Theory of the Electrical Properties of Germanium and Silicon

HARVEY BROOKS

Department of Applied Physics, Harvard University, Cambridge, Massachusetts

Introduction

During the past year substantial progress has been made in determining the energy-level structure of the valence-type semiconductors, germanium and silicon. This progress has stemmed in turn from three principal developments:

1. The exploitation of the cyclotron resonance technique by Kittel and Kip *et al.* at Berkeley and by Lax and Dexter *et al.* at M.I.T.

2. Detailed and relatively accurate calculations of the band structure of germanium and silicon from first principles by F. Herman and collaborators at R.C.A.

3. Detailed experimental studies, especially at Naval Research Laboratory, Bell Laboratories, and General Electric, on the location and properties of the energy levels due to impurities and imperfections in Ge and Si.

The first two items have given us an insight into the structure of the valence and conduction bands, whereas the last has enabled us to fill in the picture in the forbidden energy gap. With this picture, it is possible to present an account of the electrical properties of these semiconductors from a deductive point of view, proceeding from the energy level structure as a starting point. It is the aim of this paper to outline such a presentation, and to summarize our understanding of the physical mechanisms underlying the electrical properties of these elements. It will be evident that many of the general principles apply equally well to other semiconductors, and can serve as a basis for the discussion of the compounds in the companion paper by Burstein and Egli.

In Section I, we outline in a qualitative and brief fashion the general principles of energy band theory which are of most relevance to the semiconductor problem. Most of these are ideas which were already well developed in the late 1930's, but are perhaps not very well known to the worker in the field of semiconductor applications. In Section II, we discuss the motion of electrons in energy bands under the influence of external electric and magnetic fields, presenting an elementary theory of the cyclotron resonance experiment as an example. This, in turn, serves as a natural point to summarize the current picture of the continuous energy levels of Ge and Si as it has developed quantitatively from the experiments, and in the light of theory.

In Section III, the theory of localized states is discussed with special emphasis on the influence of the complex structures of the valence and conduction bands in modifying the simple theory of hydrogen-like impurities. Qualitative discussion is also given of the localized levels provided by other than hydrogenic impurities, and of the levels produced by surfaces and by dislocations.

In Section IV, we indicate how the energy level picture is to be used in the statistical mechanical calculation of the equilibrium properties of semiconductors. Since the Fermi statistics as applied to simple semiconductors has been very adequately covered in well-known texts such as Shockley's *Electrons and Holes in Semiconductors*, we concern ourselves here mainly with the modifications introduced by complex band structure, and by the existence of temperature dependent energy levels resulting from the interaction between electrons and thermal vibrations.

In Section V, we present the phenomenological theory of the electrical conductivity, Hall effect, magnetoresistance, and thermoelectric power of Ge and Si, again stressing the influence of the complex band structure. In this section we treat the collision time as an empirical parameter without reference to detailed mechanisms of scattering, which are considered in Section VI. In Section VI, mechanisms of lattice and ionized impurity

scattering are discussed in some detail, and some new results on scattering in the valence band are presented.

Finally, in Section VII, we review the optical properties of Ge and Si in terms of the different classes of allowed and forbidden transitions which may occur on the energy level diagram.

The choice of topics has been selective, rather than comprehensive. The basis of selection of particular topics has often been the author's own interests, and in general the emphasis has been on those topics for which treatments are not readily available in the literature.

I. Properties of Energy Bands

Many properties of solids can be satisfactorily treated by means of a physical model in which each electron is considered to move independently of the others. This does not mean that it is assumed not to interact with the other electrons, but only that the interaction is averaged over the motions of all the other electrons in the system, when considering the motion of a given electron. Such a picture was first introduced by Hartree in connection with the quantum mechanical description of electrons in atoms. The scheme was later modified and improved by Fock, who showed that most of the simplifications made possible by the Hartree assumption could be retained in a model which took partial account of the correlations in the motions of the electrons which were entirely ignored in the Hartree scheme.

For many purposes, the Hartree-Fock description is quite accurate for electrons in solids, especially for the description of the least tightly bound electrons, the ones which are primarily concerned in the electrical properties. The description is a self-consistent one; that is, each electron moves in an electrostatic potential which is made up of two contributions, the potential due to the ions, regarded as fixed charge distributions, and the potential arising from the charge density of all the electrons which are concerned in the self-consistent calculation. This latter charge distribution is computed from the wave functions, which in turn are determined by solution of the Schrodinger equation for each electron in the potential. Self-consistency is achieved in principle when the potential at the start and finish of the problem is the same (*1*).

The reason for the utility of the self-consistent wave functions in the case of solids is that if the wave function for each electron is the solution of a one-electron Schrodinger equation in a triply periodic potential, then the resultant electronic charge distribution, and consequently the potential, computed from these wave functions is also triply periodic with the same periods. Thus if we start the problem with a periodic potential, successive calculations of wave functions and potentials will always lead to

a periodic potential. The properties of the solutions of the wave equation for a single electron in a periodic potential present many special simplicities and general properties which we can describe independently of the particular form of the potential, except for its translational periodicity and other symmetry properties characteristic of the particular crystal structure being investigated. It is the properties of the eigenfunctions, and especially the eigenvalues, of the general periodic potential problem which we now wish to discuss (*2*).

In the first place, we find that there are additional quantum numbers (besides the energy itself) which are "good" quantum numbers or "constants of the motion," and which can therefore be used to label the various energy states. The most important of these is the so-called "reduced wave vector" $\mathbf{k}$. It can be shown that the wave functions for a periodic potential are all of the form

$$b(\mathbf{k}, \mathbf{r}) = u(\mathbf{k}, \mathbf{r}) \exp i\mathbf{k} \cdot \mathbf{r} \tag{1.1}$$

where $u(\mathbf{k}, \mathbf{r})$ is periodic with the same periodicity as the potential. The vector $\mathbf{k}$ spans a 3-dimensional space, known as reciprocal space. However, only the vectors in a limited domain of this space actually have a separate physical significance. This limited region is a unit cell of a lattice in reciprocal space, known as the reciprocal lattice. In the literature, it is usually called the "first Brillouin zone." For our purposes it may be thought of simply as a polyhedron whose shape is determined by the crystal lattice being considered. This polyhedron is illustrated for the diamond lattice, of which Ge and Si are examples, in Fig. 1 (*3*). Each point in the first BZ defines a reduced wave vector $\mathbf{k}$, which is a vector from the center of the BZ to the point. The center can always be defined because, regardless of whether the original crystal has a center of inversion, the unit cell in reciprocal space does. Each reduced wave vector within the unit cell or BZ corresponds to a different eigenstate, and usually a different eigenvalue, for the periodic potential problem. The eigenvalues of the problem are distributed in allowed and forbidden bands of energies. Within each allowed range of energies, the eigenvalue may be expressed as a continuous function of $\mathbf{k}$. The different allowed ranges may be labelled in order of increasing energy (*2*), so that the complete set of eigenvalues for the problem may be labelled by the functions

$$\mathcal{E}_n(\mathbf{k})$$

If we focus attention on one particular value of $\mathbf{k}$, then there will be an infinite set, ascending in energy, of eigenvalues for this $\mathbf{k}$. Although there is some degree of arbitrariness in the labelling of eigenvalues, we can always label them in such a way that if $\mathcal{E}_n(\mathbf{k}) < \mathcal{E}_{n'}(\mathbf{k})$ for one particular value of $\mathbf{k}$, then $\mathcal{E}_n(\mathbf{k}) \leq \mathcal{E}_{n'}(\mathbf{k})$ is satisfied for *every* value of $\mathbf{k}$ in

the BZ. *If* the eigenvalues are labelled in this way, then it can be shown that any surface of constant energy in **k**-space, i.e., a surface whose equation is

$$\mathcal{E}_n(\mathbf{k}) = \text{const.} \tag{1.2}$$

has the same symmetry as the reciprocal lattice.* Thus, for example, if

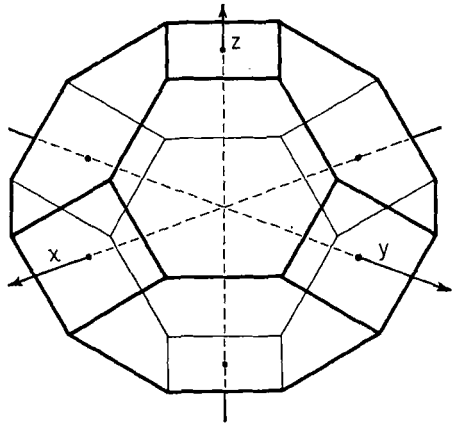

FIG. 1. First Brillouin Zone for the diamond lattice. The diamond lattice consists of two interpenetrating face-centered cubic lattices displaced by the vector $\left(\frac{a_0}{4}, \frac{a_0}{4}, \frac{a_0}{4}\right)$ relative to each other, where a_0 is the cube edge. In the Brillouin Zone the centers of the square faces are at $\mathbf{k} = (2\pi/a_0, 0, 0)$ and five other equivalent points; the centers of the hexagonal faces are at $\mathbf{k} = (\pi/a_0, \pi/a_0, \pi/a_0)$ and seven other equivalent points. The total volume of the zone is equivalent to one-half of an electronic energy level per atom in the crystal.

the crystal lattice has cubic symmetry, all the energy surfaces are left unchanged by any rotation in **k**-space which leaves a cube centered at the origin invariant.

The equal sign is quite important in the above inequality. Depending on crystal symmetry, there will be certain values of **k** for which two or more eigenvalues become equal. When this occurs, we have a "degeneracy." Degeneracy usually occurs in the BZ only for points of particularly high symmetry, for example, the center of the zone, the center of a polyhedral face, or a corner. As we move away from such points in **k**-space, the degeneracy is lifted. In terms of energy surfaces, we say that two or more surfaces touch at certain points in **k**-space and sepa-

* In general, the symmetry of the reciprocal lattice is the same as that of the crystal lattice, except that the symmetry group of the reciprocal lattice always includes the inversion in the origin. In the case of the diamond structure, this means that the energy surfaces actually have greater symmetry than the original lattice; in fact they have the same symmetry as a body-centered cubic crystal structure.

rate as we move away from these points. Actually, surfaces may touch either at points, along lines, or, in some anomalous cases, even along whole surfaces. When touching surfaces do occur, they usually have cusps or sharp angles at the point or line of contact.

At the polyhedral surfaces of the BZ and at the center, it can be shown that the energy functions $\mathcal{E}_n(\mathbf{k})$ have the property that their derivatives normal to the surface vanish. This statement is true for general points of polyhedral faces, but if it happens that two energy surfaces touch at a point on the face, then it breaks down. However, it still follows that the sum of the normal derivatives of the two degenerate energy functions is equal to zero. This situation actually represents a very special case, and if we either slightly perturb the potential so as to lower the symmetry of the crystal, or if we move a small distance away from the degenerate point, then the normal derivative vanishes again.

The condition that the normal derivative of the energy vanishes on the zone boundary implies that at some point on the surface of the BZ there must be either an extremum or a saddle point of the energy. The simplest case is that of a maximum or a minimum. In this case, the bottom or top edge of an allowed band corresponds to a reduced wave vector either at a symmetrical point on the surface of the BZ, such as the center or corner of a face, or at the central point of the BZ. More complicated situations can occur, however, and in principle maxima and minima of the energy can occur at interior points of the zone. As we shall see, this more complex situation seems to be true for both Ge and Si (*4*). In such a case, the band edge occurs, then, at an interior point of the BZ, and if the reciprocal lattice has cubic symmetry, there will be a number of equivalent points in the zone having the same maximum or minimum energy.

In semiconductors, we are always dealing with energy bands which are nearly full or nearly empty; that is, the number of electrons or missing electrons (holes) in any band is only a very small fraction of the total number that could be accommodated. This results in an important simplification of the form of the function $\mathcal{E}_n(\mathbf{k})$; namely, for the charge carriers we are usually interested in, it can be expanded in a Taylor series about the wave vector $\mathbf{k}_0$ which gives the maximum or minimum of $\mathcal{E}_n(\mathbf{k})$. In other words, since $\nabla\mathcal{E}_n(\mathbf{k}_0) = 0$, $\mathcal{E}_n(\mathbf{k})$ is a quadratic function of the components of $\mathbf{k} - \mathbf{k}_0$, where $\mathbf{k}_0$ is the value of the wave vector for which the band edge occurs.

In the most general case we have

$$\mathcal{E}_n(\mathbf{k}) - \mathcal{E}_n(\mathbf{k}_0) = \sum_{ij} a_{ij}\xi_i\xi_j$$

where $\xi_1 = k_x - k_{x0}$; $\xi_2 = k_y - k_{y0}$; $\xi_3 = k_z - k_{z0}$ (1.3)

Equation (1.3) corresponds to energy surfaces which are ellipsoids centered about the point $\mathbf{k} = \mathbf{k}_0$. In a cubic crystal, such as Ge or Si, if the point $\mathbf{k}_0$ is not on an axis or plane of symmetry, there will be 48 band edges in the BZ because of symmetry, and the complete energy surface for a given energy near the band edge will consist of 48 ellipsoids centered about the 48 equivalent points. If $\mathbf{k}_0$ is in a (100)-direction, there will be 6 ellipsoids, each an ellipsoid of revolution about the (100)-axis on which its center lies. If $\mathbf{k}_0$ is on a face of the BZ in the (100)-direction, there will in effect be only 3 ellipsoids, each consisting of two half-ellipsoids on opposite faces of the zone. Values of $\mathbf{k}$ on opposite faces of the BZ are actually equivalent to each other, so that the two half-ellipsoids are equivalent to a single ellipsoid. These two situations are sketched in two dimensions in Fig. 2. In the 3-dimensional case, the half-ellipsoids will be centered on the square faces of Fig. 1, which shows the BZ structure for the diamond lattice. From symmetry it also follows that the unique axis of each ellipsoid (two are equal) must lie along the respective (100)-axis.

Similar situations may arise for $\mathbf{k}_0$ on (111) or (110) axes. In the (111) case, there will be either 8 or 4 ellipsoids of revolution with their unique axes along the various body diagonal (111)-directions. In the case of 4 ellipsoids, they will actually consist of half-ellipsoids centered on the hexagonal faces.* In the case of (110) ellipsoids, the three principal axes can be all different.

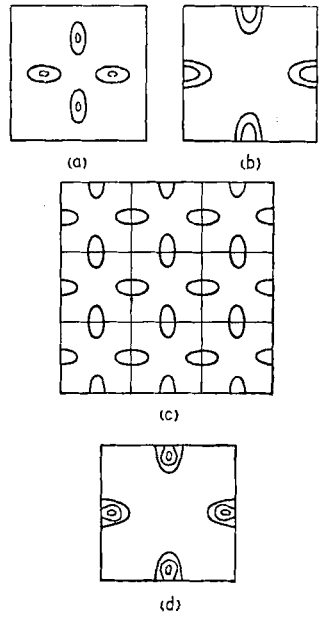

FIG. 2. Elliptical energy contours for a two dimensional simple cubic structure.

(a) Band-edge points at an interior point of the BZ.

(b) Band-edge point at the center of an edge of the BZ.

(c) Case (b) illustrated for the extended wave vector scheme to show continuity of half ellipsoids.

(d) Distorted energy surfaces which result when the band edge points are very close to the zone boundary. Under these conditions the energy ceases to be proportional to k^2 for very small excitation above the band edge.

* An alternative way of visualizing the situation described above is to consider that the domain of $\mathbf{k}$ is all space rather than just the first Brillouin zone. In this case, the functions $\mathcal{E}_n(\mathbf{k})$ may be regarded as functions in k-space which are strictly periodic, i.e. which repeat from cell to cell, each cell being obtained by translation from the original first BZ. In this picture, when an energy surface intersects the surface of a BZ, it continues on into the next cell. Thus, for example, when a half

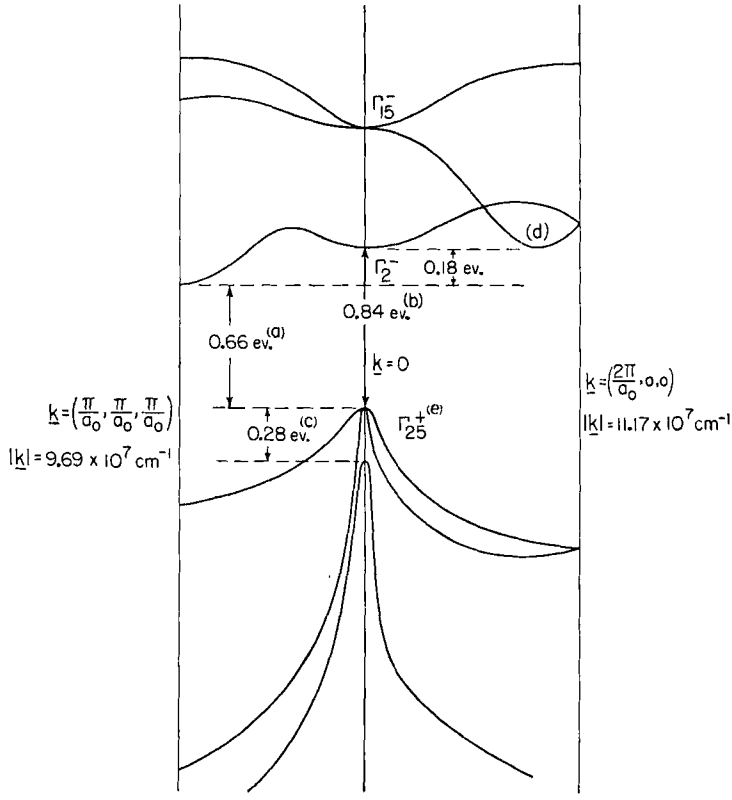

FIG. 3. Energy as a function of reduced wave vector for (100) and (111) directions, as deduced from theoretical calculations and from cyclotron resonance and optical experiments. (A) Germanium; (B) Silicon. The curves are only approximate. Accurately known energy intervals are dimensioned on the diagrams.

(a) Thermal free energy gap at 300°K.

(b) Deduced from optical data of Dash *et al.* (*67*) and of Fan *et al.* (*118*).

(c) Deduced by Kahn (*115*) from optical data of Briggs and Fletcher (*117*).

(d) Deduced by F. Herman (private communication to W. Paul) from optical data on Si-Ge alloys (*103*).

(e) Symbols denote symmetry type at $\mathbf{k} = 0$ according to the scheme of reference *2*.

If the band edge is a point of degeneracy, that is, if two or more energy surfaces are coincident at this point, then the energy surfaces

ellipsoid touches the face of a cell, it is seen to be simply half of a complete ellipsoid which extends an equal distance into the next cell. This situation is also indicated in the sketch in Fig. 2.

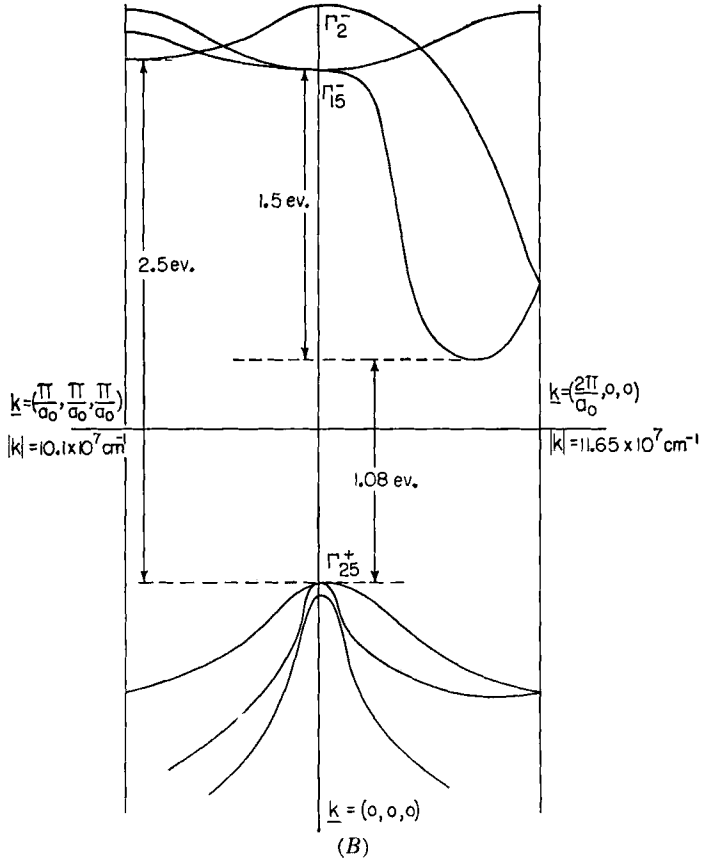

FIG. 3. (*Continued*).

cannot be described as ellipsoids or spheres, but are warped in such a way as to retain their cubic symmetry. The energy difference from the band edge is still proportional to the square of the magnitude of the wave vector difference, but is a more complicated function of the direction cosines of the vector $\mathbf{k} - \mathbf{k}_0$.

Examples of several of the cases mentioned above can be found in Ge and Si. The conduction band in Ge is thought to have prolate ellipsoids along (111)-directions. The conduction band of Si has ellipsoids in (100)-directions. The valence bands of both Ge and Si are thought to be doubly degenerate at their upper edges, corresponding to pairs of warped surfaces in contact at the point $\mathbf{k} = 0$ (*5*). The doubly degenerate band edge in the valence band is nearly triply degenerate, in fact, so that it is

thought that another energy maximum at $\mathbf{k} = 0$ lies very nearby, the separation being determined by spin-orbit interaction.*

The band structure of germanium and silicon as indicated by cyclotron resonance experiments, combined with the extensive theoretical calculations of Herman and his coworkers (*6*), are shown in Fig. 3. What is actually shown in this figure is the plot of the various $\mathcal{E}_n(\mathbf{k})$ for $\mathbf{k}$ along certain particular directions in the BZ. Close study of the figure reveals many of the general characteristics of energy surfaces described above. We see for example, that the slopes of the $\mathcal{E}$ *vs.* $\mathbf{k}$ curves vanish at the end of each curve, corresponding to the face of the BZ. The one exception to this generalization occurs for a pair of curves which become degenerate in the (100)-direction. It is fairly evident that in this case, the sum of the two slopes is equal to zero at the zone face.

We need a method for counting the total number of quantum states which lie between the band edge and a given energy within the energy band, and from this we can compute the number of states per unit energy range, or "density of states," as it is usually called. To make the first calculation, we merely compute the volume in $\mathbf{k}$-space contained inside a given energy surface. For the purposes of this calculation, of course, the "interior" of a surface is the side of the surface on which the edge point of the band lies. The number of states per unit volume of $\mathbf{k}$-space is just $V/8\pi^3$, where V is volume of the crystal. Because one electrons of each direction of spin can be accommodated in each state, the number of electrons is $V/4\pi^3$, or in other words, the largest number of electrons which can be accommodated between a band edge and an energy $\mathcal{E}$, is given by

$$n_e = \frac{V_{\mathbf{k}} V_{\mathbf{r}}}{4\pi^3} \tag{1.4}$$

where we have written coordinate space volume to distinguish it from volume in $\mathbf{k}$-space. From the fact that $\mathcal{E} - \mathcal{E}_0$ is proportional to $\mathbf{k}^2$, it follows that the volume $V_{\mathbf{k}}$ is proportional to $|\mathcal{E} - \mathcal{E}_0|^{3/2}$, irrespective of the shape of the energy surfaces. This relation is always true sufficiently close to the band edge.

II. Motion of Electrons in External Fields

The reduced wave vector $\mathbf{k}$ is a vector which in many ways behaves similarly to a momentum (*7*). If we define

$$\mathbf{P} = \hbar\mathbf{k}$$
$$\hbar = h/2\pi \tag{2.1}$$

* Spin-orbit interaction can be of great importance in removing degeneracies in the band structure. For general considerations on the effect of spin-orbit interaction on energy bands, see reference *5a*.

P is called the reduced momentum, and we can then speak of energy surfaces in **P**-space instead of **k**-space. The energy for a band may then be written as $\mathcal{E}_n(\mathbf{P})$. We know that in ordinary classical mechanics we have the relations

$$\mathbf{v} = \mathbf{P}/m = \nabla_P H(\mathbf{P}, \mathbf{q}); \quad \frac{d\mathbf{P}}{dt} = -\nabla_q H(\mathbf{P}, \mathbf{q}) \tag{2.2}$$

where $H(\mathbf{P}, \mathbf{q})$ is the Hamiltonian, or total energy expressed as a function of coordinates and momenta. In the case of solids, we identify the kinetic energy with the band energy $\mathcal{E}_n(\mathbf{P})$ and assume that the coordinate derivative $\nabla_q H(\mathbf{P}, \mathbf{q})$ just gives the negative of the applied external force. This results in the following relations, which can be demonstrated to follow for electrons in bands, provided the fields are not too large (7)

$$\mathbf{v} = \nabla_P \mathcal{E}_n(\mathbf{P}) \tag{2.3a}$$

$$\frac{d\mathbf{P}}{dt} = \mathbf{F} = -eE - \frac{e}{c}\mathbf{v} \times \mathbf{H} \tag{2.3b}$$

Equation (2.3a) tells us immediately that at a band edge the electron velocity vanishes. This is also true at $\mathbf{k} = 0$, and in a direction normal to the face of a BZ at the BZ surface. Equation (2.3b) can also be written in terms of acceleration as follows

$$\frac{d\mathbf{v}}{dt} = \nabla_P \nabla_P \mathcal{E}_n(\mathbf{P}) \cdot \frac{d\mathbf{P}}{dt} = -\nabla_P \nabla_P \mathcal{E}_n(\mathbf{P}) \cdot \left[e\mathbf{E} + \frac{e}{c} \nabla_P \mathcal{E}_n(\mathbf{P}) \times \mathbf{H} \right] \tag{2.4}$$

The tensor quantity $\nabla_P \nabla_P \mathcal{E}_n(\mathbf{P})$ behaves like the reciprocal of a mass and is usually referred to as the effective mass tensor. It can be either positive or negative, and can indeed be positive and negative for different directions at the same value of $\mathcal{E}$. At the bottom edge of a band, where $\mathcal{E}$ has a true minimum, the effective mass tensor has its three principal values all positive, whereas at the top edge of a band, the principal values are all negative. This implies that electrons near the top of a band are accelerated in the opposite direction from normal electrons. If we add up the individual accelerations of all the electrons in a fully occupied band, we find that the net acceleration of charge vanishes; i.e. no current can be produced by an electric field acting on a fully occupied band of electrons. This is why a solid having nothing but full or empty bands is an insulator at sufficiently low temperatures, in spite of the fact that the individual electrons in the band are free to move throughout the crystal. Strictly speaking, this conclusion, like Equations (2.3), is true only in the limit of weak fields, but for most practical cases it is valid.

For an almost full band, it is more convenient to discuss the behavior of the missing electrons than of those that are present. Since the resultant of the accelerations of all the electrons in the band is zero, the resultant acceleration of the incompletely occupied band is the negative of the resultant of the accelerations of the missing electrons. But since the missing electrons lie near the top of the band if the solid is reasonably close to thermal equilibrium, they have negative masses. Thus for electric fields we have two negatives; that is, the fact that the electrons are missing and that they have negative mass means that the current is in the normal direction. However, if we have a magnetic field present also, the negative mass additionally manifests itself in the velocity in the term $\mathbf{v} \times \mathbf{H}$, and this gives us a total of three negatives which reverses the behavior of an almost filled band of electrons in a magnetic field as compared with an almost empty band. The sign of the Hall voltage thus observed is opposite for nearly full and nearly empty bands. This provides one method of identifying which situation we are dealing with. These results may all be summarized in the statement that missing electrons of negative mass at the top of a band behave in every way with respect to external fields like positively charged particles having mass of the same magnitude, but positive. Such particles are called holes, and for all practical purposes may be thought of as real particles in any phenomenon which does not involve particles leaving the crystal, for this particular behavior is entirely a product of the periodic potential. The sign of the various thermoelectric effects in semiconductors is also different for electrons and holes.

Some confusion has arisen, even in the recent literature, over the physical significance of the effective mass concept and the derived concept of holes. Experiments designed to observe the motion of holes under applied fields, such as the drift mobility experiment of Haynes and Shockley (*8*), show the expected behavior of positively charged particles, but experiments designed to measure directly the inertial mass of charge carriers via the reaction on a solid body due to the acceleration of the carriers by a collapsing magnetic field (the Barnett experiment), always measure the true electronic mass (*9*). The reason is that the apparent mass of an electron in a periodic potential arises because of the possibility of Bragg reflection of an electron by the crystal. In this reflection momentum is transferred to the lattice, but the energy transfer is negligible. Thus the electron gets all the energy, but the lattice takes up just the momentum necessary to make up for the difference between $m_{eff}\mathbf{v}$ and $m_e\mathbf{v}$, where m_{eff} and m_e are the effective and actual masses of the electron, respectively. In an inertial experiment such as the Barnett experiment, the momentum given in this way to the lattice is measured

together with the electron momentum. The true momentum of the electron is just $m_e\mathbf{v}$, where $\mathbf{v}$ is the velocity given by Equation (2.3a), and m_e is the free electron mass (*10*).

A convenient illustration of the above ideas is provided by the following somewhat simplified treatment of the cyclotron resonance experiment on semiconductors (*11*). An electron in a magnetic field will, in general, spiral around the field with a certain natural frequency given in Gaussian units by eH/mc. If an electric field of this frequency is applied to the electron in a plane perpendicular to the magnetic field, the electron will gain energy from the electric field and will travel in ever increasing orbits. In a solid this situation will result in resonant energy absorption, provided that the electron can perform quite a number of orbits before being thrown out of phase with the electric field by a scattering collision, and provided the conductivity is low enough so that the rf field can penetrate below the surface. These conditions have been realized in Ge and Si at liquid helium temperatures with samples of high purity and crystal perfection. Under these conditions, the time between collisions becomes very long. Although the number of electrons present is also very small, this can be enhanced sufficiently to produce observable energy absorption either by infrared irradiation of the sample during the experiment or by applying sufficiently high rf fields to produce electron multiplication. For further details of the experimental procedure the reader is referred to the original papers (*11*).

Let us first consider electrons in the conduction band in Ge. In this case let us assume we have multiple energy ellipsoids with the energy in each given by

$$\mathcal{E}_n(\mathbf{P}) = P_1^2/2m_1 + P_2^2/2m_2 + P_3^2/2m_3 \tag{2.5}$$

where P_1, P_2, and P_3 are the components of reduced momentum along the principal axes of the ellipsoid relative to the minimum point; the three axes are in the ratio of $m_1^{1/2}$: $m_2^{1/2}$: $m_3^{1/2}$. From (2.3a) it follows that the components of velocity are

$$v_1 = P_1/m_1, \quad v_2 = P_2/m_2, \quad v_3 = P_3/m_3 \tag{2.6}$$

The equations of motion of an electron may be written

$$\begin{aligned}
\frac{dv_1}{dt} &= -\frac{eH_3}{m_1c}v_2 + \frac{eH_2}{m_1c}v_3 \\
\frac{dv_2}{dt} &= -\frac{eH_1}{m_2c}v_3 + \frac{eH_3}{m_2c}v_1 \\
\frac{dv_3}{dt} &= -\frac{eH_2}{m_3c}v_1 + \frac{eH_1}{m_3c}v_2
\end{aligned} \tag{2.7}$$

Setting $v_1, v_2, v_3 = v_{10}, v_{20}, v_{30}e^{i\omega t}$, we find the frequency conditions

$$\begin{aligned} \omega &= 0 \\ \omega^2 &= \frac{e^2H_1^2}{m_2m_3c^2} + \frac{e^2H_2^2}{m_1m_3c^2} + \frac{e^2H_3^2}{m_1m_2c^2} \\ &= \omega_1^2\alpha_1^2 + \omega_2^2\alpha_2^2 + \omega_3^2\alpha_3^2 \end{aligned} \tag{2.8}$$

The zero frequency corresponds to motion of the electron along the magnetic field direction. In the second equation ω_1, ω_2, and ω_3 are three frequencies with the field respectively along the three principal axes of the

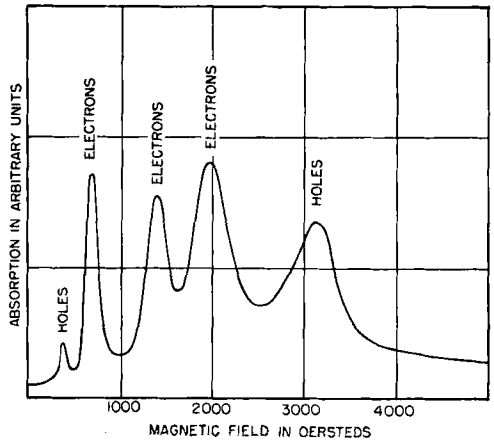

FIG. 4. Typical cyclotron resonance spectrum for germanium. Resonances due to two types of holes and one type of electron are shown. In these experiments the magnetic field is varied while the frequency of the r-f electric field is held constant.

Taken from fig. 2 of Dresselhaus, Kip, and Kittel, "Cyclotron Resonance of Electrons and Holes in Silicon and Germanium Crystals," *Phys. Rev.* **98**, 368 (1955).

ellipsoid, while α_1, α_2, and α_3 are the direction cosines of the field with respect to the ellipsoid axes. In an actual experiment, the frequencies resulting from all of the multiple ellipsoids are observed simultaneously, since there are electrons occupying each of them. For an arbitrary direction of field, these frequencies will in general be different, so that multiple peaks are obtained which shift relative to each other as the crystal orientation is changed relative to the field. A typical "cyclotron resonance spectrum" for Ge is illustrated in Fig. 4. It is obtained by varying the field at fixed frequency. It is to be noted that resonances due to holes and electrons can be observed simultaneously. If circularly polarized rf fields are used (*12*), only one sign of carrier is excited at a time, and the sign of the carrier can be distinguished by the sense of circular

polarization relative to the magnetic field direction, thus providing direct evidence for the physical reality of holes.

In the case of Ge, the experiments indicate that the energy surfaces are prolate ellipsoids along (111)-directions in the crystal. The experiment cannot distinguish between the model of 8 and 4 ellipsoids. For an arbitrary direction of field, in general, four frequencies will be observed with this model, but for the field along a (100)-direction, all the ellipsoids are equivalently oriented, and the single frequency observed is

$$\omega^2 = \tfrac{1}{3}(\omega_1^2 + \omega_2^2 + \omega_3^2) \tag{2.9}$$

The principal masses can be determined and are

$$m_1 = 1.58m_e, \quad m_2 = m_3 = 0.082m_e \tag{2.10}$$

The anisotropy ratio $m_1/m_2 = 19.3$ is an important quantity which will occur in connection with expressions for the dc galvanomagnetic effects.

In the case of Si, the energy surfaces are prolate ellipsoids located along the six (100)-directions. According to cyclotron resonance data, the effective masses are (*11*)

$$m_1 = 0.98m_e, \quad m_2 = m_3 = 0.19m_e \tag{2.11}$$

and the anisotropy ratio $m_1/m_2 = 5.15$.

The theoretical calculations of Herman (*6*) suggest that the edge of the valence band should be triply degenerate at $\mathbf{k} = 0$ with wave functions having symmetry similar to that of atomic p-functions. Including spin, the degeneracy is actually six-fold. Dresselhaus and Kittel (*5*) have pointed out, however, that this degeneracy will be partially removed by spin-orbit interaction, so that there will be one four-fold degenerate state, highest in energy and corresponding to an atomic $j = \frac{3}{2}$ state and one doubly degenerate state corresponding to an atomic $j = \frac{1}{2}$ state. At points other than $\mathbf{k} = 0$ the four-fold degenerate state will be split into two doubly degenerate branches, each of which will give rise to energy surfaces in the form of warped spheres. The $j = \frac{1}{2}$ state will have spherical energy surfaces so long as the distance in energy from the band edge is small compared with the spin-orbit splitting. The expressions for the energy eigenvalues of the two branches are given by

$$\mathcal{E} = -\frac{\hbar^2}{2m_e}[ak^2 \pm \sqrt{b^2k^4 - c^2(k_x^4 + k_y^4 + k_z^4)}] \tag{2.12a}$$

$$\mathcal{E} = -A - \frac{\hbar^2}{2m_e}ak^2 \tag{2.12b}$$

for the upper and lower band, respectively, the zero of energy being taken at $\mathbf{k} = 0$ for the upper branch. Equations (2.12) are a good approxima-

tion when

$$(a + b)k^2 \ll A \tag{2.13}$$

An energy function of the form (2.12) is much more difficult to use in Equation (2.4) than is the case for ellipsoidal energy surfaces. Lax and co-workers (*11*) have carried out the solution by a perturbation method for the case when the momentum parallel to the magnetic field is zero. The result may be expressed as follows

$$\omega = \omega_0 \left\{ A_\pm + B_\pm \left(\frac{1 - 3\cos^2\theta}{2} \right)^2 \right\} \tag{2.14}$$

where

$$A_\pm = a \pm b \mp \frac{c^2}{4b}; \quad B_\pm = \mp \frac{c^2}{8b} \tag{2.15}$$

In these equations, ω_0 is the free electron cyclotron frequency.

For warped energy surfaces, such as occur in the upper branch of the valence band, the motion in a magnetic field has been discussed by Shockley (*13*). The path of the electron in **P**-space is defined by the intersection of a plane perpendicular to **H** with a constant energy surface.

TABLE I. Effective-Mass Parameters for the Valence Bands of Germanium and Silicon

	Ge	Si
a*	13.0	4.1
b*	12.01	2.83
c*	7.1	2.33
A_+†	23.96	6.45
B_+†	−0.53	−0.24
A_-†	2.04	1.75
B_-†	+0.53	+0.24

* The parameters shown in Equation (2.12).

† The parameters in the frequency Equation (2.14).

As a result of the warping of the surface, there is a slightly different cyclotron frequency associated with each value of P_H, the component of reduced momentum along the magnetic field. The observed cyclotron resonance absorption will be the result of the superposition of the absorption lines of individual electrons with different values of P_H, so that the observed line will not have a simple resonance shape. Since the largest number of electrons will have P_H in the neighborhood of zero, it is plausible that the peak of the absorption line should occur for the cyclotron frequency corresponding to $P_H = 0$. This was assumed by Lax and coworkers in interpreting their data, and was later justified by a detailed calculation of the line shape. This calculation predicted an asymmetri-

cal line whose degree of asymmetry was quantitatively confirmed by experiment.

The valence band of Si has a structure similar to that of Ge. The constants of Equations (2.14) and (2.12) for both Ge and Si are given in Table I.

In addition to the approximations already pointed out, the equations of motion (2.4) are semiclassical in character and only represent the motion of the electron when the quantum numbers in the magnetic field are large. The average number of quanta excited at temperature T is of the order

$$n = \frac{kT}{\hbar\omega} \tag{2.16}$$

which is of order unity at liquid He temperature. Thus the semiclassical approach may be barely applicable. At the time of writing, no successful quantum mechanical approach to the problem has been reported (*14*).

For the warped energy surfaces of the valence bands, one would expect to observe not only the fundamental cyclotron frequency given by (2.14), but also certain higher harmonics of this frequency. Some evidence for such harmonics has been found, but their relative amplitude appears to be a strong function of the conditions of excitation (*15*).

III. Properties of Localized States in Semiconductors

The electronic energy levels in bands are the only types of levels which would appear in a perfect crystal of infinite extent. However, real crystals always possess various types of impurities and structural imperfections which usually give rise to new energy levels in the forbidden energy gaps. Usually these are not wholly new levels, but represent states which are split off from the bands. Whereas the wave functions corresponding to the band levels represent charge density which is spread throughout the crystal, the wave functions associated with energy levels in the forbidden range are localized in the vicinity of the imperfection (*16*). These localized levels can occur at impurities, at line imperfections or dislocations, and at grain boundaries or free surfaces. The conditions under which localized levels can occur and their general nature will be developed in the following paragraphs.

The simplest type of localized level, and the only type which is fairly well understood in Ge and Si, is that associated with a substitutional impurity from Group III or Group V of the periodic table, i.e., in the columns adjacent on either side to Ge and Si. As an illustration, let us consider the case of As in Ge. The element As contains 5 electrons outside a closed-shell configuration; it behaves chemically with a valence of 5.

When located in a normal Ge lattice site, the As ion core looks a good deal like the Ge ion core, so to a first approximation we may regard the energy level structure of the crystal as not modified by the substitution of As for Ge in a few positions. However, As does supply an additional electron which has to be accommodated in the energy level structure. Since the valence band is just full in the perfect Ge crystal, the extra electron must go into the otherwise empty conduction band. In this band level, however, the extra electron is free to wander over the whole crystal, and since the As atom would have been neutral with the extra electron, when this electron is lost to the conduction band, the As cell appears as a site of one positive electronic charge. This charge exerts a force on the extra electron in the conduction band, tending to attract it back towards the As cell. This attraction is quite weak in Ge, however, because the intervening Ge crystal between the As and the extra conduction electron behaves like a dielectric medium with a dielectric constant roughly equal to the ordinary macroscopic dielectric constant of Ge, which in this case has a value of 16. Thus at large distances from the As, the conduction electron is subject to an attractive potential (*17*) of magnitude

$$V(r) = -e^2/Kr \tag{3.1}$$

where K is the dielectric constant, and e the electronic charge.

Now for fields which are sufficiently slowly varying, it can be shown that the electron in the conduction band behaves as though the equations of motion (2.3) could be quantized exactly as are the equations of motion for an ordinary free particle in quantum theory. The resulting Schrodinger-like equation is the so-called effective mass equation, whose solutions, for spherical energy surfaces and the potential (3.1), are wave functions and energy values analagous to those of the ordinary hydrogen atom. The wave functions fall off in amplitude exponentially with the distance from the impurity, with effective radius given by

$$r = \frac{\hbar^2}{me^2} K = Ka_h \frac{m_e}{m} \tag{3.2}$$

where a_h is the radius of the first Bohr orbit in hydrogen, a convenient unit of distance in such problems. With this value of r, the total lowering of the energy below the conduction band is

$$\Delta\mathcal{E} = -\frac{e^2}{2a_h} \frac{m}{m_e} \frac{1}{K^2} \tag{3.3}$$

The quantity $e^2/2a_h$ is the ionization energy of hydrogen, and is equal to 13.62 ev. Although the derivation was only approximate, Equation (3.3) turns out to be nearly exact. Essentially, it expresses the fact that

an electron from the conduction band behaves exactly like the electron in a hydrogen atom, except with its mass equal to the effective mass and the effective charge on the nucleus being reduced by a factor of the dielectric constant. If the effective mass were just equal to the free electron mass, Equation (3.3) would give 0.053 ev for Ge and 0.094 ev for Si as the energy required to free an electron from the vicinity of the As impurity into the conduction band.

When we have ellipsoidal energy surfaces, the treatment is somewhat more complex. This problem has been given a semirigorous quantum mechanical foundation by Kittel and Mitchell, by Lampert, and by Luttinger and Kohn (*18*). The complete wave function for the electron in the field of the impurity may be written

$$\psi(\mathbf{r}) = \sum_i \{\phi_i(\mathbf{r})u_0(\mathbf{k}_i, \mathbf{r}) + \nabla\phi_i(\mathbf{r}) \cdot \mathbf{u}_1(\mathbf{k}_i, \mathbf{r})\}e^{i\mathbf{k}_i\cdot\mathbf{r}} \tag{3.4}$$

where the wave function near the edge of the band may be written

$$b_n(\mathbf{k}, \mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}}\{u_0(\mathbf{k}_i, \mathbf{r}) + i(\mathbf{k} - \mathbf{k}_i) \cdot \mathbf{u}_1(\mathbf{k}_i, \mathbf{r})\} \tag{3.5}$$

Here the subscript i refers to a particular band-edge point in reduced wave vector space, and (3.5) is valid for $\mathbf{k}$-vectors only in the immediate vicinity of the band-edge point—over the same region, in fact, in which the approximation of ellipsoidal energy surfaces is valid. The functions u_0 and $\mathbf{u}_1$ are periodic, i.e., the same in every cell of the crystal except the impurity cell, whereas the envelope function $\phi(\mathbf{r})$ is slowly varying from cell to cell. This envelope satisfies the effective mass wave equation

$$-\frac{\hbar^2}{2m_1}\frac{\partial^2\phi}{\partial x_1^2} - \frac{\hbar^2}{2m_2}\frac{\partial^2\phi}{\partial x_2^2} - \frac{\hbar^2}{2m_3}\frac{\partial^2\phi}{\partial x_3^2} + V(\mathbf{r})\phi = (\mathcal{E} - \mathcal{E}_c)\phi \tag{3.6}$$

where $\mathcal{E}_c$ is the energy of the band edge, $\mathcal{E}$ is the actual eigenvalue, and $V(\mathbf{r})$ is the disturbed part of the potential due to the presence of the impurity.

Equation (3.6) cannot be solved exactly, but has been treated by the variational method (*18*) using a trial wave function of the form

$$\phi(\mathbf{r}) = \left[\frac{ab^2}{\pi r_0^3}\right]^{1/2} \exp\{-[a^2x^2 + b^2(y^2 + z^2)]^{1/2}/r_0\} \tag{3.7}$$

where

$$r_0 = K\hbar^2/m_e e^2 \tag{3.8}$$

The energy is minimized with respect to the parameters a and b, using the masses given by (2.10) and (2.11). The results for Ge and Si are as follows:

	Ge	Si	
a	0.368	0.465	
b	0.132	0.270	
r_0/a	$43.5a_h$	$25.8a_h$	(3.9)
r_0/b	$121.1a_h$	$44.4a_h$	
$\mathcal{E} - \mathcal{E}_c$	0.00905 ev	−0.0298 ev	

In Equation (3.9), a_h is the Bohr radius of hydrogen, and the quantities r_0/a and r_0/b represent the extension of the wave functions away

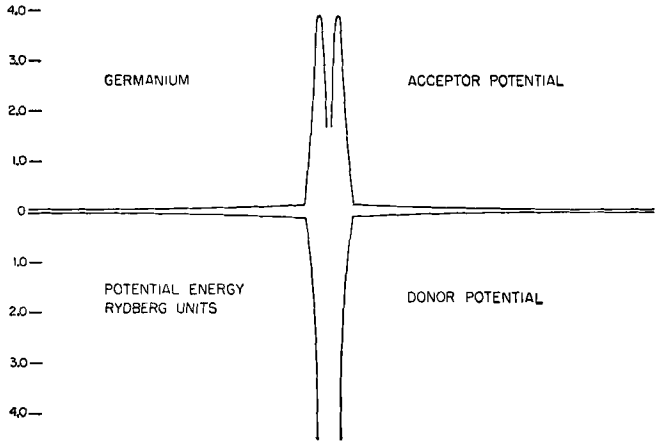

FIG. 5. Sketch of the effective potential experienced by an electron in the field of an impurity. Outside the impurity cell only the part of the potential which differs from the lattice potential is shown. Inside, the complete potential of the impurity is indicated, modified by the fact that the impurity is in a cavity in a dielectric medium of dielectric constant K. Upper figure for donor, lower figure for acceptor.

from the impurity center. It is seen that (3.3) gives approximately the numerical results shown in the last row of (3.9), if we take as the effective mass the geometric mean of the three principal masses.

The ionization energies in (3.9) are extremely small, and in practice the electron is usually separated from the impurity at all except the very lowest temperatures.

Similar techniques have been used by Luttinger and Kohn and by Kleiner (*19*) for the computation of excited states. The results have been verified experimentally in an approximate way by Burstein and coworkers, using optical techniques (*20*).

The potential energy e^2/Kr cannot be valid right down to the origin, nor can the effective mass equation (3.6) be valid inside the impurity cell. When the electron is inside the impurity, it moves as though it had the normal electronic mass and experiences the full unshielded field of

the impurity ion. Thus the total effective potential experienced by the electron is somewhat as sketched in Fig. 5.

Outside the impurity cell, this does not include the periodic part of the potential but only the disturbed part. The complete wave function for the impurity electron is then obtained by matching a solution of the form (3.4) to a solution inside the impurity cell. The matching conditions have been treated in detail for various types of band structures by Brooks and Fletcher (*21*). The matching conditions alter the eigenvalues from (3.3) or (3.9). For Group V donors, the effect is almost always to

TABLE II. Ionization Energy for Group V Impurities in Germanium and Silicon

Impurity	Germanium		Silicon	
	Theory[a]* (ev)	Experiment (ev)	Theory[b]* (ev)	Experiment (ev)
Sb	0.012	0.0096 (*22*, *23*)	0.037	0.039 (*24*)
P	0.012	0.0120 (*22*, *23*)	0.039	0.044 (*24*)
As	0.012	0.0127 (*22*, *23*)	0.043	0.049 (*24*)
Hydrogenic	0.012		0.030	

* For spherical energy surfaces, based on estimates from spectroscopic data.

[a] Calculated for $m_{eff} = \frac{1}{5}m_e$, $K = 16$.

[b] Calculated for $m_{eff} = \frac{1}{3}m_e$, $K = 12$.

increase the impurity ionization energy beyond the hydrogenic value. Rough estimates are shown for $m/m_e = 0.2$, $K = 16$, and $m/m_e = 0.33$, $K = 12$ in Table II. The first case corresponds roughly to the Ge conduction band, and the second to the Si conduction band.

Let us now turn to the consideration of Group III impurities. In this case the impurity atom contains fewer electrons outside a closed shell, so that in the first approximation we have the normal electronic energy level structure of pure germanium with one electron missing from the top of the valence band. This electron behaves like a hole, or positive electron. Since in this approximation all the electronic wave functions are band functions, the corresponding charge is spread uniformly throughout the crystal, including the impurity cell. Since the impurity atom now has one nuclear charge less than normal, it appears negatively charged, and at large distances from the impurity the hole moves in a dielectric shielded potential exactly analagous to the potential seen by the conduction electron in the donor case. Thus, so far as motion outside the impurity cell is concerned, the treatment of an "acceptor" is precisely analagous to that of a donor. We have already seen however, that inside the cell we have an electron of normal mass and no dielectric con-

stant. What becomes of the hole concept inside the impurity cell? We may look at the problem most conveniently in terms of the motion of an electron in the state of the crystal in which the acceptor level is occupied by an electron. The electron in the acceptor comes from the top of the valence band, and is therefore an electron of negative mass. Outside the impurity cell it moves in the repulsive potential of the impurity cell, but since it has negative mass, it is nevertheless bound. The wave function is of the general form of Equation (3.4) with $\phi(\mathbf{r})$ localized. This wave function must be matched at the surface of the impurity cell to a wave function inside the cell for a normal electron moving in the field of the impurity ion plus the negative charge due to the other electrons in the valence band. The resulting potential inside the cell is indicated in Fig. 2 (upper curve). It is found that the matching of wave functions can be carried out straightforwardly (*21*), and again the influence of the impurity cell itself on the hydrogenic eigenvalue is rather small, at least for the dielectric constant and effective mass of Ge and Si.

The effective mass equation satisfied by the acceptor electron is that appropriate to the degenerate valence band. The situation for warped surfaces and degenerate bands has been considered in an approximate fashion by Kittel and Mitchell and by Luttinger and Kohn (*18*). It is found that (3.6) must be replaced by a system of coupled differential equations, which result in mixing of the wave functions from the different degenerate bands. Recently a variational solution of the coupled differential equations has been given by Kohn and Schechter (*21a*). For Ge the resulting eigenvalue is:

	Calculated	Experimental (B doped)	
$\mathcal{E}_v - \mathcal{E}$	-0.0089 ev	-0.0104 ev	(3.10)

The theory agrees with experiment in predicting almost identical ionization energies for donors and acceptors in Ge. In both cases the energies are slightly too small, a fact which can probably be accounted for by the correction for the central cell. There is not a correspondingly satisfactory treatment for acceptor states in Si since in this case the ionization energy of the acceptor is of the same order of magnitude as the spin-orbit splitting.

So far the treatment has been based on the assumption that the interaction of impurities could be ignored. Because of the large orbits of impurity states, however, the orbits begin to overlap for rather low impurity concentrations. This effect is further enhanced by the lack of spherical symmetry of the wave functions, as shown by Equation (3.9).

An approximate condition may be derived for the critical mean distance between donors at which the ionization energy disappears. At this

concentration, electrons can jump from impurity to impurity, and the single impurity level is broadened out into a band. The critical separation of impurities occurs when the "impurity band" begins to overlap the conduction band. The condition is (*25*)

$$\frac{D}{r_1} = \ln 6\left(\frac{D}{r_1} + 1\right) \tag{3.11}$$

giving $D/r_1 = 3.23$, where D is the critical spacing and r_1 is given by (3.12).

$$r_1 = r_0/b \tag{3.12}$$

The critical concentration is $1/D^3$ and takes the values

$$\begin{aligned} n_D &= 1.2 \times 10^{17} \text{ donors/cm}^3 \text{ for Ge} \\ n_D &= 1.8 \times 10^{18} \text{ donors/cm}^3 \text{ for Si} \end{aligned} \tag{3.13}$$

Experimental results for donor ionization energy as a function of donor concentration for Ge are shown in Fig. 6 (*26*). The agreement between

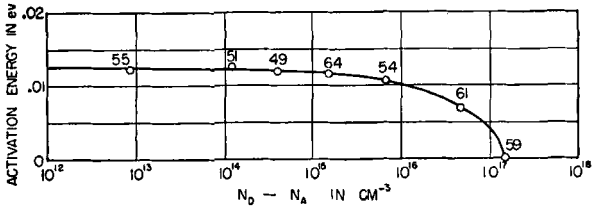

FIG. 6. Ionization energy of donors as a function of donor concentration in n-type germanium. From P. P. Debye and E. M. Conwell, *Phys. Rev.* **93**, 693, 704 (1954).

theory and experiment is seen to be excellent. Until the conduction band structures were elucidated by cyclotron resonance experiments, a considerable amount of theoretical work was done to explain the critical concentration (*27*), including the effect of the random distribution of impurities. The present results appear to indicate that the effect is a perfectly straightforward one of overlap of wave functions, and does not involve any subtle considerations.

The existence of impurity banding was first suggested on experimental grounds by Hung (*28*), and considerable experimental evidence for it has accumulated since. It results in an unexpectedly large conductivity at low temperatures. A number of attempts have been made to calculate the energy level structure and the electronic transport properties associated with impurity bands (*29*). The difficult feature of the problem arises from the random locations of the impurity atoms. The estimates

discussed in Equations (3.11) and (3.13) are based on the assumption of a regular lattice of impurities, and give surprisingly good values for the critical concentration.*

We should expect that, in addition to the ground hydrogenic state which we have computed in Equation (3.9), there should exist excited states of the impurity center. Burstein and coworkers have actually found evidence of such levels in silicon by long wavelength infrared absorption measurements at liquid helium temperature (*30*). These results will be taken up in more detail in Section VII. A theory for excited donor levels in silicon has been given independently by Luttinger and Kohn and by W. Kleiner (*19*). The spacing of the excited levels agrees very accurately with the theory, as is to be expected from the fact that the excited wave functions have p-like symmetry and are therefore much less influenced by the impurity cell than is the ground state.

When we come to consider imperfections other than Group III or Group V impurities, we must pay more attention to what happens to the wave functions inside the impurity cell. Almost no theoretical work has been done on this problem. The simplest case is, perhaps, that of a neutral impurity center which has a very different ion core potential than the rest of the atoms of the crystal. We will assume that the disturbance in potential is confined to the impurity cell itself. If the potential is nearly like the host crystal, there will be no bound level. Rather each conduction-band electron as it goes by will be scattered and will give rise to some net localization of charge. As the potential gets deeper compared with the host crystal, a bound level will eventually split off from the conduction band. For this to happen in Ge, the mean cell potential has to be of the order of 5 ev lower than for the surrounding Ge atoms (*25*). Since the impurity cell now has a negative charge after it has captured an electron, this bound level is, by definition, an acceptor level in spite of the fact that it is closer to the conduction band than to the valence band. As the impurity potential gets deeper, the bound level gradually moves towards the valence band until it finally merges with it. At this point, the valence band accommodates one extra electron. At some stage, probably before the first level merges with the valence band, a second bound level may split off from the conduction band. If this second bound level is occupied by an electron, the impurity cell can become doubly charged. The case which seems to be most common in practice is that in which two bound levels exist in the forbidden gap, one near the valence

* It should be noted that an impurity band develops for impurity concentrations considerably below those given by Equation (3.12). The most important electrical effects occur when the banding is appreciable, but before the impurity band has broadened out to overlap the bottom of the conduction band.

band and the other near the conduction band. Acceptor levels of this type occur for Au and for many of the iron group transition metals in Ge (*31*). In the case of gold in germanium, an additional level about 0.05 ev from the valence band has recently been established by Dunlap as a donor level (*32*).

When the potential of the neutral impurity is deep enough to cause the first bound level to merge with the valence band, there will be localization of negative charge near the impurity cell. This will produce a long-range Coulomb field which can bind a hole in a normal acceptor state similar to that of a Group III impurity. This will represent the normal situation when a neutral impurity of this type is added, since the valence band can then accommodate one more than the number of available electrons and will have a hole in it.

Some of the situations which occur experimentally are illustrated in Table III. Not all of these results are established experimentally beyond question.

In the case of Group II or Group VI impurities, we might expect to find He ion-like levels. In this case, however, the influence of the potential in the impurity cell is so great, that it is difficult to find any correlation between observed levels and a hydrogenic model.

It is also possible to imagine a neutral impurity for which the potential is so shallow that an additional electron would not be bound, but would go into the conduction band, although its wave function would be partially localized near the impurity. If we now consider the neutral crystal and examine the energy level of the last electron, we find that the impurity becomes positively charged when this electron is removed, but that the added charge is sufficient to bind the electron locally in a non-hydrogenic level. Such a level could even approach close to the valence band, but would still be a donor level, by definition. An example of this situation seems to be the case of gold in silicon, reported by Taft and Horn (*33*). An energy level diagram for Si is shown in Fig. 7. In addition to the donor levels mentioned, the diagram shows two levels of unknown origin found in Si by Haynes and Hornbeck (*34*). These deep-lying levels, called traps, have a remarkable property: they accept electrons, but cannot apparently capture holes directly from the valence band. This suggests that they must be positively charged even when occupied by an electron; in other words, they are double donors. In this case, a potential barrier could lie several tenths of a volt above the edge of the valence band, and the thermal energy of holes would not be sufficient to surmount it. If this energy-level picture is correct, there must exist an additional energy level corresponding to a doubly charged impurity which is merged with the valence band or is split off below it.

TABLE III. Summary of Observed Impurity Levels in Germanium and Silicon*

Element	Type	$E_C - E_T$ (ev)	$E_T - E_v$ (ev)	References
Germanium				
B	A		0.0104	G1, C1, D5
Al	A		0.0102	G1
Ga	A		0.0108	G1
In	A		0.0112	G1
P	D	0.0120		G1
As	D	0.0127		G1, D1, D5
Sb	D	0.0096		G1
Li	D	0.0093		M3
Zn	A		0.029	D2, D4, B1
Cu(1)	A,R		0.25	B2, K1, B1, D4, R3
Cu(2)	A		0.040	M1, D4, K1
Au(1)	D		0.05	D6, D7
Au(2)	A,R		0.15	D3, N1, K1, K2
Au(3)	A,T	0.18		D3, N1, D6
Ni(1)	A,R		0.23	B2, N2, M4, M5
	A,T			
Fe(1)	A,R		0.34 ± 0.02	N2, T2, T5
(2)	A,T	0.27 ± 0.02		N2, T2, T5
Co(1)	A,R		0.25 ± 0.01	N2, T3, T4
(2)	A,T	0.31 ± 0.01		N2, T2, T3
Mn(1)	A,R		0.16 ± 0.01	T6
(2)	A,T	0.35 ± 0.01		
Pt	A		0.04	D4
"Deathnium"	A,R		~0.25	H1, B2, F1
Frenkel defects	A	0.2		B3, J1, L1
Silicon				
B	A		0.045	M2, B4, M3
Al	A		0.057	M2, B4
Ga	A		0.065	M2, B4
In	A		0.16	M2, B4, N3
P	D	0.044		M2
As	D	0.049		M2, M3
Sb	D	0.039		M2
Li	D	0.033		
Au(1)	D		0.33	M2, T1
Electron traps (1)	D,T	0.57		H2
(2)	D,T	0.79		H2

Key to Symbols: D = donor level, A = acceptor level, R = recombination center, T = minority carrier trap (probably double donor or acceptor).

* Based on J. A. Burton, *Physia* **20**, 845 (1954).

B1 Burstein, E., Davisson, J. W., Bell, E. E., Turner, W. J., and Lipson, H. G., *Phys. Rev.* **93**, 65 (1954).

B2 Burton, J. A., Hull, G. W., Morin, F. J., and Severiens, J. C., *J. Phys. Chem.* **57,** 853 (1953).
B3 Brown, W. L., Fletcher, R. C., and Wright, K. A., *Phys. Rev.* **92,** 591 (1953).
B4 Burstein, E., Picus, G. S., and Sclar, N., "Optical and Photoconductive Properties of Silicon and Germanium," ONR Photoconductivity Conference Nov. 1954, to be published by Wiley, New York.
C1 Conwell, E. M., *Phys. Rev.* **94,** 1416 (1954).
D1 Debye, P. P., and Conwell, E. M., *Phys. Rev.* **93,** 693 (1954).
D2 Dunlap, W. C., Jr., *Phys. Rev.* **85,** 945 (1952).
D3 Dunlap, W. C., Jr., *Phys. Rev.* **91,** 1282 (1953).
D4 Dunlap, W. C., Jr., *Phys. Rev.* **29,** 21 (1954).
D5 Debye, P. P., *Phys. Rev.* **91,** 208A (1953).
D6 Dunlap, W. C., Jr., *Phys. Rev.* **97,** 614 (1955).
D7 Dunlap, W. C., Jr., *Bull. Am. Phys. Soc.* **30,** 12 (1955), abstract F2.
F1 Fan, H. Y., Navon, D., and Gebbie, H. A., *Physica* **20,** 855 (1954).
G1 Geballe, T. H., and Morin, F. J., *Phys. Rev.* **95,** 1085 (1954).
H1 Hall, R. N., *Phys. Rev.* **87,** 387 (1952).
H2 Hornbeck, J. A., and Haynes, J. R., *Phys. Rev.* **97,** 311 (1955).
J1 James, H. M., and Lark-Horovitz, K., *Z. physik Chem.* **198,** 107 (1951).
K1 Kaiser, W., and Fan, H. Y., *Phys. Rev.* **93,** 977 (1954).
K2 Kaiser, W., and Fan, H. Y., *Phys. Rev.* **93,** 911A (1954).
L1 Lark-Horovitz, K., *in* "Semi-Conducting Materials" (H. K. Henisch, ed.), p. 47. Academic Press, New York, 1951.
M1 Morin, F. J., and Maita, J. P., *Phys. Rev.* **90,** 337 (1953).
M2 Morin, F. J., Maita, J. P., Schulman, R., and Hannay, N. B., *Phys. Rev.* **96,** 833 (1954).
M3 Morin, F. J., and Maita, J. P., *Phys. Rev.* **96,** 28 (1954).
M4 Maesen, F. van der, Penning, P., and Wieringen, A. van, *Philips Research Rept.* **8,** 241 (1953).
M5 Maesen, F. van der, and Brenkman, J., Electrochem. Soc. May, 1954.
N1 Newman, R., *Phys. Rev.* **94,** 278 (1954).
N2 Newman, R., and Tyler, W. W., *Phys. Rev.* **94,** 1419 (1954).
N3 Newman, R., G.E. Report No. 55-RL-1257, Feb. (1955).
T1 Taft, E. P., and Horn, F. H., *Phys. Rev.* **93,** 64 (1954).
T2 Tyler, W. W., Woodbury, H. H., and Newman, R., *Phys. Rev.* **94,** 1419 (1954).
T3 Tyler, W. W., Newman, R., and Woodbury, H. H., *Phys. Rev.* **97,** 669 (1955).
T4 Woodbury, H. H., and Tyler, W. W., *Bull. Am. Phys. Soc.* **30,** 11 (1955), abstract F1.
T5 Tyler, W. W., and Woodbury, H. H., *Phys. Rev.* **96,** 874 (1954).
T6 Tyler, W. W., Newman, R., and Woodbury, H. H., *Phys. Rev.* **98,** 461 (1955).

At the present time, practically no theoretical work has been carried out to explain the position of impurity levels other than the simple hydrogenic ones. We have seen, however, from earlier arguments that the mean potential of the impurity can vary between rather wide limits and still give rise to essentially hydrogenic levels. The problem in the general case is an extremely complicated one quantitatively, although the quali-

tative features described above can be fairly readily understood and simple models can be constructed with the properties described.

Some rough qualitative considerations may be adduced to explain some features of the results with non-hydrogenic impurities. For example, it appears to be a good empirical generalization that elements which produce deep-lying acceptor levels in Ge give rise to donor levels in Si. The lattice constant of Ge is 4.2% larger than that of Si, so that an Au atom

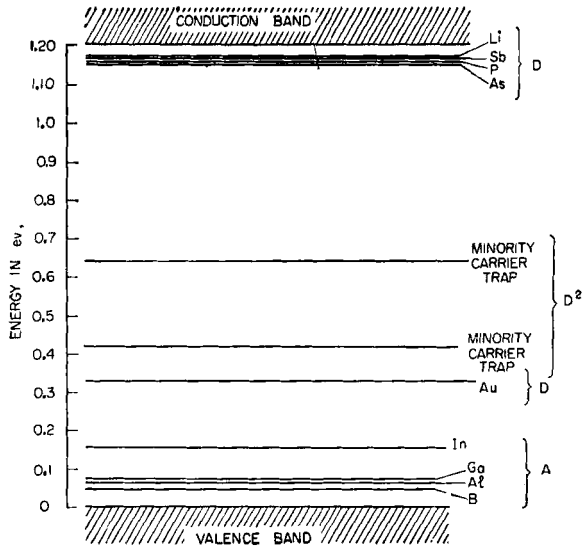

FIG. 7. Energy-level diagram for the forbidden gap in silicon, including trapping levels of unknown origin: J. A. Hornbeck and J. R. Haynes, *Phys. Rev.* **97,** 311 (1955). Since energy levels are deduced from thermal activation energies, they must be interpreted as energy levels appropriate to 0°K, and the diagram is drawn accordingly. A = acceptor level, D = donor level, D^2 = double donor (i.e., when level is occupied, it still carries one positive charge).

would be more compressed in Si. Since the gold is compressed anyway, this will have the effect of raising all the levels for Au in Si relative to those in Ge. Apparently this rise is sufficient to bring the donor level which is nearly merged with the valence band in Ge well up into the forbidden gap in Si and to cause the other two acceptor levels to merge with the conduction band in Si. There is one piece of evidence against this picture. W. Paul has made measurements of the effect of pressure on the upper of the two gold levels in n-type Au-doped Ge (*35*). These measurements indicate that an increase in pressure actually causes the gold level to move *away* from the conduction band and remain nearly fixed relative to the valence band. The two pieces of information are hard to rec-

oncile at the moment. Present theory should make it possible to make more quantitative calculations on such questions, however.

It seems surprising that so many localized levels corresponding to different states of ionization of Au can exist within the narrow forbidden gap, since the self-energy of two electrons both localized in the same cell is many times larger than the size of the gap. However, calculations for a simplified model (*21*) show that even for a variation of potential in the impurity cell up to 10 times the size of the gap, a level remains localized within the gap. Thus the experimental result is at least reasonable.

Isolated impurities constitute only one of the many types of imperfections which can occur in semiconductors as well as metals. Other types include lattice vacancies and interstitials, line imperfections or dislocations, and surface imperfections such as grain boundaries or, indeed, free surfaces. All such imperfections can give rise to localized states with energy levels in the forbidden gap. Practically nothing is known about such levels from a theoretical standpoint, and experimental information is in a much hazier state than is the case for impurities and energy bands themselves. The following facts are pertinent to this discussion:

1. Germanium can be made to flow plastically by a slip mechanism above about 500°C. Dislocations put into a Ge single crystal by bending give rise to acceptor levels whose total number is of the order of magnitude of the number of atomic sites along the dislocations (*36*).
2. The free surface of Ge probably possesses surface traps whose density is of the order of at least $10^{11}/cm^2$, and which behave essentially like acceptor levels below the center of the forbidden gap (*37*).
3. Grain boundaries in Ge, which can always be described by suitable arrays of dislocations, give rise to acceptor levels most of which lie very close to the top of the valence band (*38*). Indeed, in gold-doped Ge which is nearly insulating at liquid N_2 temperatures, grain boundaries act as short-circuit paths of thin p-type layers.
4. Vacancies and interstitials generated by radiation bombardment or by heat treatment give rise to acceptor levels in Ge, and to paired donor and acceptor levels in the middle of the forbidden gap in Si (*39*, *40*).

The level structure produced by dislocations in Ge appears to be rather complicated. Pearson, Read, and Morin (*41*) investigated the electrical characteristics of n-type Ge following plastic deformation, and concluded that dislocations produce an acceptor level about 0.2 ev below the edge of the conduction band. Gallagher and Tweet (*36*) investigated the electrical properties of gold-doped Ge in which dislocations had been introduced by bending. The doping was such that the material was high-

resistivity p-type at low temperatures before treatment. After treatment the material became lower-resistivity p-type, the data being interpretable in terms of an acceptor level about 0.05 ev above the top of the valence band. They also found that plastically bent n-type Ge remained n-type near the neutral axis, but changed to p-type near the fibers of maximum strain. This qualitative result again supports the idea that acceptor levels are introduced which lie below the center of the forbidden gap. The apparent contradiction between the two results seems to originate in differences in the heat-treatment following deformation. The low-lying acceptor state near the valence band seems to disappear after extensive annealing and is probably associated with debris of the plastic deformation in the form of isolated vacancies generated according to a mechanism originally suggested by Seitz (*42*). It is interesting to note that certain orientations of grain boundary also appear to generate acceptor states at about 0.05 ev above the valence band, whereas other orientations give acceptors lying even closer to the valence band (*43*). The conductivity associated with grain boundaries probably results from holes in the space-charge region surrounding the boundary, rather than from conduction by the surface acceptor states in the boundary itself.

It is interesting to speculate on the nature of the energy-level structure produced by extended imperfections such as dislocations. Shockley has suggested the hypothesis of "dangling bonds" (*44*) which, in the mode of description we have been using, may be regarded as neutral centers having a potential much deeper than that of the surrounding Ge lattice. The dangling bonds arise essentially from the fact that Ge atoms near the center of an edge dislocation do not have normal coordination. Thus the dislocation may be regarded as a linear array of neutral centers whose linear density depends on the degree of edge character of the dislocation. These centers are so close together, however, that they influence one another strongly when they capture electrons and so cannot be treated like isolated impurities. As they become filled with electrons, they become charged and repel further electrons, so that they cannot become fully occupied. The line of negative charge on the dislocation induces a space charge in the surrounding material, and results in severe disturbance of the whole potential distribution around the dislocation. Shockley (*44*, *45*) has suggested that extended imperfections may lead to one- or two-dimensional energy bands. There seems to be little evidence in favor of such conducting states, however, and it seems more probable that the electrons are more localized and form, in effect, linear or two-dimensional lattices of charges. This will be so both because of inevitable irregularities in the dislocations or surfaces and because of characteristics which are inherent in the nature of the one- or two-dimensional problem as com-

pared with the three-dimensional one. Read (*44*) has considered the nature of the statistical mechanical problem which occurs when we have dense distributions of acceptors on a line or surface. The general idea may be stated simply as follows. Suppose $-E_t$ is the energy gained when an electron falls from the conduction band into the first trap. Then when there are n electrons in traps, the total energy of the system is given by

$$-nE_t + \alpha e^2 n^2/K \tag{3.14}$$

where α is a number of order unity, i.e., between 1 and 10. The second term arises from the electrostatic interaction of the trapped charges. Minimizing with respect to n, we obtain

$$\begin{aligned} n &= KE_t/2\alpha e^2 \\ E_{\min} &= -KE_t^2/4\alpha e^2 \\ E_{\min}/n &= -E_t/2 \end{aligned} \tag{3.15}$$

This analysis is for a one-dimensional imperfection. Taking E_t as 0.2 ev, we find $n = 1.76K/\alpha \times 10^6$ filled traps per centimeter of dislocation length.

In (3.14) and (3.15) we have neglected the influence of the positive charge induced in the surrounding semiconductor by the line of negative charge. This can be shown merely to alter the effective value of α, however. Actually α is not quite a constant, but, as shown by Read (*44*), is given explicitly by

$$\alpha = \ln (n^{3/2}/\pi^{1/2}N_D^{1/2}) - 0.866$$

where N_D is the density of donors per cubic centimeter in the bulk semiconductor (strictly, the excess of donors over acceptors). This calculation involves the assumption that the charges are localized at the acceptor sites.

As an example, we may take $N_D = 10^{15}$ donors/cm^3, corresponding to 1.7 ohm-cm material. We find $n = 3.7 \times 10^6$ trapped electrons per centimeter of dislocation for the maximum number which can be accommodated. Since α decreases with increasing N_D, there is an increase in the number of electrons which can be accommodated with decreasing resistivity of material. For a spacing between acceptor sites of 4 Å (a typical value), the sites will be only 15% occupied in the example given, even at absolute zero. For this example α is about 3.

For the two-dimensional imperfection, the situation is somewhat more complicated, the total energy being given by

$$-nE_t + \alpha e^2 n^{3/2}/K - \frac{5\pi e^2}{12} n^3/KN_D \tag{3.16}$$

where N_D is the density of ordinary atomic impurities in the bulk of the n-type semiconductor. Because of the last term, (3.16) does not have an absolute minimum. Instead, the minimum is determined by the maximum available number of states. A calculation with plausible numbers, in fact, shows that the middle term in (3.16) is always negligible compared with the other two, so that the self-energy of the surface charge is unimportant. Thus, there is a fundamental difference between line and surface imperfections. The middle term of (3.16) is usually ignored without discussion in the literature dealing with surface states.

A type of localized atomic state which is of great practical importance is the so-called recombination center. Such centers were first introduced for Ge and Si by Hall (*46*) and by Shockley and Read (*47*).

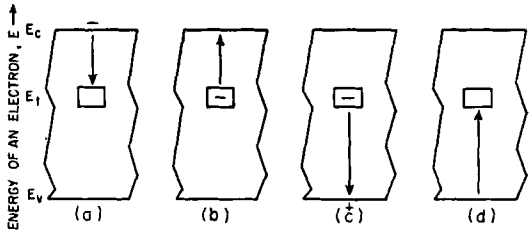

FIG. 8. The recombination mechanism of Read, Shockley, and Hall. The basic processes involved in recombination by trapping: (a) electron capture, (b) electron emission, (c) hole capture, (d) hole emission. From W. Shockley and W. T. Read, Jr., *Phys. Rev.* **87**, 835 (fig. 1) (1952).

They are needed to explain the observed recombination rate of holes and electrons in semiconductors, since direct recombination appears to be a relatively improbable process (*48*). The process envisaged is shown in Fig. 8. An acceptor center near the middle of the forbidden gap captures an electron, to become negatively charged, following which it captures a hole. The process could also occur in reverse order with a donor-type center. Very little is known about the nature or origin of recombination traps. Certain elements, dissolved in Ge, are known to increase the recombination rate drastically (*49*). Examples are Cu, Ni, Co, Fe, and Mn, but they are effective only at temperatures such that the corresponding acceptor centers are singly charged (*50*). On the other hand, there is fairly convincing evidence that structural imperfections also affect recombination. For example, Ge which is quenched from a high temperature shows increases in recombination even when great precautions are taken to keep copper away (*39*). Other experiments show a good correlation between recombination rate and crystal perfection as determined by X-ray line broadening (*51*). Although there may be considerable question about the validity of the interpretation of the X-ray data in terms of a

dislocation density, the general correlation with recombination seems unquestionable. There seems also to be a correlation between recombination rate and impurity content (*52*). Little theoretical work has been carried out on the trapping process itself. Trapping can occur theoretically either by a radiationless process in which energy is conserved by the lattice vibrations, or by a radiative process in which the energy is carried away by one or more light quanta. However, unless the trapping occurs via excited states of the center in several steps, a multiphonon process would be involved for the radiationless transition in Ge and Si (*54*), and this has very low probability. The possibility that capture in the center occurs with the emission of radiation has been investigated by Aigrain (*53*), who has searched for the recombination radiation associated with trapping in the Ge recombination level which is 0.2 ev above the valence band (*46*) and has found some evidence for its existence.* Wannier (*55*) has discussed the trapping problem on the basis that the rate determining process is the slowing down of a conduction electron to essentially zero velocity, in which case it is assumed to be captured with certainty. With this extreme assumption, a lower limit for the lifetime of an electron in the presence of empty traps can be set at 10^{-9} sec. For the deeper electron traps found in p-type Si (0.79 ev below the conduction band), Hornbeck and Haynes (*34*) estimate the mean life of an electron in the presence of empty traps (density 10^{13} cm^{-3}) at 2×10^{-8} sec corresponding to a cross-section per trap of 3×10^{-13} cm^2, which in turn corresponds to an effective trap radius of $60a_h$ or $5Ka_h$.

Such very large cross-sections appear to be associated only with so-called minority carrier traps, which are now believed to be double donors or acceptors and therefore present a strong Coulomb attraction for the approaching carrier. In the case of recombination centers, the cross-sections are generally much lower. For Ni in Ge, for example, the cross-section for holes is about 4×10^{-15} cm^2, whereas for electrons, it is about 10^{-16} cm^2, the difference reflecting the fact that Ni is an acceptor and therefore presents unit negative charge for capture of holes, but is neutral for the capture of electrons (*49*).

IV. Statistical Mechanics of Semiconductors

The application of Fermi statistics to the calculation of the population of energy levels in a semiconductor in thermal equilibrium has been discussed rather completely in the existing literature, and we therefore refer the reader to these treatments (*56*) in the interests of brevity. There

* More recent work indicates that the radiation observed may be due to transitions between the different branches of the valence band (Aigrain, private communication).

are a few points which have received some attention recently, which are worth mentioning in this review.

The first point concerns the significance of the energy which occurs in the Fermi factor when the energy levels of a semiconductor have an explicit temperature dependence. This question has been discussed by Rushbrooke in very general terms (*57*), and more specifically as applied to semiconductors by Landsberg (*58*) and by James (*59*). The argument which follows is essentially due to Landsberg.

The concept of individual particle levels, as envisioned in Fermi statistics, is only an approximation. Actually the energy of each electron depends also, at least to a slight degree, on the states of all the other electrons and on the excitation of each of the lattice oscillators, that is, on the state of vibration of the crystal. We use the subscript s to denote a given electronic energy level and the subscript j to describe the state of the rest of the crystal. Thus $\mathcal{E}_{s,j}$ represents the energy of an electron in the state s when the rest of the crystal is in the state j. Then the probability of occupation of the state s is given by a Fermi factor

$$f(E_s) = 1/[1 + \exp(E_s - E_F)/kT] \tag{4.1}$$

where

$$E_s = -kT \ln \{\Sigma_j \exp(-\mathcal{E}_{s,j}/kT)\} \tag{4.2}$$

The summation is taken over all possible states of the rest of the crystal.

The energy E_s which occurs in the Fermi factor is thus related to a partial statistical sum, or partition function, in precisely the way that the Helmholtz free energy is related to the normal partition function. The quantity E_s may thus be described as the free energy of the crystal when one electron is held in a quantum state s. Such a definition, of course, implies that we may still identify the state s through all the possible states of the rest of the crystal. In Ge and Si this identification presents no difficulty, since the effects of the rest of the crystal may be treated as a relatively small perturbation.

A more careful analysis for a system at constant pressure shows that the quantity E_s actually has the properties of a partial Gibbs free energy. In this case E_s in Equation (4.2) must be replaced by

$$E_s' = E_s - V_s\left(\frac{\partial E_s}{\partial V_s}\right)_T \tag{4.3}$$

where V_s is the volume available to the electron in state s, essentially the volume of the crystal.

The second point concerns the statistics of localized levels. Equation (4.2) or (4.3) applies to states in the allowed energy bands of the crystal. If we take into account spin degeneracy, the Fermi factor must be multiplied by 2 in order to compute the total occupation probability of a

band state. Now, band states have the property that the corresponding wave functions are spread throughout the crystal. Thus there is practically no price, in terms of extra electrostatic interaction, for putting two electrons in the same state. This is the condition for the applicability of Fermi statistics in its simple form. In the case of localized states, however, a very different situation obtains. Even though an electron may be allowed two directions of spin in a localized state, once the state is occupied by an electron of either spin, it cannot then be occupied by an electron of opposite spin, because the electrostatic repulsion of the two localized charge distributions would raise the energy of the second electron. Nevertheless the statistical analysis must take into account the double degeneracy of the level. The statistics may be formulated quite generally, using the free energy concept as before. The Fermi factor for a discrete level is of the form

$$f(E_D) = 1/[1 + \exp (E_D - E_F)/kT] \tag{4.4}$$

where

$$E_D = -kT \ln \Big[\sum_l g_l \exp (-\varepsilon_l/kT) \Big] \tag{4.5}$$

where g_l is the degeneracy of the l^{th} level, and ε_l is its energy. For a single level of multiplicity g, the effect of (4.5) is to lower the effective energy of the level by an amount $-kT \ln g$, about $-0.7kT$ in the ordinary case of spin-degenerate levels. The corresponding Fermi factor is

$$f = 1/[1 + (1/g) \exp (\varepsilon_D - E_F)/kT] \tag{4.6}$$

where ε_D represents the position of the discrete level.

Before discussing applications of (4.6), we shall consider the population of electrons in the band levels. Near the band edge of a semiconductor, we have seen that the density of states is proportional to $(E - E_0)^{1/2}$, irrespective of the shape of the energy surfaces. Thus, for example, the number of electrons in the conduction band is given by

$$n_c = \int_0^\infty 1/[1 + \exp (E - E_F)/kT] C E^{1/2} dE \tag{4.7}$$

where E represents energy measured from the edge of the conduction band. If the number n_c is known, then Equation (4.7) may be regarded as an equation for the determination of E_F. For small densities, the solution may be written

$$\exp (-E_F/kT) = (\sqrt{\pi}/2)(1/n_c) C (kT)^{3/2} [1 - (1/2\sqrt{2}) \{(\sqrt{\pi}/2)(1/n_c) C (kT)^{3/2}\}^{-1}] \tag{4.8}$$

The usual approximation in semiconductor statistics corresponds to the second term in the bracket of (4.8) being negligible.

The numerical value of the constant C depends on the energy surfaces. For the multiple ellipsoid structure, as in the conduction bands of Ge and Si, we find

$$(\sqrt{\pi}/2)C = 2\nu(2\pi/h^2)^{3/2}(m_1m_2m_3)^{1/2} \tag{4.9}$$

where ν is the number of minima, and spin-degeneracy is taken into account. Equations (4.8) and (4.9) can also be put in the form

$$n_e = 2(2\pi m_{\text{eff}}kT/h^2)^{3/2} \exp(E_F/kT) \tag{4.10}$$

where

$$m_{\text{eff}} = (m_1m_2m_3)^{1/3}\nu^{2/3} \tag{4.11}$$

As an illustration for the Ge conduction band, assuming 4 minima, we have $m_{\text{eff}} = 0.550m_e$. The quantity m_{eff} is a "density-of-states effective mass," and must not be confused with the cyclotron effective mass.

TABLE IV. The Ratios A_n/A_0 and A_p/A_0 for Ge and Si

	Ge (4 minima)	Si (6 minima)
Electrons	0.412	1.129
Holes (1)	0.2075	0.390
Holes (2)	0.0084	0.068
Total holes	0.216	0.458
Geometric mean of holes and electrons	0.299	0.719

For warped energy surfaces, such as occur with degenerate band edges, the following treatment is appropriate. The energy surfaces may always be written in the form

$$\mathcal{E} = (P^2/2m_e)\alpha(\theta,\phi) \tag{4.12}$$

where $\alpha(\theta,\phi)$ is a coefficient depending only on the direction of the reduced momentum vector $\mathbf{P}$. It is readily shown that

$$(\sqrt{\pi}/2)C = 2(2\pi m_e/h^2)^{3/2} \sum_i (1/4\pi) \int [\alpha_i(\theta,\phi)]^{-3} d\Omega \tag{4.13}$$

the integral being over the complete solid angle in momentum space, and the summation over the different degenerate surfaces. Numerical results for Ge and Si are summarized in Table IV. The figures in the table are in units of A_0, where

$$A_0 = 2(2\pi m_e/h^2)^{3/2}(kT)^{3/2} = 2.436 \times 10^{19} \text{ at } 294^\circ\text{K} \tag{4.14}$$

The table entry thus represents the ratio $\nu(m_1m_2m_3/m_e^3)^{1/2}$ or A_n/A_0.

Quite generally we have the relations

$$\begin{aligned} n &= A_n \exp (E_F - E_c)/kT \\ p &= A_p \exp (E_v - E_F)kT \end{aligned} \tag{4.15}$$

where n and p are, respectively, the number of electrons in the conduction band and the number of holes in the valence band, and A_n and A_p are obtainable from Table IV. For intrinsic material, we have

$$n = p = n_i = (A_n A_p)^{1/2} \exp [-(E_c - E_v)/2kT] \tag{4.16}$$

where E_c and E_v are the conduction and valence band energies, or, more strictly, free energies. With the aid of (4.16) and the measured intrinsic resistivity at 294°K, and also a knowledge of electron and hole drift mobilities at these temperatures (*60*, *61*), we obtain for Ge

$$E_G = E_c - E_v = 0.656 \text{ ev} \tag{4.17}$$

A similar calculation for Si gives $E_G = 1.089$ ev.

It must be emphasized that the energy gaps obtained in this way are free energy gaps. It is also possible to obtain an enthalpy of activation from the slope of a graph of ln ρ *vs.* $1/T$. Such graphs usually give good straight lines, indicating that the enthalpy of activation is nearly a constant. The gaps obtained from them are 0.785 ev for Ge and 1.21 ev for Si. The discrepancy between these figures and the free-energy gaps permits an estimate of the temperature coefficient of the gap. If the gap width is assumed to vary linearly with temperature, the variation necessary to explain the discrepancy is 4.4×10^{-4}/°C for Ge and 4.1×10^{-4}/°C for Si.

Some, but not all, of this temperature variation can be accounted for merely by the variation of lattice constant with temperature. The pressure variation of resistivity has been measured for Ge by Paul and Brooks (*62*) and by others (*63*). The equivalent temperature coefficient may be found from the relation

$$(\partial E_G/\partial T)_p = (\partial V/\partial T)_p (\partial p/\partial V)_T (\partial E_G/\partial p)_T \tag{4.18}$$

and is -0.78×10^{-4} ev/°C for Ge, much less than the observed value. The difference must be ascribed to an explicit temperature dependence of the energy gap, arising from the interaction between electrons and lattice vibrations (*64*). For Si, the pressure coefficient of the gap appears to be opposite in sign (*65*), and would predict a temperature coefficient of $+0.26 \times 10^{-4}$ ev/°C. Thus in the case of Si, we must assume that the explicit temperature dependence of the gap is by far the dominant effect.

It is of some interest to compare results deduced from resistivity measurements on intrinsic material with those obtained from optical

data. Unfortunately, interpretation of optical data is not straightforward. This is because of uncertainty of how optical absorption data should be extrapolated to zero absorption coefficient in order to obtain a true gap width. Furthermore, optical data do not measure a true energy difference. The reason for this is that during an optical transition the lattice remains frozen, as it were, in the same configuration it had in the initial state, which is not in general a configuration of thermal equilibrium for the final state. The situation is further complicated by the fact that the transition between band edge points in **k**-space is forbidden, and can only take place with simultaneous emission of absorption of a lattice vibration or phonon. Nevertheless, Bardeen and coworkers have shown (*66*) that the "forbidden" transition is sufficiently strong in Ge and Si to be observed under the usual experimental conditions, so that the threshold frequency for optical absorption gives a reasonable picture of the gap width. The "optical gap widths" are 0.62 ev for Ge and 1.05 ev for Si, in each case about 0.04 ev smaller than the corresponding free energy gaps (*67*). The values for the shift with temperature of the optically determined energy gap are -4.4×10^{-4} ev/°C for Ge, and -4.5×10^{-4} ev/°C for Si, in rather good agreement with the values deduced from free energy considerations.*

Let us now consider the theory of the energy gap in greater detail. The energy levels of the system may be written in the form

$$\mathcal{E}_{s,j} = \mathcal{E}_s + \mathcal{E}_j + \delta_{s,j} \tag{4.19}$$

where s stands for electronic quantum numbers and j for all the lattice quantum numbers. The energy $\delta_{s,j}$ represents the energy of interaction of an electron in state s with a lattice in state j. If this interaction can be treated as a small perturbation, then we may write

$$\delta_{s,j} = \Sigma_{\mathbf{q}} b_s(\mathbf{q})(n_{\mathbf{q}} + \tfrac{1}{2}) \tag{4.20}$$

where $\mathbf{q}$ is the wave-number vector which identifies a particular lattice vibration, $b_s(\mathbf{q})$ is an interaction coefficient, and $n_{\mathbf{q}}$ is the number of quanta of oscillator $\mathbf{q}$ which are excited.

If we substitute (4.20) into the expression for E_s we find

$$E_s = \mathcal{E}_s - kT\Sigma_{\mathbf{q}} \ln \{\tfrac{1}{2} \sinh [(h\nu_{\mathbf{q}} + b_s(\mathbf{q}))/2kT]\} \tag{4.21}$$

For the free energy gap, i.e. the difference in free energy between two states, labelled 1 and 2, Equation (4.21) leads to

$$E_1 - E_2 = \mathcal{E}_1 - \mathcal{E}_2 - kT \sum_{\mathbf{q}} \ln \left[\frac{\sinh (h\nu_{\mathbf{q}} + b_2(\mathbf{q}))/2kT}{\sinh (h\nu_{\mathbf{q}} + b_1(\mathbf{q}))/2kT}\right] \tag{4.22}$$

* That is, using Equation (4.16) and activation energies from reference *61*.

At room temperature, for the lattice frequencies for which the b's are appreciable, it suffices to replace the hyperbolic functions by their arguments, leading to the approximate relation

$$E_G = E_1 - E_2 \cong \varepsilon_1 - \varepsilon_2 - kT \sum_{\mathbf{q}} \ln \left[\frac{h\nu_{\mathbf{q}} + b_2(\mathbf{q})}{h\nu_{\mathbf{q}} + b_1(\mathbf{q})} \right] \tag{4.23}$$

Assuming that the b's are obtained from a perturbation calculation, it is readily found that all the b_1's are negative, corresponding to a downward displacement of the edge of the conduction band, and all the b_2's are positive, corresponding to an upward displacement of the edge of the valence band. Hence all the logarithmic terms in (4.23) are positive, and the free energy gap E_G varies linearly with the temperature with a negative coefficient, in agreement with observation.

The mean change in energy in an optical transition near the band edge may be written in the general form

$$\Delta\varepsilon = \frac{\Sigma_j(\varepsilon_{1,j} - \varepsilon_{2,j}) \exp(-\varepsilon_{2,j}/kT)}{\Sigma_j \exp(-\varepsilon_{2,j}/kT)} \tag{4.24}$$

where the statistical weighting factors are taken as those appropriate to the lower state for both states involved in the transition. This reflects the fact that the lattice remains "frozen" during the transition. Using (4.20) in (4.24), we have

$$\Delta\varepsilon = \varepsilon_1 - \varepsilon_2 - kT\Sigma_{\mathbf{q}}[b_2(\mathbf{q}) - b_1(\mathbf{q})]/[h\nu_{\mathbf{q}} + b_2(\mathbf{q})] \tag{4.25}$$

The temperature coefficient in (4.25) agrees with that in (4.24) to the first order in the ratio $b(\mathbf{q})/h\nu_{\mathbf{q}}$, a result which has already been mentioned by James (*59*). It is well to emphasize that had $\Delta\varepsilon$ been computed by averaging each $\varepsilon_{s,j}$ with the statistical weight appropriate to the particular electronic state, rather than the final one, then $\Delta\varepsilon$ would have been independent of temperature in the high temperature approximation, a result which could be obtained directly from thermodynamic relations between energy and free energy.*

As to actual magnitudes, the expressions (4.23) and (4.25) are identical for all practical purposes. The part of the temperature coefficient which is due to explicit temperature dependence of the free energy gap is 3.6×10^{-4}/°C for Ge, and 4.4×10^{-4}/°C for Si. The theory of the b's has been given by Fan (*64*). It is subject to modification due to the com-

* This discussion is, strictly speaking, valid only when the form of the normal modes does not depend on the electronic state. This is probably a good assumption for valence semiconductors but not for ionic crystals. The author is indebted to M. Lax for calling his attention to this restriction.

plex nature of the energy surfaces. This modification will be discussed in Section VI.

There is nothing in the theory presented so far which would explain the difference between the thermal and optical gaps. However, it must be remarked that the optical gap is not accurately given by (4.24). This arises primarily because the optical transition also involves a phonon (*66*) which can supply part of the energy for the transition. Thus the optical gap should be smaller than the thermal gap by just the energy of a phonon necessary to scatter an electron from the point $\mathbf{P} = 0$ in the conduction band to a minimum point on the (111)-axis. The energy difference amounts to about 0.035 ev, and this has been used by Marfarlane and Roberts (*109*) to deduce that the minima in Ge lie about ⅔ of the way out to the BZ boundary along the (111)-axis. This in turn would suggest an 8-, rather than a 4-minimum model.

Warschauer and Paul (*69*) have also made measurements of the pressure coefficient of the optical gap. For absorption coefficients between 10 and 70 cm^{-1}, the displacement averages about 8.5×10^{-12} ev/dyne/cm^2, as compared with 5.5×10^{-12} ev/dyne/cm^2 deduced by Paul and Brooks (*62*) for the thermal gap from measurements of resistivity as a function of pressure in the intrinsic range. The discrepancy between the optical and thermal data can probably be explained in terms of a change with pressure in the shape of the absorption *vs.* wavelength curve, such that extrapolation of the optical data to zero absorption coefficient would produce a smaller optical gap. As yet, however, the experiments only suggest, but do not quantitatively establish, this explanation.

Returning now to the question of impurity levels and their statistics, we consider the case of a donor level of degeneracy g below a conduction band by an energy difference ε_D. It is necessary to consider also the possibility that acceptor levels are present which, being way below the donor level, will be completely filled with electrons captured from the donors. We then obtain the formula of DeBoer and van Geel (*70*)

$$\frac{n(N_A + n)}{N_D - N_A - n} = \frac{A_n}{g} \exp\{-(\varepsilon_c - \varepsilon_D)/kT\} \tag{4.26}$$

where A_n is the quantity obtained in connection with Table IV. The data shown in Fig. 6 were obtained by Debye and Conwell (*26*) by choosing the constant A_n to give the best fit to low-temperature Hall effect data, on a variety of samples of different purity. They used Equation (4.26) in a form which amounted to assuming $g = 2$. However, they found the best fit to all data was obtained by taking $A_n = 0.125\, A_0$. Since this is only ⅓ the value of A_n computed from cyclotron data, one would be tempted to believe that g should be increased to 6, which seems unlikely,

or that there is a temperature dependence of $E_c - E_D$, which seems equally unlikely. The latter hypothesis would require a donor ionization energy which *increased* with temperature at a rate of 2.6×10^{-4} ev/°C. This is highly improbable since it would imply a donor ionization energy of the order of 0.08 ev at room temperature. It seems most likely that the criterion of optimum fit is too insensitive actually to determine the density-of-states mass in this manner.

When we have to deal with holes, a relation analagous to (4.26) holds. In this case, however, the usual situation is that there are several types of acceptor level to which an electron may be added, corresponding, for example, to different possible spin orientations of the missing electron. As a result, the symmetry of Equation (4.26) with respect to holes and electrons is preserved. If we take as the argument of the exponential on the right of (4.26), or its hole analogue, the *free* energy change resulting from the addition of an electron from the conduction band or the removal of an electron to the valence band, then we obtain the correct result. Since the state of the impurity center with four electrons per atom is usually the state of minimum degeneracy, it is usually most convenient to take this state as the reference for computing free energy changes. More generally, we can take into account possible excited levels of the impurity center by replacing the donor energy level in accordance with Equation (4.5).

Equation (4.26) exhibits three distinct regions. In the first, n is comparable to $N_D - N_A$ and is essentially independent of temperature. This is the case normally exhibited by Ge and Si at room temperature, and down to at least liquid nitrogen temperature. In the next region, n is much less than $N_D - N_A$, but larger than N_A, so that an approximate solution of (4.26) is

$$n = (N_D - N_A)^{\frac{1}{2}}(A_n/g)^{\frac{1}{2}} \exp\{-(\mathcal{E}_c - \mathcal{E}_D)/2kT\} \qquad (4.27)$$

In this region the ionization energy of the donor is actually twice the observed activation energy for conduction. In the third region, n is less than both N_A and $N_D - N_A$, and the activation and ionization energies become equal. This will always happen at sufficiently low temperature. However, because of the smallness of the activation energies, factors of $T^{\frac{3}{4}}$ and $T^{\frac{3}{2}}$ which occur on the right side of (4.27) and its counterpart tend to conceal the change in apparent activation energy. Mere graphs of $\ln \rho$ *vs.* $1/T$ are not reliable for the determination of activation energies in such cases.

Another case of interest is that which occurs when there exist several levels for a given center, the most thoroughly studied example being that of Au in Ge. In this case we will assume that the deeper level corre-

sponds to the first electron, which may go in with only one direction of spin, whereas the upper level corresponds to the addition of a second electron, which may go in with either direction of spin. If E_{T_1} and E_{T_2}

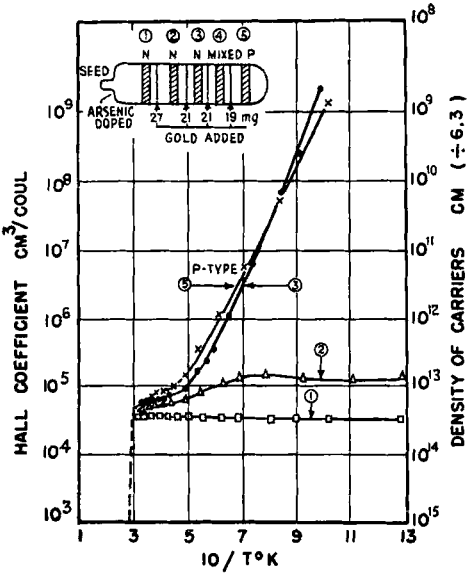

FIG. 9. Variation of Hall effect and carrier concentration with temperature for gold-doped germanium. From W. C. Dunlap, Jr., *Phys. Rev.* **97**, 614 (1955).

represent energies of the first and second trapped electrons, respectively, then the Fermi factor for an Au level is given by (*71*)

$$f = \frac{2}{1 + [1 + 4 \exp (E_{T_1} - E_F)/kT]/[1 + 4 \exp (E_F - E_{T_2})/kT]} \tag{4.28}$$

If the Fermi level is well above T_1 and above T_2, then we find

$$\frac{n(n + 2N_T - N_D)}{N_D - n} = \frac{A_n}{4} \exp \{-(E_C - E_{T_2})/kT\} \tag{4.29}$$

where N_T is the number of trapping atoms, and E_C the edge of the conduction band, while N_D is the number of conventional donors. Equation (4.29) has a solution if $2N_T > N_D$, and in general the activation energy will be equal to $E_C - E_{T_2}$. If the Fermi level is below both T_1 and T_2, then we obtain

$$\frac{p(N_D + p)}{N_T - N_D - p} = \frac{A_p}{2} \exp \{-(E_v - E_{T_1})/kT\} \tag{4.30}$$

where p is the number of holes in the valence band. More generally in Equations (4.29) and (4.30), N_D should be replaced by the excess of normal donors over normal acceptors. This equation has a solution if $N_T > N_D$, and in general the activation energy is equal to the energy $E_{T_1} - E_v$, where E_v is the energy of the edge of the valence band. Both these types of behavior have been reported by Dunlap (*31*). Typical behavior is illustrated in Fig. 9, which shows results for both p- and n-type gold doping. The type of situation described above has been recently formulated in more general terms by Landsberg (*72*).

So far we have excluded from consideration the case in which classical statistics is not a good approximation. Many cases occur in which the electron concentration in the conduction band or hole concentration in the valence band becomes high enough to bring about a transition to Fermi statistics, the so-called degenerate case. This situation has been considered extensively, especially by Lark-Horovitz and co-workers (*56*).

V. Theory of Transport Properties of Semiconductors

In the present section we shall present an elementary derivation of formulas for electrical resistance, Hall effect, and magnetoresistance effects. A more rigorous derivation would involve the use of the Boltzmann equation, which we shall omit here. The present derivation has the advantage of keeping the physics in evidence at each step, although the chain of reasoning is not complete.

We shall consider first the case of a single type of charge carrier with spherical energy surfaces. The electric field is taken in the x-direction, the magnetic field in the z-direction, and the Hall voltage or field is measured in the y-direction. The equations of motion may be written

$$\begin{aligned} dv_x/dt &= (e/m)E_x + \omega_0 v_y \\ dv_y/dt &= (e/m)E_y - \omega_0 v_x \\ \omega_0 &= eH/mc \end{aligned} \tag{5.1}$$

For convenience we have taken the carrier as a hole, and m stands for the effective mass. The solution of (5.1) may be written in the complex form

$$v_x + iv_y = (v_x + iv_y)_0 e^{-i\omega_0 t} + (e/m)(E_x + iE_y)(1 - \exp[-i\omega_0 t])/i\omega_0 \tag{5.2}$$

where $(v_x + iv_y)_0$ is the initial condition. We assume that each collision completely randomizes the direction of motion of the carrier. For conduction problems we are only interested in *acquired* motion, so hereafter it will be legitimate to omit initial conditions. If we focus attention on any particular carrier, the probability that it has survived collision for a time t is exp $(-t/\tau)$, where τ is the collision time and in this example will

be assumed to be a function only of the energy of the carrier. The mean acquired velocity is then obtained simply by averaging (5.2) with respect to the normalized probability $(1/\tau)\exp(-t/\tau)$. The mean contribution to the electric current is then obtained by multiplying the mean acquired velocity by the charge e, and this must be averaged over all the electrons in the distribution in order to obtain the total current according to the relation $j = ne\langle \mathbf{v} \rangle_{\text{AV}}$. The result is

$$j_x + ij_y = (ne^2/m)(E_x + iE_y)\langle \tau/(1 + i\omega_0\tau)\rangle_{\text{AV}} \tag{5.3}$$

where n is the total number of carriers. The average is taken with the distribution function

$$f(\mathcal{E})\, d\mathcal{E} = \frac{4}{3\sqrt{\pi}} x^{3/2} \exp(-x)\, dx$$

where

$$x = \mathcal{E}/kT \tag{5.4}$$

This is similar to a Maxwell distribution, but contains an extra weighting factor of the energy. Hereafter, unless otherwise specified, all averages used in this section will be with respect to the distribution (5.4).

The result (5.3) must be rearranged by separating into real and imaginary parts.

$$j_x = \frac{ne^2}{m}\left\{E_x\langle\tau\rangle_{\text{AV}} + E_y\omega_0\left\langle\frac{\tau^2}{1+\omega_0^2\tau^2}\right\rangle_{\text{AV}} - E_x\omega_0^2\left\langle\frac{\tau^3}{1+\omega_0^2\tau^2}\right\rangle_{\text{AV}}\right\} \tag{5.5}$$

$$j_y = \frac{ne^2}{m}\left\{E_y\left\langle\frac{\tau}{1+\omega_0^2\tau^2}\right\rangle_{\text{AV}} - E_x\omega_0\left\langle\frac{\tau^2}{1+\omega_0^2\tau^2}\right\rangle_{\text{AV}}\right\}$$

Measurements are usually made in such a way that the Hall probes, measuring E_y, are effectively open circuit; hence in (5.5) we must set $j_y = 0$. When this is done we finally obtain the relations

$$\frac{E_y}{E_x} = \omega_0 \frac{\langle\tau^2/(1+\omega_0^2\tau^2)\rangle_{\text{AV}}}{\langle\tau/(1+\omega_0^2\tau^2)\rangle_{\text{AV}}} \tag{5.6}$$

$$j_x = \frac{ne^2}{m} E_x\left\{\langle\tau\rangle_{\text{AV}} + \omega_0^2\left[\frac{\langle\tau^2/(1+\omega_0^2\tau^2)\rangle_{\text{AV}}^2}{\langle\tau/(1+\omega_0^2\tau^2)\rangle_{\text{AV}}} - \langle\tau^3/(1+\omega_0^2\tau^2)\rangle_{\text{AV}}\right]\right\} \tag{5.7}$$

If τ is independent of energy, the averages may be dropped and we have the simple results

$$j_x = \frac{ne^2\tau}{m} E_x$$

$$E_y = (1/nec) j_x H \tag{5.8}$$

Thus the electrical resistance is independent of magnetic field, whereas the Hall constant (which is the coefficient of $j_x H$ in the second Equation (5.8)) directly measures the density of carriers, n.

In general, of course, τ does depend on energy, and we cannot make the above simplification. However, it is often a good approximation to assume that $\omega_0\tau$ is small, so that we can expand the denominators in (5.6) and (5.7). The result is

$$E_x = j_x\rho_0 \left\{1 + \omega_0^2 \frac{\langle\tau^3\rangle_{AV}\langle\tau\rangle_{AV} - \langle\tau^2\rangle_{AV}^2}{\langle\tau\rangle_{AV}^2}\right\} \tag{5.9}$$

$$E_y = \frac{1}{nec}\frac{\langle\tau^2\rangle_{AV}}{\langle\tau\rangle_{AV}^2} j_x H$$

$$\sigma_0 = 1/\rho_0 = (ne^2/m)\langle\tau\rangle_{AV}$$

With the aid of the Schwarz inequality, it may be shown that the correction term always represents an increase of resistance with magnetic field. The so-called magnetoresistance is given by

$$\Delta\rho/\rho_0 H^2 = (e^2/m^2c^2) \frac{\langle\tau^3\rangle_{AV}\langle\tau\rangle_{AV} - \langle\tau^2\rangle_{AV}^2}{\langle\tau\rangle_{AV}^2} \tag{5.10}$$

It is also useful to define an auxiliary coefficient

$$\Delta\rho/\rho_0 R^2\sigma_0^2 H^2 = \langle\tau^3\rangle_{AV}\langle\tau\rangle_{AV}/\langle\tau^2\rangle_{AV}^2 - 1 \tag{5.11}$$

The theory of τ will be discussed in Section VI. Here we shall give only the results necessary for the purposes of calculating the averages in Equations (5.6)–(5.11). There are two principal sources of scattering in Ge and Si:

1. Scattering by lattice vibrations. This is the dominant form of scattering in pure materials and at high temperatures. For this type of scattering, the collision time τ decreases uniformly with increasing energy, inversely as the square root of the energy for the simplest theory.

2. Ionized impurity scattering. This is the dominant form of scattering in impure semiconductors and at low temperatures. For this type of scattering, the collision time varies as a positive power of the energy, the $+\frac{3}{2}$ power in the simplest theories. The coefficient of this power is actually a slowly varying function of energy, but this variation can be neglected for the purposes of the present analysis.

The two types of scattering will occur simultaneously at some energies, and they combine according to the law

$$1/\tau(\mathcal{E}) = 1/\tau_L(\mathcal{E}) + 1/\tau_I(\mathcal{E}) \tag{5.12}$$

where $\tau_L(\mathcal{E})$ is the mean free time for lattice scattering, and $\tau_I(\mathcal{E})$ that for impurity scattering, each considered alone.

Under some circumstances neutral impurity scattering also occurs for which $\tau(\mathcal{E})$ is independent of energy.

It must be emphasized that the existence of a τ is itself an approximation. For spherical energy surfaces centered at $\mathbf{P} = 0$, it always exists, provided energy loss in collisions can be neglected, but for nonspherical surfaces and surfaces centered at other than $\mathbf{P} = 0$, there is no case in which the assumption of a τ is entirely justified. These points will be discussed in Section VI.

In many cases it is convenient to assume that τ varies according to a power law of the form

$$\tau = \tau_0/\mathcal{E}^p \tag{5.13}$$

For example, for the simplest theory $p = \frac{1}{2}$ for scattering by lattice vibrations, and $p = -\frac{3}{2}$ for scattering by ionized impurities. When this scattering law is substituted into Equation (5.9), we obtain the following relations

$$\sigma_0 = (ne^2/m)\tau_0(kT)^{-p}\Gamma(\tfrac{5}{2} - p)/\Gamma(\tfrac{5}{2}) \tag{5.14}$$

$$R = (1/nec)\{\Gamma(\tfrac{5}{2} - 2p)\Gamma(\tfrac{5}{2})/[\Gamma(\tfrac{5}{2} - p)]^2\} \tag{5.15}$$

$$\xi = \Delta\rho/\rho_0 R^2\sigma_0^2 H^2 = \{\Gamma(\tfrac{5}{2} - 3p)\Gamma(\tfrac{5}{2} - p)/[\Gamma(\tfrac{5}{2} - 2p)]^2\} - 1 \tag{5.16}$$

where $\Gamma(x)$ is the gamma function (*73*). The combination of gamma functions which occurs in (5.15) will be called $\lambda(p)$, whereas that in (5.16) will be called $\eta(p)$. These functions will be found to be useful even for nonspherical energy surfaces. However, a difficulty occurs because $\lambda(p)$ blows up as p approaches $\frac{5}{4}$, and $\eta(p)$ blows up as p approaches $\frac{5}{6}$. This is an important problem because, as we shall see, experiments indicate that the lattice collision time varies as $\mathcal{E}^{-0.64}$ in n-type Ge, and as $\mathcal{E}^{-1.66}$ in p-type Ge. The divergence of the integrals for $\lambda(p)$ and $\eta(p)$ is, of course, merely the result of the approximation used in arriving at (5.13)–(5.15). This approximation can actually break down in three ways.

1. The lattice scattering cannot be approximated by a simple power law down to the lowest energies, but always approaches an $\mathcal{E}^{-\frac{1}{2}}$ behavior for sufficiently slow electrons (see Equations (6.15) and (6.16) in the next section).

2. The scattering law (5.13) is not valid down to zero energy, but is cut off at a certain energy by the onset of ionized impurity scattering, which varies as a positive power of the energy. Hence the integrals involved in $\langle\tau^r\rangle_{\text{AV}}$ do not actually diverge as $\mathcal{E} \to 0$.

3. The terms in $\omega_0^2\tau^2$ in the denominators of the averages in (5.6) and (5.7) become important and cannot be neglected for sufficiently small $\mathcal{E}$. They can also be seen to prevent the divergence of the integrals.

Which case occurs depends on whether the term in $\omega_0^2\tau^2$ is smaller or larger than unity at the energy for which τ reaches its maximum when impurity scattering is taken into account. If it is smaller, then the values of the three coefficients remain independent of magnetic field, whereas if larger, the coefficients show field dependence down to quite low fields.

The exact evaluation of the τ-averages would require lengthy numerical integration using the composite τ of Equation (5.12). The only case which can be integrated in terms of tabulated functions is the case of no impurity scattering with $\tau_L(\mathcal{E}) \propto \mathcal{E}^{-1/2}$. However, H. Jones (*74*) and Johnson and Lark-Horovitz (*75*) have evaluated the integral for the τ-averages corresponding to $\lambda(p)$ for the case $p = 1/2$ and variable amounts of impurity scattering. If lattice scattering predominates heavily, then the averages may be approximated fairly well with incomplete gamma functions, or by simple power series in the value of the variable $x = x_0$ for which $\tau(\mathcal{E})$ has its maximum. This variable is given by

$$x_0 = [(q/p)\Gamma(5/2 + q)\Gamma(5/2 - p)]^{1/(p+q)}[\mu_L^{(T)}/\mu_I^{(T)}]^{1/(p+q)} \quad (5.17)$$

where p and q are the energy exponents of lattice and impurity scattering, respectively, and $\mu_L^{(T)}$, $\mu_I^{(T)}$ are actual mobilities which would exist due to lattice and impurity scattering at the temperature T at which x_0 is evaluated.

As an illustration of the sensitivity of the results even to small amounts of impurity scattering, we show, in the table below values of the functions $\lambda(p)$ and $\eta(p)$ for various ratios $\mu_L^{(T)}/\mu_I^{(T)}$ and for the case $p = 0.64$, $q = 1.5$

Ratio $R = \mu_L^{(T)}\mu_I^{(T)}$	$\lambda(p)$	$\eta(p)$
0.02	1.21	0.95
0.01	1.22	1.19
0.002	1.27	1.25
0.001	1.27	1.34
10^{-4}	1.32	1.45
10^{-5}	1.34	1.56
10^{-6}	1.34	1.63
0	1.35	1.74

For comparison, it is interesting to note that when we have lattice scattering alone with $p = 1/2$, we obtain $\lambda(p) = 3\pi/8 = 1.1781$,

$$\eta(p) = 4/\pi = 1.274$$

while for ionized impurity scattering alone, with $q = 3/2$, we obtain $\lambda(p) = 315\pi/512 = 1.931$ and $\eta(p) = 32768/6615\pi = 1.572$.

Curves for the composition of lattice and impurity scattering in the calculation of $\langle\tau\rangle_{\text{AV}}$ have been given by Johnson and Lark-Horovitz (*75*) and by Conwell (*60*).

Two further generalizations of the above results are possible and must be carried out before the formulas can be applied to Ge and Si. The two effects which must be considered are the influence of anisotropic effective mass, and the simultaneous presence of electrons or holes originating in different energy bands. The fundamental idea to be applied in each case is that the different electrons in the distribution are effectively in parallel, so that their contributions to the conductivity are additive. This is in contrast to the different scattering mechanisms, which are effectively in series.

Equation (5.5) may thus be generalized immediately, simply by summing terms similar to those in brackets for all the bands involved, that is

$$j_x = \sum_i \frac{n_i e^2}{m_i} \left\{ E_x \langle\tau_i\rangle_{\text{AV}} + E_y \omega_i \left\langle \frac{\tau_i^2}{1+\omega_i^2\tau_i^2} \right\rangle_{\text{AV}} - E_x \omega_i^2 \left\langle \frac{\tau_i^3}{1+\omega_i^2\tau_i^2} \right\rangle_{\text{AV}} \right\} \tag{5.18}$$

$$j_y = \sum_i \frac{n_i e^2}{m_i} \left\{ E_y \left\langle \frac{\tau_i}{1+\omega_i^2\tau_i^2} \right\rangle_{\text{AV}} - E_x \omega_i \left\langle \frac{\tau_i^2}{1+\omega_i^2\tau_i^2} \right\rangle_{\text{AV}} \right\}$$

where it must be remembered that the cyclotron frequencies ω_i have a sign which will be taken as positive for holes and negative for electrons. As a typical illustration of the application of (5.18), we may consider the case of three bands, each with spherical energy surfaces, two being hole bands and one an electron band. We will also write the equations for the case of small magnetic field, where the denominators in (5.18) may be taken as unity. Then the Hall constant is given by

$$R_0 = \frac{-n_1\langle\mu_1^2\rangle_{\text{AV}} + n_2\langle\mu_2^2\rangle_{\text{AV}} + n_3\langle\mu_3^2\rangle_{\text{AV}}}{[n_1\langle\mu_1\rangle_{\text{AV}} + n_2\langle\mu_2\rangle_{\text{AV}} + n_3\langle\mu_3\rangle_{\text{AV}}]^2} \tag{5.19}$$

and the magnetoresistance coefficient ξ is given by

$$\xi = \frac{[n_1\langle\mu_1\rangle_{\text{AV}} + n_2\langle\mu_2\rangle_{\text{AV}} + n_3\langle\mu_3\rangle_{\text{AV}}][n_1\langle\mu_1^3\rangle_{\text{AV}} + n_2\langle\mu_2^3\rangle_{\text{AV}} + n_3\langle\mu_3^3\rangle_{\text{AV}}]}{[-n_1\langle\mu_1^2\rangle_{\text{AV}} + n_2\langle\mu_2^2\rangle_{\text{AV}} + n_3\langle\mu_3^2\rangle_{\text{AV}}]^2} - 1 \tag{5.20}$$

where n_1, n_2, and n_3 are the numbers of electrons and of each type of hole, respectively; $\mu_1 = e(\tau_1/m_1)$, etc. are the carrier mobilities as a function of energy for each band, and are averaged over the appropriate distribution of the form (5.4) for each band. The well-known formula for

the Hall effect in a semiconductor with minority carriers present is a special case of (5.19). The general theory has been developed for the multiband model by Johnson and Whitesell (*76*), and its application to Ge has been discussed by Willardson, Harman, and Beer (*77*). Both groups of authors have extended the analysis to high magnetic fields, but have confined their attention to an energy dependence of τ corresponding to $p = \frac{1}{2}$ or $p = -\frac{3}{2}$.

The limiting case of very large magnetic fields is especially simple, although it cannot be achieved in practice except in some cases at liquid nitrogen temperature or lower. In this case, for the 3-band situation used as an illustration above, the Hall constant directly measures the total charge density, independent of any mobility averages, i.e.,

$$R_\infty = (1/ec)[1/(-n_1 + n_2 + n_3)] \tag{5.21}$$

The effect of magnetic field on the resistance is given by

$$\rho_\infty/\rho_0 = \frac{[n_1\langle 1/\mu_1\rangle_{AV} + n_2\langle 1/\mu_2\rangle_{AV} + n_3\langle 1/\mu_3\rangle_{AV}][n_1\langle \mu_1\rangle_{AV} + n_2\langle \mu_2\rangle_{AV} + n_3\langle \mu_3\rangle_{AV}]}{[-n_1 + n_2 + n_3]^2} \tag{5.22}$$

where ρ_∞/ρ_0 is the ratio of the resistance at effectively infinite magnetic field to that at zero field. It is of interest that the resistance at high fields is independent of the field, and depends on the sum of the reciprocals of the mobilities. This means that in the high-field case, the different scattering mechanisms, as well as the different electrons, contribute additively to the resistance, and thus the treatment of simultaneous impurity and lattice scattering is greatly simplified. The same simplification results if measurements are conducted at sufficiently high frequency, so that $\omega\tau \gg 1$. In the absence of a magnetic field, in fact, the expression for the current at high frequency becomes

$$j_x = E_x \sum_i \frac{n_i e^3}{m_i^2 \omega^2} \left\langle \frac{1}{\mu_i} \right\rangle_{AV} \tag{5.23}$$

The theory for ellipsoidal energy surfaces has been given by Meiboom and Abeles (*78*). Here we shall present only a very oversimplified derivation along the lines of Equations (5.2) and (5.5). We consider the equations of motion of a carrier only for the very special case of the electric fields and the magnetic field along the three principal axes of one ellipsoid—the electric fields in the x- and y-directions, and the magnetic field

in the z-direction. The equations of motion are then

$$\begin{aligned} dv_x/dt &= (e/m_1)E_x + (e/m_1c)v_yH \\ dv_y/dt &= (e/m_2)E_y - (e/m_2c)v_xH \end{aligned} \tag{5.24}$$

where m_1 and m_2 are the masses for the x- and y-directions, respectively. Solving these equations and carrying out appropriate averages, as in the spherical case, we obtain finally

$$e\langle v_x\rangle_{\text{AV}} = (e^2E_x/m_1)\langle\tau\rangle_{\text{AV}} - (e^4E_xH^2/m_1{}^2m_2)\langle\tau^3\rangle_{\text{AV}} + (e^3E_yH/m_1m_2)\langle\tau^2\rangle_{\text{AV}} \tag{5.25}$$

$$e\langle v_y\rangle_{\text{AV}} = -(e^3HE_x/m_1m_2)\langle\tau^2\rangle_{\text{AV}} + (e^2E_y/m_2)\langle\tau\rangle_{\text{AV}}$$

Here we have made the approximation that the magnetic field is small, which results in the criterion

$$(eH\tau/\sqrt{m_1m_2}\,c) \ll 1 \tag{5.26}$$

We must sum the expressions on the right of (5.25) over the equivalent ellipsoids. For simplicity, we shall take the ellipsoids along the cubic axes, so that the different ellipsoids correspond merely to permuting the masses in (5.25) and (5.26). When this is carried out, we obtain the following expression for the Hall coefficient

$$R_0 = \frac{1}{nec}\frac{\langle\tau^2\rangle_{\text{AV}}}{\langle\tau\rangle_{\text{AV}}{}^2}\frac{m_1 + m_2 + m_3}{[(1/m_1) + (1/m_2) + (1/m_3)]^2}\frac{3}{m_1m_2m_3} \tag{5.27}$$

It can be shown that this expression is completely general and is valid for any arrangement of ellipsoids which has over-all cubic symmetry. Equation (5.27) can also be put in the form

$$R_0 = (1/nec)\lambda(p)[3K(K+2)/(2K+1)^2] \tag{5.28}$$

$$K = \frac{m_1}{m_2}, \quad m_2 = m_3$$

The theory for the magnetoresistance is somewhat more complicated, because for a cubic crystal there are actually three independent magnetoresistance coefficients at low magnetic fields (*85*). The simplified theory may be used to obtain two of them for the case of ellipsoids along the cubic axes, which is the case of the silicon conduction band. For the transverse case we obtain

$$\begin{aligned} \xi_{100}^{010} &= [\langle\tau^3\rangle_{\text{AV}}\langle\tau\rangle_{\text{AV}}/\langle\tau^2\rangle_{\text{AV}}{}^2][(K^2+K+1)(2K+1)/K(K+2)^2] - 1 \\ &= \eta(p)[(K^2+K+1)(2K+1)/K(K+2)^2] - 1 \end{aligned} \tag{5.29}$$

by means of (5.25). For the longitudinal magnetoresistance, corresponding to a magnetic field parallel to the electric current, the coefficient vanishes in this particular orientation, as it does in all orientations for

spherical energy surfaces. The reason for this is that when the electron is accelerated along the electric field, it moves parallel to the magnetic field, and its motion is therefore unaffected. This is not generally true in the anisotropic case simply because, for an electric field not along a principal axis, there is a component of acceleration perpendicular to the magnetic field, and the motion is therefore modified. Thus when the ellipsoids are in (111)-directions, as in Ge, there is no direction of electric field for which the longitudinal magnetoresistance vanishes. In this case the magnetoresistance coefficient, defined in analogy with (5.16), is given by

$$\xi_{100}^{100} = \eta(p)\tfrac{2}{3}[(K-1)^2(2K+1)/K(K+2)^2] \tag{5.30}$$

The corresponding transverse magnetoresistance coefficient for (111)-ellipsoids is given by

$$\xi_{100}^{010} = \eta(p)[(2K+1)^2/3K(K+2)] - 1 \tag{5.31}$$

with the symmetry relations

$$\begin{aligned} \xi_{110}^{110} &= (\tfrac{1}{2})\xi_{100}^{100} \\ \xi_{110}^{1\bar{1}0} &= (\tfrac{1}{2})\xi_{100}^{100} + \xi_{100}^{010} \\ \xi_{110}^{001} &= \xi_{100}^{001} \end{aligned} \tag{5.32}$$

Simple relations may also be derived for the case of very large magnetic fields. For the Hall effect we have

$$R = 1/nec \tag{5.33}$$

valid for either (100) or (111)-ellipsoids. For the transverse magnetoresistance we obtain, in the case of ellipsoids along the cube axes

$$\begin{aligned} \frac{\rho_\infty}{\rho_0} &= \left\langle \frac{1}{\tau} \right\rangle_{\text{AV}} \langle \tau \rangle_{\text{AV}} (\tfrac{1}{9})(m_1 + m_2 + m_3)(1/m_1 + 1/m_2 + 1/m_3) \\ &= \langle 1/\tau \rangle_{\text{AV}} \langle \tau \rangle_{\text{AV}} [(K+2)(2K+1)/9K] \end{aligned} \tag{5.34}$$

In the case of (111)-ellipsoids we obtain, along cubic axes

$$\begin{aligned} \frac{\rho_{\infty 100}^{100}}{\rho_0} &= [(2K+1)(K+2)/9K] \\ \frac{\rho_{\infty 100}^{010}}{\rho_0} &= \langle 1/\tau \rangle_{\text{AV}} \langle \tau_{\text{AV}} \rangle [(2K+1)^2/3K(K+2)] \\ \frac{\rho_{\infty 110}^{110}}{\rho_0} &= [(2K+1)^2/3K(K+2)] \end{aligned} \tag{5.35}$$

Numerical values for the mass factors in (5.28), (5.29), (5.30), (5.31), (5.34), and (5.35) are shown in Table V. The ratios on the left of (5.35) are the ratio of resistivity at high field to the zero-field resistivity. If the

collision times behave with energy according to a power law like (5.13), then we find

$$\langle 1/\tau\rangle_{AV}\langle\tau\rangle_{AV} = \gamma(p) = \{\Gamma(5/2 + p)\Gamma(5/2 - p)/[\Gamma(5/2)]^2\} \quad (5.36)$$

This factor has the value $32/9\pi = 1.131$ for $p = 1/2$, and $32/3\pi = 3.39$ for $p = 3/2$, corresponding to ideal lattice and impurity scattering, respectively. A similar factor occurs in connection with high-frequency conductivity. The basic relation is

$$\begin{aligned} \sigma_{ht}/\sigma_{dc} &= \gamma(p)(e^2/\omega^2\mu_{dc}^2)/m_{eff}^2 \\ 1/m_{eff} &= (1/3)(1/m_1 + 1/m_2 + 1/m_3) \end{aligned} \quad (5.37)$$

where μ_{dc} is the mobility calculated from the dc conductivity, σ_{dc}.

TABLE V. Mass Factors in Expressions for Magnetoresistance and Hall Constant for Various Examples*

	Ge		Si	
	$K = 19.3$	$K = 14.5$	$K = 5.16$	$K = 4.5$
	Mass factors only			
R_0 (5.28)	0.782	0.797	0.865	0.878
ξ_{100}^{010} (5.29)			1.402	1.354
ξ_{100}^{010} (5.31)	1.280	1.256		
ξ_{100}^{100} (5.30)	1.010	0.923	1.00	1.00
$\rho_{\infty 100}^{010}/\rho_0$ (5.34)			1.746	1.605
$\rho_{\infty 100}^{010}/\rho_0$ (5.35)	1.280	1.256		
$\rho_{\infty 100}^{100}/\rho_0$ (5.35)	4.83	3.79	1.000	1.000
$\rho_{\infty 110}^{110}$ (5.35)	1.280	1.256	1.330	1.286

* For low fields, each factor in the table must be multiplied by $\lambda(p)$ for Hall constant, and by $\eta(p)$ for both longitudinal and transverse magnetoresistance. For high fields, the factor in the table must be multiplied by 1 for longitudinal magnetoresistance and by $\eta(p)$ for transverse magnetoresistance (for a more general tabulation of mass factors, see reference 92).

In the absence of a magnetic field the theory of electrical resistivity may be extended rather simply to cover the case of a τ which depends on direction as well as on energy. In this case, the weighting function is proportional to

$$(v^2) \exp(-\mathcal{E}/kT)dS_P/|\nabla_P \mathcal{E}(P)| \quad (5.38)$$

where dS_P is an element of surface in **P**-space on the energy surface.

It might be thought that since, for small magnetic fields, the change in direction of a carrier is small between collisions, the collision time for

a particular carrier will depend only on its initial velocity, and that therefore the averages involved in $\lambda(p)$ and $\eta(p)$ could be carried out by using the weighting factor (5.38) and including the directional dependence of τ. This procedure is only approximate, however, since the correct treatment of the problem by means of the Boltzmann equation gives rise to additional terms, which involve the gradient of τ tangentially to the energy surface. Although the more general theory has been given in principle by Wilson and others, it has never been applied in detail to an explicit semiconductor model.

Shockley (*79*) and Seitz (*80*) have given, in principle, general solutions applicable to cases when the approximation of small magnetic field is no longer suitable. These solutions are good for arbitrary energy surfaces and arbitrarily large fields, provided only that a collision time exists. In the general case, since the carrier changes its direction of motion significantly in one mean free path, the averaging of the collision time is very complicated. The general solutions have never been applied to the semiconductor problem.

In the case of the energy surfaces appropriate to the valence bands of Ge and Si, the simple analysis for quadric energy surfaces is no longer strictly applicable. Again, although the problem has been solved in principle (*81*), the solution has never been applied in detail to the semiconductor problem. Actually, however, the energy surfaces do not deviate too much from spheres, and so we might expect the multiband theory of Equation (5.18) to be applicable.

Another theoretical quantity of considerable importance in correlating electrical properties and band structures is the thermoelectric power. The theory of the thermoelectric power of semiconductors has recently been reviewed by Herring (*82*). In terms of the notation used in discussing the galvanomagnetic effects, a general expression for thermoelectric power Q may be written

$$eTQ = \langle(\mathcal{E} - E_F)\tau\rangle_{AV}/\langle\tau\rangle_{AV} \tag{5.39}$$

or for a multiband situation

$$eTQ = \frac{\Sigma_i(n_i/m_i)\langle(\mathcal{E} - E_F)\tau_i\rangle_{AV}}{\Sigma_i(n_i/m_i)\langle\tau_i\rangle_{AV}} \tag{5.40}$$

where n_i is the number of carriers in band i, and m_i is the mass, which in the case of ellipsoidal surfaces must be taken as the reciprocal mean mass as given, for example, by (5.37). Equation (5.40) may be interpreted by the statement that the thermoelectric powers arising from each band combine as though they were voltage sources connected in parallel, each having an internal resistance appropriate to the corresponding con-

tribution to the conductivity from that band. The sign of the contribution depends, of course, on whether the band edge is above or below the Fermi energy. Equations (5.39) and (5.40) may also be readily generalized to include the effect of a directional dependence of τ, in the same manner as indicated by Equation (5.38), except that in this case the correction is rigorously valid.

The theory outlined above gives only the electronic contribution to the thermoelectric power. At low temperatures there is another and more significant contribution, which arises from the lattice vibrations (*83*). The theory is discussed in detail in reference *82*. The basic idea is the following. At low temperatures, the electrons give momentum to the lattice vibrations (phonons) at a sufficient rate for the latter to be unable to come to equilibrium among themselves. In consequence, there is a net flow of energy in the lattice, related to the electric current, but carried by the phonons rather than the electrons. Since the thermoelectric power is directly related to energy transport by the electric current, the phonons give rise to a big increase of thermoelectric power at low temperatures. The extensive analysis of Herring (*82*) has demonstrated that the experimental results can be explained in almost all detail by the theory.

For purposes of illustration, we shall now discuss the application of the above theories to n-type Ge, the material which has, perhaps, been most extensively explored from the experimental standpoint. For the galvanomagnetic effects, we shall discuss some measurements made on n-type Ge by G. Benedek (*84*) at 0°C. The experimental results were

$$\begin{aligned} \mu_H &= R_0\sigma_0 = 4480 \text{ cm}^2/\text{volt sec} \\ \xi_{100} &= 1.086 \\ \xi_{100}^{010} &= 0.477 \end{aligned} \tag{5.41}$$

The dependence of these coefficients on magnetic field was checked carefully and found to be negligible up to 2000 Gauss, and so the small field theory is applicable. As a further test of this assumption and of the validity of the theory, we show the test of the symmetry relations (5.32)

$$\begin{aligned} \tfrac{1}{2}\xi_{100}^{100} &= 0.543 & \xi_{110}^{110} &= 0.492 \\ \xi_{110}^{1\bar{1}0} &= 1.006 & \tfrac{1}{2}\xi_{100}^{100} + \xi_{100}^{010} &= 1.021 \end{aligned} \tag{5.42}$$

Since this symmetry would only apply for ellipsoids along the (111)-directions, the good agreement is striking confirmation of the cyclotron resonance data. More detailed confirmation of the angular dependence of the coefficients has been obtained by Pearson and Suhl (*85*).

It is possible to eliminate $\eta(p)$ from (5.30) and (5.31) to obtain an experimental value for the mass ratio K. The best fit is given by $K = 14.5$,

which is somewhat less than the cyclotron resonance value, but barely outside the experimental error. If we take this result, we can compute the individual coefficients if we know λ and η. Taking $p = 0.66$ to fit the drift mobility data of Morin and Maita (*86*), we must still make some assumption regarding the cut-off due to impurity scattering, as in Equation (5.17). It is found that the best fit is obtained for this particular sample by taking $\mu_L^{(T)}/\mu_I^{(T)}$ as $\frac{1}{220}$ at 0°C. Under these conditions, we obtain

$$\begin{aligned} \xi_{100} &= 1.090 \\ \xi_{100}^{010} &= 0.481 \\ \mu_H/\mu_D &= 0.980 \end{aligned} \tag{5.43}$$

Since the impurity scattering was adjusted to give agreement in this case, the only physical significance of (5.43) lies in the fact that the amount of impurity scattering required is not unreasonable. As good a fit of the experimental data cannot be obtained without including impurity scattering. Morin (*87*) reports a value of the Hall-to-drift mobility ratio which is somewhat larger than 0.98, actually about 1.05. However, his samples were somewhat purer, and taking this into account would raise the theoretical value to about 1.05 also.

An analysis of magnetoresistance data on n-type silicon has been made along similar lines by Pearson and Herring (*89*). They report a mass ratio of 4.6 to 4.9, which is in excellent agreement with the value of 5.2 obtained from cyclotron resonance data. Values of η ran from 1.12 at liquid air temperature to 1.21 at room temperature. This is to be compared with a value of 1.18 for Ge at room temperature.

For n-type Ge Equation (5.39) for the thermoelectric power may be written

$$\begin{aligned} -(e/k)Q &= (E_C - E_F)/kT + \tfrac{5}{2} - p \\ &= \ln (A_n/n) + \tfrac{5}{2} - p \\ &= \ln (A_0/n) + \ln (A_n/A_0) + \tfrac{5}{2} - p \end{aligned} \tag{5.44}$$

where A_n and A_0 are the quantities defined in connection with Table IV, n is the number of electrons in the conduction band, and p is the mobility exponent. The only unknown in this expression is the ratio A_n/A_0. This has been determined by Geballe and Hull (*90*) to be 0.65, which is larger than the theoretical value of Table IV by a factor of 1.65. Such a discrepancy suggests that there may be 8 minima instead of 4 for the conduction band, since otherwise an exponent of $p = 0$ would be required in the scattering.

When we come to deal with p-type Ge or either n- or p-type Si, the kind of analysis discussed above runs into difficulties, for in all these

cases the temperature dependence of the conductivity or drift mobility, which is proportional to $\langle\tau\rangle_{AV}$, is much stronger than $T^{-1.5}$. Results for all the cases are summarized in Table VI. Possible explanations for these anomalies will be discussed in the next section. The point we are concerned with here is the implications of these temperature dependencies for the galvanomagnetic coefficients. In all except the case of n-type Ge, if we interpret the drift-mobility results directly in terms of a mean free time $\tau \propto \mathcal{E}^{-p}$, where $p = r - 1$, r being the temperature exponent in Table VI, then we find that all the integrals for the galvanomagnetic effects diverge. If we correct for impurity scattering, of course they converge, but the precise values of λ and η become very sensitive to the

TABLE VI. Temperature Dependence of Drift and Conductivity Mobility in Germanium and Silicon

	n	p
Germanium (*91*)	$3.5 \times 10^{7}T^{-1.6}$	$9.1 \times 10^{8}T^{-2.3}$
Silicon (*61*)	$4.0 \times 10^{9}T^{-2.6}$	$2.5 \times 10^{8}T^{-2.3}$

behavior of $\tau(\mathcal{E})$ in the immediate vicinity of the point where it has a maximum, and are accordingly very sensitive to temperature and impurities and are quite large in magnitude. The greater the relative importance of impurity scattering, the smaller will be both λ and η, so that these quantities will increase with temperature and decrease with impurity content,* even when the impurity mobility calculated separately is many orders of magnitude larger than the lattice mobility. Trends of this sort have been observed in Si and Ge (*61*, *86*, *91*), but at the moment it is difficult to say whether they have any real significance.

Most theoretical explanations of the temperature dependence of mobility have the property in common that they require the energy dependence of the lattice mobility to flatten out at sufficiently low energies. This avoids the convergence difficulty, but does not essentially alter the conclusion that large values of λ and η would be predicted. Thus any reasonable hypothesis for n-type Si would require λ between 3 and 4, whereas the observed value is only slightly greater than 1 (*61*). For p-type Ge, similar conclusions apply. In this case, in a typical sample, a λ of 2.5 is anticipated, whereas the observed value is in the neighborhood of 2. However, as is explained below, this value of 2 is entirely accounted for by the two-band model, so that the extra factor due to the energy dependence of τ would make the factor too large by far. In the case of

* Provided, of course, we remain in the range where scattering is predominantly due to the lattice.

n-type Si, analysis shows that even with the large exponent one should not expect appreciable dependence of λ and η on the magnetic field. In the case of p-type Ge, owing to the presence of high-mobility holes in one of the degenerate bands, one would expect some field dependence of these ratios, even at 0°C, but this would not be enough to bring the values of λ and η into agreement with observation. From these considerations, we are forced to the conclusion that the anomalous temperature dependence of mobility cannot be associated to any substantial extent with an energy dependence of the mean free path; it must rather be associated principally with an explicit temperature dependence of the mobility, probably through the effective mass, at least in p-type Ge and Si, and in n-type Si.* In this connection, more extensive Hall and magnetoresistance measurements in Ge and Si as a function of impurity content and at several temperatures would be very desirable.

C. Herring (*92*) has carried out similar calculations using an explicit model for the scattering. Although the factors of increase for λ are somewhat smaller than indicated by the present analysis, the essential conclusions of the preceding paragraph are not changed.

The case of p-type Ge has been analyzed in detail by Willardson, Harman, and Beer (*77*), and is of special interest. Here it is important to take into account the simultaneous presence of holes of two different masses, as outlined in Equations (5.19)–(5.22). However, it is probably also a reasonable first approximation to take the energy surfaces as spherical. Calculations in Section 6 will show that to a good approximation the mobilities of the two types of holes are in inverse ratio to their effective masses, which gives a mobility ratio of 7.5 to 1. Willardson *et al.* find that their data can best be fitted by assuming $n_3/n_2 = 0.02$ and $\mu_3/\mu_2 = 7.5$. By assuming pure lattice scattering with $p = \frac{1}{2}$, it is possible to evaluate the integrals in the exact expressions (5.18) and (5.19), and so to compute the Hall effect and transverse magnetoresistance coefficients as functions of magnetic field. The results of comparison of theory and experiment are shown in Figs. 10 and 11. In the case of the magnetoresistance, the dotted curve of Fig. 11 shows the computed coefficient for the case in which only one type of hole is assumed. The dramatic influence of the 2% of high-mobility holes is thus indicated. The reason for this drastic behavior, of course, lies in the fact that in the Hall effect the mobility occurs squared, whereas in the magnetoresistance it occurs cubed. The result may be described by saying that the effect of low-mass holes is to increase the apparent value of λ from 1.18 to 1.93

* It should be emphasized that this conclusion is far from securely established. Recent evidence from high field magnetoresistance measurements in several laboratories is not entirely consistent with it.

and the apparent value of η from 1.274 to 3.04, the latter resulting in a 20-fold increase in magnetoresistance at low fields over that for a spherical surface having the mobility of the high-mass holes. It is to be emphasized again that these results are obtained with the assumption that $p = \frac{1}{2}$ for the lattice scattering. The value of 1.93 for $\lambda(p) = \mu_H/\mu_D$

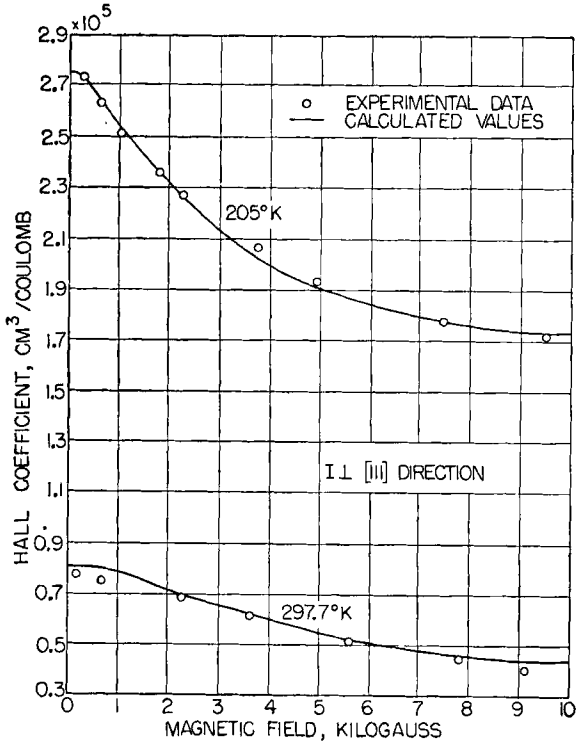

FIG. 10. Variation of Hall constant with magnetic field for p-type germanium. The solid curve is the theoretical curve computed with $n_3/n_2 = 0.02$ and $\mu_3/\mu_2 = 7.5$, assuming the mean free path to be independent of energy. Taken from fig. 1, p. 1515 of reference *77*.

agrees well with the result quoted by Morin (*87*) for holes at high temperatures, based on Prince's drift mobility measurements (*91*). A much larger value would be obtained if a larger p had been taken.

According to the interpretation of Willardson *et al.*, Morin's (*87*) anomalous temperature dependence of μ_H/μ_D for holes is simply the result of the strong field dependence of the Hall coefficient, which is produced by the low-mass holes. As the temperature is lowered, the product $\omega\tau$ becomes rapidly larger and results in the low-field approximation being less and less valid. These results illustrate the necessity of

always studying field dependence in any measurements of Hall effect or magnetoresistance.

While the mobility ratio found by Willardson *et al.* is in good agreement with theory, the percentage of low-mass holes is about two times too small, as can be seen by reference to Table IV. In fact, if we use the values of number and mobility ratio deduced from Table IV, we would

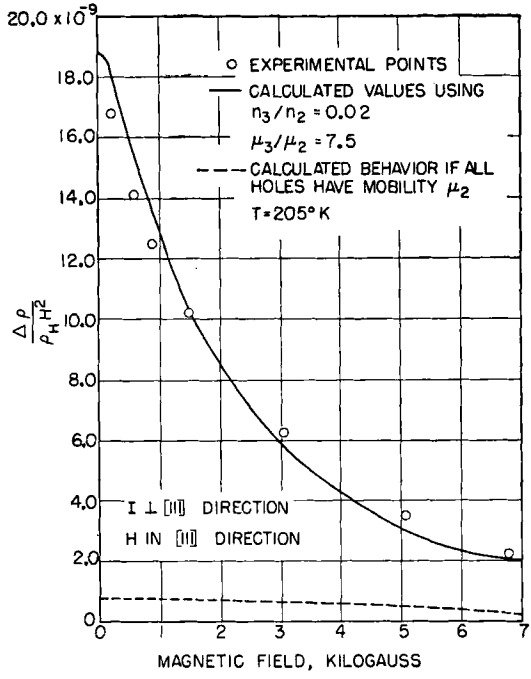

FIG. 11. Variation of magnetoresistance coefficient $\Delta\rho/\rho H^2$ with magnetic field in p-type Ge. The solid curve is the theoretical curve computed with the same constants as in fig. 10. The dotted curve is the result which would be obtained on a one-band model. From fig. 4, page 1517 of reference 77.

obtain an effective λ of 2.66, an effective η of 2.87, and an increase of the magnetoresistance over the one-mass value by a factor of 35. This discrepancy is difficult to explain, since rough arguments suggest that taking into account the warping of the energy surfaces would serve to increase the theoretical values of λ and η still more. It is interesting to note that the density-of-states mass deduced by Geballe and Hull (*90*) from thermoelectric power would agree much better with the interpretation of Willardson *et al.*, if we assumed that the low mass is given correctly by the cyclotron resonance data, but that the high mass is that deduced from thermoelectric power. This also suggests an explicit mass variation

with temperature which is in the right direction and of the correct order of magnitude to explain the deviation of the mobility law from $T^{-1.5}$.

VI. Mechanisms of Scattering

In the previous section we have introduced a phenomenological parameter $\tau(\mathcal{E})$, the collision time. It was indicated that τ has various sorts of energy variation, depending on the mechanism of scattering. In the present section, we shall discuss the validity of the concept of collision time and discuss detailed calculations of the scattering probability. In the course of the presentation, we shall indicate the derivation of the energy dependence for lattice and ionized impurity scattering and shall examine various mechanisms which might explain deviations from the simple $T^{-1.5}$ law for lattice mobility and $T^{+1.5}$ for ionized impurity mobility.

1. Lattice Scattering: The Deformation Potential

For semiconductors, the concept of a deformation potential introduced by Shockley and Bardeen (*93*) has proved very useful in gaining an insight into the mechanism of lattice scattering. We shall begin with the consideration of spherical energy surfaces in cubic crystals. In this case, the position of the edge of an energy band may be written as a linear function of the dilatation

$$E_C - E_{C_0} = E_1 \nabla \cdot \mathbf{u}(\mathbf{r}) \tag{6.1}$$

where E_C is the position of the band edge in the distorted crystal, and E_{C_0} that in the undistorted crystal, E_1 is a constant and $\mathbf{u}(\mathbf{r})$ a vector function which represents the elastic displacement of the lattice at position $\mathbf{r}$. Shockley and Bardeen (*94*) show that for long-wave vibrations of the crystal, the quantity (6.1) may be treated as an effective potential for the motion of an electron near the band edge, the electron satisfying an effective-mass Schrodinger equation in this potential exactly analagously to the case of an impurity center discussed in Section III. The displacement $\mathbf{u}(\mathbf{r})$ may be expanded in terms of elastic waves, and since the fluctuations in potential are small, we may consider the scattering effect of each elastic wave independently. This is an assumption conventionally made, which amounts to saying that only scattering processes involving one phonon are of importance. It is not always justified, but the results are probably not qualitatively changed by multiphonon effects. The perturbation in potential may thus be expanded in Fourier series

$$E_C - E_{C_0} = E_1 \sum_{\mathfrak{d},t} \mathbf{a}_t(\mathfrak{d}) \cdot \mathfrak{d} \exp\,(i\mathfrak{d} \cdot \mathbf{r} - i\omega t) \tag{6.2}$$

where $\mathfrak{d}$ is the wave vector of the phonon and t has three values representing the three directions of polarization. The summation is over values of $\mathfrak{d}$ lying in the first Brillouin Zone shown in Fig. 1. If the material were elastically isotropic, the summation would be only over longitudinal oscillations, since $\mathbf{a}$ would be perpendicular to $\mathfrak{d}$ for transverse modes. In practice, the transverse modes make some contribution, but we shall ignore this for the moment.

Each term in the potential (6.2) is itself a periodic potential and will therefore give rise to diffraction of carriers. However, since each term represents a traveling wave, we must solve the diffraction problem in a system of coordinates moving with the velocity of the elastic wave, and then transform back to the laboratory system. In the moving coordinate system, energy is conserved, and it is therefore not conserved in the laboratory system. In either system, the diffraction condition is simply

$$\mathbf{k} - \mathbf{k}' = \mathfrak{d} \tag{6.3}$$

For spherical surfaces the wave vectors in the moving system are

$$\mathbf{k} - (m/\hbar\omega)u_l^2\mathfrak{d}, \quad \mathbf{k}' - (m/\hbar\omega)u_l^2\mathfrak{d} \tag{6.4}$$

where ω is the phonon frequency and u_l its velocity. Since energies are conserved in the moving system, in the fixed system we have

$$(\hbar^2/2m)(k'^2 - k^2) = \pm\hbar\omega \tag{6.5}$$

the $\pm$ arising from the fact that terms such as (6.2) actually occur in pairs with both signs of ω because the total potential must be real. Equation (6.5) just expresses the conservation of energy, $\hbar\omega$ being the quantized energy of a phonon. It is easily seen that (6.5) still holds when m is a tensor quantity and is hence equally valid for ellipsoidal surfaces. Usually $\hbar\omega$ is small compared with the electron energy, the condition for this being

$$4u_l/v_{\mathbf{k}} \ll 1$$

i.e., the phonon velocity is much less than the electron velocity. This condition may also be written:

$$\mathcal{E} \gg 8mu_l^2 \tag{6.6}$$

and is satisfied so long as $\mathcal{E}$ exceeds about 10^{-3} ev. At this energy, however, ionized impurity scattering usually becomes dominant anyway, so that we are always justified in neglecting $\hbar\omega$ in practice.

With the potential (6.2), the scattering probability per unit time is given by

$$(2\pi/\hbar)\,|E_1\mathbf{a}_l(\mathfrak{d})\cdot\mathfrak{d}|^2 \times \text{density of final states} \tag{6.7}$$

Since only rather long wavelength phonons are involved, we may apply the equipartition theorem and assume that an energy k_0T is associated with each longitudinal mode. This immediately gives

$$|\mathbf{a}_l(\mathfrak{d}) \cdot \mathfrak{d}|^2 \simeq k_0T/\rho u_l^2 V_r \tag{6.8}$$

where k_0 is the Boltzmann constant and ρ the density of the crystal. We thus find for the probability of scattering between $\mathbf{P}$ and $\mathbf{P}'$

$$\mathbf{W}(\mathbf{P} \rightarrow \mathbf{P}') = (2\pi/\hbar)(k_0T/\rho u_l^2)E_1^2\, dS_{\mathbf{P}'}/|\nabla_{\mathbf{P}'}\mathcal{E}| \quad (1/h^3) \tag{6.9}$$

where $dS_{\mathbf{P}'}$ is the surface element in reduced momentum space. For spherical energy surfaces the probability (6.9) is seen to be independent of the angle of scattering. The same proves to be true even for ellipsoidal surfaces. Thus, the total probability of scattering to all final states of energy $\mathcal{E}$ is just proportional to the density of states, or the collision time becomes

$$1/\tau = (8\pi^3/h^4)(2m)^{3/2}\mathcal{E}^{1/2}(k_0T/\rho u_l^2)E_1^2 \tag{6.10}$$

For ellipsoidal surfaces $m^{3/2}$ is replaced by $(m_1m_2m_3)^{1/2}$, but otherwise the result is unchanged. The value $p = \frac{1}{2}$ used in Section 5, of course, follows immediately from (6.10), as does the $T^{-1.5}$-law for mobility. The expression for the mobility is, explicitly

TABLE VII. Deformation Potential Theory for n-Type Ge and Si, Assuming Dilatational Contributions to the Potential Only

	Ge	Si
Electron mobility at 300°K	3900 cm²/volt-sec	1420
Mean ρu_l^2	1.545×10^{12} dynes/cm²	1.97×10^{12}
$(m_e/m)^{5/2}$	80.8	20.4
E_{1C}	13.62	12.80 ev

$$\mu_L = \frac{2(2\pi)^{1/2}e\hbar^4\rho u_l^2}{3m^{5/2}(k_0T)^{3/2}E_1^2} \tag{6.11}$$

For spheroidal energy surfaces, it is necessary to replace m according to

$$m^{-5/2} = (m_1m_2^2)^{-1/2}(\tfrac{1}{3})(m_1^{-1} + 2m_2^{-1}) \tag{6.12}$$

Relative to the electronic mass the values for n-type Ge and Si are $(m/m_e)^{-5/2} = 80.8$ and 20.4, respectively. Numerically we may write

$$\mu_L = 3.0 \times 10^{-5}(\rho u_l^2/T^{3/2}E_1^2)(m_e/m)^{5/2} \tag{6.13}$$

where μ_L is in cm²/volt-sec, E_1 in ev per unit dilation, and ρu_l^2 in dynes/cm², having the dimensions of an elastic modulus. We obtain the results of Table VII for Ge and Si. These results assume that all the scattering

is due to dilatation, and that scattering between different ellipsoids—so-called inter-valley scattering—can be neglected, as can scattering by optical modes of vibration. Actually none of these assumptions are too well fulfilled in practice.

Herring (*92*) has considered optical mode and inter-valley scattering for n-type Ge and Si. The principle involved is the same in both cases. In the ordinary case of intra-valley scattering we have seen that only long-wavelength lattice waves are involved. These have a small quantum energy $\hbar\omega \ll k_0T$, so that we were able to apply the classical equipartition law for computing the mean square amplitude, and furthermore we could neglect the energy loss of the electron in the scattering process. For inter-valley scattering, on the other hand, the electron changes its reduced momentum by a large amount and, in consequence of the selection rule (6.3), the scattering process must involve an energetic phonon. When the scattering process involves emission of a phonon, i.e., an electron giving energy to the lattice, the electron will have less energy in the final state, and so will lie on a smaller ellipsoid with a lower density of states. In the case where a phonon is absorbed, the density of states will be increased in the final state, but, on the other hand, the phonon must be initially excited in the lattice, a situation which has a small probability when $\hbar\omega \gg k_0T$. Since the phonons involved in inter-valley scattering all have about the same energy, characteristic of the distance between valleys in momentum space, and since the scattering matrix element is probably not a strong function of the position of the electron on the initial or final ellipsoid, it is a fairly good approximation to write the inter-valley scattering in terms of a phonon process involving only one frequency. Herring (*92*) writes the total scattering probability in the form

$$1/\tau = W_0 + W_a + W_e \tag{6.14}$$

where

$$W_0 = W_1 \frac{k_0T}{\hbar\omega}\left(\frac{\varepsilon}{\hbar\omega}\right)^{1/2} \tag{6.15}$$

$$W_a = W_2 \frac{(1 + \varepsilon/\hbar\omega)^{1/2}}{\exp(\hbar\omega/k_0T) - 1} \tag{6.16}$$

$$W_e = W_2 \frac{(\varepsilon/\hbar\omega - 1)^{1/2} \exp(\hbar\omega/kT)}{\exp(\hbar\omega/kT) - 1}, \quad \varepsilon \geq \hbar\omega \tag{6.17}$$

$$W_e = 0 \qquad \varepsilon < \hbar\omega$$

In (6.14) the first term represents intra-valley scattering, the second inter-valley scattering with absorption of a phonon, and the third inter-valley scattering with emission of a phonon. The collision time obtained from (6.14) has a complex energy and temperature dependence which can give rise to temperature variation of the mobility considerably more rapid

than the $T^{-1.5}$-law obtained from (6.11). However, $d \ln \mu / d \ln T$ obtained from (6.14) is far from constant with temperature, especially if the inter-valley scattering is such as to give rise to a rather large exponent. If phonons of several different energies are actually involved the curves may resemble more closely the ideal log-log linearity which seems to be characteristic of the variation of observed mobility with temperature, even for exponents as high as 2.3. Herring has computed the functions corresponding to $\lambda(p)$ and $\eta(p)$ using the scattering law (6.14) in place of the simple exponent behavior. Because the behavior approaches $\varepsilon^{1/2}$ for low enough energies, the integrals do not diverge, and the functions do not reach such large values as when the law (5.13) is used with a cut-off by ionized impurity scattering. These questions need more investigation by means of Hall effect and magnetoresistance measurements on samples of varying impurity content.

It is interesting to observe that the theory which gives about the right exponent for n-type Ge at room temperature corresponds to about equal contributions from intra-valley and inter-valley scattering. However, the ratio μ_H/μ_D seems to have a stronger temperature dependence for this assumption than is indicated by the experimental results of Morin (*87*).

Optical phonons are quantized lattice vibrations which can occur in structures which possess more than one atom per unit cell: they correspond to nonequivalent atoms moving in opposite phase to each other, but with relatively slow change in phase from cell to cell. These modes can give rise to intra-valley scattering, but with a much higher energy change of the electron than that which occurs with normal scattering by long-wave acoustical modes considered in Equation (6.11). Formally, the theory is then almost identical to that for inter-valley scattering, and can thus also give rise to anomalous temperature dependence of the mobility. Optical mode scattering does not lend itself to treatment by the deformation potential method, and reliable estimates of its relative importance are hard to make.

For the many-valley type of structure which occurs in the conduction bands of Ge and Si, the transverse as well as the longitudinal modes can make an important contribution to normal intra-valley scattering. The different energy ellipsoids are not equivalent with respect to the stress axis, and so their band edges may be shifted differently for a given tensile stress on the crystal.

From symmetry considerations it may be shown that the only strains which shift a given ellipsoid in a cubic crystal are a pure dilatation, already considered, and a shear corresponding to a tension in the direction of the vector position of the center of the ellipsoid, with a symmetrical

compression in the two directions at right angles. The complete expression for the shift may be written

$$\Delta E_c^{(i)} = \hat{\mathbf{n}} \cdot \{E_1 \mathbb{1} + E_2(\epsilon - \tfrac{1}{3}\Delta \mathbb{1})\} \cdot \hat{\mathbf{n}} \tag{6.18}$$

where $\hat{\mathbf{n}}$ is the unit vector in the direction of the position of the center of the ellipsoid in **P**-space, $\mathbb{1}$ is the unit dyadic, Δ the dilatation, and ϵ the strain tensor of the crystal. The scattering probabilities for longitudinal and transverse vibrations are

$$W_l(\theta) = \frac{2\pi}{\hbar} |E_1 + E_2 \cos^2\theta|^2 (k_0 T/\rho u_l^2) \times \text{density of states} \tag{6.19}$$

$$W_t(\theta) = \frac{2\pi}{\hbar} E_2^2 \sin^2\theta \cos^2\theta (k_0 T/\rho u_t^2) \times \text{density of states} \tag{6.20}$$

where u_l and u_t are, respectively, the velocities of longitudinal and transverse waves, and θ is angle between the reduced momentum ($= \mathbf{P}' - \mathbf{P}$)

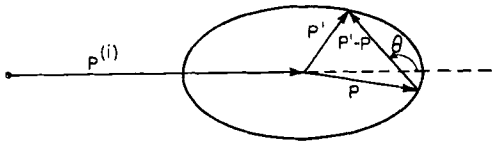

FIG. 12. Scattering process for transverse modes, showing definition of the angles used in Equation (6.20).

of the vibration and the vector $\mathbf{P}^{(i)}$ defining the position of the center of the ellipsoid and hence one of its principal axes. The situation is sketched in Fig. 12.

Thus, unlike the simple case of (6.7), the scattering has a complex angular dependence. In such cases, the existence of a collision time is doubtful; one should really go back to a rigorous solution of the Boltzmann transport equation. However, Herring has shown (*92*) that, provided the scattering is not too anisotropic, it is still a reasonable approximation to replace the complete scattering problem by a collision time. In this case, however, the collision time is defined by the integral

$$1/\tau(\mathbf{P}) = (1/h^3) \int W(\mathbf{P} \to \mathbf{P}') \frac{[\mathbf{v}(\mathbf{P}) - \mathbf{v}(\mathbf{P}')] \cdot \mathbf{v}(\mathbf{P})}{[\mathbf{v}(\mathbf{P})]^2} \frac{dS_{\mathbf{P}'}}{|\nabla_{\mathbf{P}'} \mathcal{E}(\mathbf{P}')|} \tag{6.21}$$

where $W(\mathbf{P} \to \mathbf{P}')$ is obtained from (6.19) and (6.20). Equation (6.21) represents the sum of the probabilities of scattering out of the state **P** on the ellipsoid to all possible final states on the same ellipsoid, weighted according to the relative change in velocity produced by each collision. The integral (6.21) may be evaluated simply for the case when **P** is along

the unique principal axis of the ellipsoid. Then the weighting factor becomes simply $(1 - \cos\theta)$, and we have

$$1/\tau_l = (8\pi^3/h^4)(2m)^{3/2}\mathcal{E}^{1/2}(k_0T/\rho u_l^2)[E_1^2 + \tfrac{2}{3}E_1E_2 + \tfrac{1}{15}E_2^2] \quad (6.22)$$

$$1/\tau_t = (8\pi^3/h^4)(2m)^{3/2}\mathcal{E}^{1/2}(k_0T/\rho u_t^2)[\tfrac{2}{15}E_2^2] \quad (6.23)$$

where τ_l and τ_t are the collision times associated with longitudinal and transverse waves, respectively, and both associated with an initial value **P** along the principal axis. For spheroidal surfaces $m^{3/2} = (m_1m_2^2)^{1/2}$.

E. Vogt* has carried out numerical calculations of the ratio of the collision time for **P** along the unique axis to that for **P** along an axis at right angles. This should give a good measure of the anisotropy. The degree of anisotropy depends, of course, on the relative values of the constants E_1 and E_2. In general, however, the anisotropy is not large for ellipsoids having the shape appropriate to the Ge conduction band, although it would become much larger if the energy surfaces were spherical, as Herring shows (*92*).

The value of the constant E_2 may be estimated from data on piezoresistance (*92*, *95*). This is an effect of nonhydrostatic strain on the electrical resistance, and in extrinsic material arises from two sources. In the first place, shearing strain raises some ellipsoids relative to others, resulting in different carrier populations in different ellipsoids. In the second place, the change in the ellipsoids influences the inter-valley scattering and hence the over-all collision time. If we ignore the second effect for the moment, Herring has given (*92*) a relation which permits calculation of E_2. The piezoresistance coefficients are defined in a manner analogous to elastic constants. Thus, for example, we have, for n-type Ge

$$m_{44} = -\frac{1}{\sigma}\frac{\partial\sigma_{xy}}{\partial\epsilon_{xy}} = -\frac{1}{3}\frac{E_2}{k_0T}\frac{K-1}{2K+1} \quad (6.24)$$

and for Si we obtain

$$\frac{1}{2}(m_{11} - m_{12}) = -\frac{1}{2\sigma}\left(\frac{\partial\sigma_{xx}}{\partial\mathcal{E}_{xx}} - \frac{\partial\sigma_{xx}}{\partial\mathcal{E}_{yy}}\right) = -\frac{1}{2}\frac{E_2}{k_0T}\frac{K-1}{2K+1} \quad (6.25)$$

where σ_{xx}, etc. represent components of the conductivity tensor, $\mathcal{E}_{xx}$, etc. represent components of the strain tensor, and K is the mass ratio as defined previously. The unlisted components vanish theoretically in each case. The results are as shown in Table VIII. The row labelled Transverse/Longitudinal gives the ratio of the total collision probability for scattering by transverse waves to that by longitudinal waves. The last row gives the ratio of anisotropies of the relaxation times, taken from Vogt's (*92*)

* Described in reference *92*, especially Figure 12. Note that our definitions of E_1 and E_2 differ slightly from Herring's Ξ_d and Ξ_u.

calculations for the appropriate ratio E_2/E_1. It will be noted that there are two possible solutions for E_1 depending on whether it has the same or the opposite sign from E_2. Both results are shown in the table. These results must be regarded as no more than crude order-of-magnitude estimates, in view of the neglect of the anisotropy of the relaxation time and of inter-valley scattering. It seems most likely that inclusion of inter-valley scattering would result in reduced values of E_1, and that the neglect of anisotropy is of relatively little importance for the lower row of

TABLE VIII. Calculation of E_2 from Piezoresistance Data and of E_1 from Mobility Data for n-Ge and n-Si

	n − Ge	n − Si
m_{44}	−93.0	
$\frac{1}{2}(m_{11} - m_{12})$		−79.5
E_2 (ev)	+15.3	+10.8
E_1 (ev)*	−13.05	−13.8
	or	or
	+ 2.85	+ 6.6
u_t^2/u_l^2	0.306	0.513
$\frac{\text{Transverse}}{\text{Longitudinal}}$	0.54	0.335
$\frac{\tau(0°)}{\tau(90°)}$	2.5 or 0.8	—

* In estimating E_1 use was made of Equation (6.22) with anisotropy neglected.

values in the table, and would result in a smaller E_1 for the upper row. A self-consistent calculation, using Vogt's results to estimate the effect of anisotropy, gives $E_1 = -6.5$ instead of -15.7 for the upper row. These crude calculations are nevertheless sufficient to indicate that transverse and longitudinal effects can be of the same order of magnitude.

More rigorous theories of scattering by transverse vibrations have been discussed by Herring (*82*, *92*) and by Dumke (*96*), but no detailed numerical results have been published.

The deformation potential approach can also be used to discuss p-type Ge. A detailed theory has not been given, but a few general conclusions can be reached by relatively simple arguments. For pure dilatations the whole band structure is moved up and down as a unit, much as in the theory for the conduction band. For a first approach the band structure may be approximated by concentric spheres, neglecting the warping of the energy surfaces. Even with this simplification, the calculation is straightforward but rather involved, and we shall not reproduce it here. Since there are four degenerate bands in all, inter-band scattering is im-

portant, even when one considers only acoustical modes, so that energy conservation may be assumed in the scattering process. An equation similar to (6.12) results, except that the mass factor $(m_e/m)^{5/2}$ is replaced as follows

$$\text{High mass } \left(\frac{m}{m_e}\right)^{5/2} \rightarrow \frac{1}{2}\left(\frac{m_1}{m_e}\right)^{5/2} + \frac{1}{2}\left(\frac{m_1}{m_e}\right)\left(\frac{m_2}{m_e}\right)^{3/2} \tag{6.26}$$

$$\text{Low mass } \left(\frac{m}{m_e}\right)^{5/2} \rightarrow \frac{1}{18}\left(\frac{m_2}{m_e}\right)^{5/2} + \frac{1}{2}\left(\frac{m_2}{m_e}\right)\left(\frac{m_1}{m_e}\right)^{3/2} \tag{6.27}$$

Here m_1 is the mass for the high-mass band, about $0.3m_e$ for Ge, and m_2 is that of the low-mass band, about $0.04m_e$. In each case the first term on the right represents the contribution of intra-band scattering and the second that of inter-band scattering. In the case of the low-mass holes almost the entire scattering by acoustical modes is scattering into the high-mass band rather than intra-band scattering. Furthermore, the scattering depends essentially on the density of final states. Since the final state is the same both for low-mass and high-mass holes, namely the high-mass band, this implies that the collision times for the two types of holes are approximately equal and that the mobilities are in the inverse ratio of the masses. For holes in Ge, the effective values in (6.26) and (6.27) are

$$\begin{aligned} &\text{High mass } (m_e/m)^{5/2} = 26.5 \\ &\text{Low mass } (m_e/m)^{5/2} = 232.0 \end{aligned} \qquad E_1 = 12.4 \text{ ev} \tag{6.28}$$

These are the factors by which the mobility is increased over that for electrons of normal mass. The value of E_1 deduced, of course, neglects the influence of shear, but it is based on the contribution to the mobility of the high-mass holes.

The existence of a strong piezoresistance effect in p-type Ge and Si shows that the band structure is significantly influenced by shearing strains, and we should therefore expect transverse vibrations to contribute to the lattice scattering. From symmetry considerations, the only thing a pure shear could do would be to shift the low-mass and high-mass bands relative to each other; it cannot shift the band structure as a unit. For points in **P**-space, far enough away from the origin for the separation of the degenerate bands to be large compared with the displacement of the bands which would be produced by the shear, the two bands are displaced up and down relative to each other by an amount which depends on the angle between the **P**-vector and the stress axis. For points nearer the origin in **P**-space, the bands are distorted in a complicated way, since the degeneracy is removed at the origin. However, in the piezoresistance experiments of Smith (*95*), the strains involved were of the order of 10^{-4}, which, with energy shifts of the order of 10 ev per unit

strain, would give an energy shift of only 0.001 ev, which is much less than k_0T at room temperature and so affects a negligible portion of the band. Similarly, it is shown that the shearing modes which can give rise to deformation potential scattering have a root mean square strain amplitude of about 5×10^{-4}, corresponding to an energy shift of 0.005 ev, which is larger but still reasonably small compared with k_0T at room temperature.

For points in **P**-space sufficiently far from the origin, the energy shift may be written

$$\Delta E^{(i)} = \hat{\mathbf{n}} \cdot \left[E_1 \underset{\sim}{1} \Delta \pm E_2 \left(\underset{\sim}{\epsilon} - \underset{\sim}{1} \frac{\Delta}{3} \right) \right] \cdot \hat{\mathbf{n}} \tag{6.29}$$

where the symbols have nearly the same meaning as in (6.18), i.e. $\hat{\mathbf{n}}$ represents the unit vector in the direction of **P**, $\underset{\sim}{\epsilon}$ is the strain tensor, and Δ the dilatation, or diagonal sum of the strain tensor. The positive sign corresponds to the high-mass band and the negative sign to the low-mass band. To avoid confusion, we shall adopt the sign convention that E_1 or E_2 is positive when the shift is in a direction into the band, so that a positive E represents an upward shift for a conduction band and a downward shift for a valence band.

Using the result (6.29) to compute the strains associated with longitudinal and transverse elastic waves, we may obtain expressions for the scattering matrix elements, and finally expressions for the reciprocal collision time. We may express the final result as follows: For scattering within high-mass band

$$1/\tau = (8\pi^3/h^4)(2m_1)^{3/2}\mathcal{E}^{1/2} \left\{ \frac{k_0T}{\rho u_l^2} \left(\frac{1}{2} E_1^2 + \frac{3}{5} E_1E_2 + \frac{1}{4} E_2^2 \right) + \frac{k_0T}{\rho u_t^2} \left(\frac{3}{20} E_2^2 \right) \right\} \tag{6.30}$$

and for scattering from high-mass to low-mass band, we have

$$1/\tau = (8\pi^3/h^4)(2m_2)^{3/2}\mathcal{E}^{1/2} \left\{ \frac{k_0T}{\rho u_l^2} \left[\frac{1}{2} E_1^2 + \left(\frac{1}{4} + \frac{3}{20} \sqrt{\frac{m_1}{m_2}} E_1E_2 \right) + \left(\frac{7}{80} + \frac{3}{80} \sqrt{\frac{m_1}{m_2}} \right) E_2^2 \right] + \frac{k_0T}{\rho u_t^2} \left(\frac{9}{40} E_2^2 \right) \right\} \tag{6.31}$$

and for scattering from low-mass to high-mass band we have a similar expression to (6.31), but with the sign of E_2 reversed and m_1 and m_2 interchanged; these results reduce to the previous ones when we let $E_2 = 0$.

In order to obtain an estimate of E_2 it is necessary to develop a theory for the piezoresistance effect in p-type material. Here it turns out that the principal effect arises from the anisotropy in the velocity on a sur-

face of constant energy which is set up by the strain. The greatest difficulty in the theory originates from the use of the spherical approximation for the energy surfaces in the valence band. This implies that the piezoresistance effect should be independent of orientation and in particular that the m-coefficients defined by Herring (*92*) should satisfy the isotropy relation

$$m_{44} = \tfrac{1}{2}(m_{11} - m_{12}) \tag{6.32}$$

whereas in fact we have $m_{44} = +66.0$, $\tfrac{1}{2}(m_{11} - m_{12}) = -6.0$. The best we can do within the framework of the present approximations is to take an angular average of the m's. The most appropriate average turns out to be:

$$m_{ii} = \tfrac{3}{5}m_{44} + \tfrac{1}{5}(m_{11} - m_{12}) = +38.4 \tag{6.33}$$

Setting the theoretical piezoresistance equal to this average, we obtain the relation

$$\langle m_{ii} \rangle = \left\{\frac{3}{5} - \frac{1}{25}\frac{1 + 2.285\xi + 0.83\xi^2}{1 + 1.2\xi + 1.48\xi^2}\right\}\frac{E_2}{k_0 T} = 38.4 \tag{6.34}$$

where in writing the theoretical expression we have neglected the contribution of low-mass holes to the mobility and also the contribution of inter-band scattering. Of the two terms in brackets in (6.34), the first represents the effect of the anisotropy in the velocity introduced by the strain and the second represents the effect of the anisotropy in the collision time itself. This latter effect is very small, and is the only one which depends specifically on which band we are dealing with. The symbol ξ is simply the ratio E_2/E_1, so that the shear coefficient E_2 is involved implicitly as well as explicitly in (6.34). Numerical solution of (6.34) in a self consistent manner leads to a best value of E_2 of $+1.66$ ev per unit strain.

With the use of (6.30), knowing E_2 from the above discussion, we can solve for E_1 in terms of the mobility for high-mass holes, which, following the discussion of Section 5, we can take as 1640 cm^2/volt-sec. This solution gives the following alternative values of E_1 for the valence band of Ge

$$E_1 = -13.3 \text{ ev or } E_1 = +11.3 \text{ ev}$$

where, according to the previously adopted conventions, the first case represents an upward shift of the band edge position with dilatation, and the second represents a downward shift. With either alternative solution the contribution of shear to the total scattering is small; for example, for the first solution, the total contribution of transverse waves to the scattering is only about 2.5% of the total scattering. This conclusion finds

some support in experiment, since, as may be seen by reference to Equations (6.30) and (6.31), the conclusion that the mobilities of the two types of holes are in the inverse ratio to their masses depends upon the fact that most of the scattering arises from dilatation.

The values of E_1 shown in Table VIII were computed from Equations (6.22) and (6.23) directly, without taking into account the anisotropy in the scattering. This can be taken into account rather crudely for n-Ge, taking as the true collision time $\frac{2}{3}\tau(90°) + \frac{1}{3}\tau(0°)$, where the two collision times are those calculated by Vogt (*92*) for the appropriate value of E_2/E_1, and the appropriate effective masses and elastic anisotropy. With all these assumptions, we obtain the following numerical results for Ge

Shift in conduction band edge	+ 4.5	or	− 6.5	
Shift in valence band edge	+13.3	or	−11.3	(6.35)
Difference	− 8.8		+ 4.8	

The difference in the two band shifts provides an independent check on the deductions, since it can be compared with the observed pressure shift of the band gap. For Ge this is equivalent to −4.3 ev per unit dilatation. The combination which comes closest to agreeing with the pressure shift is that shown in the first column of (6.35). The agreement is not very good.

Since the mobility does not obey the $T^{-1.5}$-law for either electrons or holes, there is an ambiguity in interpretation depending on what temperature is chosen to match the deformation potentials. If liquid nitrogen is chosen, Equation (6.35) is replaced by (6.35′)

Shift in conduction band edge	+2.33	−5.8	
Shift in valence band edge	+8.04	−6.04	(6.35′)
Difference	−5.71	+0.24	

Here the agreement with the pressure shift is somewhat improved. Considering the many omissions and approximations made in these estimates* the agreement is fairly satisfactory, and probably sufficient to fix the choice of matching sets of values in these equations.

The calculations for p-type Ge fail completely to provide an understanding of the temperature dependence of hole mobility in Ge, since all the mechanisms discussed lead to an $\varepsilon^{1/2}$ dependence of the collision probability. Furthermore, since it is believed that all the band edges lie at $\mathbf{P} = 0$, we cannot invoke inter-valley scattering as a possible mechanism.

* A theory for scattering in p-type Ge has also been given by H. Ehrenreich and A. W. Overhauser, *Bull. Am. Phys. Soc.* **30,** 10 (1955), abstract D8. It differs in its conclusions in important respects from the present analysis, especially with regard to the magnitude of scattering by transverse modes.

In the case of Ge, as we shall see in Section VII, strong evidence indicates that the third valence band lies 0.28 deeper than the band edge, so that scattering to or from this band cannot be important at ordinary temperatures. Moreover, it is probably sufficiently far away in energy so that it cannot significantly affect the curvature, at least of the high mass band. In addition to the deformation potentials already discussed, there exists the possibility of a potential associated with alterations in curvature of the bands with strain (*92–94*). This could give rise to a contribution to the collision probability which is proportional to $\mathcal{E}^{3/2}$ instead of $\mathcal{E}^{1/2}$. However, this type of scattering would have to predominate over the normal scattering in order to explain a $T^{-2.3}$-law for mobility. Scattering by optical modes could produce something resembling the observed temperature dependence. However, according to Ehrenreich and Overhauser, this type of scattering is much too small to explain the observed result.* Furthermore, all the mechanisms discussed lead to a large energy dependence of the collision time, and we have already seen in connection with Section V that it is hard to reconcile the galvanomagnetic results with other than a nearly normal energy dependence of the collision time. At the present time the anomalous temperature dependence of mobility in p-type Ge must be regarded as a major unsolved problem. It is interesting to note that the temperature law is extremely sensitive to strain, as shown by the experiments of Lawrence (*97*). This is mainly a consequence of the fact that according to the theory, the piezoresistance effect should be inversely proportional to the temperature. Lawrence's results are in qualitative agreement with this conclusion.

2. Ionized Impurity Scattering

We have seen in Section III that impurities different in valence from the host lattice give rise to long-range Coulomb fields in the semiconductor with a potential of the form $V(r) = e^2/Kr$ where K is the macroscopic dielectric constant. If we neglect what happens in the central cell this potential gives rise to a scattering probability

$$W(\mathbf{P} \rightarrow \mathbf{P}')(d\mathbf{P}'/h^3) = \frac{e^4}{K^2} \frac{4}{|\mathbf{P} - \mathbf{P}'|^4} d\mathbf{P}' \tag{6.36}$$

where $\mathbf{P}$, $\mathbf{P}'$ are the initial and final reduced momentum vectors, and the probability shown is the total probability of scattering into a volume

* These calculations were made with the deformable ion model of Bloch (*113*). This model severely overestimates the mobility in Ge due to the fact that it actually omits an important part of the perturbation in the potential resulting from the lattice vibrations. This part of the perturbation is included in the deformation potential model.

$d\mathbf{P}'$ of reduced momentum space. When (6.36) is integrated over all final states to obtain the total collision probability, the integral diverges. This is true even when the collisions are weighted with the velocity change in accordance with (6.21). The divergence arises from the small-angle scattering characteristic of a Coulomb potential. In order to obtain a finite result, it is necessary to "cut off" the Coulomb field of the impurity at some distance. In the original treatment of Conwell and Weisskopf (*88*) this was done by omitting scattering processes from (6.36) which arose from encounters between the incident particle and the impurity, corresponding to distances of closest approach greater than half the mean distance between impurity ions in the crystal. A later more rigorous formulation by Brooks and by Herring (*88*) introduced a shielding factor into the potential. This arises from the fact that the other electrons in the conduction band distribute themselves around the impurity in such a way as to cancel its field at large distances. When only one sign of impurity is present, the distance at which cancellation occurs is of the same order of magnitude as the mean distance between impurities, so that this more rigorous treatment gives about the same result as the Conwell-Weisskopf analysis. To obtain the scattering probability taking into account shielding, it is necessary to replace $\mathbf{P} - \mathbf{P}'$ in (6.36) by

$$\left\{|\mathbf{P} - \mathbf{P}'|^2 + \frac{e^2h^2n}{\pi K k_0 T}\left(2 - \frac{n}{N_D}\right)\right\}^2 \tag{6.37}$$

where n is the number of electrons in the conduction band and N_D is the number of ionized donors. Equation (6.37) is valid when only one type of carrier is present; otherwise N_D should be replaced by $N_D - N_A$, the excess of donors over acceptors. If the donors are completely ionized, the factor $(2 - n/N_D) = 1$. For material in which holes and electrons are present simultaneously, n is replaced by $n + p$, the total number of carriers.

We may integrate (6.36), putting in the velocity weighting factor and taking into account the conservation of energy in the scattering process. If the ionized impurities are randomly arranged in the lattice, they scatter independently, and the total scattering is proportional to the number of impurities. Thus we have for a spherical energy surface

$$\frac{1}{\tau} = \frac{\pi}{m^{\frac{1}{2}}} \frac{1}{(2\mathcal{E})^{\frac{3}{2}}} \frac{e^4n}{K^2} \left\{\ln (1 + b) - \frac{b}{1 + b}\right\} \tag{6.38}$$

where

$$b = \frac{8\pi m K k_0 T}{e^2h^2n} \mathcal{E} \tag{6.39}$$

In the usual situation b is very large, so that the bracket may be replaced by $\ln b$. From (6.38) and (6.39), the mobility may be calculated

as in Section V. In making this calculation, the logarithm is taken outside the integral, and its argument is replaced by the value assumed when the rest of the integrand has a maximum. This value is $\mathcal{E} = 3k_0T$, and the mobility is then given by

$$\mu_I = 2^{7/2}\pi^{-3/2}(k_0T)^{3/2}K^2e^{-3}m^{-1/2}n^{-1}\left[\ln\frac{6m(k_0T)^2K}{\pi e^2\hbar^2 n(2 - n/N_D)}\right]^{-1} \tag{6.40}$$

In the case of intrinsic material, or if donors and acceptors are present simultaneously, n^{-1} before the bracket is replaced by the reciprocal of the sum of the number of ionized donors and acceptors, and in the argument of the logarithm n is replaced by the total number of carriers. Since impurity scattering is mainly important at relatively low temperatures, the case of greatest interest is that in which there is only one type of carrier, but both donors and acceptors are present in the crystal. In this case n must be replaced by the total number of ionized donors $n + N_A$ outside the bracket, whereas inside the bracket it is replaced by

$$\begin{aligned} n \to n + \left(1 - \frac{n + N_A}{N_D}\right)(n + N_A) \\ = (1 - f_D{}^2)N_D - N_A \end{aligned} \tag{6.41}$$

where f_D is the Fermi factor for the donor levels, i.e., the probability that the donor is occupied, and N_A is the number of acceptors, assumed completely occupied at all times. It is interesting to note that (6.41) does not vanish as the number of carriers approaches zero in the conduction band, since some shielding of the donor potential results from repopulation of adjacent bound levels. The limiting value of (6.41) is

$$(1 - N_A/N_D)N_A \tag{6.42}$$

The potential distribution inside a semiconductor due to random distributions of impurity has been discussed by James and co-workers (*98*).

The result of Equation (6.40) has already been used in the discussion of combined lattice and impurity scattering, in Section V, above.

It is now of importance to inquire what sort of modifications will be introduced by nonspherical energy surfaces or by degenerate bands. This problem has not been considered in the literature, and we shall content ourselves here with some order-of-magnitude estimates. We take advantage of the fact that the dominant contribution to the scattering comes from very small scattering angles, and use the approximation that a relaxation time exists which is defined in accordance with Equation (6.21). This integral can be evaluated easily only for the special cases in which the initial state **P** is along one of the principal axes of the ellipsoid. The

result may be represented by replacing the effective mass factor in (6.38) according to

$$m^{-1/2} \rightarrow (1/2)m_1^{1/2}(1/m_2 + 1/m_3) \tag{6.43}$$

where m_1 is the principal mass in the direction of **P**, and m_2 and m_3 are the other two principal masses. For germanium this gives collision probabilities, one of which is roughly 14.1 times larger than that for the free electron mass, and the other two of which are about 1.83 times that for the free electron mass. Knowing only the collision times for the principal directions, it is not certain how they should be averaged to obtain a mobility, but two different plausible averaging methods indicate that the mobility for Ge should lie between 2.6 and 3.7 times the free electron value, corresponding to a mobility effective-mass from $1/7$ to $1/14$ the free electron mass. This is to be compared with a value of 1/1.8 estimated earlier from magnetoresistance data. It also represents a value somewhat larger than that which best fits Hall mobility data according to Debye and Conwell (*24*, *26*). One must conclude that the observed impurity scattering is somewhat higher than might be anticipated from theory. For silicon, the collision probabilities are 5 times and 1.34 times the free electron values, and the mobility is probably less than 2.25 times the free electron value. The above theory is very crude because the collision-time approximation is bad when the anisotropy is as large as indicated above.

For scattering of holes in the valence band, we need consider only the scattering within a single degenerate band. This is because the scattering comes so largely from small momentum changes, and there is a minimum momentum change for interband scattering. The quantitative criterion that interband scattering is negligible turns out to be exactly that the quantity b, defined by Equation (6.39) be much greater than unity, a condition which is always fulfilled in practice except under conditions where the semiconductor becomes degenerate at low temperature or for high carrier concentrations. The impurity scattering formula does not appear to have been critically tested for holes in Ge or Si. The consequence of all this is that the ionized impurity scattering can be computed for a single spherical band, and that the mass factor does not include the factor $1/2$ which we found in the case of lattice scattering in Equation (6.26).

One consequence of this conclusion is that, in the impurity scattering regime, the mobility of low-mass holes is about 2.8 times that of high-mass holes, and therefore the influence of the low-mass holes on Hall effect and magnetoresistance should be considerably less than in the lattice scattering range. No experiments bearing on this question are available.

It may be remarked that all the treatments of impurity scattering are relatively crude. In the first place, use is made of the so-called Born approximation, in which the scattering is treated as a small perturbation on the motion of the incident carrier; and in the second place, effects originating in the impurity cell or in its immediate vicinity have been ignored. The impurity cell would tend to increase the magnitude of the scattering and make it less strongly dependent on energy. This may possibly account for the observation of Debye and Conwell (*26*) that the ionized impurity mobility for electrons in Ge depends on temperature less strongly than $T^{1.5}$ as predicted by Equation (6.40).

Another essentially classical effect is neglected in the simple treatment of ionized impurity scattering. Although collisions between electrons have no direct effect on the resistivity if the energy surfaces are spherical, they can influence it indirectly by altering the momentum of the electrons between successive collisions with the lattice. When the scattering is by ionized impurities, the faster electrons acquire more momentum from the electric field, since they have longer mean free paths. Part of this acquired momentum tends to become redistributed among all the electrons between collisions, with the consequence that a fast electron does not actually acquire as much momentum from the field on the average as it would have in the absence of electron–electron collisions. The resulting decrease in electron mobility has been calculated in the classical limit by Spitzer *et. al.* (*99*) and results in ionized impurity scattering which is 60% of that given by the Brooks-Herring formula (6.40).* A similar effect occurs with any mechanism of scattering which is energy dependent, but in the case of lattice scattering it is very much weaker.†

If the energy surfaces are non spherical, or when two types of carriers are present simultaneously, scattering of carriers by carriers can contribute to the total resistance. Ordinarily this can occur only under conditions of carrier density and temperature for which ionized impurity scattering is the dominant mechanism. Electron–hole scattering has been discussed by Prince (*91*).

3. Other Types of Scattering

In addition to lattice scattering and ionized impurity scattering, there exist two types of neutral impurity scattering. It was first pointed out by Bardeen and Pearson (*100*) that neutral donors, owing to the large orbit of the bound electron, could present a rather large cross section for scattering in which the bound and free electron exchange places. A strictly

* For an elementary discussion of these effects, see reference *26*, p. 698.

† This correction reduces the theoretical impurity mobility and slightly improves agreement between theory and experiment.

analagous effect occurs in the scattering of electrons by hydrogen atoms. The effect has been studied by Erginsoy (*101*) who gives the formula

$$\frac{1}{\tau} = \frac{K\hbar^2}{me^2} \cdot \frac{20\hbar}{m} N_n \tag{6.44}$$

where the first factor is the radius of the hydrogenic orbit of the bound electron on the impurity, and N_n is the number of neutral impurities per cubic centimeter. The neutral impurity scattering is proportional to the dielectric constant, while ionized impurity scattering is inversely proportional to the square of the dielectric constant. Thus neutral impurity scattering becomes of much greater importance in high dielectric constant semiconductors. Neutral impurity scattering is independent of temperature, and the collision time is independent of energy. Neutral impurity scattering only becomes of importance at very low temperatures when $n \ll N_n$, i.e., when almost all the impurities are un-ionized. It is probably the factor which determines the line width in cyclotron resonance experiments. For ellipsoidal surfaces, the mass in the first factor of (6.44) should be replaced by the geometric mean mass $(m_1 m_2^2)^{1/3}$, and in the second factor by the mobility mass $(1/3)(m_1^{-1} + 2m_2^{-1})$. The theory has not been worked out for the valence band, but since the scattering is approximately isotropic it is probable that the intra-band and inter-band scattering are related as in the case of lattice scattering.

Another type of neutral impurity scattering can occur in Si-Ge alloys. A theory of this scattering has been worked out by Brooks (*102*) using the idea that statistical fluctuations in the composition of an alloy result in random displacements of the band edges much as in the deformation potential approach to lattice scattering. The order of magnitude of the band displacements can be guessed very roughly from the change in energy gap with composition, which has been studied by several workers (*103*). The principal predictions are (1) that in the range where alloy scattering predominates, the mobility should vary as $T^{-1/2}$, corresponding to a mean free path which is independent of carrier energy and temperature, and (2) that alloy scattering should become comparable with lattice scattering at room temperature in the composition range of 5 to 15 atom percent Si in Ge or vice versa. Little quantitative experimental work has been done to verify these relations, although a few percent of Si in Ge does appear to reduce the room-temperature mobility.

Another type of scattering is that arising from dislocations. This has been considered by W. T. Read (*104*) in its most important aspect, and earlier by Dexter and Seitz (*105*). The latter authors considered only the scattering resulting from the strain field of a dislocation, and found it was so small as to be of no importance in good quality Ge or Si. Read,

however, called attention to another type of effect which could occur in n-type Ge. We have already seen that the dislocation line can act as an electron acceptor and can thus give rise to a large space-charge cylinder in n-type Ge. This cylinder will have a different potential from its surroundings and can act as a very powerful scattering center. Electron mobilities in bent n-type Ge have been measured, and are in at least qualitative agreement with the predicted effect (*41*).

4. Other Effects of Lattice Vibrations

In Section III we discussed the shift in band edges due to the perturbing influence of the lattice vibrations. A quantitative theory of this effect has been given by Fan (*64*). This theory must be modified to take into account the ellipsoidal energy surfaces in the conduction band and the degeneracy of the valence band. For the conduction band, a theory analagous to Fan's gives the following expression for the downward shift

$$\Delta E = (m_1 m_2)^{1/2} \frac{k_0 T}{\rho u_l^2} E_1^2 \frac{2}{h^2} \left(\frac{3\pi^2}{\Omega}\right)^{1/3} \left[\frac{\sin^{-1} a}{a} + \sqrt{1 - a^2}\right] \tag{6.45}$$

where

$$a = \frac{m_1 - m_2}{m_1}$$

where m_1 and m_2 are the principal masses and Ω is the atomic volume. The theory has been worked out on the assumption that only dilatation contributes to the lattice interaction and that only acoustical type modes are involved. If we include optical modes, assuming the same interaction constant E_1 (which is of doubtful validity), the expression (6.45) would be doubled. For the valence band, in view of the crudeness of the calculation, it will suffice to use (6.45) with $m_1 = m_2$. We use the values $E_1 = 14.3$ ev for the conduction band in Ge, and $E_1 = 8.9$ ev for the valence band. The number for the valence band differs from that used in (6.28) in order to take account of the factor $\frac{1}{2}$ due to angular dependence which occurs in (6.26). From (6.45) we obtain for Ge

$$\begin{aligned} \frac{\Delta E}{\Delta T} &= 0.79 \times 10^{-4} \text{ ev/°C} \quad \text{for electrons} \\ &= 0.46 \times 10^{-4} \text{ ev/°C} \quad \text{for holes} \end{aligned} \tag{6.46}$$

Thus the total band-edge shift becomes 1.25×10^{-4} ev/°C, which is about one third the value which best fits the experiments, but is of the right order of magnitude.

It is rather easy to see, however, that the above method of calculation would tend to underestimate the shift. Thus, for example, it is assumed that the energy of an electron continues to increase parabolically away

from the band edge. Since this is not the case, as shown by Fig. 3, and since modes of high wave-number make an important contribution to ΔE, it is seen that the actual value will be substantially larger for electrons, and somewhat larger for holes. In particular, scattering into the (100) minima, which are probably not more than 0.18 ev above the (111) minima in Ge, and which have a higher effective mass, will make an important contribution.

In principle the interaction between electrons and lattice vibrations can lead to a change of curvature and hence of effective mass at the band edge. The effect is in the direction of lower effective mass (greater curvature) but can be shown to be less than 1% in practical cases.

Two other effects which involve the interaction of electrons with lattice vibrations are worth mentioning. The optical absorption process which is responsible for the absorption edge corresponding to the band gap in Ge and Si is an indirect process which involves a phonon as well as a photon (*66*). The theory of this type of transition will be discussed in Section VII (process 2 in Fig. 13 and in the text), and the predicted absorption coefficients are there shown to be in order-of-magnitude agreement with the observed values.

A second effect is the broadening of the energy levels of an impurity center due to the interaction of the localized electron with the lattice vibrations. This effect has recently been discussed by Lax and Burstein (*106*). As discussed further in Section VII these states appear to be very nearly hydrogen-like in their behavior, in spite of the complexity of the valence band edge from which they are derived. The mechanism of broadening may be described as follows. As a lattice wave passes over an acceptor center, a local distortion of the lattice occurs, and if the wavelength is long enough so that the distortion has constant phase over the dimensions of the impurity wave function, the energy level will be shifted by an amount which is equal to the local shift in the band edge by the deformation potential appropriate to that particular strain. For short wavelength vibrations the shifts induced in different parts of the impurity region cancel in phase, so that only phonons whose frequency corresponds to a temperature of less than 100°K will make an appreciable contribution to the broadening. Furthermore, most of the broadening will be in the ground level, since the effective radius of the impurity orbits increases as the principal quantum number, and the number of lattice modes of wavelength greater than the critical value thus decreases as the inverse cube of the principal quantum number of the excited state. Below about 50°K the broadening is due entirely to zero-point vibrations, and is hence temperature independent. The observation that the broadening begins to increase above about 50°K is direct experimental evidence for the idea

that it is only the long wavelength oscillations which play a significant role.

Lax and Burstein give a theoretical estimate of the width at half-absorption as 0.003 ev, which is to be compared with the observed width of 0.001 ev. In making this estimate the complexity of the valence band was neglected, and it was treated as a single spherical band with an effective mass such as to give the correct ionization energy for the ground state. The deformation potential for the lattice vibrations was taken as purely dilatational and adjusted to give the correct magnitude for the hole mobility at room temperature in Si. It was also found necessary to take into account roughly the "motional narrowing" which arises because the frequency of the important lattice vibrations is comparable with the frequency of motion of the hole in its orbit. The disagreement between the theoretical and experimental line width is possibly not surprising in view of the approximations involved. Also it is possible that some of the scattering for the mobility is contributed by optical modes, which would have little or no influence on the broadening. This idea is supported by the fact that if the deformation potential is computed from the hole mobility at liquid nitrogen temperature, the ratio of theoretical to experimental line width is brought down to 1.7.

5. Conclusions

The account set forth in Sections V and VI does not represent a fully worked out story, either from the theoretical or from the experimental view. On the theoretical side the approximation of using a collision time, and the incomplete working out of the consequences of elastic anisotropy and of anisotropy of the energy surfaces make all the quantitative conclusions regarding deformation potentials crude and tentative. The tentative conclusions which do emerge may be summarized roughly as follows:

1. Magnetoresistance data on n-type Ge and Si support the idea of (111) and (100) minima for the conduction band edges in these two materials, respectively. This conclusion, and the deduced mass anisotropies are reasonably consistent with the cyclotron resonance data.
2. In the case of n-type Ge, departures from the theoretical $T^{-1.5}$-law for lattice mobility are slight, and can probably be accounted for by inter-valley scattering.
3. The thermoelectric power provides a means of estimating the density of states near the band edges, but gives values which are about twice too large for both n and p Ge as compared with the values deduced from the cyclotron effective masses.
4. The Hall effect and magnetoresistance in p Ge can only be ac-

counted for by a model involving two types of holes, with a scattering law having a normal $\varepsilon^{-\frac{1}{2}}$-dependence in the lattice scattering range. It seems most likely that the observed $T^{-2.3}$-dependence of the hole mobility must be accounted for by an explicit temperature dependence of the effective mass for the high-mass holes, probably arising from interaction with phonons. This is also consistent with the thermoelectric data on p Ge.

5. Scattering by the shear component of the deformation potential is quite important in n-type Ge and Si, but is relatively unimportant in p-type Ge. Of the possible dilatational components of the deformation potentials which can be deduced from piezoresistance data, there is one set which is also consistent with the independently measured pressure coefficient of the band gap. The fit is best if low temperature mobilities are used.

6. For n Ge the theoretical ionized impurity scattering seems to be smaller than observed by at least a factor of 2.

7. The part of the shift in the band gap with temperature which is due to electron–lattice interaction is about 2 to 3 times larger than given by the simple theory, but the deviations can be accounted for qualitatively in terms of the detailed structure of the conduction bands.

8. The broadening of impurity levels in p Si is somewhat less than is consistent with the deformation potentials deduced from hole mobility on a simplified one-band model of the valence band.

VII. Optical Properties

It is convenient to classify the optical transitions which may occur in Ge and Si according to the general scheme shown in Fig. 13. This diagram shows schematically a series of $\varepsilon(\mathbf{P})$-curves for the conduction and valence bands, the direction of $\mathbf{P}$ being taken as that which includes the band edges, in particular the lowest point in the conduction band. Impurity levels are also indicated on the diagram as horizontal lines. Since from Equation (3.11) we have seen that any localized level may be expressed as a linear combination of band wave functions, the extent of the horizontal line for a localized state is an indication of the different values of $\mathbf{P}$ which enter appreciably into the linear combination. If the degree of localization is slight, as with the hydrogenic impurity levels near the band edges, then only a small range of $\mathbf{P}$ in the vicinity of the band edge point is needed to describe the state. On the other hand, if the localized state is a trapping center deep in the forbidden gap, then it is highly localized and requires nearly all the values of $\mathbf{P}$ in the band for an adequate description. Thus in the diagram, nonhydrogenic states are shown with great extension.

The possible transitions are indicated by arrows in the diagram. For simplicity we have considered only absorption processes. Arrows which join continuously represent two-step processes. These are processes in which a transition takes place via an intermediate state in such a way as to conserve energy in the over-all process, but not in the individual transitions. They are only of practical importance when the direct transition between initial and final states is forbidden by a selection rule (*66*, *107*, *108*).

1. *Allowed Transitions* (*Process 1*)

For allowed transitions in solids we have the selection rule

$$\mathbf{P} - \mathbf{P}' = (h/\lambda)\mathbf{s} \tag{7.1}$$

where λ is the optical wavelength and $\mathbf{s}$ is a unit vector in the direction of propagation of the electromagnetic wave. For a wavelength of 2×10^{-4} cm, the momentum on the right of (7.1) corresponds to an electronic energy of 3.5×10^{-7} ev, which is entirely negligible; hence, for practical purposes we may take the right side of (7.1) as equal to zero. Thus an allowed transition is one which can be represented by a vertical arrow in Fig. 13.

There are two cases of allowed transitions to be distinguished. The vertical transition at the point $\mathbf{P} = 0$ may be either forbidden or allowed, depending on the symmetry of the wave functions corresponding to the initial and final bands. The transition is allowed only if the wave functions are, respectively, even and odd with respect to reflection in the origin, i.e., if the initial and final states are of opposite parity. This is believed to be the case for the valence and conduction bands of Ge and Si. Under these conditions the transition probability is approximately constant as a function of $\mathbf{P}$ for vertical transitions near $\mathbf{P} = 0$. Since the total absorption at frequency ν is proportional to the total numbers of pairs of initial states having this frequency, it is easily shown that the absorption coefficient near the threshold varies according to

$$\mu_d = \frac{e^2(2m_{\text{red}})^{3/2}}{\pi n c m_e h^2} f_{if}[h(\nu - \nu_t)]^{1/2}$$
$$\frac{1}{m_{\text{red}}} = \frac{1}{m_c} + \frac{1}{m_v} \tag{7.2}$$

where m_c and m_v are the effective masses corresponding to the curvatures at the points $\mathbf{P} = 0$ of the conduction and valence bands, respectively, n is the refractive index, c is the velocity of light, ν_t is the threshold absorption frequency, ν is the actual frequency, and f_{if} is the oscillator

strength for the transition—a pure number of order unity. If the transition were a forbidden one, the selection rule would be violated away from the point $\mathbf{P} = 0$, and the transition would be roughly proportional to P^2, and hence to an additional power of $(\nu - \nu_t)$. Such a transition is

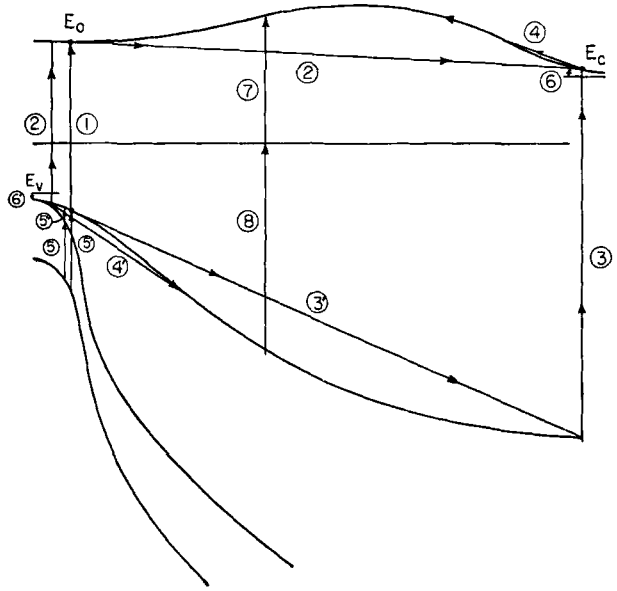

FIG. 13. Graph of energy *vs.* P for germanium, showing possible optical processes.

(1) Allowed vertical transition from the valence band to the conduction band.

(2) Indirect transition from near the edge of the valence band to near the edge of the conduction band (optical followed by phonon transition).

(3) Indirect transition from near the edge of the valence band to near the edge of the conduction band (optical transition from deep in the valence band, followed by phonon transition from the edge of the valence band into a hole).

(4) Indirect transition within the conduction band.

(4′) Indirect transition within the valence band.

(5)(5′)(5″) Direct transitions between components of the valence band.

(6) Transition from a normal donor level into the conduction band.

(7) Transition from a deep trap to the conduction band.

still considered as allowed in the sense used hereafter because it satisfies the selection rule (7.1).

A curve of absorption coefficient *vs.* $h\nu$ as determined by Dash *et al.* on very thin single crystals of Ge (*67*) is shown in Fig. 14. The plateau beyond $h\nu = 0.81$ ev is interpreted as the beginning of the allowed absorption. The absorption coefficient has been followed to above 10^5 cm^{-1} and appears to be increasing at a rate considerably faster than would be suggested by (7.2). In fact, the increase is more like the $\frac{3}{2}$-power of

$(\nu - \nu_t)$, suggesting a forbidden absorption. In any case, the minimum in the conduction band at $\mathbf{P} = 0$ appears to lie about 0.18 ev above the conduction band edge. In Si the edge for allowed absorption is more difficult to locate, but seems to be between 2.0 and 2.5 ev, and again the absorption coefficient appears to vary more nearly like $(\nu - \nu_t)^{3/2}$, although the magnitude of the absorption, as well as other evidence discussed below, strongly indicates that the transition is an allowed one.

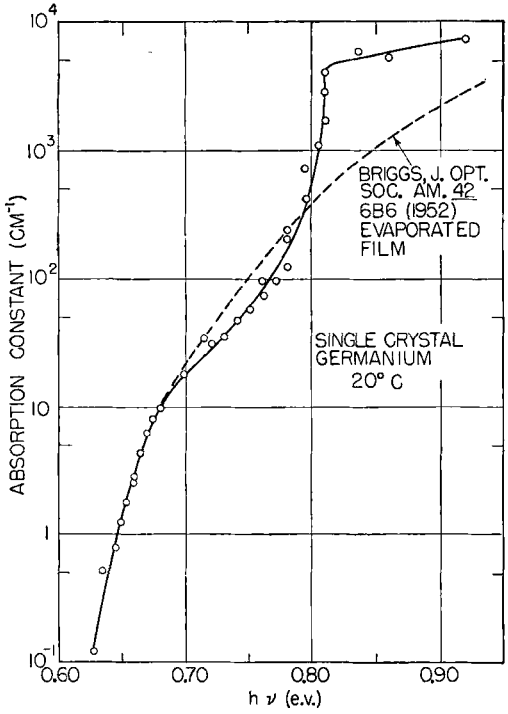

FIG. 14. Absorption coefficient of germanium as a function of frequency (W. C. Dash and R. Newman).

2. Forbidden Absorption (Processes 2 and 3)

As Fig. 14 shows, there is very strong absorption on the long wave side of the allowed threshold, and this has been interpreted by Bardeen *et al.* (*66*) in terms of a two-step process, as shown in Fig. 13. In the first step the light produces a vertical transition to a point near $\mathbf{P} = 0$ in the conduction band, and in the second step, a phonon scatters the excited electron to a position near the band edge. Another two-step process which would lead to the same final state is the process 3, in which the electron is first raised directly from the valence band to the final state in

the conduction band by an optical transition, and the resulting hole in the valence band is then scattered by a phonon to a position near the top of the valence band. The latter process is thought to be somewhat less probable than process 2.

The transition probability for any two-step quantum mechanical process is given by a formula of the general form (*108*)

$$W(i,f) = \frac{4\pi^2}{h} \left| \sum_I \frac{(i|M_1|I)(I|M_2|f)}{E_I - E_f} \right|^2 \rho(E_f) \tag{7.3}$$

where the summation is over all possible intermediate states, and where M_1 and M_2 are operators representing the constituent processes, E_I and E_f are, respectively, the energies of the intermediate and final states, and $\rho(E_f)$ is the density of final states. For the processes which we are considering, M_1 represents the operator for an optical transition and satisfies the selection rule (7.1), whereas M_2 represents the operator for the phonon which has the appropriate momentum to carry the electron from $\mathbf{P} = 0$ to one of the energy minima in the conduction band. Bardeen *et al.* (*66*) have evaluated (7.3) for the process 2 shown in Fig. 13 and obtain the formula

$$\mu_I = M_s^2(2n_s + 1) \frac{e^2(m_v m_c)^{3/2} f_{iI} \langle E_I - E_i \rangle_{AV}}{2\pi n c m_c \hbar\omega \langle (E_I - E_f)^2 \rangle_{AV}} (E_f - E_t)^2 \tag{7.4}$$

Here m_v and m_c are the "density-of-states" effective masses for valence and conduction band edges, respectively, f_{iI} is an oscillator strength, M_s^2 is a phonon matrix element squared, n_s is the number of quanta excited of the appropriate phonon wave length, and $E_f - E_t$ is the difference in energy between the actual quantum energy and the threshold energy for this type of absorption, i.e., the band gap.

Bardeen *et al.* attempt to estimate the phonon factor in (7.4) by means of the resistivity mobility. For frequencies near threshold $\langle E_I - E_i \rangle_{AV}$ cancels $\hbar\omega$ in the denominator, and $E_I - E_f = E_0 - E_C$, where E_0 is the energy at the point $\mathbf{P} = 0$, and E_C is the energy of the conduction band edge. Using $E_0 - E_C = 0.18$ ev for Ge, and 1.0 ev for Si, and assuming an allowed transition at the band center, they obtain

$$\begin{aligned} \mu_i &= 100 \text{ cm}^{-1} \quad \text{at } 0.72 \text{ ev for Ge} \\ \mu_i &= 300 \text{ cm}^{-1} \quad \text{at } 1.45 \text{ ev for Si} \end{aligned} \tag{7.5}$$

These results are to be compared with 30 cm^{-1} for Ge and 700 cm^{-1} for Si found by Dash *et al.* Considering the uncertainty in estimating the magnitude of the phonon matrix element, the theory is in quite good agreement with experiment. Actually, of course, since the change in momentum of the electron is rather large, the same phonons are not involved

in this transition as are involved in ordinary lattice scattering. If we use the deformation potential approach to estimate M_s^2 in (7.5)—a questionable procedure for short wavelength phonons—we find

$$M_s^2(2n_s + 1) \cong E_1^2 \frac{k_0 T}{\rho u_l^2} \xi \coth \xi$$
$$\xi = \frac{\hbar u_l \sigma}{2k_0 T} \tag{7.6}$$

where $\xi \coth \xi$ reduces to 1 for the phonons that participate in ordinary lattice scattering. This factor is of order unity for room temperature in Ge, but may be somewhat larger for Si. For low temperatures, however, the factor becomes large, and the phonon factor tends to become temperature independent at low temperatures.*

3. Free Carrier Absorption (Processes 4 and 4')

In the preceding paragraphs we considered processes in which the electron makes an optical transition between bands. Processes are also possible in which the electron makes an optical transition within a band and then reaches the final state via a phonon transition exactly as previously. The possibility of such processes arises essentially because the velocity operator for band states has a finite expectation value, so that nearly diagonal matrix elements for the optical transition exist. Such processes can be treated by means of Equation (7.3), and this has been done by Bardeen (*110, 111*) and by Rosenberg and Lax (*112*). However, somewhat similar results may be obtained by a semiclassical approach. In fact the absorption coefficient may be obtained from the high-frequency conductivity given in Equation (5.37).

The observed absorption in the forbidden gap has been studied by a number of workers (*114*). It is, of course, much weaker than the inter-band absorption, because the initial states are only very sparsely occupied compared with the initial states in the valence band for inter-band absorption. The actual transition probabilities from a given initial state are of the same order of magnitude as for the forbidden inter-band absorption discussed above (process 2). It is found experimentally that the absorption is proportional to the number of carriers, and that in intrinsic material the contributions of the electrons and holes are additive. Furthermore, the absorption varies about as the square of the wavelength, as is required by (5.37). In n-type Ge the magnitude of the absorption coefficient is many times that predicted for free electrons of the appropriate dc mobility. This is accounted for by the factor $\gamma(p)/m_{\text{eff}}^2$ in (5.37), which

* This discussion ignores the effect of the phonon energy on the energy gap. For a more complete treatment see reference *109*.

is about 72 for Ge (*115*). Rosenberg and Lax (*112*) have shown that additional absorption could be accounted for by inter-valley phonon induced transitions, which are not as severely limited by conservation of energy considerations as in the dc case, owing to the energy supplied by the

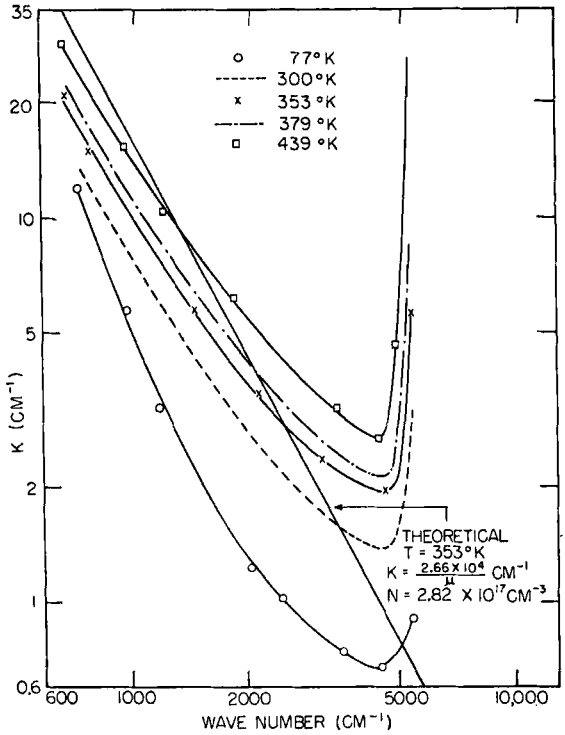

FIG. 15. Free electron absorption in n-germanium. From Seventh Quarterly Report, Purdue University, Project DA 36-039-SC-15339, fig. 15, p. 59.

photon. Another possibility is that for the photon energies involved account must be taken of the changes in the curvature of the energy surfaces away from the minimum at the band edge. One would expect, however, that such an effect would lead to a slower variation of absorption with wavelength than λ^2, whereas actually the λ^2-law seems to be well satisfied for *n*-type Ge and Si.

It is also predicted by (5.37) that the infrared absorption should vary inversely with the dc mobility. This relation has been tested experimentally by Fan and co-workers (*116*), and is found to be fairly well obeyed in the lattice scattering regime, but the behavior in the impurity scattering regime is not understood. Typical curves for *n*-type Ge are shown in Fig. 15.

4. Free Hole Absorption (Processes 5, 5', and 5'')

Another type of transition which can occur involves an allowed type of optical jump between different branches of the same degenerate band structure, as indicated by process 5 in Fig. 13. Although this is really a special case of process 1, it is much weaker and is proportional to the hole density near the top of the valence band. Two factors account for the weakness of the transition. First, the transition is forbidden at $\mathbf{P} = 0$, and the transition probability is proportional to P^2 and therefore to

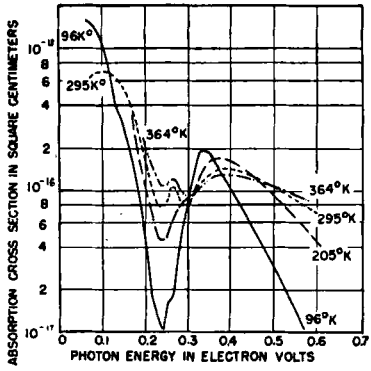

FIG. 16. Optical absorption in p-type Ge, from Briggs and Fletcher, reference *117*.

$(\mathcal{E}_v - \mathcal{E})$ for states near, but away from, the band edge. This is an inevitable feature of transitions within a degenerate band structure, because the different degenerate branches have the same parity at $\mathbf{P} = 0$. Secondly, and more important, a transition can only occur, consistent with the exclusion principle, if the final state is empty, so that the total absorption is proportional to the hole density in the upper state of the transition pair.

All this leads to an absorption which behaves like free hole absorption in that it is proportional to the number of conducting holes, but which does not show the characteristic λ^2 dependence of the electronic absorption, and is also about 10 times stronger. In fact, as was shown by Briggs and Fletcher (*117*), and as is illustrated in Fig. 16, the free hole absorption in Ge shows considerable structure, which is, moreover, very temperature dependent. The structure has been explained quite well by Kahn (*115*), using the model of a three-fold degenerate valence band, split by spin-orbit effects into an upper doubly degenerate and a lower singly degenerate branch. Kahn found that the data were consistent with the effective masses obtained from cyclotron data, and the short wavelength peak could be properly explained if the spin-orbit splitting at

$\mathbf{P} = 0$ were taken to give a separation of 0.28 ev. The transition probabilities can be computed completely in terms of the effective mass parameters, so that in principle there are no undetermined constants in the calculation other than the spin-orbit splitting. However, this calculation program has not been completely carried out. Also, there are some minor discrepancies between theory and experiment which are difficult to account for. In particular the absorption at long wavelengths appears to be less than predicted by theory.

Similar measurements on free hole absorption in Si have revealed no evidence of structure out to 12 microns (*118*), and furthermore the λ^2 relation is fairly well fulfilled. Although the spin orbit splitting is probably less than 0.05 ev for Si, and therefore one would expect the structure to appear at much longer wavelengths than in Ge, there is rather surprisingly no evidence of structure at all.

TABLE IX. Comparison of Optical and Thermal Ionization Energies for Donors and Acceptors in Silicon

Impurity	Thermal	Optical
B Acceptor	0.045 (*23*)[a]	0.046 (*20*)
Al "	0.057 (*23*)	0.067[b]
Ga "	0.065 (*23*)	0.071[b]
In "	0.16 (*23*)	0.16[b]
P Donor	0.044 (*23*)	0.046[b]
As "	0.049 (*23*)	0.056[b]

[a] Numbers in parentheses are reference numbers.

[b] Burstein, Henvis, Picus, and Shulman, unpublished results quoted in reference *30*.

5. *Absorption by Group III and V Impurities (Processes 6 and 6')*

Direct transitions between hydrogenic impurities and the appropriate conduction bands were first observed in *p*-type Si by Burstein and coworkers (*20*). This observation can only be made at low temperatures, where the acceptors are appreciably de-ionized. In this way it was possible to measure the ionization energy of acceptors quite accurately and to compare it with that determined from Hall and resistivity data at low temperatures. For boron the agreement was excellent. A more complete comparison is shown in Table IX, which also shows results for *n*-type Si obtained by the same authors. Such discrepancies as appear are probably experimental and have no theoretical significance. The optical data may be considered the more reliable. There are no optical results for Ge because of the longer wavelengths involved and attendant experimental difficulties.

An interesting biproduct of these results was the observation of transitions to excited states of the impurity centers. The positions of the excited states seem to give a rather striking confirmation of the hydrogenic model. This is shown in Table X (*20*).

TABLE X. Comparison of Optical Data on p-Type Silicon with the Hydrogenic Model

	Hydrogen Model[a]	B	Al	Ga	In[b]
$1s$	0.0460 ev	0.0460 ev	0.067	0.071	0.156
$2p$	0.0115	0.0116	0.012	0.013	0.019
$3p$	0.0051	0.0062	0.007	0.009	0.010
$4p$	0.0029	0.0029	0.003	0.003	0.004

[a] Spherical energy surfaces with $m_{eff} = 0.45m_e$.

[b] Note that energy of excited states is less reliable in this case. Data taken from thermal ionization energy.

Similar results have been obtained for donor levels. It is unfortunate that there is not a good theory with which to compare the experiments on p-type silicon; in fact, the simple hydrogenic model appears to work surprisingly well except for the ground state, despite the fact that here we have to deal with a degenerate band structure in which, furthermore, the spin-orbit splitting is insufficient to minimize the influence of the split-off band. Theoretical transition probabilities have also been computed on the hydrogenic model with spherical symmetry. The observed transition probabilities are in rough agreement with theory, except that the $1s$–$2p$ transition is relatively very much weaker than it should be, by a factor of the order of 30 (*20*).

The structure of the absorption corresponding to transitions from the acceptor level to the valence band is also of considerable interest. A theory for this for Ge has been given by Teitler, Burstein, and Lax (*119*). Since the acceptor level is comprised of band wave functions corresponding to values of $\mathbf{P}$ very close to the band edge, where $\mathbf{P} = 0$, the structure of this absorption, and the theory for it, is quite similar to that developed by Kahn for the inter-band transitions (*115*), except that the momentum distribution of holes in the final state is determined by the nature of the impurity level rather than by a Maxwell distribution.

6. Absorption by Other Types of Impurities (Processes 7 and 8)

Transitions may take place directly from deep-lying trap levels to either the valence or conduction bands. Because of the high degree of localization in the trap level, the selection rule for allowed inter-band

transitions has no relevance in this case. Also the solubilities of gold and the iron-group transition elements, which produce deep lying levels in Ge, is so small that absorption cannot be observed directly. The transitions are rather observed through the photoconductivity produced, that is, from the conductivity resulting from the excited electron or hole (*31*, *50*).

The process is illustrated in Fig. 17 for germanium doped with iron (*50*). The photoconductive response is the change in resistance of the

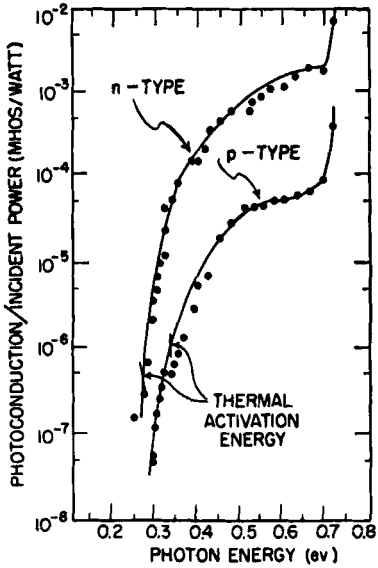

FIG. 17. Photoconductive response of iron-doped germanium, illustrating correspondence between optical and thermal thresholds, from W. W. Tyler, R. Newman, and H. H. Woodbury, *Phys Rev.* **98**, 882 (figs. 1 and 2) (1954).

sample per watt of incident light energy, a meaningful quantity, since the illuminated samples obey Ohm's law and show a response which is proportional to the intensity. The position of the thermal activation energy, as determined from resistivity data, is indicated on the response curves. In *n*-type material at liquid N_2 temperature, the upper Fe level is partially occupied and the photoconduction originates in the excitation of an electron into the conduction band. The threshold energy for such a process should and does correspond roughly to the thermal "ionization energy" of this level, although response is so broad that an accurate comparison is not possible. In *p*-type material, the photoconduction is due to excitation of an electron from the valence band into an empty lower iron level, and again the threshold is roughly correct.

There is an interesting contrast in the behavior of n and p photoconduction, however. As seen from the figure, the n-type material is much more sensitive; also it shows very much longer recovery times after the illumination is removed. This behavior may be possibly correlated with the fact that in n-type material it is difficult for electrons to retrap in the upper $1e$ iron level because the center is already negatively charged and hence possesses a "potential barrier" for electrons. No such situation occurs in p-type material, since there is no charge barrier against recapture of holes. It is found that the strong photoconductivity may also be "quenched" by simultaneous illumination in the proper wavelength band. All these effects occur at liquid N_2 temperature, where both the Fe levels are filled. This is believed to be associated with the saturation of minority carrier (hole) traps, but a detailed explanation is not available. At room temperature, where the upper level is only partially occupied, the principal effect of the Fe is to reduce the recombination lifetime for carriers injected by light and hence to reduce the photosensitivity. Under these circumstances the deep Fe level is thought to act as a recombination trap by first capturing a hole and then an electron. Unfortunately, although most of these effects may be explained qualitatively in terms of a model involving singly and doubly charged acceptor states for the iron, it has so far not been possible to use such a simple model for quantitative interpretation of the photoconductivity results.*

7. Other Optical Processes

So far the discussion has concerned itself entirely with absorption. It is clear, however, that whenever absorption is possible there must exist a converse emission process. Often the emission is difficult to observe because competing processes make it very weak. So far evidence has been found for three types of emission processes:

1. Direct recombination of holes and electrons with emission of a single photon, momentum being conserved by lattice vibrations. This effect has been observed by Haynes and Briggs (*48*) and by Newman (*120*). It is essentially the inverse of process 2 in Fig. 13, and hence the probability and spectral distribution of the emission can be predicted from absorption data with the use of the principle of detailed balancing. This has been done by van Roosebroeck and Shockley (*48*). They show that the spectral distribution should be extremely sharply

* Since the above paragraph was written, substantial progress has been made by Tyler and his collaborators. In particular they have succeeded in explaining most of the photoconduction and quenching properties of Mn doped Ge in terms of the two known levels introduced by the Mn impurity.

peaked at the long-wave limit of the absorption curve, and that the recombination rate at 300°K corresponds to a lifetime for holes of 0.22 sec. This is much longer than observed life-times, which are thought to arise from a two-step process (*47*). This value, however, seems to be in accord with the experimental results (*120*) as does the spectral distribution.

2. Recombination in Ge via a deep trap at about 0.2 ev above the valence band.

3. Recombination radiation observed (*121*) when a Si *p-n* junction is operated at very high back voltage beyond avalanche break-down (*122*). Since, in the high field, both electrons and holes will acquire sufficient energy from the field to produce secondaries, it is energetically possible for radiation to be emitted which is three times the forbidden gap, or about 3.5 ev. Nearly constant radiation output is actually observed (*121*) out to 2.4 ev, after which it falls off but is still measurable out to 3.4 ev. This explanation, however, is not established and involves some theoretical difficulties.

In this review we have been mainly concerned with electronic processes in semiconductors and have therefore confined our discussion to optically induced electronic transitions. It should be mentioned, however, that absorption in the far infra red has been found which appears to be nearly independent of carrier concentration or type and is accordingly ascribed to lattice vibrations. Since germanium and silicon are valence crystals, their lattice vibrations do not interact strongly with an external field, and hence the resulting absorption is quite weak, comparable in strength to the absorption by impurities at low temperatures (*123*).

Acknowledgments

This review has benefited from numberless conversations and discussions with individuals. I should like to thank especially Dr. B. Lax for discussions of the cyclotron resonance work, Dr. C. Herring for several discussions and for a prepublication copy of his manuscript, Dr. W. C. Dash for permission to use his data in advance of publication, Dr. E. Kane, Dr. H. Ehrenreich, and Dr. A. W. Overhauser for several interesting discussions of valence band structure and hole scattering and Dr. G. Benedek and Dr. D. Warschauer for many suggestions. Finally, I should like to thank Dr. W. Paul for reading the manuscript in its entirety, and for the correction of many errors, omissions, and obscurities.

REFERENCES

1. Hartree, D. R., *Proc. Cambridge Phil. Soc.* **24,** 89 (1928); for a good summary discussion of the quantum mechanical many body problem, especially as applied to solids, see F. Seitz, "Modern Theory of Solids," especially Chapter 6, also pp. 334–344. McGraw-Hill, New York, 1940.

2. Brouckaert, L. P., Wigner, E. P., and Smoluchowski, R., *Phys. Rev.* **50,** 58 (1936); Herring, C., *ibid.* **52,** 365 (1937); Manning, M. F., and Chodorow, M. I., *ibid.* **56,** 787 (1939), especially discussion in fine print at end of paper.
3. Taken from Adams, E. N., II, report CML-TN-P8, September 1954, Chicago Midway Laboratories.
4. Herman, F., and Callaway, J., *Phys. Rev.* **89,** 518 (1953); Herman, F., *ibid.* **93,** 1214 (1954); **95,** 847 (1954).
5. Dresselhaus, G., Kip, A. F., and Kittel, C., *Phys. Rev.* **95,** 568 (1954); **98,** 368 (1955).
5a. Elliot, R. J., *Phys. Rev.* **96,** 266, 280 (1954).
6. See references in reference *4*; Parmenter, R., unpublished; Herman, F., *Physica* **20,** 801 (1954); cf. also reference *5*.
7. Cf., for example, Shockley, W., "Electrons and Holes in Semiconductors," especially Chapters 6, 7, and 15. van Nostrand, New York, 1950.
8. Haynes, J. R., and Shockley, W., *Phys. Rev.* **75,** 691 (1949); reference *7*, p. 56.
9. Brown, S., and Barnett, S. J., *Phys. Rev.* **87,** 601 (1952).
10. For further details see Shockley, W., *Phys. Rev.* **88,** 953 (1952); Rostoker, N., *ibid.* **88,** 952 (1952).
11. Dresselhaus, G., Kip, A. F., and Kittel, C., *Phys. Rev.* **92,** 827 (1953); Lax, B., Zeiger, H. J., Dexter, R. N., and Rosenblum, E. S., *Phys. Rev.* **93,** 1418 (1954); Dexter, R. N., Zeiger, H. J., and Lax, B., *ibid.* **95,** 557 (1954); Dexter, R. N., and Lax, B., *ibid.* **96,** 223 (1954); Shockley, W., *ibid.* **90,** 491 (1953); Dexter, R. N., Lax, B., Kip, A. F., and Dresselhaus, G., *Phys. Rev.* **96,** 222 (1954); Lax, B., Zeiger, H. J., and Dexter, R. N., *Physica* **20,** 818 (1954).
12. Kip, A. F., *Physica* **20,** 818 (1954); Dresselhaus, G., Kip, A. F., and Kittel, C., *Phys. Rev.* **98,** 368 (1955).
13. Shockley, W., *Phys. Rev.* **79,** 191 (1950).
14. Cf., however, Luttinger, J. M., and Kohn, W., *Phys. Rev.* **97,** 863 (1955).
15. Lax, B., private communication.
16. James, H. M., *Phys. Rev.* **76,** 1611 (1949); Slater, J. C., *ibid.* **76,** 1592 (1949); Saxon, D. S., and Hutner, R. A., *Philips Research Rept.* **4,** 81 (1949); Koster, G. F., and Slater, J. C., *Phys. Rev.* **95,** 1167 (1954).
17. Bethe, H., M.I.T. Radiation Lab. Report No. 43-12 (1942).
18. Kittel, C. H., and Mitchell, A. H., *Phys. Rev.* **96,** 1488 (1954); Luttinger, J. M., and Kohn, W., *Phys. Rev.* **96,** 802 (1954); Lampert, M. A., *ibid.* **97,** 352, 869 (1955); Luttinger, J. M., and Kohn, W., *Phys. Rev.* **97,** 863 (1955).
19. Kohn, W., and Luttinger, J. M., *Phys. Rev.* **97,** 1721 (1955); Kleiner, W. H., *ibid.* **97,** 1722 (1955).
20. Burstein, E., Bell, E. E., Davisson, J. W., and Lax, M., *J. Phys. Chem.* **57,** 849 (1953); Burstein, E., Oberly, J. J., Davisson, J. W., and Henvis, B. W., *Phys. Rev.* **82,** 764 (1951); cf. also reference *107*.
21. Brooks, H., and Fletcher, N., unpublished.
21a. Kohn, W., and Schecter, D., *Phys. Rev.* **99,** 1903 (1955).
22. Geballe, T. H., and Morin, F. J., *Phys. Rev.* **95,** 1085 (1954).
23. Morin, F. J., Maita, J. P., Shulman, R., and Hannay, N. B., *Bull. Am. Phys. Soc.* **29,** 22 (1954).
24. Debye, P. P., and Conwell, E. M., *Phys. Rev.* **87,** 1131 (1952).
25. Brooks, H., unpublished calculations; presented at ASM Semiconductor Symposium, February 1954.
26. Debye, P. P., and Conwell, E. M., *Phys. Rev.* **93,** 693, 704 (1954).
27. Castellan, C. W., and Seitz, F., *in* "Semi-conducting Materials" (H. K. Henisch,

ed.), pp. 8–25. Academic Press, New York, 1951; cf. also discussion in reference *26*; Erginsoy, C., *Phys. Rev.* **80,** 1104 (1950); Erginsoy, C., *ibid.* **88,** 893 (1952).

28. Hung, C. S., and Gliessman, J. R., *Phys. Rev.* **79,** 726 (1950); Hung, C. S., *ibid.* **79,** 727 (1950); Hung, C. S., and Gliessman, J. R., *ibid.* **96,** 1226 (1954).

29. James, H. M., and Ginzbarg, A. S., *J. Phys. Chem.* **57,** 840 (1953); Landauer, R., and Hellund, J. C., *J. Chem. Phys.* **22,** 1665 (1954); Aigrain, P., *Physica* **20,** 978 (1954).

30. For a comprehensive summary of this work, see Burstein, E., Picus, G. S., and Sclar, N., "Optical and Photoconductive Properties of Silicon and Germanium," ONR Photoconductivity Conference, Nov., 1954, to be published by Wiley, New York.

31. Dunlap, W. C., Jr., *Phys. Rev.* **91,** 1282 (1953); Tyler, W. W., Woodbury, H. H., and Newman, R., *ibid.* **94,** 1419 (1954); Tyler, W. W., Newman, R., and Woodbury, H. H., *ibid.* **97,** 669 (1955); Woodbury, H. H., and Tyler, W. W., *Bull. Am. Phys. Soc.* **30,** 11 (1955), abstract F1; Tyler, W. W., and Woodbury, H. H., *Phys. Rev.* **96,** 874 (1954); Tyler, W. W., *ibid.* **96,** 226 (1954).

32. Dunlap, W. C., Jr., *Phys. Rev.* **97,** 614 (1955); Dunlap, W. C., Jr., *Bull. Am. Phys. Soc.* **30,** 12 (1955), abstract F2.

33. Taft, E. P., and Horn, F. H., *Phys. Rev.* **93,** 64 (1954).

34. Haynes, J. R., and Hornbeck, J. A., *Phys. Rev.* **90,** 152 (1953); Hornbeck, J. A., and Haynes, J. R., *ibid.* **97,** 311 (1955).

35. Paul, W., unpublished measurements. Samples supplied by Dunlap, W. C., Jr.

36. Gallagher, C. J., *Phys. Rev.* **92,** 846 (1953); Gallagher, C. J., and Tweet, A. G., *ibid.* **96,** 834 (1954).

37. Bardeen, J., *Phys. Rev.* **71,** 717 (1947); Bardeen, J., and Morrison, S. R., *Physica* **20,** 873 (1954).

38. Taylor, W. E., Odell, N. H., and Fan, H. Y., *Phys. Rev.* **88,** 867 (1952); Pearson, G. L., *ibid.* **76,** 459 (1949); Taylor, W. E., and Fan, H. Y., *ibid.* **78,** 335 (1950); Odell, N. H., and Fan, H. Y., *ibid.* **78,** 334 (1950); Tweet, A. G., *ibid.* **96,** 828 (1954); Mueller, R. K., *Physica* **20,** 1053 (1954).

39. Mayburg, S., and Rotundi, L., *Phys. Rev.* **91,** 1015 (1953); **95,** 38 (1954); Theurer, H. C., and Scaff, J. H., *J. Metals* **18,** 59 (1951); Fuller, C. S., Theurer, H. C., and Van Roosbroek, W., *Phys. Rev.* **85,** 678 (1952); Esaki, L., *ibid.* **89,** 1026 (1953).

40. Lark-Horovitz, K., *in* "The Present State of Physics," pp. 57–127. American Association for the Advancement of Science, Washington, D.C., 1953; Electron Bombardment: Brown, W. L., Fletcher, R. C., and Wright, K. A., *Phys. Rev.* **92,** 591 (1953); Klontz, E. E., and Lark-Horovitz, K., *ibid.* **86,** 643 (1952); Alpha Bombardment: Brattain, W. H., and Pearson, G. L., *ibid.* **80,** 846 (1950); Neutron Bombardment: Lark-Horovitz, K., *in* "Semi-conducting Materials" (H. K. Henisch, ed.), pp. 47–48. Academic Press, New York, 1951; Cleland, J. W., Crawford, J. H. Jr., Lark-Horovitz, K., Pigg, J. C., and Young, F. W., Jr., *Phys. Rev.* **83,** 312 (1951); **84,** 961 (1951); James, H. M., and Lark-Horovitz, K., *Z. physik. Chem.* **198,** 107 (1951).

41. Pearson, G. L., Read, W. T., Jr., and Morin, F. J., *Phys. Rev.* **93,** 666 (1954).

42. Seitz, F., *Advances in Phys.* **1,** 43 (1952).

43. Tweet, A. G., to be published in *Phys. Rev.*

44. Shockley, W., *Phys. Rev.* **91,** 228A (1953); Read, W. T., Jr., *Phil. Mag.* **45,** 775 (1954); **45,** 1119 (1954).

45. Shockley, W., *Phys. Rev.* **56,** 317 (1939).

46. Hall, R. N., *Phys. Rev.* **87,** 387 (1952).

47. Shockley, W., and Read, W. T., Jr., *Phys. Rev.* **87,** 835 (1952).
48. Van Roosebroek, W., and Shockley, W., *Phys. Rev.* **94,** 1558 (1954); Haynes, J. R., and Briggs, H. B., *ibid.* **86,** 647 (1952).
49. Burton, J. A., Hull, G. W., Morin, F. J., and Severiens, J. C., *J. Phys. Chem.* **57,** 853 (1953).
50. Tyler, W. W., Newman, R., and Woodbury, H., *Phys. Rev.* **96,** 882 (1954).
51. Kulin, S. S., and Kurtz, A. D., *Acta Metallurgica* **2,** 354 (1954).
52. Fan, H. Y., Navon, D., and Gebbie, H. A., *Physica* **20,** 855 (1954); also reference *49.*
53. Aigrain, D., *Physica* **20,** 1010 (1954); ONR Photoconductivity Conference, Nov., 1954.
54. Lax, M., "Influence of Lattice Vibrations on Electronic Transitions in Solids," ONR Conference on Photoconductivity, Nov., 1954, to be published by Wiley, New York.
55. Wannier, G. H., *Phys. Rev.* **91,** 207A (1953).
56. Shockley, W., "Electrons and Holes in Semiconductors," Chapters 10 and 16. van Nostrand, New York, 1950; Bragg, J. K., and Hebb, M. H., General Electric Research Lab. Report RL-468, Dec., 1950; Hutner, R. A., Rittner, E. S., and DuPre, F. K., *Philips Research Rep.* **5,** 188 (1950); Lark-Horovitz, K., *in* "The Present State of Physics," pp. 57–127. American Association for the Advancement of Science, Washington, D.C., 1953. For degenerate case, see especially Lark-Horovitz, K., and Johnson, V. A., *Phys. Rev.* **69,** 258 (1946); Johnson, V. A., and Lark-Horovitz, K., *ibid.* **71,** 374, 909 (1947); Finlayson, D. M., Johnson, V. A., and Shipley, F. M., *ibid.* **87,** 1141 (1952).
57. Rushbrooke, G., *Trans. Faraday Soc.* **36,** 1055 (1940).
58. Landsberg, P. T., to be published.
59. James, H. M., "Note on the Statistical Mechanics of Semiconductor Crystals with Electron-Vibration Coupling," Photoconductivity Conference, Nov., 1954, to be published by Wiley, New York.
60. Conwell, E. M., *Proc. I.R.E.* **11,** 1327 (1952).
61. Morin, F. J., and Maita, J. P., *Phys. Rev.* **96,** 28 (1954); Prince, M. B., *ibid.* **93,** 1204 (1953).
62. Paul, W., and Brooks, H., *Phys. Rev.* **94,** 1128 (1954).
63. Miller, P., and Taylor, J., *Phys. Rev.* **76,** 179 (1949); Bridgman, P. W., *Proc. Am. Acad. Arts Sci.* **79,** 139 (1951); **82,** 71 (1953).
64. Fan, H. Y., *Phys. Rev.* **82,** 900 (1951).
65. Paul, W., and Pearson, G. L., *Phys. Rev.* **98,** 1755 (1955).
66. Hall, L. H., Bardeen, J., and Blatt, F. J., *Phys. Rev.* **95,** 559 (1954); Bardeen, J., Blatt, F. J., and Hall, L. H., "Indirect Transitions from the Valence to the Conduction Bands," ONR Photoconductivity Conference, Nov., 1954, to be published by Wiley, New York.
67. Dash, W. C., Newman, R., and Taft, E. P., *Bull. Am. Phys. Soc.* **30,** 1, 53 (1955); Dash, W. C., and Newman, R., submitted to *Phys. Rev.*
68. Gurevich, L., *J. Phys.* (*U.S.S.R.*) **9,** 477 (1945); **10,** 67 (1946); Frederikse, H. P. R., *Phys. Rev.* **91,** 491 (1953); **92,** 248 (1953).
69. Warschauer, D., and Paul, W., discussion of paper by Fan, H. Y., Shepherd, M. L., and Spitzer, W., Photoconductivity Conference, Nov., 1954, to be published by Wiley, New York; Warschauer, D., Paul, W., and Brooks, H., *Bull. Am. Phys. Soc.* **30,** 54 (1955), abstract V10.
70. De Boer, J. H., and van Geel, W. C., *Physica* **2,** 186 (1935).
71. Brooks, H., unpublished. A more approximate account and detailed comparison with experiment are given by Dunlap, W. C., Jr., *Phys. Rev.* **97,** 614 (1955).

72. Landsberg, P. T., unpublished.
73. Cf. for example, Whittaker, E. T., and Watson, G. N., "Modern Analysis," 4th ed., Chapter 12. Cambridge Univ. Press, New York, 1935.
74. Jones, H., *Phys. Rev.* **81,** 149 (1951).
75. Johnson, V. A., and Lark-Horovitz, K., *Phys. Rev.* **82,** 977 (1951).
76. Johnson, V. A., and Whitesell, W. J., *Phys. Rev.* **89,** 941 (1953).
77. Willardson, R. W., Harman, T. C., and Beer, A. C., *Phys. Rev.* **96,** 1512 (1954).
78. Meiboom, S., and Abeles, B., *Phys. Rev.* **93,** 1121 (1954); Abeles, B., and Meiboom, S., *ibid.* **95,** 31 (1954).
79. Shockley, W., *Phys. Rev.* **79,** 191 (1950).
80. Seitz, F., *Phys. Rev.* **79,** 372 (1950).
81. Wilson, A. H., "The Theory of Metals," 2nd ed., Chapter 8. Cambridge Univ. Press, New York, 1953; also reference *87.*
82. Herring, C., *Phys. Rev.* **96,** 1163 (1954).
83. Gurevich, L., *J. Phys.* (*U.S.S.R.*) **9,** 477 (1945); **10,** 67 (1946); Frederikse, H. P. R., *Phys. Rev.* **91,** 491 (1953); **92,** 248 (1953).
84. Benedek, G., Paul, W., and Brooks, H., *Bull. Am. Phys. Soc.* **30,** 54 (1955), abstract V10.
85. Pearson, G. L., and Suhl, H., *Phys. Rev.* **83,** 768 (1951).
86. Morin, F. J., and Maita, J. P., *Phys. Rev.* **94,** 1525 (1954).
87. Morin, F. J., *Phys. Rev.* **93,** 62 (1954).
88. Brooks, H., *Phys. Rev.* **83,** 879 (1951); Conwell, E., and Weisskopf, V. F., *ibid.* **77,** 388 (1950); see also discussion in section 6.
89. Pearson, G. L., and Herring, C., *Physica* **20,** 975 (1954).
90. Geballe, T. H., and Hull, G. W., *Phys. Rev.* **94,** 1134 (1954).
91. Prince, M. B., *Phys. Rev.* **92,** 681 (1953).
92. Herring, C., *Bell System Tech. J.* **34,** No. 2 (1955).
93. Shockley, W., and Bardeen, J., *Phys. Rev.* **77,** 407 (1949).
94. Shockley, W., and Bardeen, J., *Phys. Rev.* **80,** 72 (1950).
95. Smith, C. S., *Phys. Rev.* **94,** 42 (1954); theory given in reference *92.*
96. Dumke, W. P., *Phys. Rev.* **98,** 230A (1955).
97. Lawrence, R., *Phys. Rev.* **89,** 1295 (1953).
98. James, H. M., and Lehman, G. W., "Potential Fluctuations in Homogeneous Semiconductors." Purdue Report, June, 1950.
99. Spitzer, L., Jr., and Harm, R., *Phys. Rev.* **89,** 977 (1953); Cohen, R. S., Spitzer, L., Jr., and Routly, P. M., *ibid.* **80,** 230 (1950).
100. Bardeen, J., and Pearson, G. L., *Phys. Rev.* **75,** 865 (1949).
101. Erginsoy, C., *Phys. Rev.* **79,** 1013 (1950).
102. Brooks, H., unpublished (1952).
103. Levitas, A., Wang, C. C., and Alexander, B. H., *Phys. Rev.* **95,** 846 (1954); Johnson, E. R., and Christian, S. M., *ibid.* **95,** 560 (1954).
104. Read, W. T., Jr., to be published.
105. Dexter, D. L., and Seitz, F., *Phys. Rev.* **86,** 964 (1952).
106. Lax, M., and Burstein, E., *Phys. Rev.* **91,** 208A (1953); Technical Report No. 8, ONR Contract Non-669 (04), proj. NR 017-419.
107. For an excellent review of the optical properties of Ge and Si, see chapter by Burstein, E., Picus, G., and Sclar, N., "Optical and Photoconductive Properties of Silicon and Germanium," ONR Photoconductivity Conference Nov., 1954, to be published by Wiley, New York.
108. For a discussion of two-step processes see Heitler, W., "Quantum Theory of Radiation," 3rd ed., pp. 140-141. Oxford Univ. Press, New York, 1954.
109. Macfarlane, G. G., and Roberts, V., *Phys. Rev.* **97,** 1714 (1955).

110. Bardeen, J., *Phys. Rev.* **79,** 216 (1950); cf. also reference *114.*
111. Cf. also Fan, H. Y., and Becker, M., *in* "Semi-conducting Materials" (H. K. Henisch, ed.), pp. 132–147. Academic Press, New York, 1951.
112. Rosenberg, R., and Lax, M., to be published.
113. Bloch, F., *Z. Physik* **52,** 55 (1928); **59,** 208 (1930).
114. Briggs, H. B., *Phys. Rev.* **77,** 727 (1950); Collins, R. J., and Fan, H. Y., *ibid.* **91,** 230 (1953).
115. Kahn, A., Thesis, University of California (1954); *Phys. Rev.* **97,** 1647 (1955); Teitler, R., Burstein, E., and Lax, M., to be published.
116. Spitzer, W., Collins, R. J., and Fan, H. Y., *Bull. Am. Phys. Soc.* **30,** 13 (1955), abstract F7; Collins, R. J., and Fan, H. Y., Contract DA36-039-15339, Seventh Progress Report, March 1953.
117. Briggs, H. B., and Fletcher, R. C., *Phys. Rev.* **87,** 1130 (1952); **91,** 1342 (1953).
118. Fan, H. Y., Shepherd, M. L., and Spitzer, W., "Infra-red Absorption and Energy Band Structure of Germanium and Silicon," ONR Photoconductivity Conference, Nov., 1954, to be published by Wiley.
119. Teitler, R., Burstein, E., and Lax, M., to be published, results quoted in reference *104.*
120. Newman, R., *Phys. Rev.* **91,** 1313 (1953).
121. Newman, R., *Bull. Am. Phys. Soc.* **30,** 13 (1955), abstract F13; Newman, R., G. E. Research Laboratory Report No. 55-RL-1348.
122. McKay, K. G., *Phys. Rev.* **94,** 877 (1954).
123. Lax, M., and Burstein, E., *Phys. Rev.* **97,** 39 (1955).