electroluminescent cell used in the Laboratory.*

(2) 'Carrier injection' electroluminescence. Emission of light produced by current injection across a rectifying structure, as obtained when a 'cat's whisker' is put in contact with a crystal, or in the vicinity of a P–N junction.

(3) Electrophotoluminescence (Gudden–Pohl effect). Momentary surillumination of a phosphorescent substance having been previously submitted to photo-excitation. Other effects, *e.g.* the quenching or enhancement of photoluminescence while an electric field is applied during the photoexcitation, will be discussed in the same section.

Professor Destriau has given two comprehensive reports on electroluminescence and related topics[1, 2], in which the reader will find an extensive number of references, mainly of the experimental points of view. This present paper is devoted to some theoretical aspects of the problem (see also Williams[3, 4]).

2. 'PURE' OR 'INTRINSIC' ELECTROLUMINESCENCE OF PHOSPHORS

2.1. Mechanism of Electroluminescence

The impact acceleration mechanism, proposed by Destriau[5], is generally accepted. Curie[6,7] has studied what is implied by this mechanism and has proposed a three-step excitation process, as follows:

(1) The supply of electrons from shallow donor levels into the conduction band, under the action of an applied electric field ('ionization' process, or 'process 1');

(2) The acceleration of these electrons in the conduction band; pre-breakdown avalanches ('electron chains') may occur ('acceleration' process or 'process 2');

(3) The collision of the high energy electrons with luminescence centres ('excitation' process or 'process 3'). This excitation may be followed by light emission with an efficiency η.

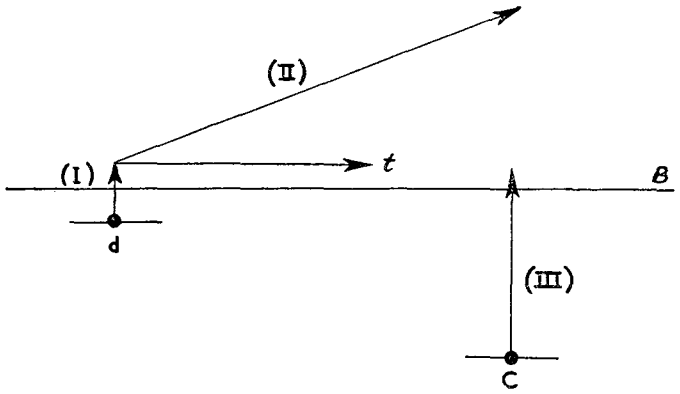


Figure 2. *The impact acceleration mechanism of electroluminescence. B = conduction band, d = shallow donor level, c = light centre. (I) Field ionization of the donor. (II) Possible acceleration of the ejected electron (t = trajectory of an electron which remains thermal). (III) Ejection of the electron initially in the centre, resulting from the impact of the accelerated electron.*

Let $f_1(E), f_2(E), f_3(E)$ be the respective probabilities of these three processes. The elementary brightness–field relation B (E) is thus:

$$B = f_1(E).\, f_2(E).\, f_3(E).\, \eta(E) \qquad \ldots (1)$$

The observed brightness–voltage law would be derived from the preceding one by an integration on the different donor levels involved (see Williams[8], and many other authors), and an averaging process on the different parts of the crystal (in which the local field is not the same) and the different components of the brightness wave[9]. This rather intricate problem has not yet been solved, but it seems that B depends mainly on the excitation term f_3, whose increase with E is so rapid (exponential) that a convenient fit with the experiments is obtained, often, by neglecting the other terms:

$$f_3 \simeq \exp(-b/E) \qquad \ldots (2)$$

where $b \simeq$ constant:

$$B \simeq f(E).\, \exp(-b/E) \qquad \ldots (3)$$

where $f(E) \simeq$ constant (Destriau[5]). In a better approximation, $f(E)$ would, of course, be an increasing function of E; the exact law $B(E)$ is probably different for different phosphors.

According to such a formula, there is no definite voltage threshold for electroluminescence: The observed threshold depends on the sensitivity of the detection apparatus.

The relation between the local field E and the applied voltage V is most intricate. Destriau[1] and Matossi[10] have studied the effect of the internal polarization of the phosphor. The field E acting in the crystals is out of phase with the field E_0 round the crystals and leads it by an angle φ:

$$\tan \varphi = \frac{2}{\kappa \rho f} \quad \ldots (4)$$

where κ is the dielectric constant of the crystals, ρ their resistivity and f the frequency.

The field values are related by:

$$E = E_0 \cos \varphi \quad \ldots (5)$$

The angle φ appears on the waves of brightness (Figure 5). Roberts[11] and Destriau[12] have given formulae which take into account the dielectric constant of the insulator.

2.2. The Excitation Process

Let us first discuss the excitation of the light centres (f_3), which is, according to this theory, the dominating process. We accept in this sub-section the existence of accelerating electrons as an experimental fact. If an electron of charge q travels a path of length x in the direction of the field since the beginning of the acceleration, its kinetic energy in the conduction band will be:

$$\epsilon = qEx \quad \ldots (6)$$

Let W be the required energy for excitation or ionization of the centres, the electron must gain an energy $\epsilon > W$, whence $x > l$ where

$$W = qEl \quad \ldots (7)$$

Let $\bar{x}$ be the average acceleration path; the proportion of electrons whose path exceeds l, *i.e.* those able to excite the centres, is according to a well-known formula:

$$f_3 = \exp(-l/\bar{x}) = \exp(b/E) \quad \ldots (8)$$

where:

$$b = W/q\bar{x} \quad \ldots (9)$$

The expression (8) relies upon the assumption of a random distribution of x-values in the range extending from 0 to ∞.

Formula (9), in connection with formula (3), allows us to derive the average path $\bar{x}$ when the experimental law $B(E)$ is known. $\bar{x}$ is usually of the order of 10^{-5} cm; l is 10 times greater, *i.e.* $f_3 \sim 1/20{,}000$.

The study of the depths of emptied traps during the Gudden-Pohl effect leads to paths $\bar{x}$ of the same order[6, 13].

These paths are much larger than the widths of the centres and the traps. This fact gives us one reason for excluding the hypothesis of a direct excitation of the centres or traps by the field. Another reason for introducing the conduction band is that well-crystallized materials are needed so as to observe electroluminescence (this condition seems to be necessary but by no means sufficient).

The mean path $\bar{x}$ is the path of an accelerating electron before its energy loss. This path is ruled by the formula[7, 14]:

$$\frac{1}{\bar{x}} = \sum_i \frac{1}{\bar{x}_i} \qquad \ldots (10)$$

where i includes all the energy loss possibilities (*i.e.* lattice scattering by optical and acoustical waves, neutral and ionized impurity scattering), except electroluminescence[7]. The energy loss produced by electroluminescence must be considered indeed as a rare accident ($l \gg \bar{x}$). According to Curie[7], the most

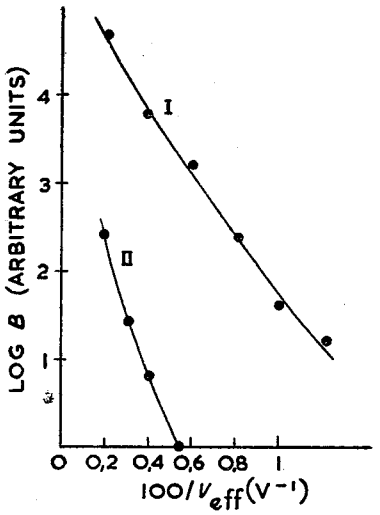


Figure 3. *Brightness versus voltage relation for a ZnS(Cu) (curve I) and a ZnS(Mn) (curve II) studied by Destriau*[1].

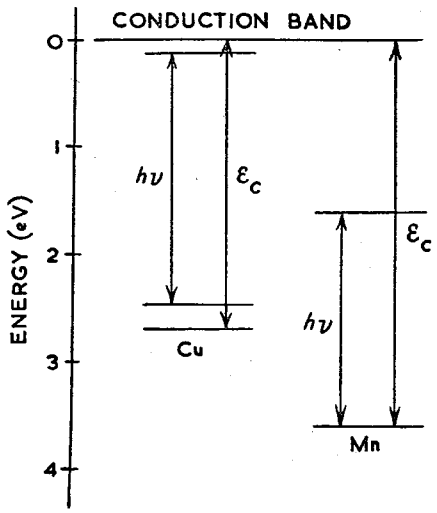


Figure 4. *A possible explanation for the higher values of b in manganese-sulphides (see text). The exact place of the manganese-level is not yet known, but it seems to be situated inside the valence band.*

important event to be considered in summation (10) is the energy loss on a trap or a defect; the above figure for $\bar{x}$ leads to a cross-section $\sigma \sim 10^{-12}$ cm^2. The capture cross-section of the trap is much smaller.

$\bar{x}$ depends possibly on the field E (Nagy[14]); it depends also on the sample temperature T (see sub-section 2.4).

Some light may be thrown on the nature of the centres by formula (9); but such studies have not yet been systematically attempted. For copper and silver activated zinc sulphides, the centres are ionized, W is probably to be identified with the depth ϵ_c of the centre below the conduction band and is little different from the emitted light quantum $h\nu$; this assumption is made in the determination of $\bar{x}$. But for manganese-activated sulphides (Figure 3), the experiments give systematically larger values of b, while $h\nu$ is correlated to yellow-red photons instead of green or blue ones.

I suggest the following scheme as an explanation (Figure 4). The emitted quantum $h\nu$ is ascribed to a transition between localized excitation levels of the Mn ions, in electroluminescence as well as in photoluminescence; the energy loss W is probably a kind of average between $h\nu$ and ϵ_c.

On the other hand, for blue and green centres in ZnS(Cu), b, and thus W, seems to vary in the same way as $h\nu$[14].

In the case of manganese, the energy loss W may occur during direct impact excitation, in the same way as for Cu or Ag centres. A quite different competitive excitation process has, however, been suggested by F. E. Williams in the Paris Symposium on Luminescence, *i.e.* resonance transfer excitation by means of electron–hole pairs. The occurrence of this mechanism, in which the potential energy of the free electron would be converted into light instead of its kinetic energy, concurrent with the above impact mechanism, would of course also explain some differences: W would be the energy required for electron-hole pair production. Figure 4 refers to the case of direct impact excitation.

The study of the brightness waves gives some evidence in agreement with the assumption of different excitation mechanisms according to the nature of the activator Mn or Cu. Figure 5 shows the contrast between the simple form of the brightness waves obtained with manganese sulphides and the rather intricate form obtained with copper- or silver-activated sulphides.

The secondary peak is ascribed[7] to the re-capture into the ionized copper or silver centres of electrons that have travelled in the sample far from their starting point. Zalm, Diemer and Klasens[15] have thus considered carriers captured by the traps; experiments by Destriau[2] seem however to show that the secondary peak appears only when crystal–crystal contacts occur. But at all events a non-symmetrical waveform seems to characterize the case where ionization of centres leads to radiative recombinations (and it is then the general rule), while simple waves are always obtained with manganese-activated zinc sulphides in which excitation alone leads to radiative recombinations.

2.3. **Kinetics of Electroluminescence**

In a photoconducting phosphor, the collision of the accelerated electron with the centre is generally accompanied by the ejection of the electron initially in the centre, because the latter electron receives from the former an impulse in the direction of the field[7]. Thus the ejected electron leaves the centre in a very short time—less than 10^{-13} sec if the electron speed is about 10^7 cm sec^{-1} compared to the life necessary for light emission 10^{-8} sec.

An experiment by Zalm[16] reinforces this conclusion. The study of the brightness waves in an oscillating field of constant direction (oscillating between 0 and a given value) shows that with ZnS(Cu) the light emission appears when the field is zero and the conduction electrons are returning into the excited part of the crystal for recombination with the empty ionized centres.

Quite different are the waves given by activators of the manganese-type; in this case, the maximum emission occurs when the applied voltage is maximum (with a shift due to the life of the excited state).

The return of electrons into ionized centres in electroluminescence is thus 'bimolecular'.

The kinetics of electroluminescence may be fundamentally different from the

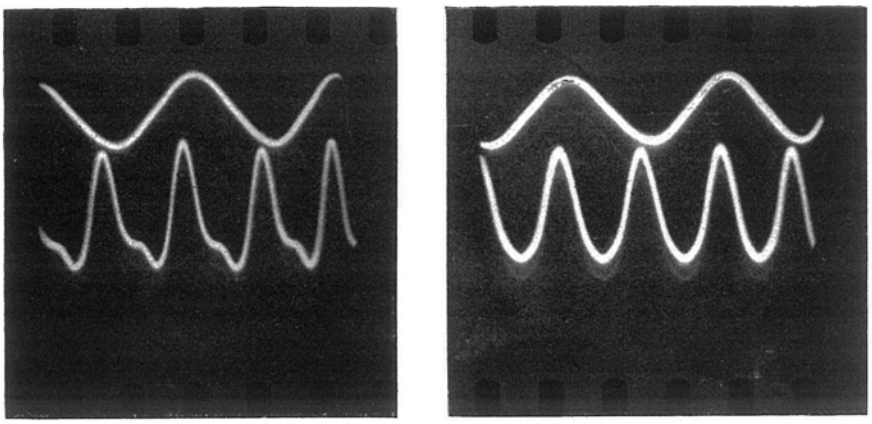


Figure 5. *Brightness waves obtained with an A.C. sinusoidal voltage for a ZnS(Cu) and a ZnS(Mn) studied by Mattler*[9].

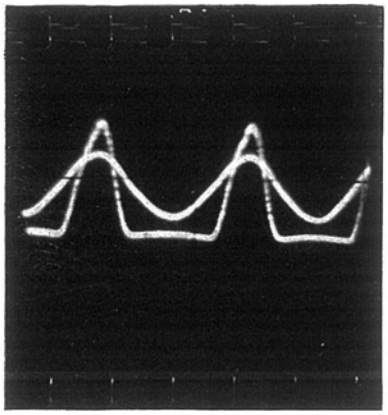


Figure 13. *Brightness waves shown by a SiC rectifier*[55]. *Sinusoidal wave: voltage. Distorted wave: brightness. Note the lack of dephazing. Light appears only when the bulk of the crystal is positive, the entrance of electrons in it being thus allowed (forward direction).*

Facing p. 255]

kinetics of phosphorescence, even on the same phosphor; lattice and impurity scattering has less influence on high speed electrons and therefore the electronic displacements are much larger in electroluminescence. Table 1 summarizes some conclusions[7, 17].

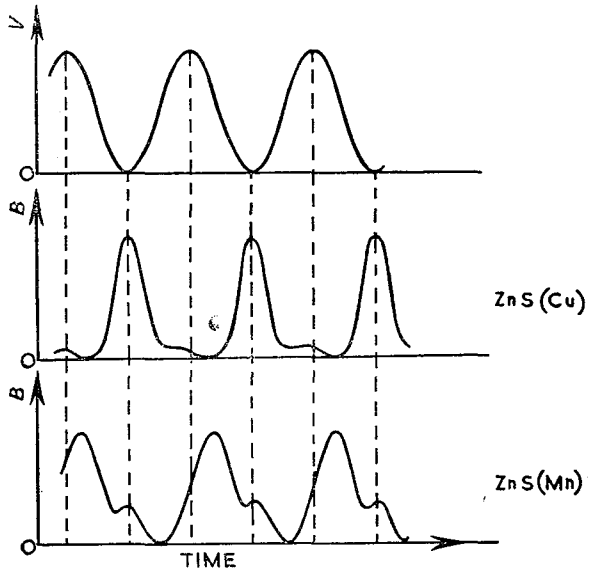


Figure 6. *Brightness waves in a field of constant direction (Zalm[16]). V = Applied voltage, B = brightness.*

Table 1. The kinetics of the luminescence processes change from 'monomolecular' to 'bimolecular' as the electronic displacements become larger.

Studied phenomenon	Electronic paths	Approximate kinetics of the phenomenon
Long-decay phosphorescence; glow experiment.	Small ($< 10^{-6}$ cm)	Monomolecular
I.R. stimulation; photoluminescence during excitation.	Intermediate	Intermediate
Perfect crystals; electroluminescence; carrier injection	Large ($> 10^{-5}$ cm)	Bimolecular

The assumption of a hyperbolic decay (bimolecular kinetics) leads to the following theoretical formula for the frequency dependence of the electroluminescent brightness[7]:

$$\frac{1}{B} = \alpha + \frac{\beta}{f} \qquad \ldots (11)$$

where α, β are constants.

The derivation of this formula uses rather over-simplified considerations; however, the experimental fit is satisfactory (Figure 7).

The ratio β/α depends on the time constant for the radiative recombinations; when it is very short (the case of the blue band), β is much larger than α and thus:

$$B \sim f \qquad \ldots (12)$$

Destriau[19] has developed direct methods for computing these time constants. The results he obtains are averages of the results for electrons in the conduction band and in the shallow traps, respectively.

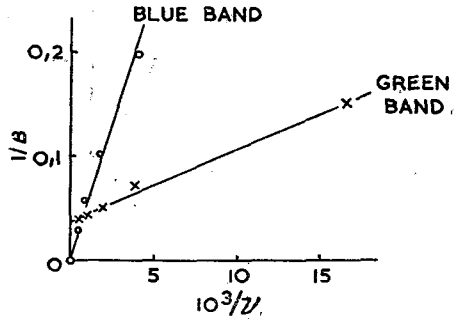


Figure 7. *Variation with frequency of the blue and green emission bands of an electroluminescent ZnS(Cu). Experimental points: Measurements by Waymouth et al.*[18]. *Straight lines: computed from formula*(11).

2.4. The Supply of Electrons to the Conduction Band

The ionization process of luminescent centres (and also of non-luminescent donor levels) may result in a 'chain' increase of the number of accelerating electrons. These chains are far from having the magnitude of the breakdown 'avalanches'. They seem, however, to be able to bring about an important multiplication of the number of free electrons (let us say a factor of 1000). The cell must, of course, have sufficient width to enable the chains to reach such a multiplication factor.

For some samples of willemite or ZnS, introduced in an insulating oil and submitted to an intense field (20,000 V across 2 cm) without contact with electrodes, a thickness of at least 1 mm was found necessary to observe light emission.

Other samples were illuminating even in small grains: this may be explained by a more intense supply of primary electrons[7].

Let us now study where these primary electrons come from.

Are they brought into the conduction band by the field, or by thermal activation? The studies of the temperature effect on brightness (see Mattler[20], and other authors) show the absence of thermal activation, when supply is due to the field.

In some cases, an increase of temperature seems to help the field ionization of donors[21], *e.g.* the low temperature part of the ZnS(Cu) curve, but the slope is too small to be ascribed to a thermal activation effect only. The ionization

probability for a donor whose depth is ϵ depends on the ratio:

$$p \sim \exp\left[-\frac{\epsilon - f(E)}{kT}\right] \qquad \ldots (13)$$

where $f(E)$ is a function of the field. Different analytic expressions may be suggested[7] for $f(E)$. This hypothesis has been considered in some detail in a recent paper by Thornton[21].

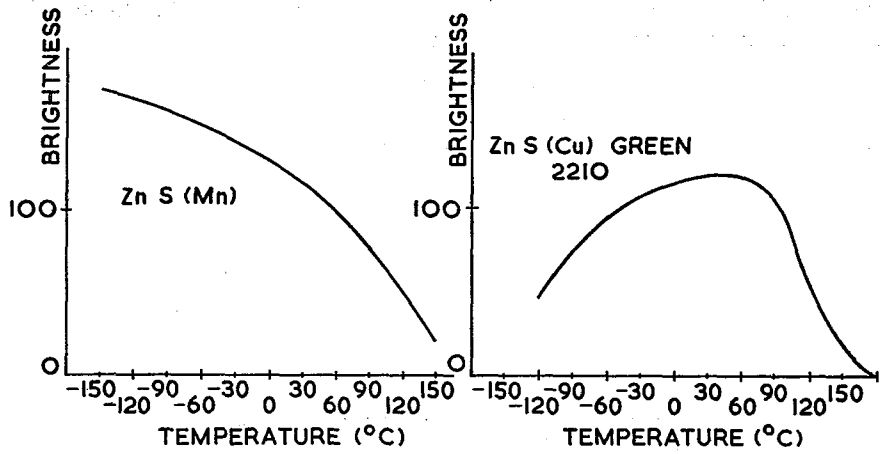


Figure 8. *Dependence of the electroluminescent brightness on temperature (from Mattler[20]). For low temperatures, the electronic paths are large and thus the electron acceleration is easy; on the other hand the electron supply is probably easier for higher temperatures. Both effects are competitive. Thermal quenching also occurs.*

The increase of the electroluminescent brightness with temperature, predicted by this mechanism, occurs often indeed, especially in the low temperature region, but a decrease is also observed. It is sometimes possible to ascribe this decrease to thermal quenching, but this interpretation seems doubtful when it occurs below room temperature. The mean acceleration path $\bar{x}$ is probably larger for low temperatures, but a simple expression, such as $x \sim T^{-1}$, would lead to:

$$b \sim T \qquad \ldots (14)$$

from formula (9), and although Zalm[16] has indeed observed variations of b with temperature, these variations are much more complex.

Figure 8 shows two characteristic forms of the brightness *versus* temperature dependence, which may be ascribed to the competition between the two above effects. It has been suggested[22] that, as an indirect action of temperature, thermal ionization of traps may result in space charge modifications resulting, in turn, in a change of the local field for a given applied voltage. The filling and the ionization of traps seems also to be correlated with electro-thermoluminescence (Gobrecht, Hahn and Gumlich).

The assumption of donor levels shallower than the centres in the sulphides, so as to be directly emptied by the field, has been proposed independently by Piper and Williams[23], and Curie[24], It is supported experimentally by the study

of conduction of single CdS crystals[25] in the dark at low temperature, which has shown the existence of such levels. According to Curie[7], the depth of these levels would be, in ZnS, of the order of 1 eV, *i.e.*, larger than the traps depths. The different calculations[26, 27] of the direct field ionization of localized levels agree that a field higher than 10^6 V cm^{-1} is required. According to Williams[8], Lanczos's calculations[26] for hydrogen atoms do not need correcting for the

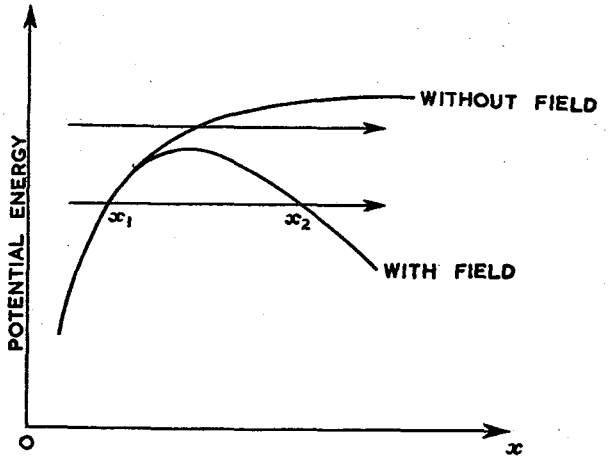


Figure 9. *Field Ionization of a localized level*[26]*: A quantum mechanical tunnel effect through the barrier* (x_1 x_2) *occurs* (*pre-ionization*), *or a passing over the barrier* (*ionization*).

surrounding medium, if the levels involved are sufficiently localized; the critical field for ionization of donors would then be:

$$E_c = 2{\cdot}4 \times 10^6 \, \epsilon^2 \text{ V cm}^{-1} \qquad \ldots (15)$$

where ϵ is the depth of the donor in eV. Franz's formula[27] takes the band gap ϵ_G into account; he obtains for a square well donor potential:

$$E_c = \frac{10^7 \, \epsilon_G^{1/2}}{8 + \log E_c \tau} \epsilon^{3/2} \qquad \ldots (16)$$

Values of the same order as Lanczos's are obtained if one chooses τ, the decay constant of the level, between 1 msec and 1 microsec. Fields of this magnitude are not available in the bulk of the phosphor, but probably do exist in the cathode potential barrier.

A Mott–Schottky barrier, with appropriate assumptions on the donor distribution, has been first considered by Piper and Williams[23] and by Diemer, Zalm and Klasens[15, 28] (see also Fischer[29], and Howard[30]). An intrinsic potential barrier[7] seems to exist in the case of electroluminescent samples illuminating without contact with the cathode. When there is a contact, however, a Mott–Schottky barrier appears; this barrier is more important than the intrinsic one, as grains of substance deposited on the metal electrodes are generally brighter than identical grains deposited between the electrodes.

Let N_S be the number of surfaces states involved; the intrinsic barrier would be as important as the Mott–Schottky if:

$$N_S \sim N_D d \qquad \ldots (17)$$

where N_D is the donor concentration per cm^3 and d is the barrier width. If we put:

$$N_D \sim 10^{18} \text{ cm}^{-3} \text{ and } d \sim 5 \times 10^{-5} \text{ cm}$$

then equation (17) leads to

$$N_S \sim 5 \times 10^{13} \text{ cm}^{-2}.$$

Thus to take the smaller, but not negligible, part played by the intrinsic barrier into account, we must assume:

$$N_S \sim 10^{12} \text{ to } 10^{13} \text{ cm}^{-2} \qquad \ldots (18)$$

According to the donor distribution assumed, different formulas have been derived for the brightness *versus* field dependence. But the results are not very sensitive to this distribution, and may be described by the expression (3), with different choices of the slowly varying function $f(E)$ before the exponential term.

It must be noted that even in the barrier, the field is possibly not high enough for direct ionization of the light centres.

2.5. The Acceleration Process

The required acceleration of conduction electrons is easier to understand if the phenomenon occurs in the high fields obtained in the cathode barrier rather than in the bulk of the material. The connection between the acceleration process in electroluminescence and similar processes occurring in breakdown theory has been independently recognized by Curie[6] and Piper and Williams[23] (see also Diemer[31]). It is well known that the magnitude of the pre-breakdown avalanches decreases rapidly with the applied field; these avalanches are practically undetectable when the field is below one-fifth of the breakdown strength. This last figure is as yet unknown in zinc sulphides, but it may be reasonably assumed that ordinary pre-breakdown avalanches require more than 10^5 V cm^{-1}. Such fields are produced in the barrier, at least in some spots[16], but the average field in the sample may be as low as 10^4 V cm^{-1}. According to the author, however, low fields may also produce electron acceleration, but the resulting avalanches (ascribed to donor levels and centres, not to valence band ionization) are quite different from the ordinary pre-breakdown avalanches.

Therefore two groups of theories must be considered; the first group assumes a surface barrier acceleration process[8, 15, 16, 23, 28, 29, 30]; the second one assumes a surface supply in primary electrons followed by a volume acceleration process[6, 7, 14]. In the first case the light emission would be localized in the immediate vicinity of the barrier; in the second case the light emission would also be localized in some relatively higher field spots with the use of applied voltages near the electroluminescence threshold, but would extend to the bulk of the sample when higher voltages are applied.

Such an extension of the enlighting region has been observed with willemite cells[7]. Destriau has indeed pointed out (private communication) that it may be derived from the formula (3). Let us look at two parts of the crystal A_1 and A_2,

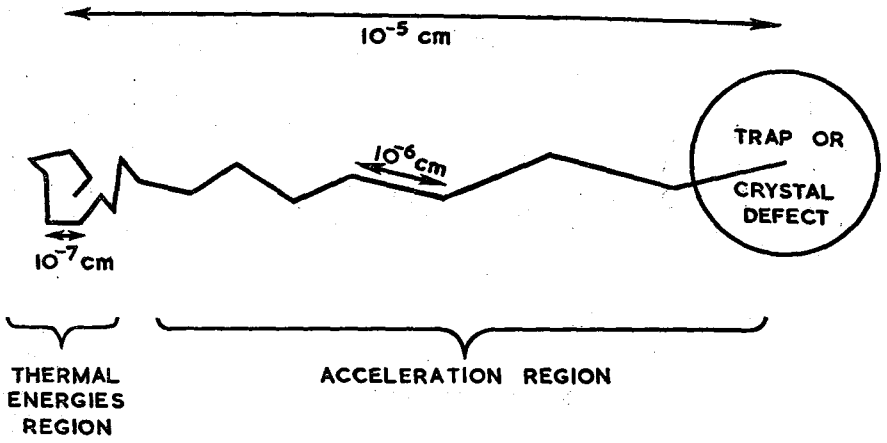


Figure 10. *The trajectory of a field-accelerating electron in the conduction band*[8].

in which, according to local inhomogeneities, the local fields are $E_1 = k_1 V$ and $E_2 = k_2 V$ with $k_1 < k_2$ (V = applied voltage). Let:

$$B_1 \sim \exp(-b/E_1) \text{ and } B_2 \sim \exp(-b/E_2) \qquad \ldots (3)$$

then the ratio

$$\frac{B_1}{B_2} \sim \exp\left[-\frac{b}{V}\left(\frac{1}{k_1} - \frac{1}{k_2}\right)\right] \qquad \ldots (19)$$

increases with V.

At first sight, microscopic observation of electroluminescent powders shows the occurrence of very bright spots[15, 32, 33], in contrast with the more regular appearance of the same powders irradiated with u.v., and therefore favours the barrier theory. The author, however, has observed such inhomogeneities only upon large imperfect crystals, smaller crystals seem to emit light from their bulk; in such crystals no bright spots seem to occur, so the uniform light appearance can hardly be ascribed to the diffusion of light.

Moreover, Destriau[34] has emphasized that microscopic observation enhances considerably the apparent brightness of the spots compared to the bulk. Let G be the magnifying power of the microscope; it is well known that the ratio of the clarities C_1 for a spot and C_2 for a continuum is:

$$\frac{C_1}{C_2} = G^2 \qquad \ldots (20)$$

The crystal is indeed punctual for direct eyesight but its image by observation across the microscope can cover many retinal cells (or many 'sensitivity specks' if the observation is made by means of a photographic plate); let N_r be the number of retinal cells covered by the whole crystal image, while we shall assume

for the sake of simplicity that one cell only is covered by the image of the bright spot. The ratio of the corresponding clarities is:

$$\frac{G_{\text{spot}}}{G_{\text{bulk}}} = N_{\text{r}} \gg 1 \qquad \ldots (21)$$

(not as large as expression (20) but > 10).

Thus the existence of localized bright spots is partly real and partly an optical appearance, while the bulk emission in some cases may not be seen across the microscope in view of the poor clarity of this apparatus.

Frankl[35] has pointed out that if a crystal which seems to show emission at localized regions is immersed in a liquid with a high optical refractive index, the emission becomes much more uniform.

Let us now study the two groups of theories previously distinguished.

(1) *The surface barrier theory*[8, 15, 16, 22, 23, 28, 29]. This hypothesis uses the extensive investigation made of the acceleration of electrons in the neighbourhood of breakdown fields (von Hippel[36], Seitz[37], Fröhlich[38], Seeger and Teller[39] *etc.*), and does not lead to any particular problem. Generally breakdown does not occur, partly as a result of the thinness of the high field barrier. (See, however, Diemer[31].)

The width of the natural barrier is:

$$d = \left(\frac{\kappa V}{2\pi N_{\text{D}} q}\right)^{1/2} \qquad \ldots (22)$$

where V is the applied voltage ($= 300$ V), N_{D} is the donor concentration ($= 10^{18}$ cm^{-3}) and κ is the dielectric constant of ZnS ($= 8$).
Thus:

$$d = 5{\cdot}1 \times 10^{-5} \text{ cm}$$

The maximum value of the field in the barrier is:

$$E = 2\,V/d \simeq 12 \times 10^{6} \text{ V cm}^{-1} \qquad \ldots (23)$$

The field E increases as $V^{1/2}$ and if expression (3) is used the brightness/voltage relation becomes:

$$B = f(V) \exp\left(-b/\sqrt{V}\right) \qquad \ldots (24)$$

This formula has been used by Zalm, Diemer and Klasens[28] and other authors[30, 40]. With some samples the experimental fit is rather good, but with others different simple formulae have been found better[14, 41, 42, 43]. Perhaps (Thornton[21]) expression (24) is valid at lower voltages, and the simple expression $\exp(-b/V)$ valid at higher voltages.

(2) *The surface supply and volume acceleration theory.* This theory assumes that a field not very much larger than 10^{4} V cm^{-1} is sufficient to produce electron acceleration, and therefore considers three zones in an electroluminescent cell:

(*a*) *A superficial zone* with a very high field (10^{6} to 10^{7} V cm^{-1}), in which the supply of electrons to the conduction band occurs;

(*b*) *An acceleration zone* with a (non-uniform) field above 10^{4} V cm^{-1}, where the electrons coming from the high field region may give rise to the light emission;

(*c*) *A dark zone* with a field lower than 10^4 V cm^{-1}, where the electron energies remain thermal.

The various experimental aspects of electroluminescence depend on the relative development of these three zones. The light emission appears at the surface or in the bulk, according to the applied voltage (see above) and to the construction of the cell, especially with reference to the shape of the electrodes and the magnitude of the conductivity of the dielectric in which the sulphide is embedded. Let ω be its resistivity; it may be seen from equation (4) and equation (5) that the bulk of the cell is dark for $\omega < 10^9\ \Omega$ cm (this is perhaps why electroluminescence is so rarely seen on cadmium sulphide) and emits light for $\omega > 10^{10}\ \Omega$ cm.

We shall now consider the possibility of electron acceleration with low fields (little above 10^4 V/cm). The discussion given in the papers[6, 7] uses the acceleration condition in the form:

$$qEl > h\nu\left[1 + \frac{2}{\exp(h\nu/kT) - 1}\right] \qquad \ldots (25)$$

where l is the free path for lattice scattering and ν the frequency of the lattice vibrations. For ZnS, $l \sim 10^{-7}$ cm for thermal electrons, whence $E > 10^5$ V cm^{-1}. Thus, in ZnS with the field of a few 10^4 V cm^{-1} used in electroluminescence, equation (22) is not satisfied for thermal electrons: these electrons remain generally thermal. In polar crystals, however, the free path increases with energy, so that an electron which, following scattering fluctuations, has gained an energy of the order of some tenths of an eV, has a free path $l \sim 10^{-6}$ cm sufficient to satisfy the acceleration condition of equation (22) with the applied field. Then this electron gains energy continuously, but stops on the average at the end of the path $\bar{x} \sim 10^{-5}$ cm (see equation (9)), which is the free path of the electron before its energy loss on a defect or a light centre.

We shall now discuss this acceleration process in another form, using Seitz's[37], Frohlich's[38] and Shockley's[44] formalism for 'hot' electrons in a crystal. This formalism relies upon the conception of a high electronic temperature depending on the field (see also Nagy[14], Goffaux[45]).

Let us first remark that if we put kT_e equal to the average energy of the accelerating electrons:

$$kT_e = qE\bar{x} \qquad \ldots (26)$$

expression (8) then becomes:

$$\exp(-l/\bar{x}) = \exp(-W/kT_e) \qquad \ldots (27)$$

therefore the probability distribution (8) is nothing else than one kind of formulation of Boltzmann's statistics. T_e is the electronic temperature, of the order of 4700° K for $\bar{x} = 10^{-5}$ cm and $E = 40{,}000$ V cm^{-1}.

Let us now study the acceleration condition:

$$\left(\frac{d\epsilon}{dt}\right)_{\text{field}} > \left(\frac{d\epsilon}{dt}\right)_{\text{phonons}} \qquad \ldots (28)$$

in order to explain the existence of such 'hot' electrons in the sulphide at the moderate fields used.

It will be useful to study first the case of a non-polar semiconducting material such as germanium. The free paths l are then very large (10^{-5} cm) even for low energies; they do not depend on ϵ, but the energy loss is relatively large:

$$\left(\frac{d\epsilon}{dt}\right)_{\text{phonons}} = \frac{\text{const. } \epsilon}{l/v} \quad \ldots (29)$$

$$\left(\frac{d\epsilon}{dt}\right)_{\text{field}} = \frac{q^2 l E^2}{mv} \quad \ldots (30)$$

If we consider a delta function distribution of energies, we put the two expressions (29) and (30) equal and we see that the (average) energy of the electrons is proportional to the field:

$$\epsilon = E \quad \ldots (31)$$

According to Seitz[37], in diamond, $\epsilon = 1$ eV for $E = 5000$ V cm^{-1}. Thus it is easy to accelerate charge carriers in non-polar crystals up to energies corresponding to an infra-red electroluminescence; but this is a stable state which is very difficult to trespass unless breakdown fields are applied. For $(d\epsilon/dt)_{\text{field}}$ decreases while $(d\epsilon/dt)_{\text{phonons}}$ increases with increasing ϵ; the effect of optical phonons makes this difficulty still greater.

In polar crystals on the contrary, the paths are small ($l \sim 10^{-7}$ cm) for low energies because optical polarization waves play their part even for thermal electrons. Thus thermal electrons are hardly accelerated. Even in the breakdown case, Fröhlich's calculations[38] show that the electronic temperature T_e is not much different from the lattice temperature T; Fröhlich describes the distribution of electrons in equilibrium with the phonons and donor levels at a given depth ΔV by:

$$\frac{1}{kT} - \frac{1}{kT_e} \simeq \frac{1}{\Delta V} \quad \ldots (32)$$

if $\Delta V = 1$ eV, $kT = 1/40$ eV, then:

$$kT_e = 1/39 \text{ eV}.$$

Since the path increases linearly with energy ϵ, while the energy loss at the end of this path remains the same $\sim h\nu$, an electron of high enough initial energy is continuously accelerated:

$$\left(\frac{d\epsilon}{dt}\right)_{\text{phonons}} = \frac{h\nu}{\tau} \sim \epsilon^{-1/2} \quad \ldots (33)$$

$$\left(\frac{d\epsilon}{dt}\right)_{\text{field}} = \frac{q^2 l E^2}{mv} \sim \epsilon^{1/2} \quad \ldots (34)$$

Let us look at the crossing point of the two curves $(d\epsilon/dt)$ phonons and $(d\epsilon/dt)$ field (Figure 11): its energy corresponds to a stable stationary state S of the electrons in non-polar crystals, but in polar crystals to an unstable state U. Thus, in a polar crystal, two groups of electrons must be distinguished: the most numerous one is composed of slow electrons, whose energies remain under the crossing point U, with a temperature $T_e \sim 300$°K (group I); on the other hand, an electron that has passed over that crossing point gains continuously energy from the field. These accelerating electrons form another group

(II); their temperatures are extremely high and would even be infinite (non-stationary state) without the interruption of the acceleration produced by the ionization of localized levels. This ionization may result or not in an electroluminescent emission, according to the nature of the sample.

However Goffaux[45] points out that the probability of a thermal electron reaching the Group II in a pure polar crystal is certainly extremely small.

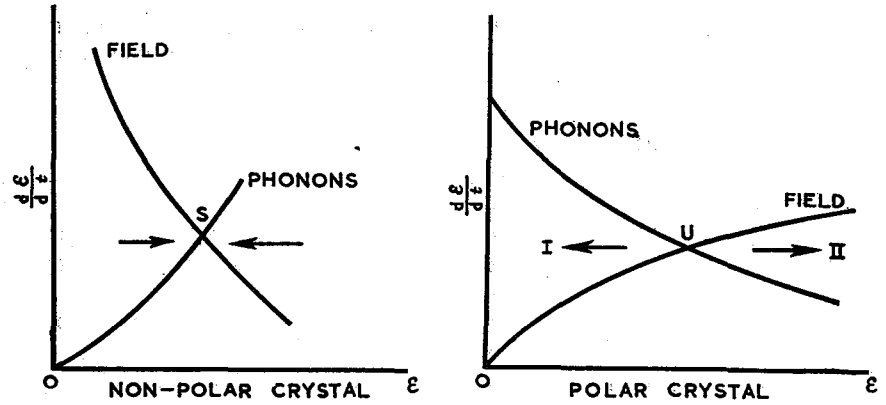


Figure 11. *In a non-polar crystal the average energy of electrons corresponds to the stable point S. In a polar crystal the crossing point U is unstable; the majority of electrons remains in region I; but an electron which, following scattering fluctuations, has attained region II, becomes continuously accelerated and may give electroluminescence.*

Zinc sulphide seems not to be a pure polar crystal, but its bonding possesses a strong covalent character. This fact possibly enhances considerably the probability of passing from one group to the other.

The effect of a strong magnetic field on electroluminescence[1, 46]. A quenching of electroluminescence by high magnetic fields is expected; an electron submitted to an electric field E and a perpendicular magnetic field H moves *in vacuo* along a cycloid whose extension along the E-direction is:

$$2R = 2\frac{m}{q} \cdot \frac{E}{H^2} \qquad \ldots (35)$$

and thus it cannot be accelerated above the energy:

$$\epsilon_{max} = qE.\,2R = 2m\frac{E^2}{H^2} \qquad \ldots (36)$$

If acceleration occurs in a field $E \sim 10^4$ V cm^{-1}, a magnetic field $H \sim 20{,}000$ oersted would give a strong quenching, while if the acceleration occurs in a barrier ($E \sim 10^6$ V cm^{-1}) a value one hundred times larger would be required. Destriau[1] used 60,000 oersted; Ince[46] was able to apply a magnetic field as high as 130,000 oersted. In both cases, no appreciable quenching effect was observed.

Ince's conclusion is that acceleration occurs thus in a barrier. However the author's opinion is that we do not at present know how to compute in a

satisfactory way the acceleration of an electron in both an electric field and a magnetic field. In the Hall effect too, the simple theory that considers the motion of a free classical electron in a vacuum is sometimes (*e.g.* in bismuth) in considerable disagreement with experiments. A high effective mass may explain the absence of quenching. Directional effects in the conduction band may also be most important: if the effective masses were so anisotropic that the motion of the electron practically occurs along one direction only, the transverse magnetic force would possibly not affect its acceleration. However this is, as yet, a mere speculation and Ince's experiment remains the most serious objection against the volume acceleration theory.

2.6. Some Conditions for Good Sensitivity

(1) A good crystallization is necessary.
(2) The defects of the crystal hinder the acceleration of electrons. Traps are such defects: willemite and the majority of ZnS phosphors giving a good Destriau effect have little afterglow.
(3) An electroluminescent substance may be formed by addition of donor levels, or by the creation of surface states or a surface barrier.

In this way the following 'recipes' are explained. Several successive heatings (superficial oxydization of the material) lower the electroluminescence thresholds A ZnS + ZnO burned compound may also be used, its luminescence being much better that that of the ZnS alone. In these two cases, ZnO acts by its shallow donor levels.

At the Symposium on Luminescence in Paris (May 1956), Lehmann described the formation of electroluminescent cells by mixing a non-electroluminescent phosphor with local heterogeneities (*e.g.* Cu_2S or another semiconducting or metallic powder), in the vicinity of which the field is enhanced. This enhancement results from geometrical considerations (refraction of field lines of force), even independently of any barrier formation.

Moreover, the usual methods of improving the luminescent response of phosphors may be modified in view of application to electroluminescence phenomena, *e.g.* the concentration optimum, and sensitization[47].

2.7. Electroluminescence of Organic Substances

Bernanose and his co-workers[48] have found that some organic substances (gonacrin, acridine orange) emit light when an electric field is applied. This phenomenon is ruled by the same empirical laws as inorganic electroluminescence; the formula (3) holds for the brightness *versus* voltage relation, waves of brightness are out of phase with voltage, *etc.*

However the value of b is somewhat higher and equation (9) leads to smaller electronic paths; $\bar{x}$ is found of the order of the molecular size. Thus the acceleration theory is perhaps still valid in the organic case.

But the higher values of b suggest the alternative possibility of a direct excitation of the molecule by the field. Bernanose[49] has discussed this mechanism in the following way: Let ψ_1 and ψ_2 be the molecular wave functions respectively in the normal and the excited state. The field E, applied in the x direction,

is considered as producing a small perturbation $-qEx$, thus the first order perturbed ψ may be written:

$$\psi = \psi_1 - \frac{\int \psi_1^* (-qEx)\, \psi_2 \, d\tau}{\epsilon_2 - \epsilon_1} \psi_2 \qquad \ldots (37)$$

$$\psi = \psi_1 - \frac{E\mu_{12}}{\epsilon_2 - \epsilon_1} \psi_2 \qquad \ldots (38)$$

μ_{12} is the matrix element of the dipole moment. The emission rate is thus:

$$\frac{E^2 \mu_{12}^2}{(\epsilon_2 - \epsilon_1)^2} \times \tau^{-1} \qquad \ldots (39)$$

where τ is the mean life of the excited state. Values of $\tau \sim 10^{-7}$ sec, whence $\mu \sim 10^{-18}$ e.s.u., lead to a satisfactory experimental fit.

CARRIER-INJECTION ELECTROLUMINESCENCE

3.1. Low Fields Phenomena

Injection of electrons into a P-type material, or of holes into an N-type one, produces a large number of recombinations; some of them may be radiative.

The effect shown by silicon carbide has been extensively studied. Lossew[50, 51] observed two types of luminescence:

(1) 'Luminescence I', which is blue-green and occurs in localized spots when the crystal is biased in the blocking direction.

(2) 'Luminescence II', which is generally yellow and extends over the surface of the crystal near the electrode when the crystal is biased in the forward direction.

'Luminescence I' is ascribed to field emission phenomena. 'Luminescence II' is quite different. Lehovec[52] *et al.* proposed the theory for 'Luminescence II' accepted at present.

The inside of the crystal is generally P-type, thus the activators are normally empty. The samples which exhibit luminescence II are assumed to be N-type at their surface. The forward direction thus corresponds to the surface negative and is accompanied by electron injection in the P-type region (hole injection in the N-type region also occurs). One observes therefore the emission spectra of the activators (Figure 12). The nature of these activators is as yet unknown; Schön[53] and Szigeti[54] have experimentally derived their energy levels spectrum.

In contrast with 'intrinsic' electroluminescence, high fields are not required, but a current must pass through the crystal, while intrinsic electroluminescence is observed even with a crystal not placed in contact with electrodes. The light output B is a linear function of the current I with a small threshold[52]:

$$B = aI - b \qquad \ldots (40)$$

Moreover, the brightness waves observed with an SiC rectifier do not show the out-of-phase character of the brightness waves in zinc sulphide electroluminescence[55] (Figure 13, on plate, facing p. 255).

Such carrier injection phenomena have also been observed[56, 57] in germanium or silicon, with artificially made P–N junctions. If the material is very pure,

direct band-to-band electron–hole recombination occurs (a phonon is generally emitted in addition to the photon, in order to conserve the total impulse). The emission spectrum $\mathcal{J}(\nu)$ is conveniently described by the theoretical formula:

$$\mathcal{J}(\nu) = \nu^3 \, e^{-h\nu/kT} \, [1 - \exp(-\alpha_\nu d)] \qquad \ldots (41)$$

where α_ν is the absorption coefficient of the material and d is the thickness of the sample. The bracket takes into account the re-absorption of the emitted light in the emitting region and depends on the geometry of the specimen.

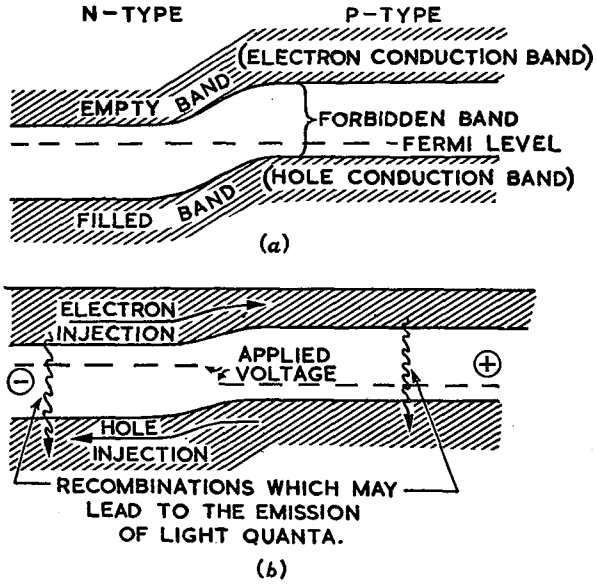


Figure 12. *Energy diagrams for a P–N junction without an applied field (A) and for current flow in the forward direction with subsequent injection of minority charge carriers (B). From Lehovec, Accardo, Jamgochian*[52].

As a result of the occurence of this term, the maximum of $\mathcal{J}(\nu)$ is obtained for:

$$h\nu \simeq \epsilon_G \qquad \ldots (42)$$

ϵ_G the energy gap ($\lambda \simeq 1{\cdot}8$ microns for germanium).

Injection phenomena may also be observed without using a P–N junction; if holes are injected through a metal point–contact into an N-type material, light appears in the neighbourhood of the contact. This was done by Aigrain[58]. The wavelength of the emitted radiation was found to be $\lambda > 2$ microns (between 2 and 6 microns). The phenomenon was first ascribed to electron–hole recombination in two steps on crystal defects, but now seems more probably due to hole acceleration electroluminescence (see sub-section 2.5, a short discussion of carrier acceleration in non-polar crystals). Light trapping effect are most important in germanium and may be avoided by an appropriate choice of the shape of the specimen (Weierstrass sphere with the cat's whisker placed at the Weierstrass point).

Injection electroluminescence has also been observed upon Group III–V semiconductors[59, 60]. The reported brightness waves are in phase with voltage.

The Kinetics of the Radiative Recombination

The radiative electron–hole recombination may be reasonably assumed to be 'bimolecular'[61, 62] in the sense of this word used in Ref. 17, *i.e.* that according to the large electronic paths in germanium (10^{-5} cm) and the occurrence of a

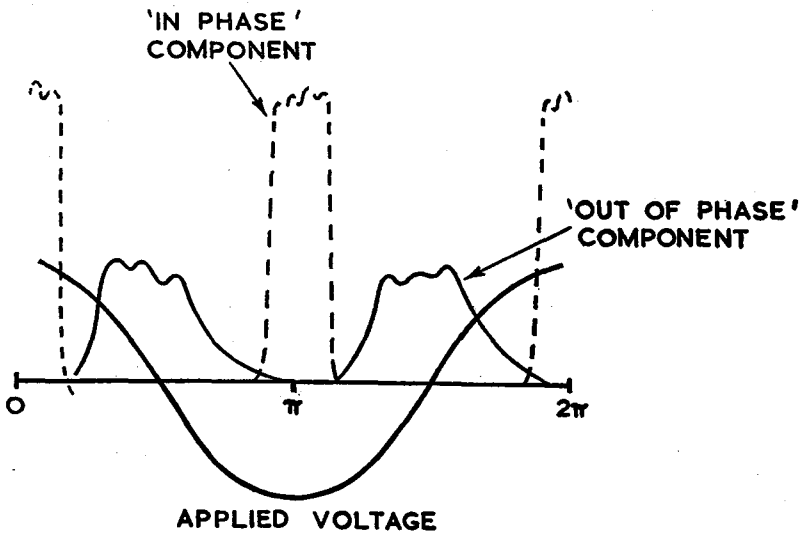


Figure 14. *The 'out of phase' and the 'in phase' components obtained by Piper and Williams*[23].

field, an electron may recombine with a large number of holes or centres. But if the majority carrier concentration is very large, the majority carrier density may appear as a constant and the bimolecular mechanism will result in an approximate first order kinetics depending on the minority carrier density[57]. The brightness B is thus proportional to the current and not to its square: see formula (40). High carrier density favours band-to-band radiative recombinations[4], low carrier density favours recombinations on crystal defects (activators in SiC, non-radiative centres in germanium).

3.2. **Field Emission Luminescence Phenomena**

Single crystals of zinc sulphide tightened between metal electrodes show, besides the normal electroluminescence which is out of phase with the applied potential, an 'in-phase' brightness wave component (Piper and Williams[23]). The spectra of the two components are the same. The 'in-phase' component appears above a higher voltage threshold and then increases faster with voltage than the normal component. In such crystals a current passes through the material; Frankl[35] reports that the in-phase output is proportional to this current. Quite the reverse of normal electroluminescence, the emission occurs with D.C. currents. This is ascribed to a field emission phenomenon: electrons

emitted by the cathode are accelerated and then excite the light centres as in intrinsic electroluminescence. But the electron supply is quite different.

Such an emission has also been reported on CdS crystals[31, 63]. Luminescence I of carborundum is probably due to a similar mechanism, in the same way as the occurrence of yellow spots found by Newman[64] in silicon P–N junctions biased in the blocking direction.

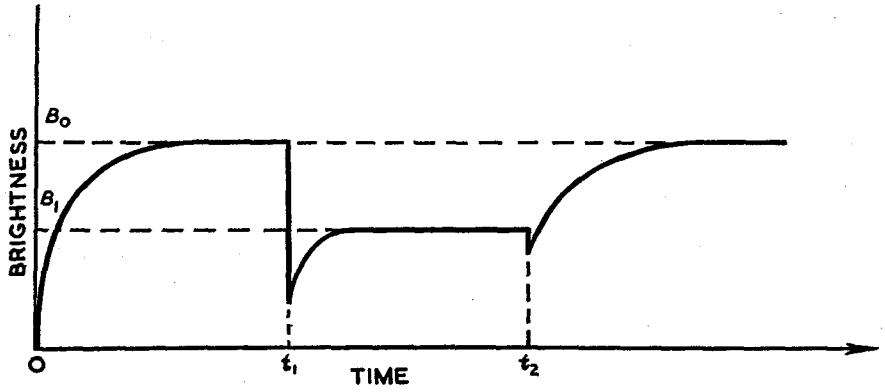


Figure 15. *The quenching effect*[68]. *The exciting radiation is applied at* t_0 *while the field is applied at* t_1 *and removed at* t_2.

4. ELECTROPHOTOLUMINESCENCE AND OTHER EFFECTS

4.1. **The Gudden–Pohl Effect**

Many phosphors which have been previously excited (u.v.-, X-, or α-excitation) show a momentary surillumination if an electric field is applied during the decay[65]. This effect differs from electroluminescence primarily by the requirement of a previous excitation; moreover even with A.C. fields, the eye observes only a flash and not a continuous light emission. *Electrophotoluminescence is ascribed to the field emptying of filled traps, while electroluminescence consists in a field-excitation of centres*[6, 7].

The fields required for electrophotoluminescence are generally smaller, but not very much, than those used for intrinsic electroluminescence. Let $\bar{x}$ be the average path introduced in sub-section 2.2; the electric field E ionizes the traps whose depths E_{traps} are less than:

$$E_{\text{trap}} < qE\bar{x} \qquad \ldots (43)$$

As in electroluminescence, a direct action of the field on the localized level does not seem to be the case.

The electrons freed from the traps may recombine with the centres emptied by the previous excitation: hence the light emission even if the field cannot excite new centres.

Re-trapping by means of the Gudden–Pohl effect has been studied by Mattler and Curie[13].

It seems that the field ionizes only a small number of traps; thermoluminescence and i.r. stimulation are still possible after the action of the field (on

the contrary, field application after thermoluminescence or stimulation does not produce any emission[1]). A large number of traps, especially among those that are directly associated with the centres, may be protected against the action of the conduction electrons by localized potential barriers around the trap–centre system (but, on the other hand, the field may empty some donor levels or

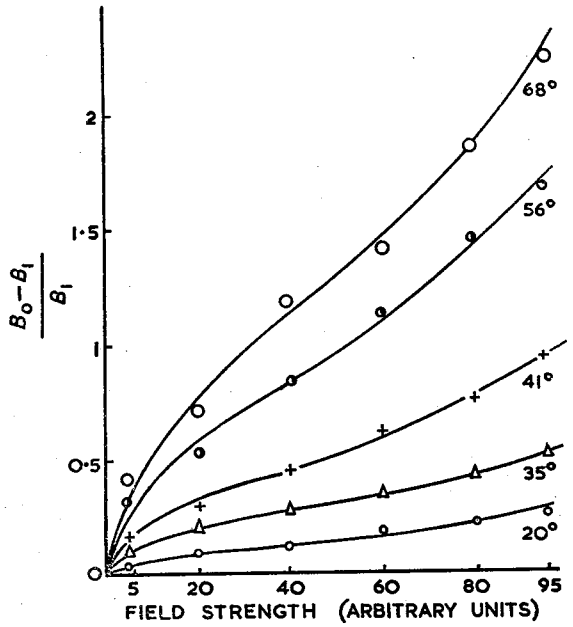


Figure 16. *Field dependence of the quenching effect at various temperatures*[69].

surface traps that are not involved in ordinary phosphorescence phenomenon but possibly act during i.r. stimulation). The shallow part of the glow curve seems to be weakened by the field, but a large re-trapping effect into the deep traps may enhance the corresponding part of the glow curve[2]. Destriau and Mattler have given evidence for directional effects[66]: immediately after the Gudden–Pohl flash and without further excitation, a second flash is observed with a field applied in a direction perpendicular to the first one.

The duration of the flash is long enough to allow the study of several brightness waves[67]. These waves differ from those obtained in pure electroluminescence by the fast decrease of their amplitude and of their shift compared to the voltage wave. This decrease is easily understood from formula (4), the shift being related to the internal conductivity of the phosphor.

4.2. The Quenching Effect of Electric Fields on Photoluminescence

Figure 15 shows the behaviour of many phosphors when a field is applied during equilibrium excitation. Let B_0 be the brightness of the plateau obtained without any field and B_1 the brightness upon field application. Destriau and

Mattler[69] have found that the curves giving the ratio $(B_0 - B_1)/B_1$ *versus* the field strength E, for a constant value of temperature T, can be superposed by a simple change in the scale of ordinates (Figure 16).

Therefore the effect arises from a simple competition between radiative and non-radiative field-induced transitions. For, if P_r and P_n^r are the respective probabilities per unit time of the radiative and non-radiative recombinations, the light efficiency η will be:

$$\eta_0 \text{ (without field)} = \frac{P_r}{P_r + P_{nr}(T)} \quad \ldots (44)$$

$$\eta_1 \text{ (with a field)} = \frac{P_r}{P_r + P_{nr}(T) + P_{nr}(E)} \quad \ldots (45)$$

The expression:

$$\frac{B_0 - B_1}{B_1} = \frac{\eta_0}{\eta_1} - 1 = \frac{P_{nr}(E)}{P_r + P_{nr}(T)} \quad \ldots (46)$$

possesses the previous property.

Figure 16 shows, moreover, that $P_{nr}(E)$ must depend on T in the following way:

$$P_{nr}(E) = f(T) \times g(E) \quad \ldots (47)$$

where f and g are two increasing functions; the increase of f must destroy the effect of the increase of $P_{nr}(T)$. The need of a thermal activation is suggested (*step-by-step effect*, similar to those described by Vigean and Curie[70, 71]).

Destriau has given evidence[72] for a near-constant character of $P_{nr}(E)$.

The brightness waves have been mostly studied with pulsed X-ray excitation: they are deeply perturbed by the field[2, 34, 73, 74, 75]. The transient effects are still more intricate and no theory of them can yet be given.

4.3. **The Enhancing Effect**

Sometimes, instead of $B_1 < B_0$ (with the symbols used in the preceding paragraph), one finds $B_1 > B_0$. The enhancing effect was first observed by G. Destriau and M. Destriau[76, 77] with X-ray excitation; the samples used showed often the quenching effect with u.v. excitation. Mattler[78] has observed an enhancement during α-ray excitation (Figure 17).

A possible enhancing effect may be foreseen from the intrinsic electroluminescence theory[79]: instead of field-excited electrons, the photoelectrons could be accelerated by the field and excite other centres.

However, for the present case (high-energy excitation on powdered samples), many features seem to rule out such a mechanism. Firstly, the enhancement is observed with so low a threshold (see Figure 19) that electron acceleration hardly appears possible. Then the linear or sub-linear dependence of the enhancing ratio B_1/B_0 *versus* voltage contrasts with the exponential dependence of the electroluminescent output. (Studies by Mattler of the frequency and temperature dependence are in progress.) Moreover the brightness waves are nothing else but the waves observed without field, amplified in a constant ratio[75, 76] (Figure 18).

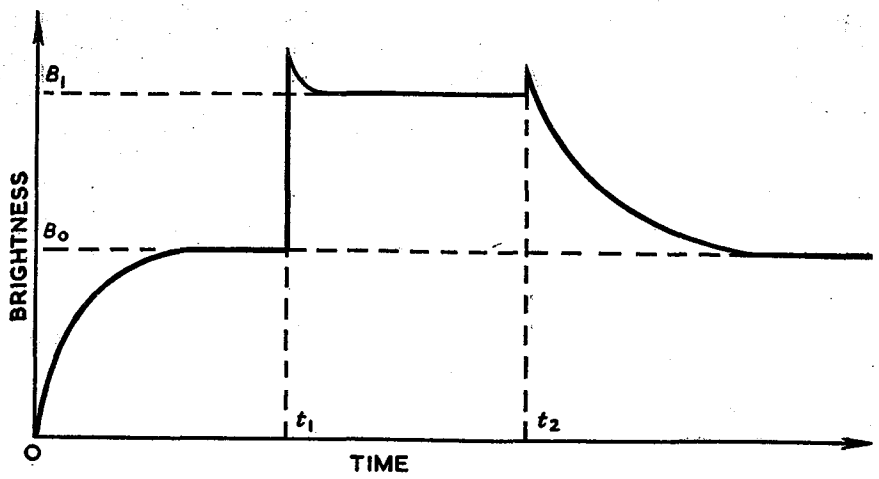


Figure 17. The *enhancing effect.*

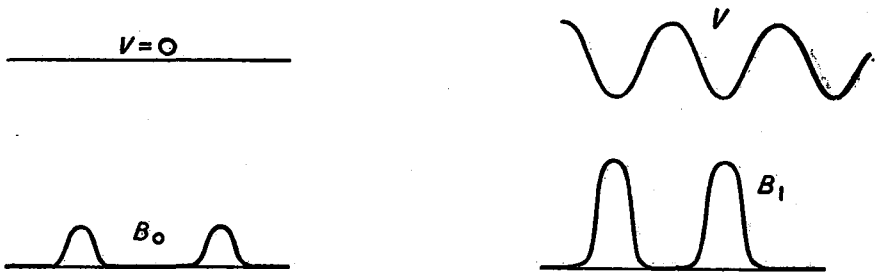


Figure 18. *Brightness waves during the enhancing effect (M. Destriau*[76]*).* B_0 *with pulsed X-ray excitation.* B_1 *with field applied.*

Excitation, by X-rays or α-rays, is ascribed to fast secondary electrons, which are able to excite the centres before any field acceleration. In the author's opinion, the drift of electrons and holes produced by the field allows these electrons to reach and excite a wider region of the phosphor. Perhaps the field diminishes the rate of energy loss for the high-speed electrons, but probably the larger the action of the field, the slower are the carriers. This mechanism is not yet plain; it is not yet possible to discuss the respective parts played by the fast electrons (impact excitation mechanism) and the slow carriers (resonance transfer mechanism[82]); but in spite of the too-small number of experiments performed until now, it seems possible to ascribe the enhancement effect observed by Destriau and Mattler to a drift effect without field acceleration.

Destriau has shown[2, 77] that the enhancing effect under X-rays becomes saturated when the applied voltage increases. But Mattler does not find any approach to saturation under α-rays. This fact resembles the behaviour of the current in an ionization chamber, which shows less tendency to saturate at

high ionization densities. The local density of secondary electrons is very large in the neighbourhood of the α path: if both phenomena actually depend on the same mechanism, this suggests the previous explanation in terms of a drift.*

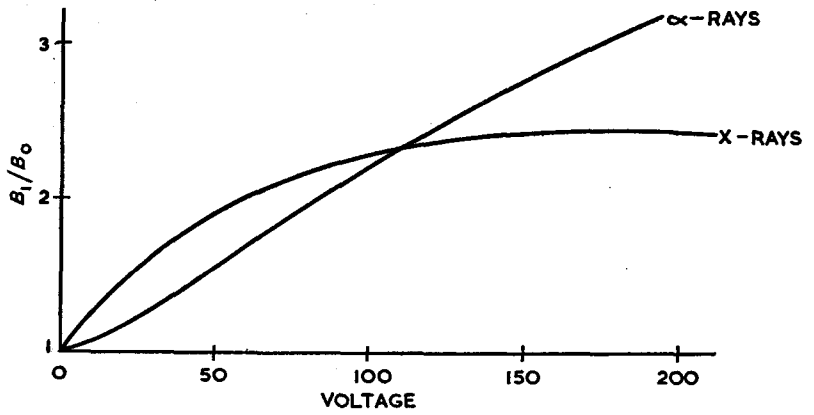


Figure 19. *Field dependence of the enhancing effect (Destriau[76], Mattler[78]).*

A *pure* enhancement effect has not yet been observed with u.v. excitation, if ordinary cells (a phosphor powder embedded in an insulator) are used.

However, Gumlich[80] has suggested the superposition of an enhancement effect to the quenching effect in the red part of the emitted spectrum, when an exciting radiation falling into the fundamental absorption band of the lattice is used.

Cusano and Williams[81] have found high light amplification (B_1/B_0 as high as 90) with transparent phosphor films deposited by evaporation, in contact with electrodes, when irradiated by u.v. and submitted to a field. Here, electron acceleration in the cathode barrier, following photo-excitation (and also injection from the cathode) may be invoked[82]. The phenomenon contrasts with Destriau's and Mattler's by the occurence of a voltage threshold, and by the fast increase of the light output for higher voltages (Figure 20).

* *Note added in proof.* At the meeting of the Electrochemical Society (Washington, May 1957), J. Mattler gives the result of preliminary studies upon the action of temperature in Destriau's enhancing effect. Quite the contrary of pure electroluminescence, the enhancing ratio B_1/B_0 increases continuously for increasing temperatures in the range studied (-150 to $+100$°C). The following simple theory has been proposed:

Let us call l the diffusion path of an electron and $L \gg l$ its mean path before capture by a centre or a defect. When no field is applied, the electron undergoes isotropic diffusions; after $n = Ll$ diffusions, it has only reached the distance

$$\bar{x}_0 = l\sqrt{(n/3)} \qquad \ldots (48)$$

from its initial place.

When a large field is applied, the travelled path $\bar{x}$ increases and tends towards L. Whence for the enhancing ratio the limit:

$$\frac{B_1}{B_0} \to \frac{L}{\bar{x}_0} \sim \sqrt{\left(\frac{3L}{l}\right)} \qquad \ldots (49)$$

L goes up for high temperatures and l goes down: both effects produce an increase of the enhancing ratio.

Both enhancing effects seem to require a phosphor of the ZnS(Mn) type *i.e.* whose centres are excited and not ionized. In the same way as in cathodoluminescence[17], it seems that in the case of too-large exciting electron densities, the ionization of centres does not lead to radiative recombinations with an appreciable probability.

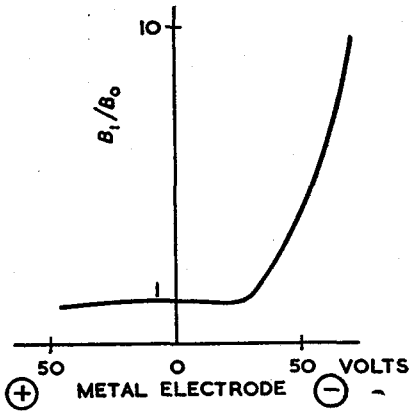


Figure 20. *Brightness versus voltage relation in William's and Cusano's enhancement effect*[82].

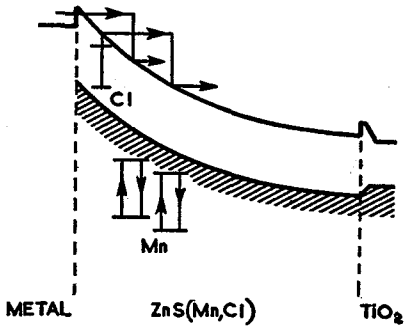


Figure 21. *Mechanism of the above enhancement effect according to Williams*[82]. *The centres of the auto-activated ZnS (Cl-centres) are photoionized; electron acceleration and multiplication—either produced by impact ionization or field emission phenomena from the electrodes—occurs as in electroluminescence; then the Mn-centres are excited. The figure refers to the case of impact excitation; resonance transfer also occurs.*

ACKNOWLEDGEMENTS

I would like to thank all my laboratory colleagues for their kind co-operation, and especially Prof. G. Destriau and Dr. J. Mattler. I am also indebted to all the people who attended the Symposium on Luminescence in Paris (May 1956), and especially to Dr. F. E. Williams (Schenectady) and Dr. R. Goffaux (Charleroi) for stimulating discussions.

REFERENCES

1. G. Destriau. *Phil. Mag.* **37**, 700 (1947).
2. G. Destriau and H. F. Ivey. *Proc. Inst. Radio Engrs., N.Y.* Special issue, p. 1911, (Dec. 1955).
3. F. E. Williams. *Advances in Electronics.* Vol. 5, p. 137 (Academic Press, New York, 1953).
4. F. E. Williams. *Phys. Rev.* **98**, 1809 (1955).
5. G. Destriau. *C. R. Acad. Sci., Paris.* **209**, 36 (1939). *Trans. Faraday Soc.* **35**, 227 (1939).
6. D. Curie. *J. Phys. Radium.* **13**, 317 (1952).
7. D. Curie. *J. Phys. Radium.* **14**, 510, 672 (1953).
8. F. E. Williams and W. W. Piper. *Brit. J, Appl. Phys.* Suppl. no .4, p. 39 (1955).
9. J. Mattler. *Symposium on Luminescence.* (Paris, May 1956); *J. Phys. Radium.* **17**, 725 (1956).

10. F. Matossi. *Phys. Rev.* **98,** 434 (1955).
11. S. Roberts. *J. Opt. Soc. Amer.* **42,** 850 (1952).
12. G. Destriau. *Meeting of the Illuminating Engineering Society,* Cleveland (Sept. 12, 1955) (see also ref. 2).
13. J. Mattler and D. Curie. *C. R. Acad. Sci., Paris.* **230,** 1086 (1950).
14. E. Nagy. *Symposium on Luminescence.* (Paris, May 1956); *J. Phys. Radium.* **17,** 773 (1956).
15. P. Zalm, G. Diemer and H. Klasens. *Philips Res. Rep.* **9,** 81 (1954).
16. P. Zalm. *Symposium on Luminescence.* (Paris, May 1956); *J. Phys. Radium.* **17,** 777 (1956).
17. M. Curie and D. Curie. *Questions actuelles en Luminescence cristalline.* (Editions de la Revue d'Optique, Paris, 1956.)
18. J. F. Waymouth, C. W. Jerome and W. C. Gungle. *Sylvania Technologist.* **5,** 54 (1952).
19. G. Destriau. *J. Phys. Radium.* **16,** 798 (1955).
20. J. Mattler. *J. Phys. Radium.* **17,** 42 (1956).
21. J. Frenkel. *Tech. Phys. U.S.S.R.* **5,** 685 (1938); W. A. Thornton. *Phys. Rev.* **102,** 38 (1956).
22. Johnson, Piper and Williams. *J. Electroch. Soc.* **103,** 221 (1956).
23. W. W. Piper and F. E. Williams. *Phys. Rev.* **87,** 151 (1952).
24. D. Curie. *Heidelberg-Mosbach Luminescence Symposium.* (July 9, 1952).
25. R. Frerichs. *Phys. Rev.* **76,** 1869 (1949).
26. C. Lanczos. *Z. Phys.* **68,** 204 (1931).
27. W. Franz. *Ann. Phys., Lpz.* **11,** 17 (1953).
28. P. Zalm, G. Diemer and H. A. Flasens. *Philips Res. Rep.* **10,** 205 (1955).
29. A. Fischer. *Phys. Verhandl.* **5,** 64 (1954).
30. B. T. Howard. *Phys. Rev.* **98,** 1544 (1955).
31. G. Diemer. *Philips Res. Rep.* **9,** 109 (1954).
32. G. Diemer. *Philips Res. Rep.* **10,** 194 (1955).
33. F. J. Waymouth and F. Bitter. *Phys. Rev.* **95,** 941 (1954).
34. G. Destriau. *Symposium on Luminescence.* (Paris, May 1956); *J. Phys. Radium.* **17,** 734 (1956).
35. D. R. Frankl. *Meeting of the Electrochemical Society, Cincinnatti,* (May 1–5, 1955); *Symposium on Luminescence* (Paris, May 1956); *J. Phys. Radium.* **17,** 731 (1956).
36. A. Von Hippel. *J. Appl. Phys.* **8,** 815 (1937); *Phys. Rev.* **76,** 127 (1949).
37. F. Seitz. *Phys. Rev.* **76,** 1376 (1949).
38. H. Fröhlich. *Proc. Roy. Soc.* A, **188,** 521 (1957).
39. R. J. Seeger and E. Teller. *Phys. Rev.* **54,** 515 (1938).
40. G. F. Alfrey and J. B. Taylor. *Brit. J. Appl. Phys.* Suppl. no. 4, p. 44 (1955).
41. B. T. Howard, H. F. Ivey and W. Lehmann. *Phys. Rev.* **96,** 799 (1954).
42. A. Luyckx and A. J. Stokkink. *Brit. J. Appl. Phys.* Suppl. no. 4, p. 57 (1955).
43. D. W. G. Ballentyne. *Symposium on Luminescence.* (Paris, May 1956); *J. Phys. Radium.* **17,** 759 (1956).
44. W. Shockley. *Bell Syst. Tech. J.* **30,** 990 (1951).
45. R. Goffaux. *Bull. Acad. Roy. Belg.* **XL,** 808 (1954). *Symposium on Luminescence.* (Paris, May 1956).
46. A. N. Ince. *Proc. Phys. Soc., Lond.* B, **67,** 870 (1954).
47. H. C. Froelich. *J. Opt. Soc. Amer.* **43,** 320 (1953).
48. A. Bernanose, M. Comte and P. Vouaux. *J. Chim. Phys.* **50,** 64, 261 (1953); *Brit. J. Appl. Phys.,* Suppl. no. 4, p. 54 (1954).
49. A. Bernanose. *J. Chim. Phys.* **52,** 396 (1955).
50. O. Lossew. *Phil. Mag.* **6,** 1024 (1928).
51. O. Lossew. *Phys. Z.* **30,** 920 (1929); *Phys. Z.* **32,** 695 (1931); *Phys. Z.* **34,** 397 (1933).
52. K. Lehovec, G. Accardo and E. Jamgochian. *Phys. Rev.* **83,** 603 (1951); *Phys. Rev.* **89,** 20 (1953).
53. M. Schön. *Z. Naturf.* **8a,** 442 (1953).
54. G. Szigeti. *Brit. J. Appl. Phys.* Suppl. no. 4, p. 56 (1955).
55. G. Curie and D. Curie. *J. Phys. Radium.* **15,** 61 (1954).
56. J. R. Haynes and H. B. Briggs. *Phys. Rev.* **86,** 647 (1952).
57. R. Newman. *Phys. Rev.* **91,** 1313 (1953).
58. P. Aigrain. *Physica.* **20,** 1010 (1954).
59. R. Braunstein. *Phys. Rev.* **99,** 1892 (1955); G. A. Wolff, R. A. Herbert and J. D. Broder. *Phys. Rev.* **100,** 1144 (1955).
60. T. S. Moss. *Symposium on Luminescence.* (Paris, May 1956); *J. Phys. Radium.* **17,** 712 (1956).
61. W. Shockley and W. T. Read. *Phys. Rev.* **87,** 835 (1952).

62. W. Shockley and W. Van Roosbroek. *Phys. Rev.* **94,** 1558 (1954).
63. K. M. Boër and U. Kummel. *Z. Phys. Chem.* **200,** 193 (1952).
64. R. Newman, W. C. Dash, R. N. Hall and W. E. Burch. *Phys. Rev.* **98,** 1536 (1955).
65. B. Gudden and R. Pohl. *Z. Phys.* **2,** 192 (1920).
66. G. Destriau and J. Mattler. *C.R. Acad. Sci., Paris.* **223,** 894 (1946).
67. G. Destriau and J. Mattler. *J. Phys. Radium.* **13,** 205 (1952).
68. G. Destriau. *J. Phys. Radium.* **4,** 32 (1943).
69. G. Destriau and J. Mattler. *J. Phys. Radium.* **11,** 529 (1950).
70. F. Vigean. *C.R. Acad. Sci., Paris.* **232,** 819 (1951).
71. F. Vigean and D. Curie. *C.R. Acad. Sci., Paris.* **232,** 955 (1951).
72. G. Destriau. *C.R. Acad. Sci., Paris.* **230,** 1061 (1950).
73. F. Matossi. *Phys. Rev.* **94,** 1151 (1954).
74. F. Matossi and S. Nudelman. *Phys. Rev.* **98,** 1545 (1955).
75. H. E. Gumlich. *J. Phys. Radium.* **17,** 117 (1956).
76. M. Destriau. *C. R. Acad. Sci., Paris.* **238,** 2298 (1954).
77. G. Destriau, J. Mattler, M. Destriau and H. E. Gumlich. *J. Electrochem. Soc.* **102,** 682 (1955).
78. J. Mattler. Short communication, *Symposium on Luminescence.* (Paris, May 1956); *J. Phys. Radium.* **17,** 758 (1956).
79. D. Curie. *C. R. Acad. Sci., Paris.* **237,** 791 (1953).
80. H. E. Gumlich. *Symposium on Luminescence.* (Paris, May 1956); *J. Phys. Radium.* **17,** 754 (1956).
81. D. A. Cusano. *Phys. Rev.* **98,** 546 (1955).
82. F. E. Williams. *Phys. Rev.* **98,** 547 (1955); *Symposium on Luminescence.* (Paris, May 1956); *J. Phys. Radium.* **17,** 742 (1956).

INDEX