THE BELL SYSTEM TECHNICAL JOURNAL

VOLUME XXXV | MAY 1956 | NUMBER 3

Copyright 1956, American Telephone and Telegraph Company

Chemical Interactions Among Defects in Germanium and Silicon

By HOWARD REISS, C. S. FULLER, and F. J. MORIN

Interactions among defects in germanium and silicon have been investigated. The solid solutions involved bear a strong resemblance to aqueous solutions insofar as they represent media for chemical reactions. Such phenomena as acid-base neutralization, complex ion formation, and ion pairing, all take place. These phenomena, besides being of interest in themselves, are useful in studying the properties of the semiconductors in which they occur. The following article is a blend of theory and experiment, and describes developments in this field during the past few years.

CONTENTS

I. Introduction 536
II. Electrons and Holes as Chemical Entities 537
III. Application of the Mass Action Principle 546
IV. Further Applications of the Mass Action Principle 550
V. Complex Ion Formation 557
VI. Ion Pairing 565
VII. Theories of Ion Pairing 567
VIII. Phenomena Associated with Ion Pairing in Semiconductors 575
IX. Pairing Calculations 578
X. Theory of Relaxation 582
XI. Investigation of Ion Pairing by Diffusion 591

XII. Investigation of Ion Pairing by Its Effect on Carrier Mobility..... 601
XIII. Relaxation Studies..... 607
XIV. The Effect of Ion Pairing on Energy Levels..... 610
XV. Research Possibilities..... 611
Acknowledgements..... 613
Appendix A — The Effect of Ion Pairing on Solubility..... 613
Appendix B — Concentration Dependence of Diffusivity in the Presence of Ion Pairing..... 617
Appendix C — Solution of Boundary Value Problem for Relaxation..... 619
Appendix D — Minimization of the Diffusion Potential..... 623
Appendix E — Calculation of Diffusivities from Conductances of Diffusion Layers..... 626
Glossary of Symbols..... 630
References..... 634

I. INTRODUCTION

The effort of Wagner[1] and his school to bring defects in solids into the domain of chemical reactants has provided a framework within which various abstruse statistical phenomena can be viewed in terms of the intuitive principle of mass action.[2] Most of the work to date in this field has been performed on oxide and sulfide semiconductors or on ionic compounds such as silver chloride. In these materials the control of defects (impurities are to be regarded as defects) is not all that might be desired, and so with a few exceptions, experiments have been either semiquantitative or even qualitative.

With the emergence of widespread interest in semi-conductors, culminating in the perfection of the transistor, quantities of extremely pure single crystal germanium and silicon have become available. In addition the physical properties, and even the quantum mechanical theory of the behavior of these substances have been widely investigated, so that a great deal of information concerning them exists. Coupled with the fact that defects in them, especially impurities, are particularly susceptible to control, these circumstances render germanium and silicon ideal substances in which to test many of the concepts associated with defect interactions.

This view was adopted at Bell Telephone Laboratories a few years ago when experimental work was first undertaken. Not only has it been possible to demonstrate quantitatively the validity of the mass action principle applied to defects, but new kinds of interactions have been discovered and studied. Furthermore new techniques of measurement have been developed which we feel open the way for broader investigation of a still largely unexplored field.

In fact solids (particularly semiconductors like germanium and silicon)

appear in every respect to provide a medium for chemical reactivity similar to liquids, particularly water. Such pehnomena as acid-base reactions, complex ion formation, and electrolyte phenomena such as Debye Hückel effects, ion pairing, etc., all seem to take place.

Besides the experiments theoretical work has been done in an attempt to define the limits of validity of the mass action principle, to furnish more refined electrolyte theories, and most importantly, to provide firm theoretical bases for entirely new phenomena such as ion pair relaxation processes.

The consequence is that the field of diamond lattice[3] semiconductors which has previously engaged the special interests of physicists threatens to become important to chemists. Semiconductor crystals are of interest, not only because of the specific chemical processes occurring in these substances, but also because they serve as proving grounds for certain ideas current among chemists, such as electrolyte theory. On the other hand renewed interest is induced on the part of physicists because chemical effects like ion pairing engender new physical effects.

The purpose of this paper is to present the field of defect interaction as it now stands, in a manner intelligible to both physicists and chemists. However, this is not a review paper. Most of the experimental results, and particularly the theories which are fully derived in the text or the appendices are entirely new. Some allusion will be made to published work, particularly to descriptions of the results of some previous theories, in order to round out the development.

The governing theme of the article lies in the analogy between semiconductors and aqueous solutions. This analogy is useful not so much for what it explains, but for the experiments which it suggests. More than once it has stimulated us to new investigations.

In our work we have made extensive use of lithium as an impurity. This is so because lithium can be employed with special ease to demonstrate most of the concepts we have in mind. This specialization should not obscure the fact that other impurities although not well suited to the performance of accurate measurements, will exhibit much of the same behavior.

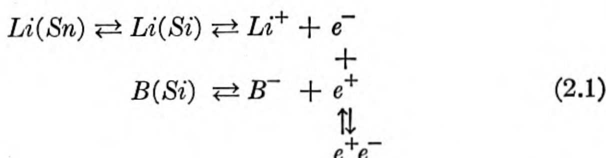
II. ELECTRONS AND HOLES AS CHEMICAL ENTITIES

Since electrons and holes[4] are obvious occupants of semiconductors like germanium and silicon, and are intimately associated with the presence of donor and acceptor impurities,[4] it is fitting to inquire into the roles they may play in chemical interactions between donors and ac-

ceptors. This question has been discussed in two papers,[5, 6] and only its principle aspects will be considered.

To gain perspective it is convenient to consider a system representing the prototype of most systems to be discussed here. Consider a single crystal of silicon containing substitutional boron atoms. Boron, a group III element, is an acceptor, and being substitutional cannot readily diffuse[7] at temperatures much below the melting point of silicon. If this crystal is immersed in a solution containing lithium, e.g., a solution of lithium in molten tin, lithium will diffuse into it and behave as a donor. Evidence suggests that lithium dissolves interstitially in silicon, thereby accounting for the fact that it possesses a high diffusivity[8] at a temperature where boron is immobile, for example, below 300°C. When the lithium is uniformly distributed throughout the silicon its solubility in relation to the external phase can be determined. Throughout this process boron remains fixed in the lattice.

If both lithium and boron were inert impurities the solubility of the former would not be expected to depend on the presence or absence of the latter, for the level of solubility is low enough to render (under ordinary circumstances) the solid solution ideal.[9] On the other hand the impurities exhibit donor and acceptor behaviors respectively, and some unusual effects might exist. We shall first speculate on the simplest possibility in this direction, with the assistance of the set of equilibrium reactions diagrammed below.*



$$\begin{array}{ccccccc} Li(Sn) & \rightleftarrows & Li(Si) & \rightleftarrows & Li^+ & + & e^- \\ & & & & & & + \\ & & B(Si) & \rightleftarrows & B^- & + & e^+ \\ & & & & & & \upharpoonleft\downharpoonright \\ & & & & & & e^+e^- \end{array} \qquad (2.1)$$

At the left lithium in tin is shown as $Li(Sn)$. It is in reversible equilibrium with $Li(Si)$, un-ionized lithium dissolved in silicon. The latter, in turn, ionizes to yield a positive Li^+ ion and a conduction electron, e^-. Boron, confined to the silicon lattice as $B(Si)$ ionizes as an acceptor to give B^- and a positive hole, e^+. The conduction electron, e^-, may fall into a valence band hole, e^+, to form a recombined hole-electron pair, e^+e^-. This process and its reverse are indicated by the vertical equilibrium at the right.

All of the reactions in (2.1), occuring within the silicon crystal are describable in terms of tansitions between states in the energy band dia-

* A glossary of symbols is given at the end of this article.

gram of silicon, exhibited in Fig. 1. The conduction band, the valence band, and the forbidden gap are shown. Lithium and boron both introduce localized energy states in the range of forbidden energies. The state for lithium lies just below the bottom of the conduction band while that for boron lies just over the top of the valence band. The separations in energy between most donors or acceptors and their nearest bands are of the order of hundredths of an electron volt while the breadths of the forbidden gaps in germanium or silicon are of the order of one electron volt.

Process 1 in Fig. 1 involving a transition between the donor level and conduction band corresponds to the ionization of lithium in (2.1). Process 2 is the ionization of boron while process 3 represents hole-electron recombination and generation. The various energies of transition are the heats of reaction of the chemical-like changes in (2.1).

Proceeding in the chemists fashion one might argue as follows concerning (2.1). If e^+e^- is a stable compound, as it is at fairly low temperatures, then its formation should exhaust the solution of electrons, forcing the set of lithium equilibria to the right. In this way the presence of boron, supplying holes toward the formation of e^+e^-, increases the solubility of lithium. In fact if e^+ is regarded as the solid state analogue of the hydrogen ion in aqueous solution, and e^- as the counterpart of the hydroxyl ion, then the donor, lithium, may be considered a base while boron, may be considered an acid. Furthermore e^+e^- must correspond to water. Thus the scheme in (2.1) is analogous to a neutralization reaction in which the weakly ionized substance is e^+e^-.

If the immobile boron atoms were replaced by immobile donors, e.g., phosphorus atoms, a reduction, rather than an increase, in the solubility

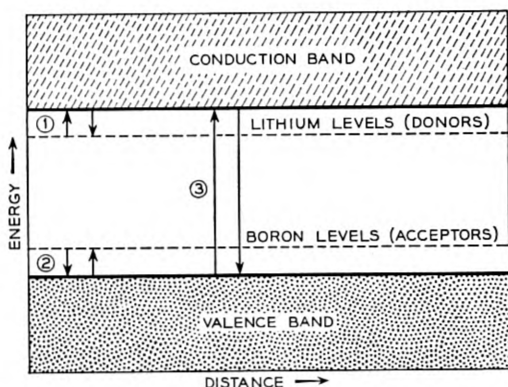


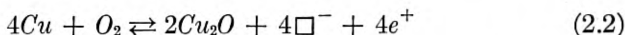
Fig. 1 — Energy band diagram showing the chemical equilibria of (2.1).

of lithium might be expected on the basis of an oversupply of electrons (i.e., by the common ion effect[10]). In that case we would have a base displacing another base from solution.

The intimate comparison between this kind of solution and an aqueous solution is worth emphasizing not so much for what it adds to one's understanding of the situation but rather for the further effects it suggests along the lines of analogy. These additional phenomena have been looked for and found, and will be discussed later in this article.

The scheme shown in (2.1) should be applicable, in principle, to other donors and acceptors and to germanium and other semiconductors as well as silicon. Furthermore the external phase may be any one of a suitable variety, and need not even be liquid. Other systems, however, are not as convenient, especially in regard to the ease of equilibration of an impurity over the parts of an heterogeneous system. The lengths to which one can go in comparing electrolytes and semiconductors are discussed in a recent paper.[11]

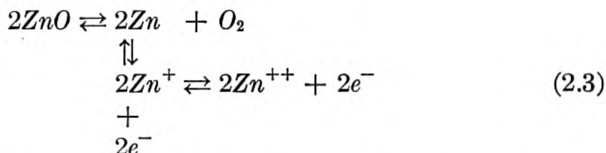
In order to quantify the scheme of (2.1) it seems natural to invoke the law of mass action.[2] Treatments in which holes and electrons are involved in mass action expressions are not new, although systems forming such perfect analogies to aqueous solutions do not seem to have been discussed in the past. For example, in connection with the oxidation of copper Wagner[12] writes



$$4Cu + O_2 \rightleftarrows 2Cu_2O + 4\square^- + 4e^+ \quad (2.2)$$

in which $\square^-$ is a negatively charged cation vacancy in the Cu_2O lattice, and e^+ is a hole. Wagner proceeds to invoke the law of mass action in order to compute the oxygen pressure dependence in this system.

In another example Baumbach and Wagner[13] and others have investigated oxygen pressure over non-stoichiometric zinc oxide. They consider the possible reactions



$$\begin{array}{lll} 2ZnO \rightleftarrows & 2Zn & + O_2 \\ & \upharpoonleft\downharpoonright & \\ & 2Zn^+ \rightleftarrows 2Zn^{++} + 2e^- & \\ & + & \\ & 2e^- & \end{array} \quad (2.3)$$

and apply the law of mass action. In (2.3) the various states of Zn are presumably interstitial.

Kroger and Vink[14] have recently considered the problem in oxides and sulfides in a rather general way. However in none of the oxide-sulfide systems has it been possible to achieve really quantitative results. In

contrast silicon and germanium offer possibilities of an entirely new order. The advent of the transistor has not only provided large supplies of pure single crystal material, but it has also made available a store of fundamental information concerning the physical properties of these substances. For example, data exists on their energy band diagrams including impurity states — also on resistivity — impurity density curves, diffusivities of impurities, etc. Furthermore, the amount of ionizable impurities can be controlled within narrow limits, and can be changed at will and measured accurately. Consequently it is reasonable to assume that experiments on germanium and silicon will be more successful than similar investigations using other materials.

At this point it is in order to examine whether or not the treatment of electrons and holes as normal chemical entities satisfying the law of mass action is altogether simple and straightforward. This problem has been investigated by Reiss[5] who found the treatment permissible only as long as the statistics satisfied by holes and electrons remain classical. The validity of this contention can be seen in a very simple manner. Consider a system like that in (2.1). Let the total concentration of donor (ionized and un-ionized) be N_D, the concentration of ionized donor be D^+, the concentration of conduction electrons be n, and that of valence band holes be p. Let N_A and A^- denote the concentrations of total acceptor and acceptor ions respectively. Finally, let α be the thermodynamic activity[15] of the donor (lithium in (2.1)) in the external phase.

Then, corresponding to the heterogeneous equilibrium in which lithium distributes itself between the two phases we can write

$$\frac{N_D - D^+}{\alpha} = K_0 \tag{2.4}$$

in which K_0 depends on temperature, but not on composition. This assumes the semiconductor to be dilute enough in donor so that the activity of un-ionized donor can be replaced by its concentration, $N_D - D^+$. For the ionization of the donor we can write the mass action relation,

$$\frac{D^+ n}{N_D - D^+} = K_D \tag{2.5}$$

and for the acceptor,

$$\frac{A^- p}{N_A - A^-} = K_A \tag{2.6}$$

while for the electron-hole recombination equilibrium

$$np = K_1 \tag{2.7}$$

In (2.5), (2.6), and (2.7) all the K's are independent of composition. To these equations is added the charge neutrality condition,

$$D^+ + p = A^- + n \tag{2.8}$$

Equations (2.4) through (2.8) are enough to determine N_D in its dependence on N_A, α, and the various K's. Together they represent the mass action approach. To demonstrate their validity it is necessary to appeal to statistical considerations.

Thus $N_D - D^+$, the concentration of un-ionized donor is really the density of electrons in the donor level of the energy diagram for the semiconductor. According to Fermi statistics this density is given by[5]

$$N_D - D^+ = N_D/\{1 + \tfrac{1}{2} \exp [(E_D - F)/kT]\} \tag{2.9}$$

in which E_D is the energy of the donor level, F is the Fermi level,[16] k, the Boltzmann constant, and T, the temperature. Furthermore, according to Fermi statistics, n, the total density of electrons in the conduction band is

$$n = \sum_i g_i/\{1 + \exp [(E_i - F)/kT]\} \tag{2.10}$$

where g_i is the density of levels of energy, E_i, in the conduction band, and the sum extends over all states in that band. Similar expressions are available for the occupation of the acceptor level and the valence band. F is usually determined by summing over all expressions like (2.9) and (2.10) and equating the result to the total number of electrons in the system. This operation corresponds exactly to applying the conservation condition, (2.8). It is obvious from the manner of its determination that F depends upon $N_D - D^+$, n, etc.

If we now form the expression on the left of (2.5) by substituting for each factor in it from (2.9) and (2.10), it is obvious that the result depends in a very complicated fashion upon F, and so cannot be the constant, K_D, independent of composition, since in the last paragraph F was shown to depend on composition. On the other hand if attention is confined to the limit in which classical statistics apply[17] the unities in the denominators of (2.9) and (2.10) can be disregarded in comparison to the exponentials, and those equations become

$$N_D - D^+ = 2N_D e^{F/kT} e^{-E_D/kT} \tag{2.11}$$

and

$$n = e^{F/kT} \sum g_i e^{-E_i/kT} \tag{2.12}$$

respectively. Moreover, from (2.11)

$$D^{+} = N_D[1 - 2e^{F/kT}e^{-E_D/kT}] = N_D \tag{2.13}$$

where the second term in brackets is ignored for the same reason as unity in the denominators of (2.9) and (2.10). Substituting (2.11) through (2.13) into (2.5) yields

$$\frac{D^{+}n}{N_D - D^{+}} = \frac{\sum_i g_i e^{-E_i/kT}}{2e^{-E_D/kT}} \tag{2.14}$$

in which the right side is truly independent of composition, since F has cancelled out of the expression. Similar arguments hold for (2.6) and (2.7). Therefore in the classical limit the law of mass action is valid, at least insofar as internal equilibria are concerned.

We have next to examine the validity of (2.4) which is really the law of mass action applied to the heterogeneous equilibrium between phases. Substitution of (2.11) into (2.4) leads to the prediction

$$\alpha = \frac{2e^{-E_D/kT}}{K_0}\{e^{F/kT}\}N_D = K\{e^{F/kT}\}N_D \tag{2.15}$$

in the classical case, if (2.4) is valid. In order to confirm (2.15) it is necessary to evaluate the chemical potentials[18] of the donor in the external phase and in the semiconductor, and equate the two. The resulting expression should be equivalent to (2.15).

Since α is the activity of the donor in the external phase its chemical potential in that phase is, by definition,

$$\mu = \mu^0(T, p) + kT \ln \alpha \tag{2.16}$$

where μ^0, the chemical potential in the standard state, may depend on temperature and pressure, but not on composition. To compute the chemical potential in the semiconductor statistical methods must once more be invoked. Thus, according to (2.13), donor atoms are nearly totally ionized in the classical case, so that the addition of a donor atom to the semiconductor amounts to addition of two separate particles, the donor ion and the electron. The chemical potential of the added atom is therefore the sum of the potentials of the ion and the electron separately. Since the ions are supposedly present in low concentration the latter can serve as an activity,[19] and in analogy to (2.16) we obtain for the ionic chemical potential

$$\mu_{D^+} = \mu_{D^+}{}^0(T, p) + kT \ln D^{+} \tag{2.17}$$

Furthermore, it is well established[20] that the Fermi level plays the role of chemical potential, μ_e, for the electron

$$\mu_e = F \tag{2.18}$$

Thus the chemical potential for the donor atom is

$$\begin{aligned} \mu_D &= \mu_{D^+} + \mu_e = \mu_{D^+}{}^0 + kT\,\ell n\,D^+ + F \\ &= \mu_{D^+}{}^0 + kT\,\ell n\,N_D + F = \mu_{D^+}{}^0 + kT\,\ell n\,\{e^{F/kT}\}N_D \end{aligned} \tag{2.19}$$

where (2.13) has been used to replace D^+ by N_D. We note that the activity of the donor atom must be

$$\{e^{F/kT}\}N_D \tag{2.20}$$

with $e^{F/kT}$ playing the role of an activity coefficient.[21]

Equating μ_D given by (2.19) to μ in (2.16) results in the equation

$$\alpha = \exp[(\mu_{D^+}{}^0 - \mu^0)/kT]\{e^{F/kT}\}N_D \tag{2.21}$$

which can be made identical to (2.15) by identifying

$$\exp[(\mu_{D^+}{}^0 - \mu^0)/kT]$$

with K of that expression. Thus in the classical case the law of mass action is applicable to the heterogeneous equilibrium.

When classical statistics no longer apply it is still possible to evaluate $N_D - D^+$, using the full expression (2.9). Therefore the solubility N_D, of the donor can still be determined if (2.4) remains valid. To decide this question it is necessary to evaluate μ_D, the chemical potential of the donor in the semiconductor under non-classical conditions. This problem is not as simple as those treated above, but it can be solved, and the detailed arguments can be found in Reference 5. Here we shall be content with quoting the results. However, before doing this the non-classical counterpart of (2.15) will be written by combining (2.9) with (2.4). The result is

$$\alpha = [K_0/\{1 + \tfrac{1}{2}\exp[(E_D - F)/kT]\}]N_D \tag{2.22}$$

and if (2.4) is valid (2.22) should be derivable by equating μ to the proper value of μ_D.

Since in the non-classical case a finite portion of the donor states are occupied by electrons, the introduction of an additional *average* donor atom is no longer equivalent to adding two independent particles whose chemical potentials can be summed. In the statistical derivation of μ_D it is therefore necessary to evaluate the total free energy of the semi-

conductor phase, and to differentiate this with respect to N_D, keeping temperature and pressure fixed.* The result is

$$\mu_D = \mu_{D^+}{}^0 + kT \ln N_D + F - kT \ln \{1 + 2\exp[-(E_D - F)/kT]\} \tag{2.23}$$

in which it has been assumed that the concentration of impurity is sufficiently low so that the solution would be ideal if the impurity could not ionize. In the classical case the exponential in the logarithm is small compared to unity and (2.23) becomes identical with (2.19), as it should. In the totally degenerate case the exponential dominates the unity and we have

$$\begin{aligned} \mu_D &= \{\mu_{D^+}{}^0 + E_D - kT \ln 2\} + kT \ln N_D \\ &= \mu_D^0 + kT \ln N_D \end{aligned} \tag{2.24}$$

which is the chemical potential of an un-ionized component of a dilute

* An interesting by-product of this derivation (discussed in Reference 5) is the fact that the Fermi level, F, is hardly ever the Gibbs free energy per electron for the electron assembly, although it is always the electronic chemical potential, in the sense that it measures the direction of flow of electrons. This arises because the Gibbs free energy is not always a homogeneous function[22] of the first degree in the mole numbers (electron numbers). Thus if the number of electrons in the assembly is N, the Gibbs free energy, G, is given by

$$G = NF + kT \sum_j \left[V\left\{\frac{\partial \omega_j}{\partial V}\right\}_{T,N} - \omega_j \right] \ln \frac{\omega_j}{h_j}$$

where the sum is over all energy levels, j, referred to an invariant standard level. V is the volume of the system, ω_j is the total number of states at the jth level, and h_j is the number of unoccupied states (holes) at the jth level. For F to be the free energy per electron the term involving the sum must vanish so that

$$F = \frac{G}{N}$$

But this can only happen when

$$\omega_j = K_j V$$

where K_j is independent of V. This requirement is formally met in the case of the free electron gas where the electrons have been treated as independent particles in a box so that

$$\omega_j = [8 m_0{}^{3/2} \pi E\, dE/2h^3]V$$

where m_0 is the electron mass, and h, Plank's constant. Since this is the case most frequently dealt with in thermodynamic problems it has been customary to think of F as the free energy per electron, although even here the truth of the contention depends on the assumption of particle in the box behavior.

At the other extreme, it is obvious that ω_j for a level corresponding to the deep closed shell states of the atoms forming a solid cannot depend at all on the external volume since they are essentially localized. In computing the free energy of the semiconductor phase it is necessary to understand carefully subtleties of this nature.

solution, as it should be for the degenerate case in which ionization is suppressed. Equating μ_D in (2.23) to μ in (2.16) yields

$$\alpha = \left\{ \frac{\frac{1}{2} \exp [(\mu_{D^+}{}^0 - \mu^0 + E_D)/kT]}{1 + \frac{1}{2} \exp [(E - F)/kT]} \right\} N_D \qquad (2.25)$$

which is identical with (2.22) if K_0 is taken to be

$$\tfrac{1}{2} \exp[(\mu_{D^+}{}^0 - \mu^0 + E_D)/kT] \qquad (2.26)$$

Thus one arrives at the conclusion that the law of mass action remains valid for the heterogeneous equilibrium even when it fails for the homogeneous internal equilibria.

This is a fairly important result since it implies that solubilities can give information on the behavior of the Fermi level and hence on the distribution of electronic energy levels, even under conditions of degeneracy.

The chemical potential specified by (2.23) is of course important in itself, for treating any equilibrium (external or internal) in which the donor may participate.

One last remark is in order. This concerns the treatment of heterogeneous equilibria involving some external phase, and the surface[23] rather than the body of a semiconductor. In such treatments it has been customary to compute the chemical potential of an ionizable adsorbed atom by summing the ion chemical potential and the Fermi level, as in (2.19). This is no more possible if the statistics of the surface states are non-classical, then it is possible when considering non-classical situations involving the body of the crystal. Care must therefore be exercised also in the treatment of surface equilibria.

The above discussion has shown that there are extensive ranges of conditions under which holes and electrons obey the law of mass action, and behave like chemical entities. In the next section some of the consequences of this fact will be developed.

III. APPLICATION OF THE MASS ACTION PRINCIPLE

Equations (2.4) through (2.8) will now be used to determine how, in the classical case, the solubility, N_D, of lithium in (2.1) depends upon N_A the concentration of boron in silicon. In the experiments to be described, the systems are classical, and the donors and acceptors therefore so thoroughly ionized that N_D can be replaced by D^+ and N_A by A^-. Insertion of (2.4) into (2.5) yields

$$D^+ n = \alpha K_D K_0 = K^* \qquad (3.1)$$

since α is maintained constant. Furthermore (2.7) can be written as

$$np = K_1 = n_i^2 \tag{3.2}$$

where n_i is obviously the concentration of holes or electrons under the condition that the two are equal. It is called the intrinsic concentration[24] of holes or electrons. The values of n_i in germanium and silicon have been determined by Morin.[25, 26] Fig. 2 gives plots of the logarithms of n_i in germanium and silicon versus the reciprocals of temperature. These results are necessary for subsequent calculations.

Since N_A and A^- are assumed equal, we may dispense with (2.6). The one remaining equation is then (2.8) which we adopt unchanged. These three relations, (3.1), (3.2), and (2.8) are sufficient to determine D^+ or N_D as a function of A^- or N_A. The only undetermined parameter in the set is K^* and this can be evaluated by measuring the solubility, D^+, in the absence of acceptor, i.e., under the condition that A^- is zero. The symbol D_0^+ is used to designate this value of D^+. In Reference 6 it is shown that

$$D_0^+ = K^*/(K^* + n_i^2)^{1/2}$$

or

$$K^* = (D_0^+)^2/2 + \{(D_0^+)^4/4 + n_i^2(D_0^+)^2\}^{1/2} \tag{3.3}$$

Eliminating K^* by the use of this relation it is further shown in Reference 6 that

$$D^+ = \frac{A^-}{1 + \sqrt{1 + (2n_i/D_0^+)^2}} + \left\{\left[\frac{A^-}{1 + \sqrt{1 + (2n_i/D_0^+)^2}}\right]^2 + (D_0^+)^2\right\}^{1/2} \tag{3.4}$$

which is the required relation between donor solubility and acceptor concentration.

Examination of (3.4) reveals several simple features, the more important of which we list below:

(1) When A^- (the acceptor doping) is sufficiently large so that $(D_0^+)^2$ in the second term can be ignored relative to the term in A^-, (3.4) reduces to that of a straight line with slope

$$D^+/A^- = \frac{2}{1 + \sqrt{1 + (2n_i/D_0^+)^2}} \tag{3.5}$$

Knowledge of this slope is equivalent to knowledge of D_0^+.

(2) Where the straight line portion of the D^+ versus A^- curve is in-

volved, the temperature dependence of the solubility, D^+, enters only through the ratio, n_i/D_0^+. If this ratio is very small, then

$$D^+ \approx A^- \tag{3.6}$$

and the solubility is independent of temperature. In this condition D^+ may approximate A^- by being either slightly less or slightly greater than the latter. Details are given in Reference 6.

(3) Whereas D^+ at small values of doping may be an increasing function of temperature, it may, depending on the system, be a decreasing function of temperature at high dopings. Thus doping may change the sign of the temperature coefficient of solubility. Because of this, doping sometimes may prevent precipitation of a donor when a semiconductor is cooled, since the latter becomes an undersaturated rather than a supersaturated solution of impurity. Details are given in Reference 6.

(4) It is also shown in Reference 6 that for the acceptor to have any effect on the solubility of the donor the concentration of A^- should satisfy the following criterion

$$A^- > (D_0^+ \quad \text{or} \quad n_i) \tag{3.7}$$

D_0^+ or n_i being used depending on which is greater. Obviously at high

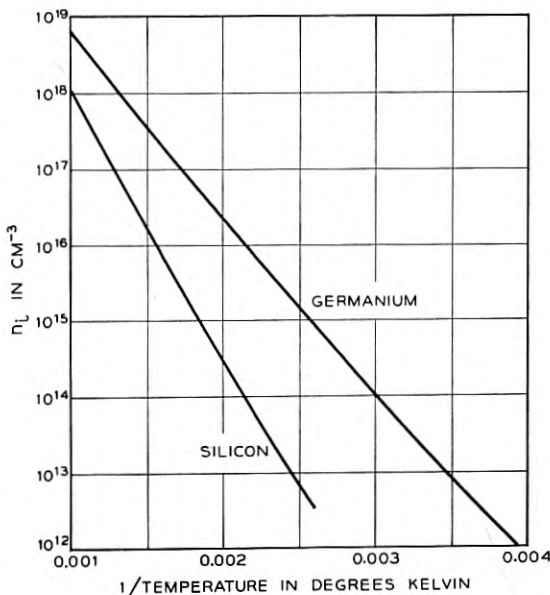


Fig. 2 — Temperature dependences of intrinsic carrier concentrations in germanium and silicon.

temperatures when n_i achieves a very large value it may not be possible to have A^- exceed n_i, and no effect due to the acceptor will be observable. This is simply a mathematical reflection of the fact that the hypothetical compound e^+e^- in (2.1) is highly dissociated at high temperatures so that the holes contributed by the acceptor cannot cause the exhaustion of electrons in the solution.

In Reference 6 the system described in (2.1) was investigated for the purpose of testing (3.4). The concentrations, D^+ and A^-, of lithium and boron respectively were determined by measuring the electrical resistivities of the crystal specimens before and after immersion in molten tin contaning lithium. Some typical results of these experiments are shown in Fig. 3 which contains three D^+ versus A^- isotherms for the temperatures 249°, 310°, and 404°C. For the case shown the tin phase contained 0.18 per cent lithium by weight.

The points in the figure represent experimental findings, while the drawn curves are based on theory. The agreement between theory and experiment is very good, in fact the overall accuracy appears to be better than 1 per cent. These isotherms are only a few of a large group obtained at different temperatures and with differently proportioned external phases. The accuracy in all of these is of the same order.

Various of the features of (3.4) listed above are apparent in the curves of Fig. 3. For example at large values of A^- the curves are straight lines, thus validating (3.5). Also, the inversion of the temperature coefficient of solubility with doping is apparent for the curves cross one another, and whereas, at low dopings (low A^-) the solubility is an increasing function of temperature, at high dopings it decreases with increasing temperature. Finally we note that D^+ remains more or less independent of A^- until A^- exceeds n_i, confirming (3.7). Values of n_i appear in the Figure.

The possible increases in solubility above D_0^+ are really quite large. For example in Fig. 3 the largest increase is of the order of a factor of 10^3. However in some experiments increases of 10^6 have been observed. These effects truly represent profound interactions between impurities which are present in highly attenuated form. Thus the number of atoms per cubic centimeter in crystal silicon is of the order of 5×10^{22} cm^{-3}. Interactions at doping levels as low as 10^{15} cm^{-3}, as appear in Fig. 3, therefore take place at atom fraction levels of about 2×10^{-8}.

In Fig. 4 we show a curve of lithium solubility at room temperature in gallium-doped germanium. The curve is wholly experimental; no attempt has been made to apply theory. The symbols D^+ and A^- are once more used for the donor and acceptor. In this case the curve again exhibits some of the general features required by (3.4). The measure-

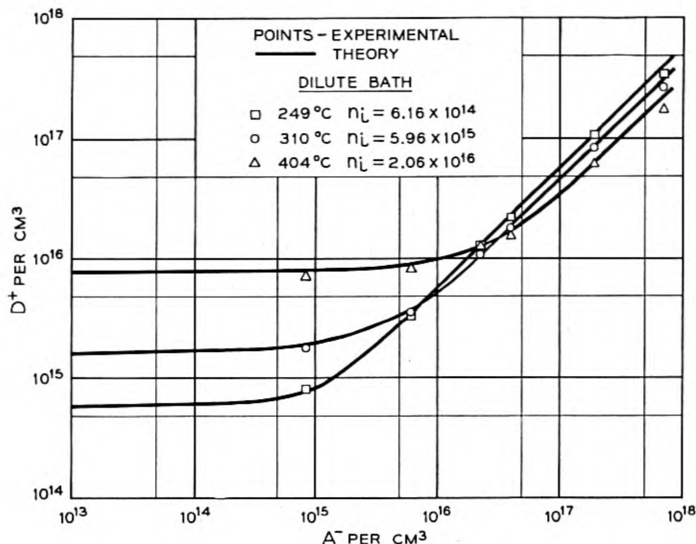


Fig. 3 — Isotherms showing the solubility of lithium D^+, in silicon as a function of boron doping A^-, for an external phase of tin containing 0.18 per cent lithium.

ments were made by saturating gallium-doped germanium crystals with lithium by alloying lithium to the germanium surface at a high temperature, and letting it diffuse in. Following this the crystals were cooled and lithium was allowed to precipitate to equilibrium. In this case the external solution is the precipitate and is of unknown composition.

If the straight line portion of the curve is used to determine D^+/A^- appearing in (3.5), the value of D_0^+ associated with the precipitate as an external phase can be computed by using the value of n_i obtained from Fig. 2 for 25°C. The latter is 3×10^{13} cm^{-3}, and the measured D^+/A^- is 0.85. Application of (3.5) then leads to a value of D_0^+ of 6.6×10^{13} cm^{-3} at 25°C. Since the highest value of D^+ measured in Fig. 4 is 5.5×10^{18} cm^{-3}, the solubility increase here shows a factor of 10^5. Interaction is already apparent at values of A^- as low as 10^{14} cm^{-3}, and since there are 4.4×10^{22} cm^{-3} atoms per cubic centimeter in pure germanium this represents interaction at levels of atom fraction as low as 2×10^{-9}.

IV. FURTHER APPLICATIONS OF THE MASS ACTION PRINCIPLE

In the last section the possibility was mentioned of inverting the sign of the temperature coefficient of solubility, and so preventing impurity

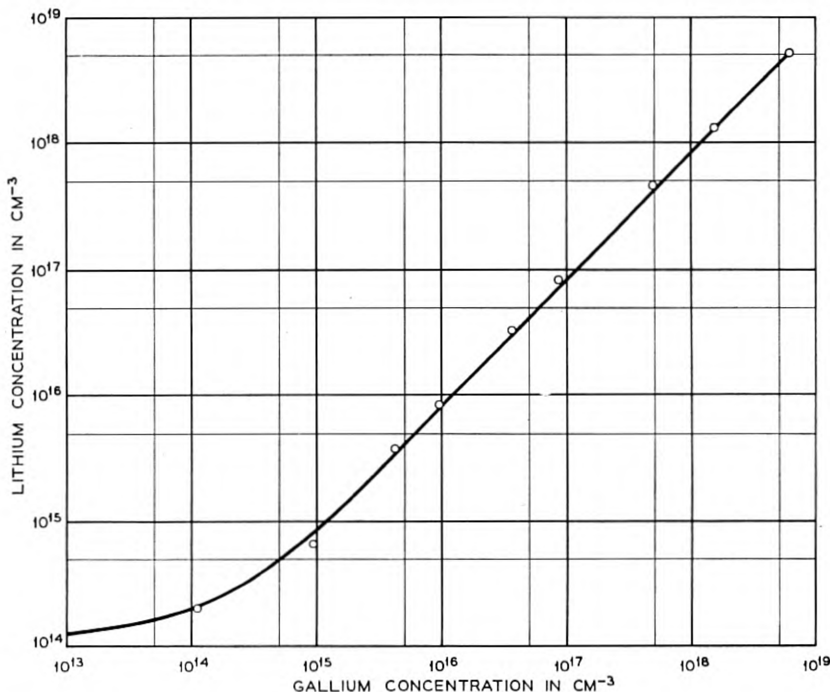


Fig. 4 — Room temperature isotherm showing the solubility of lithium in germanium as a function of gallium doping, the external phase being an alloy of lithium and germanium. The curve merely shows locus of experimental points.

precipitation which might normally occur upon cooling a crystal specimen. An experiment demonstrating this effect is described in Reference 6. Two specimens of germanium, one without added acceptor, and the other containing gallium at an estimated concentration of 1.3×10^{19} cm^{-3}, were saturated with lithium. Table I compares the changes in lithium content observed in these samples with the passage of time. After 25 days no apparent precipitation had occurred in the gallium doped specimen, while precipitation was almost complete in the other.

This result suggests a practical scheme for measuring the concentration of lithium along the solidus curve of the lithium-germanium phase diagram, i.e., the solubility of lithium in solid germanium when the external phase is also composed of germanium and lithium, and probably represents the liquidus phase. This measurement, though desirable, has not been performed before because lithium, diffused into germanium at an elevated temperature, precipitates when the specimen is cooled.

TABLE I

Ga Conc. (cm^{-3})	Li Conc. after saturation (cm^{-3})	Li Conc. after 4 days at room Temp. (cm^{-3})	Li Conc. after 25 days at room Temp. (cm^{-3})
0	1.4×10^{16}	9.0×10^{15}	1.1×10^{15}
1.3×10^{19}	8.0×10^{18}	8.0×10^{18}	8.0×10^{18}

Resistivities then measure only the dissolved lithium although the true solubility at the temperature of saturation includes the precipitated material.

However, we have seen that germanium suitably doped with gallium will not lose lithium by precipitation. Therefore the experiment might be performed in doped germanium. The only difficulty with this suggestion lies in the fact that *doping changes the solubility*. This objection can be overcome through use of (3.4). In terms of that equation D^+ would be measured in the presence of gallium whereas D_0^+, the solubility in undoped germanium, is required. But according to (3.4) if D^+, n_i, and A^- (gallium concentration) are known D_0^+ can be computed. In fact solving (3.4) for D_0^+ yields

$$D_0^+ = \frac{\dfrac{D^+(D^+ - A^-)}{2} + \sqrt{\left[\dfrac{D^+(D^+ - A^-)}{2}\right]^2 + (D^+)^2 n_i^2}}{\sqrt{n_i^2 + \dfrac{D^+(D^+ - A^-)}{2} + \sqrt{\left[\dfrac{D^+(D^+ - A^-)}{2}\right]^2 + (D^+)^2 n_i^2}}} \tag{4.1}$$

The plan is therefore self-evident. Samples of germanium of known suitable gallium contents A^- are to be saturated with lithium at various temperatures. If a judicious choice of gallium content is made the lithium will not precipitate when the specimen is cooled. Therefore the value of D^+ characteristic of the saturation temperature can be determined through resistivity measurements performed at room temperature. Taking n_i from Fig. 2 it then becomes possible to calculate D_0^+ using (4.1).

The crystal specimens employed were cut in the form of small rectangular wafers of dimensions, approximately 1 cm × 0.4 cm × 0.1 cm. On the surfaces of these, small filings of lithium were distributed densely enough so that their average separation was less than the half thickness of the specimen's smallest dimension. The filings were alloyed to the germanium specimen by heating in dry helium for 30 seconds at 530°C. Then the crystals were permitted to saturate with lithium by diffusion from the alloy at some chosen lower temperature. After the period of saturation which ranged from one half hour to as long as 168 days, de-

TABLE II

T °C.	ρ_0 ohm cm	A^- (cm^{-3})	ρ ohm (cm)	D^+ cm^{-3}	D_0^+ (cm^{-3})
25					6.6×10^{13}
100	0.0523	2.2×10^{17}	0.0735	$.9 \times 10^{16}$	2.5×10^{14}
200	0.44	1.3×10^{16}	0.90	7.8×10^{15}	4.6×10^{15}
250	0.1494	4.7×10^{16}	0 652	3.9×10^{16}	2.6×10^{16}
300	0.042	2.9×10^{17}	0.108	2.15×10^{17}	7.3×10^{16}
500	0.00614	4.5×10^{18}	0.0340	4.13×10^{18}	1.7×10^{18}
608	0.00577	5.0×10^{18}	0.049	4.78×10^{18}	2.8×10^{18}
650	0.00584	4.3×10^{18}	0.0178	3.75×10^{18}	2.4×10^{18}

pending on the temperature, the specimen surface was lapped smooth with carborundum paper. Resistivities were then measured by means of a two point probe.

Table II collects the data showing T, the temperature of saturation in degrees centigrade, ρ_0 the resistivity before saturation, A^- the gallium concentration computed from ρ_0, ρ the resistivity after saturation, and D^+ the lithium concentration computed from ρ. The final column shows D_0^+ computed using (4.1) and Fig. 2.

In Table II the 25°C value of D_0^+ has been taken as the value computed in section III in connection with Fig. 4. It might be thought (in view of a later section in this paper) that the 25° and 100°C values of D_0^+ are not as reliable as the others because at the low temperatures involved the solubility of lithium may be influenced by ion pairing as well as electron-hole equilibria. However, Appendix A shows that the possible error is small.

In Fig. 5 D_0^+ is plotted against temperature using these data. The plot is the curve labeled $Ga^- = 0$, and the open circles were obtained by inserting the measured D^+ values (crosses) into (4.1). We notice that the curve has a maximum in the neighborhood of 600°C. The occurrence of a maximum, is a necessity if D_0^+ is to pass to zero, as it must at the melting point of germanium. It is also worth noticing that D_0^+ near room temperature lies in the range of order 10^{13} cm^{-3}, but that its measurement has been effected at concentrations as high as 10^{18} cm^{-3}. This illustrates another application of the electron-hole equilibrium, namely in the determination of solubilities.

With D_0^+ in our possession it is interesting to return to (3.4) and to calculate D^+ as a function of temperature for various levels of A^-. This has been done for values of A^- equal to 10^{15}, 10^{16}, 10^{17}, and 10^{18} cm^{-3}. The curves so obtained appear in Fig. 5, labeled $Ga^- = 10^{15}, 10^{16}, 10^{17}, 10^{18}$ cm^{-3}, respectively. Their most striking common feature is the minimum which appears below 200°C. This minimum introduces a new prob-

lem in preparing samples without precipitate. Thus consider the $A^- = 10^{16}$ cm^{-3} curve. Suppose the specimen is saturated at 200°C. Then according to Fig. 5, if A^- for the specimen is 10^{16} cm^{-3}, D^+ after saturation will be 7×10^{15} cm^{-3}. However, as the sample is cooled it will tend, at first, to become supersaturated. For example it will achieve its maximum supersaturation at about 140°C. where the minimum of the 10^{16} cm^{-3} curve appears. Thereafter it will return to its undersaturated state. In fact at 25°C a concentration of 9.3×10^{15} cm^{-3} could be supported, whereas the solution contains no more than 7×10^{15} cm^{-3} lithium atoms. Some of these may have precipitated as the cooling process passed through the minimum, so that sufficient time must be provided for the process of re-solution.

If the original saturation had taken place at 250°C, the concentration

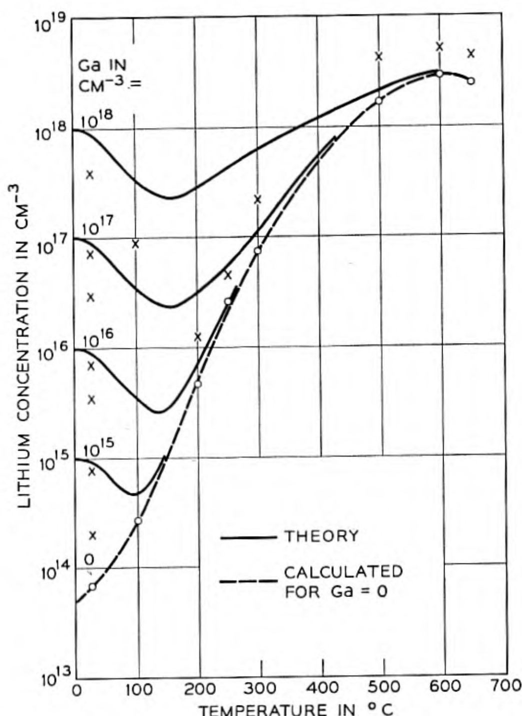


Fig. 5 — Solubility of lithium in germanium as a function of temperature for various gallium dopings. The external phase is an alloy of lithium and germanium. The broken line is the locus of the points (circles) calculated from equation (4.1) for zero gallium concentration. The values of A^- and D^+, used in applying (4.1), correspond to the points shown by X in the illustration. See Table II.

of lithium would have been 2.4×10^{16} cm^{-3}. Since this exceeds the 9.3×10^{15} cm^{-3} supportable at 25°C, such a sample would have contained some precipitate. It was important to avoid these various pitfalls in preparing the specimens used in the above study. Care was taken to insure that this was the case.

We now turn to another application of the electron-hole equilibrium. It has been emphasized that just as a fixed acceptor will increase the solubility of lithium in silicon, a fixed donor should decrease it. In fact in a crystal containing a p-n junction[27] the solubility should be above normal on the *p* side and below normal on the *n* side. The built-in field[28] which exists at the junction is a reflection of this difference in solubility, for if it were not present the concentration gradient created by the disparity in solubilities would cause the lithium to diffuse from the *p* to the *n* side until its concentration was uniform throughout the crystal. Obviously this field is in such a direction as to cause lithium ions to move back to the *p* side.*

Now in both silicon and germanium the oxide layers on the surface can react readily with dissolved lithium. As a result the surface behaves as a sink, and at temperatures as low as room temperature lithium is lost to the surface from the body of the crystal. At higher temperatures the body of the crystal can be exhausted of lithium in a few minutes. There are many experiments which one would like to perform in which the crystal must be maintained without loss of lithium at an elevated temperature for long periods of time.

The application now to be discussed involves utilization of the built-in field at a p-n junction to prevent lithium from reaching the surface where

---

* The distribution of lithium in the space charge region of a p-n junction cannot be computed by the methods advanced thus far. This is because the charge neutrality condition (2.8) is no longer valid. Instead the concentration of lithium is determined by Boltzmann's law,[29] and is given by

$$D^+ = D_*^+ \exp\,[-\,qV/kT]$$

where q is the charge on a lithium ion, V is the local electrostatic potential, and D_*^+ is the concentration where V is zero.

V itself must be determined from Poisson's equation[30]

$$\nabla^2 V = -\frac{4\pi\rho}{\kappa}$$

where ρ is the local charge density and κ is the dielectric constant of the medium. In semiconductors ρ is given in terms of V by[31]

$$\begin{aligned} \rho &= q[H + D^+ - 2n_i \sinh\,(qV/kT)] \\ &= q[H + D_*^+ \exp\,[-\,qV/kT] - 2n_i \sinh\,(qV/kT)] \end{aligned}$$

where H is the local density of fixed donors less the local density of fixed acceptors.

it can attack the oxide. Two specimens of 0.34 ohm cm p-type silicon doped with boron were cut from adjacent parts of a crystal. Each specimen was about 1 cm long, 0.2 cm wide, and 0.15-cm thick. The samples were lapped on No. 400 silicon carbide paper, etched in HF and HNO_3 and sealed in helium-flushed evacuated quartz tubes, one containing a small grain of P_2O_5 . The tubes were then heated at 1,200°C. for 24 hours. This treatment introduced an n-type layer, highly doped with phosphorus and about 0.001-cm thick, into the surface regions of the specimen in the tube containing P_2O_5 . Upon removal from the tube this specimen was lapped on the end to remove the n-skin. Complete removal was determined by testing with a thermal probe.

Small cubes of lithium (0.038 cm on a side) were placed on the ends of both samples (the lapped end of phosphorus-doped one) and alloyed to the silicon for 30 seconds at 650°C in an atmosphere of dry helium. After this treatment the various junction contours should have looked like those in Fig. 6, in which the bottom crystal is shown with the phosphorus-doped skin (cross hatched). During the alloying process a small amount of spherical diffusion of lithium occurs so that small hemispherical n-regions form with the alloy beads as origins. These are shown in Fig. 6.

Next the specimens were heated in vacuum for about six hours at 400°C. Diffusion of lithium into the body of the crystal should occur during this period. However in the sample not protected by the n-type skin lithium should leak to the oxide sink on the surface so that the n-type region due to the lithium should have the pear-shaped contour shown in the upper part of Fig. 7. If the built-in field at the p-n junction

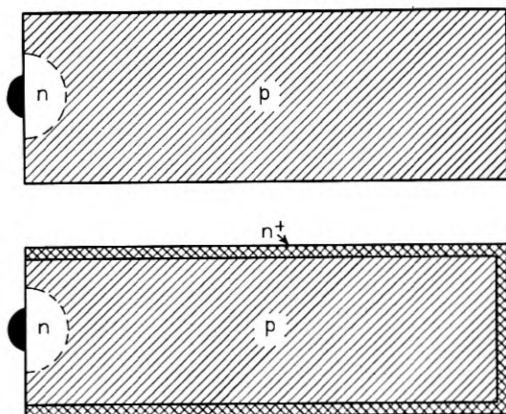


Fig. 6 — Initial stage following alloying in the diffusion experiment to demonstrate the impermeability to lithium of a heavily doped n-type skin on silicon.

formed by the phosphorus layer prevents lithium from reaching the surface, diffusion in the sample with the skin should be plane parallel with a straight front (except at the rear where the skin has been lapped off and lithium can leak out) as the p-n junction contour in the lower part of Fig. 7 indicates.

An acid staining technique[32] which reveals the junction contours should then develop a picture resembling Fig. 7. The two specimens were cut along their long axes and the stain applied to the newly exposed surfaces. The result has been photographed and is shown in Fig. 8 where the crystal on the right has the *n*-skin. The *p*-regions show up dark and the *n*, light. The result agrees with Fig. 7.

In another experiment a crystal completely enclosed in a phosphorus skin was immersed in the tin bath described in Section III. It was discovered that lithium entered the crystal with no evident difficulty, just as though the skin were absent, but once in, could not be driven out by removal of the external source and continued heating. The implication is clear. The built-in field has a rectifying action permitting the lithium to enter the crystal but not to leave. In this sense it performs the same function for the mobile lithium ions as it does for holes in a p-n junction diode.[33]

V. COMPLEX ION FORMATION

In the previous text processes involving the interaction of electrons and holes have been considered. In this section attention will be drawn,

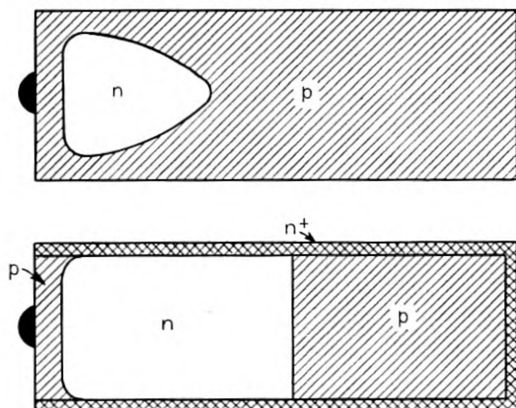


Fig. 7 — Distribution of lithium after an extended period of diffusion at a temperature lower than the alloying temperature — showing leakage out of the crystal in the one case (no-skin) and conservation in the other.

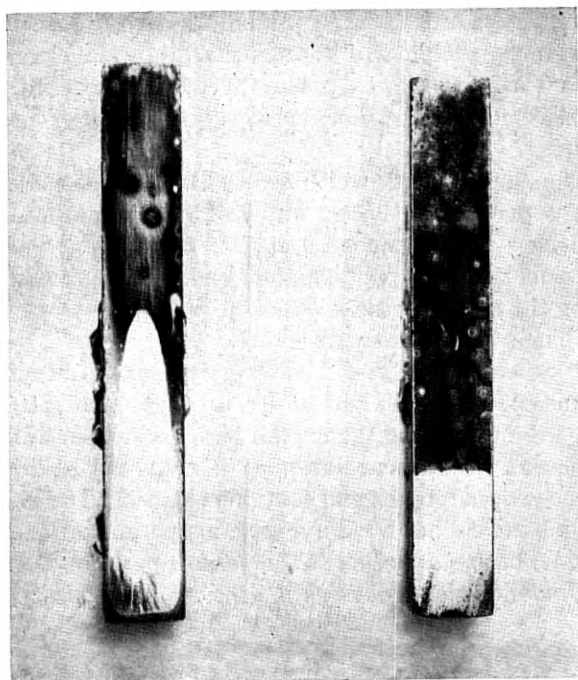


Fig. 8 — Photograph of experimental situation described schematically in Fig. 7.

to the possibility of interactions between the donor and acceptor ions themselves. For example, in (2.1) direct interaction of Li^+ and B^- above 600°C may be possible, especially in view of the mobility of Li^+. Such a reaction was indicated in the work of Reiss, Fuller, and Pietruszkiewicz.[34]

Fig. 9 is of assistance in understanding the nature of these observations. In it are shown plots of the solubility of lithium in silicon. In this case the situation is similar to that involved in the germanium curves of Fig. 5 because the external phase is composed of silicon and lithium and is probably of the liquidus composition. It is formed by simply alloying lithium to the silicon surface. In Fig. 9, Curve A, illustrates how solubility depends on temperature when the silicon is undoped. Curve B, unlike A, is not an experimental plot, i.e., it is not supposed to represent the locus of the points through which it seems to pass. Instead it has been calculated from the theory expounded below. The points themselves are experimental and represent solubility measurements on silicon doped with boron to the level 1.9×10^{18} cm^{-3}.

Curve A possesses a maximum (just as the D_0^+ curve of Fig. 5) in the neighborhood of 650°C. A marked disparity is apparent between solubilities in undoped and doped silicon, the solubility in the latter being greater. Below 500°C this disparity is easily understood. It stems from the electron-hole equilibrium considered previously. However the high solubility in doped silicon at high temperatures is not explicable on this basis since the crystal becomes intrinsic, and e^+e^- is mostly dissociated. To account for this phenomenon Reiss, Fuller, and Pietruszkiewicz invoked the idea of interaction between Li^+ and B^-. They presented the following argument.

At low temperatures lithium ions occupy the interstices of the silicon

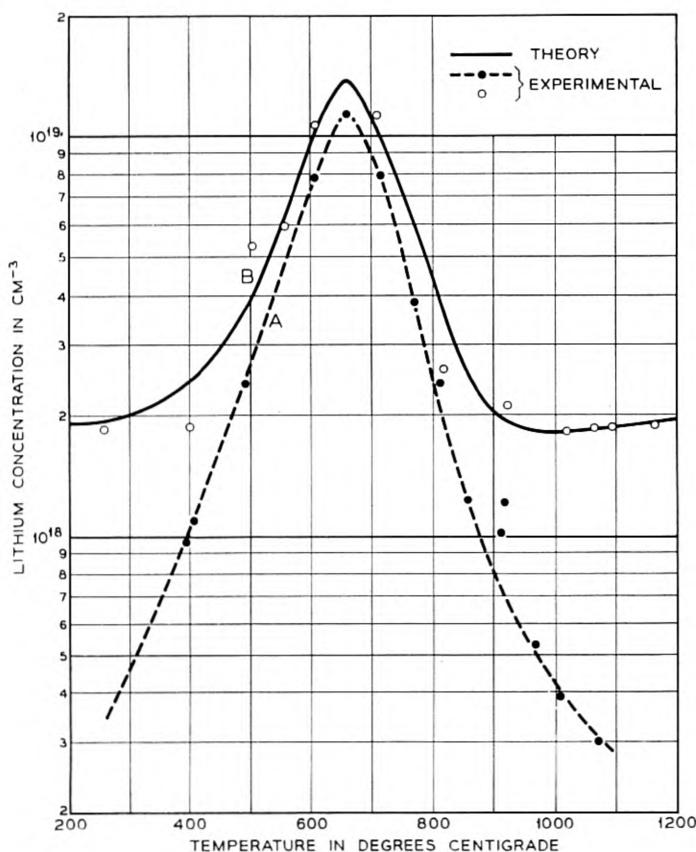


Fig. 9 — Plots showing the solubility of lithium in silicon as a function of temperature. The external phase is an alloy of lithium and silicon. Curve A is for undoped silicon. The locus of the points in B is for silicon doped with about 1.9×10^{18} cm^{-3} boron.

lattice as in Fig. 10. In an interstitial position lithium can approach an oppositely charged boron, but the interaction will be, at the most, coulombic so that an ion pair will form (see later sections). A covalent bond is unable to appear not only because there are no electrons available for it, but also because the lithium ion cannot move to a position where it can satisfy the tetrahedral symmetry inherent in sp^3 hybridization.[35] Calculations (of the sort appearing in the later sections of this paper) show that at high temperatures, at the ion densities involved, ion pairs of the kind depicted in Fig. 10 are completely dissociated.

Suppose, however, that as temperature is raised vacancies dissolve in the silicon lattice, and that one such vacancy occupies a position near

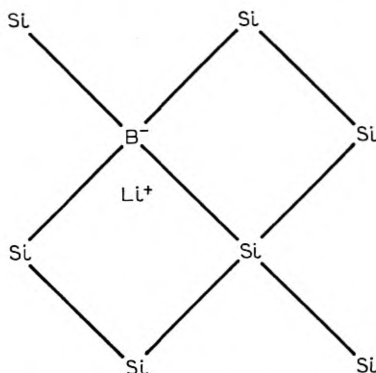


Fig. 10 — Schematic diagram of a silicon lattice showing a lithium ion in an interstitial position near a substitutional boron ion, as it occurs in an ion pair.

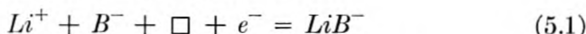
a boron ion, as in Fig. 11, a slight modification of Fig. 10 in which the dots represent electrons (dangling bonds). Unpaired electrons such as these might capture an electron from the valence band of silicon so that the vacancy acquires a negative charge and behaves like an acceptor. It is reasonable to suppose that the positive lithium ion will move into this negative vacancy, in the tetrahedral position, and form a covalent bond as in Fig. 11. The lithium-boron complex so formed retains a negative charge and is thus a complex ion. If the specimen were extrinsic at these high temperatures, there would still appear to be as many net acceptors as before the addition of lithium.*

If the LiB^- compound is stable enough (a question to which we shall

* It is possible that rapid cooling may quench some of these LiB acceptors into the crystal at room temperature. If this is so it should be possible to investigate the associated energy level by Hall measurements in the interval of time before the complexes anneal out. Similar phenomena might be observed in germanium.

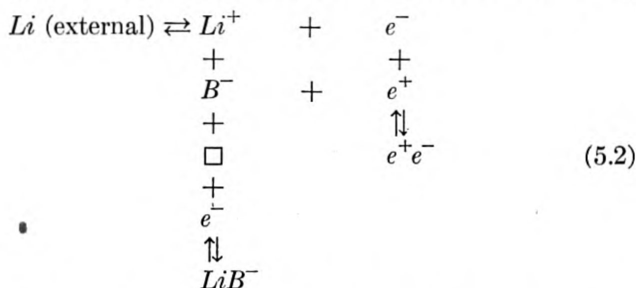
return below) to hold the lithium atom, the solubility of lithium will be determined principally by the density of boron atoms. At low temperatures, vacancies are reabsorbed and the lithium atoms return to their interstitial positions, at quenched-in densities corresponding to the temperatures of equilibration. However, boron acceptors now appear to be compensated since interstitial lithium behaves as a donor. This renders it feasible to measure the concentration of lithium by the determination of resistivity.

The overall reaction may be written in the form



$$Li^{+} + B^{-} + \square + e^{-} = LiB^{-} \tag{5.1}$$

in which $\square$ represents a vacancy. This equilibrium can be grafted onto (2.1) so that the latter becomes (ignoring un-ionized lithium and boron)



$$\begin{array}{lcccc} Li \text{ (external)} \rightleftarrows & Li^{+} & + & e^{-} \\ & + & & + \\ & B^{-} & + & e^{+} \\ & + & & \upharpoonleft\downharpoonright \\ & \square & & e^{+}e^{-} \\ & + & & \\ & e^{-} & & \\ & \upharpoonleft\downharpoonright & & \\ & LiB^{-} & & \end{array} \tag{5.2}$$

The original vertical equilibrium involving holes and electrons loses its significance at high temperatures, and the new vertical reaction becomes important, for both $\square$ and e^{-} appear in increased concentrations. In this way a certain amount of symmetry, insofar as temperature is concerned, is introduced into the problem, i.e., as one equilibrium ceases to dominate

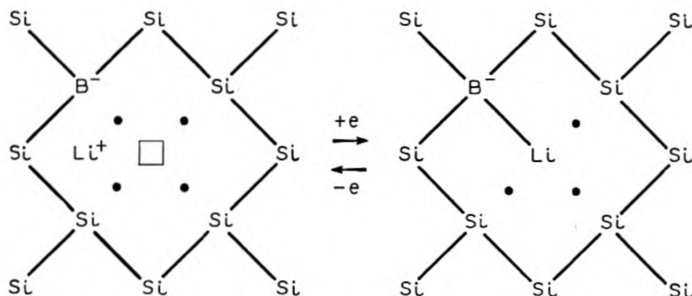


Fig. 11 — Schematic diagram illustrating the reaction in (5.1). The square represents the center of a vacancy and the dots, electrons left unpaired by the occurrence of the vacancy.

the system the other begins to take effect. This symmetry, of course, is necessary for explaining the symmetrical locus of the points around Curve B in Fig. 9.

The scheme (5.2) can be treated quantitatively by applying the mass action principle, but now the symbol D^+ can not be used for the solubility of lithium since the totality of dissolved lithium is distributed between LiB^- and Li^+, and the symbol only applies to the latter. We therefore denote the total concentration of lithium by N_D, and the concentration of LiB^- by C. Then

$$N_D = D^+ + C \tag{5.3}$$

The same argument applies to boron, so that its total concentration will be designated by

$$N_A = A^- + C \tag{5.4}$$

The problem then reduces to specifying N_D as a function of N_A. To accomplish this, to (3.1) and (3.2) is added the mass action expression going with (5.1)

$$\frac{C}{D^+A^-n} = \gamma e^{-(\beta/T)} = \pi \tag{5.5}$$

where γ and β are constants. It has been assumed that the vacancy concentration follows a temperature law of the form $\gamma^* \exp[-\beta^*/T]$ where γ^* and β^* like γ and β are constants. This permits the equilibrium constant when multiplied by the vacancy concentration to assume the form $\gamma \exp[-\beta/T]$ shown in (5.5). In place of (2.8) a new conservation condition,

$$D^+ + p = C + A^- + n \tag{5.6}$$

is introduced. The combination (3.1), (3.2), (5.3), (5.4), (5.5) and (5.6) can be solved so that N_D, the lithium solubility appears as a function of the total boron concentration N_A. Thus

$$N_D = \frac{N_A}{1 + \sqrt{1 + (2n_i/N_D{}^0)^2}} + \sqrt{\left\{\frac{N_A}{1 + \sqrt{1 + (2n_i/N_D{}^0)^2}}\right\}^2 + (N_D{}^0)^2 + \frac{\pi N_A (N_D{}^0)^2[1 + \sqrt{1 + (2n_i/N_D{}^0)^2}]}{2 + \pi (N_D{}^0)^2[1 + \sqrt{1 + (2n_i/N_D{}^0)^2}]}} \tag{5.7}$$

In this equation $N_D{}^0$ like D_0 in (3.4) is the solubility of lithium in undoped silicon, i.e., in silicon from which boron is absent.

All the parameters in (5.7) are independently measurable save π

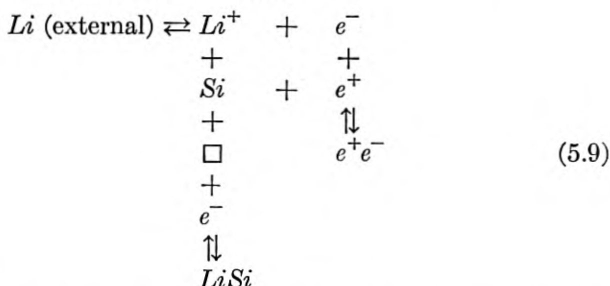
which can be known for all temperatures when γ and β have been determined. Reiss, Fuller, and Pietruszkiewicz used two of the points near Curve B in Fig. 9, above 1,000°C, to define values of N_D for use in (5.7). Then π was computed from (5.7) at these two temperatures. From these values of π, γ and β were determined, and from these, in turn, π was calculated for all temperatures down to 200°C. Using π, N_D was computed from (5.7) over the entire experimental range of temperature. The result is Curve B of Fig. 9 which fits the experimental points very well.

Another check on the validity of the theory (which has not yet been accomplished) would be the following. At high temperatures (5.7) reduces to

$$N_D = N_D{}^0 + \left\{\frac{\pi(N_D{}^0)^2[1 + \sqrt{1 + (2n_i/N_D{}^0)^2}]}{2 + \pi(N_D{}^0)^2[1 + \sqrt{1 + (2n_i/N_D{}^0)^2}]}\right\} N_A \quad (5.8)$$

i.e., N_D is a linear function of N_A with the slope (in brackets) depending upon π. Measurement of this slope at one temperature would thus provide an independent evaluation of π.

A little thought concerning the scheme outlined in (5.2) leads one to wonder why the introduction of boron really increases the solubility of lithium because the same mechanism could be applied to the case in which boron is absent, i.e., to Curve A of Fig. 9. Thus, if B^- is replaced by a silicon atom in Figs. 10 and 11, the entire scheme can be adopted unchanged, except that Si replaces B^-. Thus



and one wonders why LiB^- should be more stable than $LiSi$. A possible answer is the following:

The tetrahedral covalent radius of boron is 0.88 Å.[36] This is to be contrasted with the tetrahedral radius of silicon which is 1.17 Å.[36] When boron is substituted in the silicon lattice it therefore produces considerable local compressive strain. This strain is partially relieved when a vacancy is formed adjacent to the boron. Thus the energy required to form a vacancy near a boron ion in silicon is less than is required for its

formation near a silicon atom. Hence the endothermal heat of formation of LiB^- in (5.2) is reduced substantially (by the amount of the released energy of elastic strain) below the heat of formation of $LiSi$. This accounts for the greater stability of the former.

The compressive strain around a substitutional boron in germanium is also illustrated by ion pairing studies to be described later in Section XII. Its action in that case keeps the ions which form a pair from approaching each other as closely as they otherwise might. Although really quantitative studies of pairing have not yet been performed in silicon, the lattice parameters of germanium and silicon are sufficiently close to render it fairly certain that the same strain exists in the latter as in the former. This lends support to the previous argument.

Before closing this section there is another related topic which is worth mentioning. This concerns part of the explanation of the retrograde solubility observable in the curves of Figs. 5 and 9, i.e., the occurrence of the maxima. The solubilities along these curves are given by (3.3) in the form

$$D_0^+ = K^*/(K^* + n_i^2)^{1/2}$$

Suppose that at low temperatures K^* is an increasing function of temperature and considerably larger than n_i. Then we have the approximation

$$D_0^+ = (K^*)^{1/2} \tag{5.10}$$

in which the solubility D_0^+ must increase with temperature. If n_i increases more rapidly than K^* with temperature, a point will be reached at which n_i^2 in the denominator of the (3.3) in its special form above, exceeds K^* by so much that the latter can be ignored. When this is so another approximation holds,

$$D_0^+ = \frac{K^*}{n_i} \tag{5.11}$$

in which D_0^+ decreases with temperature since n_i increases more rapidly than K^*. Since (5.10) predicts an increase in solubility with temperature at low temperatures and (5.11) a decrease at higher temperatures a maximum occurs somewhere between. The maximum may not be due to this cause alone, however. For example K^* contains the activity, α, in the external phase, and this may vary with temperature in an erratic manner.

In any event the influence of the electron-hole equilibrium on D_0^+ in both silicon and germanium cannot be ignored. The fact that the distribution coefficients of donors and acceptors in silicon are usually some

ten-fold greater than in germanium may be due to the smaller width of the forbidden gap in the latter. This makes for greater values of n_i and according to (3.3) smaller values of D_0^+.

VI. ION PAIRING

The preceding text drew an analogy between semiconductors and aqueous solutions — phenomena such as neutralization, common ion effects, and complex formation have been discussed. Another feature of "wet" chemistry which has appealed to chemists concerns the influence of coulomb forces among ions on the properties of solutions. This subject is of peculiar interest because such forces are well understood, and considerable progress can be made in the quantitative prediction of their effects.

The first really successful theoretical treatment of coulomb forces in solution is the Debye-Hückel theory.[37] This treatment recognizes the long range character of coulomb forces, and endeavors to account for their effects in terms of a communal interaction involving all of the ions in solution. The theory has now been shown to include certain statistical inconsistencies[38] which, however, are of small consequence in dilute solutions where theory and experiment are in excellent agreement.

The central feature of the Debye-Hückel theory is the concept of the ionic atmosphere, i.e., the time average excess concentration of ions of opposite sign which accumulates in the neighborhood of a particular ion. The radius of this atmosphere is measured (order of magnitude-wise) by the now famous Debye length.

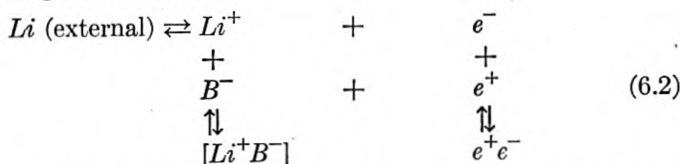
$$L = \sqrt{\frac{k\kappa T}{8\pi q^2 N}} \tag{6.1}$$

in which κ is the dielectric constant of the medium, q is the charge on an ion, and N is the (in this case identical) concentration of both positive and negative ions. As κ decreases or N increases, L becomes smaller so that the atmosphere is more tightly gathered in. As this process continues a stage is reached in which the atmospheres of some of the ions may be best thought of as being fully constituted by a single ion of opposite sign, i.e., an *ion pair* forms. This pair-wise interaction is so intense relative to the communal interaction mentioned above, that insofar as the paired ions are concerned it may be regarded as the only interaction influencing the distribution of the pairs themselves. Unpaired ions may still be treated by the communal Debye-Hückel theory but their concentration must be considered as the true concentration of ions reduced by the

concentration of pairs since the latter possess effectively no fields. In any event when pairing occurs the Debye-Hückel effects are relatively second order, since, even normally, they represent quite small deviations from ideal solution behavior. Under pairing conditions it is desirable, in the first approximation, to focus one's attention on the pairing interaction.

While developing the aqueous solution analogy inherent in our semiconductor model it is natural to inquire whether or not a system like (2.1), in which at least one of the ions can move, will show effects due to coulomb interaction. A preliminary calculation using (6.1) indicates that if coulomb effects are to be observed they are likely to be of the ion pairing variety rather than of the Debye-Hückel type because the dielectric constants of semiconductors are low relative to that of water, e.g., 12 for silicon[39] and 16 for germanium[40] as against 80 for water.[41] The dominance of ion pairing stems, as it will become clear later, from still another feature peculiar to semiconductors. This is the closeness with which two ions of opposite sign can approach one another in semiconductors. In any event experiments are not yet at the stage of sensitivity necessary for the accurate measurement of the small Debye-Hückel effects so that we are virtually compelled to ignore such phenomena.

Fig. 10 is a picture of an ion pair in boron-doped silicon. Corresponding to this process one may sketch in another vertical equilibrium in (2.1) to yield (ignoring un-ionized Li)



$$\begin{array}{lcccc} Li\ \text{(external)} \rightleftarrows & Li^+ & + & e^- & \\ & + & & + & \\ & B^- & + & e^+ & \quad (6.2) \\ & \upharpoonleft\downharpoonright & & \upharpoonleft\downharpoonright & \\ & [Li^+B^-] & & e^+e^- & \end{array}$$

where $[Li^+B^-]$ stands for the ion pair in which the individual ions maintain their polar identities and the binding energy is coulombic. The ion pair is a compound in a statistical sense since as will be seen later the distance between the ions of a pair is distributed over a range of values. The interaction between Li^+ and B^- is to be distinguished from that shown in (5.2). The latter occurs at high temperatures whereas the former is presumably limited to low temperatures, below 300°C.

The quantitative aspects of ion pairing were first considered by Bjerrum[42] and later by Fuoss[43] who placed Bjerrum's theory on a somewhat more acceptable basis. Fuoss's theory, however, suffers from some of the same limitations as Bjerrum's. Nevertheless the Bjerrum-Fuoss theory is capable of satisfying experimental data over broad ranges of conditions. In the next section we present a brief resumé of this theory together with relevant criticism and its relation to a more refined theory due to Reiss.

VII. THEORIES OF ION PAIRING

Fuoss begins by considering a solution of dielectric constant κ, conaining equal concentrations, N, of ions of opposite sign. When equilibrium has been achieved each negative ion will have another ion (most probably positive) as a nearest neighbor, a distance r away from it. Fuoss discounts the possibility that the nearest neighbor will be another negative ion, and proceeds to calculate what fraction of such nearest neighbors lies in spherical shells of volumes, $4\pi r^2\,dr$, having the negative ions at their origins. If this fraction is denoted by $g(r)\,dr$, it may be evaluated as follows.

In order for the nearest neighbor to be located in the volume, $4\pi r^2\,dr$, two events must take place simultaneously. First the volume, $4\pi r^3/3$, enclosed by the spherical shell must be devoid of ions, or else the ion in the shell would *not* be the nearest neighbor. Since $g(x)dx$ is the probability that a nearest neighbor lies in the shell, $4\pi x^2\,dx$, the probability that a nearest neighbor does not lie in this shell is $1 - g(x)dx$. From this it is easily seen that the chance that the volume $4\pi r^3/3$ is empty is

$$E(r) = 1 - \int_a^r g(x)\,dx \tag{7.1}$$

where a is the distance separating the centers of the two ions of opposite sign when they have approached each other as closely as possible.

The second event which must take place is the occupation of the shell $4\pi r^2\,dr$ by *any* positive ion. The chance of this event depends on the time average concentration of positive ions at r. This concentration is bound to exceed the normal concentration N by an amount depending on r, because of the attractive effect of the negative ion at the origin. It may be designated by $c(r)$. The probability in question is then

$$4\pi r^2 c(r)\,dr \tag{7.2}$$

The chance $g(r)\,dr$ that the nearest neighbor lies in the shell $4\pi r^2\,dr$ is therefore the product of (7.1) by (7.2), i.e., the product of the probabilities of the two events required to occur simultaneously. This leads to the relation

$$g(r) = \left(1 - \int_a^r g(x)\,dx\right) 4\pi r^2 c(r) \tag{7.3}$$

an integral equation whose solution is

$$g(r) = \exp\left[-4\pi \int_a^r x^2 c(x)\,dx\right] 4\pi r^2 c(r) \tag{7.4}$$

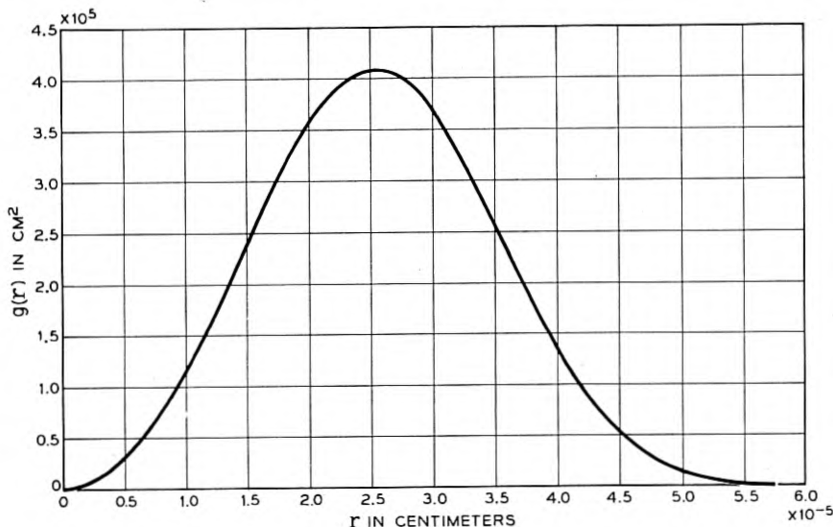


Fig. 12 — Distribution of nearest neighbors in a random assembly of particles for a concentration of 10^{16} cm^{-3}.

That (7.4) solves (7.3) is easily demonstrated by substitution of the latter into the former.

If there were no forces of attraction between ions then $c(r)$ would equal N, and if a is take equal to zero (7.4) reduces to

$$g(r) = 4\pi r^2 N \exp(-4\pi r^3 N/3) \tag{7.5}$$

This function is plotted in Fig. 12 for the case $N = 10^{16}$ cm^{-3}. Note that the position of the maximum, the most probable distance of location of a nearest neighbor, occurs near the value of r equal to $(3/4\pi N)^{1/3}$. This is the radius of the average volume per particle when the concentration is N, i.e. the volume, $1/N$.

In order to write $g(r)$ for the case of coulombic interaction it is necessary to compute $c(r)$ under these conditions. Fuoss (after Bjerrum) reasoned as follows. If a theory can be constructed which depends only upon the characteristics of *near* nearest neighbors (nearest neighbors at small values of r) then the force of interaction experienced by the nearest neighbor can be assumed to originate completely in the coulomb field of the negative ion at the origin. This is predicated on the argument that both positive and negative ions develop atmospheres of opposite sign which are superposed when the two ions are close to one another. The result is a cancellation of the net atmosphere leaving nothing for the two

ions to interact with but themselves. Thus the potential energy of interaction, for near nearest neighbors will be

$$-\frac{q^2}{\kappa r} \tag{7.6}$$

For small values of r, therefore, $c(r)$ can be derived from Boltzmann's law and is given by

$$c(r) = h \exp [q^2/\kappa kTr] \tag{7.7}$$

where h is a constant. Guided by the requirement that $c(r)$ should equal N at infinite distance from the central negative ion, h was set equal to N giving, finally,

$$c(r) = N \exp [q^2/\kappa kTr] \tag{7.8}$$

The assumption that a theory could be developed depending only on near nearest neighbors proved reasonable, but the choice of $h = N$ in (7.8) leads to certain logical difficulties. Thus the average volume dominated by a given negative ion is evidently $1/N$. If (7.8) is summed over this volume the result, representing the number of positive ions in $1/N$, should be unity since there are equal numbers of positive and negative ions. Unfortunately, the result exceeds unity by very large amounts except for very small values of N, i.e., for very dilute solutions. We shall return to this point later.

If (7.8) is inserted into (7.4) the resulting $g(r)$ has the form typified by Fig. 13. First, there is an exponential maximum occurring at $r = a$, followed by a long low minimum, and this by another maximum which like the one in Fig. 12 occurs, not far from $r = (3/4\pi N)^{1/3}$, if N is not too large. For small values of N the minimum occurs at

$$r = b = q^2/2\kappa kT \tag{7.9}$$

The function $g(r)$ is actually normalized in (7.4) so that the area under the curve is unity. The second maximum corresponds to the most probable position for a nearest neighbor in a random assembly, i.e., to the maximum in Fig. 12. Essentially the first maximum has been grafted onto Fig. 12 by the interaction at close range which makes it probable that short range neighbors will exist. At high values of N the region under the first maximum becomes so great that enough area is drained (by the condition of normalization) from the second maximum to make it disappear entirely. At this point the minimum is replaced by a point of inflection. More will be said concerning this phenomenon later.

Fuoss chooses to define all sets of nearest neighbors inside the mini-

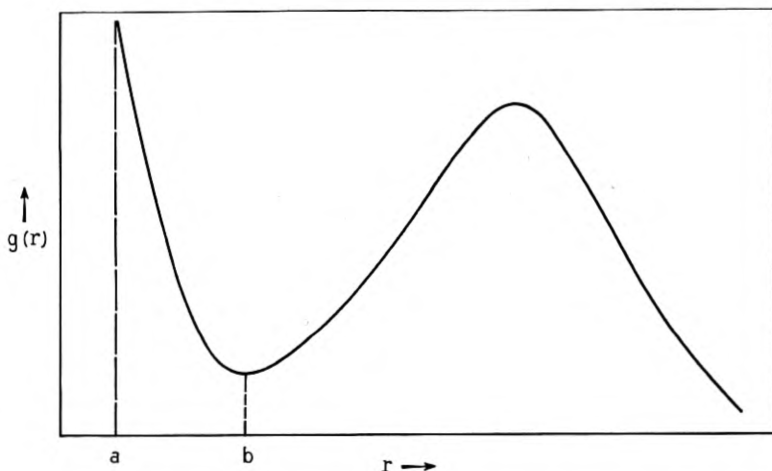


Fig. 13 — Schematic distribution of neighbors in an assembly of particles when forces of interaction are present. Repulsive forces are reflected in the appearance of a distance a, of closest approach of two particles, attractive forces by the exponential maximum at a.

mum, i.e., inside $b = q^2/2\kappa kT$, as ion pairs, and the rest as unpaired. No thought is given to the small fraction of nearest neighbors which involves ions of like sign, as it must be small inside $r = b$. Nor is any thought given to the possibility that a given positive nearest neighbor may be the nearest neighbor of two negative ions simultaneously. Such a coincidence would be very improbable at a distance short enough to be within $r = b$. Thus if the entire theory can be made to depend on what happens inside b, its foundations are reasonable, except for the choice of $h = N$.

To obviate this difficulty Fuoss had further to devise a means of performing all calculations under conditions where the choice of $h = N$ was not inconsistent. He assumed (following Bjerrum) that paired and unpaired ions were in dynamic equilibrium and that the law of mass action could be applied to this equilibrium. Thus if P represents the concentration of pairs, $N - P$ denotes the concentration of unpaired ions of one sign and the mass action expression is

$$\frac{P}{(N - P)^2} = \Omega \tag{7.10}$$

where Ω is an equilibrium constant independent of concentration. At infinite dilution, where the assignment $h = N$ is valid, Ω should be the same as at higher concentrations. Therefore (7.4) can be used to evalu-

ate Ω at infinite dilution, and the value so obtained employed at higher concentrations.

Besides the inconsistency of the choice, $h = N$, the form (7.4) contains another objectionable feature. This is revealed by a more rigorous treatment devised recently by Reiss,[44] and has to do with the factor,

$$\exp\left[-4\pi \int_a^r x^2 c(r)\, dx\right],$$

in (7.4). It can be shown that this factor is inconsistent with the supposition that the nearest neighbor to a given negative ion interacts only with that ion and no other. Fortunately, in Fuoss's scheme $g(r)$ given by (7.4) needs to be used only at infinite dilution, and then only for such values of r as lie inside b. Under this condition and in this range the exponential factor in question can be replaced by unity from which it deviates only slightly. Thus the form of $g(r)$ used eventually is

$$g(r) = 4\pi r^2 N \exp\left[q^2/\kappa kTr\right] \qquad (7.11)$$

Ω is computed as follows. At infinite dilution P tends toward zero so that (7.10) becomes

$$\frac{P}{N} = \Omega N \qquad (7.12)$$

But P/N is the fraction of ions paired which by definition is the fraction of nearest neighbors lying inside $r = b$. From the definition of $g(r)$, P/N is evidently given by

$$\frac{P}{N} = \int_a^b g(r)\, dr = 4\pi N \int_a^b r^2 \exp\left[q^2/\kappa kTr\right] dr \qquad (7.13)$$

which upon substitution in (7.12) yields

$$\Omega = 4\pi \int_a^b r^2 \exp\left[q^2/\kappa kTr\right] dr \qquad (7.14)$$

The evaluation of Ω in this way permits one to base the entire theory on the distribution of *near* nearest neighbors, so that all the assumptions which demand this procedure are validated.

Using the computed Ω in (7.10) P can be evaluated, and also $N - P$ which as the concentration of *free* ions of one species measures the thermodynamic activity of that species. In this manner it is possible to calculate the *equilibrium* effects of coulomb interaction insofar as solution properties are concerned. To treat transport phenomena such as ionic mobility in an applied electric field Fuoss assumes that paired ions repre-

senting neutral complexes are unable to respond to the applied field and so do not contribute to the overall mobility. The mobility of unpaired ions is assumed to be μ_0, the mobility observable at infinite dilution. The apparent mobility μ at any finite concentration is then μ_0 reduced by the fraction P/N of ions paired. Thus

$$\mu = [1 - (P/N)]\mu_0 \tag{7.15}$$

The Bjerrum-Fuoss theory when applied to real systems reproduces the experimental data very well, although the parameter a, the distance of closest approach, needs to be determined from the data itself.

The concept of a pair defined in terms of the minimum occurring at b, becomes rather vague when that minimum vanishes in favor of a point of inflection. At this stage triplets and other higher order clusters form and the situation becomes very complicated.

In Reference 44, Reiss has developed a more refined theory of pairing. Instead of avoiding the use of an inconsistent $g(r)$ by introduction of the mass action principle, an attempt is made to provide a rigorous form for $g(r)$, which proves to be the following

$$g(r) = \exp\,[-4\pi r^3 N/3]\; 4\pi r^2 h \exp\,[q^2/\kappa kTr] \tag{7.16}$$

in which

$$h = 1 \bigg/ \left[\int_a^\infty \exp\,[-\,4\pi r^3 N/3]\; 4\pi r^2 \exp\,[q^2/\kappa kTr]\, dr\right. \tag{7.17}$$

It is also shown that the activity of an ionic species, measured by $N - P$ in the Bjerrum-Fuoss theory, is measured by $\sqrt{hN}$ in the more rigorous theory. The distribution (7.16) suffers neither from an inability to conserve charge in the volume $1/N$ (as does (7.4)) nor from any inconsistency involving the interaction of a nearest neighbor with other ions than the one to which it is nearest neighbor [as does (7.4)].

When $\sqrt{hN}$ computed by (7.17) is compared with $(N - P)$ computed according to (7.10) and (7.14), for arbitrary values of κ, a, T, and N, the results are almost identical. This shows the virtue of the Bjerrum-Fuoss theory, and in fact, suggests that in most cases it should be used for calculation rather than the more refined theory, for the latter involves rather complicated numerical procedures.

The refined theory can also be adapted to the treatment of transport phenomena.[45] Thus in place of $g(r)$ it is possible to write a distribution function $\Gamma(\vec{r})$, specifying the fraction of nearest neighbors lying in the volume element $d\vec{r}$, in a system in the steady state rather than at equilibrium. In the presence of an applied field the distribution loses its spheri-

cal symmetry and it must be defined in terms of the volume elment $d\vec{r}$, lying at the vector distance $\vec{r}$, rather than in terms of the spherical shell of volume, $4\pi r^2 dr$. In reference (44) it is shown that

$$\Gamma(\vec{r}) = \exp[-4\pi r^3 N/3] c(\vec{r}) \tag{7.18}$$

where $c(\vec{r})$ is the density function in the non-equilibrium case, and is determined by the equation

$$\frac{kT}{q} \nabla^2 c + c\nabla^2 \psi + \nabla c \cdot \nabla \psi = 0 \tag{7.19}$$

after suitable boundary conditions have been appended. The quantity ψ, designates the local electrostatic potential, determined by the ions as well as the applied field. These equations are restricted specifically to the semiconductor case in which the negative ion is unable to move.

The current carried by nearest neighbors in the volume element $d\vec{r}$ in unit volume of solution is

$$J(\vec{r}) = -\exp[-4\pi r^3 N/3] c(\vec{r}) \mu_0 \nabla[\psi + (kT/q)\, \ell n\, c(\vec{r})] \tag{7.20}$$

Using these equations it proves possible in reference 45 to provide a more refined version of (7.15) in which the mobility of nearest neighbors inside $r = b$ need not be considered zero, nor those outside $r = b$ be considered perfectly free and possessed of the mobility μ_0. In fact the average mobility of a nearest neighbor separated by a distance r from its immobile partner proves to be

$$\bar{\mu} = \frac{\mu_0}{2(1-F)} \left(\left[\frac{\varepsilon^2}{3r^2} + \frac{4\varepsilon}{3r} + 2 \right] \exp(-\varepsilon/r) + 2F\left(\frac{\varepsilon}{3r} - 1\right) \right) \tag{7.21}$$

where

$$\varepsilon = q^2/\kappa kT \tag{7.22}$$

and

$$F = \left(\frac{\varepsilon^2}{2a^2} + \frac{\varepsilon}{a} + 1 \right) \exp(-\varepsilon/a) \tag{7.23}$$

For values of r greater than ε (7.21) can be approximated by

$$\frac{\bar{\mu}}{\mu_0} = \frac{1}{2} \left(\frac{\varepsilon^2}{3r^2} + \frac{4\varepsilon}{r} + 2 \right) \exp(-\varepsilon/r) \tag{7.24}$$

and is therefore a function of ε/r. Fuoss's b corresponds to $r = \varepsilon/2$ or to $\varepsilon/r = 2$. Fig. 14 contains a plot of $\bar{\mu}/\mu_0$ versus r for $T = 400°K$, $a = 2.5 \times 10^{-8}$ cm, $q = 4.77 \times 10^{-10}$ statcoulombs, and $\kappa = 16$. Note that

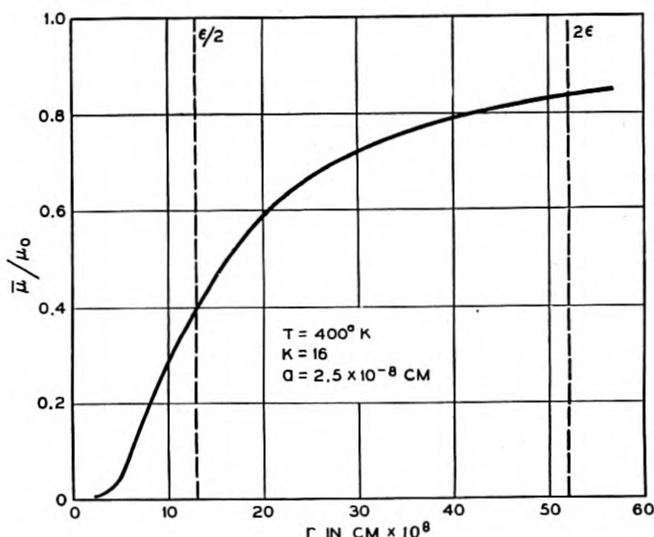


Fig. 14 — Average mobility (calculated from the refined theory of pairing) of a mobile ion in a pair as a function of the distance from its immobile neighbor. The example shown corresponds to a substance having $a = 2.5 \times 10^{-8}$ cm $\kappa = 16$ at a temperature of 400°K.

at $r = \varepsilon/2 = b$, $\bar{\mu}/\mu_0$ is near 0.5 which is the average value of Fuoss's $\bar{\mu}/\mu_0$ for ions taken from either side of $r = b$. Therefore a certain symmetry with respect to $r = b$ does exist, tending to justify Fuoss's model. According to (7.24) $\bar{\mu}/\mu_0$ is 0.8 by the time $r = 3\varepsilon/2 = 3b$, independent of the value of a. In other words an ion located a short distance beyond b *does* have practically complete mobility as the Bjerrum-Fuoss theory assumes.

The refinement of (7.15) which occurs can be written as follows

$$\mu = \left\{ \frac{2h}{(1 - F)} \int_a^\infty \left[\left(2r^2 + \frac{4\varepsilon r}{3} + \frac{\varepsilon^2}{3} \right) + 2F \left(\frac{\varepsilon r}{3} - r^2 \right) \exp\,(\varepsilon/r) \right] \exp\,(-\,4\pi r^3 N/3)\, dr \right\} \mu_0 \tag{7.25}$$

Comparison of μ/μ_0 computed from (7.25) with $1 - (P/N)$ appearing in (7.15) over wide ranges of conditions again reveals an excellent correspondence and further substantiates the Bjerrum-Fuoss theory. Since calculations employing the latter are so much simpler it is expedient to regard the cruder theory as an accurate approximation to the more refined one. This practice will be followed from now on.

VIII. PHENOMENA ASSOCIATED WITH ION PAIRING IN SEMICONDUCTORS

In this section we shall discuss some of the phenomena which are to be expected in semiconductors when ion pairing takes place. At the time of writing several of these phenomena have been investigated quantitatively in germanium and casually in silicon. A report on these studies will be given in the later sections of this paper.

In the meantime it is fitting to inquire into the peculiarities which arise because a semiconducting medium rather than a dielectric liquid is involved. The possible means of detecting and measuring ion pairing in semiconductors are numerous, and many of them do not have counterparts in aqueous solution. This implies that a host of new phenomena are to be expected, many of which are peculiar to semiconductors.

Some distinctions between semiconductors and liquids are apparent at once. Thus ions are not always mobile in semiconductors at temperatures where ion pairing is pronounced. Lithium is exceptional in this respect, being mobile in germanium and silicon down to very low temperatures. In fact ion pairing has been observed in germanium containing lithium down to dry ice temperatures, and even below. Another difference is the low dielectric constant of semiconductors as compared with water. Furthermore, in semiconductors, charge balance need not be maintained by the ions themselves, but may be effected by the presence of holes or electrons. Although charged the latter entities need not be considered in pairing processes since, as particles, they possess effective radii of the order of their thermal wavelengths which may exceed 20 Angstroms at the temperatures involved. At these distances very little coulomb binding energy would be available. Under certain rare conditions the screening effect of these mobile carriers may make some contribution. This may be particularly the case when *relaxation processes* (to be discussed later) are carried out in poorly compensated specimens of semiconductor, since such processes involve phenomena between ions separated by large distances.

A very obvious distinction is the fact that ions in a semiconductor occupy a lattice, and cannot therefore move through a continuum of positions, as in the case of liquid solutions. Furthermore the lattice may introduce elastic strain energy into the binding energy of a pair. This influence will alter the value of a, the distance of closest approach, when the latter is chosen so as to achieve the best fit between theory and experiment. As the extent of pairing is extremely sensitive to the magnitude of a, its measurement provides a useful tool for exploring the state of strain in the neighborhood of an isolated impurity. We shall demonstrate

this application later in connection with the strain in the neighborhood of a substitutional boron in germanium.

Aside from its bearing on the minimum distance a, the existence of the lattice will be ignored in the following considerations.

The values of a, typical of semiconductors, are generally of the order of 2 Angstroms as against 6 to 8 Angstroms for ions in liquids. This results from the fact that liquid ions are generally solvated. The consequence to be expected, and indeed found, is that ion pairing will be far more pronounced in semiconductors than in liquids of comparable dielectric constant.

The fact that ions have limited mobilities in semiconductors can be turned to advantage by choosing a system such as lithium and boron in silicon in which only one species of ion, in the case mentioned, lithium, is mobile. Under these conditions it is possible to obviate the clustering phenomenon, mentioned previously, which appears in liquids at high ion concentrations. Clustering is prevented because the immobile ions are uniformly distributed in a random manner, having been grown into the crystals at high temperature where pairing and related processes are unimportant. The obvious complications attending cluster formation can therefore be avoided.

Of course, mobility, being limited to a single species of ion is also an advantage in the theory of the transport phenomena, in such systems.

It is convenient to list some of the effects due to pairing which are to be expected in semiconductors. We do so in the following compilation.

(A) *Equilibrium Phase Relations*

From (6.2) it is apparent that the pairing equilibrium should affect the solubility of lithium in silicon. The same must be true for germanium doped with an acceptor. Although such effects probably occur, they are accompanied by influences arising from the other possible equilibria. As a result the situation is somewhat complex and it is not easy (see Appendix A) to produce experimental conditions under which pairing will be evident. For this reason quantitative investigations along these lines have not yet been attempted.

(B) *Variation of Energy Levels*

When an ion pair is formed of a donor and acceptor, both the donor and acceptor levels are altered. Thus the proximity of the negative acceptor ion increases the difficulty of return to the donor state for an electron, (i.e. the donor level is raised). Likewise the acceptor level is

lowered. In ion pairs it is in fact to be expected that the donor level will be moved up into the conduction band and the aceptor level down into the valence band.* This change in energy level structure should be apparent in Hall coefficient measurements at low temperature. Experiments of this sort have been conducted and are reported in this paper. Under certain conditions this phenomenon may be useful for the elimination of trapping[46] levels from the forbidden gap.

(*C*) *Change of Carrier Mobility*

Ion pairs possess dipolar fields, and consequently, scattering cross-sections very much smaller than those of point charges. The addition of lithium to a sample under such conditions that more than half the added lithium becomes paired should therefore *increase* rather than *decrease* the mobility of holes. The latter effect is the one to be expected in the absence of pairing. In other words not only carriers but also the scatterers are removed by compensating the acceptor with donor. Experiments of this sort have been performed. They are described later in this paper. Since they allow us to measure the degree of pairing with good accuracy they have been very valuable in validating the theory, and also in exploring the nature of the potential function in the neighborhood of an isolated acceptor.

(*D*) *Relaxation Times*

A semiconductor containing unpaired donors and acceptors at one temperature can be cooled to a lower temperature, and the impurities should then pair. If the temperature is lowered sufficiently, the pairing process will be slow enough to be followed, kinetically, by observing any parameter (such as carrier mobility) sensitive to pairing. Experiments of this sort have been performed and will be described later.

The process of pairing can be characterized by a calculable relaxation time, which depends on the acceptor concentration, the diffusivity of the mobile donor, the dielectric constant, and the charges on the ions among other things. The measured time can therefore be used as a means of determining any one of these parameters.

(*E*) *Diffusion*

It is evident that pairing should reduce the diffusivity of a mobile donor. Studies of diffusion in the presence of an immobile acceptor should

* A rough calculation indicates that about 0.5 e.v. would be required to place an additional electron on an ion pair.

therefore reveal the action of pairing. Experiments of this sort have been performed and will also be described in this paper.

The reduction in the diffusivity of a donor such as lithium may be desirable in certain places.

(F) *Direct Transport*

Diffusion studies suffer from the defect that ion pairing produces a concentration dependent diffusivity. (See Appendix B). For this resaon a very desirable measurement would involve determining the amount of a mobile donor like lithium transported by an electric field through a uniformly saturated specimen of semiconductor. This flux, together with information concerning the level of saturation, should provide a direct measure of the mobility of lithium under homogeneous conditions. Formula (7.15) or its refinement (7.25) could then be applied directly to the results.

The above list is by no means complete, for there are still other techniques available for measurement, for example nuclear and paramagnetic resonance. Enough has been given however to indicate the wide range of phenomena which ion pairing in solids can affect. In liquids, only A and F are of any consequence. It is important to realize that not only do these phenomena serve as tools for the study of ion pairing, but that ion pairing, when properly understood, can serve as a tool for the study of the phenomena themselves.

IX. PAIRING CALCULATIONS

The evaluation of Ω according to (7.14) presents somewhat of a problem because the integral must be arrived at numerically. Fortunately, the literature contains tables[47] of the integral in what amounts to dimensionless form. The transformation

$$\xi = q^2/\kappa kTr \tag{9.1}$$

is introduced and then Ω is shown to be given by

$$\Omega = 4\pi[q^2/\kappa kT]^3\, Q(\alpha) \tag{9.2}$$

where

$$\alpha = q^2/\kappa kTa \tag{9.3}$$

and $\log_{10} Q(\alpha)$ is tabulated in Table III.

In a specimen in which the numbers of donors and acceptors are un-

TABLE III

α	$\log_{10} Q(\alpha)$	α	$\log_{10} Q(\alpha)$
2.0	$-\infty$	18.0	2.92
2.5	−0.728	20.0	3.59
3.0	−0.489	25.0	5.35
4.0	−0.260	30.0	7.19
5.0	−0.124	35.0	9.08
6.0	0.016	40.0	11.01
7.0	0.152	45.0	12.99
8.0	0.300	50.0	14.96
9.0	0.470	55.0	16.95
10.0	0.655	60.0	18.98
12.0	1.125	65.0	21.02
14.0	1.680	70.0	23.05
16.0	2.275	75.0	25.01
		80.0	27.15

equal* (7.10) may be written as

$$\frac{P}{(N_A - P)(N_D - P)} = \Omega \tag{9.4}$$

where N_A and N_D are, respectively, the total densities of acceptors and donors.

This equation has the following solution for P/N_D, the fraction of donors paired.

$$\frac{P}{N_D} = \frac{1}{2}\left(1 + \frac{1}{\Omega N_D} + \frac{N_A}{N_D}\right) - \sqrt{\frac{1}{4}\left(1 + \frac{1}{\Omega N_D} + \frac{N_A}{N_D}\right)^2 - \frac{N_A}{N_D}} \tag{9.5}$$

Inspection of (9.5) reveals that for given N_A and Ω, P/N_D is a decreasing function of increasing N_D.

Very often, P/N_D is measured in an experiment, and from this it is desired to calculate a, the distance of closest approach. For such purposes the form (9.5) is not very convenient. In fact an entirely different procedure is to be preferred. Suppose P/N_D is denoted by θ, and θ is substituted into (9.4), into which (9.2) has been inserted. We obtain

$$\log_{10} Q(\alpha) = \log_{10}\left[\frac{1}{4\pi}\left(\frac{\kappa kT}{q^2}\right)^3 \frac{\theta}{(N_A - \theta N_D)(1 - \theta)}\right] \tag{9.6}$$

A knowledge of θ thus suffices to determine $\log_{10} Q(\alpha)$, from which, in turn, α can be determined by interpolation in Table III. Then (9.3) can be used for the evaluation of a.

* This is a situation which cannot arise in liquids, since there, charge balance must be maintained by the ions themselves. It can occur when the ions are of different charge, but then things are complicated by the formation of triplets, etc., in addition to pairs.

TABLE IV

T°K	Ω (cm³)	T°K	Ω (cm³)
100	2.2×10^{2}	400	2.3×10^{-17}
150	6.45×10^{-7}	500	1.54×10^{-18}
200	3.42×10^{-11}	600	3.0×10^{-19}
225	1.28×10^{-12}	700	1.03×10^{-19}
250	8.79×10^{-14}	800	4.7×10^{-20}
300	1.61×10^{-15}		

Experiments which will be described later indicate that in germanium, gallium and lithium can approach as close as 1.7×10^{-8} cm. Using this value of a, and $\kappa = 16$, $q = 4.77 \times 10^{-10}$ statcoulombs, the values of Ω appearing in Table IV were computed from (9.2)

With these values, P/N_D, the fraction of donors paired can be computed from (9.5) as a function of temperature and N_A for the simplest case, i.e., the one for which $N_A = N_D$. Fig. 15 contains plots showing these dependences. It must be remembered that all other things remaining the same P/N_D will be greater than the values shown in Fig. 15 when $N_D < N_A$.

A rather important integral to which reference shall be made later is

$$I(r_2, r_1) = \int_{r_1}^{r_2} x^2 \exp (q^2/\kappa kTx)\, dx \tag{9.7}$$

The integral appearing in (7.14) is a special case of (9.7) with $r_1 = a$, and $r_2 = b$. $I(r_2, r_1)$ has been evaluated over a considerable range. To facilitate matters the transformation

$$x = (q^2/\kappa kT)\, \lambda \tag{9.8}$$

has been employed. In this notation r_1 and r_2 transform to ρ_1 and ρ_2, and

$$I(r_2, r_1) = (q^2/\kappa kT)^3 \int_{\rho_1}^{\rho_2} \lambda^2 \exp (1/\lambda)\, d\lambda = (q^2/\kappa kT)^3 i(\rho_2, \rho_1) \tag{9.9}$$

Figs. 16 and 17 contain plots of $i(\rho_2, 0.05)$ out to $\rho_2 = 5$. The choice of ρ_1 equal to 0.05 was rather unfortunate since for $\kappa = 16$, and $T = 300$°K it corresponds to $\rho_1 = 2.5 \times 10^{-8}$ cm. Since acceptors like gallium possess values in respect to lithium as low as 1.7×10^{-8} cm $i(\rho_2, 0.05)$ is not much use in these cases. The choice 0.05 was made before the experimental data on gallium was available. Below we shall describe a method for extending $i(\rho_2, \rho_1)$ to cases where r_1 is less than 2.5×10^{-8} cm.

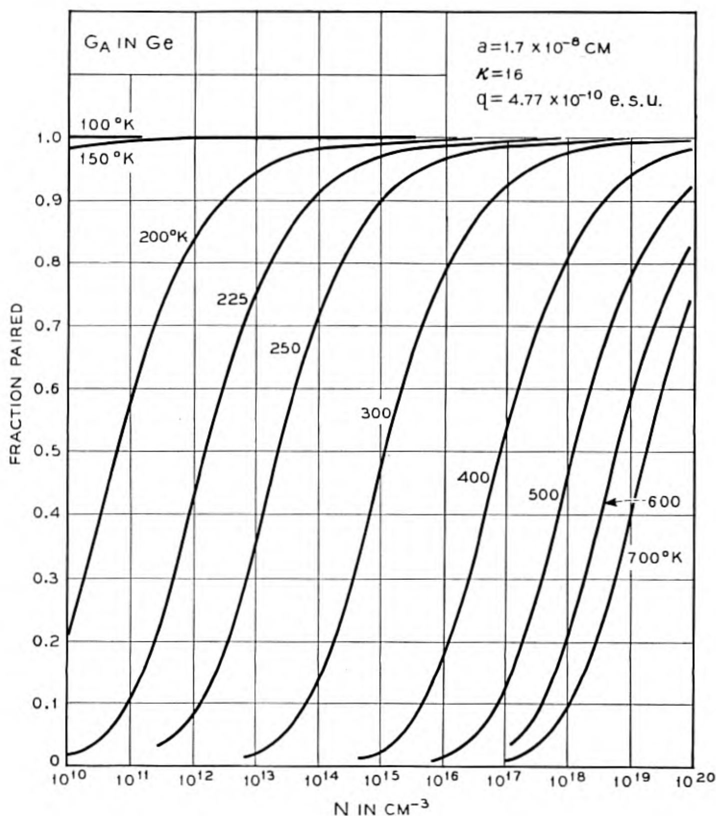


Fig. 15 — Fraction of ions paired, assuming equal densities of positive and negative ions, calculated as a function of temperature and concentration from equation (9.5). The situation illustrated might apply to gallium and lithium in germanium in view of the choice of a and κ.

Fig. 16 covers the range from $\rho_2 = 0.05$ to 0.08 and involves a logarithmic scale because of the sharp variation of i in this range. (This points up the sensitivity of the degree of pairing to the magnitude of a.) Fig. 17 extends the curve to $\rho_2 = 5$. When ρ_2 exceeds 5, $i(\rho_2, 0.05)$ can be obtained from the formula

$$i(\rho_2, 0.05) = 3865 + \frac{(\rho_2)^2}{2} + \frac{(\rho_3)^3}{3} \tag{9.10}$$

In order to determine $i(\rho_2, \rho_1)$ when $\rho_1 \geqq 0.05$, the following formula may be used.

$$i(\rho_2, \rho_1) = i(\rho_2, 0.05) - i(\rho_1, 0.05) \tag{9.11}$$

Finally for cases in which $\rho_1 < 0.05$, Table III can be used. Thus

$$i(\rho_2, \rho_1) = Q(1/\rho_1) - Q(20) + i(\rho_2, 0.05) \tag{9.12}$$

where $1/\rho_1$, and 20 are α values in Table III.

X. THEORY OF RELAXATION

In Section VIII attention was drawn to the fact that ion pairing in semiconductors can be made to occur slowly enough so that its kinetics can be followed. It is possible to characterize these kinetics by a relaxation time τ, which we shall endeavor to calculate in the present section.

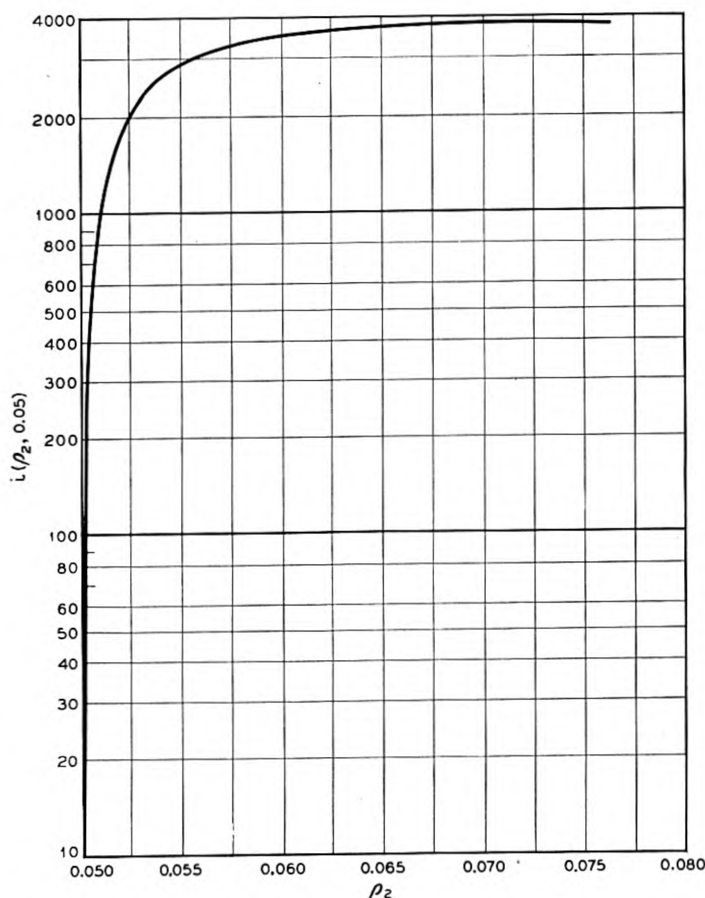


Fig. 16 — Plot, for small values of ρ_2 of $i(\rho_2, 0.05)$ from (9.9).

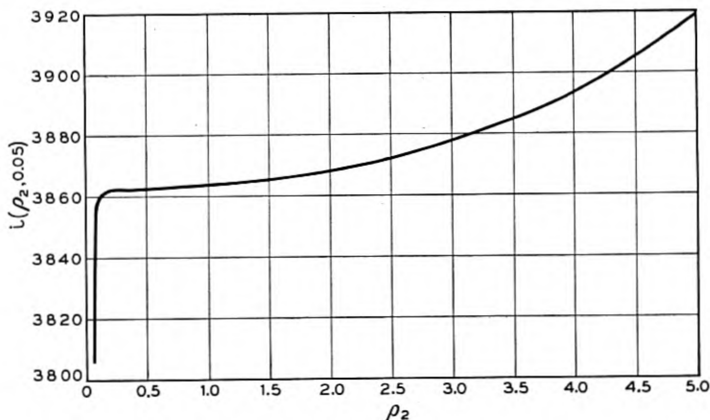


Fig. 17 — Plot, for larger values of ρ_2, of $i(\rho_2, 0.05)$ from (9.9).

Suppose a system is first maintained at a temperature high enough to prevent pairing, and then, at an instant designated as zero time, is suddenly chilled to a temperature at which pairing takes place. One thereby has a system which would normally contain pairs but which finds itself with donors and acceptors which are uniformly and randomly distributed. Since the donors are assumed mobile, a process ensues whereby they drift toward acceptors until an equilibrium is established in which each acceptor develops an atmosphere of donors with density $c(r)$, given by (7.7).

This final state in which the atmosphere is fully developed is the paired state characteristic of the lower temperature. The relaxation time to be defined must measure the interval required for the near completion of the above process.

In order to acquire physical feeling for the phenomenon, we begin with some simple considerations. In particular a system will be dealt with containing equal numbers of positive and negative ions. This restriction can be lifted later.

Now, to a first approximation the pairing phenomenon may be regarded as a trapping process in which mobile, positive donor atoms are captured by the negative acceptors. Thus, suppose each acceptor is imagined to possess a sphere of influence of radius R, beyond which its force field may be considered negligible, and inside which a positive ion is to be regarded as captured. This picture immediately emphasizes certain subtleties which require discussion before further progress can be made.

In the crudest sense one might reason that the probability of an encounter between a positive ion and a negative trap would depend on the

product of the densities of both. These densities must be equal because when a positive ion is trapped the resulting ion pair is neutral so that a trap is eliminated simultaneously. If these equal densities are designated by n, we arrive at the second order rate law

$$-\frac{dn}{dt} = k_2 n^2 \tag{10.1}$$

where k_2 is a suitable constant, and t is time.

This law would be perfectly valid if the mean free path of a mobile positive ion were large compared to the distance between ions and the probability of sticking on a first encounter were small. The trapping *cross-section* rather than the movement prior to trapping would determine the trapping rate. In this case the rate would certainly depend on the concentrations of both the traps and the ions being trapped.

On the other hand, in our case, not only is the mean free path of a positive ion much smaller than the distance between ions, but the sticking probability is high. A given ion must *diffuse* or make many random jumps before encountering a trap and upon doing so is immediately captured. Therefore, the rate of reaction is diffusion controlled.

Because of the random jump process a given mobile ion is most likely to be captured by its *nearest neighbor* during the first half of relaxation, and relative to the degree of advancement of the trapping process, the density of traps may be considered constant. This leads to first order kinetics rather than second,* i.e., to

$$-\frac{dn}{dt} = k_1 n \tag{10.2}$$

where n is the density of untrapped ions.

By definition k_1 is the fraction of ions captured in unit time, i.e., the probability that one ion will be captured per unit time. Its reciprocal must be the average lifetime of an ion. This lifetime

$$\tau = \frac{1}{k_1} \tag{10.3}$$

shall be defined as the *relaxation time* for ion pairing. A rough calculation of τ can be made quickly. Thus, suppose that the initial concentrations of donors and acceptors are equally N. About each fixed acceptor can be described a sphere of volume, $1/N$. On the average this sphere should be occupied by one donor which according to what has been said above, will eventually be captured by the acceptor at the center. In the mind, all

* The phenomenon stems from the fact that first and second order processes are almost indistinguishable during the first half of the reaction, but also from the fact that the diffusion control prevents the process from being a true second order one, although its departure from second order may be small.

the spheres can be superposed so that an assembly of donors N in number is contained in the volume $1/N$, at the density N^2. The problem of relaxation is then the problem of diffusion of these donors to the sink of radius R, at the center of the volume. The bounding shell of the sphere may be considered impermeable, thus enforcing the condition that each donor shall be trapped by its nearest neighbor. Since the diffusion problem has spherical symmetry the radius, r, originating at the center of the sink at the origin may be chosen as the position coordinate. At $r = R$, the density, ρ, of diffusant may be considered zero. The radius, L, of the volume, $1/N$, is so large compared to R, that in the initial stages of diffusion L may be regarded as infinite.

In spherical diffusion to a sink from an infinite field, a true steady state is possible, and this steady state is quickly arrived at when the radius, R, of the sink is small.[48] Under this condition concentration is described by

$$\rho = A - \frac{B}{r} \tag{10.4}$$

where A and B are constants. Furthermore at early times n is still N, the initial concentration at $r = L \approx \infty$, so that

$$\rho(\infty) = N^2 \tag{10.5}$$

In addition we know that

$$\rho(R) = 0 \tag{10.6}$$

These boundary conditions suffice to determine A and B in (10.4), and yield

$$\rho = N^2\left[1 - \frac{R}{r}\right] \tag{10.7}$$

Now the rate of capture ($-(dn/dt)$ in (10.2)) is obviously measured by the flux of ions into the spherical shell of area, $4\pi R^2$, which marks the boundary of the sink. This flux is given according to Fick's law[49] by

$$4\pi R^2 D_0 \left(\frac{\partial\rho}{\partial r}\right)_{r=R} = -\frac{dn}{dt} \tag{10.8}$$

where D_0 is the diffusivity of the donor. Substituting (10.7) into (10.8) yields

$$4\pi N^2 R D_0 = -\frac{dn}{dt} \tag{10.9}$$

During the initial stages of trapping the right side of (10.2) may be

written as k_1N, i.e.,

$$k_1 N = -\frac{dn}{dt} \tag{10.10}$$

Equating the left sides of (10.9) and (10.10) gives

$$k_1 = 4\pi NRD_0$$

or

$$\tau = \frac{1}{k_1} = \frac{1}{4\pi NRD_0} \tag{10.11}$$

It now remains to choose a value for the capture radius, R. A reasonable guess may be made as follows: Around each acceptor there is a coulomb potential well of depth

$$V = -q^2/\kappa r \tag{10.12}$$

Since the average thermal energy is kT, it seems reasonable to regard an ion as trapped when it falls to a depth kT in this well. Thus, inserting kT on the left of (10.12) and R for r on the right leads to

$$R = q^2/\kappa kT \tag{10.13}$$

and upon substitution in (10.11) we obtain

$$\tau \approx \frac{\kappa kT}{4\pi q^2 ND_0} \tag{10.14}$$

This result, obtained by crude reasoning, is actually quite close to the more rigorous value derived below. Furthermore, the above derivation is useful in providing insight into the physical meaning of the relaxation time.

The chief difficulty with the preceding lies in the arbitrary choice of R, and is a direct consequence of the long range nature of coulomb forces. Another difficulty arises because the distribution of donors about acceptors is eventually specified by (7.7) so that at $r = R = q^2/\kappa kT$

$$\frac{\partial c}{\partial r} = -\frac{h}{e}\left\{\frac{\kappa kT}{q^2}\right\} \tag{10.15}$$

Since this slope has a negative value the trap exhibits some aspects of a source rather than a sink which could only produce a positive concentration gradient. This last objection will not be serious when h is very small since, then the final value of $c(r)$ beyond $r = q^2/\kappa kT = R$ will be effectively zero, as would be required for a perfect sink.

The last point raises still another question: What happens when the sink is not perfect, i.e. where the equilibrium state does not involve complete pairing?

All these difficulties can be removed by a more sophisticated treatment of the diffusion problem. Thus, retain the sphere of volume, $1/N$, enclosing N donors at the density N^2. However, the equations of motion of these donors are altered to account for the fact that besides diffusing they drift in the field of the acceptor at the origin. Thus the flux density of donors will be given by

$$J^*(r, t) = -D_0 \left\{ \frac{\partial \rho}{\partial r} + \left\{ \frac{q^2}{\kappa k T r^2} \right\} \rho \right\} = -D_0 \left\{ \frac{\partial \rho}{\partial r} + \frac{R}{r^2} \rho \right\} \tag{10.16}$$

where R has been substituted for $q^2/\kappa kT$. Equation (10.16) is obtained by adding to the diffusion component,

$$-D_0 \frac{\partial \rho}{\partial r}$$

of the flux density, the drift component,

$$-\frac{\mu_0 q}{\kappa r^2} \rho,$$

where μ_0 is the mobility of a donor ion and $-q/\kappa r^2$ is the field due the acceptor at the origin. The Einstein relation[50]

$$\mu_0 = qD_0/kT \tag{10.17}$$

has also been used to replace μ_0 with D_0.

The spherical shell bounding the volume, $1/N$, of radius

$$L = \left(\frac{3}{4\pi N} \right)^{1/3} \tag{10.18}$$

is regarded as impermeable, so we obtain the boundary condition

$$J^*(L, t) = 0. \tag{10.19}$$

Furthermore an arbitrary inner boundary, $r = R$, is no longer defined but use is made of the real boundary, $r = a$, i.e., the distance of closest approach, at which is applied the condition

$$J^*(a, t) = 0 \tag{10.20}$$

As before, the initial condition may be expressed as

$$\rho = N^2 \qquad t = 0 \qquad a < r < L \tag{10.21}$$

The continuity equation,[51] in spherical coordinates takes the form

$$\frac{1}{r^2}\frac{\partial}{\partial r}\{r^2 J^*\} = -\frac{\partial \rho}{\partial t} \tag{10.22}$$

Substitution of (10.16) into (10.22) gives, finally,

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left\{r^2\frac{\partial \rho}{\partial r} + R\rho\right\} = \frac{1}{D_0}\frac{\partial \rho}{\partial t} \tag{10.23}$$

Equations (10.23), (10.21), (10.20) and (10.19) form a set defining a boundary value problem, the solution of which is $\rho(r, t)$, from which, in turn, $J^*(r, t)$ can be computed. It then remains to compute (dn/dt) in (10.2) from J^*. The former is not simply $4\pi R^2 J^*$ (as in (10.8)) because now J^* is not defined unambiguously, being a function of r. $J^*(R, t)$ might be employed but then the method is no less arbitrary than the simple one described above.

Fortunately, nature eliminates the dilemma. It is a peculiarity of spherical diffusion, when the sink radius is much smaller than the radius of the diffusion field, that after a brief transient period, $4\pi r^2 J^*(r)$, except near the boundaries of the field, becomes practically independent of r, and depends only on t. This feature is elaborated in Appendix C. Since in our case the radius of the field is of order, L, and the effective radius of the sink is of order, R, and $L \gg R$, it may be expected that this phenomenon will be observed. In fact its existence has been assumed previously in the derivation of (10.4).

Under such conditions it does not matter how the radius of the sink is defined so long as $4\pi R^2$ is multiplied by $J^*(R)$ and not the value of J^* at some other location.

The boundary value problem, (10.23), (10.21), (10.20), (10.19) is solved in Appendix C, and it is shown there that the value of $4\pi r^2 J^*(r)$ obtained after the transient has passed is closely approximated by

$$4\pi r^2 J^*(r) = -\frac{4\pi q^2 N^2 D_0}{\kappa k T} e^{-t/\tau} \tag{10.24}$$

with

$$\tau = \frac{\kappa k T(N - M)}{4\pi q^2 N^2 D_0} \tag{10.25}$$

where

$$M = 1/4\pi \int_a^L r^2 \exp\,[q^2/\kappa k T r]\, dr \tag{10.26}$$

The close connection between M defined by (10.26) and h defined by (7.17) is apparent. Thus in (7.17) when $r = L$, $\exp[-4\pi r^3 N/3]$ is e^{-1}, and for larger values of r this exponential quickly forces the convergence of the integral. Therefore the values of h and M will be almost equal. This is not surprising since they are meant to be the same thing, i.e., the average concentration, $c(\infty)$, of donors at infinite distance in the equilibrium atmosphere of an acceptor. Both quantities are computed so as to conserve charge in this atmosphere.

At large values of N, M proves to be much smaller than N so that (10.25) reduces to (10.14), validating the crude treatment, for τ in (10.24) is obviously the relaxation time. This is easily seen by writing

$$-\frac{dn}{dt} = -4\pi r^2 J^*(r) = \frac{4\pi q^2 N^2 D_0}{\kappa k T} e^{-t/\tau} \tag{10.27}$$

from which one derives by integration

$$n = M + (N - M)e^{-t/\tau} \tag{10.28}$$

According to (10.28) at $t = 0$, $n = N$, the correct initial density for unpaired ions. At $t = \infty$, $n = M$, also the correct density, i.e., the density at large values of r, when equilibrium is achieved. Obviously τ plays the role of the relaxation time, since by differentiation of (10.28)

$$-\frac{d(n - M)}{dt} = \frac{(n - M)}{\tau} \tag{10.29}$$

which is to be compared with (10.2) and (10.3).

Values of M can be computed using formulas (9.10), (9.11), and (9.12) and Figs. 16 and 17 since the integral in (10.26) is one of the i integrals Fig. 18 shows some values of M, computed in this way for the temperatures 206°, 225°, 250°, and 300°K, for a semiconductor where the value of $a = 2.5 \times 10^{-8}$ cm, $\kappa = 16$, and q $= 4.77 \times 10^{-10}$ statcoulombs. The plots are of M versus N. Note that the values of M are generally much less than N, the disparity increasing with lower temperatures and larger N.

It is also possible to calculate τ for the above system in its dependence upon N and T. To do this the value of D_0 must be known as a function of temperature. Fuller and Severiens[52] have measured the diffusivities of lithium in germanium and silicon down to about 500°K. These data plot logarithmically against $1/T$ as excellent straight lines. In Fig. 19, we show an extrapolation of the line for lithium in germanium down to the neighborhood of 200°K. From this figure it is possible to read values of

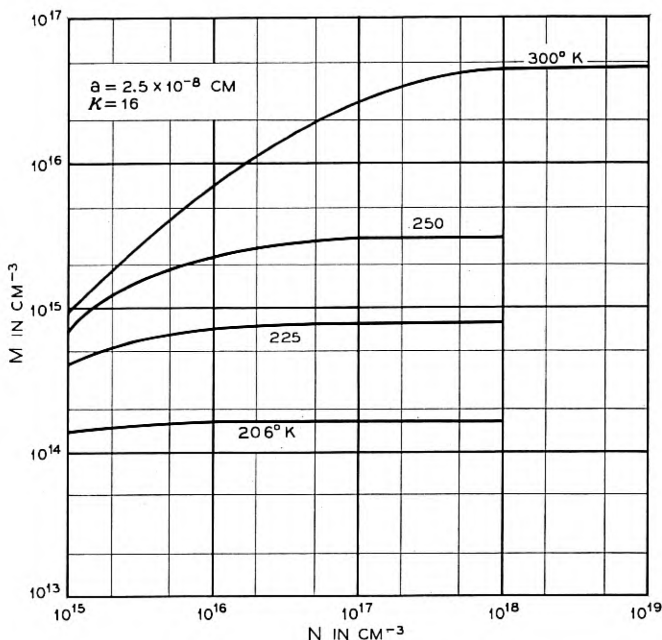


Fig. 18 — Dependence of constant M defined by (10.26) on temperature and concentration, for particular values of a and κ.

D_0 for germanium to which the system of Fig. 18 refers, since κ has been chosen at 16.

Using Figs. 18 and 19, Fig. 20 was computed. It shows τ plotted in seconds versus N for the same temperatures appearing in Fig. 18. These curves show that at values of N as low as 10^{15} cm^{-3} relaxation times are short enough to be observable down to 200°K, being at the most some 50 hours in extent. The value of N makes a big difference. For example at 200°K the relaxation time is only 4 minutes with $N = 10^{18}$ cm^{-3}. Presumably, at 10^{18} cm^{-3}, relaxation could be observed down to much lower temperatures.

It is interesting to note that insofar as M hardly appears in τ, the latter is independent of the distance of closest approach, a. Since a is to some extent empirical this is a fortunate circumstance, and the measurement of τ may provide an accurate means of determining, N, D_0, κ, or q, whichever parameter is regarded as unknown. Furthermore κ as a macroscopic parameter has real meaning in τ since the forces involved may be regarded as being applied over the many lattice parameters separating the drifting donor from its acceptor.

This section will be closed by indicating how the restriction to systems containing equal numbers of donors and acceptors might be lifted. Thus, suppose N_A exceeds N_D. Then there will be $N_A - N_D$ mobile holes maintaining charge neutrality. To a first approximation these will screen the $N_A - N_D$ uncompensated acceptor ions so that the N_D donors will see effectively only N_D acceptors. Thus in first approximation τ can be computed for this system by replacing N in the preceding formulas by N_D.

Of course it is possible that there will be a further effect. Thus the mobile holes will probably shield some of the compensated acceptors as well. This will lead to a further (probably small) reduction in τ, over and above that obtained by replacing N by N_D. We shall not go into this in the present paper, because in most of the experiments performed N_D was near N_A. In the few exceptions the crude correction, suggested above, can be used.

XI. INVESTIGATION OF ION PAIRING BY DIFFUSION

Most of the theoretical tools required for the study of ion pairing have now been provided, and attention will be turned to experiments which

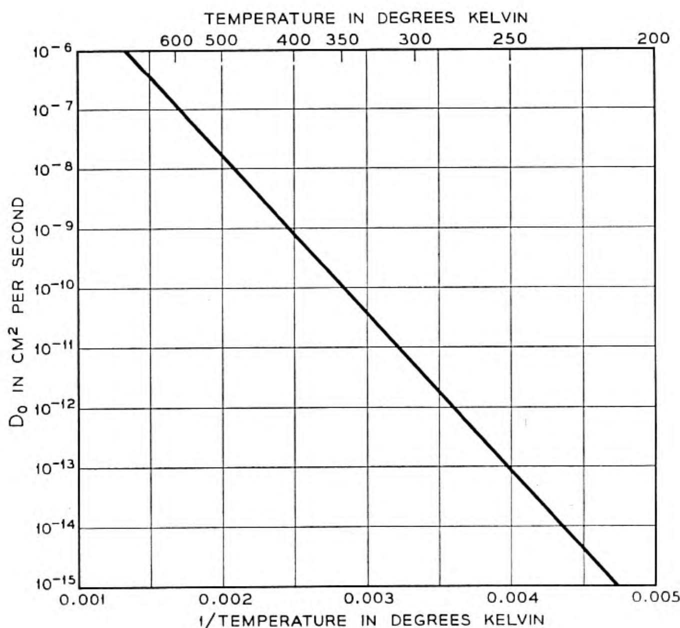


Fig. 19 — Diffusivity of lithium in germanium extrapolated from the data of Fuller and Severiens.

have been performed in this field. A fairly large group of these exist, and it remains to describe them in detail. We shall begin with the study of the diffusion of lithium in p-type germanium.

At the outset a matter having to do with the *diffusion potential* demands attention. This is the potential which arises, for example, in p-type material, because the mobility of a hole is so much greater than the mobility of a lithium ion. In consequence, holes diffuse into regions containing high concentrations of lithium more rapidly than lithium ions can diffuse out to maintain space charge neutrality. As a result such re-

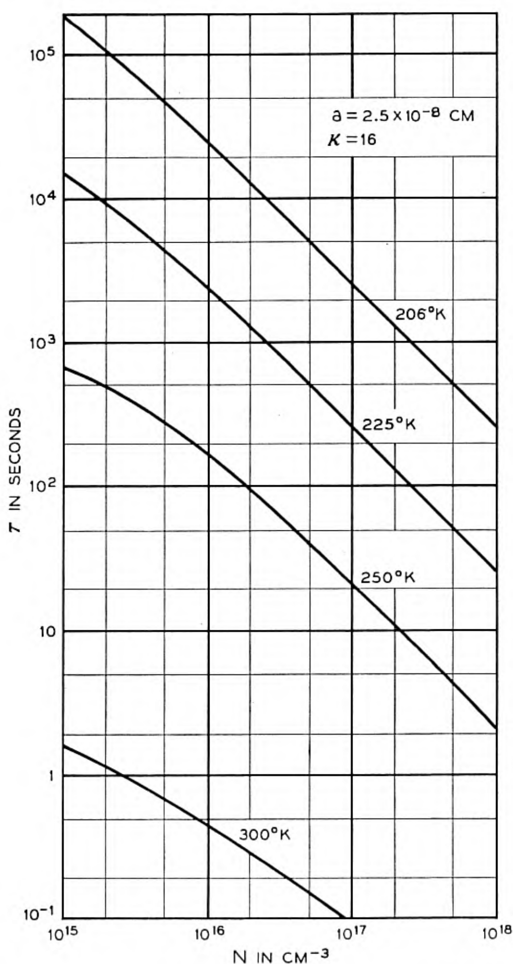


Fig. 20 — Relaxation time as a function of temperature and concentration computed from equation (10.25) using the data of Figs. 18 and 19.

gions develop positive potentials and a field exists tending to expel lithium. This causes the lithium to drift as well as diffuse so that Fick's law[49] is no longer valid.

The most that can be done toward the elimination of diffusion potentials is to minimize them so that no local space charge exists. At equilibrium, this corresponds to the condition[53]

$$N_D - N_A = 2n_i \sinh(qV/kT) \tag{11.1}$$

where V is the local electrostatic potential. It is always permissible to assume that fast moving electrons and holes are in equilibrium relative to diffusing ions. If a material which is p-type everywhere is being considered, (11.1) can be simplified to

$$N_A - N_D = n_i \exp\,[-qV/kT] \tag{11.2}$$

In Appendix D it is proved that (11.2) will be valid everywhere within a region where N_A is constant and greater than N_D, provided that N_D does not fluctuate through ranges of the order N_A in a distance less than

$$\ell = \sqrt{\frac{\pi\kappa kT}{q^2 N_A}} \tag{11.3}$$

Under most conditions of experiment ℓ will be of the order of 10^{-5} cm. Unfortunately many of the experiments described in this section (particularly those performed at 25°C.) involve diffusion layers as thin as 10^{-6} cm. As a result space charge will exist and the diffusion potential will not always be minimized. Even if it is minimized so that (11.2) is satisfied the residual field will still aid diffusion and lead to higher apparent diffusivities. Therefore the effect cannot be ignored even when minimization has been achieved.

In the absence of space charge the drift component of flux density due to the field is easily computed. It will be given by

$$-\mu \frac{\partial V}{\partial x} N_D \tag{11.4}$$

According to (11.2)

$$-\frac{\partial V}{\partial x} = \frac{kT}{q(N_A - N_D)} \frac{\partial N_D}{\partial x} \tag{11.5}$$

so that (11.4) becomes

$$\begin{aligned} -\frac{\mu kT}{q}\left(\frac{N_D}{N_A - N_D}\right)\frac{\partial N_D}{\partial x} &= -\frac{\mu_0 kT}{q}\left(1 - \frac{P}{N_D}\right)\left(\frac{N_D}{N_A - N_D}\right)\frac{\partial N_D}{\partial x} \\ &= -D_0\left(1 - \frac{P}{N_D}\right)\left(\frac{N_D}{N_A - N_D}\right)\frac{\partial N_D}{\partial x} \end{aligned} \tag{11.6}$$

where (7.15) and the Einstein relation[50] have been used, and D_0 is the diffusivity in the absence of pairing.

P/N_D in (11.6) can be evaluated using (9.5) so that the coefficient preceding $(\partial N_D/\partial x)$ contains N_D as the only variable.

In Appendix B it is shown that ion pairing itself leads to severe departures from Fick's law.[49] In fact the diffusion flux density in the presence of pairing is given by

$$-\frac{D_0}{2}\left(1+\frac{\frac{1}{2}\left(N_D-N_A+\frac{1}{\Omega}\right)}{\sqrt{\frac{1}{4}\left(N_D-N_A-\frac{1}{\Omega}\right)+\frac{N_D}{\Omega}}}\right)\frac{\partial N_D}{\partial x} \tag{11.7}$$

Here again the diffusivity is specified by the factors preceding $(\partial N_D/\partial x)$ and, though variable, depends only on N_D, the local concentration of diffusant. Adding the two coefficients appearing in (11.6) and (11.7) the value of the diffusivity, D, in the presence of both pairing and diffusion potential is obtained. Thus

$$D=\frac{D_0}{2}\left(1+\frac{\frac{1}{2}\left(N_D-N_A+\frac{1}{\Omega}\right)}{\sqrt{\frac{1}{4}\left(N_D-N_A-\frac{1}{\Omega}\right)^2+\frac{N_D}{\Omega}}}+2\left(1-\frac{P}{N_D}\right)\left(\frac{N_D}{N_A-N_D}\right)\right) \tag{11.8}$$

It is obvious from (11.8) that even in the absence of space charge D is an extremely complicated function of N_D, and will be much more complex if space charge needs to be considered. When $N_D \ll N_A$ (11.8) reduces to

$$D=D_0\left(\frac{1}{1+\Omega N_A}\left(1+\frac{N_D}{N_A}\right)\right)\approx\frac{D_0}{1+\Omega N_A} \tag{11.9}$$

Comparison with equation (B15) shows that when (11.8) is true (i.e., in the absence of space charge) the diffusion potential may be ignored for $N_D \ll N_A$. Comparison of (B14) with (B15) shows how much D can vary with N_D when ion pairing occurs.

The proper study of diffusion in the presence of ion pairing should be augmented by a mathematical analysis, accounting for the concentration dependent diffusivity. Since this dependence is complicated the resulting boundary value problem must be solved numerically, and this

represents a formidable task. Although work along these lines is being done we shall content ourselves, in this article, with a less quantitative approach. The following plan has been followed.

A rectangular wafer of semiconductor uniformly doped with acceptor to the level, N_A, is uniformly saturated with lithium to a level, N_D, slightly less than N_A. Thus, the resulting specimen is well compensated but not converted to n-type. Lithium is then allowed to diffuse out of the specimen, and because of the thinness of the wafer, this process may be regarded as plane-parallel diffusion normal to its large surfaces. Low resistivity p-type layers therefore develop near the surfaces. If the thin ends of the wafer are put in contact with a source of current, current will flow parallel to its axis, so that the equipotential surfaces will be planes normal to this axis. The flow of current will be one dimensional because the inhomogeneity in lithium distribution occurs in the direction normal to its flow (see Fig. 21).

If two probe points are placed at a fixed distance apart on the broad surface of the wafer (see Fig. 21), then the conductance measured between them is a reflection of the total number of carriers in the low resistivity layers, i.e., a measure of the total amount of lithium which has diffused out. A more detailed connection between this conductance and diffusivity is derived in Appendix E. For the moment, however, attention will be confined to the description of the general plan of experiment.

According to the formulas derived in the early parts of this section, and also to (B14) and (B15), the diffusivity is something like $D_0/2$ in the

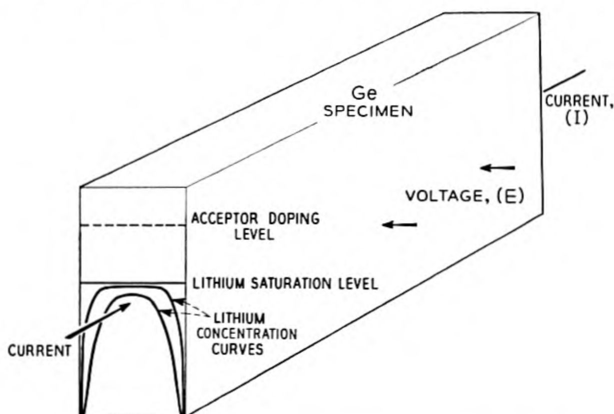


Fig. 21 — Diagram illustrating measurement of dependence of diffusivity on ion pairing (see Section XI).

bulk of the wafer where N_D almost equals N_A, but is as low as $D_0/(1 + \Omega N_A)$ near the surface where $N_D \ll N_A$. If ΩN_A is very much larger than unity as it will be under conditions where appreciable pairing occurs, the diffusivity will, therefore, be much smaller near the surface than at the high end of the diffusion curve, deeper within the specimen. The surface will then offer resistance to diffusion, and it may be expected that the measured value of the diffusivity will correspond more closely to the slow process near the surface rather than to the faster process occurring deeper in the semiconductor. Of course this cannot be entirely true because the resistance at the surface coupled with the lack of resistance inside the wafer will tend to steepen the concentration gradient near the surface. This will give the impression of a diffusivity somewhat higher than the one corresponding to the surface.

If the current flowing in the wafer under the conditions of measurement is I, and the potential measured between the points is V, then the conductance between the points is

$$\Sigma = I/V \tag{11.10}$$

In Appendix E it is shown (under the assumption that D is constant) that

$$\Sigma/\Sigma_0 = 1 + \frac{2.256\vartheta\sqrt{D}}{d}\left(\frac{\Sigma_\infty N_D^\circ}{\Sigma_0 N_A}\right)\sqrt{t} \tag{11.11}$$

where Σ_0 is the conductance after the specimen is saturated with lithium, but before any lithium has diffused out, and Σ_∞ is the conductance before lithium has been added. N_A is the uniform concentration of acceptor, and N_D° is the initial uniform concentration of lithium, while d is the thickness of the wafer. ϑ is a correction factor which arises because the mobility of holes varies from point to point in the wafer, as the density of lithium varies. There are two extreme types of variation.

The first takes place in a specimen in which, at room temperature (where the conductance measurement is made) ion pairing is complete. Then the local density of impurity scatterers[54] will be $N_A - N_D$. At the other extreme no ion pairing occurs, and the density of scatterers is $N_A + N_D$.

The nature of ϑ depends on how much pairing is involved. In Fig. 22 ϑ has been evaluated in its dependence on N_D° for the extreme cases mentioned. Furthermore it has been assumed then that N_D is given by a Fick's law solution of the diffusion problem, and that diffusion begins in a nearly compensated specimen.

The first thing to notice is that ϑ is not very different from unity in

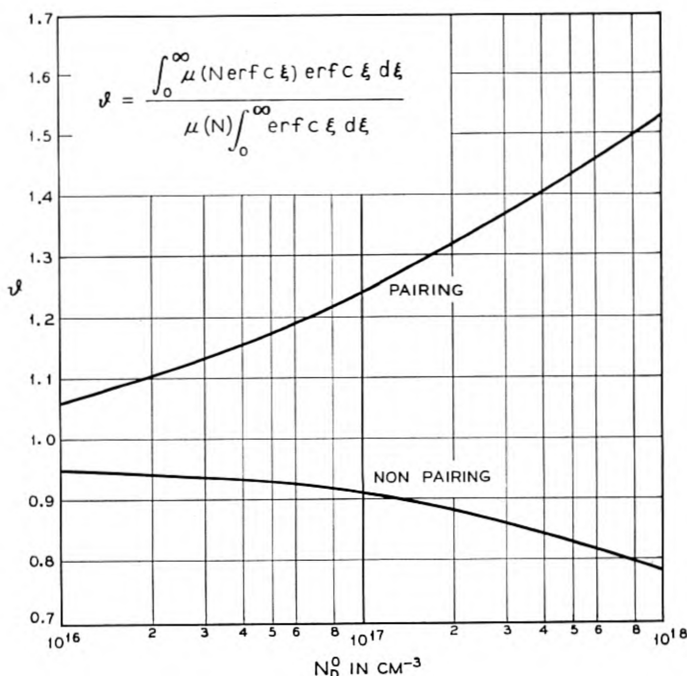


Fig. 22 — Plots of correction factor ϑ, required to compensate for the dependence of hole mobility on the density of scattering centers along a diffusion curve. ϑ is plotted against the initial density of donor and is shown for the two extreme cases of pairing and no pairing.

either extreme, and therefore closer to unity in some intermediate situation. In any event the correct value of ϑ can be read from Fig. 22 if the experiments involve either extreme at the measurement temperature. This has, in fact, been approximately the case in our experiments, in which pairing is almost complete at the temperature where conductances have been measured.

According to (11.11) a plot of Σ/Σ_0 against $\sqrt{t}$ should be a straight line of slope

$$S = \frac{2.256\vartheta\sqrt{D}}{d}\left(\frac{\Sigma_\infty N_D{}^\circ}{\Sigma_0 N_A}\right) \tag{11.12}$$

Measurement of S therefore affords a measure of D. Of course the apparent D obtained in this manner can never represent anything beyond some average quantity having the general significance of a diffusivity. This follows from the previous discussion concerning the non-constancy

of D. The only exception to this statement occurs in connection with high temperature experiments (above 200°C.) where both pairing and the diffusion potential are of little consequence. The mere fact that Σ/Σ_0 plots as a straight line against $\sqrt{t}$ is not evidence for the constancy of D. In Appendix E it is shown that a straight line will result, even when ion pairing is important, provided that the diffusion potential is based on the no-space-charge condition, i.e. provided that D varies only through its dependence on N_D.

On the other hand, the last statement implies that the existence of a straight line relationship is evidence that the diffusion potential has at least been minimized.

The most careful experiments were performed in germanium doped to various levels with gallium, indium, and zinc as acceptors. The germanium specimens were cut in the form of rectangular wafers of approximate dimensions (1.25 cm × 0.40 cm × 0.15 cm). Fresh lithium filings, were evenly and densely spread on one surface of the wafer, and alloyed to the germanium by heating for 30 seconds at 530°C in an atmosphere of dry flowing helium. Then the other surface was subjected to similar treatment.

After this the specimen was sealed in an evacuated pyrex tube and heated at a predetermined temperature for a predetermined period of time. The temperature was chosen, according to Fig. 5, so that the saturated specimen would still be p-type and just barely short of being fully compensated. Also attention was paid to the problem of avoiding precipitation on cooling. The time of saturation was determined from an extrapolation of the known lithium diffusion data, in germanium, of Fuller and Severiens[52] which is plotted in Figure 19 for the range extending from about 0° to 300°C.

After saturation the sealed tube was dropped into water and cooled. It was opened and the wafer ground on both sides, first with No. 600 Aloxite paper, and then with M 303½ American Optical corundum abrasive paper. The final thicknesses of the specimens ranged from 0.025 to 0.075 cm, the thinnest samples being used for the runs at the lowest temperature.

If the specimen is quite thin and highly compensated it is possible in principle to measure very small diffusivities (as low as 10^{-14} cm^2/sec) within a period of several hours. This is so because the low resistivity layer formed near the surface, although thin, will carry a finite share of the current in thin compensated specimens. On the other hand, additional difficulties arise. Diffusion layers as small as 100Å may be involved. If the surface is microscopically rough, diffusion will not be plane-parallel

and the measured diffusivity will appear larger than the real diffusivity. This condition can be partially corrected by etching the surface chemically until it is fairly smooth.

When dealing with such thin layers, the no-space-charge assumption becomes invalid and the diffusion potential ought really to be considered. Considering all the difficulties, i.e., concentration dependence of diffusion coefficient, possible existence of space charge, and roughness of surface, it is apparent that only qualitative effects are to be looked for in the diffusivities which have been measured.

The most that can be predicted is that for specimens containing a given amount of acceptor, the measured D (some average quantity) should be less than D_0, the disparity increasing with decreasing temperature. At high temperatures D should converge on D_0. Furthermore, at a given temperature D should decrease with an increase in concentration of acceptor. These tendencies are in line with the idea that reduction of temperature or increase of doping leads to an increase in pairing.

Runs were carried out on specimens etched with Superoxol[55] at the temperatures 25°, 100°, and 200°C. In the 25°C run the wafer was allowed to remain in the measuring apparatus under the two probe points in air, and Σ was measured from time to time. At 100°C the specimen was immersed in glycerine containing a few drops of HCl, the temperature of the bath being controlled. Periodic removal from the bath facilitated the measurement of Σ. At 200°C glycerine was again used as a sink for lithium, the sample being removed periodically for measurement.

Fig. 23 illustrates some typical plots of Σ/Σ_0 versus $\sqrt{t}$. They are all satisfactorily straight. Fig. 24 shows a plot of log D_0 against $1/T$, extrapolated from the data of Fuller and Severiens.[52] In this illustration, values of log D (obtained from the above measurements by determining the slopes S and employing (11.12)) are also plotted at the temperatures of diffusion. For ϑ the case of complete pairing was assumed.

The first thing to note is that the points for log D all lie below log D_0 except at 200°C and satisfy the qualitative requirement outlined above.* Moreover they drop further below log D_0 as the temperature is reduced, while at 200°C they have almost converged on log D_0.

The results for zinc are particularly interesting. Zinc is supposed to have a double negative charge in germanium.[64] Hence we would expect very intense pairing to occur. This is indicated in the diffusion data where the sample containing zinc at the rather low level, $N_A = 2.7$

---

* The long range nature of the interaction forces becomes evident when one considers that the diffusivities are being altered by impurity (acceptor) concentrations of the order of 1 part per million.

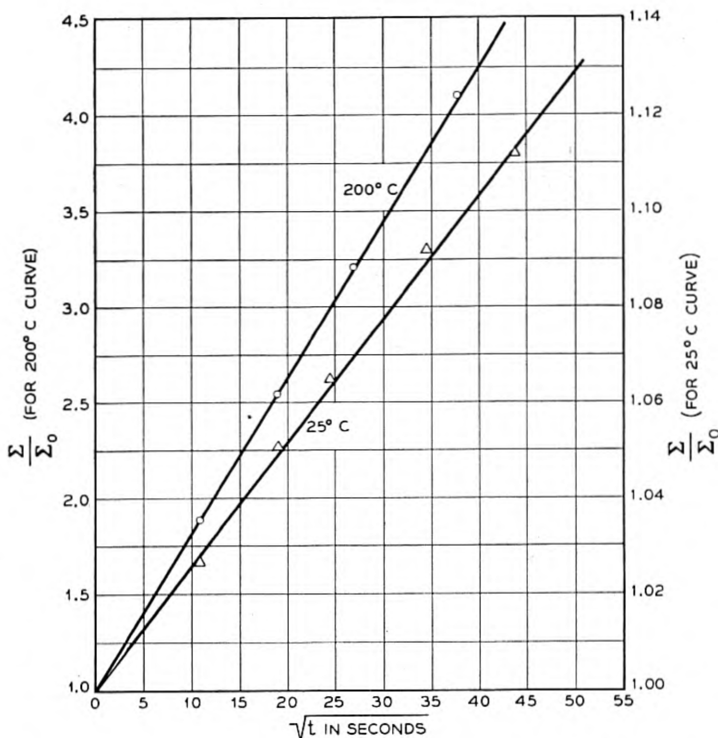


Fig. 23 — Curves illustrating the observed linear dependence of Σ/Σ_0 on the $\sqrt{t}$.

$\times\ 10^{16}\ \text{cm}^{-3}$, shows a large reduction in diffusivity even at temperatures as high as 200°C.

The difficulties discussed in this section serve to emphasize the importance of a direct transport experiment in which lithium atoms *uniformly* distributed throughout germanium or silicon, uniformly doped with acceptor, are caused to migrate by an electric field, and their mobilities measured. Because of the uniform dispersion of solutes the mobility will be constant everywhere. Furthermore no diffusion potential will be involved, and also the refined formula (7.25) can be applied. There are, however, many difficulties associated with the performance of this type of measurement.

In closing it may be mentioned that a few much less careful experiments of the kind described here have been performed in boron-doped silicon. The results indicate ion pairing in a qualitative way but more definite experiments are needed.

XII. INVESTIGATION OF ION PAIRING BY ITS EFFECT ON CARRIER MOBILITY

In Section VIII attention was called to the fact that ion pairing should influence the mobility of holes, because each pair formed, reduces the number of charged impurities by two. Thus, a specimen previously doped with acceptor, might, if sufficient lithium is added, exhibit an increase in hole mobility, even though the addition of lithium implies the addition of more impurities. This effect has been observed in connection with the Hall mobility of holes in germanium.

Two specimens of germanium were cut from adjacent positions in a single crystal doped with gallium to the level 3×10^{17} cm^{-3}. One of these was saturated with lithium through application of the same procedure

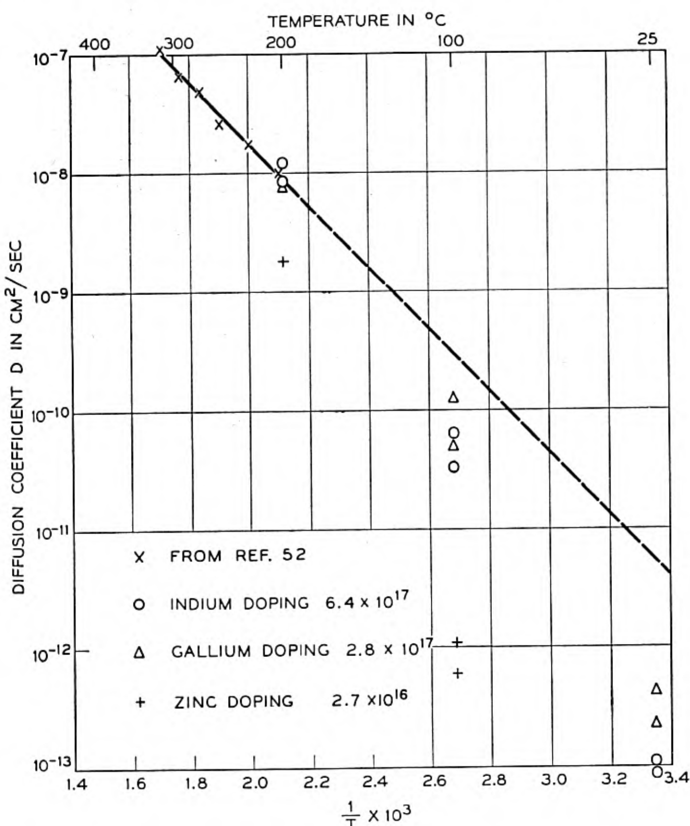


Fig. 24 — Plot of diffusivity of lithium in undoped germanium as a function of temperature — also showing points for apparent diffusivities of lithium in variously doped specimens.

employed in section V. Hall mobilities of the two specimens were measured[65] down to below 10°K. Cooling was carried out slowly to permit as much relaxation into the paired state as possible (see Section X). In Fig. 25 plots of the Hall mobilities versus temperature of both specimens are presented. Curve A is for the sample containing 2.8×10^{17} cm^{-3} lithium. It therefore contained about 5.8×10^{17} cm^{-3} total impurities as compared to the control sample whose curve is shown as B in Figure 25 and which contained only 3×10^{17} cm^{-3} impurities.

The lithium doped bridge exhibits by far the higher Hall mobility for holes (except at very low temperatures where poorly understood phenomena occur). In fact at 40°K the sample containing lithium shows a hole mobility 16 times greater than that of the control at the corresponding temperature. Rough analysis of the relative mobilities at T = 100°K indicate $\sim 2 \times 10^{17}$ cm^{-3} scattering centers in the control sample and 5×10^{15} cm^{-3} scattering centers in the sample containing pairs.

This experiment has been repeated with other specimens doped to different levels with gallium and even with other acceptors, and leaves no doubt that a mechanism which is most reasonably assumed to be pairing, is removing charged impurities from the crystal.

The phenomenon we have just described suggests an excellent method for testing the ion pairing formula derived in Sections VII and XI, for it

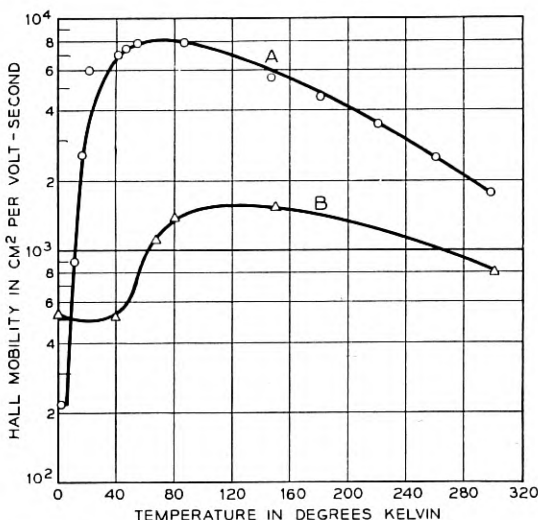


Fig. 25 — Plot of Hall mobility as a function of temperature for germanium containing 3×10^{17} cm^{-3} gallium. Curve A is for a sample containing 2.8×10^{17} cm^{-3} lithium.

enables us to determine at what temperature, at given values of N_A and N_D, P/N_D is exactly 0.5. Thus consider the fact that, all other things being equal, the control bridge and the one containing added lithium will exhibit equal Hall mobilities at a given temperature when the concentrations of charged impurities are identical in both of them. Now the concentration of such impurities in the control is simply N_A. The concentration in the bridge containing lithium is

$$N_A + N_D - 2P \tag{12.1}$$

The quantity $2P$ is removed from $N_A + N_D$, because each time a pair forms two charged scatterers are eliminated. The condition that the scattering densities in both bridges be equal is then simply

$$N_A = N_A + N_D - 2P$$

or

$$\frac{P}{N_D} = 0.5 \tag{12.2}$$

Therefore if plots of Hall mobilities versus temperature such as those appearing in Figure 25 are continued until they cross, the temperature of crossing marks the point at which P/N_D is 0.5.

In Fig. 26 typical crossings of this kind are shown. They are for two different gallium doped germanium crystals, one containing 3×10^{17} cm^{-3} gallium and the other 9×10^{15} cm^{-3}. The curves for the controls and lithium saturated samples in each case are shown as plots of the logarithm of Hall mobility against logarithm of absolute temperature. The lines plotted in this manner are straight. The lithium content of the bridge containing 9×10^{15} cm^{-3} gallium was 6.1×10^{15} cm^{-3} while that in the bridge with 3×10^{17} cm^{-3} gallium was 2.8×10^{17} cm^{-3}. All of these concentrations were obtained from Hall coefficient measurements in the controls and the lithium doped specimens.

As the temperature is increased the mobilities of the samples with lithium are reduced and approach the mobilities of the controls. This happens because pairs dissociate and more charged impurities appear. Finally when P/N_D is exactly 0.5 the curves cross. In Fig. 27 we notice that mobility measurements were not performed right up to the cross point, but that the straight lines have been extrapolated. This procedure was adopted of necessity, because of the high diffusivity of lithium. Thus, reference to Fig. 5 shows that the solubility in doped germanium decreases to a minimum as the temperature is raised from room temperature, and there is danger of precipitation. For this reason the measure-

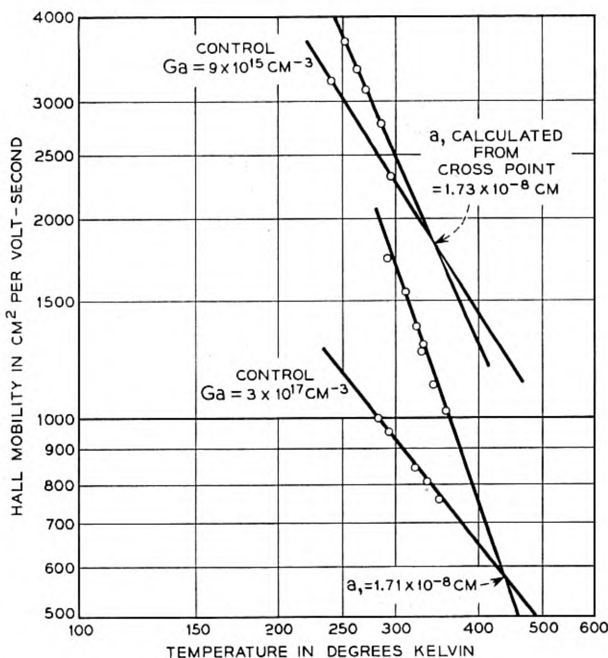


Fig. 26 — Illustration of cross over phenomenon for germanium samples containing gallium. Sample 314 contains 9×10^{15} cm^{-3} gallium and sample 302 contains 3×10^{17} cm^{-3}. Samples 316 and 301 are the corresponding samples to which lithium has been added.

ments were not carried to high temperatures.* In addition the value of the Hall coefficient was carefully checked at each temperature to see if it had changed. Since the reciprocal of the Hall coefficient[66] measures the carrier density any reduction in its value would have implied loss of compensation, or precipitation of lithium.

Over the measured points no appreciable variation of Hall coefficient was noted. Fortunately, the pairing relaxation time is quite small (less than a second) at the high temperatures involved so that it wasn't necessary to hold the samples at these temperatures for long periods in order to achieve pairing equilibrium. The times involved were too short for the occurrence of phase equilibrium characterized by precipitation.

The above discussion points up some of the care that must be taken to obtain reliable measurements. Another factor which enters the picture is the possible existence of a precipitate in the lithium doped bridge.

* In boron-doped germanium the cross-over was actually observed — no extrapolation having been necessary, because the temperature of intersection was sufficiently low.

During the course of our experiments it was discovered that precipitates have a profound effect on carrier mobility, reducing it so severely, that the mobility of the lithium doped bridge may never even rise above that of the control. Great care must be exercised in the preparation of suitable bridges to avoid the presence of precipitated lithium. Thus it may be necessary to saturate the bridge at a very low temperature (see Section IV, Figure 5) so that it is somewhat undersaturated at room temperature. This means that diffusion periods of weeks may be involved.

In Fig. 26 the sample with $N_A = 9 \times 10^{15}$ cm^{-3}, and $N_D = 6.1 \times 10^{15}$ cm^{-3} has $P/N_D = 0.5$ at 348°K, while the sample with $N_A = 3 \times 10^{17}$ cm^{-3} and $N_D = 2.8 \times 10^{17}$ cm^{-3} is half-paired at 440°K. This is to be expected, the more heavily doped specimen remaining paired up to higher temperatures. Using (9.6) and (9.3) it is possible to calculate a, the distance of closest approach of a gallium and lithium ion, from each of the measured cross points.

Thus in (9.6) we set $\theta = 0.5$, and N_A, N_D and T to correspond to each of the cases described. Having $\log_{10} Q(\alpha)$, α can be determined by interpolation in Table III and a then determined from (9.3). Of course κ is taken to be 16. Carrying through this procedure in connection with Fig. 26 leads to the satisfying result that $a = 1.71 \times 10^{-8}$ cm for the heavily doped sample and 1.73×10^{-8} cm for the lightly doped one. The values of Ω appearing in Table IV based on $a = 1.7 \times 10^{-8}$ cm therefore correspond to gallium.

Not only is this result satisfying because the two a's agree so well even though the samples involved were so different in constitution, but also because it is expected on the basis of the addition of known particle radii. Thus according to Pauling[36] the tetrahedral covalent radius of gallium is 1.26×10^{-8} cm while the ionic radius of lithium is 0.6×10^{-8} cm. Since gallium is presumably substitutional in a tetrahedral lattice we use its tetrahedral covalent radius, and since lithium is probably interstitial we use the ionic radius. The sum of the two is 1.86×10^{-8} cm which compares very favorably with the values of a quoted above.

This result constitutes good evidence that lithium is interstitial, for if it were somehow substitutional we might expect a to be something like a germanium-germanium bond length which is 2.46×10^{-8} cm. Such a value of a would lead to profoundly different crossing temperatures (of the order of 100° lower) so that it is not very likely.

One further point needs mention. This is the fact that as the two ions approach very closely, the concept of the uniform macroscopic dielectric constant, κ, loses its meaning. In fact, the binding energy should be increased (as though κ were reduced). Crude estimates of the magnitude

of this effect based on a dielectric cavity model show it to be of the order of some 10 or 15 percent of the energy computed on the assumption of the dielectric continuum, the increased binding energy showing up as a reduced value of a. This may account for the fact that the observed a, at 1.7×10^{-8} cm is less than the theoretical value, 1.86×10^{-8} cm.

The above example shows the ion pairing phenomenon in action as a structural tool, useful in investigating isolated impurities. In particular the demonstration that lithium is interstitial is interesting. The values of a have much more meaning as independent parameters in solids than they have in liquids, where a given ion may be surrounded by a sheath of solvating solvent molecules. Under the latter conditions the value of a can only be determined through application of the ion pairing theory itself.

Of course, certain unusual situations arise in solids also, and values of a (determined from ion pairing) are valuable indications of structural peculiarities.

Similar experiments have been performed on specimens doped with indium and boron. The results of all our investigations on the cross-over phenomenon are tabulated in Table V. In the table the first column lists the acceptor involved, and the second and third the appropriate concentrations of impurities. The fourth column contains the cross-over temperature, while the fifth, the measured value of a determined from it. The last column lists the values of a to be expected on the basis of the addition of tetrahedral covalent radii to the ionic radius of lithium — all of which appear in Pauling.[36]

The reliability of the measurements are in the order gallium, aluminum, boron, and indium. The principal reason for this is that the indium crystal was not grown specially for this work and was somewhat nonuniform. Of the two values obtained for a we tend to place more confi-

TABLE V

Acceptor	Acceptor conc. (cm^{-3})	Lithium conc. (cm^{-3})	Cross-over Temp. (°C.)	Measured a (cm)	Pauling a (cm)
B	7.0×10^{16}	5.9×10^{16}	338	2.05×10^{-8}	1.48×10^{-8}
B	7.0×10^{16}	5.54×10^{16}	320	2.27×10^{-8}	1.48×10^{-8}
B	7.0×10^{16}	5.85×10^{16}	330	2.16×10^{-8}	1.48×10^{-8}
Al	9.5×10^{15}	9.0×10^{15}	350	1.68×10^{-8}	1.86×10^{-8}
Ga	3.0×10^{17}	2.8×10^{17}	440	1.71×10^{-8}	1.86×10^{-8}
Ga	9.0×10^{15}	6.1×10^{15}	348	1.73×10^{-8}	1.86×10^{-8}
In	3.3×10^{17}	1.9×10^{17}	476	1.61×10^{-8}	2.04×10^{-8}
In	3.3×10^{17}	2.68×10^{17}	426	1.83×10^{-8}	2.04×10^{-8}

dence in 1.83×10^{-8} than in 1.61×10^{-8} cm. More work is necessary, however, before a real decision can be made.

A feature of Table V is the fact that gallium, aluminum, and indium exhibit orthodox behavior, i.e., the measured a's are in both cases slightly less than those expected on the basis of the addition of radii. The internal consistency of the theory gains support from the fact that gallium and aluminum behave similarly as the Pauling a's tabulated in Table V predict. In fact if 1.83×10^{-8} cm is taken as the more reliable indium value the three cases fail to match the Pauling radii by about the same amount, a result which implies that the disparity is due to the same cause, i.e., failure of the dielectric continuum concept.

Another feature of Table V is the fact that boron is out of line to the extent that the measured *a exceeds* the Pauling a by 50 per cent. A possible explanation is the following. The tetrahedral radii of boron and germanium are poorly matched (0.88 Å and 1.26 Å, respectively). The strain in the boron-germanium bond may appear as a distortion of the germanium atom in such a way as to increase the effective size of the boron ion. This strain was mentioned before in Section V where it was invoked to explain the stability of LiB^- complex in silicon.

XIII. RELAXATION STUDIES

The relaxation time discussed in Section X has been studied experimentally. The following procedure was used. A specimen was warmed to 350°K where a considerable amount of pair dissociation occurred, and then cooled quickly by plunging into liquid nitrogen. It was then rapidly transferred to a constant temperature bath, held at a temperature where pair formation took place at a reasonable rate, and the change in sample conductivity (as pairing took place) was measured as a function of time.

The principle upon which this measurement is based is the following. At a given temperature the occurrence of pairing does not change the carrier concentration, only the carrier mobility. As a result the measurement of conductivity is effectively a measurement of relative mobility. During relaxation the densities of charged impurities are changed, at the most, by amounts of the order of 50 per cent. Over this range, the mobility may be considered a linear function of scatterer density. The dependence of conductivity on time, as pairing takes place, must be of the form

$$\sigma = \sigma_\infty - \Phi\, e^{-t/\tau} \tag{13.1}$$

where σ_∞ is the conductivity when $t = \infty$, and τ is the relaxation time defined in section X while Φ is some unknown constant, depending among

other things on the initial state of the system. Equation (13.1) is based on the assumption that the number of charged scatterers decays as a first order process, and that σ is a linear function of this number, relative to the exponential dependence on time.

The first order character of pairing is fortunate for it renders the measurement of τ independent of a knowledge of Φ, i.e. independent of the initial state of the system. This is not only fortunate from the point of view of calculation but from experiment, since it is almost impossible to prepare a specimen in a well defined initial state.

The unimportance of Φ is best seen by plotting the logarithm of $\sigma_\infty - \sigma$ against time. According to (13.1) this plot is specified by

$$\log (\sigma_\infty - \sigma) = \log \Phi + \frac{t}{\tau} \tag{13.1}$$

Thus the reciprocal of its slope measures τ, and Φ is not involved. Fig. 27 illustrates the data for a typical run plotted in this manner. The sample is one containing about 9×10^{15} cm^{-3} gallium and the experiment was performed at 195°K (dry ice temperature). Notice that the curve is absolutely straight out to 3500 minutes, demonstrating beyond a doubt that the process is first order. The relaxation time computed from its slope is 1.51×10^5 seconds as against a value calculated by the methods

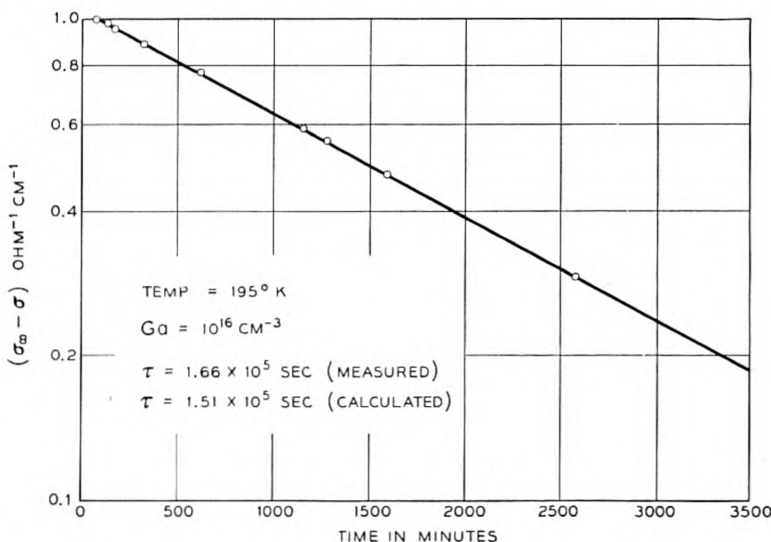


Fig. 27 — Plot of log $(\sigma_\infty - \sigma)$ as a function of time showing first order kinetics of pairing.

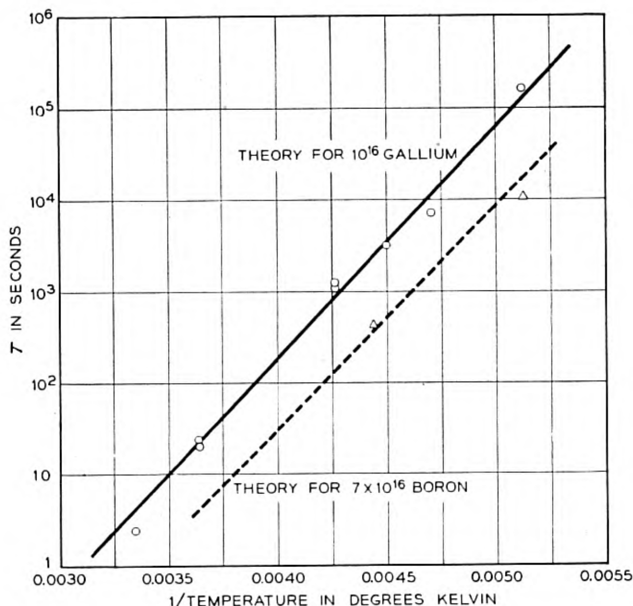


Fig. 28 — Plots of logarithm of relaxation time versus reciprocal temperature showing agreement between theory and experiment.

of section X of 1.66×10^5 seconds. The result is in good agreement with theory.

Studies of the kind illustrated in Fig. 27 have been carried out in samples doped to various levels and also at various temperatures. Boron and indium have been used as doping agents, as well as gallium. Relaxation times have been measured over the range extending from about a second to hundreds of thousands of seconds. In each case straight line plots were obtained and the agreement between calculated and measured τ's has been as good as in the example illustrated by Fig. 28. Relaxation connected with dissociation has also been measured with equally satisfactory results.

Some of these data are shown in Fig. 29 where log τ is plotted as a function of reciprocal temperature for gallium and boron at two different values of doping. The drawn curves are theoretical obtained from Fig. 20 while the points shown are experimental. It is seen that agreement is nearly perfect. The relaxation time, true to the demands of theory, does not seem to depend on the kind of acceptor used for doping, i.e., it is independent of a, the distance of closest approach.

The data in Fig. 28 actually can be used to measure the diffusivity of

lithium. As must be the case from the above mentioned agreement, the values of D_0 computed from them agree with the diffusion data of Fuller and Severiens[52] almost perfectly. This is a very quick and sensitive method (also probably exceedingly accurate) for determining diffusivities. For example the work already completed, in effect, represents the determination of diffusivities of the order of 10^{-16} cm^2/sec within a matter of an hour, and, no doubt, smaller diffusivities could be determined by doping more heavily with acceptor.

XIV. THE EFFECT OF ION PAIRING ON ENERGY LEVELS

It was predicted in Section VIII that ion pairing would drive the electronic energy states of donors and acceptors from the forbidden energy region. In this section it will be demonstrated by low temperature Hall effect measurements that the addition of lithium to gallium-doped germanium does indeed result in the removal of states from the forbidden gap rather than in the simple compensation which occurs when a non-mobile donor such as antimony is added.

At low temperatures where carrier concentration, p, is less than the donor concentration, it can be expressed in the form[67]

$$p = \frac{N_A - N_D}{N_D} \left(\frac{2\pi m_p kT}{h^2} \right)^{3/2} \exp\left[-E_A/kT\right] \tag{14.1}$$

where N_A and N_D are the concentrations of acceptor and donor states, respectively, m_p, the effective mass of free holes, h, Plank's constant, and E_A the ionization energy of the acceptor. The values of m_p and E_A are known for the group III acceptors.[68]

Lithium was added to a specimen of germanium known to contain 1.0×10^{16} cm^{-3} gallium atoms and a negligible amount of ordinary donors. Carrier concentrations for this specimen were determined from Hall coefficient measurements. The logarithm of this concentration is shown in Fig. 29 plotted against reciprocal temperature. The high temperature limit of this plot fixes $N_A - N_D$ at 1.15×10^{15} cm^{-3}.

At low temperatures the curve exhibits an extended linear portion to which (14.1) should apply. Evaluating (14.1) with $p = 4.0 \times 10^{13}$ cm^{-3} at $1/T = 0.06$ deg^{-1} and $N_A - N_D = 1.15 \times 10^{15}$ cm^{-8} we find that $N_D = 2.6 \times 10^{14}$ cm^{-3} and $N_A = 1.4 \times 10^{15}$ cm^{-3}.

Therefore, the density of apparent acceptor states has been decreased by $1.0 \times 10^{16} - 1.4 \times 10^{15} = 8.6 \times 10^{15}$ cm^{-3}. The added concentration of lithium was 1.0×10^{16} cm^{-3} $- 1.15 \times 10^{15}$ cm^{-3} $= 8.85 \times 10^{15}$ cm^{-3}, *almost identical with the loss in concentration of acceptor states.* This implies (as would be expected) that the lithium is almost totally paired.

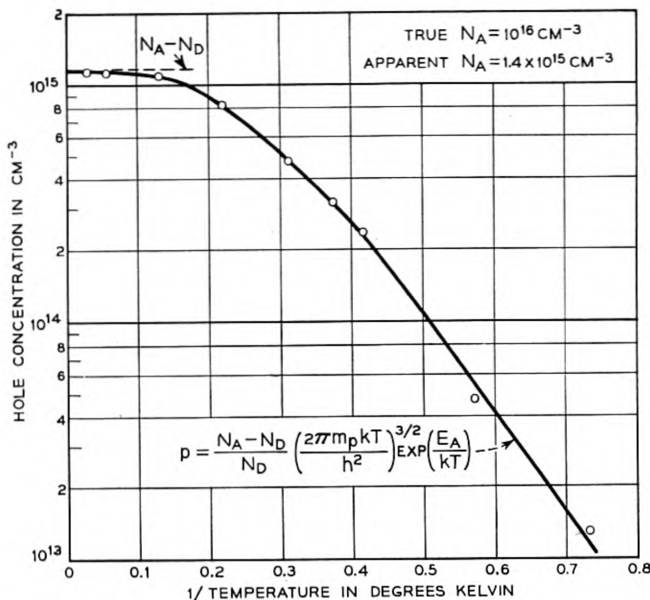


Fig. 29 — Plot of hole concentration as a function of reciprocal temperature for a sample containing ion pairs.

An even more striking result appears. From the above results the density of lithium atoms involved in pairs is 8.85×10^{15} cm^{-3} − 2.6×10^{14} cm^{-3} = 8.6×10^{15} cm^{-3}, *the same number by which the density of acceptors has been decreased!* There can be little question that ion pairing is the mechanism responsible for the removal of states.

In closing it is worth pointing out that the density of unpaired lithiums 2.6×10^{14} cm^{-3}, is certainly not characteristic of the low temperatures at which the above Hall measurements were performed. Obviously a density characteristic of some higher temperature has been quenched into the specimen. At the low temperature involved the unpaired density would be effectively zero.

XV. RESEARCH POSSIBILITIES

The fields described in the preceding text have been hardly touched, even by this long paper, and it does not seem fitting to close without some speculation concerning the possibilities of future work.

In the first place, there are other donors and acceptors besides lithium which are reasonably mobile in germanium or silicon, e.g. copper, iron,

zinc or gold. To some extent the methods of this paper can be applied to these. Furthermore, returning to lithium, there are impurities both mobile and immobile which introduce more than one energy state into the forbidden gap. The phase relations of lithium in the presence of these should be extremely interesting since the corresponding mass action equations are more complicated. Analogues of dibasic[69] acids and bases should exist.

In the case of ion pairing doubly charged acceptors like zinc in germanium[64] should be extremely interesting, since large amounts of pairing should persist up to very high temperatures. In fact such studies represent excellent means of testing for the existence of doubly charged ions. There is also the question of what happens to the two energy levels when an acceptor like zinc pairs with a single lithium ion. Are both levels driven from the forbidden gap or do they split under the perturbation?

Then there is the problem of ion *triplets* — a possibility with impurities like zinc — which is unexplored both theoretically and experimentally. Also, very strange diffusion effects must occur in the presence of doubly charged ions, to say nothing of the effect which uncompensated mobile holes might have on relaxation processes.

The field of ion pairing in silicon is relatively unexplored.

All of the phenomena discussed in this paper must occur in the group III–V compounds, more or less complicated by additional effects.

The question of the formation of the LiB^- complex in both germanium and silicon needs further study. It should behave as an acceptor and its electronic energy state might be revealed by suitable quenching techniques.

Non ionic reactions between group V donors and group III acceptors very likely occur, i.e., a real III–V covalent bond may be formed between such atoms dissolved in germanium or silicon at high temperatures. This possibility could be investigated by looking for changes in carrier mobility or impurity energy levels upon extended heating — in much the same way that ion pairing has been studied. If found, the phenomenon may provide an excellent technique for measuring the diffusitivities of all classes of impurities even at fairly low temperatures.

Such compounds may possess strange energy levels and be responsible for unexplained traps and recombination centers.

The effect of stress on the extent of ion pairing may well be profound since there will be a tendency for such stress to concentrate at imperfections. Stress studies on ion pairing may therefore be useful for further investigating the strain about an isolated impurity.

Ion pairing between lithium ions and acceptor centers located in dislocations or vacancies should occur. In the first case the dislocation

would be the analogue of the polyelectrolyte molecule in the aqueous solution.

An interesting question, in the diffusion of substitutional acceptors, concerns whether the ion or the neutral atom is responsible for diffusion. It is possible that the neutral atom, less securely bonded to the lattice, is the chief agent. This might be determined by changing the ratio of neutral atoms to ions by suitably doping with other donors or acceptors.

Doping apparently affects the concentration of vacancies which have acceptor properties and therefore the rate of diffusion.[70, 71]

Other interesting effects concerning the distribution of an impurity between two different kinds sites in the lattice[72] are also possible.

These and many other fascinating fields still require exploration. We hope to investigate some of them in the near future.

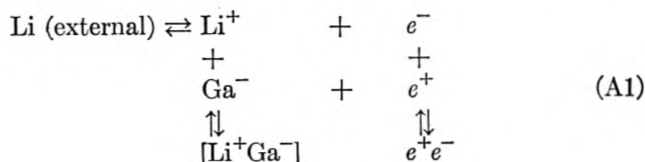
ACKNOWLEDGMENTS

The authors are greatly indebted to A. J. Pietruszkiewicz, Jr., for assistance in carrying out experimental work relating to solubility and diffusion and to J. P. Maita for help with experimental work on Hall effect and an ion-pair relaxation. Thanks are due N. B. Hannay for many helpful comments during the course of the work and during preparation of the manuscript. Thanks are also due Miss M. C. Gray for the evaluation of the integrals in Section VII and to F. G. Foster for the photograph of Fig. 8. Finally the authors would like to thank the editors of the Bell System Technical Journal for providing space so that all of the important features of our subject could be treated in one article.

Appendix A

THE EFFECT OF ION PAIRING ON SOLUBILITY

In Section VIII attention was called to the fact that ion pairing should have some effect on lithium solubility but that it would be difficult to achieve conditions under which the effect would be observable. Now, this point will be enlarged upon. Consider an equilibrium like (2.1) except imagine it to take place in germanium with gallium as the immobile acceptor. (This because germanium with gallium has been studied in ion pairing investigations.)



$$\begin{array}{llcl} \text{Li (external)} \rightleftarrows & Li^+ & + & e^- \\ & + & & + \\ & Ga^- & + & e^+ \\ & \upharpoonleft\downharpoonright & & \upharpoonleft\downharpoonright \\ & [Li^+Ga^-] & & e^+e^- \end{array} \qquad (A1)$$

where $[Li^+Ga^-]$ represents an ion pair, whose concentration we denote by P. N_A and N_D will be the total densities of acceptor and donor respectively and A^- and D^+ the densities of acceptor and donor ions in the unpaired state.

As in the main text, n and p will represent the concentrations of holes and electrons. The following relations are then to be expected on the basis of definition, mass action, and charge balance.

$$N_A = A^- + P \tag{A2}$$

$$N_D = D^+ + P \tag{A3}$$

$$D^+n = K^* \tag{A4}$$

$$np = n_i^2 \tag{A5}$$

$$\frac{P}{A^+D^-} = \Omega \tag{A6}$$

$$D^+ + p = A^- + n \tag{A7}$$

Equations (A4), (A5), and (A7) are just reproductions of (3.1), (3.2), (2.8), while (A6) is the same as (9.4). The problem is to express the solubility of lithium, N_D, as a function of N_A. Manipulation of the preceding set of equations gives this result as

$$N_D = \frac{(N_A - A^-)(1 + \Omega A^-)}{\Omega A^-} \tag{A8}$$

with A^- given by the solution of

$$\frac{N_A - A^-}{\Omega A^-} = \frac{A^-}{1 + \sqrt{1 + \left(\frac{2n_i}{D_{0+}}\right)^2}} + \sqrt{\left(\frac{A^-}{1 + \sqrt{1 + \left(\frac{2n_i}{D_0^+}\right)^2}}\right)^2 + (D_0^+)^2} \tag{A9}$$

where D_0^+ is defined by (3.3). Equation (A9) generally needs to be solved numerically for A^-.

To see what these relations predict in a special case consider the solubility of lithium in gallium-doped germanium at 300°K. At this temperature the values of n_i and D_0^+ and Ω are

$$\begin{aligned} n_i &= 2.8 \times 10^{13}\ \text{cm}^{-3} \\ D_0^+ &= 7 \times 10^{13}\ \text{cm}^{-3} \\ \Omega &= 1.61 \times 10^{15}\ \text{cm}^{-3}. \end{aligned} \tag{A10}$$

TABLE AI — TEMPERATURE = 300°K

N_A (cm^{-3})	N_D (cm^{-3})	N_D^* (cm^{-3})	$P = N_A - A^-$ (cm^{-3})
10^{14}	1.25×10^{14}	1.25×10^{14}	0.15×10^{14}
10^{15}	0.94×10^{15}	0.875×10^{15}	0.44×10^{15}
10^{16}	0.985×10^{16}	0.875×10^{16}	0.77×10^{16}
10^{17}	0.990×10^{17}	0.875×10^{17}	0.92×10^{17}
10^{18}	0.995×10^{18}	0.875×10^{18}	0.97×10^{18}

The value of n_i is taken from Figure 2, of D_0^+, from Figure 5, and of Ω, from Table IV. Using (A10) together with (A9) and (A8) leads to the results tabulated in Table AI. In this table, N_D^* represents the solubility for the case $\Omega = 0$, i.e., the solubility if there were no ion pairing. The main feature to be obtained from the Table is that N_D is not very much larger than N_D^*, no matter how large the value of N_A. This is true in spite of the fact that the last column which lists P shows that at $N_A = 10^{18}$ cm^{-3} P is about 98% of N_D so that pairing of the donor is virtually *complete*.

The result is not limited to the special conditions of doping and temperature chosen in compiling Table AI, but must be quite general. One can arrive at this conclusion in the following way.

By subtracting (A3) from (A2) we obtain

$$N_A - N_D = A^- - D^+. \tag{A11}$$

The quantities A^- and D^+ appear in equations (A4) and (A7), while n and p, appearing in (A4) and (A7) are related by (A5). These three equations are sufficient for the determination of D^+ in its dependence on A^-. That this is the case is immediately obvious when (A4), (A5), and (A7) are recognized as reproductions of (3.1), (3.2) and (2.8). In fact this means that the desired relationship between D^+ and A^- is nothing more than equation (3.4) which itself is predicated on (3.1), (3.2), and (2.8). But then it is known according to (3.6), that D^+ can at the most be slightly greater than A^-, although most likely less. This assumes of course that we deal with dopings sufficiently high so that (3.5) applies. On the other hand at low dopings (3.4) tells us that D^+ will be D_0^+. Therefore if we work with a system in which in the absence of pairing the electron-hole equilibrium has driven the value of N_D close to N_A (as it has in this system — see N_D^*) the introduction of pairing cannot drive it much higher, since according to (A11) if D^+ cannot get higher than A^-, N_D cannot exceed N_A. This is evident in Table AI where N_D comes very close to N_A but never exceeds it.

When N_A is very small so that D^+ equals D_0^+ and does exceed A^- by

a large amount, there can be no visible increment in solubility as a result of pairing because P can never exceed N_A which by definition is small.

The physical reason for these limitations is the following. Suppose N_D is driven close to N_A by the hole-electron equilibrium so that in terms of carriers (holes and electrons) the specimen is very closely compensated. Then if by the formation of pairs, additional donors are caused to enter the crystal, the electrons they donate cannot be absorbed by holes because very few of the latter are present. Thus the following two sketched equilibria will oppose each other

$$\begin{array}{lllll} \text{Li (external)} \rightleftarrows & \text{Li}^+ & + & e^- & \\ & + & & & \\ & \text{Ga}^- & & & \text{(A12)} \\ & \upharpoonleft\downharpoonright & & & \\ & [\text{Li}^+\text{Ga}^-] & & & \end{array}$$

the one involving electrons attempting to drive lithium out of solution because of the build-up of electron concentration, and the pairing equilibrium attempting to bring lithium into solution in order to form pairs. Thus the pairing process will not be as efficient a solubilizer as might be thought at first.

This point can be illustrated by considering a situation in which the germanium crystal not only contains gallium to the level, N_A but also an immobile donor, to the level $N = 0.99\, N_A$. Thus, the crystal is almost compensated before any lithium has been added. Nevertheless, there are still N_A gallium ions so that even though the hole-electron equilibrium, working on the differential, $0.01\, N_A$, cannot increase the solubility of lithium, the pairing process might. To investigate this situation equations (A2) to (A7) can be adopted with the simple change that $(A^- - N)$ replaces A^- in (A7).

Taking the situation covered by (A10) at 300°K, Table AII was compiled. Here again N_D^* is the solubility for $\Omega = 0$.

If only the hole-electron effect were operative, then we could not expect to drive N_D much beyond $N_A - N$. In the 10^{16} case $N_A - N$ is 10^{14} cm^{-3} and in the 10^{17} case it is 10^{15} cm^{-3}. The values of N_D^* in Table AII thus confirm this argument. Furthermore, N_D is in neither case much greater than N_D^* showing that despite the fact that there were, respec-

TABLE AII — TEMPERATURE 300°K

N_A (cm^{-3})	N (cm^{-3})	N_D (cm^{-3})	N_D^* (cm^{-3})	P (cm^{-3})
10^{16}	0.99×10^{16}	3.2×10^{14}	1.26×10^{14}	3×10^{14}
10^{17}	0.99×10^{17}	1.6×10^{16}	0.88×10^{15}	1.6×10^{15}

tively, 10^{16} and 10^{17} cm^{-3} gallium ions available for pairing, the pairing process did very little to increase the solubility.

If the constant Ω is exceedingly large as is probably the case for a multiply charged acceptor, it is possible that ion paring will have a measurable effect on solubility.

Appendix B

CONCENTRATION DEPENDENCE OF DIFFUSIVITY IN THE PRESENCE OF ION PAIRING

In Section VIII it was mentioned that the diffusivity of a mobile donor like lithium is concentration dependent when the donor participates in a pairing equilibrium with an immobile acceptor. In this appendix we propose to investigate the nature of the dependence.

Consider a semiconductor, uniformly doped to the level, N_A, with acceptor. Let the local density of mobile donor be $N_D(x)$, x being the position coordinate. If $P(x)$ is the local pair concentration, then the local density of free diffusible ions is $(N_D - P)$. The flux of these diffusing ions then depends upon the gradient (assuming Fick's law[49]) of $(N_D - P)$. Thus, if D_0 is the diffusivity of free donor, i.e. the diffusivity in the absence of pairing, then the flux density is

$$f = -D_0 \frac{\partial(N_D - P)}{\partial x} \tag{B1}$$

If we apply (9.4) to the present case we can write

$$\Omega = \frac{P}{(N_A - P)(N_D - P)} = \frac{-(N_D - P) + N_D}{[(N_A - N_D) + (N_D - P)](N_D - P)} \tag{B2}$$

from which it is possible to solve for $(N_D - P)$. Thus

$$N_D - P = \frac{1}{2}\left(N_D - N_A - \frac{1}{\Omega}\right) + \sqrt{\frac{1}{4}\left(N_D - N_A - \frac{1}{\Omega}\right)^2 + \frac{N_D}{\Omega}} \tag{B3}$$

Substitution of (B3) into (B1) yields

$$f = -\frac{D_0}{2}\left[1 + \frac{\frac{1}{2}\left(N_D - N_A + \frac{1}{\Omega}\right)}{\sqrt{\frac{1}{4}\left(N_D - N_A - \frac{1}{\Omega}\right)^2 + \frac{N_D}{\Omega}}}\right]\frac{\partial N_D}{\partial x} \tag{B4}$$

If ion pairing was not thought of, the flux density would have been written in terms of the gradient of the total concentration, N_D.

$$f = -D \frac{\partial N_D}{\partial x} \tag{B5}$$

where D is the diffusivity. Comparison of (B5) with (B4) leads to the relation

$$D = \frac{D_0}{2}\left[1 + \frac{\frac{1}{2}\left(N_D - N_A + \frac{1}{\Omega}\right)}{\sqrt{\frac{1}{4}\left(N_D - N_A - \frac{1}{\Omega}\right)^2 + \frac{N_D}{\Omega}}}\right] \tag{B6}$$

so that D depends on the local concentration, N_D, of diffusant.

It is interesting to explore the limiting forms of D when $N_D \ll N_A$ and when $N_D = N_A$. In the latter case (B6) reduces to

$$D = \frac{D_0}{2}\left[1 + \frac{\frac{1}{2\Omega}}{\sqrt{\frac{1}{4\Omega^2} + \frac{N_A}{\Omega}}}\right] \tag{B7}$$

while (B3) becomes

$$N_A - P + \frac{1}{2\Omega} = \sqrt{\frac{1}{4\Omega^2} + \frac{N_A}{\Omega}}. \tag{B8}$$

Substituting the left side of (B8) for the denominator involving the radical in (B7) leads to

$$D = \frac{D_0}{2}\left[1 + \frac{1}{2(N_A - P)\Omega + 1}\right] \tag{B9}$$

But according to (B2), when $N_A = N_D$,

$$(N_A - P)\Omega = \frac{P}{N_A - P} \tag{B10}$$

so that (B9) becomes

$$D = \frac{D_0}{2}\left[1 + \frac{1}{\frac{2P}{N_A - P} + 1}\right] \tag{B12}$$

Now in case the degree of pairing is high (which is, of course, the case we are interested in) P will be almost equal to N_A so that

$$\frac{2P}{N_A - P} \tag{B13}$$

will be a very large number. If this is so the second term in brackets in (B12) can be set equal to zero and we have

$$D = \frac{D_0}{2}. \tag{B14}$$

In the other extreme with $N_D \ll N_A$ (B6) becomes

$$D = \frac{D_0}{2}\left[1 + \frac{\frac{1}{2}\left(\frac{1}{\Omega} - N_A\right)}{\sqrt{\frac{1}{4}\left(\frac{1}{\Omega} + N_A\right)^2}}\right] = \frac{D_0}{1 + \Omega N_A} \tag{B15}$$

Since ΩN_A can exceed unity by a large amount it is evident that the relation in (B15) predicts a large reduction in diffusivity towards the front end of a diffusion curve where $N_D \ll N_A$, and (B14) a smaller reduction in D_0 where N_D may be close to N_A. That part of the medium near the front of the diffusion curve acts therefore like a region of high resistance, confining the diffusant to the back end where the resistance is low.

Appendix C

SOLUTION OF BOUNDARY VALUE PROBLEM FOR RELAXATION

In Section X equations (10.23), (10.21), (10.20), and (10.19) defined a boundary value problem which we reproduce here, except that (10.20) and (10.19) have been written more completely with the aid of (10.16). Thus

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial \rho}{\partial r} + R\rho\right) = \frac{1}{D_0}\frac{\partial \rho}{\partial t} \tag{C1}$$

$$\frac{\partial \rho}{\partial r} + \frac{R}{r^2}\rho = 0, \qquad r = L, \qquad r = a \tag{C2}$$

$$\rho = N^2, \qquad t = 0, \qquad a < r < L \tag{C3}$$

In principle this problem is soluble by separation of variables.[56] Thus we define

$$\rho(r, t) = G(r)\, S(t) \tag{C4}$$

which upon substitution into (C1), yields the two ordinary differential equations

$$\frac{d}{dr}\left[r^2\frac{dG}{dr} + RG\right] + \eta^2 G = 0 \tag{C5}$$

$$\frac{d\,\ell n\, S}{dt} + \eta^2 D_0 = 0 \tag{C6}$$

where η^2 is an arbitrary positive parameter.

The allowable values of η are determined by (C2) which can now be

replaced by

$$\frac{dG}{dr} + \frac{R}{r^2} G = 0, \qquad r = L, \qquad r = a \tag{C7}$$

Equation (C6) can be solved immediately to give

$$S_\eta(t) = e^{-\eta^2 D_0 t} \tag{C8}$$

and if we assign the subscript η to the G going with η the most general solution of (C1) and (C2) will be

$$\rho = \sum_\eta A_\eta G_\eta(r) e^{-\eta^2 D_0 t} \tag{C9}$$

where the A_η are arbitrary constants so determined that (C3) is satisfied.

Equation (C9) shows that in reality there exists, for this problem, a spectrum of relaxation times, $1/\eta^2 D_0$. After a brief transient period many of the higher order terms will decay away and eventually only the first two terms will have to be considered. Finally when equilibrium is attained only the first term will survive.

The last statement implies that $\eta = 0$, is an allowable eigenvalue, i.e., that the first term is independent of time. That this is so can be proved by solving (C5) for $\eta = 0$, and substituting the result in (C7). Thus

$$G_0(r) = \exp\left(\frac{R}{r}\right) \tag{C10}$$

and this does satisfy (C7). ρ can then be approximated after the transient by

$$\rho = A_0 \exp\left(\frac{R}{r}\right) + A_1\, G_1(r) e^{-\eta_1^2 D_0 t} \tag{C11}$$

from which it is obvious that the relaxation time dealt with in section X is

$$\tau = \frac{1}{\eta_1^2 D_0} \tag{C12}$$

In principle it should be possible to evaluate G_1 by the straightforward solution of (C5) and determination of the second eigenvalue through substitution of this solution in (C7). In fact this represents a rather unpleasant task since G is a confluent hypergeometric function.[57] Therefore we shall follow an alternative route based on the assumption that by the time (C11) applies the flux $4\pi r^2 J^*(r)$, where J^* is given by (10.16), is almost independent of r. The reader is referred to some related papers[58, 59]

for the justification of this view. Briefly it is permissible, after a short transient period, in spherical diffusion, whenever the dimensions of the diffusion field are large compared to the dimension of the sink. This results from the fact that in spherical diffusion from an infinite field[48] a real steady state is reached after a brief transient period. In contrast, in plane-parallel diffusion to a sink from an infinite field,[60] a steady state is never reached.

Substituting (C11) into (10.16) then yields

$$J^* = -D_0 A_1 e^{-\eta_1^2 D_0 t} \left(\frac{dG_1}{dr} + \frac{R}{r^2} G_1 \right) \tag{C13}$$

Multiplying J^* by $4\pi r^2$ and demanding that the product be independent of r, leads to the relation

$$r^2 \frac{dG_1}{dr} + RG_1 = \delta \tag{C14}$$

where δ is constant. The solution of (C14) is

$$G_1 = \exp\left(\frac{R}{r}\right) + \frac{\delta}{R} \tag{C15}$$

This is a sufficient approximation for G_1.

The constants η_1, A_0, A_1, and δ must now be determined. To accomplish this we note that (C2) which specifies that the boundaries at $r = a$ and $r = L$, are impermeable is equivalent to the condition that ions be conserved with the interval (a, L), or that

$$4\pi \int_a^L r^2 \rho \, dr = N \tag{C16}$$

After infinite time ρ is specified by the first term of (C11) and when this is inserted into (C16) the result is

$$A_0 = NM \tag{C17}$$

where M is defined by (10.26).

Substitution of (C17) and (C15) into (C11) gives

$$\rho = NM \exp(R/r) + \left(A_1 \exp(R/r) + \frac{A_1 \delta}{R} \right) e^{-\eta_1^2 D_0 t} \tag{C18}$$

Now (C3) applied to (C18) demands

$$NM + A_1 = 0 \tag{C19}$$

$$\frac{A_1 \delta}{R} = N^2 \tag{C20}$$

Of course this presumes that the approximation contained in (C18) is valid down to very small values of time. This assumption is well founded as the transient does vanish after a rather short time.

Inserting (C19) and (C20) in (C18) then gives us

$$\rho = NM \exp (R/r) + N[N - M \exp (R/r)]e^{-\eta_1^2 D_0 t} \tag{C21}$$

in which only η_1 remains to be determined.

Substitution of (C21) into (C16), recalling the definitions of M and L, shows that it already satisfies (C16) for any time, t. Thus (C16) cannot be used for determining η_1.

On the other hand we note from (C21) that as soon as r becomes of order, R, ρ becomes almost independent of r, being given

$$\rho = N\{N + (N - M)e^{-\eta_1^2 D_0 t}\} \tag{C22}$$

Since L is of the order 10R or greater, this means that throughout most of the volume, $1/N$ (in fact throughout 0.999 $1/N$) ρ is independent of r. Effectively, the entire volume $1/N$ has been drained of ions, i.e., they have been trapped. The total ion content at time t, may then be taken as the product of ρ, given by (C22), with $1/N$, that is,

$$N + (N - M)e^{-\eta_1^2 D_0 t} \tag{C23}$$

The time rate of change of this content must be given by the flux $4\pi r^2 J^*$.

$$\begin{aligned} \frac{d}{dt} &[N + (N - M)e^{-\eta_1^2 D_0 t}] \\ &= -\eta_1^2 D_0 (N - M)e^{-\eta_1^2 D_0 t} = 4\pi r^2 J^*(r, t) \\ &= -4\pi R N^2 D_0 e^{-\eta_1^2 D_0 t} \end{aligned} \tag{C24}$$

in which (C21) has been substituted into (10.16) to pass from the third to the fourth expression. Comparing the second and fourth term of (C24) reveals

$$\eta_1^2 D_0 = \frac{4\pi R N^2 D_0}{(N - M)} = \frac{4\pi q^2 N^2 D_0}{\kappa k T(N - M)} \tag{C25}$$

or

$$\tau = \frac{1}{\eta_1^2 D_0} = \frac{\kappa k T(N - M)}{4\pi q^2 N^2 D_0} \tag{C26}$$

the value quoted in (10.25).

APPENDIX D

MINIMIZATION OF THE DIFFUSION POTENTIAL

In Section V the statement was made that equation (11.2) was a valid approximation everywhere within a p type region, provided that N_D did not fluctuate through ranges of order N_A in shorter distances than

$$\ell = \sqrt{\frac{\pi\kappa kT}{q^2N_A}} \tag{D1}$$

This statement will now be proved.

The electrostatic potential is determined by the space charge equation[31]

$$\frac{d^2V}{dx^2} = -\frac{4\pi q}{\kappa}\,[N_D(x) + p(x) - N_A] \tag{D2}$$

where we assume that the material is everywhere p-type so that the electron density, n, does not enter the right side of (D2). Furthermore, the mobility of holes is so much greater than that of donor ions that the former may be considered to always be at equilibrium with respect to the distribution of the latter. Boltzmann's law[29] may then be applied to p. The result is

$$p = N_A \exp\,[-qV/kT] \tag{D3}$$

where the potential is taken to be zero when $p = N_A$.

Choose an arbitrary point, x_0, where the potential is V_0 and investigate (D2) in its neighborhood. We wish to determine the conditions under which the right side of (D2) may be approximated by zero, i.e., the "no-space-charge condition," in this neighborhood. The limits of the neighborhood will be defined such that

$$|\,V - V_0\,| = |\,u\,| \leqq kT/2q \tag{D4}$$

so that, in it, the exponential in (D3) can be linearized

$$p = N_A \exp\,[-\,qV_0/kT]\left(1 - \frac{qu}{kT}\right) \tag{D5}$$

Then (D2) becomes

$$\frac{d^2u}{dx^2} = \frac{4\pi q}{\kappa}\left\{N_A\,[1 - \exp\,(-\,qV_0/kT)] - N_D(x) + \left[\frac{qN_A}{kT}\exp\,(-\,qV_0/kT)\right]u\right\} \tag{D6}$$

The no space charge condition in the defined region is therefore

$$u = \left(\frac{kT \exp (qV_0/kT)}{qN_A}\right) N_D(x) + \left(\frac{kT}{q}\right) \frac{\exp (-qV_0/kT) - 1}{\exp (-qV_0/kT)} \quad \text{(D7)}$$

To simplify notation define

$$\exp[-qV_0/kT] = \gamma_0 \quad \text{(D8)}$$

Next expand both N_D and u in Fourier series

$$N_D = \sum_{s=0}^{\infty} A_s \sin sx + B_s \cos sx \quad \text{(D9)}$$

$$u = \sum_{s=0}^{\infty} \alpha_3 \sin sx + \beta_s \cos sx \quad \text{(D10)}$$

Substitution of (D9) and (D10) into (D6) and equating coefficients of like terms leads to the set of relations

$$\beta_0 = \frac{kT}{qN_A\gamma_0} [N_A(\gamma_0 - 1) + B_0] \quad \text{(D11)}$$

$$\alpha_s = \frac{4\pi q}{\kappa} \left(\frac{\ell^2/4\pi^2\gamma_0}{1 + (s^2\ell^2/4\pi^2\gamma_0)}\right) A_s \quad \text{(D12)}$$

$$\beta_s = \frac{4\pi q}{\kappa} \left(\frac{\ell^2/4\pi^2\gamma_0}{1 + (s^2\ell^2/4\pi^2\gamma_0)}\right) B_s \quad \text{(D13)}$$

Now the wavelength of the sth component in (D9) is

$$\lambda_s = 2\pi/s \quad \text{(D14)}$$

If N_D contains no important components of wavelength shorter than

$$\frac{\ell}{\sqrt{\gamma_0}} \quad \text{(D15)}$$

the B_k for such components may be set equal to zero. But then the only terms which appear in (D12) and (D13) are terms where the denominators which (with the aid of (D14)) may be written as

$$\kappa \left(1 + \frac{\ell^2}{\gamma_0\lambda_s^2}\right) \quad \text{(D16)}$$

may be set equal to κ. Thus we have in place of (D12) and (D13)

$$\alpha_s = \frac{q\ell^2}{\kappa\pi} A_s = \frac{kT}{qN_A\gamma_0} A_s \quad \text{(D17)}$$

$$\beta_s = \frac{q^2\ell^2}{\kappa\pi} B_s = \frac{kT}{qN_A\gamma_0} B_s \tag{D18}$$

The requirement that N_D contain no Fourier terms of wavelength shorter than (D15) is obviously the condition that N_D never pass from its maximum to its minimum value in a distance shorter than D(15). Since we are assuming that N_D may at places be of order N_A, and at others, of order zero, this amounts to the condition that N_D does not fluctuate over ranges comparable with N_A in distances shorter than (D15).

The use of (D11), (D17), and (D18) in (D10) yields

$$\begin{aligned} u &= \frac{kT}{qN_A\gamma_0}\left[N_A(\gamma_0 - 1) + \sum_{s=0}^{\infty} (A_s \sin sx + B_s \cos sx)\right] \\ &= \frac{kT}{q}\frac{(\gamma_0 - 1)}{(\gamma_0)} + \frac{kT}{q\gamma_0}\frac{N_D}{N_A} \end{aligned} \tag{D19}$$

which by reference to the definition (D8) for γ_0 proves to be identical with (D7), the no-space-charge condition.

Equation (D19) is only true when N_D does not fluctuate through ranges of order, N_A, in distances smaller than $\ell/\sqrt{\gamma_0}$. This distance depends on γ_0 and thus on the point where $V = V_0$, whose neighborhood is being explored. Thus, we may say that there will be no space charge at all points whose V_0 is such as to fix γ_0 at a value such that

$$\gamma_0 > \frac{\ell^2}{\lambda_{\min}^2} \tag{D20}$$

where $\lambda_{\min}$ is the minimum wavelength which needs to be considered in the Fourier expansion of N_D. In terms of the definition of γ_0 this means

$$V_0 < \frac{kT}{q} \ln \frac{\lambda_{\min}^2}{\ell^2} \tag{D21}$$

Thus, at all points where V_0 is less than the right side of (D21) the no space charge approximation will hold. (D21) shows, that in the limit when $\lambda_{\min}$ goes toward zero, i.e. when the infinite series must be used for N_D, the right side of (D21) will approach $-\infty$ and V_0 will satisfy (D21) hardly anywhere. Thus space charge will exist almost everywhere.

In most diffusion problems the extremes of potential will occur in regions where there is no space charge. Thus in one extreme N_D may equal $0.9\ N_A$ and in the other it may equal zero. If there is no space charge in these extremes we may write for them

$$N_A - N_D = p = N_A \exp(-qV/kT) \tag{D22}$$

in which (D3) has been used. Setting N_D equal to zero in one extreme

yields $V = 0$. In the other extreme $N_D = 0.9\, N_A$ so that we get

$$V = \frac{kT}{q}\, \ell n\ 10 \tag{D23}$$

This therefore is the largest value which V_0 may assume in our case. Inserting the expression in D21 in place of V_0 we end with the relation

$$10 < \frac{\lambda_{\min}^2}{\ell^2} \tag{D24}$$

Thus provided that in the distribution being considered

$$\lambda_{\min} > 3.5\ell \tag{D25}$$

there will be no space charge anywhere.

At high temperatures $0.1\, N_A$ may be less than n_i. Under these conditions (D24) should be replaced by

$$\frac{N_A}{n_i} < \frac{\lambda_{\min}^2}{\ell^2} \tag{D26}$$

and in the limit that n_i becomes very large it is obvious that (D26) will always be satisfied. The rule to be enunciated for the cases we shall be interested in is the one given in section XI, i.e. that no space charge will exist provided that $\lambda^2_{\min}$ is no less than order, ℓ.

Appendix E

CALCULATION OF DIFFUSIVITIES FROM CONDUCTANCES OF DIFFUSION LAYERS

In this appendix equation (11.12) will be derived. In the first place we note that the dependence of N_D on position x, and time t, will be of the form $N_D(x/\sqrt{t})$ at any stage of the diffusion process. This results from a theorem due to Boltzmann[61] that when the dependence of D upon x and t is of the form $D(N_D)$, i.e., the dependence is through N_D, and a semi-infinite region extending from $x = 0$ to $x = \infty$ is being considered, then, in the case of plane parallel diffusion, the only variable in the problem will be $x/\sqrt{t}$.

Although the wafers considered in Section XI are of finite thickness d, the stages of diffusion investigated are such that the two regions of loss near the surfaces have not contacted each other. As a result the system behaves like two semi-infinite regions backed against one another, and the preceding arguments hold. The conductance Σ, defined in section XI will be proportional to the integral of the product of the local carrier

density by the local mobility. Thus

$$\Sigma = \omega \int_0^{d/2} \mu(x, t)[N_A - N_D(x, t)]\, dx \tag{E1}$$

where ω is a proportionality constant and $\mu(x, t)$ is the local mobility. An upper limit of $d/2$ rather than d is used because of symmetry. The local mobility will vary because N_D, and therefore the local density of charged impurity scatterers,[54] varies. Let $N_D{}^0$ be the initial uniform density (before any diffusion out) of donors, and write ($E1$) as

$$\begin{aligned} \Sigma &= \omega \int_0^{d/2} \mu(x, t)[N_A - N_D{}^0 + N_D{}^0 - N_D(x, t)]\, dx \\ &= \omega \int_0^{d/2} \mu(x, t)[N_A - N_D{}^0]\, dx + \omega \int_0^{\infty} \mu(x, t)[N_D{}^0 \\ &\qquad - N_D(x, t)]\, dx \end{aligned} \tag{E2}$$

The second integral on the right of (E2) is given the upper limit ∞, because in the experiments we wish to perform $N_D{}^0 - N_D$ becomes zero long before x reaches $d/2$.

Now in the first integral on the right of (E2) we may set $\mu(x, t)$ equal to the constant value μ_0, which it assumes in the bulk of the wafer, because the breadth of the depletion layer near the surface (in which $\mu(x, t)$ departs from μ_0) is small compared to $d/2$. The same thing cannot be done in the second integral since the integrand vanishes beyond the depletion layer and the total contribution *comes* from that layer. We thus obtain

$$\begin{aligned} \Sigma &= \omega\mu^0(\mathrm{N}_A - \mathrm{N}_D{}^0)\, d/2 \\ &\qquad + \omega \int_0^{\infty} \mu\left(\frac{x}{\sqrt{t}}\right)\left[\mathrm{N}_D{}^0 - \mathrm{N}_D\left(\frac{x}{\sqrt{t}}\right)\right] dx \end{aligned} \tag{E3}$$

In the integral in (E3) both μ and N_D are represented as functions of $x/\sqrt{t}$, the latter because of what has been said above, and the former, because it is a function of the latter. Defining

$$\nu = x/2\sqrt{Dt} \tag{E4}$$

in which D is constant, and substituting in (E3) gives finally

$$\Sigma = \omega\mu_0(\mathrm{N}_A - \mathrm{N}_D{}^0)d/2 + 2\omega\sqrt{Dt} \int_0^{\infty} \mu(\nu)[\mathrm{N}_D{}^0 - \mathrm{N}_D(\nu)]d\nu \tag{E5}$$

Since the definite integral is a constant (E5) shows that Σ is a linear function of $\sqrt{t}$, a fact mentioned in section XI.

In order to make use of the measured dependence of Σ on $\sqrt{t}$ to determine diffusivities, the functions $\mu(\nu)$ and $N_D(\nu)$ must be specified. For the latter we shall assume the Fick's law solution[62]

$$N_D = N_D^0 \operatorname{erf} \nu \tag{E6}$$

going with constant D, and $N_D = 0$ as a boundary condition at $x = 0$ at the surface. (In section XI the limitations of this assumption in the presence of ion pairing and diffusion potential are discussed.) The ν dependence of μ is more complicated. In general, we shall be concerned with electrical measurements in two extreme cases. In the first case ion pairing, under the condition of measurement, is everywhere complete so that the local density of scatterers will be given by

$$N_A - N_D(\nu) \tag{E7}$$

In the other case ion pairing will be entirely absent, so that the local scatterer density, will be specified by

$$N_A + N_D(\nu) \tag{E8}$$

In all experiments N_A will be only slightly greater than N_D^0 so that it may be replaced by this quantity. Doing this, and substituting (E6) into (E8) and (E9) gives

$$N_D^0 \operatorname{erfc} \nu = N(\nu) \tag{E9}$$

for the scattering density in the ion pairing case, and

$$N_D^0(1 + \operatorname{erf} \nu) = N(\nu) \tag{E10}$$

for the no pairing case.

Since almost all our experiments have been in germanium we now specialize our attention to that substance. However, the procedure invoked below can be applied to silicon as well.

The dependence of hole mobility, μ, on scattering density, N, for germanium at room temperature is shown in Fig. 30 taken from Prince's data.[63] The integral in (E5) assumes the form

$$\mathrm{N}_D^0 \int_0^\infty \mu(\mathrm{N}(\nu)) \operatorname{erfc} \nu \, d\nu. \tag{E11}$$

Choosing $N(\nu)$ as either (E9) or (E10) and using Fig. 30 together with a table of error functions makes the numerical evaluation of (E11) possible. Since $N(\nu)$ given by (E9) or (E10) depends on N_D^0, so will the integral.

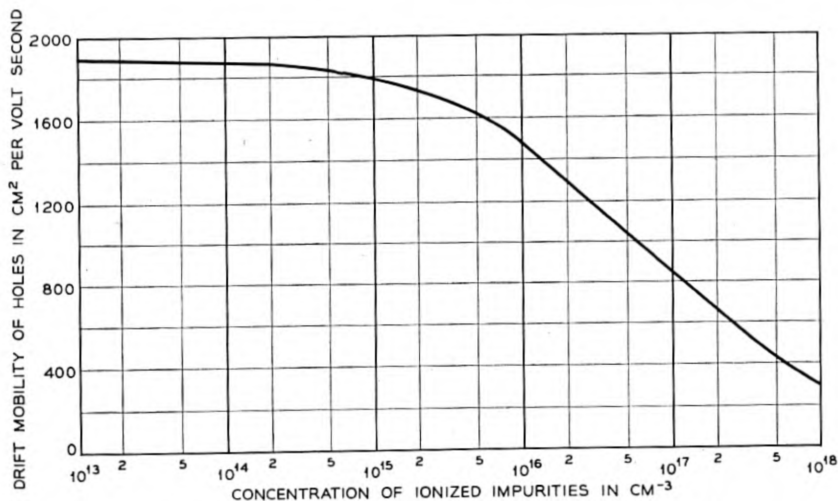


Fig. 30 — Plot of hole-drift mobility in germanium as a function of ionized impurity concentration after Prince.

The numerical evaluation has been performed for a range of N_D^0 in both the pairing and non-pairing cases. In this manner it has been possible to evaluate the "correction factor" ϑ defined by the following equation

$$\int_0^\infty \mu(\nu) \operatorname{erfc} \nu \, d\nu = \vartheta\mu_\infty \int_0^\infty \operatorname{erfc} \nu \, d\nu \tag{E12}$$

$$= \vartheta\mu_\infty(0.563)$$

where μ_∞ is the mobility in the presence of N_A scatterers. Fig. 22 contains plots (for germanium) of $\vartheta(N_D^0)$ versus N_D^0 for both the pairing and non-pairing cases. It is seen that ϑ is never much different from unity.

Equation (E5) can now be written as

$$\Sigma = \omega\mu_0(N_A - N_D^0)\, d/2 + \omega\mu_\infty[1.128\vartheta N_D^0\sqrt{D}]\sqrt{t} \tag{E13}$$

Defining

$$\Sigma_0 = \omega\mu_0(\mathrm{N}_A - \mathrm{N}_D^0)d/2 \tag{E14}$$

$$\Sigma_\infty = \frac{\omega\mu_\infty \mathrm{N}_A D}{2} \tag{E15}$$

it is obvious that Σ_0 is the conductance before any donor has diffused

out and Σ_∞ after all the donor has been diffused out. With these definitions (E16) becomes

$$\Sigma/\Sigma_0 = 1 + \frac{2.256\vartheta\mu_\infty\sqrt{D}}{\mu_0 d}\left(\frac{N_D{}^0}{N_A - N_D{}^0}\right)\sqrt{t} \tag{E17}$$

Calling the slope of this curve S leads to the result

$$S = \frac{2.256\vartheta\mu_\infty\sqrt{D}}{\mu_0 d}\left(\frac{N_D{}^0}{N_A - N_D{}^0}\right) \tag{E18}$$

or using (E14) and (E15)

$$D = \left(\frac{Sd\,\Sigma_0\,N_A}{2.256\vartheta\,\Sigma_\infty\,N_D{}^0}\right)^2 \tag{E19}$$

This is equivalent to equation (11.12).

Glossary of Symbols

a	distance of closest approach of two ions of opposite sign
A	constant in expression for ρ in section on relaxation theory
A^-	concentration of ionized acceptors
A_0	A_η going with $\eta = 0$
A_1	A_η going with η_1
A_η	constant preceding the ηth eigenfunction in solution of the relaxation problem
A_s	coefficient of sin sx in Fourier expression for N_D
b	$q^2/2\kappa kT$, position of minimum in $g(r)$
B	constant in expression for ρ in section on relaxation theory
B^-	boron ion
$B(Si)$	un-ionized boron in silicon
B_s	coefficient of cos sx in Fourier expression for N_D
$c(r)$	concentration of positive ions in atmosphere of a negative ion
C	concentration of LiB^-
d	thickness of wafer in diffusion experiment
D	diffusivity of donor ion in the most general sense
D_0	diffusivity of donor ion in the absence of pairing
D^+	concentration of ionized donors
$D_0{}^+$	value of D^+ in the absence of acceptor
$D_*{}^+$	concentration of mobile donor ions where $V = 0$
e^-	conduction band electron
e^+	valence band hole

e^+e^-	recombined hole-electron pair
E	energy level in electron gas
E_D	ionization energy of a donor
E_A	ionization energy of an acceptor
E_i	energy level in conduction band
$E(r)$	chance that volume $4\pi r^3/3$ will not contain an ion
f	flux density
F	Fermi level — also constant in equation (7.21)
g_i	density of states of energy E_i in conduction band
$g(r)$	nearest neighbor distribution function at equilibrium
G	Gibbs free energy of electron assembly
Ga^-	gallium ion in germanium
G_η	space dependent part of relaxation eigenfunction
G_0	G_η for $\eta = 0$
G_1	G_η for $\eta = \eta_1$
h	Plank's constant — also used for normalizing constant in $c(r)$
h_j	number of holes in the jth energy level
H	net local density of fixed donors
$i(\rho_2, \rho_1)$	$\varepsilon^3 I(r_2, r_1)$
I	field current in diffusion measurement
$I(r_2, r_1)$	integral for ion pairing calculations taken between r_1 and r_2
$\vec{J}(\vec{r})$	current in the atmosphere of a nearest neighbor
J^*	flux density of ions being trapped
k	Boltzmann's constant
k_1	first order rate constant in relaxation theory
k_2	second order rate constant in relaxation theory
K_0	distribution coefficient of donor between semiconductor and external phase
K_1	electron-hole recombination equilibrium constant
K_A	ionization constant of acceptor
K_D	ionization constant of donor
K_j	constant relating ω_j to volume, V
K^*	product of K_D, K_0, and α
ℓ	screening length for diffusion potential
L	Debye length — also used for radius of volume, $1/N$
Li^+	lithium ion
$Li(Sn)$	lithium in molten tin
$Li(Si)$	un-ionized lithium in silicon
$LiSi$	lithium-silicon complex
LiB	un-ionized LiB^-

LiB^-	lithium-boron complex ion in semiconductor
$[Li^+B^-]$	lithium-boron ion pair
$[Li^+Ga^-]$	lithium-gallium ion pair
m_0	normal mass of electron
m_p	effective mass of a hole
M	normalizing constant in relaxation theory
n	concentration of conduction electrons — also used for density of untrapped ions in relaxation
n_i	intrinsic concentration of electrons
N_A	total acceptor concentration
N_D	total donor concentration
$N_D{}^0$	total solubility of donor in undoped semiconductor—also used for initial density of donors in diffusion experiments
N	ion concentration in an electrolyte solution—also used for initial value of n in relaxation—also used for concentration of immobile donors in Appendix A
$N_D{}^*$	solubility of donor in absence of ion pairing in Appendix A
p	concentration of holes
P	concentration of ion pairs
q	charge on an ion
$Q(\alpha)$	tabulated integral for computing Ω
r	distance between positive and negative ions in a pair
r_1	a particular value of r
r_2	a particular value of r
R	capture radius of an ion in relaxation
S	slope of Σ/Σ_0 versus $\sqrt{t}$ curve
S_η	time dependent part of relaxation eigenfunction belonging to eigenvalue η
t	time
T	temperature
u	$V - V_0$
V	electrostatic potential — also used for volume — also used for potential difference between probe points — also used for potential energy of a positive in neighborhood of negative ion
V_0	electrostatic potential where $x = x_0$
x	variable of integration — same as r also rectangular position coordinate
x_0	special value of x.

α — ε/a—also used for thermodynamic activity of donor in external phase
α_s — coefficient of sin sx in Fourier expression for u
β — constant in exponential in LiB^- equilibrium constant
β^* — constant in exponential in expression for vacancy concentration
β_0 — β_s for $s = 0$
β_s — coefficient of cos sx in fourier expression for u
γ — pre-exponential factor in LiB^- equilibrium constant
γ^* — pre-exponential factor in expression for vacancy concentration
γ_0 — $\exp[-qV_0/kT]$
$\Gamma(\vec{r})$ — non-equilibrium nearest neighbor distribution function
δ — constant appearing in Appendix C
ε — $q^2/\kappa kT$
η — eigenvalue in relaxation problem
η_1 — second eigenvalue in set of η
θ — fraction of donor paired
ϑ — correction factor for variable carrier mobility
κ — dielectric constant
λ — x/ε
λ_s — $2\pi/s$, wavelength of sth component of fourier series
λ_{min} — wavelength of component of fourier series for N_D, having minimum wavelength
μ — chemical potential of donor in an external phase — also used for mobility of donor ion — also used for local carrier mobility
μ^0 — chemical potential of donor in external phase in standard state
μ_{D^+} — chemical potential of donor ion
$\mu_{D^+}{}^0$ — chemical potential of donor ion in the standard state
μ_e — chemical potential of an electron
μ_D — chemical potential of donor atom in semiconductor
μ_D^0 — chemical potential of donor atom in standard state
μ_0 — mobility of donor atom at infinite dilution — also used for carrier mobility in diffusion experiments before diffusion
μ_∞ — carrier mobility in diffusion experiments after all diffusant has diffused out
ν — $x/2\sqrt{Dt}$
ξ — ε/r
π — LiB^- equilibrium constant
ρ — resistivity of gallium-doped germanium after saturation with

lithium — also used for local charge density in Poisson's equation — also used for density of diffusing positive ions in relaxation

ρ_0	resistivity of gallium-doped germanium before saturation with lithium
ρ_1	r_1/ε
ρ_2	r_2/ε
σ	conductivity during relaxation
σ_∞	conductivity in relaxed state
Σ	conductance between probe points
Σ_0	conductance before diffusion begins in diffusion experiments
Σ_∞	conductance after diffusion is over in diffusion experiments
τ	relaxation time
Φ	constant in relaxation formula for conductivity
Ψ	local electrostatic potential in ionic atmosphere
ω	proportionality constant connecting conductance between probe points with integral over carrier concentration
ω_j	number of states in jth level of electronic energy diagram
Ω	ion pairing equilibrium constant
$\square$	vacant lattice site in covalent crystal
$\square^-$	negatively charged cation vacancy

REFERENCES

1. Wagner, C., Z. Phys. Chem., **B21,** p. 25, 1933, **B32,** p. 447, 1936.
2. Taylor, H. S., and Taylor, H. A., Elementary Physical Chemistry, p. 343, Van Nostrand, 1937.
3. Shockley, W., Electrons and Holes in Semiconductors, p. 6, Van Nostrand, 1950.
4. Shockley, W., Electrons and Holes in Semiconductors, Van Nostrand, 1950.
5. Reiss, H., J. Chem. Phys., **21,** p. 1209, 1953.
6. Reiss, H., and Fuller, C. S., J. Metals, **12,** p. 276, 1956.
7. Fuller, C. S., and Ditzenberger, J. A., J. App. Phys., **25,** p. 1439, 1954.
8. Fuller, C. S., and Ditzenberger, J. A., Phys. Rev., **91,** p. 193, 1953.
9. MacDougall, F. H., Thermodynamics and Chemistry, p. 143, Wiley, 1939.
10. Miller, F. W., Elementary Theory of Qualitative Analysis, p. 102, Century Company, New York, 1929.
11. Fuller, C. S., Record of Chemical Progress, **17,** No. 2, 1956.
12. Wagner, C., and Grünewald, K., Z. Phys. Chem., **B40,** p. 455, 1938.
13. von Baumbach, H. H., and Wagner, C., Z. Phys. Chem., **22B,** p. 199, 1933.
14. Kröger, F. A., and Vink, H. J., Physica, **20,** p. 950, 1954.
15. MacDougall, F. H., Thermodynamics and Chemistry, p. 258, Wiley, 1939.
16. Shockley, W., Electrons and Holes in Semiconductors, p. 231, Van Nostrand, 1950.
17. Mayer, J. E., and Mayer, M. G., Statistical Mechanics, p. 120, Wiley, 1940.
18. MacDougall, F. H., Thermodynamics and Chemistry, p. 137, Wiley, 1939.
19. Lewis, G. N., and Randall, M. C., Thermodynamics, p. 258, McGraw Hill, 1923.

20. Mayer, J. E., and Mayer, M. G., Statistical Mechanics, p. 121, Wiley, 1940.
21. MacDougall, F. H., Thermodynamics and Chemistry, p. 261, Wiley, 1939.
22. MacDougall, F. H., Thermodynamics and Chemistry, p. 25, Wiley, 1939.
23. Engell, H. J., and Houffe, K., Z. Electrochem., **56**, p. 366, 1952, **57**, p. 762, 1953.
24. Shockley, W., Electrons and Holes in Semiconductors, p. 15, Van Nostrand, 1950.
25. Morin, F. J., and Maita, J. P., Phys. Rev., **94**, p. 1525, 1954.
26. Morin, F. J., and Maita, J. P., Phys. Rev., **96**, p. 28, 1954.
27. Shockley, W., Electrons and Holes in Semiconductors, p. 86, Van Nostrand, 1950.
28. Shockley, W., Electrons and Holes in Semiconductors, p. 88, Van Nostrand, 1950.
29. Fowler, R. H., Statistical Mechanics, p. 48, Cambridge, 1929.
30. Slater, J. C., and Frank, N. H., Introduction to Theoretical Physics, p. 212, McGraw Hill, 1933.
31. Shockley, W., B.S.T.J., **28**, p. 435, 1949.
32. Fuller, C. S., and Ditzenberger, J. A., J. App. Phys., May, 1956.
33. Shulman, R. G., and McMahon, M. E., J. App. Phys., **24**, p. 1267, 1953.
34. Reiss, H., Fuller, C. S., and Pietrusczkiewicz, A. J., J. Chem. Phys. (in press).
35. Eyring, H., Walter, J., and Kimball, G. E., Quantum Chemistry, p. 231, Wiley, 1946.
36. Pauling, L., The Nature of the Chemical Bond, p. 179, Cornell, 1942.
37. Debye, P., and Huckel, E., Phys. Z., **24**, p. 195, 1923.
38. Kirkwood, J. G., J. Chem. Phys., **2**, p. 767, 1934.
39. Briggs, H. B., Phys. Rev., **77**, p. 287, 1950.
40. Briggs, H. B., Phys. Rev., **77**, p. 287, 1950.
41. Wyman, Phys. Rev., **35**, p. 623, 1930.
42. Bjerrum, N., Kgle. Danske Vidensk. Selskab., **7**, No. 9, 1926.
43. Fuoss, R. M., Trans. Faraday Soc., **30**, p. 967, 1934.
44. Reiss, H., J. Chem. Phys. (in press).
45. Reiss, H., J. Chem. Phys. (in press).
46. Shockley, W., and Read, W. T., Jr., Phys. Rev., **87**, p. 835, 1952, Haynes, J. R., and Hornbeck, J. A., Phys. Rev., **90**, p. 152, 1953, **97**, p. 311, 1955.
47. Harned and Owen, The Physical Chemistry of Electrolytes, p. 123, A. C. S. Monograph, 1950.
48. Carslaw, H. S., and Jaeger, J. C., Conduction of Heat in Solids, p. 209, Oxford, 1948.
49. Glasstone, S., Textbook of Physical Chemistry, p. 1231, Van Nostrand, 1940.
50. Shockley, W., Electrons and Holes in Semiconductors, p. 300, Van Nostrand, 1950.
51. Slater, J. C., and Frank, N. H., Introduction to Theoretical Physics, p. 186, McGraw Hill, 1933.
52. Fuller, C. S., and Severiens, J. C., Phys. Rev., **95**, p. 21, 1954.
53. Shockley, W., B.S.T.J., **28**, p. 435, 1949.
54. Shockley, W., Electrons and Holes in Semiconductors, p. 258, Van Nostrand, 1950.
55. Theuerer, H. C., U. S. Pat. No. 2542727.
56. Margeneau, H., and Murphy, G. M., The Mathematics of Physics and Chemistry, p. 213, Van Nostrand, 1943.
57. Margeneau, H., and Murphy, G. M., The Mathematics of Physics and Chemistry, p. 72, Van Nostrand, 1943.
58. Reiss, H., and La Mer, V. K., J. Chem. Phys., **18**, p. 1, 1950.
59. Reiss, H., J. Chem. Phys., **19**, p. 482, 1951.
60. Carslaw, H. S., and Jaeger. J. C. Conduction of Heat in Solids, p. 40, Oxford, 1948.
61. Boltzmann, L., Ann. Phys., **53**, p. 959, 1894.
62. Carslaw, H. S., and Jaeger, J. C., Conduction of Heat in Solids, p. 41, Oxford, 1948.
63. Prince, M. B., Phys. Rev., **92**, p. 681, 1953, **93**, p. 1204, 1954.
64. Tyler, W. W., and Woodbury, H. H., Bull. Am. Phys. Soc., **30**, No. 7, p. 32, 1955.
65. Debye, P. P., Conwell, E. M., Phys. Rev., **93**, p. 693, 1954.

66. Shockley, W., Electrons and Holes in Semiconductors, Chapter 8, Van Nostrand, 1950.
67. Shockley, W., Electrons and Holes in Semiconductors, Chapter 16, Van Nostrand, 1950.
68. Geballe, T. H., and Morin, F. J., Phys. Rev., **95**, p. 1085, 1954.
69. Conant, J. B., The Chemistry of Organic Compounds, p. 196, Macmillan, 1939.
70. Valenta, M., and Ramasastry, C., Symposium on Semiconductors, Meeting I.M.D., and A.I.M.E., Feb. 20, 1956.
71. Longini, R. E., and Green, R., Phys. Rev. (in press).

Single Crystals of Exceptional Perfection and Uniformity by Zone Leveling

By D. C. BENNETT and B. SAWYER

(Manuscript received January 23, 1956)

The zone-leveling process has been developed into a simple and effective tool, capable of growing large single crystals having high lattice perfection and containing an essentially uniform distribution of one or more desired impurities. Experimental work with germanium is discussed, and the possibility of broad application of the principles involved is indicated.

INTRODUCTION

The first publication describing the concept of zone melting appeared about four years ago.[1] As there defined, the term zone melting designates a class of solidification techniques, all of which involve the movement of one or more liquid zones through an elongated charge of meltable material. This simple concept has opened a whole new field of possibilities for utilizing the principles of melting and solidification.

The first zone melting technique to gain widespread usage was one for zone refining germanium by the passage of a number of liquid zones in succession through a germanium charge. This process may be quite properly compared to distillation, the essential difference being that the change in phase is from solid to liquid and back, instead of from liquid to vapor and back. The zone refining technique has been eminently successful in the purification of germanium. Harmful impurity concentrations are of the order of one part in 10^{10}. This is mainly because all the impurities whose segregation behavior in freezing germanium has been measured have segregation coefficients (see equation 1) differing from 1 by an order of magnitude or more.[2] During the zone refining operation, these impurities collect in the liquid zones and are swept with them to the ends of the charge, which may be later removed.

[1] Pfann, W. G., Trans. A.I.M.E., **194,** p. 747, 1952.
[2] Burton, J. A., Impurity centers in Germanium and Silicon, Physica, **20,** p. 845, 1954.

This paper deals with a second zone melting process, zone leveling,[1, 3] which has gained usage somewhat more slowly than zone refining, but which has proved to be a highly effective tool for distributing desired impurities uniformly throughout a charge. For this process, only one liquid zone is used and its composition is adjusted to produce the desired impurity concentration in the material which is solidified from the liquid zone. Appropriate precautions are taken to insure the production of single crystals, if the material is desired in this form.

Since the invention of zone leveling, the process has been developed into a precision tool and as such it has become a preferred practical method for growing germanium single crystals of uniform donor or acceptor content. It is the purpose of this paper to discuss the technical development of this process, which has had two chief objectives: (1) the attainment of the greatest possible uniformity of donor and/or acceptor impurity distribution in the crystal; and (2) the attainment of a germanium crystal lattice with a minimum of imperfections of all kinds. The presentation will cover the principles involved, the means developed and results achieved toward these objectives in that order.

The first applications of the principles of zone melting have been in the field of semiconductor materials processing, chiefly because there are no other known refining techniques capable of meeting the extremely stringent purity requirements necessary for material to be used in semiconductor devices. Nevertheless, it is clear that these relatively simple and very effective zone melting techniques are beginning to find a wide variety of useful applications throughout the general fields of metallurgy and chemical engineering.

BASIC PRINCIPLES

The basic concept, theory and experimental confirmation of zone leveling have been well covered in previous publications.[1, 3] Accordingly, the intention here is only to repeat the salient points of the theory with a special emphasis on the assumptions involved since it will be necessary to refer to them.

Fig. 1 is a schematic drawing of a zone leveling operation showing a liquid zone of constant volume containing a solute whose concentration is C_L. As the zone moves a distance Δx an increment of germanium is melted at the right end, and another is frozen at the left end. The concentration of solute in the newly frozen Δx of solid solution is C_S. The distribution coefficient k is now conveniently defined as the ratio

[3] Pfann, W. G., and Olsen, K. M., Physical Review, **89**, p. 322, 1953.

of these solute concentrations:

$$k = \frac{C_S}{C_L} \tag{1}$$

When $k < 1$, the freezing interface may be regarded as a filter permitting only a fraction k of the solute concentration in the liquid to pass into the growing solid and rejecting the rest to remain in the liquid. If the unmelted charge of solvent is pure — that is, if no solute passes into the zone at the melting interface it is readily seen that the liquid zone will be gradually depleted of its solute impurity content during passage through the charge.

An expression for the solute concentration in the solid, C_S, deposited there by the passage of one zone, for the case of "starting charge into pure solvent" has been derived[1] based on the following assumptions:

(1) The liquid volume is constant (both cross section of charge and zone length l are constant).

(2) k is constant.

(3) Mixing in the liquid is complete (i.e. concentration in the liquid is uniform).

(4) Diffusion in the solid is negligible.

The expression is

$$C_S = kC_{LO}\, e^{-kx/l} \tag{2}$$

where C_{L_0} is the initial concentration of impurities in the liquid, l is the zone length, and x is the distance moved by the solidifying interface. A set of C_S versus x/l curves is shown in Fig. 2 for various k's. From this figure it is readily seen that when k is small the decay of C_S is slow (i.e., the depletion of C_L is slight).

Largely because of this consideration, most of the practical work reported in this paper has utilized solutes in germanium having low segre-

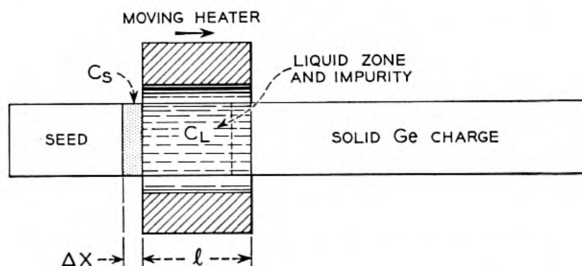


Fig. 1 — Schematic of zone leveling operation.

gation coefficients (usually antimony, whose $k = 0.003$ as donor, and indium whose $k = 0.001$ as acceptor). However, the principles of zone leveling are broad and capable of application to any solvent-solute system within the range of solubilities of its solid and liquid phases. The general method of attack[1] is first to find that composition of the liquid zone which will deposit the desired solid solution. Secondly, if one or more of the segregation coefficients involved is not small, the liquid zone must be maintained at its proper composition by admixing to the solid charge the same solutes that the zone will deposit in its product. Thus the solutes that are removed from the liquid zone at the freezing end will be replenished at the melting end.

The above mathematical treatment leads one to expect an essentially uniform solute distribution throughout a zone leveled crystal for the case under discussion in which k is small and the zone moves through a charge of pure solvent as indicated in Fig. 2. Irregular variations of C_S along the length or over the cross-section of the ingot are not predicted. The

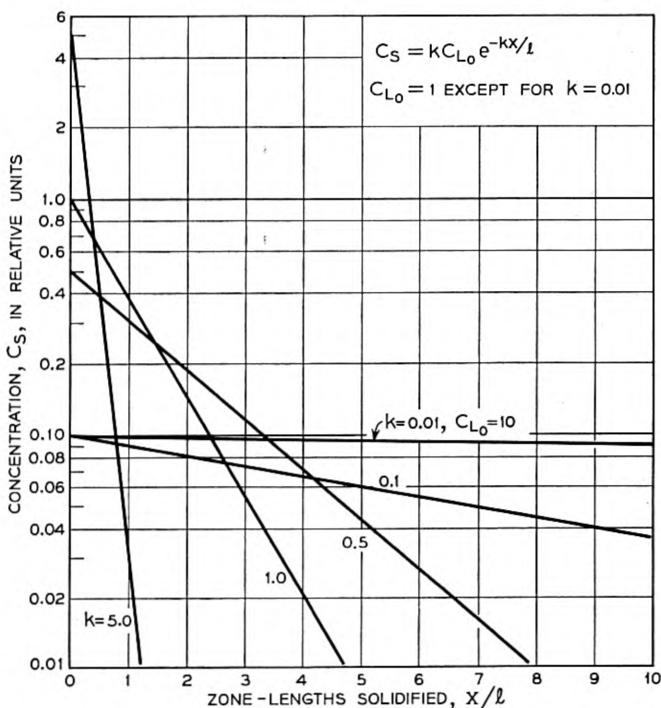


Fig. 2 — Solute concentration curves predicted for zone leveling with a starting charge of solute into pure solvent.

treatment is not concerned with lattice imperfections in the ingot such as dislocations, lineage, or grain boundaries. The predictions the theory does make have been well verified by experiment insofar as it has been possible to meet the assumptions enumerated above. However, as with most assumptions, their validity is sensitive to the experimental conditions, particularly in the cases of the first three. Much of the development effort, especially that toward improving resistivity uniformity, has been directed toward controlling the process so that these assumptions will be as nearly valid as possible.

Early experiments in zone leveling yielded crystals good enough to meet device requirements of that time. However, as semiconductor devices were designed to meet tighter design requirements, the demands on the germanium material grew more critical. Under these circumstances, it became necessary to examine the requirements on the product of the process and what precautions would be necessary to insure that its operation was under sufficient control. Accordingly, we shall discuss first the requirements on semiconductor material and then those critical aspects of the leveling operation which must be controlled to insure quality of the final product.

REQUIREMENTS ON GERMANIUM FOR SEMICONDUCTOR USES

The basic electrical bulk property of a germanium crystal is its conductivity or the reciprocal of that quantity, its resistivity. For a great majority of semiconductor uses, an extrinsic conductivity[4] is required in addition to the $\frac{1}{50}$ ohm^{-1} cm^{-1} intrinsic conductivity that results at room temperature from thermal excitation of electron-hole pairs in pure germanium. An extrinsic conductivity may be either n-type or p-type. Both of these may be produced by trace impurities distributed throughout the crystal, the n-type by donor impurities and the p-type by acceptor impurities. At room temperature donors give rise to conduction electrons and the acceptors to conduction holes which are free to move within the germanium crystal. If both donors and acceptors are present in the same crystal, the resulting electrons and holes recombine, leaving essentially the extrinsic conductivity contributed by the excess of one over the other, that is by $| N_D - N_A |$.

The fundamental requirement is, then, to control the net donor and the acceptor balance, $| N_D - N_A |$, to a predetermined value throughout the crystal. For most applications, the conductivity is to be increased by one or two orders of magnitude above the 27°C intrinsic value. An idea of the donor or acceptor concentrations involved may be acquired

[4] Shockley, W., Electrons and Holes in Semiconductors.

from noting that a conductivity of 1/5 ohm^{-1} cm^{-1} (that is, a conductivity increased by one order of magnitude) corresponds to a $N_D - N_A$ concentration of 7 parts per billion.

The next most commonly measured bulk property of germanium is the lifetime of minority carriers,[5] i.e., the time constant for decay by recombination of a surplus population of minority carriers artificially introduced into the crystal. Minority carriers are holes in n-type germanium or electrons in p-type germanium. This time constant may be regarded reasonably as a figure of merit for the crystal, being an indication of its freedom both from certain chemical impurities and from crystal faults, since these act as catalysts to the electron-hole recombination reaction. Normally, the highest possible lifetime is desired. Thus it becomes important to take extreme precautions during handling and processing of the germanium to avoid contamination, particularly by such known recombination center elements as nickel and copper[6] and it is also important to avoid crystal lattice faults such as dislocations, lineage, and grain boundaries.

Another observable quantity has recently been gaining acceptance as a more definite indication of mechanical crystal perfection than the minority carrier lifetime measurement. This is the etch pit density count, ϵ, (see Fig. 3) which is observed microscopically on an oriented (111) surface of a Ge crystal that has been etched three minutes in an agitated CP-4 etch (20 parts by volume concentrated HNO_3, 12 parts concentrated HF, 12 parts concentrated acetic acid, and 1/2 part Br_2). There is strong evidence[7] that the etch pits are formed at the intersections of dislocations with the surface of the crystal. While an etch pit count probably indicates only certain edge dislocations which intersect the surface of the crystal, it is at least a relative indication of the total dislocation density, and thus appears to be a highly useful index of crystal lattice perfection.

In the last year, evidence of a strong correlation has been observed[8] between certain electrical properties of alloy junctions, especially the breakdown voltage, and the etch pit density of the material on which the alloy junction is made. Accordingly, material to be used for alloy junction transistors is now selected on the basis of its maximum etch pit count and its freedom from lineage, twin, and grain boundaries.

The usual device test requirements on n- or p-type Ge material vary

[5] Valdes, L. B., Proc. I.R.E., **40**, p. 1420, 1952.

[6] Vogel, F. L., Read W. T., and Lovell, L. C., Phys. Rev., **94**, p. 1791, 1954.

[7] Vogel, F. L., Pfann, W. G., Corey, H. E., Thomas, E. E., Physical Review, **90**, p. 489, 1953.

[8] Zuk, P., and Westberg, R. W., private communication.

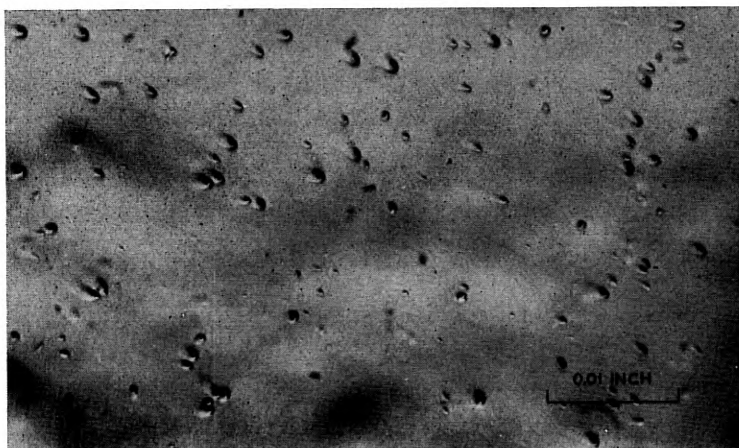


Fig. 3. — Microphotograph of Typical Etch Pits on (111) Plane.

from device to device, but may be summarized as follows:

(1) *Composition* — The donor-acceptor balance $N_D - N_A$ must be accurately controlled so that the resistivity, ρ, of the crystal is uniform and falls within acceptable tolerance limits.

(2) *Macro Perfection* — The crystal shall contain no grain boundaries, lineage, or twinning.

(3) *Micro Perfection* — The etch pit density, ϵ, must be lower than a certain empirically determined maximum.

(4) *Lifetime of Minority Carriers* — τ, must usually be above a certain minimum, although in many cases this minimum may be as low as a few microseconds.

Assuming macro perfection a consideration of these requirements leads directly to the two general objectives mentioned in the introduction of this paper: composition uniformity and control, and crystal lattice perfection. A third objective, high chemical purity, might also be inferred from the lifetime requirement, but the results obtained by zone refining raw material and by fairly standard laboratory techniques of cleaning and baking of furnace parts at high temperature have been satisfactory. Hence this objective has required little development effort. We proceed to a discussion of critical aspects of zone leveling in the light of the two major development objectives.

COMPOSITION UNIFORMITY AND CONTROL

The experimental development work described in this paper has been concerned with the distribution of two trace impurities, indium and anti-

mony, in a pure element, germanium. The traces are generally desired in concentrations varying from 1 to 100 parts per billion, ($\rho = 35$ to 0.35 ωcm). These amounts are too small to be detected by chemical or spectrographic means, but are readily detectable by electrical resistivity measurements. Although this application of zone leveling is very specific, it should be possible, as we have already suggested, to apply the experimental results to be described to more general systems. The subject of uniformity is conveniently discussed in two sections: (a) longitudinal resistivity uniformity, and (b) cross-sectional uniformity.

(a) *Longitudinal Composition Uniformity*

It has already been shown, by (2), that if the k is small, the variation in C_S over four or five zone lengths should be slight. This should be true either if a charge of pure germanium is used, or if a charge containing the same impurity present in the liquid zone is used, provided that the charge concentration of this impurity is of the same order of magnitude as that sought in the product. Where the solute has a small k, the leveling action of the zone is strong and the large C_L that is required is relatively unaffected by variations of the order of C_S.

The primary cause of observed variations in the *longitudinal* resistivity is fluctuation of the volume of the liquid zone. If this volume increases for any reason, the solute dissolved in it will be diluted. On the other hand, if the volume decreases, which can occur only when some of the liquid freezes and if k is small, most of the zone's solute will be concentrated in the smaller volume. Thus for small k's the concentration of solute in the liquid zone, C_L, varies inversely with the zone's volume. If C_L is to be constant, the volume must be constant, i.e. assumption (1) must be valid.

Unfortunately, the zone volume is directly affected by many variables, namely temperature fluctuation and drift, fluctuation in growth rate, variation in the cross-section of the unmelted charge, variation in the inert gas flow, and even cracks in the unmelted charge. For optimum control of longitudinal resistivity uniformity, it is, therefore, necessary to control all of these variables. The remainder of this section will consider their control.

Toward minimizing the effect of temperature variation on the zone volume, it is important to consider both the means of overall temperature control and the design of the temperature field which melts the liquid zone. It is clear that variation of the temperature field as a whole will directly affect the length of the liquid zone. Accordingly, it will be important to use a precision temperature controller in order to maintain a

constant zone length. The controller used here is a servo system that cycles the power on and off about ten times a second, adjusting the on fraction of the cycle according to the demands of a control thermocouple. The sensitivity of the controller is ±0.2°C at 940°C. With a liquid zone about 4 centimeters long and a temperature gradient of about 10°C per centimeter at the solidification interface, this degree of control should introduce longitudinal resistivity variations no greater than ±0.3 per cent.

When other requirements permit, it is possible to design a temperature contour to minimize the effects of control fluctuations. When the temperature gradients at the ends of the liquid zone are small, a slight change in the general temperature of the system will cause a relatively large change in the position of the solid-liquid interface. On the other hand, when the gradient is steep, the shift in position of the interface will be small. It is with this consideration in mind that a temperature gradient of about 130°C/cm is provided at the melting end of the liquid zone (Fig. 4). A steep gradient has the added advantage that it provides a large heat flux which is capable of supplying or removing the heat of solidification even at relatively fast leveling rates. Thus, a steep temperature gradient serves effectively to localize a solid-liquid interface. Other considerations, soon to be discussed, dictate that a small temperature gradient (about 10°C/cm) must be used at the freezing end of the zone. Accordingly, high precision of temperature control is required to properly stabilize the position of this solid-liquid interface.

Variation in the cross-section of the liquid zone may be controlled by using a boat with uniform cross-section, and by using as charge material which has been cast into a mold of controlled cross-section. Less precise control is obtained by using ingots from the zone refining process which were produced in a boat matched to the zone leveler boat. Even when care is used to maintain a uniform height of the zone refined ingot,[9] the control is less precise than in a casting.

A constant and uniform growth rate is important toward obtaining uniform longitudinal resistivity because segregation coefficients vary with growth rate.[10] This is especially true in the case of the k for antimony. Under steady state conditions, the growth rate is the rate at which the boat is pulled through the heater. A stiff pulling mechanism is required in order that the slow motion be steady. In the apparatus described here, a syncronous motor, operating through a gear reduction to drive a lead screw, has served to pull the boat smoothly over polished quartz rods.

[9] Pfann, W. G., J. Metals, **5**, p. 1441, 1953.
[10] Burton, J. A., Kolb, E. D., Slichter, W. P., Struthers, J. D., J. Chem. Phys., **21**, p. 1991, Nov., 1953.

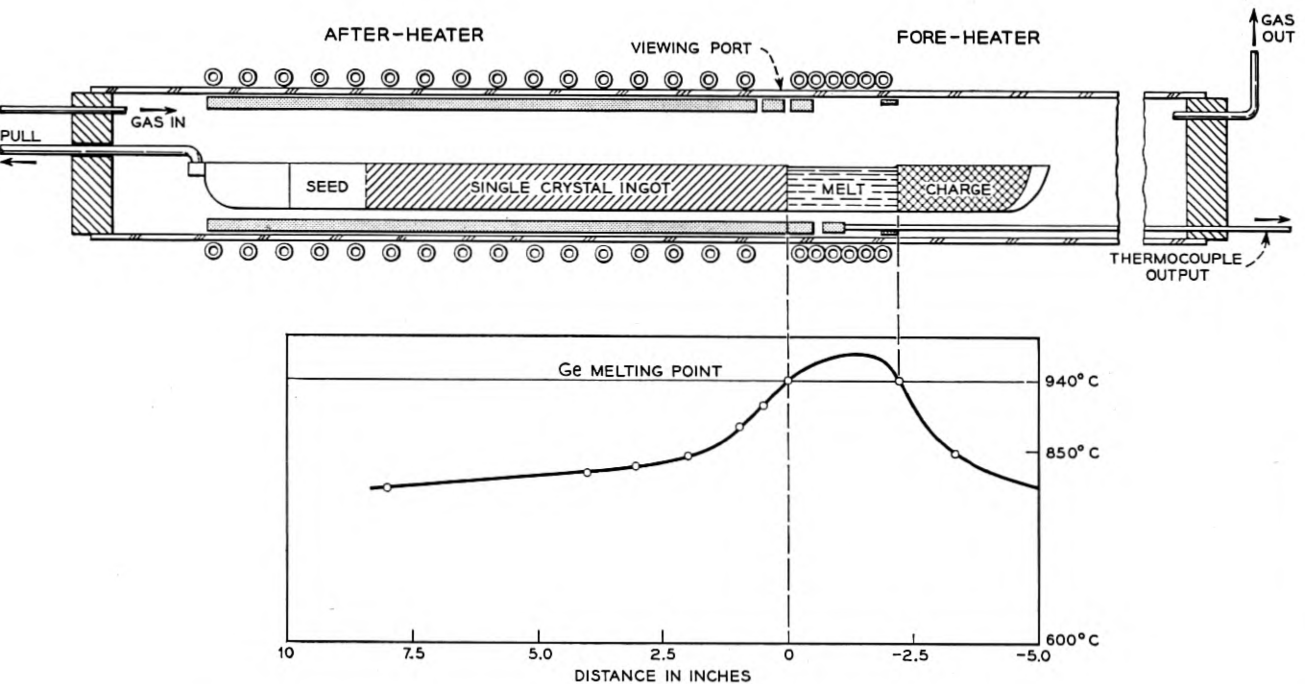


Fig. 4 — Axial temperature curve for the germanium zone leveler with after heater.

The true growth rate may be affected by factors that cause variations from steady state growth such as temperature and gas flow fluctuations. The need to control these variables has already been mentioned because of their effect on zone volume; their effect on growth rate is thus a second reason for their control.

Cracks or similar discontinuities in the unmelted charge act as barriers to heat flow. Thus they cause a local rise in temperature and lengthening of the liquid zone as the crack approaches the zone, until it is closed by melting. The resulting transient increase in liquid volume (and in ρ of the product) may be of the order of 10 per cent.

(b) *Cross-Sectional Composition Uniformity*

Difficulty may be expected in controlling the cross-sectional uniformity of the zone leveled ingot chiefly when the third assumption is invalid, i.e., when C_L throughout the liquid is non-uniform. As shown in the next paragraph, the true C_L must always rise locally near the solidifying interface due to the solute diffusion which is necessary when $k < 1$. However, it is possible to improve the validity of assumption 3 both by slowing the growth rate and by stirring the liquid zone.

One can form an estimate of a theoretically reasonable growth rate in terms of the rate of diffusion of impurities in liquid germanium. It should be noted that movement of a liquid zone containing a solute whose segregation coefficient is small implies a general movement by diffusion of essentially all the solute atoms away from the solidifying interface at a speed equal to the rate of motion of the zone. Even slow zone motion corresponds to a high diffusion flux of the solute through the liquid. As a consequence, the solute concentration must rise in front of the advancing solidification interface to a concentration $C_{L'}$ (see Fig. 5) until a concentration gradient is reached sufficient to provide a diffusion flux equal to the growth rate. Fick's Law of diffusion is useful here to calculate the extent of the rise in $C_{L'}$ at the growth interface, assuming the liquid to be at rest. The ratio of the maximum concentration to the bulk concentration may be taken from Fig. 5. If the maximum is to be no greater than 10 per cent above the mean, a maximum growth rate of 2×10^{-7} mils per second or 7×10^{-7} inches/hour would be required. Clearly, this rate is far too slow to provide an economical means of growing single crystals. For a practical process, it will be necessary to use non-equilibrium conditions at growth rates that must result in appreciable concentration differences within the liquid zone. Of course, the slower the growth rate the smaller will be the diffusion gradient and the higher will be the expected cross-sectional uniformity.

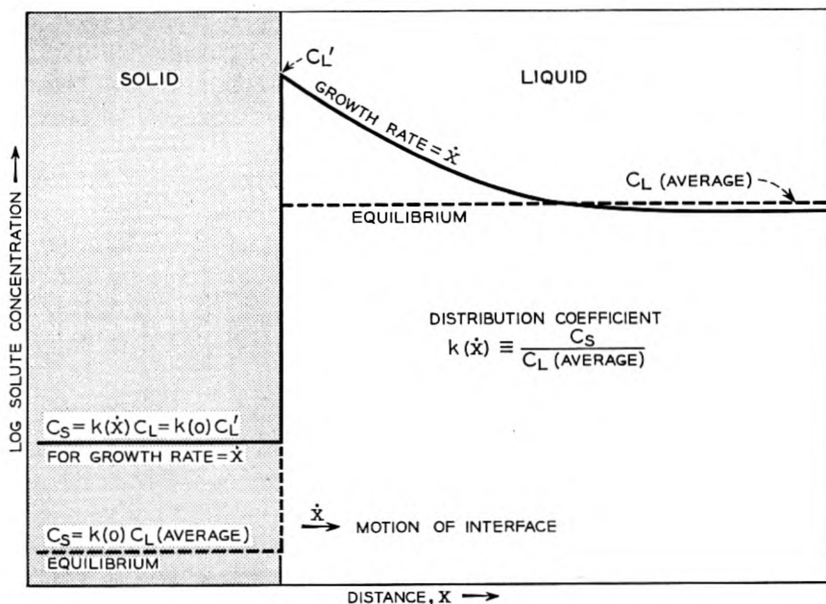


Fig. 5 — Solute concentration in solid and liquid at equilibrium and at finite growth rates.

If the liquid were static, that is, without any currents, it should be possible to obtain a uniform, controlled solute concentration in the solid even at appreciable growth rates, merely by adjusting the average concentration in the liquid to arrange that the C_L obtained at the growing interface will be the desired one. Instead of working with the equilibrium distribution coefficient k_0, one works with an effective distribution coefficient $k(\dot{x})$ for the given growth rate, $\dot{x}$:

$$k(\dot{x}) = \frac{C_S}{C_L(\text{ave})} \tag{3}$$

In practice, however, the situation is complicated by the existence of convection currents in the liquid zone. It is true that these currents tend to stir the liquid zone and thereby to minimize the concentration gradient within it. However, the currents are not uniform over the growing interface and they carry liquid of varying concentrations past the interface, causing fluctuations in C_S. Since these convection currents cannot be eliminated, one turns to the alternative of using forced stirring of the liquid zone. Such a forced stirring is readily available when RF induction heating is used by allowing the RF field to couple directly with the

liquid zone.[11] The resulting stirring currents are shown schematically in Fig. 6. It is seen that the liquid is moved from the center of the zone along its axis toward both ends. There it passes radially outward across the interface and returns along the outside of the zone to its center. These stirring currents are faster than convection currents and tend to minimize the rise of C_L at the solidification interface and to improve the uniformity of C_L and of crystal growth conditions in general over the freezing interface.

CRYSTAL LATTICE PERFECTION

A single edge dislocation in germanium may be regarded as a line of free valence bonds. The dislocation line is believed to have about 4×10^6 potential acceptor centers per centimeter, producing a space charge in the neighboring germanium and strongly modifying its semiconductor properties.[12] A lineage boundary (a term found useful to designate a low angle grain boundary) is a set of regularly spaced dislocations, and may be regarded as a surface of p-type material. Since the basic electrical properties of a semiconductor, resistivity (and also minority carrier lifetime) are drastically out of control at dislocations and arrays of dislocations, it is easy to understand why these lattice imperfections are undesirable in crystals to be used for most semiconductor purposes.

The attainment of high perfection in germanium lattices may conveniently be discussed in two parts: first, the growth of a single crystal of high perfection and, second, the preservation of the crystal's perfection during its cooling to room temperature.

The problem of growing a single crystal in the zone leveler is basically one of arranging conditions so that the liquid germanium solidifies only

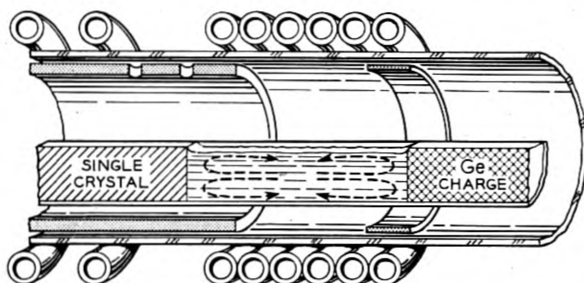


Fig. 6 — Stirring currents in liquid induced by RF induction heater.

[11] Brockmeir, K., Aluminium, **28,** p. 391, 1952.
[12] Read, W. T., Phil. Mag. **45,** p. 775, 1954.

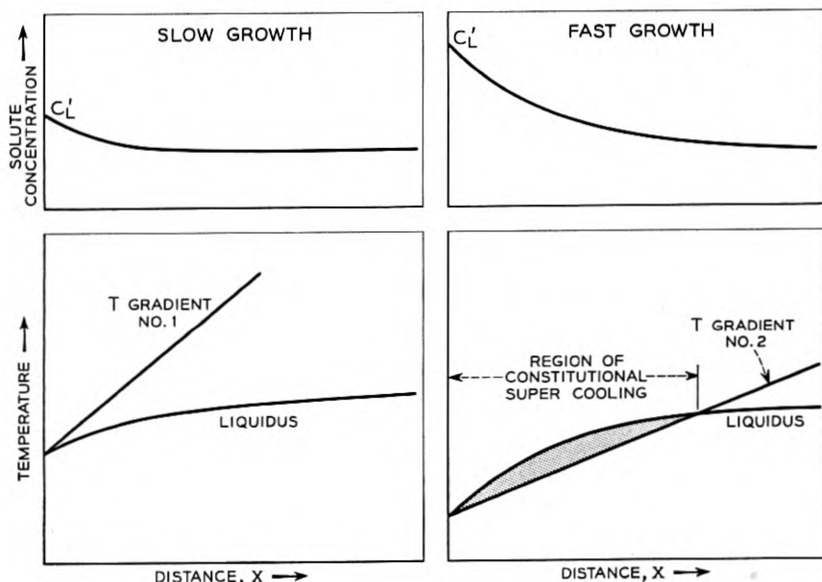


Fig. 7 — Schematic solute concentration and temperature curves in liquid, near freezing interface, illustrating constitutional supercooling. The left edge of each diagram represents the solid-liquid interface.

on the single crystal germanium seed. In order to achieve this situation, it is essential that no stable nuclei form. Thus, not only must the temperature of the liquid zone be above its freezing point everywhere except at the interface, but the liquid must also be free of foreign bodies that can ct as nuclei. Furthermore, temperature fluctuations are to be avoided.

The requirement that the liquid temperature be above its freezing point necessitates a slow growth rate because of what has been termed "constitutional supercooling."[13] This phenomenon can best be described with the aid of Fig. 7. The freezing point of a liquid is depressed by increasing concentration of solutes having k's less than unity. Because of the rise in C_L near the solidifying interface, the freezing point is more depressed in this region than that in the bulk of the liquid zone as shown in Fig. 7.

It has also been shown[14] for crystals growing in one dimension that the temperature gradient in the liquid decreases for increasing growth rates. The temperature gradients for two growth rates are plotted on Fig. 7. It can be seen that where the growth rate is slow and the temperature

[13] Chalmers, B., J. Metals, **6,** No. 5, Section 1, May, 1954.
[14] Burton, J. A., and Slichter, W. P., private communication.

gradient is steep, the temperature of the liquid is above its liquidus (freezing point curve) throughout the liquid, and no stable nuclei can form. However, increasing the growth rate decreases the temperature gradient, while it depresses the liquidus. If the temperature gradient is reduced to that indicated for fast growth, a region of constitutional supercooling will exist in front of the solidifying interface where nuclei can form and grow. The freezing of such a crystallite onto the growing crystal marks the end of single crystal growth.

A foreign body may also initiate polycrystalline growth. A natural site for nucleation by foreign bodies is the wall of the boat, close to the growth interface. Here the liquid germanium is in contact with foreign matter at temperatures approaching its freezing point. It was found by D. Dorsi that germanium single crystals could be grown satisfactorily in a smoked quartz boat, at growth rates up to 2 mils per second. However, uniformity considerations mentioned previously make it desirable to zone level at much slower rates.

It is believed that scattered dislocations may be produced in a single crystal germanium lattice by three chief mechanisms. They may be propagated from a seed into the new lattice as it grows; they may result from various possible growth faults; but probably the most important mechanism in this work is plastic deformation of the solid crystal. The first cause may be minimized by selecting the most nearly perfect seeds available, the second by using slow growth rates, and the third by minimizing stresses in the crystal.

The first hint that plastic deformation in the crystal might be an important source of dislocations came from the study of crystals pulled from the melt by the Teal-Little technique. Frequently when sections of crystals grown in the [111] direction were etched in CP_4 the pits were arrayed in a star pattern, Fig. 8(a), in which the pits appeared on lines — not randomly distributed. This coherent pattern suggested strongly that the lines were caused by dislocations in slip planes which had been active in plastic deformation of the crystal. The slip system of germanium has been determined to be the $<110>$ directions on $\{111\}$ planes.[15] If the periphery of the crystal is assumed to be in tension, it is possible to calculate the relative shear stress pattern in each slip system of the 3 $\{111\}$ planes which intersect the (111) section plane. The results of these calculations are summarized in Fig. 8(b) which shows a polar plot of the largest resolved shear stresses for these planes and also their traces in the section plane. The agreement with the observed star pattern is striking.

---

[15] Treuting, R. G. Journal of Metals, **7**, p. 1027, Sept., 1955.

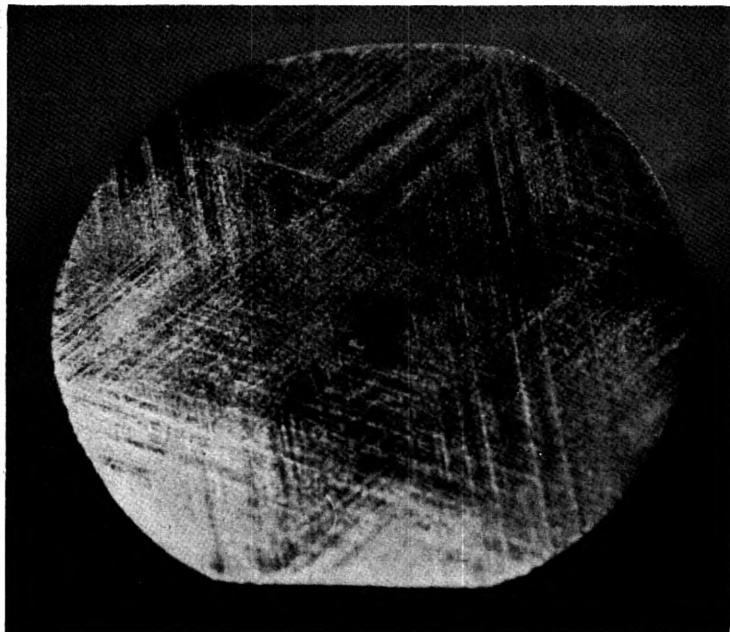


Fig. 8(a) — Star Pattern on (111) plane (etched cross-section of crystal pulled from melt).

The peripheral tension assumed in the above paragraph may be seen to be qualitatively reasonable upon consideration of the heat flow pattern of the crystal during growth. Heat must enter the crystal by conduction through its hottest surface, the growing interface, which is a 940°C isotherm. It must leave through all the other surfaces by radiation and conduction. Therefore, these surfaces must be cooler than their adjacent interiors, and cross-sections of the crystal must have cooler peripheries than cores because of the heat escaping from the peripheral surfaces. Due to thermal contraction the cooler periphery must be in tension and the core in compression.

In zone leveled crystals the distribution of etch pits on a (111) section was not dense or symmetric enough to display a star pattern. However, it was reasoned that since thermal contraction stresses appeared to play a major role in the production of dislocations in pulled crystals through plastic deformation in the available slip systems, the same mechanism might be playing a significant role in zone leveled crystals.

The only stresses in a zone leveled ingot other than those due to the weight of the crystal itself must be those due to non-uniformities in

thermal contraction. Consider a small increment of the length of a newly formed zone leveled crystal as heat flows through it from its hotter to its colder ends while the crystal moves slowly through the apparatus. Heat flows in by conduction from the higher temperature germanium adjacent to it. Heat leaves not only by conduction out the other end, but also by conduction and radiation from the ingot surface. Because of this latter heat loss, there is a radial component as well as a longitudinal component to the temperature gradient. The cooler surface contracts resulting, as above, in peripheral tension and internal compression. Clearly if the radial component of heat flow could be eliminated, there would be no peripheral contraction. Accordingly, the most desirable temperature distribution is one whose radial heat flow is zero, i.e., a case of purely axial or one dimensional heat flow, which implies a uniform temperature gradient along the axis of the ingot. In practice, it is difficult to obtain a uniform axial temperature gradient except for the special case of a very small one. This may be obtained fairly easily by the use of an ap-

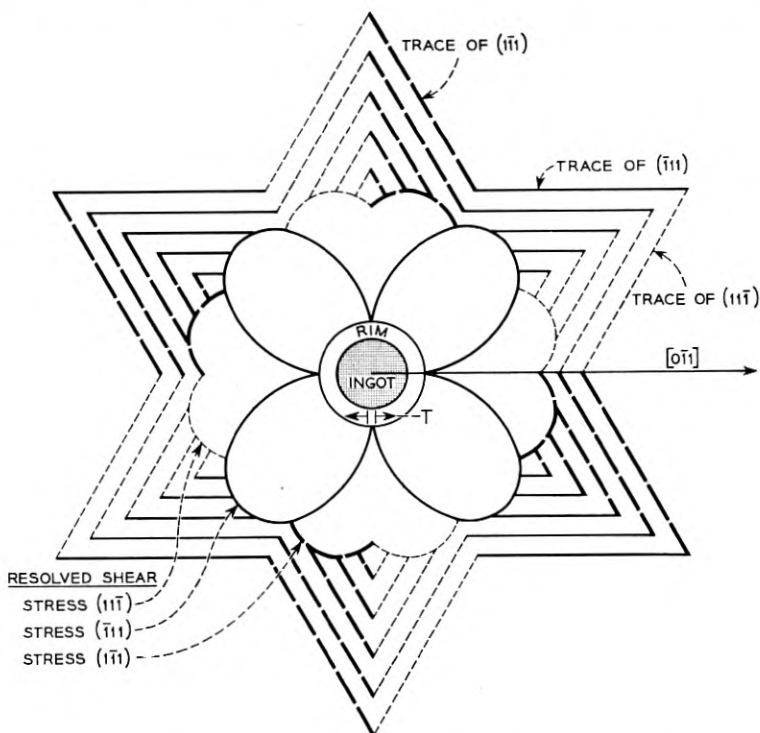


Fig. 8(b) — Resolved shear stress and slip-plane traces on (111) Plane.

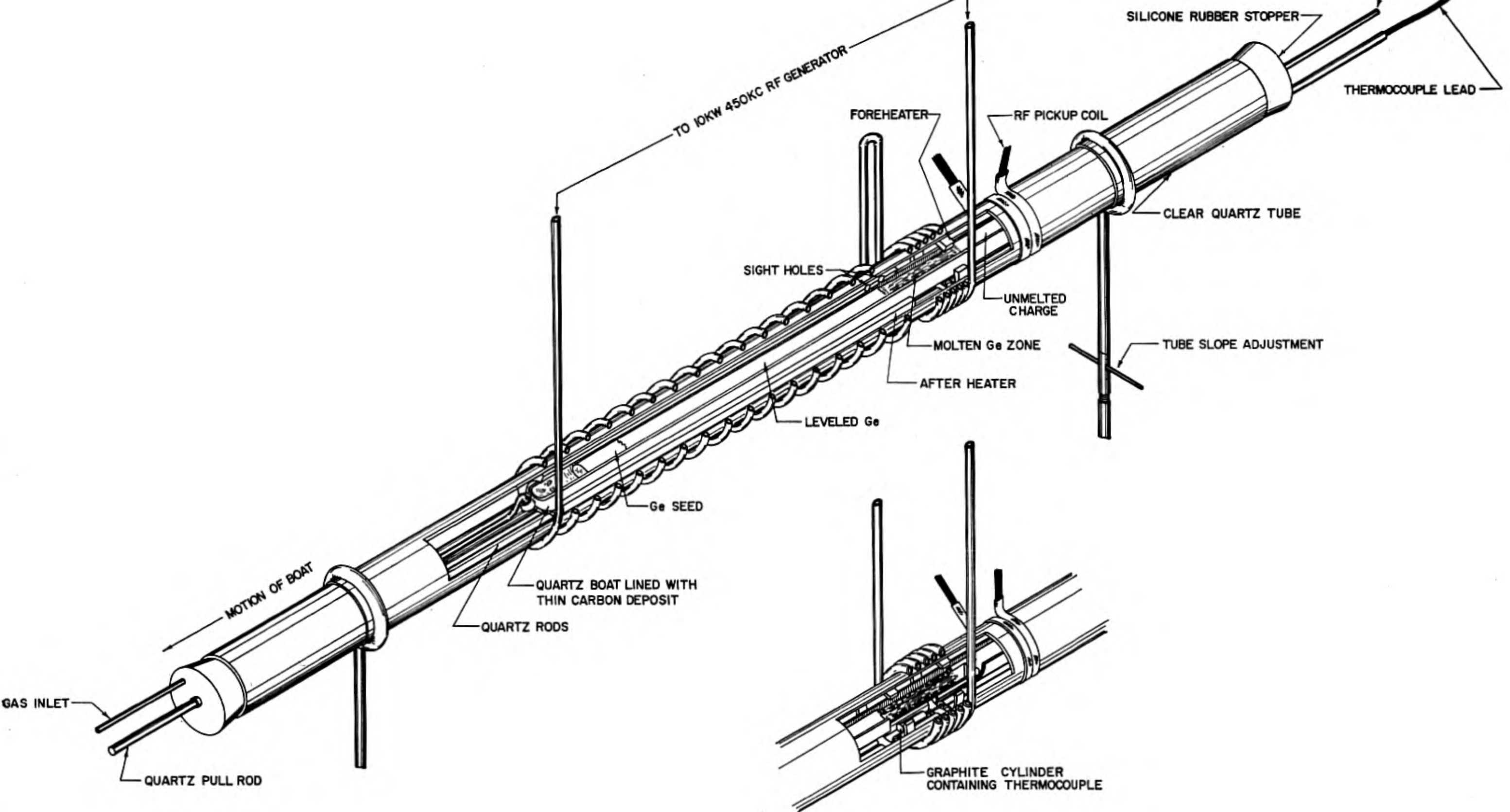


Fig. 9 — Two views of a zone leveler. (a) With after-heater. (b) Without after-heater.

propriate heater. The heater designed for this purpose is called an after-heater and is shown in Figs. 4 and 9.

The after-heater reduces the heat loss by radiation and radial conduction from the crystal maintaining the entire crystal at a temperature only slightly below its melting point throughout its growth. After zone leveling has been completed, the entire ingot is cooled slowly and uniformly. Of course, a finite temperature gradient must exist at the liquid-solid interface. The gradient at the interface of the leveler shown in Figs. 4 and 9 is about 10°C per centimeter and the maximum gradient, about ¼ inch into the solid, is 30°C per centimeter. The gradient decreases slowly to nearly zero within the after-heater, as can be seen in the measured temperature curve of Fig. 4.

A ZONE LEVELING APPARATUS AND TECHNIQUE FOR GERMANUIM

The apparatus required for zone leveling is basically simple. A single crystal seed, the desired impurities, and a germanium charge, are held in a suitable container in an inert atmosphere. Provision is supplied for either moving a heater along the charge or the charge container through a heater. The heater may be either an electric resistance type or a radio frequency induction type. The resistance heater offers the advantage of economy while the induction heating offers the advantage of direct inductive stirring of the melted zone by the RF field, which, as mentioned previously, is helpful in attaining uniformity of impurity distribution, and is therefore to be preferred for critical work.

Schematic drawings of an RF powered zone leveler following in general the original design by K. M. Olsen are shown in Fig. 9 in two useful configurations. The outer clear quartz tube serves to support the inner members of the apparatus and also to contain the inert atmosphere for which nitrogen, hydrogen, helium, or argon, can serve. For this apparatus, a quartz boat is used to contain the germanium, since it permits inductive stirring of the liquid germanium by the RF field. The auxiliary fore and after heaters, which are made of graphite, have special purposes discussed in the two preceding sections. A typical boat used in this apparatus is about 16″ long, is smoked on the inside, and is made of thin-walled clear quartz of 1″ I.D. and of semi-circular cross-section. A normal charge of zone refined Ge and seed is about 12 inches long and weighs about 500 gm. A photograph of the assembled apparatus appears in Fig. 10.

For the best results in crystal perfection and resistivity uniformity, the apparatus is run with the full length after-heater and at a slow pull

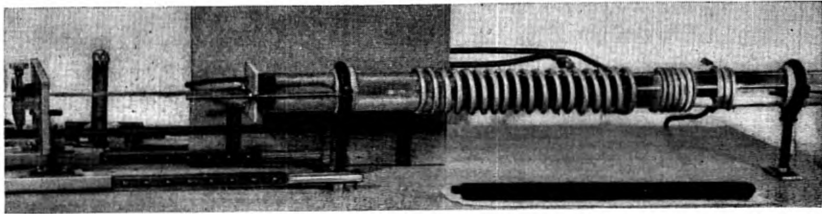


Fig. 10 — Zone leveler.

rate, 0.09 mils per second (approximately 1″ in three hours). For somewhat less critical demands a pull rate 10 times faster is used, with a shortened after-heater or none at all.

If it is desired to reproduce a resistivity obtained in the zone leveler, it is very convenient to reuse the solidified zone containing the impurity addition that yielded the desired resistivity. This solid zone, if undamaged (when cut from the finished ingot), will contain all of the solute that was not deposited during the ingot run. When it is remelted next to a seed the solute will redissolve into the liquid to yield very nearly

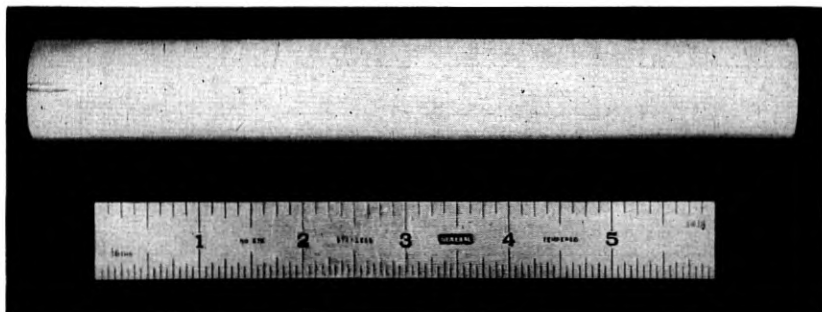


Fig. 11. — Photograph of zone leveled single crystal ingot.

the same C_L, provided that the zone volume is accurately reproduced. In this way it is readily possible to resume leveling as before and hence virtually to reproduce a desired resistivity. For the small k solutes, *In* and *Sb*, discussed in this paper the loss of C_L in one leveling run is so small as to be insignificant compared to other sources of error in this quantity.

PILOT PRODUCTION RESULTS

The capabilities of the zone leveling equipment and techniques just described may be evaluated with reasonably good accuracy on the basis of the measurement results obtained from more than 300 single crystal ingots so produced. Over 200 of these crystals were grown in the after-heater at the "slow" growth rate of 0.09 mils per second. The rest were grown with a short after-heater or none at all at a growth rate about ten times greater.

The ingots to be measured (see Fig. 11) were usually 4–6 inches long after removing seeds and solidified zones (i.e., 2–3 zone lengths), and were cut into 1 inch lengths. The ρ, τ, and ϵ measurements were taken on the flat ends of these segments. The results of the observations will be summarized and discussed in terms of the four device test requirements described earlier.

(1) *Compositional Uniformity*

The resistivity measurements were taken with a calibrated 4-point probe technique[16] at five locations on each ingot cross-section (center, top, bottom and each side). The spacing between adjacent points of the probe was 50 mils. Accordingly, these measurements would be insensitive to ρ fluctuations in the material of this order or smaller. However, an investigation by potential probing techniques, of Ge filaments cut from zone leveled ingots[17] indicates that ρ fluctuations in zone leveled material are

[16] L. B. Valdes, Proc. I.R.E., **42**, p. 420, 1954.
[17] Erhart, D. L., private communication.

TABLE I — AVERAGE RESISTIVITY VARIATIONS

(A) Along length axis. Grand Length Average ± 10%.

Growth Rate Mils per Second	n-Type		p-Type		Average ± %
	± %	No. of Ingots	± %	No. of Ingots	
0.9	9.9	27	10.9	33	10.4
0.8	7.6	12	17.4	16	13.2
0.09	9.0	108	9.3	137	9.2

(B) Over Cross-Section

Growth Rate Mils per Second	n-Type		p-Type		Average ± %
	± %	No. of Ingots	± %	No. of Ingots	
0.9	9.5	22	8.5	30	8.9
0.8	8.3	12	6.9	14	7.5
0.09	4.3	93	2.3	122	3.2

generally coarse — changing over distances 2 to 5 times larger in dimension than the 50 mil dimension in question. Thus the ρ data summarized here should give a reasonably valid representation of the true ρ variations in the ingots measured.

Table I summarizes the resistivity variations recorded as percentages of the mean resistivity of each ingot. These variations are separated into those observed (a) along the length axis and (b) over the cross-section, for the different growth conditions and resistivity types.

It is readily seen that the average variation along the length, about ±10 per cent, is larger than the average cross-sectional variation. The variations are not systematic along the length of the ingot and are chiefly due to fluctuation in the length of the liquid zone. An appreciable part of this variation is due to the effect, mentioned earlier, of discontinuities in the unmelted charges between 1 inch lengths of crystals that were being releveled. A smaller length variation of ρ, about ±7 per cent, was observed in those ingots grown from continuous charges.

Part B of the table shows that the variation of ρ over the cross-section is sensitive to the growth rate in the range covered. For slow growth, it is small, and one would reasonably expect that if further improvement in ρ variation were required, it should first be sought by improving the control of the zone length.

(2) *Macro Perfection*

Macro perfection of the pilot production product is extremely high. There were essentially no cases of polycrystallinity, or twinning, except

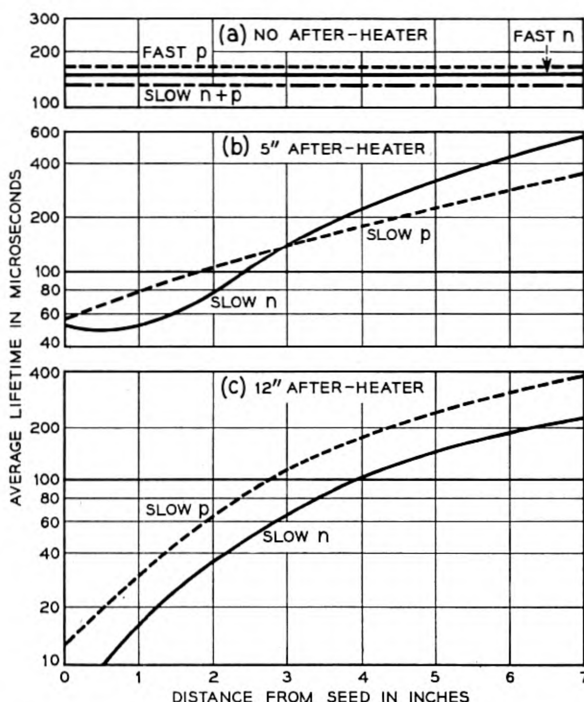


Fig. 12 — Average minority carrier lifetime plotted against distance from seed for 2–8 ohm cm crystals grown with 12″, 5″ and no after-heaters.

for clearly attributable causes such as power or equipment failure. There were few cases of lineage in the short after-heater and virtually none in the full after-heater, while lineage is not uncommon in ingots grown with no after-heater.

(3) *Micro Perfection*

Table II summarizes the etch pit density, ϵ, measurement results. In general, it can be seen that with the after-heater one can expect etch pit counts of the order of 1,500 pits per cm^2 which is lower than results without an after-heater by about an order of magnitude (and lower than

TABLE II — AVERAGE ETCH PIT DENSITIES, ϵ

	Growth Rate Mils per Second	ϵ Ave	σ	No. of Ingots
(12″ after-heater)	0.09	1560	770	39
(5″ after-heater)	0.09	3800	1600	3
	0.9	7000	1900	3
No after-heater	0.9	11000	6600	6

ϵ's of pulled Ge crystals by about two orders of magnitude). The lowest average count that has been observed is 40 pits per cm^2. This crystal was found to have the smallest X-Ray rocking-curve widths observed in germanium at Bell Telephone Laboratories — very nearly the theoretically ideal widths. The perfection indicated is exceptional — comparable to that of selected quartz crystals.

(4) *Lifetime of Minority Carriers*

τ data are summarized in Fig. 12 in which are plotted averages of the τ measurements on the ingot sections against distance from the seed. One sees a systematic rise in τ along the length axis of an ingot grown slowly in the after-heater. This is interpreted to indicate that the ingot is being slowly contaminated with chemical recombination centers during its long wait inside the after-heater at high temperatures. If improvement were needed in lifetime, it should be sought first by increasing the chemical cleanliness precautions, which were nonetheless strict in this work.

SUMMARY

A zone leveler has been developed to provide growth conditions suitable for the production of quality germanium single crystals. The crystals are nearly uniform and have exceptionally high lattice perfection. Similar levelers are in use in production.

The apparatus developed has been used to supply germanium single crystals for experiments and for the pilot production of a variety of point contact, alloy, and diffusion transistors. The machine operating at slow growth rate with an after-heater can produce one 6-inch 250-gm crystal per day. For less critical demands, it can produce several longer crystals per day.

Evaluation of the product indicates that resistivity variation on a cross-section of the ingot can be ± 3 per cent and that along the length axis it can be controlled to ± 7 per cent if a continuous charge is used. Furthermore, the crystals contain no grain boundaries or lineage and the scattered etch pit densities average about 1,500 per cm^2. Thus, the zone leveling process has proved to be simple, efficient, and capable of more than meeting the present specifications for quality germanium single crystals.

ACKNOWLEDGMENTS

The authors are indebted for the help and cooperation of many people, especially that of L. P. Adda and D. L. Erhart who guided the evaluation of zone leveled material summarized above, and that of F. W. Bergwall through whose patient effort and suggestions the machine worked.

Diffused p-n Junction Silicon Rectifiers

By M. B. PRINCE

(Manuscript received December 12, 1955)

Diffused p-n junction silicon rectifiers incorporating the feature of conductivity modulation are being developed. These rectifiers are made by the diffusion of impurities into thin wafers of high-resistivity silicon. Three development models with attractive electrical characteristics are described which have current ratings from 0 to 100 amperes with inverse peak voltages greater than 200 volts. These devices are attractive from an engineering standpoint since their behavior is predictable, one process permits the fabrication of an entire class of rectifiers, and large enough elements can be processed so that power dissipation is limited only by the packaging and mounting of the unit.

1.0 INTRODUCTION

1.1 The earliest solid state power rectifier, the copper oxide rectifier, was introduced in the 1920's. It found some applications where efficiency, space, and weight requirements were not important. In 1940 the selenium rectifier was introduced commercially and overcame to a great extent the limitations of the copper oxide rectifier. As a result, the selenium rectifier has found wide usage. In early 1952 a large area germanium[1] junction diode was announced which showed further improvements in efficiency, size, and weight. In addition it shows promise of greater reliability and life as compared to the earlier devices. However, all of these devices have one drawback in that they cannot operate in ambient temperatures greater than about 100°C.

Also in 1952, the silicon alloy[2] junction diode was announced and was shown to be capable of operating at temperatures over 200°C. However it was a small area device and could not handle the large power that the other devices could rectify. During the past three years development has been carried on by several laboratories in improving the size and power capabilities of these alloy diodes. In early 1954 the gaseous diffu-

[1] Hall, R. N., Proc. I.R.E., **40,** p. 1512, 1952.
[2] Pearson, G. L., and Sawyer, B., Proc. I.R.E., **40,** p. 1348, 1952.

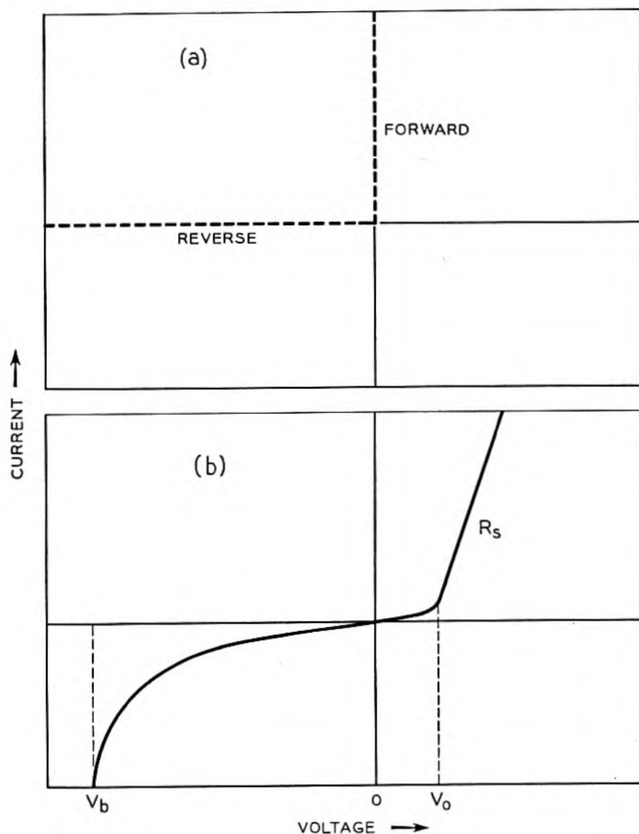


Fig. 1 — (a). Ideal rectifier. (b). Semiconductor rectifier.

sion technique[3] for producing large area junctions in silicon was announced. This technique lends itself very readily to controlling the position of junctions in silicon. An early rectifier[3] made by this technique was one half cm^2 in area and conducted 8 amperes at one volt in the forward direction and about 2 milliamperes at 80 volts in the reverse direction. The series resistance of this device was approximately 0.07 ohms.

1.2 In order to understand quantitatively the problems associated with power rectifier development, consider Fig. 1(a) which shows what an engineer would like in the way of an ideal rectifier. It will pass a large amount of current in the forward direction without any voltage

[3] Pearson, G. L., and Fuller, C. S., Proc. I.R.E., **42,** No. 4., 1954.

drop and will pass no current for any applied voltage in the reverse direction. At present no device with this characteristic exists. A typical semiconductor rectifier has a characteristic of the type shown in Fig. 1(b). In these devices there is a forward voltage, V_0, that must be developed before appreciable current will flow and a series resistance, R_s, thru which the current will flow. In the reverse biased direction there is a current that will flow due to body and surface leakage and that usually increases with reverse voltage. At some given reverse voltage, V_B, the device will break down and conduct appreciable currents. To have an efficient rectifier, V_0 and R_s should be as small as possible and V_B should be as large as can be made; also, the reverse leakage currents should be kept to a minimum. According to semiconductor theory, V_0 depends mainly upon the energy gap of the semiconductor, increasing with increasing energy gap. R_s consists of two parts; body resistance of the semiconductor and resistance due to the contacts to the semiconductor. The higher the resistivity of the semiconductor, the higher is the body resistance part of R_s. The leakage currents in the reverse direction depend to some extent on the energy gap of the semiconductor, being smaller with larger energy gap; and V_B depends most strongly on the resistivity of the semiconductor, being larger for higher resistivity material. Another factor that is important in the choice of the semiconductor is the ability of devices fabricated from the semiconductor to operate at high temperatures; high temperature operation of devices improves with larger energy gap semiconductors. Thus there are two compromises to be made in choosing the material (energy gap) and resistivity of the semiconductor.

1.3 This paper reports on a special class of rectifiers in which improved performance has been obtained. These devices are made by using the diffusion technique with silicon. The diffusion process permits both accurate geometric control and low resistance ohmic contacts, which in turn makes it possible to reduce R_s to very small values independent of the resistivity of the initial silicon. Therefore, high resistivity material can be used to obtain high V_B. An explanation of this result is given in Section 3. Silicon permits small reverse currents and high temperature operation. Its only drawback is that $V_0 \simeq 0.6$ volts. Rectifiers made of silicon with the diffusion technique are able to pass hundreds of amperes per square centimeter continuously in the forward direction in areas up to 0.[illegible] square centimeter. One type of device whose area is 0.06 cm^2 readily conducts ten amperes with less than one volt forward drop. The forward current voltage characteristic of this family of rectifiers follows an almost exponential characteristic indicating that

R_s is extremely small (<0.05 ohms). Although the measured reverse currents are greater than those predicted by theory for temperatures up to 100°C, the reverse losses are low and do not affect the efficiency appreciably.

1.4 The diodes made by the diffusion of silicon are very attractive from an engineering standpoint for several reasons. First of all, their behavior is predictable from the theory of semiconductor devices, as are junction transistors. This makes it possible to design rectifiers of given electrical, thermal, and mechanical characteristics. Secondly, rectifier elements of many sizes are available from the same diffused wafers making it possible to use the same diffusion process, material, and equipment for a range of devices. Thirdly, large enough elements can be processed so that the power dissipation in the unit is limited only by the thermal impedance of mount and package.

2.0 DIFFUSION PROCESS

2.1 It will be shown in 3.2 that the forward characteristic of these devices is practically independent of the type (n or p) and resistivity of the starting material. The reverse breakdown voltage of a silicon p-n junction depends primarily on the resistivity of the lightly doped region. With these two considerations in mind; that is, to fabricate rectifiers having the desirable excellent forward characteristic and at the same time high reverse breakdown voltage, high resistivity silicon is used as the starting material for the diffused barrier silicon rectifiers. Single crystal material has been found to give a better reverse characteristic than multicrystalline material. Also, it has been found that p-type material has yielded units with a better reverse characteristic than n-type material. Therefore, in the remainder of this paper, we will limit discussion to rectifiers made from high resistivity, single crystalline, p-type silicon. We will designate this material as π type silicon.

2.2 In addition to the fine control one has in the diffusion process (see 2.4), the process lends itself admirably to the semiconductor rectifier field in as much as the distribution of impurities in this process results in a gradual transition from a degenerate semiconductor at the surface of the material to a non-degenerate semiconductor a short distance below the surface. This condition permits low resistance ohmic metallic contacts to be made to the surfaces of the diffused silicon.

In order to create a p-n junction in the π silicon, it is necessary to diffuse donor impurities into one side of the slice. Although several donor type impurities have been diffused into silicon, all the devices discussed

in this paper were fabricated by using phosphorus as the donor impurity. In order to make the extremely low resistance contact to the π side of the junction that is desirable in rectifiers, acceptor impurities are diffused into the opposite side of the π silicon slice. Boron was selected from the several possible acceptor type impurities to use for the fabrication of these devices. A configuration of the diffused slice is shown in Figure 2.

2.3 It will be shown in Section 3 that there are limits to the thicknesses of the three regions, $N+$, π, $P+$, due to the nature of the operation of these rectifiers. With present techniques, it is necessary to keep

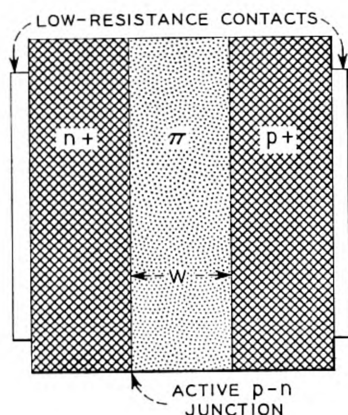


Fig. 2 — Diffused silicon rectifier configuration.

the thickness of the π region to the order of two or three mils (thousandths of an inch).

2.4 In the diffusion process of introducing impurities in silicon for the purpose of creating junctions or ohmic contacts, the diffusant is deposited on the silicon and serves as an infinite source. The resulting concentration of the diffusant is given by

$$C = C_0 \left[1 - \frac{2}{\sqrt{\pi}} \int_0^{x/\sqrt{4Dt}} e^{-y^2}\, dy \right] \qquad (1)$$

$$= C_0 \operatorname{erfc} y$$

where C = concentration at distance x below surface
C_0 = concentration at surface
D = diffusion constant for impurity at temperature of diffusion

t = total time of diffusion

$$y = \frac{x}{\sqrt{4Dt}} = \text{variable of integration}$$

A plot of $C/C_0 = \text{erfc}\, y$ versus y is given in Fig. 3. C_0 is the surface solubility density and depends upon the temperature of the diffusion process.[4] At some depth, x_j, the concentration C equals the original impurity concentration where the silicon will change conductivity type resulting in a junction. In order to obtain desirable depths of the diffused layers, $N+$ and $P+$, it is necessary to diffuse at temperatures in the range of 1000°C to 1300°C for periods of hours. With such periods it is obvious that the diffusion process lends itself to easy control and reproducibility.

3.0 CONDUCTIVITY MODULATION

3.1 It is well known that the series resistance of a power rectifier is the most important electrical parameter to control and should be made as small as possible for several reasons. The series resistance consists essentially of two parts; the body resistance of the semiconductor and the contact resistance to the semiconductor. In the early stages of rectifier development both parts of the series resistance contributed about equally to the total series resistance. However, methods were soon found to reduce the contact resistance. It then became apparent that in order to reduce the body resistance, the geometry would have to be changed and the resistivity chosen carefully. By going to larger, thinner wafers it was possible to reduce this body resistance. However, the cost of pure silicon made it important that conductivity modulation (described below) be incorporated in these devices as a method for reducing the body resistance. Our initial attempts were successful due to the fact that higher lifetime of minority carriers could be maintained in the extremely thin wafers that were used as compared to the lifetime remaining after the diffusion process in thicker wafers.

3.2 A complete mathematical description of the I–V characteristic for the conductivity modulated rectifier is practically impossible due to the fact that the equations are transcendental. However, it is easy to understand the operation of the device physically.

When the device is biased in the forward direction, electrons from the heavily doped $N+$ region are injected into the high resistivity π region. If the lifetime for these electrons in the π region is long enough, the electrons will diffuse across the π region and reach the $P+$ region

[4] Fuller, C. S., and Ditzenberger, J. A., J. Appl. Phys., **25**, p. 1439, 1954.

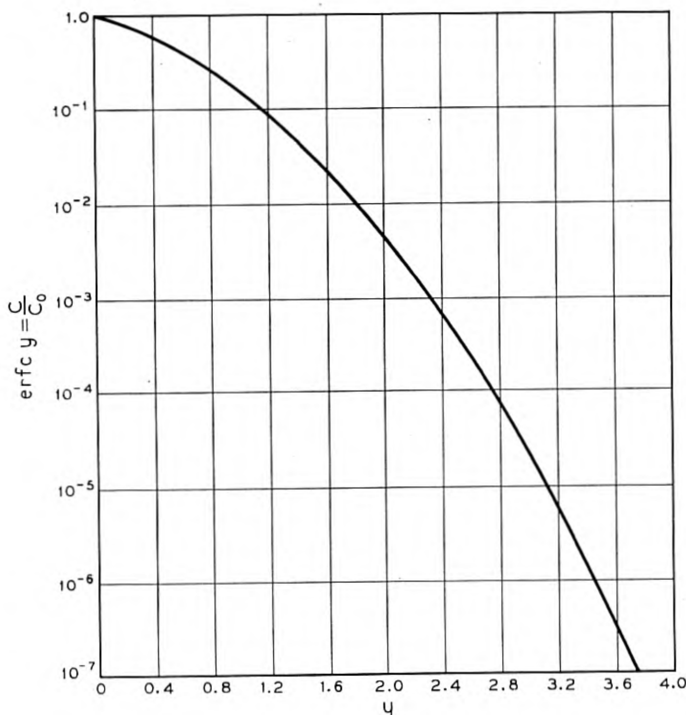


Fig. 3. — Error function complement.

with little recombination. To maintain electrical neutrality, holes are injected into the π region from the $P+$ region. These extra mobile carriers (both electrons and holes) reduce the effective resistance of the π layer and thus decrease the voltage drop across this layer. The higher the current density, the higher is the injected mobile carrier densities and therefore, the lower is the effective resistance. It is for this reason that the process is termed conductivity modulation. This effect tends to make the voltage drop across the π region almost independent of the current, resistivity, and semiconductor type.

When the junction is biased in the reverse direction, a normal reverse characteristic with an avalanche breakdown is expected and observed.

3.3 The forward characteristic of a typical unit is plotted semilogarithmically in Fig. 4. The best fit to the low current data can be

expressed as

$$I = I_0 e^{qV/NkT} \tag{2}$$

where I = current thru unit
I_0 = constant
q = charge of electron
V = voltage across unit
k = Boltzmann's constant
T = absolute temperature
and $1 < N < 2$.

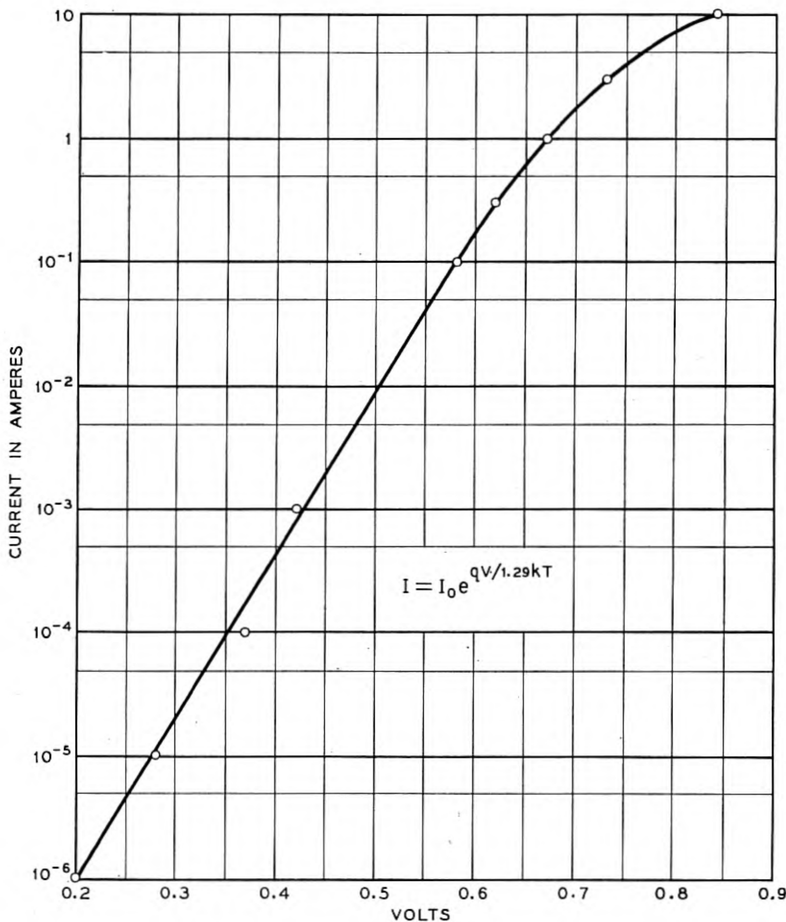


Fig. 4 — Forward characteristic of silicon power rectifier.

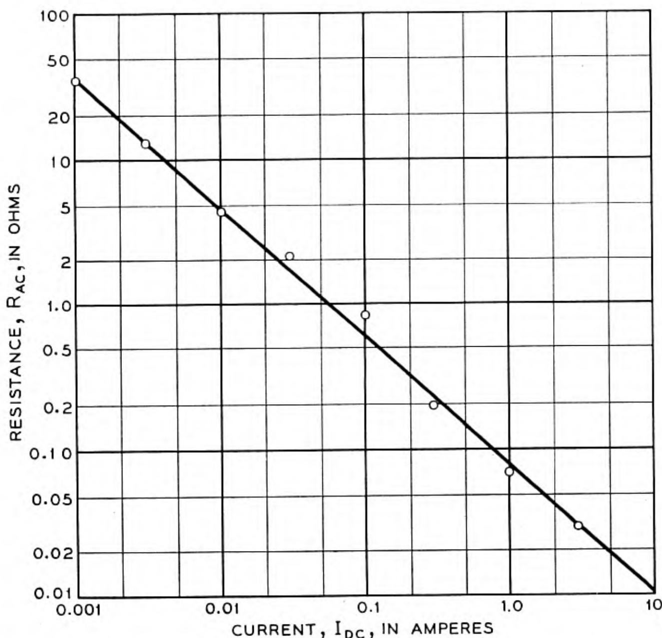


Fig. 5 — Small signal resistance versus dc forward current.

The departure of the high current data from the exponential characteristic is due to the contact resistance. Another interesting measurement of the forward characteristic is given in Fig. 5 where the small signal ac resistance is plotted as a function of the forward dc current for a typical rectifier element. The departure from the simple rectifier theory[5] where $N = 1$ is not surprising inasmuch as p-n junctions made by various methods and of different materials almost always have $N > 1$. Several calculations have been carried out using different assumptions and all indicate that the forward characteristic is independent of the type and resistivity of the middle region as long as the diffusion length for minority carriers is the order of or larger than the thickness of the region.

3.4 In order to go to higher reverse breakdown voltages (>500 volts) it is necessary to use still higher resistivity starting material. It might be expected that intrinsic silicon will be used for the highest reverse breakdown voltages when it becomes available. However, in this case

[5] Shockley, W., B.S.T.J., **28**, p. 435, 1949.

thick wafers are necessary since the reverse biased junction space charge region extends rapidly with voltage for almost intrinsic material, and high lifetime is necessary in order to get the conductivity modulation effect in these thick wafers. Therefore at present it is necessary to compromise the highest reverse breakdown voltages with the lowest forward voltage drops, in a similar manner to that discussed in Section 1. However this is now done at a different order of magnitude of voltage and current density.

4.0 FABRICATION OF MODELS

4.1 It has been pointed out in Section 1.2 that a low series resistance, R_s, is desirable and that it is composed of two parts; the body resistance and the contact resistance. In Section 3 a method for reducing the body resistance was described. The contact resistance can also be made very low. It has been found to be very difficult to solder low temperature solders (M.P. up to 325°C) to silicon with any of the standard commercial fluxes. However, it is quite easy to plate various metals to a surface of silicon from an electroplating bath or by an electro-less process[6] to which leads can readily be soldered. Some metals used for plating contacts are rhodium, gold, copper, and nickel. This type of contact yields a low contact resistance. Another technique that has shown some promise for making the necessary extremely low resistance contact is the hydride fluxing method.[7]

4.2 A wafer which may be about one inch in diameter is ready to be diced after it is prepared for a soldering operation. Up to this point all the material may undergo the same processing. Now it is necessary to decide how the prepared material is to be used; whether low current (~1 amp) devices or medium or high current (~10–50 amps) devices are desired. The common treatment of all material for the entire class of rectifiers is one reason these devices are highly attractive from a manufacturing point of view.

The dicing process may be one of several techniques; mechanical cutting with a saw, breaking along preferred directions, etching along given paths with chemical or electrical means after suitable masking methods, etc. In the case of mechanical damage to the exposed junctions, the dice should be etched to remove the damaged material. The dice are cleaned by rinses in suitable solvents and are then ready for

[6] Brenner, A., and Riddell, Grace E. J., Proc. American Electroplaters' Society, **33**, p. 16, 1946, **34**, p. 156, 1947.
[7] Sullivan, M. V., Hydrides as Alloying Agents on Silicon, Semiconductor Symposium of the Electrochemical Society, May 2–5, 1955.

assembly into the mechanical package designed for a given current rating.

4.3 The dice may be tested electrically before assembly by using pressure contacts to either side. Pressure contacts have been considered for packaging the units; however, this type of contact was dropped from development due to mechanical chemical, and electrical instabilities.

4.4 The drawbacks of the pressure contact make it important to find a solder contact that does not have the same objections. The solder used should have a melting point above 300°C, be soft to allow for different coefficients of expansion of the silicon and the copper connections, wet the plated metal, and finally, be chemically inactive even at the high temperature operation of the device. These requirements are met with many solders in a package that is hermetically sealed. This combination of a solder and a hermetically sealed package has been adopted for the intermediate development of the diffused silicon power rectifiers.

5.0 ELECTRICAL PERFORMANCE CHARACTERISTICS

5.1 Before describing the electrical properties of these diodes, let us consider some of the physical properties of a few members of the class.

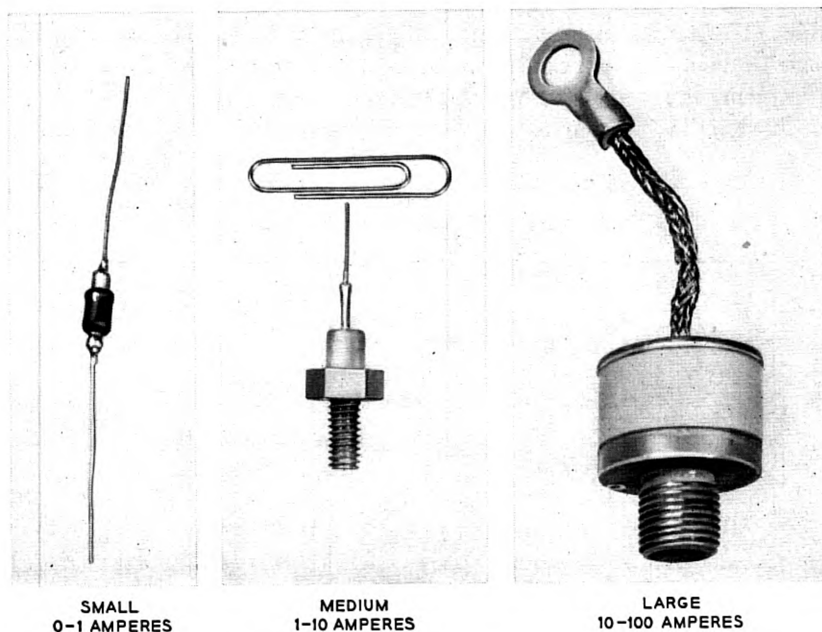


Fig. 6 — Development silicon rectifiers.

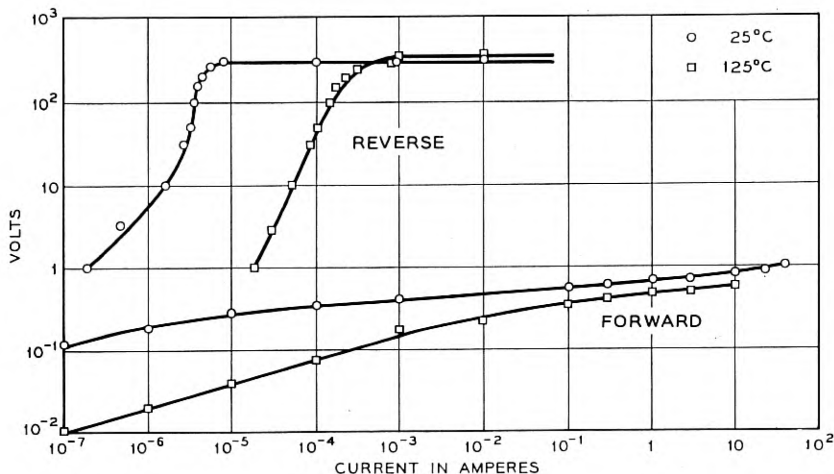


Fig. 7 — I-V characteristic of medium size rectifier.

Fig. 6 shows a picture of three sizes of units that will be discussed in this section together with the range of currents that these units can conduct. The actual current rating will depend upon the ability of the device to dispose of the heat dissipated in the unit. A description of how the rating is reached is given in Section 6.

The smallest device has a silicon die that is 0.030″ by 0.030″ in area

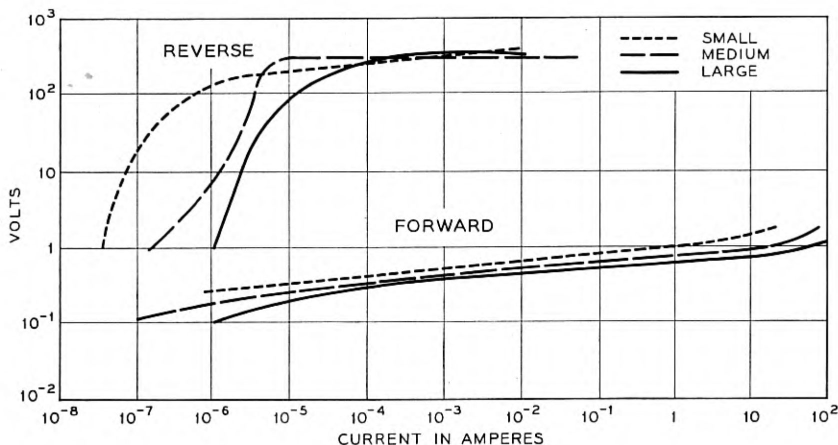


Fig. 8 — I-V characteristics of development rectifiers.

and all the units have dice about 0.005″ thick. The medium size device has a wafer 0.100″ by 0.100″ in area. The largest device has a element 0.250″ by 0.250″ in area. It is obvious that a range of die size could have been chosen for any of these rectifiers. However, electrical and thermal considerations have dictated minimum sizes and economic considerations have suggested maximum sizes. The actual sizes are intermediate in value and appear to be satisfactory for the given ratings.

5.2 Of fundamental importance to users of these rectifiers are the forward and reverse current — voltage characteristics. These characteristics of the medium size unit are shown in Fig. 7 for two temperatures, 25°C and 125°C, using logarithmic scales. It can be seen that in the forward direction at room temperature, 25°C, more than 20 amperes are conducted with a one volt drop in the rectifier. At the higher temperature more current will be conducted for a given voltage drop. In the reverse direction, this particular unit can withstand inverse voltages as high as 300 volts before conducting appreciable currents (>1 ma) even at 125°C. A comparison of the current-voltage characteristics for the three different size units is shown in Fig. 8 where again the information is plotted on logarithmic scales. This information was obtained at 25°C. One can observe that the reverse leakage current varies directly as the area of the device and the forward voltage drop varies inversely as the area. These relations are to be expected; however, the reverse characteristics indicate that surface effects are probably effecting the exact shape of the curves. The changes in the forward characteristics can be attributed to the contacts and the internal leads of the packages. The breakdown voltage can be adjusted in any size device by the proper choice of starting material and therefore no significance should be placed on the different breakdown voltages in Fig. 8.

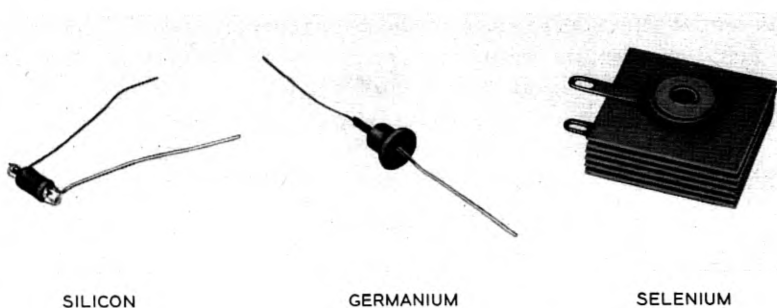


Fig. 9 — Semiconductor rectifiers of different materials.

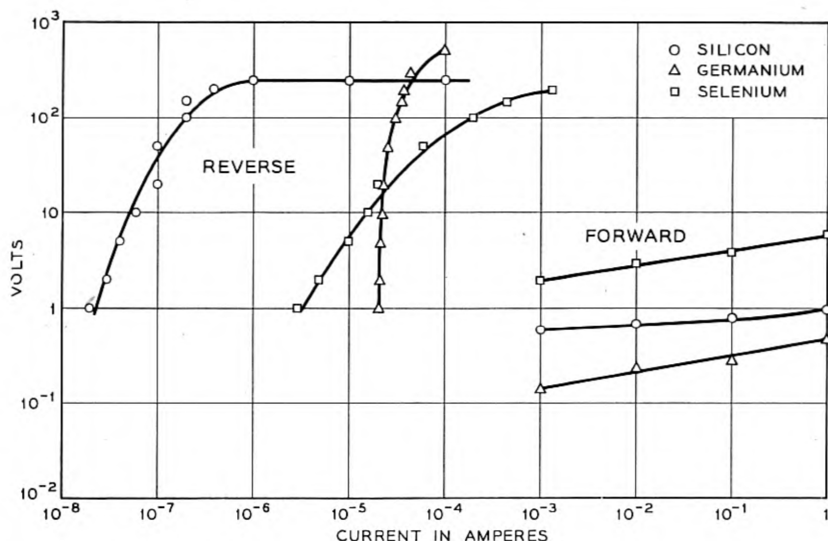


Fig. 10 — Rectifier characteristics at 25°C.

It is quite interesting to compare these units with germanium and selenium rectifiers that are commercially available. To make the comparison as realistic as one can, we have chosen to compare the smallest silicon unit with a commercially available germanium unit and a six element selenium rectifier stack rated at 100 milliamperes. The comparative size of these units can be seen in Fig. 9. Curves of the forward and reverse characteristics at 25°C are given in Fig. 10. Similar curves taken at 80°C are given in Fig. 11 and at 125°C in Fig. 12. It can be seen that the forward characteristic is best for the germanium device at all temperatures and that the reverse currents are least for the silicon rectifier. The selenium rectifier is a poor third in the forward direction. However, if one has to operate the device at 125°C, only the silicon device will be satisfactory in both the forward and reverse directions.

5.3 Capacitance measurements of all the silicon units have been made at different reverse voltages and temperatures. The temperature dependence is negligible. However, as expected in semiconductor rectifiers, the capacitance varies inversely with the voltage according to the relation VC^N = constant where $2 < N < 3$. Measurements are given in Fig. 13 for a group of medium size units. The other units made from the same resistivity material have capacitances that vary directly as their areas.

5.4 The reverse breakdown voltage, V_B, of these devices is controlled by the choice of resistivity of the starting material and the depth of diffusion of the junction. By keeping the resistivity of the initial p-type silicon above 20 ohm-cm., it is possible to keep V_B above 200 volts. Units have been made with V_B greater than 1,000 volts. The deeper diffusion causes the junction to be more "graded"[5] and therefore require a greater voltage for the breakdown characteristic. This is in line with the capacitance measurements where the exponent indicates that the junction is neither a purely abrupt junction which would result in an exponent of two nor a constant gradient junction which would result in an exponent of three.

5.5 Another interesting measurement, which is related to the lifetime of minority carriers in the high-resistivity region and the frequency response, is the recovery time of these devices. During a forward bias on a p-n junction, excess minority carriers are injected into either region. When the applied voltage polarity is reversed, these excess minority carriers flow out of these regions, giving rise initially to a large reverse current until the excess carriers are removed. The magnitude and time variation of this current will depend to some extent upon the level of the forward current but mostly upon the circuit resistance. If one adjusts the circuit resistance such that the maximum initial current in

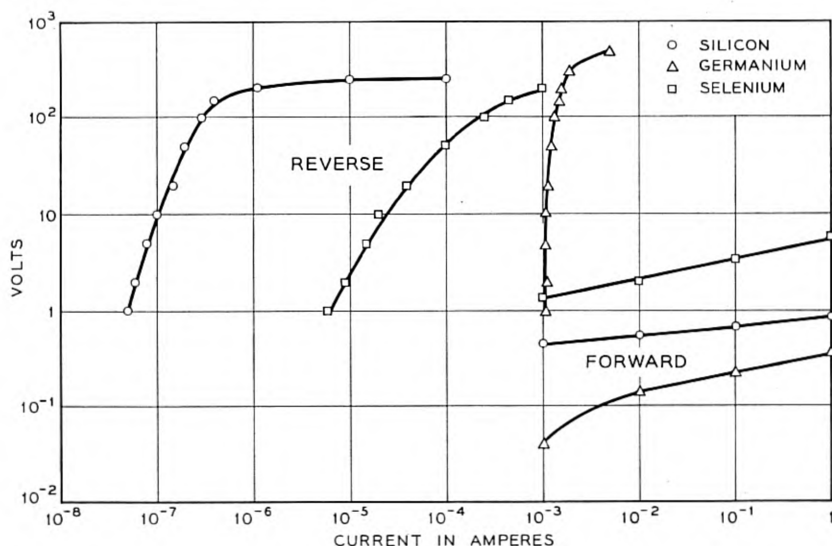


Fig. 11 — Rectifier characteristics at 80°C.

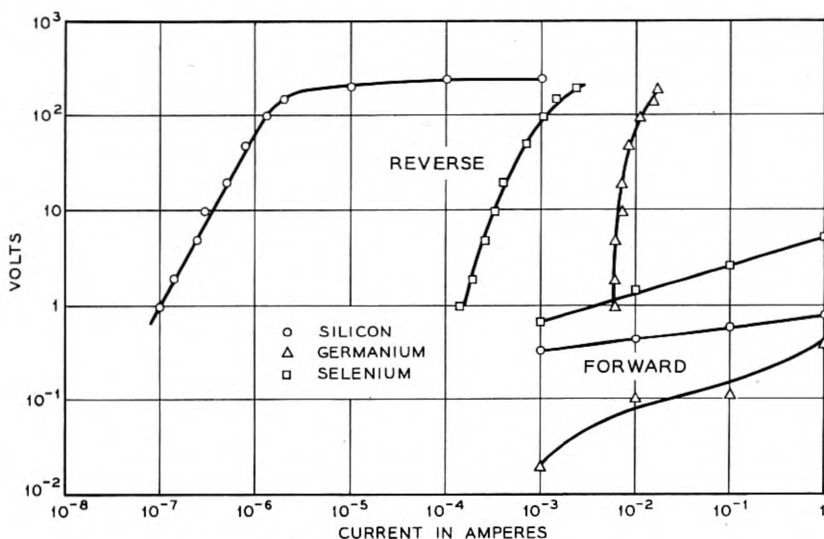


Fig. 12 — Rectifier characteristics at 125°C.

the reverse direction is equal to the forward current before reversing the polarity of the junction, then the reverse current will have a constant magnitude, limited by the circuit resistance, for a time known as the recovery time before it decays to a small steady-state value. Fig. 14 shows graphically this effect. The recovery time in diffused junctions is found to be in the range of less than 0.1 microsecond to more than 4

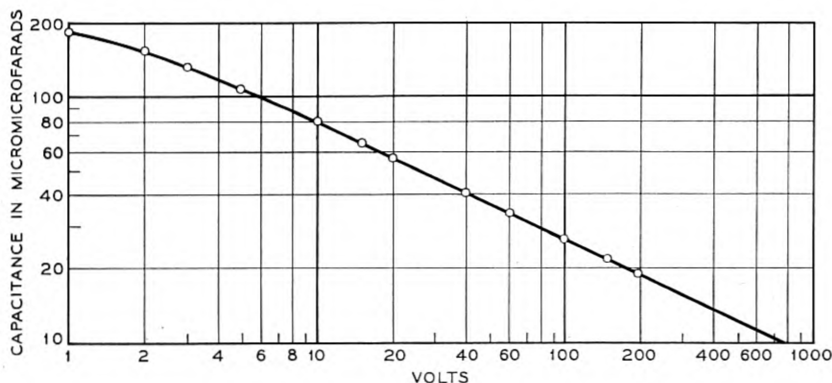


Fig. 13 — Capacitance versus reverse voltage in medium size rectifer.

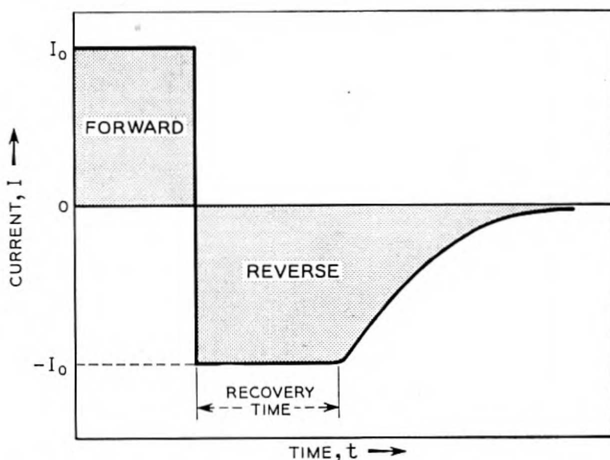


Fig. 14 — Recovery effect in silicon rectifiers.

microseconds. It can be shown that the longer recovery times are associated with higher lifetimes of minority carriers. More interesting, however, is the fact that these devices will have their excellent rectification characteristics to frequencies near the reciprocal of the recovery time. Measurements have been made of the rectification ability of typical small and medium size units by using the circuit shown in Fig. 15. The results of normalized rectified current versus frequency are given in Fig. 16 and it is seen that these units could be used to rectify power up to 1 kc/sec without any appreciable loss of efficiency.

5.6 It is interesting to note that many of the electrical measurements made with the diffused barrier silicon rectifiers are self-consistent and can be related to simple concepts of semiconductor theory. As an example, experimental measurements indicating variations of recovery time of units are related to variations in minority carrier lifetime which in turn are related to experimental variations in the forward characteristic

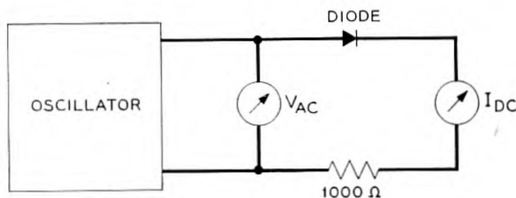


Fig. 15 — Rectification measuring circuit.

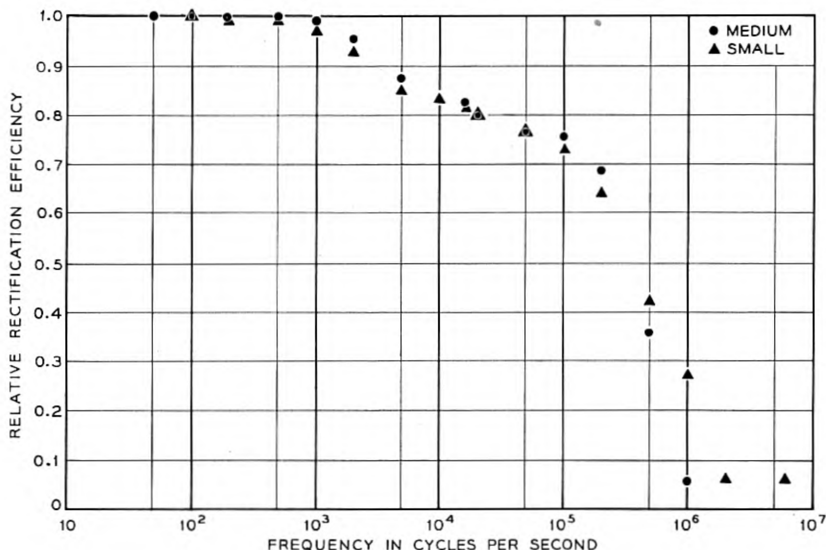


Fig. 16 — Relative rectification efficiency versus frequency.

of these same devices. Such relationships among the measurable parameters of these devices make it possible to design and control the electrical characteristics of the units and therefore make them extremely attractive from an engineering point of view.

6.0 MECHANICAL AND THERMAL DESIGN

6.1 In order to have a device that is usable for more than experimental purposes, it is necessary that it be packaged in a mechanically stable structure and that the heat generated in the combined unit should not lead to a condition where the device no longer has its desirable characteristics. In earlier sections of this paper several mechanical requirements of a satisfactory package have been suggested. These may be repeated at this point. First, pressure contacts are not satisfactory; second, oxidizing ambients are to be avoided; third, approximately one watt per ampere of forward current is generated and must be disposed; and fourth, the package must be electrically satisfactory. The first requirement is met by using soldered contacts. Since these rectifiers are usable at temperatures over 200°C, a solder was chosen that has a melting point over 300°C. The second requirement necessitated the use of a hermetic seal structure. If the seal is truly hermetic, no gases can

enter or leave the package and thus no changes of the device due to the enclosed gas should occur as long as the gas does not react with the silicon, solder or package. However, no seal is absolutely vacuum tight and thus care should be used in choosing a package design so that minimum effects should occur to the electrical properties during the use of the device. The third requirement of the disposal of the internally developed heat suggested the use of copper due to its high thermal conductivity. However, a small package alone is capable of dissipating only a small amount of heat without reaching a temperature that is too high for the device. This necessitates the use of cooling fins in conjunction with the device to make use of its electrical properties. This thermal requirement demands a package to which thermal fins can be attached. This is met by having the package contain a bolt terminal to which thermal fins can be attached or by which the unit can be mounted to a chassis for cooling. The fourth requirement consists of two parts; the package must have two leads that are electrically separated from one another and the leads must be sufficiently heavy to conduct the maximum currents. The first of these requirements is met by using glass-to-metal seals in the package and the second is met by using copper leads of sufficiently heavy cross-section. The resulting packages for the units discussed in this paper are shown in Fig. 6. It should be remembered that the packages are only intermediate development packages and that further work will probably alter these both in size and in shape. However, all the requirements mentioned will be applicable to any package.

6.2 The units pictured in Fig. 6 have a range of dc current ratings associated with them. The lower rating of each device corresponds to the maximum rating of the next smaller device. Of course, the larger units could be used for smaller current applications; however, such use would be like using a freight car to haul a pound of coal. The maximum rating of each device has been arbitrarily chosen for it to operate with a reasonable sized cooling fin at an ambient of 125°C and no forced air or water cooling. It is known that the ratings could be increased by either method of forced cooling. It has been found that a copper convection cooling fin is able to dissipate 8 milliwatts per square inch per degree centigrade. This cooling rate is obtained from the difference between the average temperature of the fin and the ambient temperature over the effective exposed area of the fin. For example, a copper fin 3½ inches square when mounted so that both surfaces are effective for cooling will be able to dissipate ten watts and at the same time prevent the temperature of the fin from exceeding 50°C above the ambient temperature. Another thermal drop is found between the junction and the

base of the package. This temperature difference depends mostly on the material of the base and its geometry. In the devices presented this drop is not more than 15°C at the maximum rated current. Thus the largest drop in temperature occurs between the cooling fin and the ambient which means that the design of the cooling fin is the controlling factor in the operating junction temperature of the rectifier.

6.3 It is possible to use the devices without an attached cooling fin. In this case, the maximum current is limited essentially by the size of the package. The small rectifier package is designed for ½ watt dissipation and therefore the maximum current that should be rectified is about 500 milliamperes. The medium size unit will comfortably rectify 1 ampere without any additional cooling and the large rectifier unit will conduct 3 amperes under the same conditions.

7.0 RELIABILITY AND LIFE MEASUREMENTS

7.1 One of the desired properties of any device is that it should operate satisfactorily at its rating for a long period of time. The above general statement contains many implications which should be made specific for the devices under consideration in this paper. By stating that these devices should operate satisfactorily we mean that they should not age during operation; that is, the forward and reverse characteristics at any temperature should not change with time. The statement implies that a rating has been established for the units. Furthermore, a "long period of time" has to be defined. There are applications where a few hours is considered a long time as in some military applications. However, in most Bell System applications, a long period of time may be 20 years or approximately 200,000 hours. Clearly, in the short time since these rectifiers have been developed, it is impossible to make a fair statement as to their reliability and their life expectancy. However, it is possible to present some results of some early experiments and describe where and how the units have lived and died. It is this information that we will present in this section. It is a common experience that during the early development of any new component, there are many units that do not satisfy all the requirements of the desired end product. These units will generally deteriorate very rapidly on life testing due to some electrical or mechanical instability. The units used for life testing have been screened to remove the above mentioned unstable devices.

7.2 The life tests consist of four types; shelf tests at room temperature and at 150°C, forward characteristic tests, reverse characteristic

tests, and load tests. The last tests are really the important tests; however, these require the dissipation of large quantities of power in the load to test only a few devices. Therefore only a few units were tested in this condition and the majority tested under other conditions. The several units under load test have been operating for six months with no noticeable change in their characteristics. These devices are the small and medium size development units. The large rectifiers would require about 10 kilowatts of dissipation each in a load to give them a fair load test.

The shelf tests at room temperature and at a temperature of 150°C have been running for six months and have indicated that most of the units remain practically constant. There have been some units that improve on standing but there is no method of predicting which ones will improve. Some units get worse on standing; however, most of these can be predicted from the initial tests since these units usually have a noisy reverse characteristic near the reverse breakdown voltage. The units that change differ only in their reverse characteristic; the forward characteristic changes are not detectable indicating that the contacts are stable. The changes in the reverse characteristic are probably due to the trapping of ions and vapors on the surface of the devices during the packaging operation. Another source of these variations is due to the non-hermeticity of the glass-to-metal seals allowing gases to diffuse into the package where they may cause changes in the reverse characteristic. These leaks have been found in many early units and new assemblies are being tried at present.

The forward characteristic life test was considered a good test since the device is subject to practically all the internal power dissipation without requiring the relatively high load dissipation. It is tests of this nature that allow one to rate the various size devices. The medium size rectifiers that ran at 15 amperes in this test failed after three months of testing; whereas no units running at 5 and 10 amperes have failed during the six months since the tests have started although their reverse characteristics have changed slightly. It should be noted that most of the change of reverse characteristic occurred during the first test period of two weeks. These changes are probably due to the causes mentioned in the above paragraph.

Reverse characteristic tests have been running for several months on a group of 10 small rectifiers which we feel have a better gas tight seal than the other development units. The voltage has been adjusted on these units such that they are pulsed into the breakdown region with a

maximum current of one milliampere. None of these units show any appreciable change.

7.3 All of those tests in the past sub-section had to do with continuous dc or ac power being supplied to the units under test. However, in actual operation the units may be subject to voltage pulses due to power line pulses, accidental shorts, etc. In order for the rectifier to be useful, it should be able to take an overload for a period of time sufficiently long to allow a protective device to operate. Pulse tests have been performed on the medium size rectifier. These devices are able to withstand over 300 amperes for times of the order of 50 microseconds. However, the fastest circuit breakers operate in about 20 milliseconds and for this period, these units can stand only approximately 50 amperes before failing. Since these units have such a low forward resistance at the operating currents (Fig. 7), any small increase in voltage across the diode will change the current through the device to a very large quantity. Therefore series protective resistances may be necessary where the possibility of short-circuiting the device is high. Such operation would reduce the efficiency of the unit and is to be avoided if possible. Another type of protection may be afforded through the use of a high impedance, high current inductor. This type of protection is quite bulky and heavy and suitable only for stationary apparatus. Another common possibility of burnout of the devices occurs when using a capacitance input in conjunction with the rectifier. When the circuit is turned on, large currents will flow to charge up the capacitors and consequently burn out the rectifiers. One possible protection from such operation is the use of a series resistance in conjunction with a time delay relay. The series resistance will limit the initial capacitor charging current and the time delay relay will short out the resistance after the capacitors have reached near their maximum charge.

7.4 Dissection of burned out units have indicated that the failure takes place through small spots on the device. This can be explained by the fact that some small areas of the device have slightly better forward characteristics. These areas will tend to conduct most of the forward current. Therefore most of the power will be dissipated there and these areas will become even more conducting leading to a channeling of the forward current through these spots with the consequent burnout. The best way to avoid such mishaps would be to make a more uniform device. Experiments are in process along this line. Another less satisfactory method would be the control of contact resistance such that the current would be limited in any particular area by the contact resistance. Similar ideas must be considered when paralleling these diffused junction

silicon rectifiers. It is possible to use these devices in parallel if one adjusts the lead resistances such that no one unit will be allowed to conduct much more than its share of the current.

7.5 As a conclusion to this section, it should be noted that these rectifiers are expected to have a long life when operated within their ratings. They are able to operate for short periods of time (seconds) at five times their rated currents. Since the rectifiers have an extremely small series resistance, they should be protected against accidental surges and turning on to a capacitance input filter.

8.0 SUMMARY

8.1 The development rectifiers described in the article are silicon diffused p-n junction rectifiers. These devices together with associated cooling fins can be used to rectify a complete range of currents from 0 to 50 amperes in a single phase, half wave rectifier circuit. They can be used in more complex rectification circuits to yield even more dc current. Also, they are able to withstand at least 200 volts peak in the inverse direction and operate satisfactorily at temperatures as high as 200°C. Furthermore, one process of diffusion and plating is sufficient for all the devices of the class. This makes it possible for one diffusion and plating line to feed material for all the rectifiers in a manufacturing operation.

8.2 The rectifiers discussed behave according to the theory of semiconductor devices which makes it possible to design them for given electrical, thermal, and mechanical characteristics. One failure to meet ideal theory of a p-n junction is with the forward characteristic.

8.3 The diffused silicon type of rectifier has been compared with germanium and selenium units and has better reverse characteristics at all temperatures. In the forward direction, the germanium units have a smaller voltage drop for any given current than the silicon rectifiers but the silicon devices are capable of operating at much higher temperatures, thereby permitting higher overall current densities than the germanium devices.

8.4 The diffused silicon rectifiers are capable of use in any rectifier application where dc currents up to the order of 100 amperes are required and where inverse peak voltages up to 200 volts are encountered. Another imoortant use for these devices will be in the magnetic amplifier application where the low reverse currents of silicon will enable large amplification factors to be realized. Since the forward characteristics of these devices are so uniform, they can be used in voltage reference circuits that require voltages near 0.6 volts and in circuits uti-

lizing the exponential character of the forward characteristic. However, as is to be expected from devices with the characteristics described in this paper, the most immediate application will be found in power supplies.

ACKNOWLEDGMENTS

It is obvious that the work reported in this paper is not the result of one man's labor. Much of the stimulus and many of the ideas are those of K. D. Smith. Other members of the Semiconductor Device Department who have contributed considerably to the development of these devices are R. L. Johnston, R. Rulison, and R. C. Swenson. D. A. Kleinman, J. L. Moll and I. M. Ross have been most helpful in discussing the theoretical aspects of these devices. The author wishes to thank H. R. Moore for his suggestions on protecting the silicon rectifiers against large overloads.

The Forward Characteristic of the PIN Diode

By D. A. KLEINMAN

(Manuscript received January 18, 1956)

A theory is given for the forward current-voltage characteristic of the PIN diffused junction silicon diode. The theory predicts that the device should obey a simple PN diode characteristic until the current density approaches 200 amp/cm². At higher currents an additional potential drop occurs across the middle region proportional to the square root of the current. A moderate amount of recombination in the middle region has little effect on the characteristic. It is shown that the middle region cannot lead to anomalous characteristics at low currents.

INTRODUCTION

In some diode applications it is desirable to have a very low ohmic resistance as well as a high reverse breakdown voltage. A device meeting these requirements, in which the resistance is low because of heavily doped P^+ and N^+ contacts and the breakdown voltage is high because of a lightly doped layer between the contacts, has been described by M. B. Prince.[1] The device is shown schematically in Figure 1a and consists of three regions, the P^+ contact, the middle P layer, and the N^+ contact. The device is called a PIN diode because the density P of uncompensated acceptors in the middle region is much less than P^+ or N^+ and in normal forward operation much less than the injected carrier density.[2]

We shall let the edge of the P^+P junction in the middle region be $x = 0$, and the edge of the PN^+ junction in the middle region be $x = w$. Thus the region $0 \leqq x \leqq w$ is space charge neutral and bounded at each end by space charge regions whose width is of the order of the Debye length

[1] Prince, M. B., Diffused *p-n* Junction Silicon Rectifiers, B.S.T.J., page 661 of this issue.

[2] A device with similar geometry has been discussed by R. N. Hall, Proc. I.R.E., **40,** p. 1512, 1952.

$$\lambda = (K/\beta e P)^{1/2} \sim 1.5 \times 10^{-5} \text{ cm.} \tag{1}$$

where K is the dielectric constant, e is the electronic charge, and β is the constant

$$\beta = e/kT = \mu_n/D_n = \mu_P/D_P \tag{2}$$

which at room temperature is 38.7 volt^{-1}. We shall denote points in the P^+ and N^+ contacts on the edges of the space charge regions by oo and ww respectively. Thus n_{oo} is the electron density in the P^+ contact at the junction, and n_o is the electron density at the same junction in the middle region. Similarly p_{ww} is the hole density at the junction in the N^+ contact and p_w is the hole density at the junction in the middle region. We shall denote equilibrium carrier densities in the three regions by n_{P^+}, n_P, p_p, p_{N^+}. Typical values for the parameters characterizing

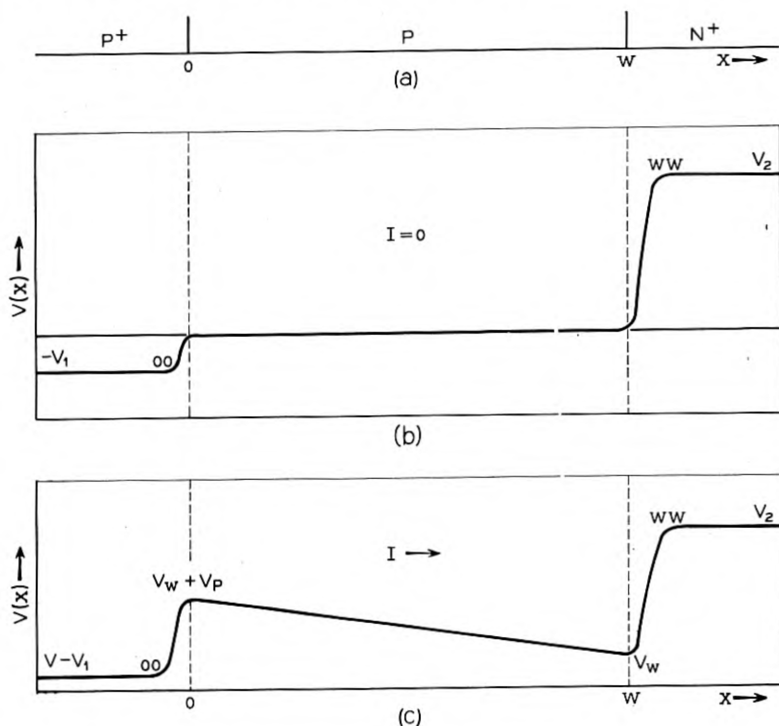


Fig. 1 — Schematic representation of the PIN diode with the P+ and N+ contacts regarded as extending to infinity. (b) shows the electrostatic potential in equilibrium and (c) shows the potential when a forward current flows.

the device are

$$\begin{aligned} w &\sim 2 \times 10^{-3} \text{ cm} \\ P &\sim 10^{15} \text{ cm}^{-3} \\ N^+, P^+ &\sim 10^{18} \text{ cm}^{-3} \\ L_n, L_p &\sim 10^{-4} \text{ cm} \end{aligned} \tag{3}$$

where L_n, L_p are minority carrier diffusion lengths in the contacts.

The present treatment makes three distinct approximations. The first is to neglect the voltage drop in the contacts. The highest currents ordinarily used are of the order of 500 amp/cm^2 which should produce an ohmic drop in the contacts of about 1 volt/cm. Since the entire diode has a length of about 0.01 cm we are neglecting only about 0.01 volts in this approximation.

The second approximation is to regard the Debye length as small compared to w and the diffusion lengths L_n, L_p. If L_n, L_p are as small as the typical values given in (3) the error made in this approximation is not completely negligible. Nevertheless, we use the approximation because it enables us to regard the device as three relatively large neutral regions and two relatively narrow space charge regions. The behavior of the device can then be determined by solving for the diffusion and drift of carriers in the neutral regions subject to boundary conditions connecting the carrier densities across the space charge layers.

The third approximation is to neglect any increase in majority carrier density in the contacts due to injection of minority carriers. This approximation is valid until the current density approaches 5×10^4 amp/cm^2, which is well above anticipated operating currents. It is conceivable that in some junctions all the current may flow through small active spots at which the current density is very high, perhaps exceeding the above figure. In such cases the current flow is two or three dimensional and the present analysis would not apply.

It is also necessary to assume some law for carrier recombination. We shall assume that recombination in the contacts is linear in the injected minority carrier density

$$\frac{dI_n}{dx} \sim \frac{n - n_{P^+}}{\tau} \tag{4}$$

Modification of the theory to suit other recombination laws is simple in principle, although considerable analytical complications might be encountered. It seems most likely that in silicon PN junctions the recombination actually is nonlinear. It can be shown that if the recombi-

nation follows some power ν of the injected density

$$\frac{dI_n}{dx} \sim n^\nu \tag{5}$$

the forward characteristic of a simple PN junction is of the form

$$\exp\left[\tfrac{1}{2}\beta(\nu + 1)V\right] \tag{6}$$

Thus nonlinear recombination can account for the observation that in silicon diodes the slope of V versus log I is usually much less than β. Our purpose here is not to study this interesting effect, but to study those effects which are due to the presence of the middle region. Therefore, we assume linear recombination for the sake of simplicity. In the last section we give a brief consideration of what to expect in the case of nonlinear recombination in the contacts. Recombination in the middle region will also be assumed to be linear in the injected carrier density, but this assumption is not critical, since it turns out that a moderate amount of recombination in the middle region does not change the qualitative behavior of the device.

BASIC EQUATIONS

Fig. 1(b)[3] shows the electrostatic potential $V(x)$ for the equilibrium case $I = 0$. The potential is constant except in the space charge layers. If we call the potential of the middle region zero, the P^+ and N^+ contacts are at the potentials $-V_1$ and V_2 respectively, where

$$\begin{aligned} \beta V_1 &= \ell n\,(P^+/p_P) \\ \beta V_2 &= \ell n\,(N^+/n_P) \end{aligned} \tag{7}$$

Figure 1c shows the potential when a forward current I flows and a forward bias V is produced across the device. We shall define the potential so that the N^+ contact remains at V_2, which puts the P^+ contact at potential $V - V_1$. The potential at a point x is then given by

$$V(x) = V_2 - \int_{ww}^{x} E(x)\,dx \tag{8}$$

where $E(x)$ is the electric field assumed zero in the contact regions $x > ww$ and $x < oo$. The applied bias V consists of three terms

$$V = V_0 + V_P + V_w \tag{9}$$

[3] This potential distribution has been discussed by A. Herlet and E. Spenke, Zeits. f. Ang. Phys., **B7, H3,** p. 149, 1955.

where V_0 is the forward bias across the junction at $x = 0$, V_P is the potential drop in the middle region, and V_w is the forward bias across the junction at $x = w$. In this notation $V(0) = V_w + V_P$ and $V(w) = V_w$.

The total current density is constant

$$I_n(x) + I_p(x) = I \tag{10}$$

We shall denote electric current densities by eI_n, eI_p, so that I_n, I_p, I have the dimensions of (particles/cm^2-sec). At $x = 0$ and $x = w$ the minority carrier currents must flow into the contacts by diffusion, which gives the boundary conditions

$$\begin{aligned} I_p(w) &= I_{ps}\left\{\frac{p_{ww}}{p_{N^+}} - 1\right\} \\ I_n(0) &= I_{ns}\left\{\frac{n_{oo}}{n_{P^+}} - 1\right\} \end{aligned} \tag{11}$$

where I_{ps}, I_{ns} are saturation current densities

$$I_{ps} = \frac{p_N{}^+ D_p}{L_P}, \qquad I_{ns} = \frac{n_P{}^+ D_n}{L_n} \tag{12}$$

The order of magnitude of the saturation current density is given by

$$e(I_{ns} + I_{ps}) \sim 3 \times 10^{-10} \text{ amp/cm}^2 \text{ in } Si$$

based on the typical values of (3). Equations (11) contain the assumptions of linear recombination and small injection into the contacts as discussed in the introduction.

In the middle region the current densities satisfy

$$\begin{aligned} I_p(x) &= D_p\left\{-\frac{dp}{dx} + \beta p E\right\} \\ I_n(x) &= D_n\left\{\frac{dn}{dx} + \beta n E\right\} \end{aligned} \tag{13}$$

Let us assume these equations remain valid in the space charge regions.[4] Since these space charge regions are narrow I_n and I_p can be considered constant and the solution of (13) in the space charge regions is

$$\begin{aligned} p(x) &= e^{-\beta V(x)}\left\{p_{ww}e^{\beta V_2} - \frac{I_p(w)}{D_P}\int_{ww}^{x} e^{\beta V(x)}dx\right\} \\ n(x) &= e^{\beta V(x)}\left\{n_{00}e^{-\beta(V-V_1)} + \frac{I_n(0)}{D_n}\int_{00}^{x} e^{-\beta V(x)}dx\right\} \end{aligned} \tag{14}$$

[4] Shockley, W., B.S.T.J., **28,** p. 435, 1949.

Since $\lambda/L_p \ll 1$ we can write for the junction at $x = w$

$$p(w) = e^{-\beta V_w} \left\{ p_{ww} e^{\beta V_2} - \frac{(p_{ww} - p_N^+)}{L_p} \int_{ww}^{w} e^{\beta V} dx \right. \qquad (15)$$
$$\left. = p_{ww} e^{\beta(V_2 - V_w)} \left\{ 1 - 0\left(\frac{\lambda}{L_p}\right) \right\} \right.$$

where $0(\lambda/L_p)$ means a term of order λ/L_p. Thus we see that if we may neglect λ/L_p and λ/L_n we have the following simple boundary conditions at the junctions

$$\begin{aligned} n_{oo} &= n_o(n_P^+/n_P)e^{\beta V_0} \\ p_o &= p_P\, e^{\beta V_0} \\ n_w &= n_P\, e^{\beta V_w} \\ p_{ww} &= p_w(p_N^+/p_P)e^{\beta V_w} \end{aligned} \qquad (16)$$

It is clear that in order to divide the device into three neutral regions we must also be able to neglect λ/w.

Finally, we have the condition of space charge neutrality

$$p - n = P \qquad (17)$$

It can be shown that the term $K^{-1}\, dE/dx$ is of order $(\lambda/L)^2$ or $(\lambda/w)^2$ and therefore negligible in our approximation. Therefore (17) is the Poisson equation for the middle region in our approximation. When we use (17) we are not saying that $E(x)$ is constant but only that $K^{-1}\, dE/dx$ is negligible compared to $p(x)$ and $n(x)$. The basic equations then are (10), (11), (13), (16), (17).

Large Injection, No Recombination

In this section we consider current densities of the order of magnitude of those that flow in normal operation of the diode as a power rectifier. These currents inject large densities of electrons and holes into the middle region greatly increasing its conductivity. The result is that the voltage drop V_P is small even though the normal resistivity of the middle region is high. For this reason the device has been called a conductivity modulated rectifier. Also in this section we shall neglect recombination in the middle region, which makes $I_n(x)$ and $I_p(x)$ constant and greatly simplifies the analysis. The effect of recombination is to remove carriers and increase the drop across the middle region. Therefore, it is desirable to keep recombination in the middle region as low as possible.

Under conditions of large injection we can say

$$n \gg P, \qquad p \gg P$$
$$n_{oo} \gg n_P{}^+ \qquad p_{ww} \gg p_N{}^+ \tag{18}$$

so that (11) becomes

$$I_n = I_{ns}(n_{oo}/n_P{}^+)$$
$$I_p = I_{ps}(p_{ww}/p_N{}^+) \tag{19}$$

and (17) becomes

$$n(x) = p(x) \qquad 0 \leqq x \leqq w \tag{20}$$

Equation (16) becomes

$$n_{oo} = n_o(n_P{}^+/n_P)e^{\beta V_0}$$
$$n_o = p_P e^{\beta V_0}$$
$$n_w = n_P e^{\beta V_w}$$
$$p_{ww} = n_w(p_N{}^+/p_P)e^{\beta V_w} \tag{21}$$

Equations (13) can be written

$$\beta E = \frac{I_n + bI_p}{2D_n n}$$
$$\frac{dn}{dx} = \frac{I_n - bI_p}{2D_n} \tag{22}$$

where $b = D_n/D_p$. Combining (19) and (21) gives the equations

$$n_o = n_i(I_n/I_{ns})^{1/2}$$
$$n_w = n_i(I_p/I_{ps})^{1/2} \tag{23}$$

where $n_i^2 = n_P p_P$ is a constant, and also

$$\beta V_0 = \tfrac{1}{2}\, \ell n\, \frac{n_P}{p_P} \frac{I_n}{I_{ns}}$$
$$\beta V_w = \tfrac{1}{2}\, \ell n\, \frac{p_P}{n_P} \frac{I_p}{I_{ps}} \tag{24}$$

From the first equation (22) we have

$$\beta V_P = \frac{I_n + bI_p}{2D_n} \int_0^w \frac{dx}{n(x)} \tag{25}$$

Upon invoking the second equation of (22) we get

$$\beta V_P = \frac{I_n + bI_p}{I_n - bI_p} \ell n \frac{n_w}{n_o} \tag{26}$$

and

$$n_w = n_o + \frac{I_n - bI_p}{2D_n} w. \tag{27}$$

We see that V_p is always positive in sign whatever the sign of $I_n - bI_p$.

We now define a parameter

$$\gamma \equiv n_o/n_w \tag{28}$$

and a device constant

$$R \equiv I_{ns}/I_{ps} \tag{29}$$

Then from (23) and (10)

$$\begin{aligned} I_n/I_p &= R\gamma^2 \\ I_n &= \frac{R\gamma^2}{1 + R\gamma^2} I \qquad I_p = \frac{1}{1 + R\gamma^2} I \end{aligned} \tag{30}$$

Combining (23), (27) and (30) gives the equation for γ as a function of total current

$$\begin{aligned} \gamma &= 1 - \frac{I_n - bI_p}{2D_n} \frac{w}{n_w} \\ &= 1 - \sqrt{\frac{I}{I_0}} \frac{(\gamma/\gamma_\infty)^2 - 1}{\sqrt{1 + b(\gamma/\gamma_\infty)^2}} \end{aligned} \tag{31}$$

where

$$\gamma_\infty^2 \equiv b/R \tag{32}$$

and I_0 is a unit of (particle) current density characteristic of the device

$$I_0 = \frac{4D_p^2 n_i^2}{w^2 I_{ps}} = 4 \left(\frac{L_p}{w}\right)^2 \frac{N^+ D_p}{L_p} \tag{33}$$

A typical value for $e\,I_0$ in a silicon diode is

$$e\,I_0 \sim 200 \text{ amp/cm}^2 \tag{34}$$

based on (3).

From (26) the potential drop in the middle region can be written

$$\beta V_P = -\frac{\gamma^2 + \gamma_\infty^2}{\gamma^2 - \gamma_\infty^2} \ln \gamma \tag{35}$$

From (24) and (30)

$$\beta(V_0 + V_w) = \ln \frac{I}{I_0} + \ln \frac{\gamma}{1 + b(\gamma/\gamma_\infty)^2} + \ln \frac{I_0}{I_{ps}} \tag{36}$$

Thus the total applied bias V as a function of total current density I is given by

$$\beta V = \ln \frac{I}{I_0} - \frac{\gamma^2 + \gamma_\infty^2}{\gamma^2 - \gamma_\infty^2} \ln \gamma + \ln \frac{\gamma}{1 + b(\gamma/\gamma_\infty)^2} + \ln \frac{I_0}{I_{ps}} \tag{37}$$

where $\gamma(I)$ is the (positive) solution of (31).

Thus far we have referred the problem of the $V - I$ characteristic to the problem of calculating $\gamma(I)$ from (31). We see that in the limits of high and low current γ approaches the limits

$$\begin{aligned} \gamma &\rightarrow 1 & I &\ll I_0 \\ \gamma &\rightarrow \gamma_\infty & I &\gg I_0 \end{aligned} \tag{38}$$

and in general lies between these limits. A good approximate solution is readily obtained by replacing (31) with the quadratic equation

$$\begin{aligned} \gamma &= 1 - z[(\gamma/\gamma_\infty)^2 - 1] \\ z &= (I/I_0)^{1/2} (1 + b)^{-1/2} \end{aligned} \tag{39}$$

which has the solution

$$\gamma = \frac{\sqrt{\gamma_\infty^4 + 4(1 + z)z\gamma_\infty^2} - \gamma_\infty^2}{2z} \tag{40}$$

A plot of this solution is shown in Fig. 2 as a function of z for $\gamma_\infty = \frac{1}{2}$, $\gamma_\infty = 2$. Since $\gamma(I)$ is bounded by unity and γ_∞, which usually will be of order unity, we can reject some of the dependence of V upon γ and retain only its essential dependence upon I. This appears in the first and second terms of (37). By means of (31) this second term can be written

$$\beta V_p = \left[\frac{\ln \gamma}{\gamma - 1} \frac{(\gamma/\gamma_\infty)^2 + 1}{\sqrt{1 + b(\gamma/\gamma_\infty)^2}}\right] \sqrt{\frac{I}{I_0}} \tag{41}$$

Retaining only the essential dependence on I we write this equation

$$\beta V_p = C(I/I_o)^{1/2} \tag{42}$$

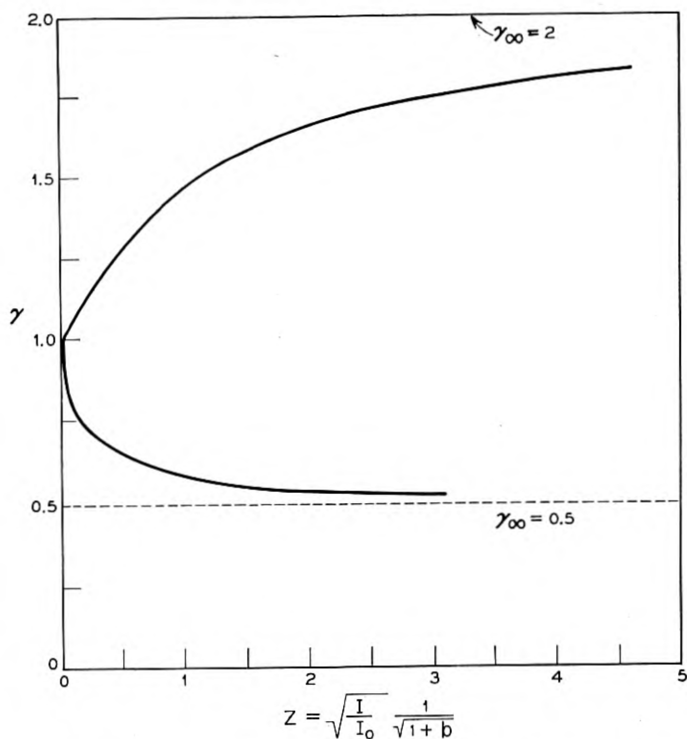


Fig. 2 — The function $\gamma(z)$ given by equation (40) for two choices of γ_∞.

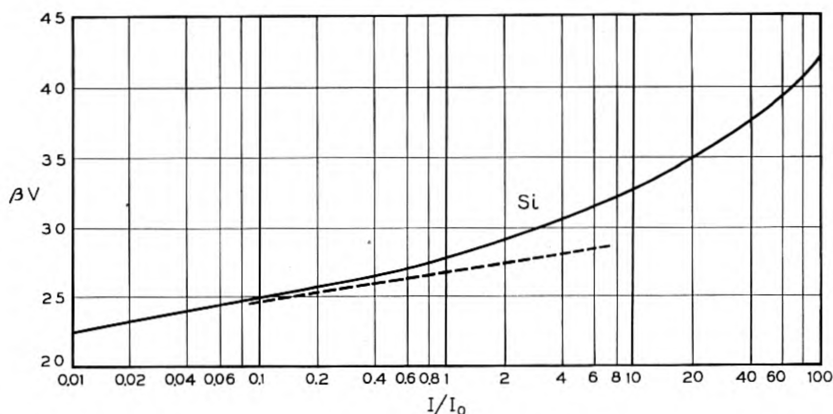


Fig. 3 — The voltage-current characteristic of the PIN diode according to equation (44). The dashed line represents an ideal *PN* diode and $eI_0 \sim 200$ amp/cm^2 in silicon.

where C is a constant representing the slowly varying coefficient of $(I/I_o)^{1/2}$ in (41). We choose C such that (42) becomes exact at high current density when βV_p is large

$$C = \frac{\ell n\, \gamma_\infty}{\gamma_\infty - 1} \frac{2}{\sqrt{b+1}} \tag{43}$$

When we regard the third and fourth terms of (37) together as a constant βV_c we obtain the simplified voltage-current characteristic

$$\beta V = \ell n \frac{I}{I_0} + C \sqrt{\frac{I}{I_0}} + \beta V_c \tag{44}$$

In this approximation it is unnecessary to evaluate $\gamma(I)$ from (31).

Fig. 3 shows plots of βV versus I/I_o calculated from (44). For plotting the curves the value $C = 1.1$ was used. To choose a value for βV_c we put $\gamma = 1$, which gives

$$\ell n \frac{\gamma}{1 + b(\gamma/\gamma_\infty)^2} \to \ell n \frac{1}{1+R} \quad \gamma \to 1 \tag{45}$$

so that

$$\beta V_c \to \ell n[I_o/(I_{ns} + I_{ps})] \tag{46}$$

which has the value 27 in silicon according to the values in (3). The dotted line is the asymptote approached by the curve at low current densities

$$\beta V \to \ell n \frac{I}{I_{ns} + I_{ps}} \qquad I \ll I_0 \tag{47}$$

This is the characteristic of a simple PN junction when

$$I \gg I_{ns} + I_{ps} .$$

We return now to the question of when the large injection conditions (18) are satisfied. Let us suppose I is much less than I_0 so that $\gamma \sim 1$, $I_n/I_p \approx R$. It follows from (30) and (23) that

$$n_o \approx n_w \approx n_i[I/(I_{ns} + I_{ps})]^{1/2} \tag{48}$$

Now let us set $n_o \gg P$ which gives a condition on the current density

$$I \gg (P/n_i)^2\,(I_{ns} + I_{ps}). \tag{49}$$

Setting $n_{oo} \gg n_p^+$, $p_{ww} \gg p_N^+$ gives

$$I \gg I_{ns} + I_{ps} . \tag{50}$$

Usually $P \gg n_i$ so that (49) includes (50). When numbers are put in

from (3) we get the condition for large injection

$$eI \gg 0.07 \text{ amp/cm}^2 \text{ in } Si \tag{51}$$

Since this current in (51) is much less than eI_o, we may quite properly speak of large injection $n \gg P$ and small currents $I \ll I_o$ at the same time.

Let us denote by

$$I_{CM} = (P/n_i)^2 (I_{ns} + I_{ps}) \tag{52}$$

the current density at which conductivity modulation starts to be important. Then we may distinguish three ranges of current: (a) very small current $I < I_{CM}$ for which large injection analysis does not apply; (b) low current $I_{CM} < I < I_o$ for which large injection analysis applies, but the voltage drop V_p in the middle region is negligible; (c) large current $I > I_o$ for which V_P is sizable. The treatment of this section has covered ranges (b) and (c). Range (c) (as treated here) does not extend to infinity but only up to current densities of the order

$$\frac{eN^+D_p}{L_p} \sim 8 \times 10^4 \text{ amp/cm}^2$$

so that the diffusion currents in the contacts may be treated as a small injection.

Small Injection, No Recombination

In this section, we shall cover ranges (a) and (b) in current density. We must go back to the basic equations, but we shall make use of two facts that have come out of the large injection analysis: (a) βV_p is negligible when $I \ll I_o$; (b) $\gamma = n_o/n_w \approx 1$ which means $n(x)$ and $p(x)$ are essentially constant in the middle region $0 \leqq x \leqq w$ when $I \ll I_o$. When we set

$$n_o = n_w, \qquad p_o = p_w \tag{53}$$

equations (16) give us

$$n_{oo} = n_P{}^+ e^{\beta(V_0+V_w)}$$
$$p_{ww} = p_N{}^+ e^{\beta(V_0+V_w)} \tag{54}$$

Then (11) gives

$$I = I_n + I_p = (I_{ns} + I_{ps}) [e^{\beta(V_0+V_w)} - 1] \tag{55}$$

Now $V_o + V_w$ is the total applied bias when V_P can be neglected; there-

fore we obtain the characteristic

$$\beta V = \ln\left(\frac{I}{I_{ns} + I_{ps}} + 1\right) \tag{56}$$

which is valid until I approaches I_o. Of course we would not have obtained this ideal characteristic of a simple PN junction had we taken recombination into account; our result depends upon the constancy of $n(x)$ and $p(x)$ in the middle region. For the case of no recombination in the middle region (56) and (44) cover ranges (a), (b) and (c). Instead of (44) the more exact expression (37) could be used requiring the evaluation of $\gamma(I)$ from (31). It seems that the extra refinement is of no help in understanding the device and unnecessary in treating experimental data. Therefore, we shall adopt (44) and the approximations leading to it as a model for treating the more complicated recombination case. That is, we shall seek a generalization of (44) which takes recombination into account in a sufficiently good approximation.

Large Injection with Recombination

We are interested in determining the effect of recombination in the middle region upon the operating characteristics of the device. Therefore we go immediately to the large injection case $n = p$. Equation (16) become

$$\begin{aligned} n_w &= n_P e^{\beta V_w} \qquad & p_{ww} &= n_w(p_N{}^+/p_P)e^{\beta V_w} \\ n_o &= p_P e^{\beta V_0} \qquad & n_{oo} &= n_o(n_P{}^+/n_P)e^{\beta V_0} \end{aligned} \tag{57}$$

which gives

$$\beta(V_o + V_w) = \ln(n_w n_o/n_i^2) \tag{58}$$

We shall assume that recombination is linear in the injected carrier density to simplify the calculation. It will be possible, later to approximate bimolecular recombination by using an appropriate value for the lifetime τ corresponding to the injected carrier density. Therefore we write

$$\frac{dI_n}{dx} = -\frac{dI_p}{dx} = \frac{n}{\tau} \tag{59}$$

Eliminating $I_n(x)$ by use of (13) gives the equation for $n(x)$

$$\frac{d^2n}{dx^2} = \frac{n}{L^2} \tag{60}$$

where L is the effective diffusion length in the middle region

$$L = [2D_n \tau/(b + 1)]^{1/2} \tag{61}$$

The solution of (60) may be written

$$n(z) = \frac{n_0 \sinh (w - z) + n_w \sinh z}{\sinh \omega} \tag{62}$$

where $z = x/L$ is the position variable and $\omega = w/L$ is the length of the middle region in units of L. Fig. 4 shows several of these solutions for the case $n_o = n_w$.

In equation (60) and the solution (62) we have neglected the equilibrium carrier densities n_p, p_P. The criterion for the validity of this approximation is

$$\sinh \tfrac{1}{2}\omega \ll (n_o/P), (n_w/P) \tag{63}$$

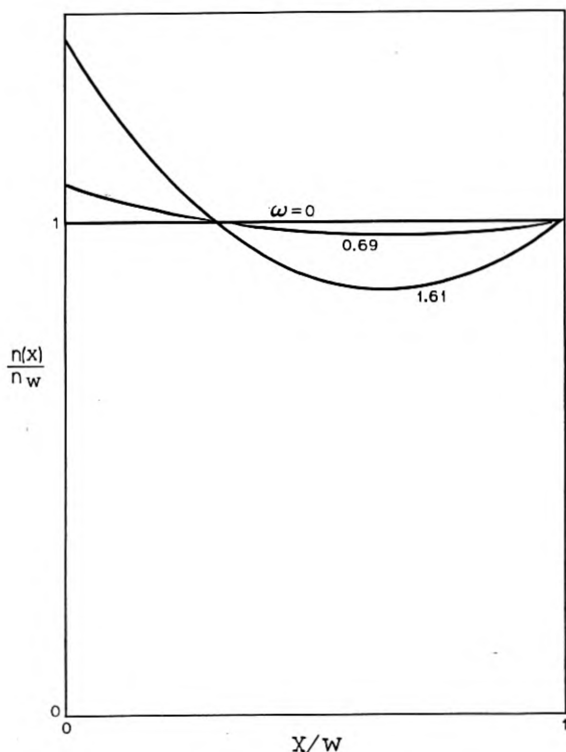


Fig. 4 — The carrier density according to equation (62) for the case $n_0 = n_w$ and several values of ω.

arrived at by considering the minima in the solutions for $\omega \gg 1$. This is really a criterion for conductivity modulation, so we shall assume henceforth that it is satisfied.

We now modify (13) by setting $n = p$ and eliminating $E(x)$ by use of (22)

$$I_n(x) = \frac{bI + 2D_n n'(x)}{b + 1}$$
$$I_p(x) = \frac{I - 2D_n n'(x)}{b + 1} \tag{64}$$

where $n'(x) = dn/dx$. Inserting these currents into (22) gives $E(x)$ and integrating gives the potential drop V_P in the middle region

$$\beta V_p = \frac{bI}{(b + 1)D_n} \int_0^w \frac{dx}{n} - \frac{b - 1}{b + 1} \ln \frac{n_w}{n_0} \tag{65}$$

This is the generalization of (26) for linear recombination.

The direct evaluation of (58) and (65) in terms of the total current I leads to a very complicated expression for the applied voltage. It will be shown in the next section that this result reduces in its simplest approximate form retaining only the essential dependence on ω to the formula

$$\beta V \approx \ln \frac{I}{I_{ns} + I_{ps}} + C \sqrt{\frac{I}{I_0(\omega)}} \tag{66}$$

which is identical with (44) except that the characteristic current density is a function of ω

$$I(\omega) = I_0 g(\omega)$$
$$g(\omega) = \frac{(\omega/2)^2}{\left[\cosh \frac{\omega}{2} \tan^{-1}\left(\sinh \frac{\omega}{2}\right)\right]^2} \tag{67}$$
$$= 1 - \frac{\omega^2}{6} + \frac{\omega^4}{48} - \cdots$$

Fig. 5 shows a plot of $g(\omega)$. These results show that if $\omega < 1$ as we might expect in a good diode recombination has no significant effect on the forward voltage-current characteristic in the conductivity modulation range of operation.

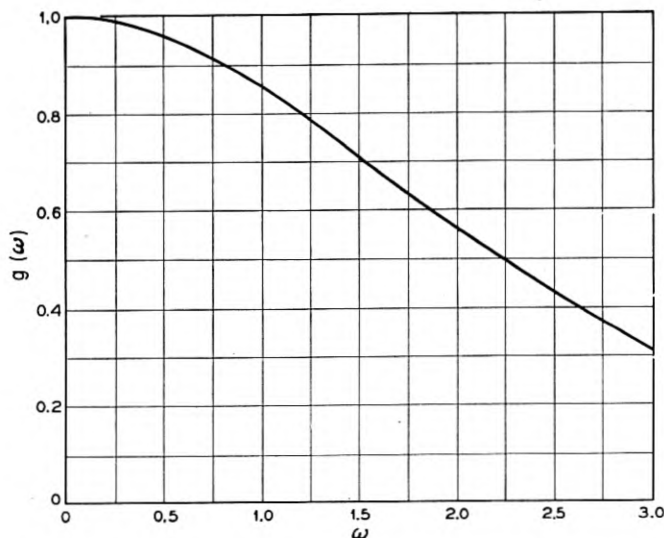


Fig. 5 — The function $g(\omega)$ of equation (67).

Analysis

We denote

$$\xi = \frac{n_w}{n_i}, \qquad \zeta = \frac{n_0}{n_i} \tag{68}$$

From (11) and (67)

$$I_p(\omega) = I_{ps}\xi^2, \qquad I_n(0) = I_{ns}\zeta^2 \tag{69}$$

By means of (62) and (64) we eliminate I_n and I_p and obtain the equations

$$\begin{aligned} (b + 1)I_{ps}\xi^2 &= I - I_r(\xi \cosh \omega - \zeta) \\ (b + 1)RI_{ps}\zeta^2 &= bI + I_r(\xi - \zeta \cosh \omega) \end{aligned} \tag{70}$$

where I_r is a (particle) current density

$$I_r = \frac{2D_n n_i}{L \sinh \omega} \tag{71}$$

In principle we could solve (70) for ξ and ζ as functions of I with R and ω as parameters; this would determine βV through (58) and (65) and complete the problem. First we shall rewrite these equations in terms of γ as in the analysis of the second section.

If we eliminate I from equations (24) we get

$$bI_{ps} + \frac{I_r}{\xi}\frac{b\cosh\omega + 1}{b+1} = RI_{ps}\gamma^2 + \gamma\frac{I_r}{\xi}\frac{\cosh\omega + b}{b+1} \tag{72}$$

which can be solved for ξ

$$\xi = \frac{I_r}{I_{ps}}\frac{\cosh\omega + b}{b+1}\frac{\gamma_0 - \gamma}{R\gamma^2 - b} \tag{73}$$

where

$$\gamma_0 = \frac{b\cosh\omega + 1}{\cosh\omega + b} \tag{74}$$

Substituting (73) into (70) gives the equation satisfied by γ

$$\left(\gamma - \frac{R\gamma^2\cosh\omega + 1}{R\gamma^2 + \cosh\omega}\right)(\gamma - \gamma_0) = \frac{I}{I_{00}}\frac{[(\gamma/\gamma_\infty)^2 - 1]^2}{R\gamma^2 + \cosh\omega} \tag{75}$$

where I_{00} is a characteristic (particle) current density

$$I_{oo} = I_o\left[\frac{\omega}{\sinh\omega}\right]^2\frac{\cosh\omega + b}{b+1} \tag{76}$$

Now the solution of (75) has two branches which as $I \to 0$ approach values given by

$$\text{a)}\quad \gamma \to \gamma_0$$
$$\text{b)}\quad \gamma \to \frac{R\gamma^2\cosh\omega + 1}{R\gamma^2 + \cosh\omega} \tag{77}$$

As I increases the first branch remains positive and approaches γ_∞ as $I \to \infty$. The second branch becomes negative and approaches $-\gamma_\infty$. Therefore, we choose that branch which satisfies

$$\gamma(0) = \gamma_0 = \frac{b\cosh\omega + 1}{b+1}$$
$$\gamma(\infty) = \gamma_\infty = (b/R)^{1/2} \tag{78}$$
$$\gamma > 0$$

On this branch γ always lies between γ_0 and γ_∞, and γ never approaches the quantity in (77b). Therefore we replace $R\gamma^2$ by b (as if $\gamma = \gamma_\infty$) in

the first factor on the left of (75), and obtain the simpler form

$$\gamma - \gamma_0 = -\sqrt{\frac{I}{I_{00}}}\frac{(\gamma/\gamma_\infty)^2 - 1}{\sqrt{R\gamma^2 + \cosh\omega}} \tag{79}$$

which is the generalization of (31).

The drop βV_P in the middle region given by (65) can be written

$$\beta V_P = \frac{b-1}{b+1}\,\ell n\,\gamma + \frac{2}{b+1}\sqrt{\frac{I}{I_{00}}}\sqrt{R\gamma^2 + \cosh\omega}\,F_\omega(\gamma) \tag{80}$$

where $F_\omega(\gamma)$ comes from $\int dx/n$ and is defined

$$\begin{aligned} F_\omega(\gamma) &= \int_0^1 \frac{\omega du}{\gamma \sinh[\omega(1-u)] + \sinh[\omega u]} \\ &= \frac{\ell n\left|\dfrac{1+Q}{1+Q}\right| - \ell n\left|\dfrac{1+e^\omega Q}{1-e^\omega Q}\right|}{\sqrt{1 - 2\gamma\cosh\omega + \gamma^2}} \end{aligned} \tag{81}$$

or

$$2\,\frac{\tan^{-1} e^\omega Q - \tan^{-1} Q}{\sqrt{2\gamma\cosh\omega - 1 - \gamma^2}}$$

The first form applies when $\gamma > e^\omega$, or $\gamma < e^{-\omega}$, and the second applies when $e^{-\omega} < \gamma < e^\omega$, and Q is the quantity

$$Q = \frac{1 - \gamma e^{-\omega}}{\sqrt{|1 - 2\gamma\cosh\omega + \gamma^2|}} \tag{82}$$

It can readily be shown that when $\omega \to 0$

$$F_0(\gamma) = \frac{\ln\gamma}{\gamma - 1} \tag{83}$$

Thus when $\omega = 0$ (80) reduces to

$$\begin{aligned} \beta V_P &\to \frac{\ell n\,\gamma}{\gamma - 1}\frac{b-1}{b+2}(\gamma - 1) + \frac{2}{b+1}\sqrt{\frac{I}{I_0}}\sqrt{R\gamma^2 + 1} \\ &= \left[\frac{\ell n\,\gamma}{\gamma - 1}\frac{(\gamma/\gamma_\infty)^2 - 1}{R\gamma^2 + 1}\right]\sqrt{\frac{I}{I_0}} \end{aligned} \tag{84}$$

which is identical with (41). It is also clear that (79) reduces to (31) as the recombination goes to zero. Finally we write from (58)

$$\beta(V_0 + V_w) = \ell n\,\gamma\xi^2 = \ell n\,\frac{I}{I_{ps}} + \ell n\,\frac{\gamma}{R\gamma^2 + \cosh\omega} \tag{85}$$

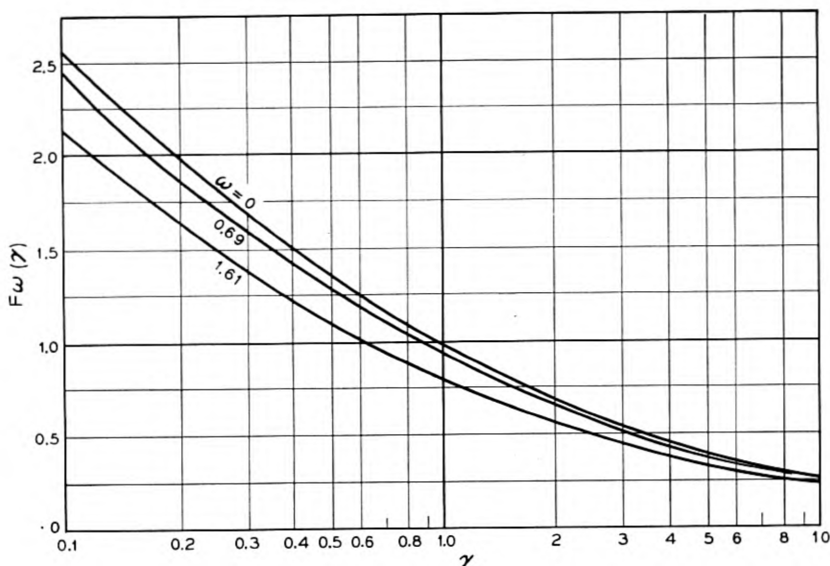


Fig. 6 — The function $F_\omega(\gamma)$ of equation (81) for several values of ω.

which reduces to (36) when $\omega = 0$. Thus the whole theory reduces correctly in the case $\omega = 0$.

The function $F_\omega(\gamma)$ is plotted in Fig. 6 for several values of ω including $\omega = 0$. The expansion of $F_\omega(\gamma)$ to order ω^2 is

$$F_\omega(\gamma) = \frac{\ln \gamma}{\gamma - 1} - \frac{\omega^2}{4} f(\gamma)$$

$$f(\gamma) = \frac{(\gamma + 1) - 2\gamma \dfrac{\ln \gamma}{\gamma - 1}}{(\gamma - 1)^2} \tag{86}$$

$$= 1 - 2\gamma \ln \frac{1}{\gamma} + \cdots$$

Our next step is to eliminate from (80) and (85) unimportant dependencies on I which would be difficult or impossible to detect experimentally. If in (85) we let $\gamma = 1$, $\cosh \omega = 1$ we get

$$\beta(V_0 + V_w) \approx \ln \frac{I}{I_{ns} + I_{ps}} \tag{87}$$

In (80) we drop the first term (as if $\gamma = 1$) and in the second term we

put $R\gamma^2 = b$ (as if $\gamma = \gamma_\infty$) and $F_\omega(\gamma) = F_\omega(1)$,

$$\beta V_p \approx \frac{2}{b+1} \sqrt{\frac{I}{I_{00}}} \sqrt{b + \cosh \omega}\, F_\omega(1) \tag{88}$$

In this way we retain the correct form of dependence on ω, but throw out the dependence on I that comes from $\gamma(I)$. It can be shown from (81) that

$$\begin{aligned} F_\omega(1) &= \frac{\tan^{-1}\left(\sinh \frac{\omega}{2}\right)}{\sinh \frac{\omega}{2}} \\ &= 1 - \frac{\omega^2}{12} + \frac{\omega^4}{180} + \cdots \end{aligned} \tag{89}$$

Thus we define the characteristic (particle) current density of the device

$$\begin{aligned} I_0(\omega) &= \frac{(b+1)I_{00}}{(b + \cosh \omega)F_\omega(1)^2} \\ &= I_0 \left[\frac{\omega}{F_\omega(1) \sinh \omega}\right]^2 = I_0 g(\omega) \end{aligned} \tag{90}$$

and (88) can be written

$$\beta V_p \approx \frac{2}{\sqrt{b+1}} \sqrt{\frac{I}{I_0(\omega)}} \tag{91}$$

This formula corresponds to (42) with $C = 2/\sqrt{b+1}$. In the spirit of the present theory the exact value of this constant is not important, so we may replace $2/\sqrt{b+1}$ in (91) by C. Then the sum of (87) and (91) gives the total applied bias (66).

Non Linear Recombination

In this section we shall consider the forward characteristic of a *PIN* diode in which the current densities at the contacts obey the law

$$\begin{aligned} I_n &= I_{ns} \left(\frac{n_{00}}{n_p^+}\right)^a \\ I_p &= I_{ps} \left(\frac{p_{ww}}{p_N^+}\right)^a \end{aligned} \tag{92}$$

where I_{ns} and I_{ps} are characteristic of the device and a is a number be-

tween 0 and 1. We see that (30) must be replaced by

$$I_n/I_p = R\gamma^{2a}$$
$$I_p = \frac{I}{I + R\gamma^{2a}} \qquad I_n = \frac{R\gamma^{2a} I}{1 + R\gamma^{2a}} \tag{93}$$

and (23) must be replaced by

$$\begin{aligned} n_0 &= n_i(I_n/I_{ns})^{1/2a} \\ n_w &= n_i(I_p/I_{ps})^{1/2a} \end{aligned} \tag{94}$$

The equation for γ is now

$$\gamma = 1 - \left(\frac{I}{I_1}\right)^{1-(1/2a)} \frac{(\gamma/\gamma_\infty')^{2a} - 1}{[1 + b(\gamma/\gamma_\infty')^{2a}]^{1-(1/2a)}} \tag{95}$$

where $\gamma_\infty' = (b/R)^{1/2a}$ and

$$I_1 = I_0(I_{ps}/I_0)^{(a-1/2a-1)} \tag{96}$$

is a characteristic (particle) current density of the device. We now obtain βV_p from (26)

$$\beta V_P \approx C'(I/I_1)^{1-(1/2a)} \tag{97}$$

where C' is a slowly varying function

$$C' = \frac{(\gamma/\gamma_\infty')^{2a} + 1}{[1 + b(\gamma/\gamma_\infty')^{2a}]^{1-(1/2a)}} \frac{\ell n\,\gamma}{\gamma - 1} \tag{98}$$

similar to the coefficient in brackets in (41). From (21) and (94) we get

$$\beta(V_0 + V_w) = \frac{1}{2a}\,\ell n\,\frac{I_n I_p}{I_{ns} I_{ps}} \tag{99}$$

If now $\gamma \sim 1$ we get

$$\frac{I}{I_{ns} + I_{ps}} = e^{a\beta(V_0+V_w)} \tag{100}$$

This shows how we must choose a to agree with the low current characteristic. On the basis of experience with silicon diodes we would choose $a \sim 0.6$, which would give

$$\beta V_P \sim C'(I/I_1)^{0.17} \tag{101}$$

The characteristic current density would be

$$eI_1 \sim 200 \times (I_{ps}/I_0)^{-2}\text{amp/cm}^2 \text{ in } Si \tag{102}$$

The value to use for I_{ps} is very uncertain, but it certainly is much less than I_0, so $I_1 \gg I_0$. Thus we would not expect to observe βV_P, and the characteristic should have the form

$$I \sim I_s e^{a\beta} \tag{103}$$

up to the highest attainable currents.

We have shown in this section how the law of recombination in the contacts affects the dependence of V_P upon I. In particular if $a = \frac{1}{2}$ there is no dependence of V_P upon I, which means that the conductivity due to injection increases just as rapidly as the current. We may conclude from (97) that the smaller the value of a the more effective is conductivity modulation in keeping down the drop V_P in the middle region.

Discussion

We have considered the *PIN* structure of Fig. 1 having typical parameters given in (3). We find that the presence of the middle region causes no significant deviation in the voltage-current characteristic from that of a simple *PN* diode until very high current densities are reached, of the order of 200 amp/cm^2 in silicon. In particular the middle region is not responsible for an anomalous slope in the plot of V versus log I. We find that recombination in the middle region can be accounted for by replacing the characteristic current density eI_0 of the device with $eI_0 g(w/L)$ where $g(w/L) < 1$ is shown in Fig. 5. Thus qualitatively there is no change in the form of the voltage-current characteristic due to recombination in the middle region, although the effect of $g(w/L)$ is to make the voltage drop somewhat higher than if recombination were absent.

We have suggested that the anomalous slope of V versus log I usually observed in silicon diodes might be due to non-linear recombination. If the recombination obeys a power law chosen to give a typical (anomalous) $V - I$ characteristic for a *PN* diode, we have shown that the *PIN* diode should manifest the same characteristic up to extremely large current densities many times eI_0. Thus the drop across the middle region should be even more negligible with non-linear than with linear recombination.

I am pleased to acknowledge my great benefit from discussions with M. B. Prince and I. M. Ross.

A Laboratory Model Magnetic Drum Translator for Toll Switching Offices

By F. G. BUHRENDORF, H. A. HENNING and O. J. MURPHY

(Manuscript received January 24, 1956)

A laboratory model magnetic drum translator, capable of serving as a one-to-one alternative to the card translator, has been built to study the problems arising from the prospective use of microsecond pulse apparatus in a telephone office environment. Electron tube amplifiers and germanium diode logic circuits supplement the drum information storage unit to provide the functional operations required. Results of preliminary laboratory tests indicate the feasibility of equipment of this kind for telephone switching control.

INTRODUCTION

The magnetic drum is one of the most widely used of the modern large-capacity digital-data storage devices. It is used as a memory unit in many of the present-day large-scale digital computers and in other applications such as inventory control of airline ticket reservations and traffic control of airplanes in flight. Two of the properties of drums as storage media have been considered particularly advantageous. One is the capacity to store up to several hundred thousand bits of information in a compact space at a low cost per bit; the other is the ability to keep the information in an easily alterable but nonvolatile form unaffected by power failure or other interruptions of operation. In terms of the speed with which information may be stored or recovered, drum memories fall near the middle of the present-day spectrum; they are very much faster than punched paper tape or groups of telephone relays but are considerably slower than cathode-ray tube or ferromagnetic-core storage devices. All of the information stored on a drum may be read out during the course of one complete revolution and, similarly, new information may be entered anywhere in the storage space within the time of one revolution; thus the access time is ordinarily of the order of a few tens of milliseconds.

It has already been pointed out[1] that automatic telephone switching

offices bear a generic resemblance to digital computers and it is therefore not surprising that the magnetic drum has engaged the attention of telephone engineers, since the speed and flexibility of such a device offers much promise in connection with forward-looking telephone office design. One system has already been described[2, 3] involving the use of magnetic drums for telephone switching control applications in an entirely new form of telephone office; it is the purpose of this article to describe another application of less complexity which could function in cooperation with equipment in existing telephone offices.

The standards of reliability and ruggedness which must be met by any equipment proposed for Bell System use are in some respects a good deal higher than those imposed on other commercial sytems such as digital computers. Thus when a new type of apparatus such as a magnetic drum and its associated electronic components is considered for a telephone job, it is necessary to determine whether the apparatus is capable of being designed to meet these stringent requirements. This was judged to be the most important objective of the undertaking about to be described, and it strongly influenced the choice of experimental application for the drum.

The program which the designers set for themselves to determine the possible suitability of the magnetic drum type of equipment might be summarized as follows:

(1) Choose an existing telephone application in which a magnetic drum system can receive a satisfactory work-out without disordering the system.

(2) Design a magnetic drum system to work cooperatively with existing office equipment, using existing power facilities. Assume that the design is aimed at practical application so that due regard is given to operating economies, and protection against power failures.

(3) Construct a full-scale model following the design, and test the model in the chosen environment long enough to determine the failure rate and the reasons for each failure.

(4) Evaluate the results in order to determine the sphere of usefulness, and the proper design philosophy for applying magnetic drum systems of any kind in existing telephone offices.

One telephone switching application which meets the qualifications of (1) above exists in the new No. 4A toll switching offices. Here, due to the demands of nationwide dialing, a large-scale translation function is required to convert destination codes into information which will properly route each call. The volume of information which must be stored for translation purposes, and the relatively rapid access desired, fall close to the optimum parameter values of magnetic drum systems. The action

takes place in cooperation with crossbar and other relay-type switching equipment typical of the present-day telephone office, thus providing an environment suitable for observing the behavior of fast pulse circuits in the presence of electrical disturbances. Finally, there exists a relatively new piece of apparatus which now performs the translation function, namely the card translator. Thus, if an exact one-to-one alternative for the card translator were constructed employing a magnetic drum, full advantage could be taken of the testing procedures already developed and a comparison could be made against a norm of performance; furthermore, a field trial would be possible, if desired, with a minimum of interference with normal operation of the telephone plant.

It was decided, therefore, to build a full-scale magnetic drum translator which could substitute for a card translator in order to obtain laboratory experience with apparatus of this type and to determine its adaptability to telephone standards and practices. The completed equipment is shown in Fig. 1. The equipment on the one frame illustrated is the equivalent in function and capacity of one card translator with its associated table. This magnetic drum apparatus is not aimed at replacing the card translator, which is a well-engineered device known to give satisfactory service in day-to-day operation. For evaluation purposes in this article, however, it is assumed to be competing with the card translator.

The following sections describe the design features and operating details of the translator which was constructed. A brief description of the card translator and that portion of the 4A office in which the drum translator must function has been included to provide the necessary background for the description. It will become evident that the requirement of interchangeability which necessitates a one-to-one equivalence with the card translator has imposed on the drum translator a number of restrictions which are not inherent in it. These tend to prevent full exploitation of the speed and code advantages which might be realized with the drum. Furthermore, the rapidity with which all of the information on the drum is presented on a continuous read-out basis would permit a type of centralized operation which will be touched on briefly and which would seem to offer apparatus economies not attained in the test model. None of these factors, however, impairs the usefulness of conclusions which may be drawn from test results concerning reliability.

SURVEY OF MAGNETIC RECORDING PRINCIPLES EMPLOYED IN THE TRANSLATOR

All magnetic drums have certain features in common: they consist of a means of moving a thin shell of magnetically-hard material rapidly

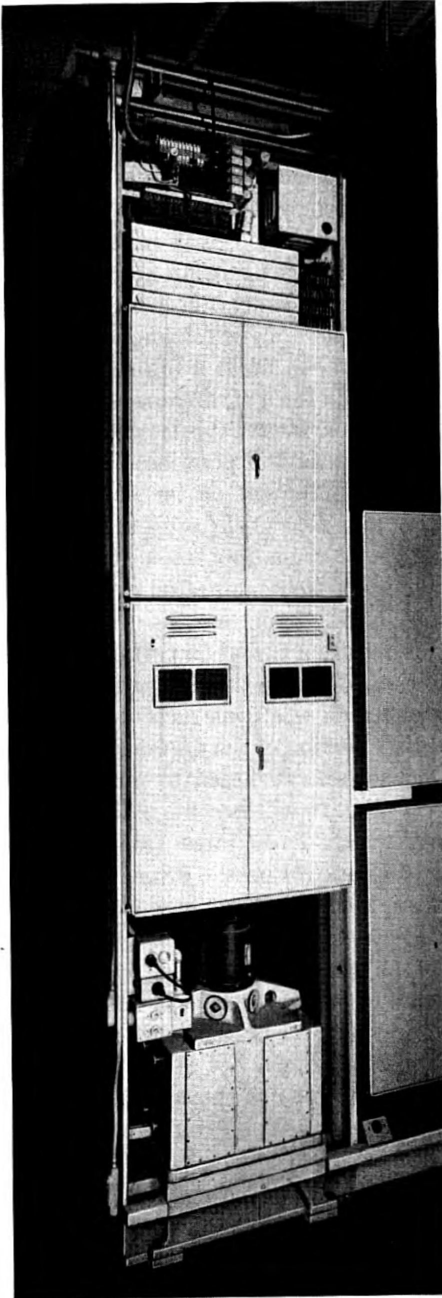


Fig. 1 — Magnetic drum translator, laboratory installation.

past one or more heads used for writing or reading digital data. Usually, as in the translator, the same head is used for both functions. In most drum-system designs the pole-tips of the heads are close to the recording surface but do not touch it, and the heads themselves bear a resemblance to those used in conventional magnetic sound recording, giving therefore, a "longitudinal" polarization to the medium as sketched diagrammatically in Fig. 2. There is very little further resemblance to sound recording, since digital information is stored in a binary or two-valued code which, on the translator drum, is represented by the two possible polarities of saturation of the magnetic medium. To one of these polarities is assigned the code value "*0*," and this condition prevails except where the opposite polarity is inserted to represent the code value "*1*."

It should be mentioned that several other systems have been devised which employ the two directions of saturation, sometimes accompanied by a general background of magnetic neutrality, to effect a greater concentration of digital information than that used in the translator. Systems other than the one chosen for this application were, for the most part, considered to be less reliable.

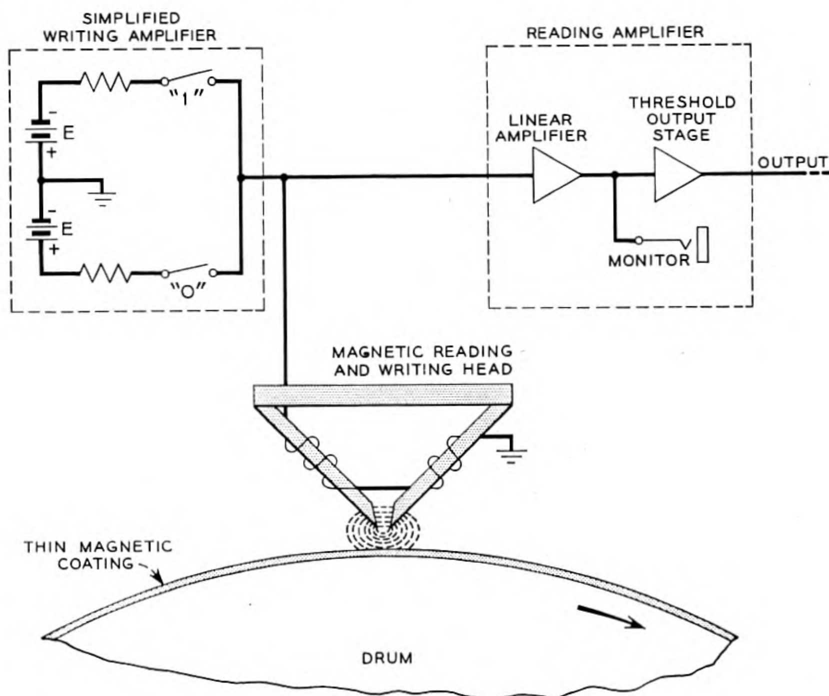


Fig. 2 — Simplified diagram of magnetic drum digital data storage system.

In order to facilitate an understanding of the action of the translator as a whole, a simplified account of the magnetic recording and reproducing process will now be given.

Magnetic Drum Geography

The circumferential strip of the drum surface which moves under the pole-tips of any magnetic head is commonly known as a *track*. On each track will be written magnetic perturbations or spots symbolizing "*1's*." It is essential that these spots be precisely located so that they may be readily removed or "altered." For this purpose a synchronizing track or some equivalent distribution of equally spaced identifying marks associated with the drum is provided. With the aid of the electronic circuits, the magnetic spots are restricted to a modular spacing defined by the synchronizing marks, and this module is spoken of as a "*slot*." On the drum surface, each intersection of track and slot is known as a "*cell*" and a cell may contain only one magnetic mark and therefore only one bit of information. As a matter of economics, the cell density should be as great as possible. The density which may be attained is determined by the degree of interference which can be tolerated among neighboring cells.

Writing Operations

The first step in preparing the drum to receive a recording is to uniformly magnetize the tracks to saturation in the polarity arbitrarily chosen to represent the code-value "*0*." This is a preconditioning operation required only when a drum is newly placed in service. Referring to Fig. 2, this may be done, for the typical head and track shown, by closing the switch marked "*0*" for the duration of at least one complete revolution of the drum. Enough current must flow through the windings of the head to establish the magnitude of fringing flux, from the pole-tips, required to saturate the thin magnetic coating. In the case of the translator drum, the coating is about ⅓ milli-inch thick; the clearance between pole-tips and recording surface is about 2 milli-inches; the interpole gap is also about 2 milli-inches at the tips, and about 20 ampere-turns of energization are required.

With the track thus preconditioned, there is virtually no output voltage from the head since the magnetization is essentially uniform and there is no changing flux threading the head to induce a voltage in the windings.

Whenever a "*1*" is to be written, a pulse of current from an electronic writing amplifier (indicated, for convenience, on Fig. 2 as a switch) is

caused to flow through the windings of the head in a direction opposite to that taken by the preconditioning current. This pulse lasts for only two or three microseconds, and movement of the drum surface is negligibly small while the current persists. The peak value of the current pulse is sufficient to magnetize to saturation in the opposite direction that portion of the track which lies directly under the pole-tips at that instant. Areas of the track far-removed in each direction from the pole-tips of the head are, of course, unaffected by this operation, and remain at saturation in the original polarity. A region of transition in magnetization therefore extends in each direction along the track from the area directly under the pole-tips.

Fig. 3 illustrates some of the wave forms resulting from writing into and reading from four adjacent cells on one track of the drum. Line A shows the pulses of writing current which were applied to the windings on the head. These were caused to appear at precisely spaced distances along the track by the combined operation of the synchronizing system and an "administration" circuit. In cells 1 and 3 the writing current polarity is chosen so as to write "*1's*." Cell 2 remains in its original preconditioned state. In cell 4 a "*1*" was previously written but is now altered to a "*0*" by a writing current pulse of the same polarity as that chosen for the preconditioning operation.

Line B in Fig. 3 illustrates the resultant magnetic state of the drum surface as viewed by the reading head. The polarization portrayed as resulting from writing a "*1*" is a bell-shaped curve. When a "*1*" is selectively altered to a "*0*" the area of track directly under the pole-tips will be carried to saturation in the original preconditioned polarity. The whole cell area, however, cannot be affected so strongly, owing to the hysteresis properties of the coating material, and there will remain traces of the "*1*" type of magnetization near the cell edges, as indicated by the solid line in cell 4.

There is no difficulty in rewriting a "*1*" in a cell which has been subjected to the above described treatment. The procedure is that outlined for the original writing of a "*1*" and the results are practically indistinguishable from those obtained by writing in a virgin cell.

Reading Operations

On subsequent revolutions of the drum, the passage, under the pole-tips, of the magnetic irregularities created by writing "*1's*" will induce a change of flux through the windings of the head. The change is, of course, a function of distance along the drum surface but since the drum is rotating continuously at a substantially uniform speed the change

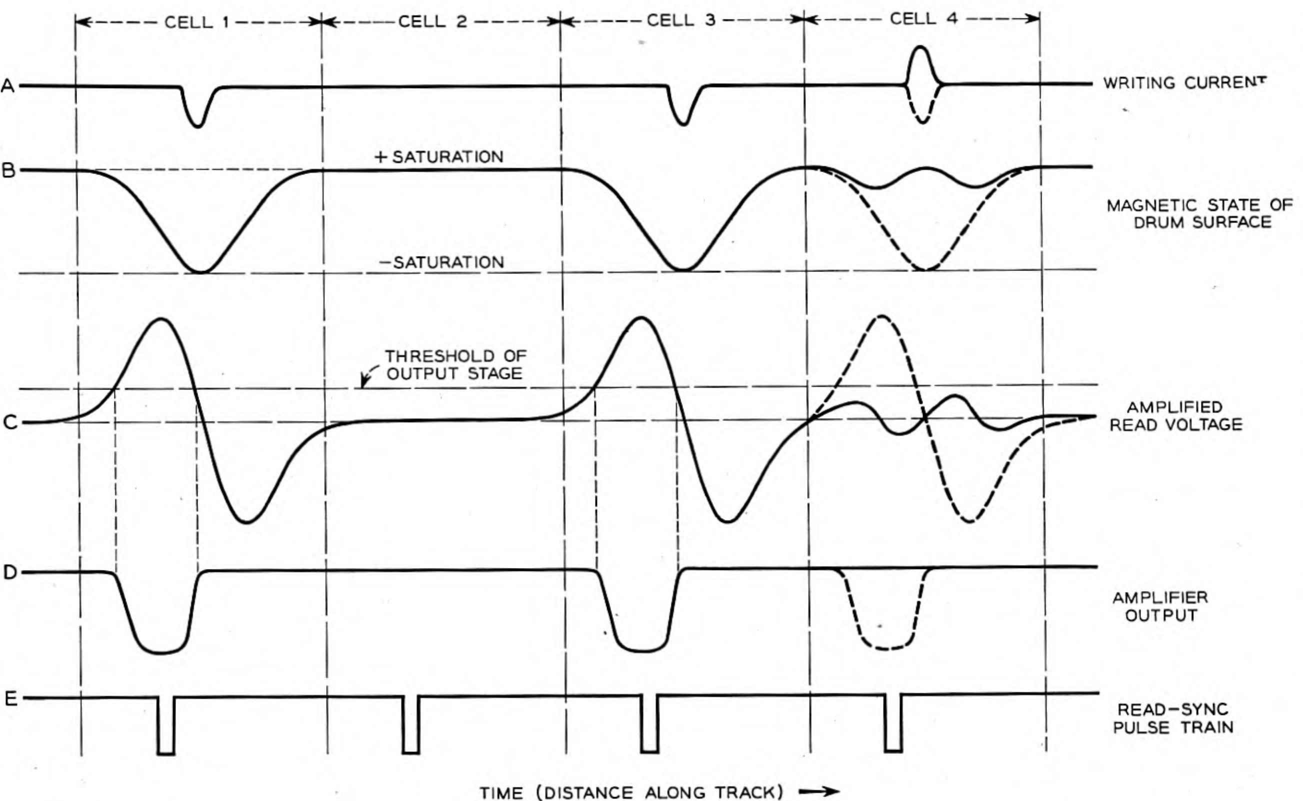


Fig. 3 — Magnetic writing and reading waveforms.

may also be represented as a function of time. This time-rate-of-change of flux within the coils of the head generates a voltage which is of the order of 50 millivolts peak-to-peak in the case of the translator. This voltage, after amplification, appears as shown in line C of Fig. 3. The trace shown is that which appears at the "linear output" monitor jack of a translator reading amplifier, and includes a phase inversion, characteristic of a three stage amplifier. Such a curve is readily recognized as being quite similar in shape to the first derivative of the normal error-function and hence we may infer that the magnetic condition of the drum surface, at least as interpreted by the head, may be portrayed by a bell-shaped curve, previously mentioned, similar to the error-function itself.

The residual magnetic irregularity pictured in cell 4 resulting from writing a "*0*" over a "*1*" will induce a voltage in the winding of the head having a different amplitude and wave shape from that occasioned by reading a "*1*." It is sketched out approximately to scale in Fig. 3 and is seen to be a smaller twinned-version of the "*1*" signal. Its amplitude ordinarily lies in the range of $\frac{1}{10}$ to $\frac{1}{3}$ of that of the "*1*" signal, and for about the middle third of the cell its instantaneous polarity is opposite to that which a "*1*" signal would have. These facts suggest at least two means of discriminating between the voltage signals obtained for the two code values: (a) on the basis of amplitude difference, and (b) on the basis of instantaneous polarity difference determined or sampled within a particular epoch in each cell.

The method adopted for the translator is that of simple amplitude threshold. The threshold value indicated by the dotted line in Fig. 3, is set so that the strongest of the residual signal outputs never exceeds it while, at the same time, the greatest possible proportion of the positive-going lobe of a "*1*" signal is allowed to produce an output. The threshold output stage of the amplifier is also arranged for limiting and this has the effect of blunting the peaks of the applied signals. The over-all result of these actions is shown by the shape of the signals in line D of Fig. 3.

Cell packing may be of major economic importance in a large installation. The general effect of making recordings closer and closer together is that the presence or absence of one of the recordings in a series has an increasing influence on the size and shape of the signals reproduced from its neighbors on either side. In the translator, the cells are spaced 20 milli-inches center-to-center along the track and the influence of action in one cell on the amplitude of reproduction from neighboring cells is never more than about 10 per cent. The trace of line C, Fig. 3, is drawn for this cell spacing and shows a slight inflection at the transition between the output voltage occasioned by reading cell 3, and the voltage obtained

from the "*1*" which was originally written in cell 4. In many applications a much larger "influence factor" may be tolerable, but this usually requires greater elaboration of the signal detecting devices. The cell size is also influenced by physical constants such as design of the head, properties of the medium, and dimensional clearances. A discussion of such factors is outside the scope of this paper but it is not unreasonable to hope for an improvement of two-to-one in packing factor in future designs.

Reading Synchronization

The magnetic drum used for the translator provides 80 tracks. About sixteen microseconds is required for each cell in a track to pass under its head. Information occupying the same slot on the drum (so-called because of its obvious relationship to the term "time-slot" commonly used in the digital computer field) is presented at the various heads essentially, but not exactly, simultaneously. Departure from exact simultaneity is occasioned by small variations in the shapes and amplitudes of the output waves shown typically as line C in Fig. 3, and by small time-variations occurring in the writing process, as applied to the various tracks.

To achieve exact simultaneity, as required for certain subsequent operations of the translator circuitry, narrow "Read Synchronizing" pulses are produced by the synchronizing circuit previously mentioned. These pulses are located, within the time boundaries of the cells, so that they fall approximately at the center of the broad output pulses from the reading amplifiers and thus permit the latter to be sampled. This relationship is indicated in lines D and E of Fig. 3. Similar pulses, slightly displaced in time, are used to control the writing operations, and are designated "Write Synchronizing" pulses. The necessity for the time-shift is apparent from an examination of lines A and E of Fig. 3.

This condensed explanation of the technology of magnetic drum digital data storage devices, particularly as applied to the translator drum, should serve as sufficient background for the description of the translator wherein the drum is but one part of a large ensemble of apparatus.

THE JOB WHICH THE CARD TRANSLATOR NOW DOES

It will be advantageous to examine very briefly the card translator and its functions in the No. 4A toll switching system so that the analogous operation of the magnetic drum equivalent may be more readily explained. A more detailed description is given in Reference 4.

The demands of nationwide toll dialing require a very extensive rep-

ertoire of translations between destination codes and routing instructions, and it must be possible to change the routing instructions with ease. The card translator fulfills these requirements. Each individual translation item is contained on a metallic card; the output code of routing instructions is in the form of selectively enlarged perforations in the perforated field of the card, arranged so as to be read by photoelectric means, and the input code, which identifies the card for purposes of selection, appears in the form of tabs projecting downward from the bottom edge. Each card is capable of holding a total of 154 bits of information, input and output, and somewhat over 1,000 cards are stacked in a bin in each card translator mechanism.

It is possible to classify the elements of any translator into three broad categories: the memory unit, the translation selecting unit, and the translation delivery unit. In the card translator the memory unit is, of course, the group of cards; the translation selecting unit consists of code bars, electro-mechanically actuated, for displacing a selected card sufficiently so that it may be "read." It also contains a network of relays which perform the function of checking the authenticity of the input codes applied to the code bars. The translation delivery unit consists, in the main, of a number of output channels, each originating with a light beam for probing one of the code elements (a bit of output information) on the card. Each output channel contains a photo-transistor, a transistor amplifier, a cold cathode gas tube circuit which has been designated a "channel output detector" and a register relay. The register relays perform work functions and therefore are located separately from the translator; some are in the decoders, others in the markers.

In the 4A office, the card translator is one of several items of common control equipment which cooperate to establish the talking connections. Other items are the sender, the decoder, and the marker. The sender receives and registers and subsequently transmits the decimal digits of the called designation; the decoder receives the code digits (from 3 to 6 in number) from the sender and submits them to the translator for conversion into information needed for the proper routing of the call; and the marker selects an outgoing trunk and establishes a transmission path by operating the crossbar switches. Since this common control equipment is associated with any one call for only the short interval necessary to establish the talking-circuit connection, its speed of operation is a matter of considerable importance.

It is obvious that the decoder is the intermediary between the translator and the remainder of the office. Each decoder, of which there are a maximum of 18 in a large office, has exclusively associated with itself a

card translator mechanism; each of these mechanisms contains an identical repertory of translations. Each decoder also has available, through connectors, a common pool of translators containing a large quantity of less-often used information. In order to better understand the duties that a magnetic drum translator must be expected to perform it will now be convenient to follow, in a highly abbreviated manner, a typical operation of the decoder and its associated card translator.

The first translation on an incoming call is performed using the first three decimal digits accumulated by a sender. As soon as three digits are available the sender connects to a decoder which immediately signals its individual translator to perform certain mechanical chores in preparation for selecting a card. There are several sequencing signals between the decoder and translator during the complete cycle of a translation (several of these signals must be synthesized by the drum translator); acting on one of these signals from the translator, the decoder passes the input code from the sender, adding certain supplemental information of its own.

The three decimal digits of the input code are in checkable combinations of two leads energized in each of three groups of five leads connected to the translator. The supplementary information supplied by the decoder is in a similar checkable combination on six leads. None of the remaining leads in the total of 38 is energized, since the translation being described involves only three code digits.

In the translator, the input code actuates the card selecting mechanism and also operates relays whose contacts are wired with a checking network which confirms that the input code, and the responsive operation of the code bars, is an authentic combination. This is done by establishing a path to operate a "code bar check" relay, CBK. (This relay retains the same identity in the magnetic drum translator.)

Acting upon the authenticity check, the card translator proceeds to select a card, and signals the decoder to begin timing for a possible non-appearance. When the card is in a position to be read, the decoder is signaled on two "index" channels, IND. The decoder now "reads" the card by applying 130 volt battery to the coils of its register relays; the required relays operate through the ionized cold-cathode gas tubes in the translator, and lock up, extinguishing the gas tubes.

The first card dropped may provide information sufficient for completing the connection; in this circumstance the decoder will then call in a marker. The first card, however, may specify that more digits are required and the decoder will so instruct the sender. The sender, unless it already has the necessary digits, is then dismissed by the decoder which also instructs the translator to restore itself to normal.

Six-digit translations are obtained in a manner similar to that described above except that the checking network on the relays is switched to check for six rather than three digits. In some instances the decoder must refer to one of the translators in the common pool of "foreign area translators" in order to obtain the required information. Frequently, several different cards must be dropped successively before a route is finally established for the outgoing call.

With the above description as a background, we may proceed to discuss the magnetic drum translator.

THE ANALOGOUS FUNCTIONS OF THE MAGNETIC DRUM TRANSLATOR

The magnetic drum translator is essentially a device which performs a translation by making a selection from a recurrent pattern of electrical pulses generated by a magnetic drum unit. A schematic diagram of the magnetic drum translator, as arranged for direct substitution for a card translator, is shown in Fig. 4. In this diagram, the system is divided into three principal functional components: (a) the drum memory assembly which produces (from the outputs of 80 reading amplifiers and a timing unit) a repetitive pattern of electrical pulses representing all the translations on the drum, both input codes and corresponding output codes; (b) the translation selecting unit which reads that portion of the pulse pattern representing input codes and acts to identify the unique code group which matches the incoming information from the decoder; (c) the translation delivery unit which, under control of the translation selecting unit, gates-out the particular pulses of the corresponding output code from the continuous stream of microsecond pulses, and converts them into signals capable of operating the register relays in the decoder.

To maintain direct interchangeability, two items of apparatus were adopted virtually without change from the card translator. These are the CODE CHECK RELAYS which accept and check input information, and the CHANNEL OUTPUT DETECTORS comprising cold-cathode gas tubes and associated transformers. This allows input and output terminal facilities to the decoder to be the same for both translators.

It should be noted that the magnetic drum memory assembly differs significantly in one functional respect from the binful of cards in the card translator. When a selected card is being read by the photo-electric cells in the output channels, no other cards are available. In the drum translator, all translations are continuously available and if a number of translation selecting and translation delivery circuits are employed, all may obtain translations from a common drum memory assembly at the same time without interference. This feature could not be demonstrated in the

test set-up as planned, but it would have been incorporated in any test which included more than one decoder in an office. In such an arrangement, the various units illustrated in Fig. 2, except the drum memory assembly, would be furnished to each decoder. One drum memory assembly (and an emergency standby) would supply the pattern of electrical

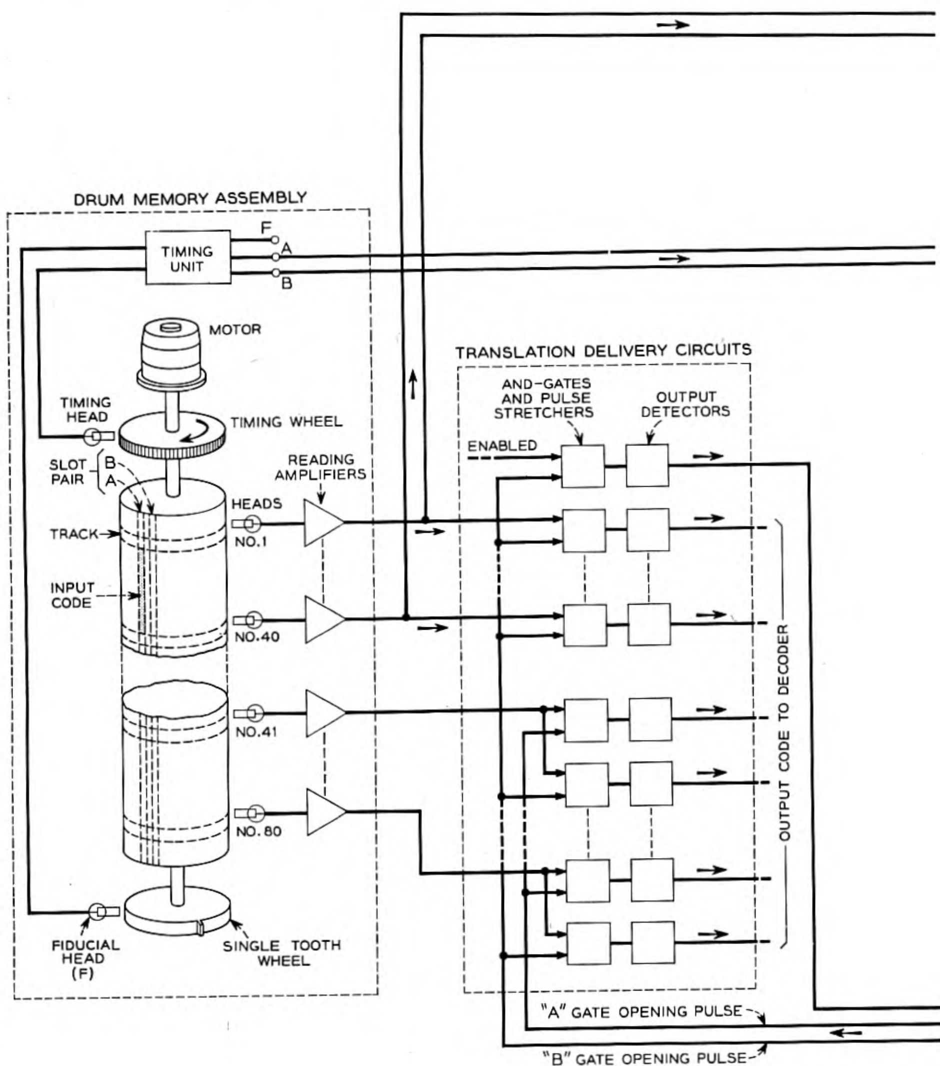
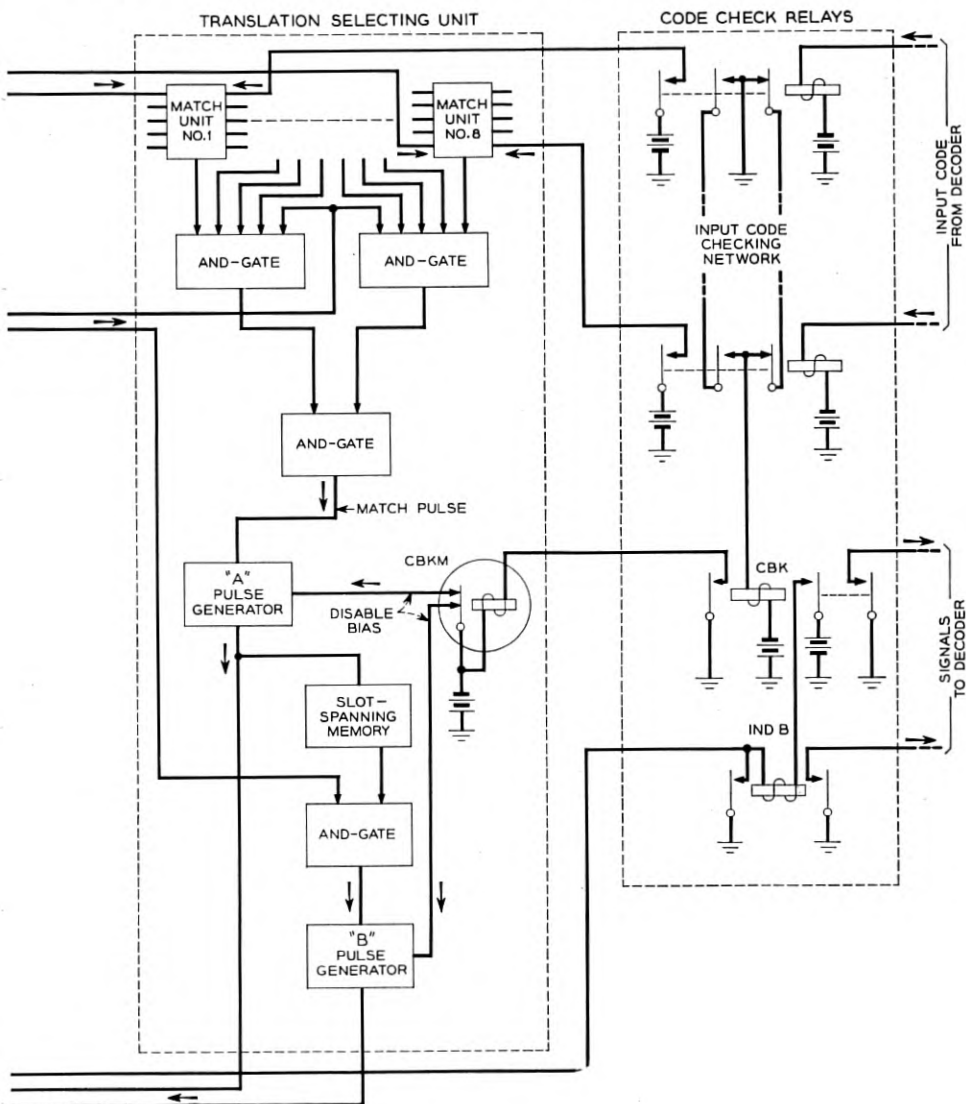


Fig. 4 — Magnetic drum

pulses to all translation selecting and translation delivery circuits in multiple. The object of such an arrangement, naturally, is to employ the magnetic drum system in the most economical manner. A further extension along the same lines would involve relay switching of the pulse circuits



translator block diagram.

to give access to the emergency drum memory, or to a "foreign area" memory where such extra memory capacity is necessary.

Let us now return to the discussion of Fig. 4 and consider the assignment of the translation information to the drum surface where it is stored. Recall that the drum surface is effectively divided into a grid by the coordinates of tracks, each passing under an individual write-read magnetic head, and "slots," each defined by the appearance of a timing pulse in a rhythmic train synchronized from the drum itself, and that the "cells," at the coordinate intersections, each accommodate one bit of code information.

Since each card in the card translator accommodates 38 bits of input code and 116 bits of output, about 160 cells, divided in the ratio of one cell for input to every three cells for output, must be assigned to each translation item. One simple and direct assignment would be to place the entire translation item in a single slot composed of 160 cells. With this layout the slot containing the desired translation would be identified by reading, or "matching" the input code, and during this same interval the output information in the same slot would be gated-out to the translation delivery circuits. A 1,000-translation drum would then be long and narrow, and far too many reading amplifiers would be required. Another evident arrangement would be to assign the entire input code to the first of each group of four slots proceeding under the heads, with the output code following in the next three slots. Such an allocation would require only 40 reading amplifiers but the drum necessary for the desired capacity, with the cell-spacing chosen, would have been larger in diameter than the mechanical designers cared to undertake in their first trial. A logical choice, therefore, was to place each translation item in a pair of adjacent slots, and this was done, although it was later recognized that other, more sophisticated, arrangements might offer certain advantages.

In Fig. 4, the apparent location of one translation item is sketched in relation to the drum surface. This sketch is not drawn to scale, since the slot width is actually only 0.020 inch, and the track width is comparable. It is also geographically inaccurate; actually the cells of any one slot are positioned in four quadrants on the drum, the associated heads being positioned in four stacks for mechanical reasons. However, all of the cells in a time slot pass under all of the heads at the same instant and the presentation of Fig. 4 was adopted for the sake of clarity.

Note, then, that the input code and one-third of the output code are recorded in the first or A slot of a slot-pair passing under the reading heads, and that the remaining two-thirds of the output code occupies

the B slot which immediately follows. The parallel (simultaneous) presentation of the entire input code to the translation selecting unit permits that unit to indicate, by a pulse, that the translation item is the one desired and to gate-out the output code in the same slot while it is still passing under the heads. Having thus identified the first slot of a translation item, it is a simple matter to provide the facility for gating-out the remaining information recorded in the next succeeding slot.

It will be seen, from the circuit arrangement shown, that the translation selecting unit also receives a portion of the output code recorded in the second slot of each pair. It is therefore necessary to distinguish between the A and B slots of a pair. This is most conveniently done by the Timing Unit, which is provided with two outputs, the pulses defining the slots appearing alternately at these outputs. One output lead is chosen to define all the A slots and it is routed to the translation selecting unit to provide a portion of the pulse-pattern required for complete and proper identification of an input code.

The action of the magnetic drum translator in making a translation may now be traced by following the block diagram of Fig. 4. The decoder, of course, gives the same preliminary signals as for the card translator, but these are ignored by the drum translator, because it is continuously presenting all 1024 translations at the rate of 30,000 per second and need not take any preparatory steps, provided its relays have returned to normal after the last translation. The normal state of the relays is checked by means of a circuit through their contacts; if this circuit is complete, the decoder receives the signal to apply the input code as soon as it seizes the translator. A more elaborate checking arrangement could have made this signal conditional upon other tests, such as a "standard translation," to determine that the electronic circuitry (in bulk) was functioning properly, but it was not considered worthwhile to do so in the system described here.

The decoder, then, furnishes the input code of the desired translation item, causing certain of the relays labeled CODE CHECK RELAYS in Fig. 4 to operate. Contacts on these relays are interwired to provide the same checking network as in the card translator, and a check on the authenticity of the input code will be evidenced by operation of the relay labeled CBK. This event is signaled to the decoder so that it may start its "no-card" timer action. When CBK closes, it also operates a chatter-free mercury-contact relay, CBKM, in the translation selecting unit, permitting that unit to produce an output at the appropriate time. Each code-check relay which operates applies a positive voltage to one of the input terminals of a "match" unit in the translation selecting unit. For each of

these input terminal there is a complementary terminal to which are applied negative-going pulses from one of the drum memory reading amplifiers. As will be explained later, advantage is taken of this complementary arrangement to obtain a signal indicating a match between either, (1) an operated code relay and a pulse from the reading amplifier, or (2) a nonoperated relay and no pulse from the reading amplifier. All of these signals, from 40 sections of the match units, are combined in a cascade of "AND" gates; when all indicate a match, the translation selecting unit delivers an output "match" pulse.

Since this match pulse is not strong enough to enable 40 gates in the output channels, it is passed to a "pulse generator" (a regenerative pulse repeater) which produces, virtually coincident in time, a powerful "A" gate-opening pulse. Note that both the "A" and the similar "B" pulse generators are enabled to operate only when the input code is authentic, as evidenced by the operated code check relay CBKM.

In an unrestricted magnetic drum translator design this identifying pulse would cause immediate registry of part of the desired information. Here, however, is evidenced one of the penalties for having a direct one-for-one substitution for a card translator. The decoder and card translator function in a definite sequence; one of the steps in this sequence is initiated by the IND signal from the translator which informs the decoder that the selected card is properly "indexed" so that it may be "read." Therefore, in the case of the drum translator, to preserve this sequence, the selected translation is permitted to pass unheeded, except that the IND signal is synthesized from the identifying B gate-opening pulse. This operation closes one relay, INDB, through a special output channel (topmost one in Fig. 4) provided for the purpose. The decoder, thus notified that the desired translation is available, applies battery to its register relays, and the output channels are completely enabled for a subsequent registry of the desired information.

The output information is usually registered during its next passage, one drum-revolution after initial identification of the item. The action of identifying the translation is again as described above, and there remains only to follow the operation in the output channels. Even before the translation selecting unit has initiated the identifying gate-opening pulse, reading amplifiers which are required to deliver an output code have each commenced delivery of a pulse to their corresponding gate terminals in the AND gate and pulse stretcher units. (See Fig. 4). When these pulse signals have reached a stable maximum, the gate-opening pulse (A or B depending on the slot which is being read at the moment) is free to pass through the gates and to trigger the pulse stretchers. The

latter devices, each containing a single transistor in a monostable circuit arrangement, deliver 12-volt pulses lasting about a millisecond. The pulse stretchers from which an output code is not required are not triggered, owing to the absence of pulses from the corresponding reading amplifiers.

The remainder of the output channel, as previously stated, is borrowed directly from the card translator, and the action is similar. In the output detector, a transformer steps-up the 12-volt pulse signal to a voltage more than sufficient to establish a discharge in the control gap of a cold-cathode gas tube. Since the decoder has applied voltage through a relay coil to the main gap, the discharge transfers, and the resultant current flow operates the relay. The operated relay, which may be in the decoder, registers the code and locks to ground through an auxiliary contact. This action also extinguishes the gas tube, thereby extending its life.

Except for relay operation, all of the activity described here for two drum revolutions repeats itself for every subsequent drum revolution for as long as the code check relay CBKM remains operated. However, once the code is registered, no further use is made of the pulses in the output channels.

When the decoder has made use of the translation, it transmits a signal which is used in the code-check relay system to indicate when all relays are properly restored. In the card translator this signal is also used to restore the selected card, but in the drum translator this operation, of course, is not required.

Administration Equipment

To utilize the magnetic drum translator as described above, it is obvious that some means for writing-in the translations is as necessary to the drum as a card punch is to the card translator. Although a selective writing, or "Administration Unit" was required, a highly efficient design was not essential to the experiment. Consequently there was constructed a separate, portable aggregation of essential basic electronic circuits, arranged for manual control, but designed with a view to possible extension to fully automatic operation. This equipment will be described in a later section.

EQUIPMENT AND CIRCUIT DESIGN DETAILS OF THE TRANSLATOR

General Description

The entire translator is mounted on an 11-foot by 32-inch bay and has been made to conform to telephone central office practices as far as pos-

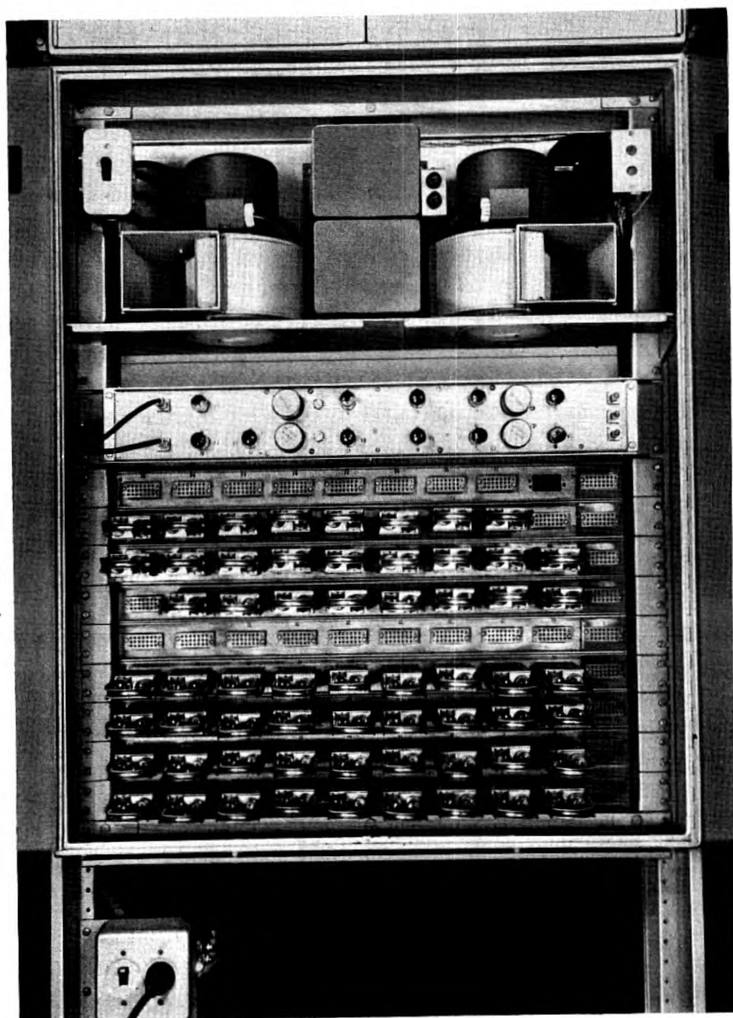


FIG. 5 — Lower casing containing partial complement of reading amplifiers, timing unit, filament transformers and blowers. Receptacle at right end of each amplifier mounting strip allows Administration unit to connect directly to magnetic heads associated with those amplifiers.

sible; except for the presence of the drum unit at the base of the rack, its appearance is not unlike that of other racks found in central offices.

Mounted directly above the drum unit is a casing of conventional design (shown open in Fig. 5) which houses the reading amplifiers, timing unit, filament transformers, and a self-contained forced-air ventilating

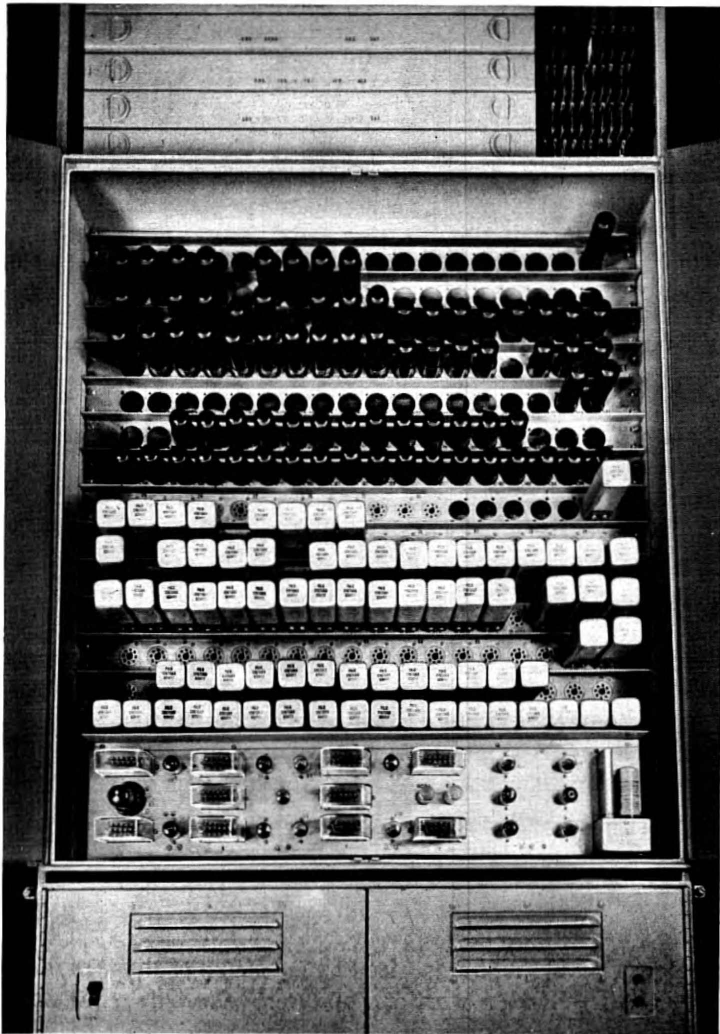


Fig. 6 — Upper casing containing translation selecting unit, and partial complement of pulse stretchers and channel detectors.

system. A second casing, (Fig. 6), located directly above the first, houses the translation selecting unit, pulse stretchers, and channel output detectors. The various plug-in components used in these sections are shown in Fig. 7. At the top of the rack are located the code-check input relays, fuses and terminal blocks.

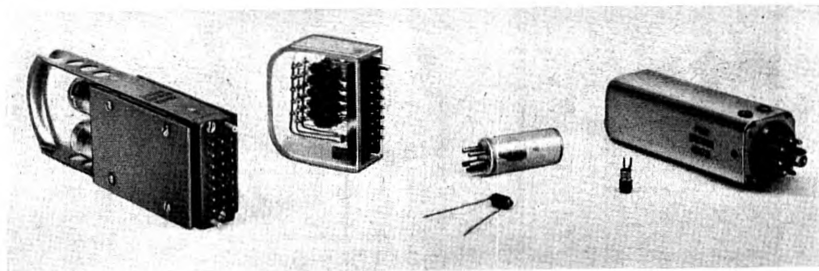


Fig. 7 — Plug-in units. Left to right, reading amplifier, match unit varistor cluster, individual varistor, match and-gate, transistor, and pulse stretcher.

In wiring the rack, use of individually-shielded conductors was held to a minimum. The cable between the drum unit and the reading amplifiers was composed of standard switchboard wire, shielded as a unit by removable sheet-metal enclosures, thus greatly reducing the bulk as compared to the usual bundle of coaxial cables.

The remainder of the wiring, which carries relatively high-level signals from unit to unit within the frame was also in the form of cables of switchboard wire; this type of wiring was tried as an experiment for microsecond pulse work, and was found to be successful in this instance.

Under normal conditions the entire translator, with the exception of the tube filaments and drum drive motor, operates from the standard plant batteries of +130 and −48 volts. Commercial 60-cycle power is normally used for filaments and motor; the motor is duplex and is designed to transfer automatically to the 48-volt plant battery in case of power failure, and the same provision would have to be made for the filaments in the event of a telephone plant installation.

Magnetic Drum Unit

The magnetic drum unit is located at the bottom of the rack, as shown in Fig. 1; a close-up view with one of the covers removed is shown in Fig. 8. A mounting casting supports the machine directly on the floor, straddling the lower member of the rack so that no load is imposed on the rack structure. The drum rotates about a vertical axis and is housed in two cast-iron end-bells spaced by a cast-iron shell. The end-bells carry the bearings for the drum, and serve to mount the motor, while the shell-casting rigidly locates the magnetic heads, each very close to the drum surface. This design requires a minimum of floor space, insures accurate bearing alignment, provides a convenient location for the magnetic heads, and permits the use of tightly-fitting gasketed covers to exclude

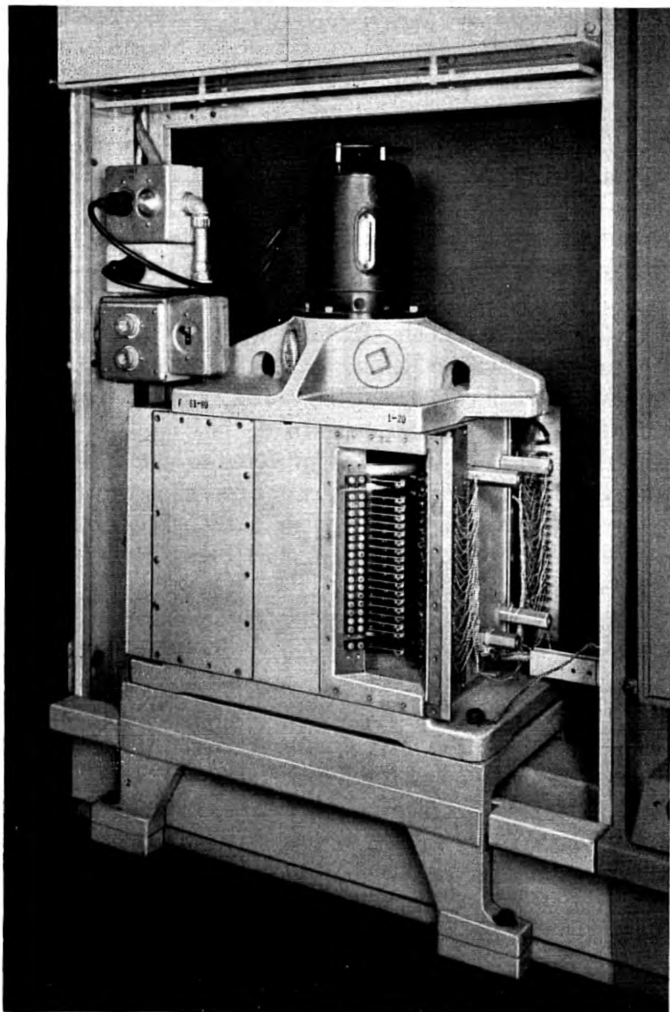


Fig. 8 — Magnetic drum unit partly uncovered to show magnetic heads and wiring terminals.

dirt and foreign material from the magnetic drum surface and the bearings. The $\frac{1}{16}$-hp motor drives the drum through a spring-diaphragm coupling.

The drum is comprised of a stress-relieved iron casting of high dimensional stability, a press-fitted steel shaft, and a $\frac{1}{16}''$ thick brass outer shell which carries the magnetic recording medium. Since both drum and

housing are of similar materials, and have almost identical temperature-expansion coefficients, it is expected that pole-tip-to-drum clearance will remain unchanged under normal conditions of service. The drum, which is 12.8″ in diameter, 10″ long, and weighs 150 pounds, is dynamically balanced and runs without sensible vibration.

Commercial super-precision angular-contact ball bearings, two at each end, are used to mount the drum in its housing. The lower bearings are arranged to share the thrust load imposed by the weight of the drum, and the upper bearings are mounted opposing each other, and are pre-loaded one against the other. The upper bearings serve only as radial constraints, the outer races being free to move axially. This type of construction results in a finished unit having a total runout of only a few ten-thousandths of an inch without the necessity of machining the drum on its own bearings. For the experimental installation, the bearings were grease-packed at assembly and can be expected to function satisfactorily during any reasonable test period. If, however, such a drum unit were made a permanent part of the telephone plant, other provisions have been considered which would insure adequate lubrication over a much more extended period.

The magnetic coating used on the drum is an electro-deposited alloy of cobalt and nickel (90 per cent Co–10 per cent Ni) approximately 0.0003″ thick. This coating was selected because of its hardness, strength, uniformity, and desirable magnetic characteristics. The thickness of the coating is such as to result in a satisfactory cell-size without undue sacrifice in output. The purpose of the brass sleeve mentioned previously is to form a nonmagnetic surface between the magnetic coating and the cast-iron core since, if the coating were applied directly to a ferro-magnetic material, its effectiveness would be greatly reduced by the shunting effect of the base material. The brass sleeve also serves to facilitate plating the drum, since brass, unlike cast-iron, is amenable to the electroplating process.

Read-Write Heads

One of the read-write heads is shown in Fig. 9. The magnetic structure consists of three rectangular bars of laminated material, arranged in the form of a triangle (as schematically represented in Fig. 2). Two legs of this triangle carry single-layer coils which are series-connected. These two legs also serve as pole-tips, being pointed at the end and separated by an air gap. The third leg serves to complete the magnetic circuit and, in assembly, is butted tightly against the other members by means of a leafspring.

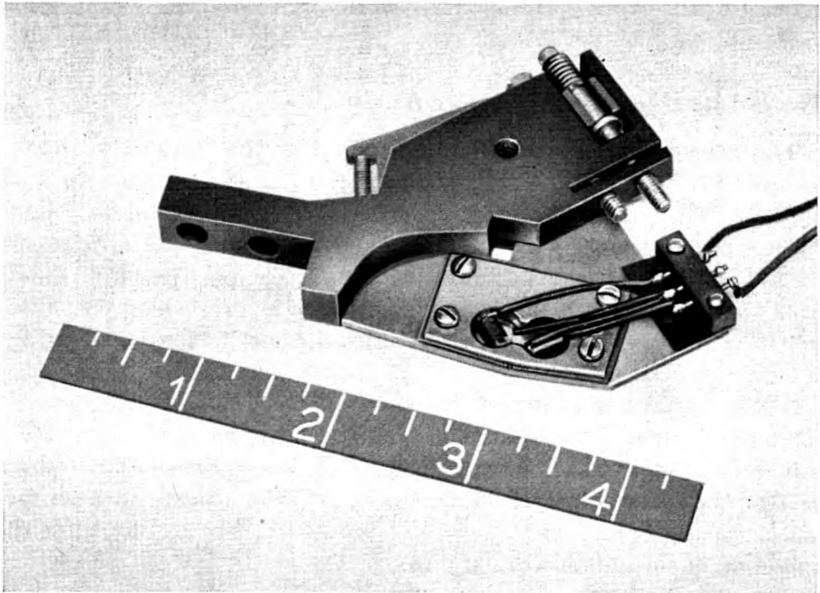


Fig. 9 — Magnetic head and mounting bracket showing means of adjustment.

The magnetic structure is assembled on a nickel-silver plate to which have been soldered two copper shoes which serve to locate the pole pieces and shield the pole-tips, thereby focusing the recording flux to some degree. After adjustment of the pole-tips, the assembly is clamped in a sandwich by means of a second, smaller nickel-silver plate. As is evident from the illustration, this magnetic assembly is in turn assembled to a mounting bracket which contains facilities for precisely adjusting the clearance between pole-tips and drum surface.

The pole-tips of the head are 0.050″ wide and the tracks are on 0.10″ centers, leaving a nominal value of 0.050″ between tracks to allow for misalignment of heads and for flux-spreading. Heads which are physically adjacent in each of the four corner stacks are mounted on 0.40″ centers, but the stacks are offset with respect to one another, thereby interlacing the tracks on the drum.

The read-write heads have been designed expressly for use in high-speed digital recording. Very thin laminations are used and this, coupled with carefully prescribed manufacturing techniques, results in a head having a satisfactory frequency response for the very short pulses employed. When used as a transducer to convert electrical pulses to mag-

netic flux, it is capable of responding faithfully to frequencies approaching ten megacycles per second.

The Timing Wheels and Associated Heads

The synchronizing pulses derived from the drum originate from a 512-tooth soft-steel gear mounted at the top end of the drum. In combination with a polarized reproducing head, the gear generates a timing signal which provides means for permanently locating the various cells used to store information on the drum surface. The polarized head differs from those used on the drum proper, being of a form which is conventional in tone-generators where, as in this instance, a sinusoidal output is desired.

A second gear is mounted at the bottom of the drum, carrying a single tooth of the same proportions as the teeth on the upper gear. In combination with a polarized reproducing head, otherwise quite similar to those used on the drum proper, this single tooth provides a signal once per revolution of the drum which (as will be shown later) is necessary for the operation of the administration unit.

The Reading Amplifier

One of the 80 plug-in reading amplifiers is pictured at the far left in Fig. 7. It employs two twin-triode vacuum tubes, and consists of a three-stage ac-coupled linear broad-band feedback amplifier, followed by a threshold output stage.

As shown in the circuit schematic of Fig. 10, the two halves of V1 and the left-hand half of V2 constitute the linear broad-band amplifier. A suitable choice of coupling elements insures that the amplification will diminish, with decreasing frequency, at a controlled rate for frequencies below a few hundred cycles per second. It is unnecessary to provide amplification at low frequencies, since the signals to be handled have no low-frequency components, and it is undesirable to do so from the standpoint of hum pickup. There is about 20db of feedback in the important part of the frequency range and the amplifier is thus substantially stabilized against variations of gain due to change in operating voltages and aging of tubes. The over-all operating voltage gain of the linear stages, with feedback, is about 56 db; the 3 db points are approximately 300 c/sec and 700 kc/sec.

The grid of the fourth stage of the reading amplifier is coupled to the output of the linear amplifier and is biased to about twice the plate-current cut-off value. The output signal from the plate of this stage, occa-

sioned by reading a "*1*", will be a negative-going pulse of approximately 40-volt amplitude from a standing potential equal to the plate supply, +130 volts. As a precaution against false signals, an externally-mounted plate-feed resistor is provided to establish at the output a condition corresponding to that of no signal present when the amplifier is removed from its receptacle.

Timing Unit

The timing unit accepts an approximately sinusoidal timing-wave signal from the upper timing head, and converts this signal into two pulse-trains, each having 1,024 narrow pulses per drum revolution, designated as A sync and B sync, alternating in time and available on separate outputs for controlling all the rest of the circuit action of the translator. A block-schematic indicating how the pulse trains are produced is shown in Fig. 11.

The general procedure for converting from a sine-wave to a synchronous train of short pulses, two per cycle of input, may be traced through the upper channel of the drawing. The signal, as represented by voltage trace 1, is amplified and clipped until a steep-sided square wave is obtained; this wave, trace 2, is applied to a push-pull phase inverter from which a pair of oppositely-phased outputs is obtained. Each of the two outputs is then differentiated by means of an R-C network, and the nega-

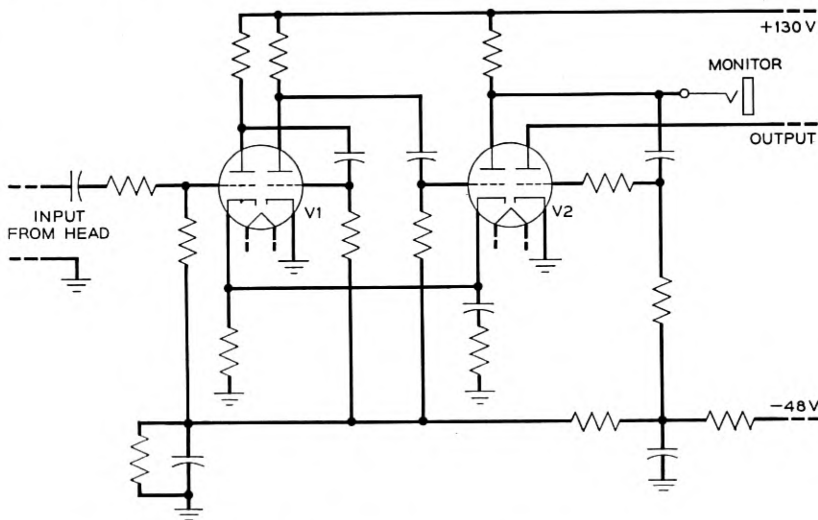


Fig. 10 — Reading amplifier circuit.

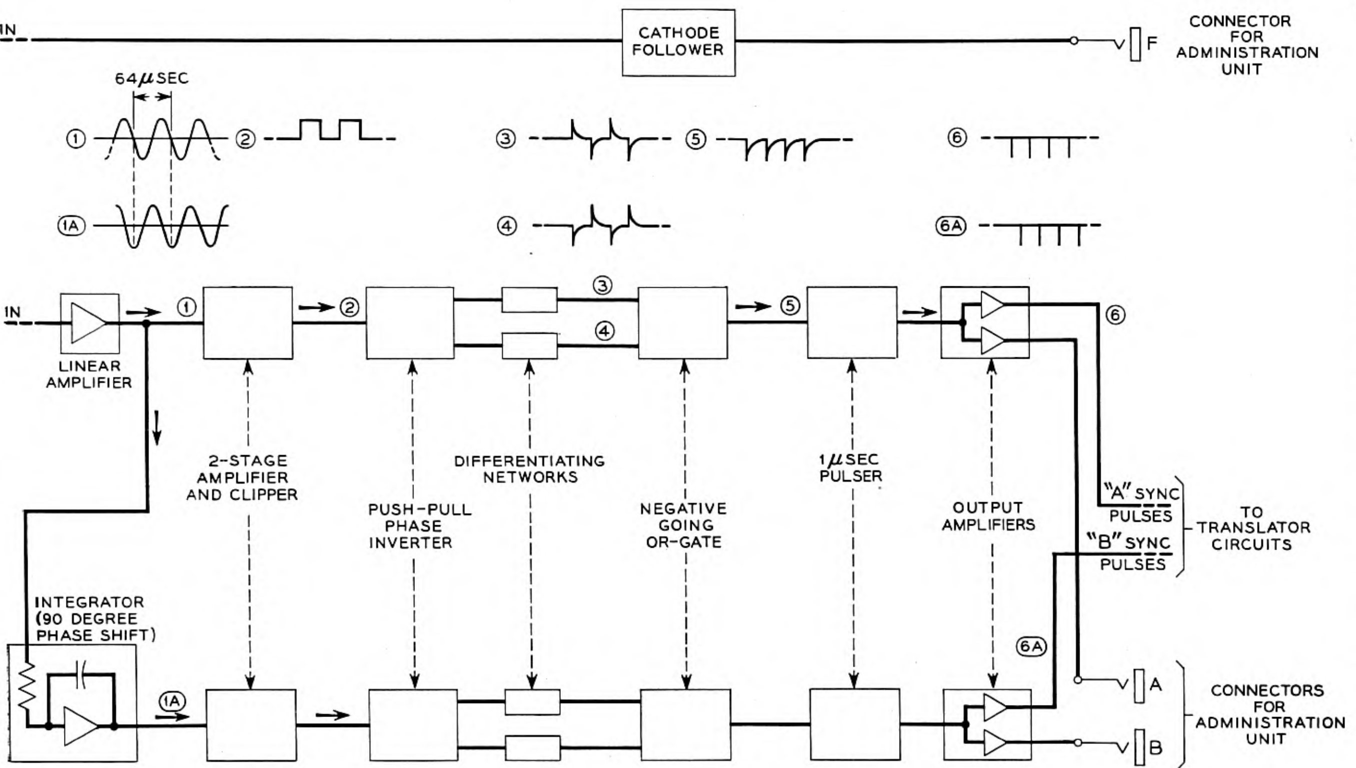


Fig. 11 — Timing unit block diagram.

tive-going spikes, traces 3 and 4, are combined in a negative-going OR gate of crystal diodes.

These spikes, trace 5, are used to trigger a cathode-coupled single-shot multivibrator, designed to give a rectangular pulse of about one microsecond duration. The multivibrator drives a pair of identical output stages: one furnishes the required A sync pulses to other equipment in the translator bay, and the other delivers its output to a coaxial connector so that, when required, the pulses may be furnished to the administration unit.

The B sync pulse-train is produced in the lower channel shown in Fig. 11. After some linear amplification, a part of the original input sine-wave is applied to a vacuum tube integrator circuit. The constants of the integrator are such that it provides very nearly a quarter-period of phase shift even if the drum varies from its nominal speed. The output of the integrator is then treated in the same manner as that described for the direct input, with the result that the required B sync pulses are produced.

The timing unit also contains a third channel which accepts the once-per-revolution signal from the special head adjacent to the single-tooth wheel. The output of this channel provides the fiducial signal, on a low-impedance basis, for administrative operations.

The Translation Selecting Unit

This unit, which appears as the bottom panel in the photograph, Fig. 6, performs a number of successive steps in making its selection. These are: (1) recognition of a match between input information from a decoder seeking a translation, and the unique corresponding information from the drum, selected from the flow of continuously-presented information; (2) production of a gate-opening pulse whose leading edge is substantially coincident in time with the leading edge of the particular A sync pulse corresponding to the entry for which the match occurred; (3) activation of a slot-spanning pulse circuit to bridge the time interval until the next-following B slot; (4) production, at a separate output, of another gate-opening pulse whose leading edge is substantially coincident in time with the leading edge of the identified B sync pulse. These actions will now be considered individually.

(1) *Recognition of Match*

Responsibility for this function is divided among a group of eight match-units operating with their associated differential amplifiers. Each match-unit is capable of comparing the inputs from five code-relays with the potentially-matching outputs of five reading amplifiers.

A circuit schematic of one of the units, with its associated differential

amplifier and some of the connected apparatus, is shown in Fig. 12. The uppermost channel on this diagram is typical of all five channels. Resistors R1 to R5 are proportioned so that the potential at point C assumes a value of +115 volts for either of the two acceptable conditions of match: (1) code-relay unoperated and reading amplifier not drawing plate current, or (2) code relay operated and reading amplifier drawing a pulse of plate current. Whenever either of the two possible conditions of mismatch exists, the potential at point C assumes a value about 15 volts higher or lower, depending on the nature of the mismatch. Resistor R6 is introduced for protective purposes only. Varistor VR1 limits the nega-

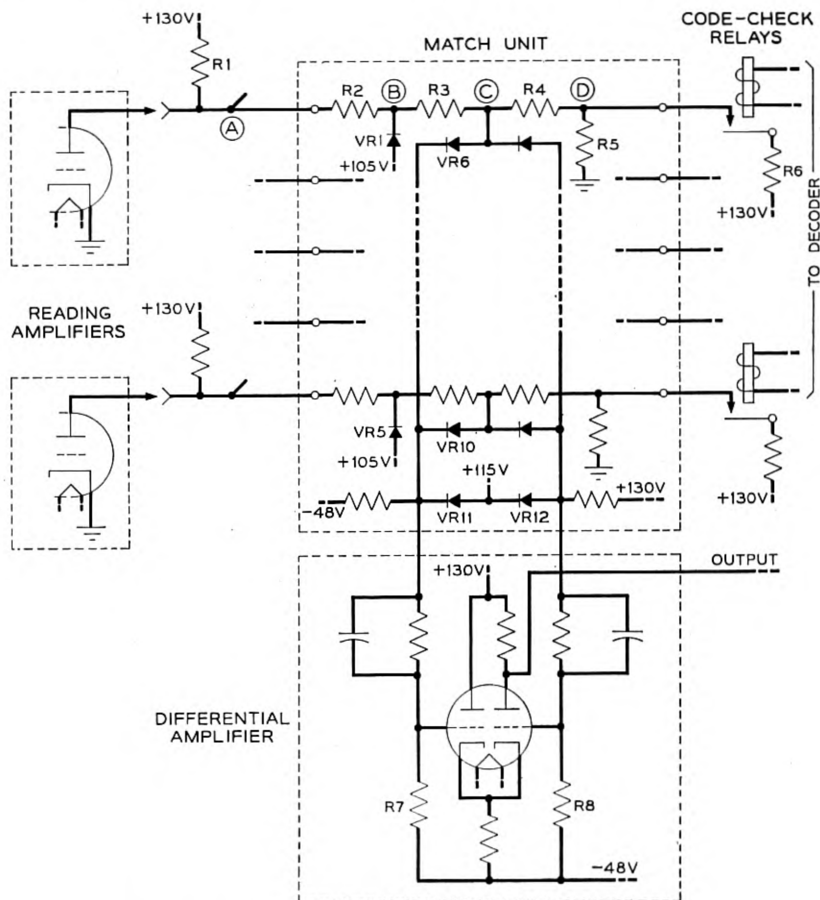


Fig. 12 — Match unit and differential amplifier circuit.

tive voltage excursion at point B, during a pulse, so that it never goes below +105 volts. This establishes the uniform pulse amplitude among the forty match channels which is necessary for proper functioning of the unit.

To detect and recognize the voltage conditions at the five junction points, two varistor gates and a differential amplifier are employed. One gate, comprising six varistors including VR6 and VR11, will transmit the type of mismatch signal which is more positive than +115 volts. This signal is dc-coupled to the left-hand grid of the differential amplifier as illustrated in Fig. 12. The type of mismatch signal which is less positive than +115 volts is blocked by this gate but is transmitted through the other gate to the right-hand grid. The threshold for this discriminating action is established by application of a fixed nominal potential of +115 volts to varistors VR11 and VR12.

At match, the output of each of the two gates presents a potential of +115 volts to the differential amplifier. The differential amplifier is biased (by inequality of R7 and R8) so that for this condition the right-hand triode is conducting, and the output potential is lower than the plate supply voltage. Positive-going mismatch signals on the left-hand grid, or negative-going signals on the right-hand grid are then equally effective in cutting off the right-hand triode, causing the output voltage to rise to plate supply potential signifying a mismatch.

The outputs from the differential amplifiers of the eight match units are combined with the A sync pulses in a system of AND gates, as illustrated in Fig. 4. A match-pulse output from this system thus signifies that conditions for match have been uniquely determined for 40 pairs of items. Thus the match unit, in total, is capable of distinguishing between all binary combinations of 40 bits or approximately 10^{12} items although when a self-checking code is employed, as in the translator application, many of these combinations are inadmissible.

(2) *The A Gate-Opening Pulse*

Occurrence of the match-pulse, as just described, indicates that the 40 items constituting one-half the contents of one of the A slots match the incoming input code; it is then desired to spill out from the other half of this same A slot the information which is also appearing at amplifier outputs at that instant. This is done by means of gates opened by the action of a gate-opening pulse, triggered by the match pulse.

The A gate-opening pulse is only a few microseconds in duration and normally is produced only once per revolution of the drum; a quiescent blocking-oscillator was chosen as the type of circuit best suited for this purpose. Whenever the code-check relays are operated in an authentic

code combination, relay CBKM is operated, removing a disabling bias from the driver stage of the blocking oscillator. When in this condition, each occurrence of the match pulse will trigger the blocking oscillator, thereby producing the A gate-opening pulse once per drum revolution.

(3) *Slot-Spanning Pulser*

Whenever an A gate-opening pulse has acted to permit read-out of information from half of the proper A slot, it is also desired to read out all the information from the next-following B slot. The first step toward doing this is to cause the A gate-opening pulse to trigger a single-shot multivibrator whose characteristic period is long enough to just bridge the time until the next slot appears. The output of this pulser is combined with the B sync pulses in an AND gate so that the selected B pulse, corresponding to the wanted B slot, can be used to trigger another gate-opening blocking-oscillator just as the match pulse was used to trigger the A gate-opening blocking-oscillator.

(4) *The B Gate-Opening Pulse*

The outputs of all the reading amplifiers must be gated for the B slot. Hence the B gate-opening pulse must operate twice as many gates as the A gate-opening pulse and must be correspondingly more powerful. This requirement is met by using the same circuit design with parallel output tubes.

Pulse Stretchers and Channel Detectors

Fig. 13 presents a simplified schematic of one of the translator output channels, together with certain of the relays in the decoder. Package-wise, the pulse stretchers combine two functions: that of an AND gate with two inputs and a threshold feature, and that of a single-shot multivibrator for amplifying and lengthening the short input pulse from the gate. A single point-contact transistor provides the necessary gain for the monostable action. The inputs to the AND gate come from sources which supply negative-going pulses from a standing potential of +130 volts. When one or the other, but not both, of these sources supplies a pulse, a larger portion of the current being supplied to resistor R1 must be drawn from the non-active source; this extra demand causes a small voltage drop which becomes evident at the gate output. The resultant weak false signal is prevented from affecting the transistor pulser by the action of threshold diode VR1 which is normally back-biased a few volts by the potential divider R2, R3. Small negative-going signals from the gate will not overcome the bias and will therefore be greatly attenuated; normal gate-output pulses, occasioned by coincidence of pulses at both inputs will,

however, overcome the bias and will be transmitted to the transistor monostable circuit.

When triggered at the base, the transistor delivers a pulse of about one millisecond duration to the load represented by the input transformer and the channel detector gas tube and thus provides the drive required to initiate ionization in the control gap of the gas tube. When brought into action, the transistor serves as a switch to connect capacitor C to collector supply resistor R6. The voltage change, occasioned by the resultant flow of current in R6, is communicated to the transformer primary through a blocking capacitor and a current limiting resistor. As capacitor C charges, the voltage at the transistor emitter will approach the collector supply potential at an approximately exponential rate. When the diminishing flow of emitter current can no longer maintain the transistor in its low-impedance mode, it reverts to its pre-triggered condition, and the timing capacitor C is then discharged, primarily through forward-conducting varistor VR2 and resistors R5 and R4.

Owing to the necessity of using early-production samples of the type of point-contact transistor chosen for this application, the associated circuitry for biasing the emitter into the normal non-conducting state is somewhat more elaborate than that which might have sufficed with later samples whose characteristics were more closely controlled.

The principal components of the channel detector are a step-up trans-

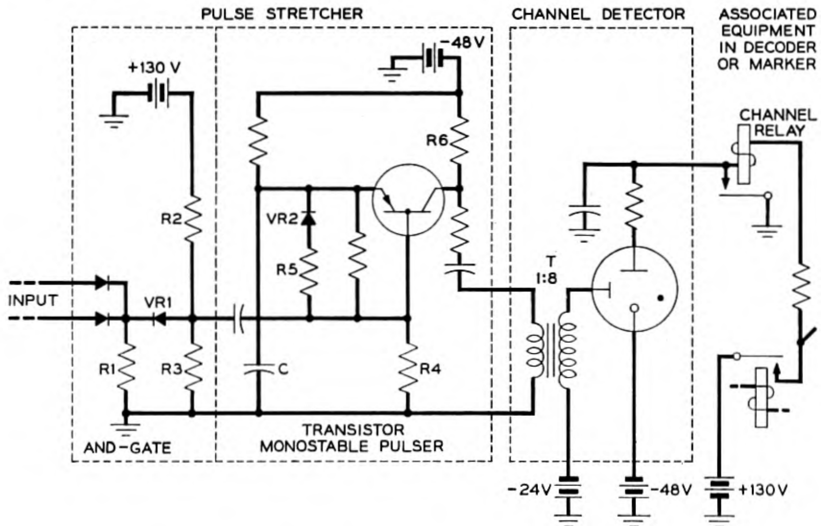


Fig. 13 — Pulse stretcher and channel detector circuit.

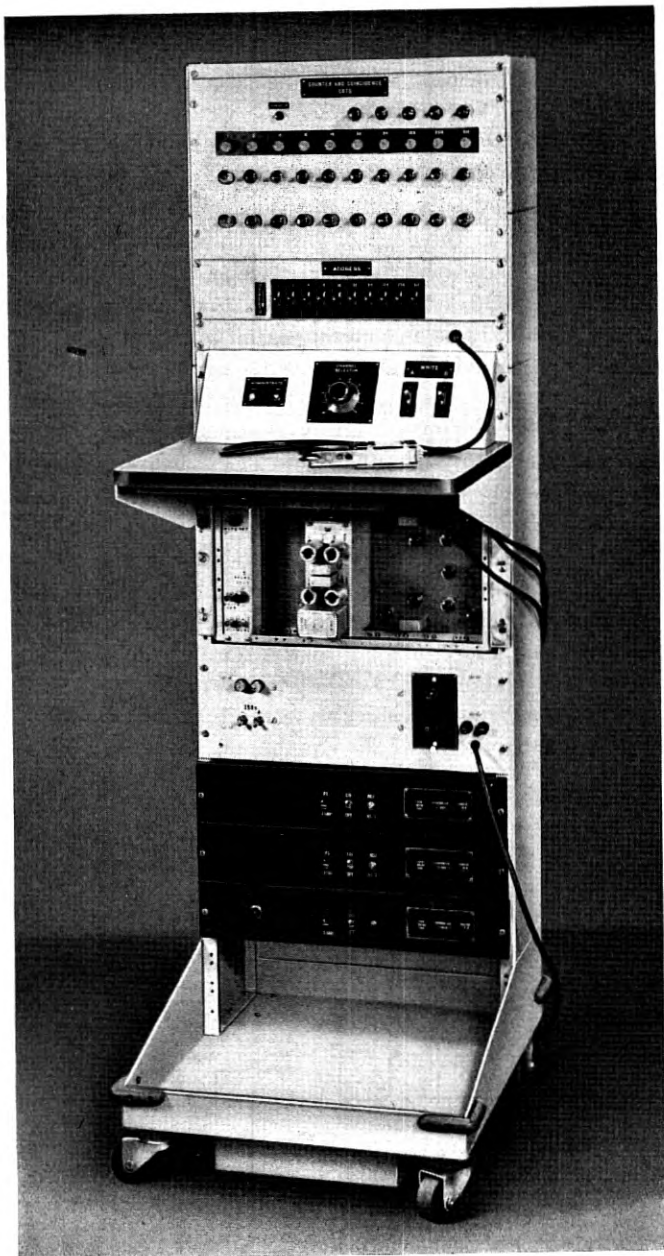


Fig. 14 — Administration unit. Three co-ax leads entering under shelf bring A, B and F pulses from translator. Cable leading to plug with bail-handle resting on shelf serves to connect writing amplifier output to magnetic heads in translator. Bottom cable connects to 60-cycle source which supplies all power.

former designed for the audio frequency range, and a cold-cathode gas tube. The starter-anode of the gas tube has a dc bias of about +24 volts with respect to its cathode to reduce the value of pulse voltage required to ionize it. When +130 volt battery is applied via the winding of the channel relay to the main anode of the gas tube, ionization established in the starter gap by the pulse stretcher signal will transfer to the main gap and cause the relay to operate. Closure of one of the relay make-contacts serves to divert the winding current from the gas tube directly to ground, thereby extinguishing the tube and prolonging its life. Other contacts, not shown, make the registered information available.

Components

A full complement of the electronic apparatus described in the last few sections utilizes plug-in components in the following quantities:

Twin-triode electron tubes	186
Cold-cathode gas tubes	121
Germanium varistors	552
Point-contact transistors	120

Only one type of each of these components is used in the translator; this uniformity greatly simplifies the maintenance problem and imposed little if any handicap on the circuit designs.

ADMINISTRATION EQUIPMENT

Whenever it is desired to add, or to change, a translation item on the drum, the auxiliary administration unit pictured in Fig. 14 is connected to the translator by three shielded cables, shown leaving the rack just under the shelf, and a ten-conductor cable, shown with its plug resting on the shelf. The shielded cables convey the A and B sync pulses and the once-per-drum-revolution fiducial F pulse to the administrator. The ten-conductor cable, with plug, is used to establish paths extending directly to magnetic heads on the drum. During the recording of any one complete translation item on the drum, this plug is successively shifted to each of nine multi-connector jacks located in the amplifier compartment of the translator.

The manual controls are located just above the shelf. At the right are the two keys for ordering a writing operation, one for the A slot and another for the B slot of the chosen pair. If either key is lifted, it will order the entry of a magnetic mark (write "*1*"). If depressed, the key will order the removal of a mark (write "*0*"). It is obvious that the translation is

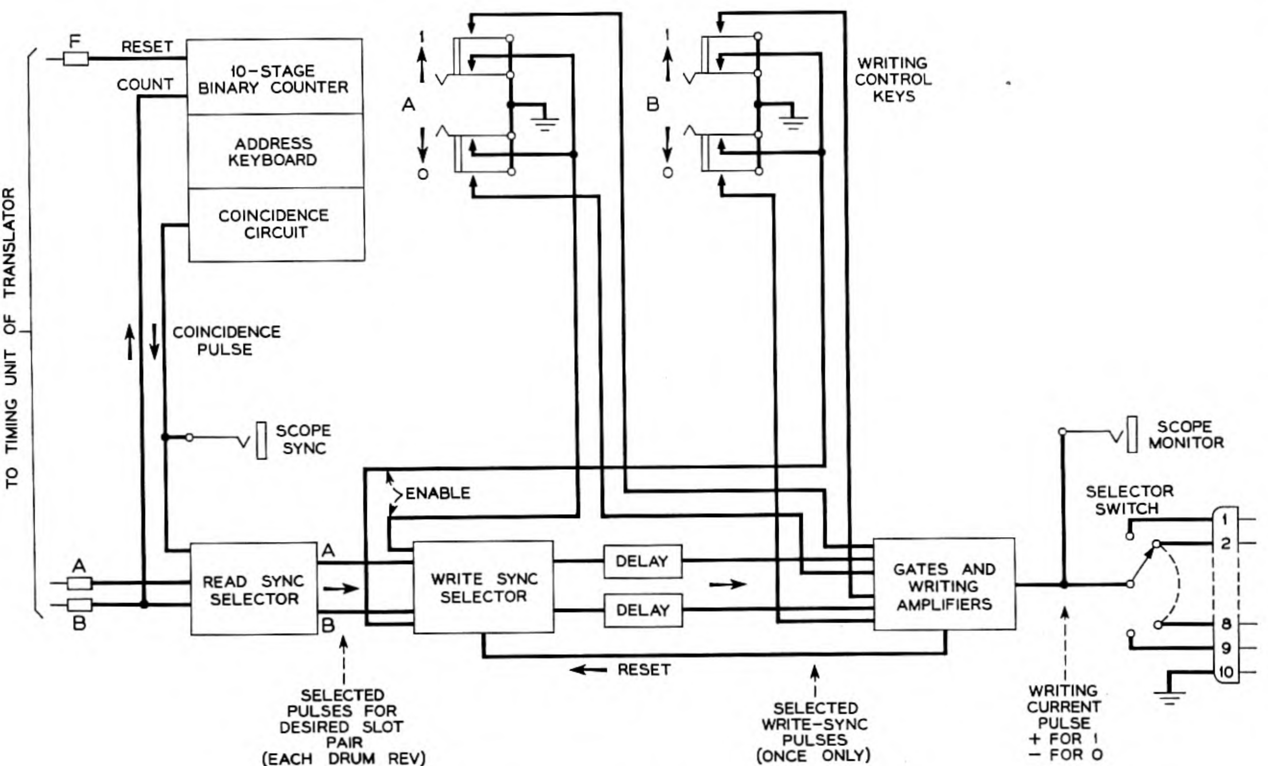


Fig. 15 — Administration unit block diagram.

inserted piecemeal by working in each track successively. The manual switching operation of connecting a single pair of writing amplifiers to each of eighty magnetic heads, in turn, is accomplished partly by setting the nine-position switch shown at the center of the panel, and partly by shifting the plug of the ten-conductor cable. At the left are two signal lights which serve as alarms to warn the operator of possible incorrect functioning of the equipment.

The operation of the administration unit can best be traced with the aid of the schematic block-diagram of Fig. 15. A ten-stage binary counter is supplied with B sync pulses from the translator; the 1,024 possible states of the counter are traversed in the course of exactly one revolution of the translator drum. The F pulse from the translator will, midway between two B pulses, set all counter stages to zero, once per revolution. After the first such reset, however, if the counter is working properly, it will always have returned to the zero condition just before the occurrence of the F pulse, by having counted 1,024 B pulses; under these conditions the F pulse, though still initiating reset action, does not change the state of the counter. The basis for the alarm signals mentioned above is a circuit arranged to detect if a change of state is occasioned by the F pulse.

Associated with the counter is a coincidence circuit with a keyboard on which may be set up any "address" between 0 and 1,023. When the count of B pulses equals the address set up on the keyboard, the coincidence circuit delivers a pulse which persists until the next B pulse alters the count; this coincidence pulse spans the time of occurrence of an A pulse, and is used in the read sync selector to gate-out a "selected" A pulse uniquely assigned to the address set up on the keyboard. A slot-spanning pulser, triggered by the selcted A pulse, gates-out the associated "selected" B pulse.

These selected pulses, which occur once per revolution of the drum, are passed through gates under control of bistable electron-tube pairs which can be set by the manual writing keys and are re-set by the writing action itself. This insures that the desired action takes place only once per key operation, instead of repeating, once per drum-revolution, as long as the keys are held operated. The manually-gated unique selected A or selected B sync pulse is then slightly delayed in time to become a selected write-sync pulse. It is passed on through further gates under direct control of the writing keys, and is employed as an input to a writing amplifier.

A pair of writing amplifiers is provided, one to write "*1*" and the other to write "*0*"; the circuits are identical quiescent blocking-oscillators sharing a common output transformer, and one or the other is triggered into

action by the write-sync pulses. The output transformer supplies the writing current pulses, under control of the selector switch, to the chosen magnetic head. Arrangements are provided for synchronizing an oscilloscope to display the writing current pulses or the voltage outputs from the head at the selected address, as required.

When a new translation item is to be entered, or an existing one altered, the address corresponding to the desired slot-pair is determined from a card-index, or ledger, listing all items on the drum. The address keyboard is then set to the assigned number, thereby singling-out the desired slot pair so that the writing operation can proceed as described above. During this procedure, the monitoring oscilloscope may be used for verifying the new entry, two cells at a time. Over-all verification is accomplished by exercising the translator through facilities already available in the toll switching office. There is nothing about this procedure which precludes the use of automatic facilities for performing the administration. There is also no fundamental need to take the translator out of routine service during the administration operation, since each writing operation disables the equipment for only a few microseconds and would rarely delay a translation by as much as one drum revolution.

CONCLUSION

After short preliminary tests, the equipment described and pictured was installed in the switching systems laboratory at Bell Laboratories. A rapid-transfer arrangement permitted direct interchangeability with a card translator in a skeletonized model of a toll switching office.

A testing program was then begun entailing continuous 24-hours-per-day operation of the magnetic drum translator for approximately one year. After an initial shakedown period during which wiring faults and other minor troubles were recognized and cleared, many millions of translations were handled with only a small proportion of failures. The accumulated data on failure rate and cause was significant, being one of the primary objectives of the experiment. An analysis of the data indicated the desirability of certain simple design changes in the existing circuitry and established a basis for the selection of future designs.

If, in the future, consideration is given to the design of equipment of this type for some specific application, new electronic developments must also be taken into account. Many more types of transistors are now available than when the present design was undertaken, and some of the newer types have capabilities which make them obvious candidates for many of the jobs now done in the translator with electron tubes. Such a substitution would not only increase reliability and decrease power con-

sumption, but since transistors are essentially current-operated devices they would seem to be particularly suitable for working with microsecond pulses in the environment of existing relay-equipped offices where the majority of interference transients are capacitively-propagated voltage-disturbances.

Evaluation of the magnetic drum reveals it to be a safe and very reliable means of storing several hundred thousand bits of information. During the course of these tests, the drum functioned perfectly, and the translations that were recorded at the beginning of the test were retained until near the end, when they were deliberately altered. During this interval of nearly continuous operation there was no detectable deterioration, or change in the signals obtained from the drum.

The results obtained from the tests of this particular drum translator indicate that the associated circuitry, working with microsecond pulses, can be designed to measure up to the exacting standards demanded for telephone office apparatus, whether the application be that of a magnetic drum translator or some other type of equipment.

REFERENCES

1. W. D. Lewis, Electronic Computers and Telephone Switching, Proc. I.R.E., **41,** pp. 1242–1244; Oct., 1953.
2. W. A. Malthaner and H. E. Vaughan, An Automatic Telephone System Employing Magnetic Drum Memory, Proc. I.R.E., **41,** pp. 1341–1347; Oct., 1953.
3. J. H. McGuigan, Combined Reading and Writing on a Magnetic Drum, Proc. I.R.E., **41,** pp. 1438–1444; Oct., 1953.
4. L. N. Hampton and J. B. Newsom, The Card Translator for Nationwide Dialing, B. S. T. J., **32,** pp. 1037–1098; Sept., 1953.

Tables of Phase of a Semi-Infinite Unit Attenuation Slope

By D. E. THOMAS

(Manuscript received February 24, 1956)

Five and seven place tables of the integral

$$B(x_c) = \frac{1}{\pi} \int_{x=0}^{x=x_c} \log \left| \frac{1+x}{1-x} \right| \frac{dx}{x}$$

which gives the phase associated with a semi-infinite unit slope of attenuation, are now available in monograph form. The usefulness of this integral and its tabulation are discussed.

H. W. Bode[1] has shown that on the imaginary axis, the values of the imaginary part of certain functions of a complex variable may be obtained from the corresponding values of the real part, and vice versa. This theorem was immediately recognized as a powerful tool in the communications and network fields. The most generally useful function which was given by Bode for use in applying this theorem to the solution of communications problems, is the phase associated with a semi-infinite unit slope of attenuation. This is given by the integral[2]

$$B(x_c) = \frac{1}{\pi} \int_{x=0}^{x=x_c} \log \left| \frac{1+x}{1-x} \right| \frac{dx}{x} \tag{1}$$

where: $B(x_c)$ is the phase in radians at frequency f_c,

$$x = \frac{f}{f_0}, \; x_c = \frac{f_c}{f_0} < 1.0$$

and f_0 = the frequency at which the semi-infinite unit slope begins

The usefulness of Integral (1) is illustrated by some of the communication problems which stimulated its accurate tabulation.

[1] Bode, H. W., Network Analysis and Feedback Amplifier Design, D. Van Nostrand Co., Inc., New York, 1945, Chap. XIV.

[2] Ibid: Chap. XV, pp. 342–343.

When the development program on deep sea repeatered submarine telephone cable systems was reactivated at the close of World War II, one of the first problems to present itself was the determination of the delay distortion of a transatlantic repeatered cable system. The only means then known of obtaining an answer to this problem was by computing the minimum phase of the system from its predictable attenuation characteristic, using Bode's straight line approximation method,[3] and then determining the delay distortion from the non-linear portion of this minimum phase. However, the non-linear phase is such a small part of the total phase, that a five figure accuracy tabulation of Integral (1) was needed for a satisfactory determination of the non-linearity. The necessary table was therefore compiled. A numerical computation was used to evaluate the integral because of the simplicity of its integrand. The minimum phase of the projected transatlantic repeatered telephone cables was then computed using this table and the anticipated delay distortion was determined from the non-linear portion of this minimum phase.

About this time the delay equalization of coaxial cable systems for television transmission became a pressing problem. Bode's technique proved to be the simplest means for determining the delay to be equalized and so the existing phase table was immediately put to use in the coaxial cable delay equalization program.

The increasing use of the tables led to a decision to publish them in The Bell System Technical Journal.[4] In order to make the tables more generally useful, the published paper included a tabulation of the phase in radians as well as in degrees. The radian tables can, for example, be used to determine the reactance characteristic associated with a given resistance characteristic of a minimum reactance impedance function.

Because of the demand for higher accuracy which occasionally arose after the publication of the five place tables, it was decided to undertake the computation of seven-place tables. These tables were also computed numerically using intervals selected to give at least ±1 accuracy in the final figure. The complete tables require forty-nine pages for tabulation. Since it is probable that only a fraction of the Journal readers would need these tables, it did not seem desirable to publish the actual tables in the Journal. They are therefore being published in original monograph form as Bell System Monograph 2550[5] entitled "Tables of Phase of a Semi-Infinite Unit Attenuation Slope." The phase is tabulated in the

[3] Ibid: Chap. XV.

[4] Thomas, D. E., Tables of Phase Associated with a Semi-Infinite Unit Slope of Attenuation, B.S.T.J., **26,** pp. 870–899, Oct., 1947.

[5] This Monograph will be available about June 15, 1956.

monograph both in degrees and radians for values of f greater than f_0 as well as for f less than f_0. The tabular intervals are 0(0.001) 0.600 (0.0005) 0.9000 (0.0001) 0.9940 (0.00005) 0.99800 (0.00001) 1.00000. These intervals were selected to permit linear interpolation for intermediate values of the phase to an accuracy of the same order as the accuracy of the tabulated values, i.e., ± 1 in the last place. The original JOURNAL article discussed the construction of the tables and the errors involved in the numerical evaluation of Integral (1), described and illustrated the use of the tables, and gave five-place tabulations of the integral. This entire article is therefore included in Monograph 2550 for completeness along with the newer seven-place tables.

B. A. Kingsbury[6] has pointed out that the Integral (1) which is tabulated in the phase tables in question is useful in other than the communications and network fields. A bibliography covering other possible fields of interest is given in an article by Murakami and Corrington.[7]

ACKNOWLEDGMENT

The author is indebted to R. W. Hamming of the Mathematical Research Department who supervised the computation of the seven place tables, to Miss R. A. Weiss who planned, programmed, ran, and checked the IBM computations of the tables and to Miss J. D. Goeltz who computed the ten-figure accuracy check points required for the construction of the tables. He also wishes to acknowledge the support and encouragement given to the project by R. L. Dietzold and P. H. Richardson, and the continued interest and helpful comments of B. A. Kingsbury.

[6] Kingsbury, B. A., private communication.
[7] Murakami, T., and Corrington, M. S., Relation Between Amplitude and Phase in Electrical Networks, R.C.A. Review, **9**, pp. 602–631, Dec., 1948.

Bell System Technical Papers Not Published in This Journal

ANDERSON, P. W.,[1] and SUHL, H.[1]

Instability in the Motion of Ferromagnets at High Microwave Power Levels, Phys. Rev., Letter to the Editor, **100,** pp. 1788–1789, Dec. 15, 1955.

ANDRUS, J., see Bond, W. L.

BEACHELL, H. C., see Veloric, H. S.

BECK, A. C.,[1] and MANDEVILLE, G. D.[1]

Microwave Traveling Wave Tube Millimicrosecond Pulse Generators, I.R.E. Trans., **MTT-3,** pp. 48–51, Dec., 1955.

BENEDICT, T. S.[1]

Single-Crystal Automatic Diffractometer — Part II, Acta Cryst., **8,** pp. 747–752, Dec. 10, 1955.

BENNETT, W. R.[1]

Application of the Fourier Integral in Circuit Theory and Circuit Problems, I.R.E. Trans., **CT-2, 3,** pp. 237–243, Sept., 1955.

BIONDI, F. J.[1]

Corrosion-Proofing Electronic Parts Against Ozone, Ceramic Age, **66,** p. 39, Oct., 1955.

BOND, W. L.[1]

Single-Crystal Automatic Diffractometer — Part I, Acta Cryst., **8,** pp. 741–746, Dec. 10, 1955.

[1] Bell Telephone Laboratories, Inc.

BOND, W. L.,[1] AND ANDRUS, J.[1]

Photographs of the Stress Field Around Edge Dislocations, Phys. Rev., Letter to the Editor, **101,** p. 1211, Feb. 1, 1956.

BOYLE, W. S., See Germer, L. H.

BOYLE, W. S.,[1] and HAWORTH, F. E.[1]

Glow-to-Arc Transitions, Phys. Rev., **101,** pp. 935–938, Feb. 1, 1956.

BOZORTH, R. M.[1]

The Physics of Magnetic Materials, Elec. Engg., **75,** pp. 134–140, Feb. 1956.

BRIDGERS, H. E.[1]

A Modern Semiconductor — Single Crystal-Germanium, Chem. and Engg. News, **34,** p. 220, Jan., 1956.

BURRUS, C. A.,[1] and GORDY, W.[5]

Millimeter and Submillimeter Wave Spectroscopy, Phys. Rev., **101,** pp. 599–603, Jan. 15, 1956.

CHYNOWETH, A. G.[1]

Dynamic Method for Measuring the Pyroelectric Effect with Special Reference to Barium Titanate, J. Appl. Phys., **27,** pp. 78–84, Jan., 1956.

CUTLER, C. C.[1]

Spurious Modulation of Electron Beams, Proc. I.R.E., **44,** pp. 61–64, Jan., 1956.

DAVIS, H. M., see Wernick, J. H.

DUNCAN, R. A.,[1] and STONE, J. A., JR.[1]

A Survey of the Application of Ferrites to Inductor Design, Proc. I.R.E., **44,** pp. 4–13, Jan., 1956.

---

[1] Bell Telephone Laboratories, Inc.
[5] Duke University.

FEHER, G.,[1] FLETCHER, R. C.,[1] and GERE, E. A.[1]

Exchange Effects in Spin Resonance of Impurity Atoms in Silicon, Phys. Rev., Letter to the Editor, **100,** pp. 1784–1785, Dec. 15, 1955.

FELDMANN, W. L., see Pearson, G. L.

FEWER, D. R.[1]

Design Principles for Junction Transistor Audio Power Amplifiers, I.R.E. Trans., **AU-3,** pp. 183–201, Nov.–Dec., 1955.

FLASCHEN, S. S.,[1] and VAN UITERT, L. G.[1]

New Low Contact Resistance Electrode, J. Appl. Phys., Letter to the Editor, **27,** p. 190, Feb., 1956.

FLETCHER, R. C., see Feher, G.

FRY, T. C.[1]

Mathematics as a Profession Today in Industry, Am. Math. Monthly, **63,** pp. 71–80, Feb., 1956.

FULLER, C. S., see Reiss, H.

GEBALLE, T. H., see Hrotowski, H. J.

GERE, E. A., see Feher, G.

GERMER, L. H.,[1] and BOYLE, W. S.[1]

Short Arcs, Nature, Letter to the Editor, **176,** p. 1019, Nov. 26, 1955.

GERMER, L. H.,[1] and BOYLE, W. S.[1]

Two Distinct Types of Short Arcs, J. Appl. Phys., **27,** pp. 32–39, Jan., 1956.

GIANOLA, U. F.[1]

Photovoltaic Noise in Silicon Broad Area p-n Junctions, J. Appl. Phys., **27,** pp. 51–53, Jan., 1956.

GORDY, W., see Burrus, C. A.

---

[1] Bell Telephone Laboratories, Inc.

HAGELBARGER, D. W., see Pfann, W. G.; Shannon, C. E.; and Williams, H. J.

HAGSTRUM, H. D.[1]

Electron Ejection from Metals by Positive Ions, Appl. Sci. Res. **B5,** Nos. 1–4, pp. 16–17, 1955.

HAWORTH, F. E., see Boyle, W. S.

HERRING, C.,[1] and VOGT, E.[1]

Transport and Deformation Potential Theory for Many-Valley Semiconductors with Anisotropic Scattering, Phys. Rev., **101,** pp. 944–961, Feb. 1, 1956.

HERRMANN, D. B., see Williams, J. C.

HOLDEN, A. N.,[1] MERZ, W. J.,[1] REMEIKA, J. P.,[1] and MATTHIAS, B. T.[1]

Properties of Guanidine Aluminum Sulfate Hexahydrate and Some of its Isomorphs, Phys. Rev., **101,** pp. 962–967, Feb. 1, 1956.

HOROTOWSKI, H. J.,[1] MORIN, F. J.,[1] GEBALLE, T. H.,[1] and WHEATLEY, G. H.[1]

Hall Effect and Conductivity of InSb, Phys. Rev., **100,** pp. 1672–1677, Dec. 15, 1955.

INGRAM, S. B.[1]

The Graduate Engineer His Training and Utilization in Industry Elec. Engg., **75,** pp. 167–170, Feb., 1956.

KAPLAN, E. L.[1]

Transformation of Stationary Random Sequences, Mathematica Scandinavica, **3,** FASCl, pp. 127–149, June, 1955.

LEWIS, H. W.[1]

Superconductivity and Electronic Specific Heat, Phys. Rev., **101,** pp. 939–940, Feb. 1, 1956.

MANDEVILLE, G. D., see Beck, A. C.

[1] Bell Telephone Laboratories, Inc.

MATTHIAS, B. T., see Holden, A. N.

MERZ, W. J., see Holden, A. N.

MILLER, L. E.[1]

Negative Resistance Regions in the Collector Characteristics of the Point-Contact Transistor, Proc. I.R.E., **44,** pp. 65–72, Jan., 1956.

MOLL, J. L.,[1] and ROSS, I. M.[1]

The Dependence of Transistor Parameters on the Distribution of Base Layer Resistivity, Proc. I.R.E., **44,** pp. 72–78, Jan., 1956.

MONTGOMERY, H. C., See Pearson, G. L.

MORIN, F. J., see Hrotowski, H. J.

MUMFORD, W. W.,[1] and SCHAFERMAN, R. L.[1]

Data on the Temperature Dependence of X-Band Fluorescent Lamp Noise Sources, I.R.E. Trans., **MTT-3,** pp. 12–16, Dec., 1955.

NESBITT, E. A., see Williams, H. J.

OLMSTEAD, P. S.[1]

QC Concepts Useful in OR, Ind. Qual. Cont., **12,** pp. 11, 14–17, Oct., 1955.

OWENS, C. D.[1]

Stability Characteristics of Molybdenum Permalloy Powder Cores, Elec. Engg., **74,** pp. 252–256, Feb., 1956.

PEARSON, G. L.,[1] MONTGOMERY, H. C.,[1] and FELDMANN, W. L.[1]

Noise in Silicon p-n Junction Photocells, J. Appl. Phys., **27,** pp. 91–92, Jan., 1956.

PFANN, W. G.,[1] and HAGELBARGER, D. W.[1]

Electromagnetic Suspension of a Molten Zone, J. Appl. Phys., **27,** pp. 12–17, Jan., 1956.

[1] Bell Telephone Laboratories, Inc.

QUINLAN, A. L.[3]

Roll-Welding Precious Metals for Telephone Contacts, Elec. Engg., **75,** pp. 154–157, Feb., 1956.

REISS, H.,[1] and FULLER, C. S.[1]

The Influence of Holes and Electrons on the Solubility of Lithium in Boron-Doped Silicon, J. of Metals, **12,** p. 276, Feb., 1956.

REMEIKA, J. P., see Holden, A. N.

ROSS, I. M., see Moll, J. L.

SCHAFERMAN, R. L., see Mumford, W. W.

SCHAWLOW, A. L.[1]

Structure of the Intermediate State in Superconductors, Phys. Rev., **101,** pp. 573–580, Jan. 15, 1956.

SCHAWLOW, A. L.,[1] and TOWNES, C. H.[4]

Effect on X-Ray Fine Structure of Deviations from a Coulomb Field near the Nucleus, Phys. Rev., **100,** pp. 1273–1280, Dec. 1, 1955.

SHANNON, C. E.,[1] and HAGELBARGER, D. W.[1]

Concavity of Resistance Functions, J. Appl. Phys., **27,** pp. 42–43, Jan, 1956.

SIMKINS, Q. W.,[1] and WOGELSONG, J. H.[1]

Transistor Amplifiers for Use in a Digital Computer, Proc. I.R.E., **44,** pp. 43–54, Jan., 1956.

SNOKE, L. R.[1]

Specific Studies on the Soil-Block Procedure for Bioassay of Wood Preservatives, Appl. Microbiology, **4,** pp. 21–31, Jan., 1956.

SOUTHWORTH, G. C.[1]

Early History of Radio Astronomy, Sci. Mo., **82,** pp. 55–66, Feb., 1956.

---

[1] Bell Telephone Laboratories, Inc.
[3] Western Electric Company.
[4] Columbia University.

STONE, H. A., see Duncan, R. A.

SUHL, H., see Anderson, P. W.

THOMAS, E. E.[1]

Tin Whisker Studies — Observation of some Hollow Whiskers and Some Sharply Irregular External Forms, Letter to the Editor, Acta Met., **4,** p. 94, Jan., 1956.

TOWNES, C. H., see Schawlow, A. L.

TOWNSEND, M. A.[1]

A Hollow Cathode Glow Discharge with Negative Resistance, Appl. Sci. Research, **Sec. B, 5,** pp. 75–78, 1955.

VALDES, L. B.[1]

Frequency Response of Bipolar Transistors with Drift Fields, Proc. I.R.E., **44,** pp. 178–184, Feb., 1956.

VAN UITERT, L. G., see Flaschen, S. S.

VELORIC, H. S.,[1] and BEACHELL, H. C.[6]

Absorption Isotherms, Isobars and Isoteres of Diborane on Palladium on Charcoal and Boron Nitride, J. Phys. Chem., **60,** p. 102, Jan., 1956.

VOGELSONG, J. H., see Simkins, Q. W.

VOGT, E., see Herring, C.

WEIBEL, E. S.[1]

Strains and the Energy in Thin Elastic Shells of Arbitrary Shape for Arbitrary Deformation, Zeitchrift f. Mathematik and Physik, **6,** pp. 153–189, May 25, 1955.

WERNICK, J. H.,[1] and DAVIS, H. M.[7]

Preparation and Inspection of High-Purity Copper Single Crystals, J. Appl. Phys., **27,** pp. 144–153, Feb., 1956.

[1] Bell Telephone Laboratories, Inc.
[6] University of Delaware.
[7] Penn State University.

WHEATLEY, G. H., see Hrotowski, H. J.

WILLIAMS, H. J.,[1] HEIDENREICH, R. D.,[1] and NESBITT, E. A.[1]

Mechanism by which Cobalt Ferrite Heat Treats in a Magnetic Field, J. Appl. Phys., **27,** pp. 85–89, Jan., 1956.

WILLIAMS, J. C.,[1] and HERRMANN, D. B.[1]

Surface Resistivity of Non-Porous Ceramic and Organic Insulating Materials at High Humidity with Observations of Associated Silver Migration, I.R.E. Trans., **PGRQC-6,** pp. 11–20, Feb., 1956.

WOOD, MRS. E. A.[1]

A Heated Sample-Holder for X-Ray Diffractometer Work, Rev. Sci. Instr., **27,** p. 60, Jan., 1956.

[1] Bell Telephone Laboratories, Inc.

Recent Monographs of Bell System Technical Papers Not Published in This Journal*

ANDERSON, P. W., and HASEGAWA, H.

Considerations on Double Exchange, Monograph 2532.

BAKER, W. O., see Winslow, F. H.

BARSTOW, J. M.

The ABC's of Color Television, Monograph 2529.

BEMSKI, G.

Lifetime of Electrons in p-type Silicon, Monograph 2534.

BENNETT, W. R.

Application of the Fourier Integral in Circuit Theory, Monograph 2533.

BRATTAIN, W. H., see Pearson, G. L.

BROWN, W. L.

Surface Potential and Surface Charge Distribution from Semiconductor Field Effect Measurements, Monograph 2501.

BULLINGTON, K.

Characteristics of Beyond-the-Horizon Radio Transmission, Monograph 2494.

BULLINGTON, K., INKSTER, W. J., and DURKEE, A. L.

Propagation Tests at 505 mc and 4,090 mc on Beyond-Horizon Paths, Monograph 2503.

* Copies of these monographs may be obtained on request to the Publication Department, Bell Telephone Laboratories, Inc., 463 West Street, New York 14. N. Y. The numbers of the monographs should be given in all requests.

DURKEE, A. L., see Bullington, K.

FREYNIK, H. S., see Gohn, G. R.

GELLER, S., and THURMOND, C. D.

On the Question of the Existence of a Crystalline SiO, Monograph 2536.

GOHN, G. R., GUERARD, J. P., and FREYNIK, H. S.

The Mechnical Properties of Wrought Phosphor Bronze Alloys, Monograph 2531.

GUERARD, J. P., see Gohn, G. R.

HASEGAWA, H., see Anderson, P. W.

HAYNES, J. R., see Hornbeck, J. A.

HORNBECK, J. A., and HAYNES, J. R.

Trapping of Minority Carriers in Silicon, Monograph 2368.

INKSTER, W. J., see Bullington, K.

LEWIS, H. W.

Search for the Hall Effect in a Superconductor. II. Theory, Monograph 2523.

LINVILL, J. G., and MATTSON, R. H.

Junction Transistor Blocking Oscillators, Monograph 2487.

LOGAN, R. A.

Precipitation of Copper in Germanium, Monograph 2524.

LOGAN, R. A., and SCHWARTZ, M.

Restoration of Resistivity and Lifetime in Heat-Treated Germanium Monograph 2525.

MATTSON, R. H., see Linvill, J. G.

MAYS, J. M., see Shulman, R. G.

MCCALL, D. W., see Shulman, R. G.

MOLL, J. L.

Junction Transistor Electronics, Monograph 2537.

PEARSON, G. L., and BRATTAIN, W. H.

History of Semiconductor Research, Monograph 2538.

SANDSMARK, P. I.

Ellipticity on Dominant-Mode Axial Ratio in Nominally Circular Waveguides, Monograph 2539.

SCHWARTZ, M., see Logan, R. A.

SHULMAN, R. G., MAYS, J. M., and MCCALL, D. W.

Nuclear Magnetic Resonance in Semiconductors. I, Monograph 2528.

THURMOND, C. D., see Geller, S.

VAN UITERT, L. G.

Low Magnetic Saturation Ferrites for Microwave Applications, Monograph 2504.

VAN UITERT, L. G.

Dc Resistivity in the Nickel and Nickel Zinc Ferrite System, Monograph 2540.

WEIBLE, E. S.

Vowel Synthesis by Means of Resonant Circuits, Monograph 2541.

WINSLOW, F. H., BAKER, W. O., and YAGER, W. A.

Odd Electrons in Polymer Molecules, Monograph 2486.

YAGER, W. A., see Winslow, F. H.

Contributors to This Issue

Donald C. Bennett, B.S. 1949 and M.S. 1951, Rensselaer Polytechnic Institute; Battelle Memorial Institute, 1951–1952; Bell Telephone Laboratories, 1952–. Mr. Bennett has been engaged in the development of processes for producing single crystals suitable for use in transistors. He is a member of the American Institute of Mining and Metallurgical Engineers.

F. G. Buhrendorf, B.S.M.E. and M.E., Cooper Union Inst. Tech. 1925. Bell Telephone Laboratories 1925–. Mr. Buhrendorf's early Laboratories work included the design of switchboard apparatus and sound recording and reproducing equipment; among the latter were the Mirrophone and the stereophonic equipment demonstrated at the New York World's Fair. During World War II he was concerned with the design of mechanical components of a number of radar systems, particularly antenna drives and range units. After the war he resumed his work on high-quality sound reproduction and more recently has devoted his efforts to the design of magnetic drum units for digital data storage and special machinery for the purification and production of single-crystal semiconductors. He is a New York State Professional Engineer.

Calvin S. Fuller, B.S. 1926 and Ph.D. 1929, University of Chicago. Bell Telephone Laboratories, 1930–. His early work was on organic insulating material, after which he made studies of plastics and synthetic rubber including investigations of the molecular structure of polymers and the development of plastics and rubbers. Since 1948 Dr. Fuller has concentrated on semiconductor research and the development of semiconductor devices. His work led to a technique of diffusing impurities into the surface of a silicon wafer, a preparation basic to the Bell Solar Battery and other silicon devices. He is a member of the A.C.S., an associate member of the A.P.S. and a member of the A.A.A.S.

H. A. Henning, B.S. in Electrochemical Engineering, Pennsylvania State College 1926; Columbia University 1930–33. Bell Telephone

Laboratories, 1926–. Mr. Henning's early Laboratories work was connected with the development of high-quality sound recording and reproducing equipment and techniques. During this interval he developed the 9A disc phonograph reproducer. Other pre-war experience included development of telephone voice recorders, noise reduction studies of the dynamics of teletype equipment, and design of coin collector slug rejectors and coin disposal relays. During World War II he was concerned with improvements to the sound power telephone, and later with development of specialized magnetic sound recording-reproducing systems. After the war he resumed his work on high quality sound recording equipment and supervised the design of the 2A lateral disc feedback recorder. More recently he has been concerned with the principles and design of magnetic drum digital data storage and apparatus. He is currently engaged in investigating the application of square hysteresis loop magnetic cores to digital computer systems.

David, A. Kleinman, S.B. in Chemical Engineering, 1946, S.M. in Mathematics, 1947, Massachusetts Institute of Technology; Ph.D. in physics, Brown University, 1952. Dr. Kleinman joined Bell Telephone Laboratories at Murray Hill in July, 1953. Since then he has studied theory of transistor devices and has been engaged in research in the band theory of solids in the Solid State Electronics Research Department. He is a member of the American Physical Society.

F. J. Morin, B.S. and M.S., University of New Hampshire, 1939 and 1940; University of Wisconsin, 1940–1941; Bell Telephone Laboratories, 1941–. During World War II, Mr. Morin was involved in research on elemental and oxide semiconductors and the development of thermistor materials. Since that time he has worked on fundamental investigations into the mechanism of conduction in silicon, germanium and oxide semiconductors. Mr. Morin is a member of the American Chemical Society and the American Physical Society.

O. J. Murphy, B.S. in Electrical Engineering, University of Texas, 1927; Columbia University, 1928–31. Bell Telephone Laboratories, 1927–. Mr. Murphy's early Laboratories projects included studies of voice-operated switching devices, effects of transmission delay on two-way telephone conversation, and voice-frequency signaling systems. During World War II he was concerned with design and development of the M-9 electrical gun director and related projects. After the war he resumed his research work on signaling systems and more recently has

concentrated on the design of magnetic drum digital data storage apparatus and circuits. He is a member of the A.I.E.E., a senior member of the I.R.E., and is a licensed professional engineer.

M. B. PRINCE, A.B., Temple University, 1947; Ph.D., Massachusetts Institute of Technology, 1951; Bell Telephone Laboratories, 1951–1956; National Semiconductor Products, 1956–. Between 1949–51 he was a research assistant at the Research Laboratory of Electronics at M.I.T. where he was concerned with cryogenic research. At Bell Telephone Laboratories, Dr. Prince was concerned with the physical properties of semiconductors and semiconductor devices and was associated with the development of silicon devices, including the Bell Solar Battery and the silicon power rectifier. Dr. Prince is a member of the I.R.E., the American Physical Society, and Sigma Xi.

HOWARD REISS, B.A., New York University, 1943; Ph.D., Columbia University, 1949; Instructor and Assistant Professor in Chemistry, Boston University, 1949–51; Head of the Fundamental Research Section, Celanese Corporation, 1951–52; Bell Telephone Laboratories, 1952–. Dr. Reiss is engaged in the theoretical chemistry of defects in semiconductors. He is a member of the American Chemical Society, the American Physical Society, Sigma Xi and Phi Lamda Upsilon.

BALDWIN SAWYER, B.E., Yale University, 1943; D.Sc., Carnegie Institute of Technology, 1952; Manhattan Project, University of Chicago, 1943–1946; Instructor and Research Associate in Physics, Carnegie Institute of Technology, 1948–1951; Bell Telephone Laboratories, 1951–. Dr. Sawyer's first work at the Laboratories was on the development of semiconductor devices, especially the silicon alloy junction diode. Since 1953 he has been in charge of a group at Allentown concerned with the growth, measurement and characterization of germanium and silicon crystals for use in semiconductor devices. He is a member of the American Physical Society, the American Institute of Mining and Metallurgical Engineers, Tau Beta Pi, Sigma Xi, and an associate of the I.R.E.

DONALD E. THOMAS, B.S. in E.E., Pennsylvania State University, 1929; M.A., Columbia University, 1932; Bell Telephone Laboratories, 1929–. Mr. Thomas specialized in the development of repeatered submarine cable systems until 1940 when he became engaged in the development of sea and airborne radar. In 1942 he entered military service where he was active in electronic countermeasures research and development.

Following the war he took part in the development and installation of the first deep-sea repeatered submarine telephone cable system between Key West and Havana. During this period he also served as a civilian member of the Department of Defense's Research and Development Board Panel on Electronic Countermeasures. At present Mr. Thomas is engaged in characterization and feasibility evaluation of research models of semiconductor devices. He is a senior member of the I.R.E. and a member of Tau Beta Pi and Phi Kappa Phi.