Shallow Impurity States in Silicon and Germanium

W. KOHN

Carnegie Institute of Technology, Pittsburgh, Pennsylvania

I. Introduction

The great importance of impurity states in determining the electrical properties of semiconductors has been recognized for a long time.[1] There are two varieties of these states, the so-called donor and acceptor states. In a donor state an electron is localized around an impurity from which

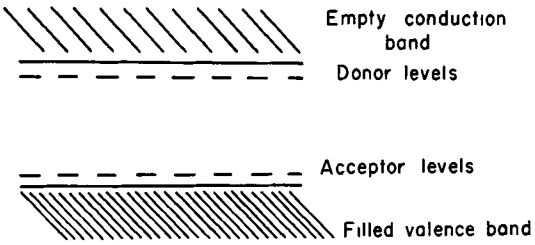

FIG. 1. Donor and acceptor levels in the band gap of a semiconductor (schematic).

it can be released into the conduction band with a rather small expenditure of energy. In this way the impurity atom (donor) can make an electron available for electrical conduction (Fig. 1). Similarly, an acceptor state is an *available* electronic state around an impurity which can accept an electron from the valence band. When this transition occurs an electron is taken from the valence band, or a hole is produced (Fig. 1). An electrical current can now be carried by the valence band. Since the energies involved in these processes are of the order of 0.01 to 0.1 ev, and hence are in general much smaller than the energy required to excite an electron directly from the valence band into the conduction band (of the order of 1 ev), the electrons and holes which are thermally excited from impurity states generally dominate the electrical properties of semicon-

[1] N. F. Mott and R. W. Gurney, "Electronic Procession in Ionic Crystals," p. 166. Oxford Univ. Press, London and New York, 1940. G. L. Pearson and J. Bardeen, *Phys. Rev.* **75**, 865 (1948). Previous reviews dealing with impurities in silicon and germanium are: J. A. Burton, *Physica* **20**, 845 (1954) and E. Burstein, G. S. Picus, and N. Sclar, *in* "Photoconductivity Conference held at Atlantic City, Nov. 4–6" (R. G. Breckenridge, ed.), p. 353. Wiley, New York, 1954.

ductors at all but the highest temperatures. Quite apart from their role as sources of charge carriers, however, impurity states have become the object of both experimental and theoretical studies in their own right, and their properties have been mapped out and interpreted in considerable detail in recent years.

Most progress in this field has been accomplished by studying germanium and silicon whose impurity content can now be controlled over an impressive range extending from about one part in 10^9 to one part in 10^3. The high purity attainable in these materials has also made it possible to elucidate the electronic structure of the host crystals in an especially detailed way. This in turn has enabled one to interpret the properties of the impurity states more fully. For this reason we shall, in this article, focus our attention principally on impurity states in silicon and germanium.

Some impurities in silicon and germanium, such as Cu and Au, can bind carriers with quite substantial energies, of the order of 0.5 ev. The orbits associated with these states are presumably rather small, of the order of a few angstroms and their structure is not very well understood. Apart from giving a brief summary of their properties in the Appendix, we shall not discuss such "deep" states further in the present article. Instead we shall limit ourselves to the "shallow" states, whose orbits are of the order of 50 A in diameter, and whose binding energies are between about 0.01 to 0.1 ev.

It may seem very surprising at first sight that the orbit of a trapped electron or hole wending its way through hundreds of crystal cells can be described at all simply. The theory shows, however, that the larger the orbit the more accurately it can be understood!

To appreciate this point let us consider a donor state for a moment. Such a state can be pictured roughly as follows. Assume a neutral impurity atom is introduced into the lattice and imagine that one electron is removed from it, leaving a positive ion behind. This ion will polarize the semiconductor so that *at large distances* it produces an electrostatic potential

$$\mathfrak{U} = \frac{e}{\kappa r}. \tag{1}$$

Here κ is the static dielectric constant. When the electron is brought back, one of two things may happen. It may be energetically favorable for this electron to occupy an orbit rather near the impurity ion. In this case Eq. (1) does *not* apply and we have a rather complicated state of affairs in the immediate vicinity of the impurity ion. In brief, we obtain a "deep" impurity state. On the other hand, the lowest energy state

may have a large orbit over most of which Eq. (1) applies. One would then guess that such a state can be described by the Schroedinger equation

$$\left(-\frac{\hbar^2}{2m^*}\nabla^2 - \frac{e^2}{\kappa r}\right)F(\mathbf{r}) = EF(\mathbf{r}) \tag{2}$$

where m^* is an appropriate effective mass.[2]

It is clear that Eq. (2) will apply more accurately the larger the orbit, since then the region near the impurity ion, where Eq. (1) does not hold, plays an insignificant role. Here we see one of the main reasons why we can form an accurate picture of the shallow impurity states: The potential is basically Coulomb-like and hence the impurity states are very closely related to the states of the hydrogen atom. A second reason is slightly more subtle. We shall see later that the concept of an effective mass holds only for a state whose wave function $F(\mathbf{r})$ has a slow spatial variation. This principle also favors the description of shallow impurity states with large orbits.

The analogy with the hydrogen atom suggests the existence not only of a lowest bound state, but of a whole spectrum of excited bound states. Such states have been clearly identified experimentally.[3] Since their orbits are even larger than that of the ground state, equations of the type (2) are particularly applicable to them.

Our understanding of the physical properties of impurity states is still being rapidly expanded. However, a great deal of insight into these fascinating "atoms" has already been gained. They are very delicate structures. They may be disrupted by an energy of the order of 0.01 ev and their orbits have the relatively immense dimensions in the range from 10^{-7} to 10^{-6} cm. Nevertheless, their structure is determined by the same quantum-mechanical laws and the same Coulomb attraction which govern their robust counterparts, the mesic atoms, with binding energies of about 10^6 ev and orbital diameters of about 10^{-12} cm. The range of energies, dimensions, and physical conditions in which hydrogen-like atoms occur in nature is truly remarkable.

The plan of the present article is as follows. First we shall review the available experimental information concerning the shallow impurity states in silicon and germanium (Part II, Sections 1–3). This information has been obtained by studying the electrical properties of silicon and germanium containing impurities, by infrared absorption experiments, by electron and nuclear spin resonance experiments, and by magnetic susceptibility measurements. Following this, we shall present the theory

[2] The actual band structures of silicon and germanium are rather complex, which requires a modification of the kinetic energy term (see Sections 5 and 9).

[3] E. Burstein, E. E. Bell, J. W. Davisson, and M. Lax, *J. Phys. Chem.* **57**, 849 (1953).

which has been developed to account for the empirical facts and shall discuss its successes and failures (Parts III–VI, Sections 4–15). Finally the properties of the deep-lying impurity states in silicon and germanium will be summarized in the Appendix. Throughout this article we shall discuss only isolated impurity states for which overlap between nearest neighbors can be neglected.

II. Empirical Properties

In this part we shall describe briefly the most important empirical information concerning the shallow impurity states in silicon and germanium. We shall return to these data in subsequent parts where they will be discussed in the light of the theoretical developments.

1. Energy Levels

a. Ionization Energies

The ionization energy of a shallow impurity "atom" is the energy of its ground state relative to that of the nearest band edge; or, physically speaking, the minimum energy required to liberate a bound carrier from its impurity ion. We shall denote this (positive) energy by E_0.

The ionization energy has been determined empirically in two ways. In one of these the number n of thermally released, conducting carriers is measured by means of the Hall constant, which is inversely proportional to n. Since the temperature dependence of n is governed primarily by an appropriate Boltzmann factor, which under the usual experimental conditions is $\exp(-E_0/kT)$,[4] this is a convenient method of determining E_0. Early measurements of this kind were made by Debye and Conwell[5] and an extensive analysis of Hall effect data has been carried out by the Bell Telephone Laboratories group.[6,7,8] The results are listed in Table I under the heading "Thermal."

The other method makes use of infrared irradiation at low temperatures. One observes the photon energy required to raise a carrier from the ground state to an excited bound state and adds to this the theoretically calculated energy of the excited state. It is believed that the latter, which is a small fraction of the total ionization energy, can be calculated quite

[4] W. Shockley, "Electrons and Holes in Semiconductors," Section 16.4. Van Nostrand, New York, 1950.

[5] P. P. Debye and E. M. Conwell, *Phys. Rev.* **93**, 693 (1954).

[6] T. H. Geballe and F. J. Morin, *Phys. Rev.* **95**, 1085 (1954). Due to an oversight, the temperatures in their graph are too high by a factor of 10.

[7] F. J. Morin, J. P. Maita, R. G. Shulman, and N. B. Hannay, *Phys. Rev.* **96**, 833 (1953). The value of 0.039 ev given there for P is a misprint.

[8] F. J. Morin, private communication.

reliably. Such studies have been carried out particularly by Burstein and his co-workers.[9,10] Their results are given in Table I under the heading "Optical." An alternative infrared technique is to determine the threshold energy for photoconductivity. Such experiments have been performed[3] but yield only very rough estimates of E_0.

TABLE I. IONIZATION ENERGIES

	Impurity element		E_0 (ev)[a] Thermal	Optical
In silicon	Li	Donors	0.033 (7)	
	P		0.044 (7)	0.045 (11)
	As		0.049 (7)	0.053 (10)
	Sb		0.039 (7)	0.043 (10)
	Bi		0.069 (8)	
	B	Acceptors	0.045 (7)	0.046 (9)
	Al		0.057 (7)	0.067 (9)
	Ga		0.065 (7)	0.071 (9)
	In		0.16 (7)	0.154 (9)
In germanium	P	Donors	0.0120 (6)	
	As		0.0127 (6)	
	Sb		0.0096 (6)	
	B	Acceptors	0.0104 (6)	
	Al		0.0102 (6)	
	Ga		0.0108 (6)	
	In		0.0112 (6)	

[a] Numbers in parentheses refer to footnote references in text.

It will be seen from Table I that, even though the results of the thermal and optical methods are in fair agreement, there remain discrepancies which, in some cases, are 10 to 20%. It appears very unlikely that the deviations are due entirely to experimental error.

The following trends are apparent from this table.

1. All shallow donors (except Li) are group V elements; all shallow acceptors are group III elements.[10a]

[9] E. Burstein, G. S. Picus, B. Henvis, and R. Wallis, *J. Phys. Chem. Solids* **1**, 65 (1956).

[10] G. S. Picus, E. Burstein, and B. Henvis, *J. Phys. Chem. Solids* **1**, 75 (1956).

[10a] This suggests the following "chemical" models: A group V donor enters the lattice substitutionally. It contributes four of its five valence electrons to saturate all covalent bonds. The fifth electron is weakly bound by the Coulomb field of the resulting positive donor ion. Similarly, a group III acceptor enters the lattice substitutionally. Its three valence electrons saturate all but one of the covalent bonds. The remaining "hole" in the bond structure circulates in a large orbit around the net negative charge in the vicinity of the acceptor ion. The energy required to remove this hole to infinity is the ionization energy of the acceptor.

[11] G. S. Picus, private communication. These data supersede those quoted in ref. 10.

TABLE II. SPECTRA OF EXCITED STATES IN SILICON

Impurity element		Energies relative to ground state[a] (in units of 0.01 ev)						
P	Donors	3.45 (11)	3.95 (10)	4.26 (11)	4.46 (10)			
As		4.21 (10)	4.74 (10)	5.06 (10)	5.29 (10)			
Sb		3.18 (11)	3.65 (10)	3.99 (10)				
B	Acceptors	3.0 (13)	3.5 (13)	3.85 (13)	3 97 (13)	4.14 (13)	4.21 (13)	4.27 (13)
Al		5.50 (13)	5.87 (13)	6.43 (13)	6.51 (13)	6.72 (13)	6.84 (13)	
Ga		5.85 (13)	6.22 (13)	6.43 (13)	~6.8 (13)	6.95 (13)	7.25 (13)	
In		14.1 (10)	14.6 (12)	15.1 (12)				

[a] Numbers in parentheses refer to footnote references in text.

2. The ionization energies of all donors and acceptors are very similar in germanium, about 0.011 ev.

3. The ionization energies in silicon are larger than in germanium and show a much wider spread, varying from 0.03 to 0.16 ev.

The ionization energies are discussed further in Sections 6, 7, and 10.

b. Spectra of Excited States

Transitions of carriers from the ground state of the impurity to discrete excited levels were first observed by Burstein and coworkers[3] with

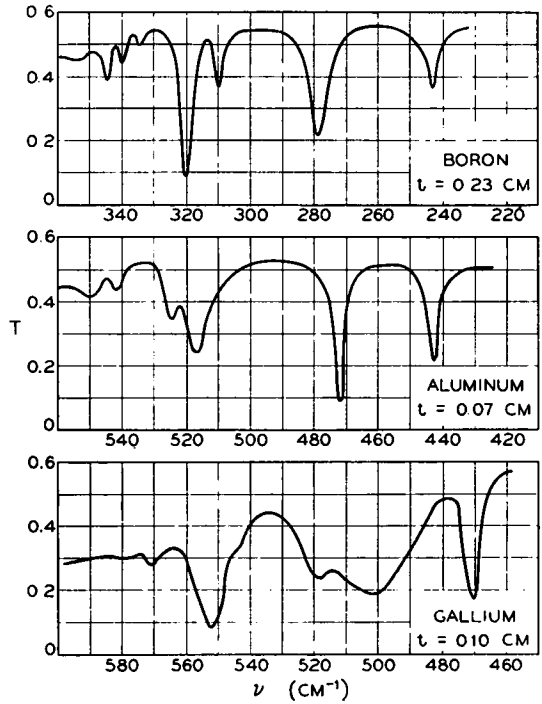

FIG. 2. Infrared absorption lines of B, Al, and Ga acceptors in silicon. T is the intensity of the transmitted radiation. (Courtesy H. J. Hrostowski.[13])

the use of infrared absorption. Further studies of such transitions have been carried out since by R. Newman,[12,12a] the group at the Naval Research Laboratories[9,10] and H. Hrostowski and R. H. Kaiser.[13] The excited states are of great theoretical interest. As a consequence of their

[12] R. Newman, *Phys. Rev.* **99**, 465 (1955).

[12a] R. Newman, *Phys. Rev.* **103**, 103 (1956).

[13] H. J. Hrostowski and R. H. Kaiser, *Bull. Am. Phys. Soc.* [2] **2**, 66 (1957); also H. J. Hrostowski, private communication.

very large orbits, they are insensitive to the complex state of affairs in the immediate vicinity of the impurity ion and should, therefore, be described well by an effective mass theory. Table II gives the experimental values of the energies of the absorbed photons, that is, the differences in energy between the ground state and the various excited states. Only impurity states in silicon have been investigated by this method so far because the radiation needed in the case of germanium lies in an experimentally difficult region ($\lambda > 100\ \mu$). Figure 2 shows typical absorption

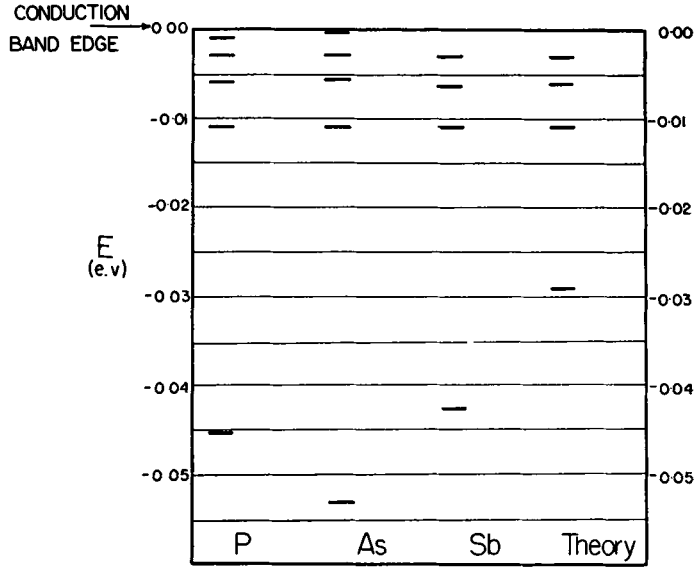

FIG. 3. Energy levels of donor states in silicon, experiment and theory. The position of the conduction band edge relative to the bound levels cannot be fixed very accurately by experiment and is here taken at 0.0109 ev above the first excited state, as predicted by theory.

plots, obtained by Hrostowski.[13] The positions of the excited levels relative to each other are of greater significance than the positions of the excited levels relative to the ground state. They are displayed in Figs. 3 and 4. The figures also show the positions of the ground states and, approximately, those of the nearest band edge. We note a striking similarity between the excited levels of different donors and, to a lesser extent, between the excited levels of different acceptors.

At helium temperature most of the lines have a width between half power points in the range from 0.5 to 1.0 $\times 10^{-3}$ ev.[13] At nitrogen temperature the width increases to about 2 $\times 10^{-3}$ ev in the case of B.[14] On the other hand, the lines associated with In broaden almost to the point of

[14] M. Lax and E. Burstein, *Phys. Rev.* **100**, 592 (1955).

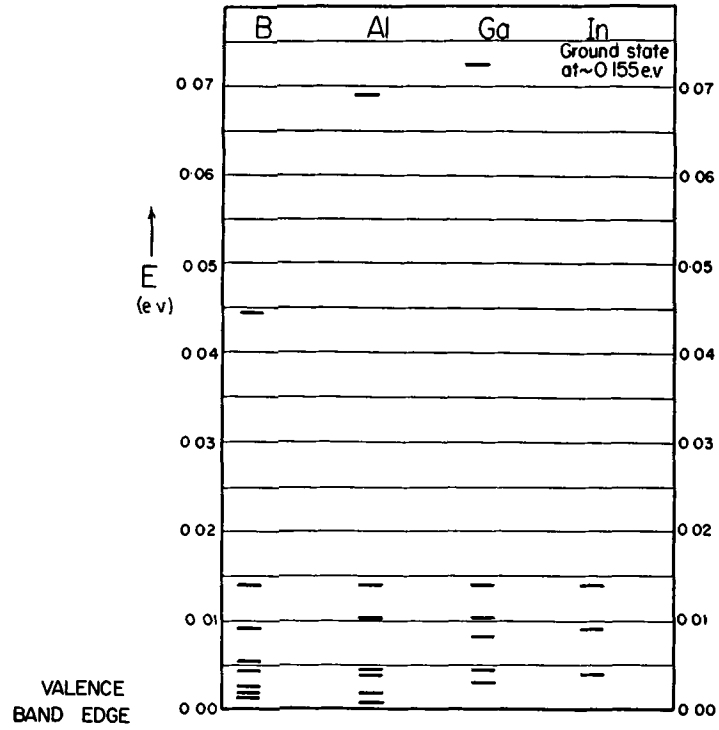

FIG. 4. Experimental energy levels of acceptor states. The position of the valence band edge is here taken to be 0.014 ev below the first excited state. The resulting ground state energies differ slightly from those quoted in Table I.

disappearing[14a] at nitrogen temperature. The line widths are interesting because they reflect the interaction of the bound carriers with the lattice vibrations. They are discussed further in Section 15.

2. SPIN RESONANCE

a. Electron Spin Resonance

The spin resonance of bound donor electrons in silicon was first observed by Fletcher *et al.*[15,16] and has since been extensively studied.[17,18]

[14a] This behavior is probably not typical of shallow impurity states and connected with the large ionization energy (0.16 ev) of In acceptors.

[15] R. C. Fletcher, W. A. Yager, G. L. Pearson, A. N. Holden, W. T. Read, and F. R. Merritt, *Phys. Rev.* **94**, 1392 (1954).

[16] R. C. Fletcher, W. A. Yager, G. L. Pearson, and F. R. Merritt, *Phys. Rev.* **95**, 844 (1954).

[17] A. Honig and A. F. Kip, *Phys. Rev.* **95**, 1686 (1954); A. Honig, *Phys. Rev.* **96**, 234 (1954).

[18] G. Feher, *Phys. Rev.* **103**, 834 (1956) and private communication.

(No resonance has yet been observed either for bound electrons in germanium or for bound holes in silicon and germanium.)

These measurements have thrown a great deal of light on the wave functions of the donor states, for, as we shall see, the resonance frequencies, the line widths, and the spin-lattice relaxation times all depend crucially on the hyperfine interaction of the donor electron with the nuclear magnetic moments within its orbit. The hyperfine interaction, in turn, gives us information about the square of the electronic wave function at the positions of the nuclei.

In their original experiments, Fletcher and co-workers found that the resonance pattern of silicon containing P donors consisted of an extremely well-resolved doublet (width $\approx$ 3 oersted, separation $\approx$ 42 oersted); the pattern of As was a quartet; and that of Sb was a superposition of a strong sextet and a weaker octet. When these observations were coupled with the fact that the naturally occurring nuclei of P have spin $\frac{1}{2}$, those of As have spin $\frac{3}{2}$, and those of Sb have spins $\frac{5}{2}$ and $\frac{7}{2}$, it was clear that the observed structures were the result of the hyperfine interaction of the donor electrons with their donor nuclei. At fixed frequency, the separation of successive lines arising from this interaction is given by

$$\Delta H = \frac{8\pi}{3} \frac{\mu_D}{I_D} |\psi(0)|^2, \tag{2.1}$$

where μ_D and I_D are the magnetic moment and spin of the donor nucleus, respectively, and $|\psi(0)|^2$ is the probability density of the donor electron at the position of the donor nucleus. Thus the measurements of

TABLE III. PROBABILITY DENSITIES AT DONOR NUCLEI IN SILICON[a]

Donor	$\lvert\psi(0)\rvert^2$ (cm^{-3})
Li	0.0032×10^{24} (18)
P	0.44×10^{24} (16)
As	1.80×10^{24} (16)
Sb	1.20×10^{24} (16)

[a] Numbers in parentheses refer to footnote references in text.

ΔH yield the important quantity $|\psi(0)|^2$. The results are given in Table III. The table also includes the value of $|\psi(0)|^2$ for Li obtained by Feher using a double resonance technique (see Section 2b). A trace of the electron resonance lines obtained with P donors is shown in Fig. 5.

Each resonance line of the hyperfine multiplets was found to have a width between half power points of about 3 oersteds. Fletcher *et al.*[16] observed that this width could be associated with so-called inhomoge-

neous broadening[19] which suggested strongly that it originated in further, unresolved hyperfine interactions of the electron. Now 4.7% of natural silicon is present in the form of the isotope Si^{29}, which carries a magnetic moment. Therefore, the resonance field of the electron should depend not

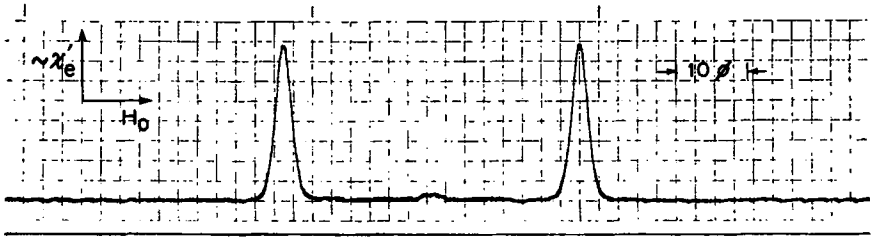

FIG. 5. Electron spin resonance lines of P donor states, at a frequency of 9000 mc. The doublet corresponds to the two possible orientations of the P nuclear spin. The width of each line is the result of the interactions of the electron with the moments of Si^{29} nuclei. (Courtesy G. Feher.[20])

only on the orientation of the spin of the donor nucleus, but also on the position and orientation of each Si^{29} spin within its orbit. Since the interaction with the Si^{29} moments is rather weak—as we shall see in more detail in Section 7c,—it offered a natural explanation of the observed widths. Table IV contains a list of the empirical line widths of the spin resonances of various donors.

TABLE IV. WIDTHS OF DONOR RESONANCES DUE TO Si^{29}

Donor	Width between half-value points of absorption line (oersted)[a]
Li	3.0 (20)
P	2.9 (20)
As	3.6 (20)
Sb	2.7 (20)

[a] Numbers in parentheses refer to footnote references in text.

Definitive experimental results concerning the spin lattice relaxation times of *isolated* donor atoms, are still lacking. The observed times are very long at low temperatures and, therefore, extremely sensitive to small disturbances such as neighboring impurities or a small number of conduction electrons. There are two types of relaxation processes:[21] 1. Those involving the simultaneous flip of an electron spin and a nuclear

[19] A. M. Portis, *Phys. Rev.* **91**, 1071 (1955).
[20] G. Feher, private communication.
[21] D. Pines, J. Bardeen, and C. P. Slichter, *Phys. Rev.* **106**, 489 (1957).

spin (donor or Si^{29}), which are characterized by the time T_x, and 2. those in which only the electron spin flips, which are characterized by the time T_e. T_x presumably originates in the hyperfine interaction, modulated by the lattice vibrations. The origin of T_e is somewhat uncertain. To give some idea of the times involved, we quote results from the experiments of Abragam and Cambrisson.[22] Working with about 10^{17} As atoms/cc, at 9000 mc and 2°K, they find $T_x = 10$ min and $T_e = 20$ min.

b. Double Resonance

A new double resonance technique has been applied successfully by Feher[18] to donor states in silicon. To illustrate the method we consider

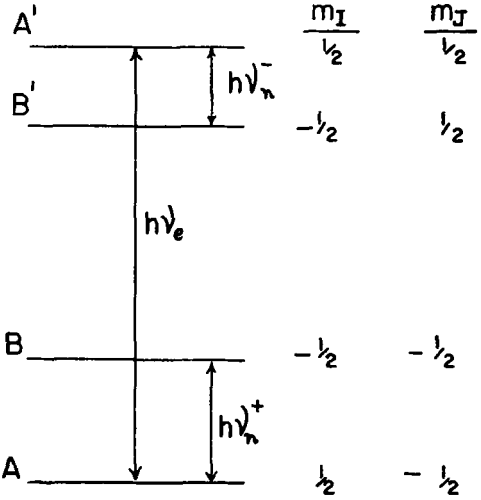

FIG. 6. Magnetic level scheme for a double resonance experiment on a P donor state (schematic); $h\nu_e$ corresponds to an electron spin flip, $h\nu_n^+$ and $h\nu_n^-$ to nuclear spin flips. (After G. Feher.[18])

the electron associated with a P donor (nuclear spin, $I = \frac{1}{2}$) and ignore for the moment the effect of the Si^{29} nuclei. When the system consisting of electron spin and nuclear spin is placed in a strong static magnetic field, it has the energy level spectrum shown in Fig. 6. The transition from A to A', corresponding to a microwave frequency, is associated with one line of the hyperfine doublet of the electron spin resonance; that from B to B' is associated with the other line. The separations $h\nu_n^+$ and $h\nu_n^-$, which lie in the radio-frequency range, and correspond to nuclear spin flips are given by

$$h\nu_n^{\pm} = 2\mu_P \left(H \pm \frac{8\pi}{3} \mu_0 |\psi(0)|^2 \right). \tag{2.2}$$

[22] A. Abragam and J. Combrisson, *Compt. rend.* **243**, 576 (1956).

Here μ_P is the magnetic moment of the P nucleus, H the external field, and $\pm(8\pi/3)\mu_0|\psi(0)|^2$ is the effective field at the nucleus arising from the electron spin. Suppose that the electron resonance $h\nu_e$ is partially saturated. The signal arising from these transitions is proportional to the difference in population between A' and A. If the nuclear transition $h\nu_n^+$ (or $h\nu_n^-$) is also induced the population difference is altered. Hence the signal strength of the electron spin resonance is changed. In this way the nuclear frequencies ν_n^+ and ν_n^- can be determined by observing the electron resonance as a function of the nuclear frequency. The value of $|\psi(0)|^2$ can then be determined using Eq. (2.2).

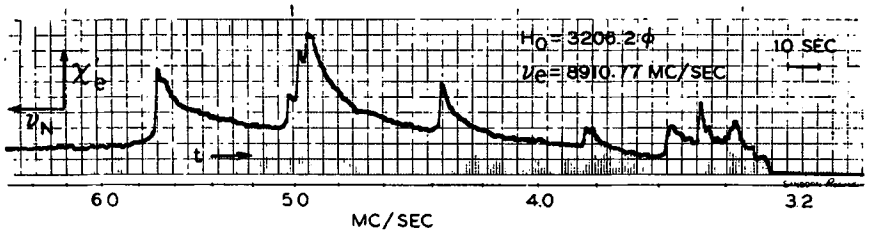

FIG. 7. Double resonance. Electron spin resonance absorption as a function of the "nuclear" frequency ν_n. Each line corresponds to the resonance frequency of a Si^{29} nucleus at a particular site relative to the P donor. (Courtesy G. Feher.[20])

Feher was able by this method to determine $|\psi(0)|^2$ for Li donors, which has the rather small value given in Table III. Ordinary techniques of electron resonance were not adequate for resolving the hyperfine multiplet.

Another interesting application of the technique of double resonance involves the Si^{29} nuclei. Again let us take a P donor and consider a particular Si^{29} nucleus at site $\mathbf{r}_l$ within the orbit of the donor electron. Clearly the nucleus has resonance frequencies given by

$$h\nu_l^{\pm} = 2\mu_{Si}\left(H \pm \frac{8\pi}{3}\mu_0|\psi(\mathbf{r}_l)|^2\right) \tag{2.2'}$$

in a magnetic field, depending on the orientation of the electron spin; μ_{Si} is the magnetic moment of the Si^{29} nucleus. As before, a nuclear signal given by (2.2') changes the intensity of the electron absorption line if the electron spin resonance is partly saturated. In this way the values of $|\psi(\mathbf{r}_l)|^2$ can be obtained for all lattice sites $\mathbf{r}_l$. Figure 7 shows the electron resonance signal as a function of the nuclear frequency ν for the case of a P donor. Each line yields the value of $|\psi|^2$ at *some* lattice site $\mathbf{r}_l$. Thus in a sense the entire wave function of the donor electron is mapped. The only remaining problem is the assignment of the various observed fre-

quencies to specific lattice sites. We shall return to a more detailed analysis of such experiments in Section 7c.

3. Static Magnetic Susceptibility

Impurity atoms exhibit both a temperature independent diamagnetism and a paramagnetism varying as $1/T$. As a very rough estimate we expect, per impurity atom,

$$\chi_{\text{dia}} \sim -\frac{e^2}{6m^*c^2}(r^2)_{\text{Av}} \sim \begin{cases} -10^{-26} \text{ cgs units in silicon} \\ -10^{-25} \text{ cgs units in germanium} \end{cases} \tag{3.1}$$

and

$$\chi_{\text{para}} \sim \chi_{\text{free spin}} \sim \frac{6.23}{T} \times 10^{-25} \text{ cgs units;} \tag{3.2}$$

here m^* is an average effective mass and T is measured in °K. For comparison, the diamagnetic susceptibilities of the host materials are

$$\begin{aligned} \chi_{\text{Si}} &= -2.7 \times 10^{-7} \text{ cgs volume units} \\ \chi_{\text{Ge}} &= -5.6 \times 10^{-7} \text{ cgs volume units.} \end{aligned} \tag{3.3}$$

Present techniques for measurement of the magnetic susceptibility have an accuracy of 1 or 2%. Therefore, to distinguish clearly the contributions of the impurity states requires concentrations of the order of 10^{16} to 10^{17} impurities per cm^3. Unfortunately, overlap effects begin to set in just in this range. As a result, no firm conclusions about the magnetic characteristics of *isolated* impurity states can as yet be drawn even though both diamagnetic and temperature dependent paramagnetic effects of the correct order of magnitude have been clearly established by experiment.[23] It is to be hoped that techniques will soon be improved by a factor of about 3 to permit experimentation on more dilute samples. The theory of the magnetic susceptibilities is discussed in Section 13.

III. Structure of Donor States

4. Conduction Bands of Silicon and Germanium

A donor state may be regarded as a wave packet which, as we shall see in the next section, is mainly composed of Bloch waves chosen from near the bottom of the conduction band. Thus a knowledge of the behavior of the conduction band near its lowest point (or points) is required for a quantitative understanding of these states. Fortunately, we have acquired a very detailed picture of the band structure of silicon and germanium, during the last few years, primarily thanks to the successful

[23] A. Van Itterbeck and W. Duchateau, *Physica* **22**, 649 (1956); also R. Bowers, private communication.

execution of cyclotron resonance experiments in these materials.[24] We shall summarize the relevant information obtained from these experiments.

a. Silicon

Figure 8 shows schematically the energy band structure of silicon along the (1,0,0) direction in **k** space.[25] The point A represents one of the 6 equivalent minima of the conduction band which are located at the

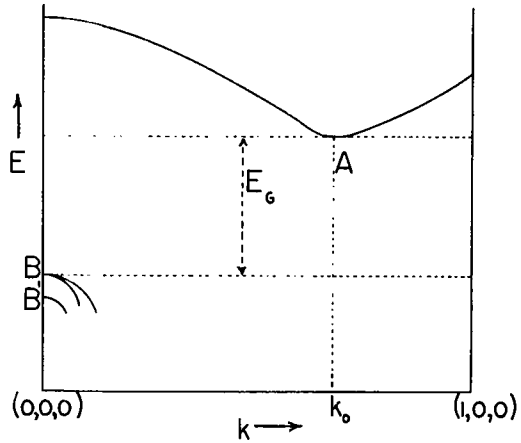

FIG. 8. Energy band structure of silicon in the (1,0,0) direction (schematic).

points $(k_0,0,0)$, $(-k_0,0,0)$, . . . $(0,0,-k_0)$. The exact magnitude of k_0 is not yet known, but there are indications that the conduction band minima are about three-quarters of the way between $\mathbf{k} = 0$ and the zone boundaries.[26] Two Bloch waves, corresponding to the two possible orientations of the electron spin, belong to each point such as A. Near one of these minima, say the one on the (0,0,1) axis, the energy of the Bloch waves is given by an expression of the form

$$E = \frac{\hbar^2}{2m_l}(k_z - k_0)^2 + \frac{\hbar^2}{2m_t}(k_x^2 + k_y^2). \tag{4.1}$$

Here the zero of energy is taken at the bottom of the conduction band;

[24] For a general discussion see G. Dresselhaus, A. F. Kip, and C. Kittel, *Phys. Rev.* **98**, 368 (1955).

[25] F. Herman, *Phys. Rev.* **95**, 847 (1955); *Physica* **20**, 801 (1954).

[26] G. G. McFarlane and V. Roberts, *Phys. Rev.* **98**, 1965 (1955); W. Kohn, *Phys. Rev.* **98**, 1561 (1955) and unpublished work.

the longitudinal and transverse "masses" m_l and m_t have the following values:[27]

$$\begin{aligned} m_l &= (0.98 \pm 0.02)m_0 \\ m_t &= (0.19 \pm 0.01)m_0. \end{aligned} \tag{4.2}$$

b. Germanium

The energy band structure of germanium along the (1,1,1) direction in **k** space is shown schematically in Fig. 9.[25] It is known with certainty from the behavior of the cyclotron resonance pattern that the conduction band minima lie on the (1,1,1) axis and the equivalent axes. There is

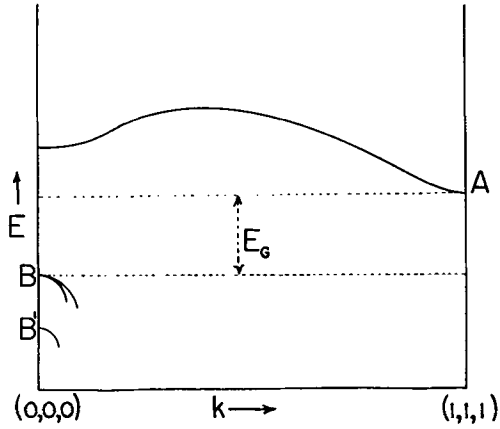

FIG. 9. Energy band structure of germanium in the (1,1,1) direction (schematic).

other strong evidence that the minima occur at the zone boundary,[28] as shown in Fig. 9. Since points such as (1,1,1) and $(-1,-1,-1)$ are then equivalent, because they differ by a reciprocal lattice vector, there are four equivalent conduction band minima. If we take the (1,1,1) direction as a new z axis and denote the distance in **k** space of the point A from $\mathbf{k} = 0$ by k_0, the energy in the vicinity of A is again given by an expression of the form (4.1). The constants m_l and m_t have the following values in germanium:[29]

$$\begin{aligned} m_l &= (1.60 \pm 0.008)m_0 \\ m_t &= (0.0813 \pm 0.002)m_0. \end{aligned} \tag{4.3}$$

[27] R. N. Dexter, B. Lax, A. F. Kip, and G. Dresselhaus, *Phys. Rev.* **96**, 1222 (1954) and private communication from Dr. Zeiger.

[28] D. K. Stevens, J. W. Cleland, J. H. Crawford, Jr., and H. C. Schweinler, *Phys. Rev.* **100**, 1084 (1955).

[29] R. C. Fletcher, W. A. Yager, and F. R. Merritt, *Phys. Rev.* **100**, 747 (1955).

5. Effective Mass Theory of Donor States

In this section we shall show that, with certain reservations, the wave function of a donor state in silicon or germanium can be expressed in the form

$$\psi(\mathbf{r}) = \sum_{j=1}^{N} \alpha_j F_j(\mathbf{r}) \varphi_j(\mathbf{r}). \tag{5.1}$$

Here N is the number of equivalent minima, which is six in silicon and probably four in germanium; the α_j are certain numerical coefficients; $\varphi_j(\mathbf{r})$ is the Bloch wave at the jth minimum; and the functions $F_j(\mathbf{r})$ are "hydrogen-like" envelope functions. If the energy of the conduction band near the jth minimum has the form

$$E_j(\mathbf{k}) = \sum_{\alpha,\beta=1}^{3} D_j^{\alpha\beta} k_\alpha k_\beta \tag{5.2}$$

$\mathbf{k}$ being reckoned from the position of the mimimum, the functions $F_j(\mathbf{r})$ satisfy so-called effective mass equations,

$$\left[E_j\left(\frac{1}{i}\nabla\right) - \frac{e^2}{\kappa r}\right] F_j(\mathbf{r}) = EF_j(\mathbf{r}). \tag{5.3}$$

Here κ is the static dielectric constant and E is the energy of the donor state relative to the conduction band minimum. In silicon, for example, the function F corresponding to the minimum on the (0,0,1) axis satisfies the equation

$$\left[-\frac{\hbar^2}{2m_l}\frac{\partial^2}{\partial z^2} - \frac{\hbar^2}{2m_t}\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right) - \frac{e^2}{\kappa r}\right] F(\mathbf{r}) = EF(\mathbf{r}). \tag{5.4}$$

[See Eqs. (4.1) and (4.2).]

It is convenient to adopt the following normalizations:

$$\frac{1}{\Omega}\int_{\text{cell}} |\varphi_j(\mathbf{r})|^2 d\mathbf{r} = 1, \tag{5.5}$$

$$\int |F(\mathbf{r})|^2 d\mathbf{r} = 1, \tag{5.6}$$

where Ω is the volume per unit cell of the crystal.

Before going on to the derivation of the wave function (5.1), let us give a qualitative picture of what it means.

Equation (5.4) evidently is very similar to the wave equation for the hydrogen atom, the difference being that the free electron mass m_0 is replaced by two different masses, m_l and m_t, and e^2 has been replaced by

the much smaller quantity e^2/κ. As a result the lowest solution of Eq. (5.4) has a mean "Bohr radius" of the order of

$$a^* = \frac{\hbar^2}{m^*(e^2/\kappa)} = a_0 \frac{\kappa}{(m^*/m_0)} \tag{5.7}$$

where m^* is an appropriate average between m_t and m_l and a_0 is the normal Bohr radius. Since m^*/m_0 is substantially less than 1 in both materials [see Eqs. (4.2) and (4.3)] and κ has the values[30]

$$\kappa_{Si} = 12.0 \tag{5.8}$$
$$\kappa_{Ge} = 16.0 \tag{5.9}$$

reasonable choices of m^* lead to the following magnitudes for a^*:

$$a_{Si}^* \approx 20 \text{ A} \tag{5.10}$$
$$a_{Ge}^* \approx 45 \text{ A}. \tag{5.11}$$

These are enormous orbits, extending over several thousand crystal cells.

Because of the heavier longitudinal mass, the wave functions $F_j(\mathbf{r})$ do not have spherical symmetry but are somewhat compressed in the direction associated with the heavy mass. This compressive effect is however much smaller than the mass ratio m_l/m_t might suggest [see Eqs. (6.5)].

The Bloch waves $\varphi_j(\mathbf{r})$ have the well-known form

$$\varphi_j(\mathbf{r}) = u_j(\mathbf{r})e^{i\mathbf{k}_j\cdot\mathbf{r}} \tag{5.12}$$

in which k_j is the position of the jth minimum and $u_j(\mathbf{r})$ is a completely periodic function. The functions $u_j(\mathbf{r})$ vary quite rapidly within each cell, particularly near the nuclei where they resemble atomic wave functions.

If we now consider one term

$$\chi_j(\mathbf{r}) = F_j(\mathbf{r})\varphi_j(\mathbf{r}) \tag{5.13}$$

in the impurity wave function (5.1), we see that it can be well approximated by

$$F_j(\mathbf{r}_0)\varphi_j(\mathbf{r}) \tag{5.14}$$

in a region consisting of just a few lattice cells. Here $\mathbf{r}_0$ is some point within this region. This shows that *locally* an electron described by (5.13) moves as though it were in the Bloch state $\varphi_j(\mathbf{r})$. However, if the electron is followed over distances of the order of a^*, its wave function is slowly modulated by the function $F_j(\mathbf{r})$ as a result of the weak Coulomb attraction of the donor ion. In brief, (5.13) can be visualized as the short wave-

[30] H. B. Briggs, *Phys. Rev.* **77**, 287 (1950).

length pattern, $\varphi_j(\mathbf{r})$, modulated in amplitude by the long wavelength envelope function $F_j(\mathbf{r})$. (See Fig. 10.)

Thus far we have discussed only *one* term of the sum (5.1). We shall show shortly that in fact any one term of this type represents an approximate solution of the Schroedinger equation with some energy E. As a result of the equivalence of the N conduction band minima, there obviously will be similar solutions, having the same energy E, associated with the other minima. The actual solutions of the impurity state problem are

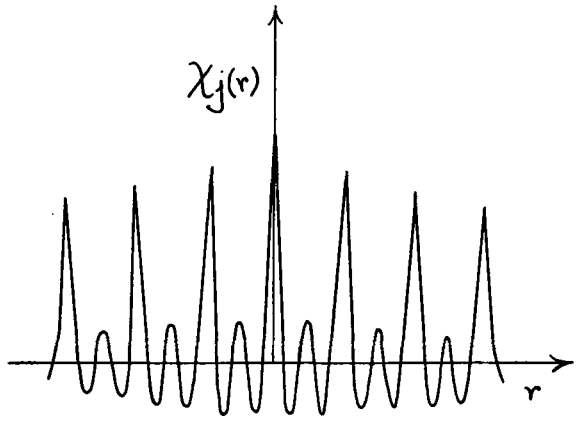

FIG. 10. Schematic representation of the donor wave function $\chi_j(\mathbf{r})$.

given approximately by linear combinations of the degenerate functions which have the correct symmetry properties. For example, the ground-state wave function has the complete tetrahedral symmetry of the Hamiltonian of the impurity state. This means that the coefficients α_j in (5.1) are all equal if the phases of $\varphi_j(\mathbf{r})$ and $F_j(\mathbf{r})$ are chosen in a natural way.[31]

We now turn to the derivation of Eq. (5.1) for the impurity state wave function.

a. Single Band Minimum at $\mathbf{k} = 0$

Equations of the type (5.3) were first obtained by Wannier[32] with the aid of the functions which have become known generally as Wannier functions. A treatment of the impurity state problem along these lines may be found in a recent article by Slater.[33] Here we shall base the discussion on the more familiar Bloch waves.

For simplicity we consider first a semiconductor whose conduction

[31] $F_j(0)$ and $u_j(0)$ real and positive for all j.

[32] G. Wannier, *Phys. Rev.* **52**, 191 (1937).

[33] J. C. Slater, *in* "Encyclopedia of Physics" (S. Flügge, ed.), Vol. XIX, p. 1. Springer, Berlin, 1956. See especially Sections 25 and 26.

band has a single minimum at $\mathbf{k} = 0$. We shall use the reduced zone scheme in which $\mathbf{k}$ is confined to the first Brillouin zone and the various energy bands are denoted by an index n. The lowest conduction band will be labeled by $n = 0$, and we shall take its minimum as our zero of energy.

In the absence of any impurities, the Hamiltonian for a conduction electron is

$$H_0 \equiv -\frac{\hbar^2}{2m_0}\nabla^2 + V(\mathbf{r}) \tag{5.15}$$

in which $V(\mathbf{r})$ is an effective periodic potential. The eigenfunctions are the Bloch waves

$$\psi_{n,\mathbf{k}}(\mathbf{r}) \equiv \frac{1}{V^{\frac{1}{2}}} u_{n,\mathbf{k}}(\mathbf{r}) e^{i\mathbf{k}\cdot\mathbf{r}} \tag{5.16}$$

which we shall take to be normalized to unity in a large cube of volume $V = L^3$. The $u_{n,\mathbf{k}}(\mathbf{r})$ are periodic functions. We denote the energy corresponding to $\psi_{n,\mathbf{k}}$ by $E_n(\mathbf{k})$.

If now one of the atoms of the perfect lattice is replaced by a positive singly charged impurity ion, the Hamiltonian for an extra electron will be

$$H = H_0 + U(\mathbf{r}) \tag{5.17}$$

where, *at large distances* from the impurity ion, one can write

$$U(\mathbf{r}) = -\frac{e^2}{\kappa r} \tag{5.18}$$

κ being the static dielectric constant. The reason for using the *static* dielectric constant is that the frequency of the impurity electron will turn out to be much lower than that of other electrons in the crystal, whose polarization gives rise to κ in a homopolar material.[34] The true $U(\mathbf{r})$ will be a rather complicated potential in the immediate vicinity of the impurity ion, but we shall suppose, and to some extent verify later, that this region does not play an important role and shall use (5.18) for all values of $\mathbf{r}$.

We consider now the motion of the impurity electron as governed by the Hamiltonian H in (5.17). We expand its wave function ψ in terms of the eigenfunctions $\psi_{n,\mathbf{k}}$ of the unperturbed Hamiltonian H_0:

$$\psi = \sum_{n,\mathbf{k}} A_n(\mathbf{k})\psi_{n,\mathbf{k}}. \tag{5.19}$$

[34] For a discussion of the role of the dielectric constant in this problem, from the point of view of the entire many electron system, see W. Kohn, *Phys. Rev.* **105**, 509 (1957).

Here $\mathbf{k}$ runs over all points inside the first Brillouin zone which are consistent with periodic boundary conditions imposed on the surface of our large cube:

$$k_\alpha L = 2\pi l_\alpha$$

l_α being an integer. Substituting (5.18) into the Schroedinger equation

$$(H_0 + U)\psi = E\psi \tag{5.20}$$

and taking the scalar product with $\psi_{n,\mathbf{k}}$, we obtain

$$(E_n(\mathbf{k}) - E)A_n(\mathbf{k}) + \sum_{n',\mathbf{k}'} (n\mathbf{k}|U|n'\mathbf{k}')A_{n'}(\mathbf{k}') = 0 \tag{5.21}$$

where

$$\begin{aligned}(n\mathbf{k}|U|n'\mathbf{k}') &\equiv \int \psi^*_{n,\mathbf{k}} \left(-\frac{e^2}{\kappa r}\right) \psi_{n'\mathbf{k}'} d\mathbf{r} \\ &= \frac{1}{V} \int u^*_{n,\mathbf{k}} u_{n',\mathbf{k}'} e^{i(\mathbf{k}'-\mathbf{k})\cdot\mathbf{r}} \left(-\frac{e^2}{\kappa r}\right) d\mathbf{r}.\end{aligned} \tag{5.22}$$

Let us study this matrix element a little more closely. We note first that $u^*_{n,\mathbf{k}} u_{n',\mathbf{k}'}$ is a periodic function and can be written as a Fourier series with the use of the reciprocal lattice vectors $\mathbf{K}_\nu$:

$$u^*_{n,\mathbf{k}} u_{n',\mathbf{k}'} = \sum_\nu C^\nu_{n\mathbf{k};n',\mathbf{k}'} e^{i\mathbf{K}_\nu \cdot \mathbf{r}} \tag{5.23}$$

which gives

$$(n\mathbf{k}|U|n'\mathbf{k}') = \sum_\nu C^\nu_{n\mathbf{k};n'\mathbf{k}'} \left(-\frac{4\pi e^2}{V\kappa|\mathbf{K}_\nu + \mathbf{k}' - \mathbf{k}|^2}\right). \tag{5.24}$$

From the orthonormality of the $\psi_{n,\mathbf{k}}$ we find by integrating (5.23) under the condition $\mathbf{k} = \mathbf{k}' = 0$,

$$C^0_{n,0;n',0} = \begin{cases} 1 & n = n' \\ 0 & n \neq n'. \end{cases} \tag{5.25}$$

It is clear that if the attractive force of the impurity ion is imagined to become weaker and weaker, the wave function ψ will more and more resemble a Bloch wave near the conduction band minimum. Since the actual Coulomb force is strongly reduced by the dielectric constant, we anticipate that the system of Eqs. (5.21) will have solutions in which $A_n(\mathbf{k})$ is

negligible unless $n = 0$ and $\mathbf{k}$ is small. This suggests that in Eqs. (5.21) we display the terms with $n = 0$ in a prominent way:

$$(E_0(\mathbf{k}) - E)A_0(\mathbf{k}) + \sum_{\mathbf{k}'} (0\mathbf{k}|U|0\mathbf{k}')A_0(\mathbf{k}') + \sum_{\substack{n' \neq 0 \\ \mathbf{k}'}} (0\mathbf{k}|U|n'\mathbf{k}')A_{n'}(\mathbf{k}') = 0 \quad (5.26a)$$

$$(E_n(\mathbf{k}) - E)A_n(\mathbf{k}) + \sum_{\mathbf{k}'} (n\mathbf{k}|U|0\mathbf{k}')A_0(\mathbf{k}') + \sum_{\substack{n' \neq 0 \\ \mathbf{k}'}} (n\mathbf{k}|U|n'\mathbf{k}')A_{n'}(\mathbf{k}') = 0 \qquad (n \neq 0). \quad (5.26b)$$

As a first approximation let us set $A_n(\mathbf{k}) = 0$ for $n \neq 0$ and replace $E_0(\mathbf{k})$ and $(0\mathbf{k}|U|0\mathbf{k})$ by their expressions for small $\mathbf{k}$ and $\mathbf{k}'$:

$$E_0(\mathbf{k}) \to \frac{\hbar^2}{2m^*} k^2 \quad (5.27)$$

$$(0\mathbf{k}|U|0\mathbf{k}') \to -\frac{4\pi e^2}{V\kappa|\mathbf{k}' - \mathbf{k}|^2}. \quad (5.28)$$

Substituting into (5.26a) we find

$$\left(\frac{\hbar^2}{2m^*} k^2 - E\right) A_0(\mathbf{k}) - \frac{4\pi e^2}{V\kappa} \sum_{\mathbf{k}'} \frac{1}{|\mathbf{k} - \mathbf{k}'|^2} A_0(\mathbf{k}') = 0. \quad (5.29)$$

This will be recognized as the Schroedinger equation in momentum space of an electron of mass m^* moving in the attractive Coulomb potential—$e^2/\kappa r$. We note, however, that in (5.29) $\mathbf{k}$ and $\mathbf{k}'$ are restricted to lie within the first Brillouin zone. If this restriction is ignored, the solutions of (5.29) are almost zero for $\mathbf{k}$ outside the first Brillouin zone in any case, so that the restriction is not of great importance. For the time being, therefore, we replace (5.29) by the equation

$$\left(\frac{\hbar^2}{2m^*} k^2 - E\right) A_0(\mathbf{k}) - \frac{4\pi e^2}{V\kappa} \sum_{\text{all } \mathbf{k}'} \frac{1}{|\mathbf{k} - \mathbf{k}'|^2} A_0(\mathbf{k}') = 0. \quad (5.30)$$

To transform this equation into a Schroedinger equation in coordinate space we introduce the Fourier transform of A_0 in the usual way, namely,

$$F(\mathbf{r}) = \frac{1}{V^{\frac{1}{2}}} \sum_{\text{all } \mathbf{k}} A_0(\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{r}}. \quad (5.31)$$

Multiplying (5.30) by $e^{i\mathbf{k}\cdot\mathbf{r}}$ and summing over $\mathbf{k}$, we obtain

$$\left(-\frac{\hbar^2}{2m^*}\nabla^2 - \frac{e^2}{\kappa r}\right)F(\mathbf{r}) = EF(\mathbf{r}) \tag{5.32}$$

a hydrogen-like so-called effective mass equation. The energy spectrum of (5.32) is of course

$$E_{n_0} = \frac{1}{n_0^2}\frac{(e^2/\kappa)^2}{2\hbar^2 m^*}, \qquad n_0 = 1, 2, \ldots . \tag{5.33}$$

The eigenfunctions are familiar from the theory of the hydrogen atom. For example, the normalized function for the ground state is

$$F(\mathbf{r}) = \frac{1}{(\pi a^{*3})^{\frac{1}{2}}} e^{-r/a^*} \tag{5.34}$$

where the "effective Bohr radius" a^* is

$$a^* = \frac{\hbar^2\kappa}{m^* e^2}. \tag{5.35}$$

The total wave function of the impurity electron is, of course, not just $F(\mathbf{r})$, but rather, in the present approximation

$$\begin{aligned}\psi(\mathbf{r}) &= \sum_{\mathbf{k}} A_0(\mathbf{k})\psi_{0,\mathbf{k}}(\mathbf{r}) \\ &= \frac{1}{V^{\frac{1}{2}}}\sum_{\mathbf{k}} A_0(\mathbf{k})u_{0,\mathbf{k}}(\mathbf{r})e^{i\mathbf{k}\cdot\mathbf{r}}\end{aligned} \tag{5.36}$$

[see Eq. (5.18)]. Anticipating again that $A_0(\mathbf{k})$ will be confined to the vicinity of $\mathbf{k} = 0$, we may write

$$u_{0,\mathbf{k}} = u_{0,0} + \mathbf{k}_\alpha(\partial u_{0,\mathbf{k}}/\partial k_\alpha) \tag{5.37}$$

in (5.36) and, to a lowest approximation, neglect all but the leading term $u_{0,0}$. The result is

$$\begin{aligned}\psi(\mathbf{r}) &\approx \frac{1}{V^{\frac{1}{2}}} u_{0,0}(\mathbf{r}) \sum_{\mathbf{k}} A_0(\mathbf{k})e^{i\mathbf{k}\cdot\mathbf{r}} \\ &= u_{0,0}(\mathbf{r})F(\mathbf{r})\end{aligned} \tag{5.38}$$

which is the form we wanted to derive [see Eq. (5.13)].

We now want to verify the two presumptions on which our derivation was based, namely that $A_0(\mathbf{k})$ extends over only a small fraction of the first Brillouin zone and that the neglected amplitudes $A_n(\mathbf{k})$, $n \neq 0$ are in fact negligible.

If we invert Eq. (5.31) and use the solution (5.34) for $F(\mathbf{r})$, we find

$$A_0(\mathbf{k}) = \frac{8\pi^{\frac{1}{2}}}{V^{\frac{1}{2}}a^{*\frac{5}{2}}} \frac{1}{[k^2 + (1/a^*)^2]^2} \tag{5.39}$$

so that $A_0(\mathbf{k})$ extends appreciably in momentum space only to values of k less than or equal to

$$\bar{k} \equiv 1/a^*. \tag{5.40}$$

This must be compared to the mean value of k at the zone boundary, obtained by replacing the first Brillouin zone by a sphere of the same volume

$$k_{ZB} = \sqrt[3]{24\pi^2}/a = 6.2/a \tag{5.41}$$

where a is the lattice parameter. In silicon and germanium, a^* has the approximate values 20 and 45 A, respectively, whereas a has the values 5.42 and 5.62 A. Therefore,

$$\bar{k}/k_{ZB} \approx \begin{cases} 0.04 \text{ in silicon} \\ 0.02 \text{ in germanium.} \end{cases} \tag{5.42}$$

At the zone boundary $A_0(\mathbf{k})$ has dropped to the fraction $(\bar{k}/k_{ZB})^4$ of its maximum value, which is quite negligible for both materials.

An estimate of $A_n(\mathbf{k})$ for $n \neq 0$ can be made from Eq. (5.26b). Treating $A_n(\mathbf{k})$ as small quantity, we have

$$A_n(\mathbf{k}) = -\frac{1}{E_n(\mathbf{k}) - E} \sum_{\mathbf{k}'} (n\mathbf{k}|U|0\mathbf{k}')A_0(\mathbf{k}') \tag{5.43}$$

which gives, in order of magnitude,

$$|A_n| \sim \frac{E_0}{\Delta E} \frac{a}{a^*} |A_0|. \tag{5.44}$$

Here E_0 and ΔE are the ionization energy and band gap energy respectively. Thus in the present theory the ratio $|A_n|/|A_0|$ is a very small quantity, of the order of 10^{-3} to 10^{-2}, in silicon and germanium.

b. Several Conduction Band Minima

We mentioned in Section 4 that in both silicon and germanium the conduction band has several minima located at equivalent points in $\mathbf{k}$ space. This requires some relatively minor modifications of the simple effective mass theory which has just been developed.[35–38]

[35] C. Kittel and A. H. Mitchell, *Phys. Rev.* **96**, 1488 (1954).
[36] M. Lampert, *Phys. Rev.* **97**, 352 (1955).
[37] W. Kohn and J. M. Luttinger, *Phys. Rev.* **97**, 1721 (1955); *ibid.* **98**, 915 (1955).
[38] W. Kleiner, *Phys. Rev.* **97**, 1722 (1955).

In the present case we can find approximate solutions of the Schroedinger equation consisting of Bloch waves associated with the vicinity of any *one* of the conduction band minima, say the one located at $\mathbf{k}_j$. We find, in complete analogy with the derivation leading to Eq. (5.38), that such solutions are given approximately by

$$\chi_j(\mathbf{r}) = F_j(\mathbf{r})\varphi_j(\mathbf{r}) \tag{5.45}$$

where φ_j is the Bloch wave at the conduction band minimum $\mathbf{k}_j$ and $F_j(\mathbf{r})$ is an envelope function satisfying the effective mass equation

$$\left[E_j\left(\frac{1}{i}\nabla\right) - \frac{e^2}{\kappa r}\right] F_j(\mathbf{r}) = EF_j(\mathbf{r}); \tag{5.46}$$

$E_j(\mathbf{k} - \mathbf{k}_j)$ is the energy of the conduction band near $\mathbf{k} = \mathbf{k}_j$.

When two effective masses are identical, as in silicon and germanium, Eq. (5.46) has axial symmetry and its solutions can be labeled by a magnetic quantum number m. Solutions corresponding to different values of $|m|$ are no longer degenerate as in the case of an isotropic mass, but the twofold degeneracy arising from $+$ and $-m$ remains.

Consider now, for example, the lowest energy solution of (5.46), which has $m = 0$. Clearly there also are $(N - 1)$ equivalent solutions, associated with the $(N - 1)$ other conduction band minima, all of which have the same energy in the effective mass approximation. This N-fold degeneracy is reduced, however, when corrections to the effective mass formalism are taken into account. The degeneracy which remains can be ascertained solely from the symmetry of the true Hamiltonian of the impurity state—invariance under the tetrahedral group T_d[39]—without detailed knowledge of the nature of the corrections. This may be achieved by the following well-known method. The N degenerate functions clearly form the basis for a representation R of the group T_d. This representation may be "reduced" into the irreducible representations A_1, A_2, E, T_1, and T_2 of the group T_d,[39]

$$R = n_1A_1 + n_2A_2 + n_3E + n_4T_1 + n_5T_2 \tag{5.47}$$

where the n_i are integers. Let us call d_i the dimensionality of the representation associated with n_i. Regardless of the detailed nature of the corrections to the effective mass formalism, the N originally degenerate levels will then split into n_1 levels with degeneracy d_1, n_2 with degeneracy d_2, etc.

Let us illustrate these remarks by using silicon, for which $N = 6$,

[39] H. Eyring, J. E. Walter, and G. E. Kimball, "Quantum Chemistry," p. 388. Wiley, New York, 1944.

as an example. For definiteness, let us take the arbitrary phase factors of the χ_j in such a way that all χ_j have the same value at $\mathbf{r} = 0$.[31] We write the linear combinations corresponding to the various irreducible representations in the form

$$\psi^{(i)} = \sum_{j=1}^{6} \alpha_j^{(i)} \chi_j \qquad i = 1, \ldots 6 \tag{5.48}$$

where $j = 1, \ldots 6$ are taken to correspond to the minima $(k_0,0,0)$, $(-k_0,0,0)$, $\ldots$ $(0,0,-k_0)$, respectively. Elementary group theory then yields the following linear combinations corresponding to the different irreducible representations:

$$\begin{array}{lll}
\alpha_j^{(1)}: & 1/\sqrt{6}(1,1,1,1,1,1) & (A_1) \\
\alpha_j^{(2)}: & \tfrac{1}{2}(1,1,-1,-1,0,0) & \left.\begin{array}{l}\\ \\ \end{array}\right\}(E) \\
\alpha_j^{(3)}: & \tfrac{1}{2}(1,1,0,0,-1,-1) & \\
\alpha_j^{(4)}: & 1/\sqrt{2}(1,-1,0,0,0,0) & \left.\begin{array}{l}\\ \\ \\ \end{array}\right\}(T_1) \\
\alpha_j^{(5)}: & 1/\sqrt{2}(0,0,1,-1,0,0) & \\
\alpha_j^{(6)}: & 1/\sqrt{2}(0,0,0,0,1,-1). &
\end{array} \tag{5.49}$$

No tetrahedrally symmetric perturbation can lift the double degeneracy of $\psi^{(2)}$ and $\psi^{(3)}$ or the triple degeneracy of $\psi^{(4)}$, $\psi^{(5)}$, and $\psi^{(6)}$. Therefore, we shall in reality have a singly, a doubly, and a triply degenerate level associated with the lowest eigenvalue of the effective mass equation. The spacing between the three levels will, of course, be the smaller, the better the effective mass theory accounts for the true state of affairs.

The situation is similar for the higher eigenvalues of the effective mass equation. Further details can be found in the article of reference 37.

c. Matrix Elements for Radiative Transitions

We have seen in Section 1b, that a great deal of information has been obtained from infrared absorption experiments in which a donor electron makes a transition from the ground state to an excited state. The transition rates are proportional to the squares of the matrix elements of momentum which we shall now evaluate.

We begin with the case of a simple band having a minimum at $\mathbf{k} = 0$, in the vicinity of which the energy is given by

$$E(k) = \frac{\hbar^2}{2m^*} k^2. \tag{5.50}$$

Let us denote the ground state by $\psi^{(0)}$ and the excited state of interest

by $\psi^{(s)}$. To calculate the momentum matrix elements it is essential to use the expressions

$$\psi^{(l)} = \sum_{\mathbf{k}} A_0{}^{(l)}\mathbf{k}\psi_{0,\mathbf{k}}(\mathbf{r}), \qquad l = 0 \text{ or } \varepsilon \tag{5.51}$$

for the wave functions rather than the less accurate expressions of the form $F^{(l)}(\mathbf{r})u_{0,0}(\mathbf{r})$.[40] Making use of the fact that $P_\alpha = (1/i)\partial/\partial x_\alpha$ is a periodic operator we find,

$$\begin{aligned}(\psi^{(0)},p_\alpha\psi^{(s)}) &= \left(\sum_{\mathbf{k}} A_0{}^{(0)}(\mathbf{k})\psi_{0,\mathbf{k}},p_\alpha \sum_{\mathbf{k}'} A_0{}^{(s)}(\mathbf{k}')\psi_{0,\mathbf{k}'}\right)\\ &= \sum_{\mathbf{k}} A_0{}^{(0)x}(\mathbf{k})A_0{}^{(s)}(\mathbf{k})(\psi_{0,\mathbf{k}},p_\alpha\psi_{0,\mathbf{k}})\\ &= \frac{m}{m^*}\sum_{\mathbf{k}} A_0{}^{(0)*}(\mathbf{k})\hbar k_\alpha A_0{}^{(s)}(\mathbf{k})\\ &= \frac{m}{m^*}(F^{(0)}(\mathbf{r}),p_\alpha F^{(s)}(\mathbf{r})).\end{aligned} \tag{5.52}$$

In view of the commutation relationship

$$[x_\alpha,H] = \left[x_\alpha, \frac{p^2}{2m}\right] = \frac{i\hbar}{m}p_\alpha \tag{5.53}$$

in which H is the total Hamiltonian appearing in (5.20), we can obtain the alternative result

$$\begin{aligned}(\psi^{(0)},p_\alpha\psi^{(s)}) &= \frac{m}{i\hbar}(\psi^{(0)},[x_\alpha,H]\psi^{(s)})\\ &= \frac{m}{i\hbar}(E^{(s)} - E^{(0)})(\psi^{(0)},x_\alpha\psi^{(s)})\\ &= \frac{m}{i\hbar}(E^{(s)} - E^{(0)})(F^{(0)}(\mathbf{r}),x_\alpha F^{(s)}(\mathbf{r})).\end{aligned} \tag{5.54}$$

Here $E^{(0)}$ and $E^{(s)}$ are the energies of $\psi^{(0)}$ and $\psi^{(s)}$, respectively.

It should be noted that neither (5.53) nor (5.54) require a knowledge of the Bloch wave at $\mathbf{k} = 0$.

In the case in which there are N equivalent conduction band minima

[40] If the latter type of wave functions are used one obtains instead of (5.52) the result $(F^{(0)}(\mathbf{r}), p_\alpha F^{(s)}(\mathbf{r}))$; i.e., one loses the essential factor m/m^*.

we may let the ground and excited states have the wave functions $\psi^{(0)}$ and $\psi^{(e)}$ of the form

$$\psi^{(i)} = \sum_{j=1}^{N} \alpha_j^{(i)} F_j^{(i)} \varphi_j. \tag{5.55}$$

A simple extension of the derivation just given leads to the result

$$(\psi^{(0)}, p_a \psi^{(e)}) = \frac{m}{i\hbar} (E^{(e)} - E^{(0)}) \sum_{j=1}^{N} \alpha_j^{(0)*} \alpha_j^{(e)} (F_j^{(0)}, x_a F_j^{(e)}). \tag{5.56}$$

6. Numerical Results and Comparison with Experiments

We shall now compare the predictions of the simple effective mass theory, developed in the preceding section, with experiment.

a. Energy Levels

To obtain the energy levels in the effective mass approximation we must determine the eigenvalues of Eq. (5.4),

$$\left[-\frac{\hbar^2}{2m_l} \frac{\partial^2}{\partial z^2} - \frac{\hbar^2}{2m_t} \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right) - \frac{e^2}{\kappa r} \right] F(\mathbf{r}) = EF(\mathbf{r}) \tag{5.4}$$

in which the parameters m_l, m_t and κ have the values given in Eqs. (4.2), (4.3), (5.8), and (5.9). These values have been determined by entirely independent experiments and will not be regarded as "adjustable."

Since (5.4) has axial symmetry it can be partly separated and reduced to a two-dimensional partial differential equation. The latter must then be solved approximately, for example by variational methods.[35–38] The solutions are, of course, rather different from simple hydrogenic ones because of the large ratio of m_l/m_t in both silicon and germanium. Nevertheless, it is convenient to label them by the familiar atomic notation (1*s*, etc.) which indicate the hydrogenic functions into which they go in the limit where $m_l/m_t \to 1$. The magnetic quantum number m refers, of course, to the remaining axis of symmetry. The results are given in Table V. Let us compare them with the experimental data.

In germanium only the ionization energy of the 1*s* state has been measured. For the different impurities, it lies between 0.0096 and 0.0127 ev (see Table I). This is in rather good agreement with the effective mass value of 0.0092 ev.

In silicon the effective mass value of 0.029 ev for the ground state is in good agreement with the ionization energy of Li (0.033 ev), but agrees poorly with those of P, As, Sb, and Bi (0.045 to 0.067 ev). Since the orbits in silicon are smaller than those in germanium, it is not surprising that

TABLE V. DONOR SPECTRA IN EFFECTIVE MASS APPROXIMATION

State	Energies in units of 0.01 ev[a]	
	Silicon	Germanium
$1s$	-2.9 ± 0.1 (37)	-0.92 ± 0.02 (42)
$2p$, $m = 0$	-1.09 ± 0.02 (41)	-0.45 ± 0.02 (42)
$2s$	-0.88 ± 0 06 (37)	
$2p$, $m = \pm 1$	-0.59 ± 0.01 (41)	-0.160 ± 0.003 (42)
$3p$, $m = 0$	-0.57 ± 0.06 (41)	-0.235 ± 0.02 (42)
$3p$, $m = \pm 1$	-0.29 ± 0.005 (41)	-0.085 ± 0.005 (42)

[a] Numbers in parentheses refer to footnote references in text.

the effective mass theory is less successful. We shall return to this comparative failure of the theory in the following Section 7.

Turning now to the excited states in silicon (see Table II), we note the following experimental values of the differences in energy between successive excited states:

P	As	Sb		
0.050	0.053	0.047	ev	
0.031	0.032	0.034	ev	(6.1)
0.020	0.023		ev	

It is encouraging to observe that the first two differences in energy are very similar for all donors whose excited states have been studied. This would, of course, be expected from the effective mass theory. Before comparing the experimental results with the calculations in more detail, we first note that the matrix elements for radiative transitions from the ground state vanish in the effective mass approximation unless the envelope function of the excited state in question has odd parity [see Eq. (5.56)]. Approximate calculations of the total intensity give the results listed in Table VI.[41] Only limited reliance can be placed on these results,

TABLE VI. CALCULATED OPTICAL ABSORPTION INTENSITIES IN SILICON

Transition	Relative intensity
$1s \rightarrow (2p, m = 0)$	4.0
$1s \rightarrow (2p, m = \pm 1)$	10.6
$1s \rightarrow (3p, m = 0)$	0.4
$1s \rightarrow (3p, m = \pm 1)$	3.1

[41] W. Kohn, *Phys. Rev.* **98**, 1856 (1955).
[42] W. Kohn, unpublished.

for the matrix elements involve the $1s$ function, which, as we have seen in connection with the $1s$ energy, is understood only imperfectly. The small intensity $1s \rightarrow (3p,\ m = 0)$ is probably significant however. Moreover, since the $(3p,\ m = 0)$ level should be almost coincident with the "strong" $(2p,\ m = \pm 1)$ level (see Table V), it is quite likely that it would escape detection. Consequently, we are led to compare the first two experimental spacings (6.1) with the following theoretical energy differences:

$$\begin{aligned}(2p,\ m = \pm 1) - (2p,\ m = 0) &: 0.050 \text{ ev} \\ (3p,\ m = \pm 1) - (2p,\ m = \pm 1) &: 0.030 \text{ ev}.\end{aligned} \tag{6.2}$$

The agreement is strikingly good and is one of the brightest points in the field of impurity states. The comparison between experiment and theory is also shown in Fig. 3.

It is easy to present two good reasons why the effective mass theory should apply better to the excited states than to the ground state. First, their orbits are larger; second, their envelope functions have odd parity and thus vanish at the donor nucleus. For both these reasons the orbits of the excited state penetrate quite negligibly into the region near the donor ion where our expression $-e^2/\kappa r$ for the potential energy presumably breaks down.

b. Wave Functions

The wave function of greatest interest is that of the ground state. It has the form

$$\psi(\mathbf{r}) = \sum_{j=1}^{N} \alpha_j F_j(\mathbf{r}) \varphi_j(\mathbf{r}) \tag{6.3}$$

[see Eq. (5.1)].

Let us begin by discussing the coefficients α_j. As we have seen, the ground state is N-fold degenerate in the effective mass theory, corresponding to N different sets of α_j. This degeneracy is partly lifted when corrections to the effective mass formalism are taken into account. However, it is not certain on *a priori* grounds to which representation the resulting lowest level belongs. One would be inclined to expect that the completely symmetrical, nondegenerate state A_1 [see Eq. (5.49)], in which all α_j are equal has the lowest energy. Fortunately there is a crucial test of this assumption. We can see from Eq. (5.49) that only the state A_1 has a nonvanishing wave function at the donor nucleus, which produces a large hyperfine interaction with the spin of the donor electron. In silicon, the experimental observation of this interaction, which has been mentioned in Section 2a and will be discussed further below, provides conclusive proof that the ground state has the complete tetrahedral sym-

metry characteristic of A_1. Donor spin resonance has not been observed in germanium, although it has been looked for[43] vigorously. Hence, no clear-cut decision can be made in this case. One might wonder if the assumption of a degenerate ground state belonging to a multi-dimensional representation of T_d would account for the failure of spin resonance experiments in germanium.

We now turn to the envelope functions F_j. We consider any one of the conduction band minima and take the direction defined by its $\mathbf{k}$ vector as the z axis. To a very good approximation the corresponding normalized wave function can be represented by

$$F = \frac{1}{(\pi a^2 b)^{\frac{1}{2}}} e^{-\sqrt{(x^2+y^2)a^2+z^2/b^2}}.^{35-37} \tag{6.4}$$

This functional form, which is of course exact in the limit of equal effective masses, was also found to be very accurate in the limit $m_l/m_t \to \infty$.[37] A variational calculation gives the following values for a and b:

$$\begin{aligned} a_{Si} &= 25.0 \times 10^{-8} \text{ cm}, & b_{Si} &= 14.2 \times 10^{-8} \text{ cm} \\ a_{Ge} &= 64.5 \times 10^{-8} \text{ cm}, & b_{Ge} &= 22.7 \times 10^{-8} \text{ cm}. \end{aligned} \tag{6.5}$$

We see that the wave functions extend further in the x and y directions, which correspond to the smaller transverse mass, than in the z direction, which corresponds to the larger longitudinal mass; however, the distortion is relatively small. Even in germanium, which has a mass ratio of 20 to 1, the ratio of a to b is only 3 to 1. Naturally we obtain a tetrahedrally symmetrical structure when the full wave function (6.3) is considered, which may be pictured as approximately spherical with a "radius" given roughly by the larger length a.

Finally we come to the Bloch wave functions φ_j. They have already been discussed in Section 4. We want to add only one important item of information. By studying the rate at which partially aligned Si^{29} nuclei relax by means of the hyperfine interaction with conduction electrons, Shulman and Wyluda[44] obtained an estimate of the square of the wave functions associated with silicon conduction band at the positions of the Si nuclei. Their results are best expressed in terms of a dimensionless quantity which measures the extent to which the conduction band wave functions are concentrated at the lattice sites $\mathbf{r}_l$:

$$\eta \equiv \frac{|\varphi_j(\mathbf{r}_l)|^2}{|\varphi_j(\mathbf{r})|^2_{Av}} = 186 \pm 18. \tag{6.6}$$

We are now ready to compare the theoretical value of $|\psi|^2$ at the posi-

[43] G. Feher, private communication.

[44] R. G. Shulman and B. J. Wyluda, *Phys. Rev.* **103**, 1127 (1956).

tion of a P donor in silicon, with the experimental value obtained from the hyperfine splitting of the donor electron resonance (Section 2a).[45] Using (6.3) with all $\alpha_j = \sqrt{\frac{1}{6}}$, and closing a, b, and η as given by Eqs. (6.5) and (6.6), we obtain

$$\begin{aligned} |\psi(0)|^2 &= 6|F_j(0)|^2|\varphi_j(0)|^2 \\ &= 6\,\frac{1}{(\pi a^2 b)}\cdot\eta \\ &= 0.042 \times 10^{24}\ \text{cm}^{-3}. \end{aligned} \tag{6.7}$$

This is much smaller than the experimental value, namely 0.44×10^{24} cm^{-3} (Table III). The difference represents a serious failure of the effective mass theory. We shall see in the following section how most of this discrepancy can be removed by a simple correction to the formalism.

7. Corrections to the Effective Mass Formalism

a. General Considerations

We have seen some of the successes and failures of the effective mass formalism in the preceding section. They are roughly what we might have expected. For states having large orbits, which do not penetrate into the immediate vicinity of the donor ion appreciably, the theory agrees well with experiment; for states having smaller orbits and substantial amplitudes near the donor ion, notably the ground states in silicon, the agreement is poor.

What are the chief reasons for the failures of the effective mass formalism? We have based our discussions on the Hamiltonian

$$H = -\frac{\hbar^2}{2m}\nabla^2 + V(\mathbf{r}) - \frac{e^2}{\kappa r} \tag{7.1}$$

[see Eqs. (5.17), (5.15), (5.18)] and have shown that it leads to an effective Hamiltonian,

$$H_{\text{eff}} \equiv E\left(\frac{1}{i}\nabla\right) - \frac{e^2}{\kappa r} \tag{7.2}$$

[45] The reason for concentrating on P ($Z = 15$) is that there exists a mathematical limit where the effective mass wave function (6.3) should become exact everywhere, even at the position of the P nucleus. This is the limit in which the charge on the P nucleus is imagined to be only infinitesimally larger than that on the Si nuclei (i.e., $Z_P = 14 + q$). In this limit the orbits become very large and the wave function is, locally, just like a superposition of silicon conduction band functions. On the other hand, consider another donor such as As ($Z = 33$). Here even in the limit in which the Coulomb attraction is imagined to be very weak ($Z = 32 + q$), the donor wave function in the cell of the As ion will resemble conduction band wave functions of Ge ($Z = 32$) more than those of Si. This fact is not allowed for in the effective mass expression (6.3).

[see Eq. (5.46)]. The errors involved in passing from (7.1) to (7.2) were discussed in Section 5, and are very small indeed for silicon and germanium, being almost certainly less than 1%. The trouble must therefore lie with the original Hamiltonian (7.1). Unfortunately a thorough analysis of these difficulties has not been published yet; therefore, we must be content with some rather incomplete remarks.

The first question which may be asked is whether the impurity electron can be described adequately by *any* single particle Hamiltonian. After all, it interacts constantly and strongly with a large number of electrons which lie within its orbit. It is by no means obvious that the actual many-particle problem can be replaced adequately by a single particle equation. A partial answer to this question was given by the author,[46] who showed that the entire many-electron Hamiltonian can be reduced to a single electron effective mass Hamiltonian of the type H_{eff} of Eq. (7.2) in the hypothetical case in which the charge on the impurity nucleus exceeds that of the other crystal nuclei by an infinitesimal amount. The same result can also be derived[47] for sufficiently highly excited states in the case in which the donor nucleus possesses a finite excess charge. This is a satisfactory result in view of the good agreement between the effective mass theory and experiments involving the excited donor states.

For the ground state of the donor, the situation near the donor ion is quite complicated. However, for $r \gtrsim r_s$ [$(4\pi/3)r_s^3$ = atomic volume] it can be shown[47] that $F(\mathbf{r})$ satisfies the effective mass equation

$$\left[E\left(\frac{1}{i}\nabla\right) - \frac{e^2}{\kappa r}\right] F(\mathbf{r}) = EF(\mathbf{r}) \qquad r \gtrsim r_s, \tag{7.3}$$

provided only that the fractional change of $F(\mathbf{r})$ over a lattice spacing is small. Here E is the *actual* (observed) energy. It is to be expected that (7.3) will break down for $r \lesssim r_s$. For example, the concept of the dielectric constant is meaningful only for sufficiently large distances. Moreover, the donor ion may distort the lattice significantly in its immediate vicinity.

b. Corrected Wave Functions[48]

For definiteness let us now consider a P donor in silicon. On the basis of Eq. (7.2) we can make the following simple correction to the effective

[46] W. Kohn ref. 34.

[47] Unpublished calculations by the author.

[48] The following discussion is based on W. Kohn and J. M. Luttinger, *Phys. Rev.* **97**, 883 (1955).

mass wave function in the *exterior* region $r \gtrsim r_s$: Instead of solving (7.3) over *all* space with the boundary conditions $F(\infty) = 0$ and $F(0) =$ finite, which determine the eigenvalue E, we solve (7.3) in the exterior region using only the one boundary condition $F(\infty) = 0$ and the *observed* energy E_{obs}. For simplicity let us also replace m_l and m_t by a single mean mass m^*. The latter may for example, be chosen so as to give the same energy, in the effective mass approximation, as m_l and m_t.[49] We then have in the exterior region

$$r \gtrsim r_s: \quad \psi = \frac{1}{\sqrt{6}} F(\mathbf{r}) \sum_{j=1}^{6} \varphi_j(\mathbf{r}) \tag{7.4}$$

$$\left[-\frac{h^2}{2m^*} \nabla^2 - \frac{e^2}{\kappa r} \right] F(\mathbf{r}) = E_{obs} F(\mathbf{r}) \tag{7.5}$$

$$F(\infty) = 0. \tag{7.6}$$

Now E_{obs} (-0.045 ev) is substantially lower than the effective mass energy (-0.029 ev). Therefore, the solution of (7.5) and (7.6) falls off more rapidly than the effective mass wave function for large r, and since E_{obs} is not an eigenvalue of the effective mass equation, $F(\mathbf{r})$ approaches infinity as r becomes small. Both these effects tend to concentrate the donor electron near the donor ion much more than in the effective mass formalism.

The asymptotic behavior of $F(\mathbf{r})$ for large values of $\mathbf{r}$ is

$$F(\mathbf{r}) \sim e^{-r/b^*} \tag{7.7}$$

where b^* is related to the "Bohr radius" a^* associated with the effective mass approximation by the equation

$$b^* = a^*(E_{\text{eff mass}}/E_{obs})^{\frac{1}{2}}. \tag{7.8}$$

We shall treat the interior region, $r \lesssim r_s$, rather crudely. Let us write the Schroedinger equation for ψ in the form

$$\left(-\frac{h^2}{2m} \nabla^2 + V + U \right) \psi = E_{obs} \psi \tag{7.9}$$

where V is the effective periodic potential for the Bloch electrons near the conduction band minima and U is the perturbation caused by the extra charge on the P nucleus. Energies are measured from the bottom

[49] Other ways of choosing m^*, for example by the condition that $F(0)$ should have the same value as that calculated with m_l and m_t give very nearly the same result.

of the conduction band. If this extra charge is imagined to become infinitesimal, (7.9) clearly reduces to

$$\left(-\frac{h^2}{2m}\nabla^2 + V\right)\psi = 0 \tag{7.10}$$

and the solution of (7.10) with the correct tetrahedral symmetry is

$$\psi = C\sum_{j=1}^{6}\varphi_j(\mathbf{r}). \tag{7.11}$$

With the actual extra charge e on the P nucleus, U and E_{obs} are still small compared to V. Thus we shall write the solution of (7.9) in the form

$$\psi = F(\mathbf{r})\sum_{j=1}^{6}\varphi_j(\mathbf{r}) \tag{7.12}$$

where $F(\mathbf{r})$ is not expected to vary appreciably over the distance r_s. Rough estimates[42,48] indicate a variation of probably no more than 10–20%. Although (7.12) still has the form of an effective mass wave function, it should be emphasized that we should not expect $F(\mathbf{r})$ to satisfy an effective mass equation for small r. As we lack detailed knowledge of $F(\mathbf{r})$ for $r \lesssim r_s$, we can only set

$$r \lesssim r_s\colon \quad F(\mathbf{r}) \approx F(\mathbf{r}_s) \tag{7.13}$$

where $|\mathbf{r}_s| = r_s$.

It may appear that there is a contradiction between this picture of a "slowly" varying $F(\mathbf{r})$ and the statement following Eq. (7.6) that $F(\mathbf{r})$ approaches infinity for small r. Actually there is no difficulty. The increase of $F(\mathbf{r})$ as $r \to r_s$ from the outside is quite rapid on the scale of a^*, the "Bohr radius" of the effective mass equation, but is slow over dimensions of the order of r_s.

A schematic picture of the envelope function $F(\mathbf{r})$, corrected for deviations from the effective mass formalism, is shown in Fig. 11.

For other donors in silicon, such as Li, As, Sb, and Bi, the discussion of the exterior region is the same as for P. One merely has to use the observed energy of the ground state for each donor. This makes $F(r_s)$ a monotonic function of E_{obs}. The passage into the interior region is much more difficult, however, for this region is now occupied by an ion which is completely different from the atoms of the host crystal and the form of the donor wave function is quite different from the expression (7.12). An extremely qualitative treatment of the interior region for these donors is given in ref. 48.

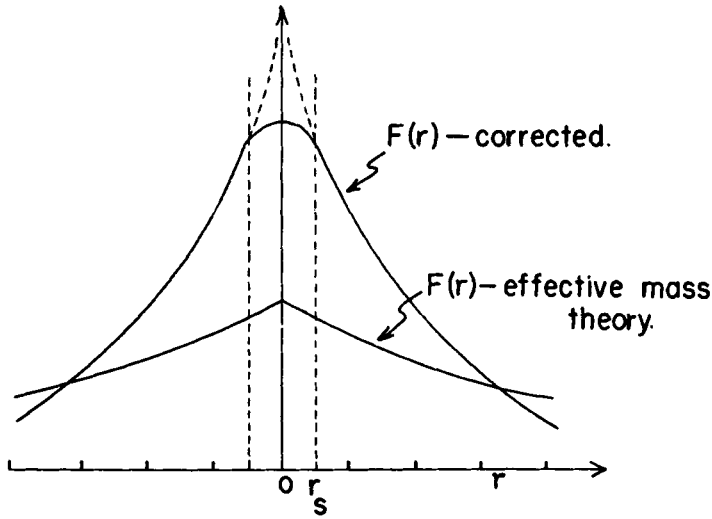

FIG. 11. Comparison of the corrected envelope function and the effective mass envelope function for a P donor (schematic).

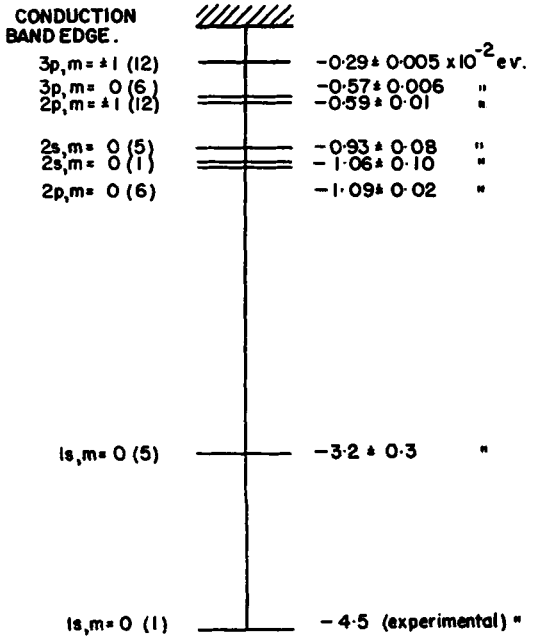

FIG. 12. Theoretical spectrum of a P donor. The numbers in parentheses refer to number of (approximately) degenerate states, spin degeneracy not included.

We remarked in Section 5b, that in the effective mass approximation there are 6 degenerate wave functions corresponding to the lowest energy eigenvalue. Now we have seen that the completely symmetrical state (belonging to representation A_1) has in fact an energy that differs quite appreciably from the effective mass value. The other 5 states have wave functions which vanish at the donor nucleus. One would, therefore, expect them to be less affected by the breakdown of the effective mass formalism near this nucleus. Their energies should then be more nearly equal to the effective mass energy.

Similarly, of the six "$2s$"-like states only one is expected to be depressed appreciably, by an amount which can be calculated from the observed "$1s$" depression by means of the quantum defect method. (See ref. 48.) The resulting spectrum for a P donor is shown in Fig. 12.

c. Comparison with Experiment

In Section 6b we calculated $|\psi(0)|^2$ for a P-donor electron by means of the effective mass formalism and found it to be 10 times smaller than the experimental value. We have also noted in Section 6a that the observed ionization energy is about 50% larger than that calculated by the effective mass theory. We shall now see that the first large discrepancy disappears when the experimental ionization energy, rather than the theoretical value, is properly included.

Following the discussion of the preceding subsection we have

$$\begin{aligned} |\psi(0)|^2 &= \left| \frac{1}{\sqrt{6}} F(0) \sum_{j=1}^{6} \varphi_j(0) \right|^2 \\ &= 6|F(0)|^2 |\varphi_j(0)|^2 \\ &\approx 6|F(\mathbf{r}_s)|^2 |\varphi_j(0)|^2. \end{aligned} \tag{7.14}$$

Now when the properly normalized[50] $F(r)$ is calculated from Eqs. (7.5), (7.6) we find that $F(\mathbf{r}_s)^2$ is 9.8 times as large as the value of $|F(0)|^2$ calculated in the effective mass approximation. This large enhancement is produced by the relatively small shift in energy because $F(\mathbf{r})$ tends to become infinite for small r. When this correction is applied to the effective mass result (6.7) one obtains

$$|\psi(0)|^2_{\text{corrected}} = 9.8|\psi(0)|^2_{\text{eff. mass}} = 0.41 \times 10^{24}\ \text{cm}^{-3}. \tag{7.15}$$

This agrees very well with the experimental result, namely 0.44×10^{24}

[50] For normalization purposes one can use the singular solution of (7.5), (7.6) for *all* r, since it is square integrable and the "interior" region makes a negligible contribution because of its small volume.

cm^{-3}, in fact somewhat better than one has reason to expect. For example, the choice of r_s as the dividing line between interior and exterior regions is somewhat arbitrary. Nevertheless, it is gratifying that this major failure of the effective mass equation can be remedied so simply and naturally.

For other donors the very rough discussion of $|\psi(0)|^2$ given in ref. 48 is found to be in reasonable agreement with experiment. Let us take the case of Li for example. Here the ionization energy is very close to the effective mass value (0.033 and 0.029 ev, respectively) so that the "enhancement factor" 9.8 of Eq. (7.15) is absent. Furthermore, Li is a very light atom so that the density of the donor electron is not as concentrated near its nucleus as near a silicon or phosphorus nucleus. Taking both these factors into account, the theoretical estimate for $|\psi(0)|^2$ was 0.002×10^{24} cm^{-3} which may be compared with the recently measured value of 0.0032×10^{24} cm^{-3}.

Let us next ask how well the theory can account for the effects of the hyperfine interaction of the donor electron with the Si^{29} nuclei (see Section 2). For this purpose we can use the exterior form of the wave function

$$\begin{aligned}\psi(\mathbf{r}) &= \frac{1}{\sqrt{6}} F(\mathbf{r}) \sum_{j=1}^{6} \varphi_j(\mathbf{r}) \\ &= \frac{1}{\sqrt{6}} F(\mathbf{r}) \sum_{j=1}^{6} u_j(\mathbf{r}) e^{i\mathbf{k}_j \mathbf{r}} \end{aligned} \tag{7.16}$$

for all donors, where $F(\mathbf{r})$ is of course computed with the ionization energy appropriate to the donor in question. The resonance frequency of a Si^{29} nucleus located at $\mathbf{r}_l$ is determined by $|\psi(\mathbf{r}_l)|^2$ according to Eq. (2.2′). With the phase convention of footnote 39a, all six $u_j(\mathbf{r})$ are equal at each lattice site, say $u(\mathbf{r}_l)$, so that we can remove the common factor $u(\mathbf{r}_l)$ from the sum in (7.16) and write

$$\psi(\mathbf{r}_l) = \frac{1}{\sqrt{6}} F(\mathbf{r}_l) u(\mathbf{r}_l) \left(\sum_{j=1}^{6} e^{i\mathbf{k}_j \mathbf{r}_l} \right). \tag{7.17}$$

Here we see an interesting *interference effect* between the six wavepackets of which $\psi(\mathbf{r})$ is composed: The magnitude of $\psi(\mathbf{r}_l)$ depends not only on $F(\mathbf{r}_l)$ and $|u(\mathbf{r}_l)| = \eta^{\frac{1}{2}}$ [see Eq. (6.6)] but also on the extent to which the six terms in (7.17) interfere with each other. This interference is determined by the magnitude k_0 of the six $\mathbf{k}_j$, that is, by the exact location of the minima of the conduction band along the (1,0,0) and equivalent axes.

Since k_0 is not yet well known, D. Schechter, in unpublished work, has plotted

$$|\psi(\mathbf{r}_l)|^2 = \tfrac{1}{6}|F(\mathbf{r}_l)|^2\eta\left(\sum_{j=1}^{6} e^{\mathbf{k}_j\cdot\mathbf{r}_l}\right)^2 \tag{7.18}$$

as function of k_0, for lattice points located near the donor ion. The comparison with the experimental results of Feher[18] is complicated by the fact that one does not know, of course, which experimental resonance line corresponds to a specific $\mathbf{r}_l$. The general magnitudes of the experimental and theoretical values of $|\psi(\mathbf{r}_l)|^2$ are in good agreement, but a more detailed analysis is still in progress. Such an analysis might test not only the theoretical picture of the donor states but could also pin down the exact position k_0 of the minima of the conduction band.

Let us turn next to the *width* of the electron spin resonance arising from the hyperfine interaction of the electron spin with the magnetic moments of the Si^{29} nuclei (see Section 2a). The shift of the resonance field of a particular donor electron arising from its interaction with the moments of Si^{29} within its orbit is given by

$$\Delta' H = \frac{8\pi}{3}\frac{\mu_{Si}}{I_{Si}}\sum_t |\psi(\mathbf{r}_t)|^2 m_t. \tag{7.19}$$

Here μ_{Si} and I_{Si} are the magnetic moment (-0.555 nuclear magneton) and spin ($\frac{1}{2}$) of Si^{29}, t runs over all lattice sites *occupied* by Si^{29} nuclei and m_t is the magnetic quantum number of the Si^{29} nucleus at $\mathbf{r}_t$; $\Delta' H$ differs from donor to donor because of differences in $\mathbf{r}_t$ and m_t. If one uses the fact that $\mathbf{r}_t$ and m_t are random, one finds the expression

$$\Delta H_{rms} = [\overline{(\Delta' H)^2}]^{\frac{1}{2}} = f^{\frac{1}{2}}\left(\frac{8\pi}{3}\mu_{Si}\right)\left[\sum_l{}' |\psi(\mathbf{r}_l)|^4\right]^{\frac{1}{2}} \tag{7.20}$$

for the root-mean-square value of $\Delta' H$. Here f is the natural fractional abundance of the Si^{29} isotope (0.0468) and the sum now runs over *all* lattice sites except that of the donor.

Equation (7.20), which again depends on k_0, must be compared to the observed root-mean-square widths of the electron resonance lines. A preliminary comparison of this type has been reported[51] and appears promising. It was mentioned in Section 2a that the hyperfine *splitting* of the electron resonance arising from interaction with the donor nucleus is much larger (42 oersted for P) than the *width* of each multiplet (≈ 3 oersted) associated with its interaction with all Si^{29} nuclei in its orbit.

[51] W. Kohn, ref. 26.

This indicates a much weaker interaction with the Si^{29} nuclei. The difference is due to the following two causes: first, the envelope function $F(\mathbf{r})$ becomes exceptionally large at $\mathbf{r} = 0$, as was mentioned in the discussion following Eq. (7.6); second while the planewave terms $\exp(i\mathbf{k}_j \cdot \mathbf{r}_l)$ are all in phase at $\mathbf{r}_l = 0$, they will interfere substantially at most other lattice sites. It appears at present that the line widths of Sb, P, and A (Table IV) can be fitted within about 25% by a choice of k_0 in the vicinity of $0.7k_{max}$. Here k_{max} is the value of k at the zone boundary. Experimental and theoretical work on this question is still in progress and the quoted estimate of k_0 must be regarded as very tentative.

IV. Structure of Acceptor States

In this chapter we shall review our current understanding of the shallow acceptor states in silicon and germanium. While these states are analogous to the donor states in a general way, we do not have nearly as detailed a picture of them. The reason for this is twofold. In the first place the theory of the states involves the structure of the valence band near its top in an essential way and we shall see in Section 8 that this band is much more complicated and less well known than the structure of the conduction bands. Secondly, spin resonance experiments, which have shed so much light on the nature of donor states, have not been successful in the case of acceptor states. Fortunately the infrared absorption spectra of these states, which were first observed by Burstein and co-workers,[3] have been determined recently by Hrostowski and Kaiser[13] with remarkable precision. Their measurements promise to be an important aid in the further elucidation of the nature of the acceptor states.

8. Valence Bands of Silicon and Germanium

Acceptor states are wave packets made up largely of Bloch waves chosen from near the top of the valence band. To understand them one requires some information about the structure of this band. This information, which has for the most part been extracted from cyclotron resonance experiments,[52,53] will now be summarized. A more detailed analysis of the valence bands of silicon and germanium can be found in articles by F. Herman,[25] Dresselhaus *et al.*,[24] and E. Kane.[54]

a. Silicon

The valence band of silicon is shown schematically in Fig. 8. Its highest point is located at $\mathbf{k} = 0$. If it were not for spin orbit coupling,

[52] G. Dresselhaus, A. F. Kip, and C. Kittel, *Phys. Rev.* **98**, 368 (1955).
[53] B. Lax, H. J. Zeiger, and R. N. Dexter, *Physica* **20**, 818 (1954).
[54] E. O. Kane, *J. Phys. Chem. Solids* **1**, 83 (1956).

this point would be sixfold degenerate, including the double degeneracy arising from spin. The simplest way to understand this degeneracy is to consider the tight binding limit, in which the wave functions corresponding to the highest point go over into atomic $3p$ functions. The spin orbit coupling lifts the degeneracy partially and leads to the situation shown in Fig. 8. The top of the valence band remains at $\mathbf{k} = 0$ and is fourfold degenerate, corresponding to atomic $J = \frac{3}{2}$ states. Slightly below it is a twofold degenerate point corresponding to $J = \frac{1}{2}$. The energy splitting λ between the two points has not yet been measured directly. Theoretical estimates give

$$\lambda \approx 0.035 \text{ ev} \tag{8.1}$$

with an uncertainty of perhaps 20%. For small values of $\mathbf{k}$, the energy E of the valence bands can no longer be expressed as a power series in k_x, k_y, and k_z. Instead, it has been shown[52,55] that it is a solution of a 6×6 secular equation of the following form:

$$\text{Det } |D_{jj'}{}^{\alpha\beta} k_\alpha k_\beta - \lambda \epsilon_j \delta_{jj'} - E\delta_{jj'}| = 0, \qquad j, j' = 1, \ldots 6. \tag{8.2}$$

Here summation of α and β over 1 to 3 is implied; ϵ_j is defined by

$$\epsilon_j = \begin{cases} 0 & j = 1, \ldots 4 \\ 1 & j = 5, 6 \end{cases} \tag{8.3}$$

the coefficients $D_{jj'}{}^{\alpha\beta}$ are certain numerical constants characteristic of the material and of the dimension $\hbar^2/m_0$; E is the energy reckoned from the top of the valence band. The 6×6 matrix $D = D_{jj'}{}^{\alpha\beta} k_\alpha k_\beta$ is Hermitian and has the following structure:[56]

$$D = \left\| \begin{array}{cccc} P/2 & R & S & 0 \\ R^* & P/6 + 2Q/3 & 0 & S \\ S^* & 0 & P/6 + 2Q/3 & -R \\ 0 & S^* & -R^* & P/2 \\ -iR^*/\sqrt{2} & -(P-2Q)/3\sqrt{2} & i\sqrt{\tfrac{3}{2}}\, R & i\sqrt{2}\, S \\ iR^*/\sqrt{2} & -i\sqrt{\tfrac{3}{2}}\, R^* & i(P-2Q)/3\sqrt{2} & iR/\sqrt{2} \end{array} \right. \left. \begin{array}{cc} iR/\sqrt{2} & -i\sqrt{2}\, S \\ -i(P-2Q)/3\sqrt{2} & i\sqrt{\tfrac{3}{2}}\, R \\ -i\sqrt{\tfrac{3}{2}}\, R^* & -i(P-2Q)3\sqrt{2} \\ -i\sqrt{2}\, S^* & -iR^*/\sqrt{2} \\ (P+Q)/3 & 0 \\ 0 & (P+Q)/3 \end{array} \right\| \tag{8.4}$$

[55] W. Shockley, *Phys. Rev.* **78**, 173 (1950).
[56] J. M. Luttinger and W. Kohn, *Phys. Rev.* **97**, 869 (1955).

where

$$\begin{aligned} P(\mathbf{k}) &= \frac{\hbar^2}{m} k^2 + (L + M)(k_x^2 + k_y^2) + 2Mk_z^2, \\ Q(\mathbf{k}) &= \frac{\hbar^2}{2m} k^2 + M(k_x^2 + k_y^2) + Lk_z^2, \\ R(\mathbf{k}) &= -(iN/\sqrt{3})(k_x - ik_y)k_z, \\ S(\mathbf{k}) &= (1/\sqrt{12})[(L - M)(k_x^2 - k_y^2) - 2iNk_xk_y]. \end{aligned} \tag{8.5}$$

For silicon, the constants have the following values:[57]

$$\begin{aligned} (L + 2M)/3 &= (-5.04 \pm 0.12)\frac{\hbar^2}{2m} \\ (L - M)/3 &= \pm(1.24 \pm 0.46)\frac{\hbar^2}{2m} \\ N &= (-7.49 \pm 0.40)\frac{\hbar^2}{2m}. \end{aligned} \tag{8.6}$$

Unfortunately, the sign of $(L - M)/3$ is not determined by classical cyclotron resonance experiments.

For $\mathbf{k} = 0$, (8.2) has a fourfold root at $E = 0$ and a twofold root at $E = -\lambda$, as it should. For sufficiently small values of k,[58] it can be factored approximately into the following 4×4 and 2×2 equations:

$$\text{Det } |D_{jj'}{}^{\alpha\beta}k_\alpha k_\beta - E\delta_{jj'}| = 0 \qquad j, j' = 1, \ldots 4 \tag{8.7a}$$

and

$$\text{Det } |D_{jj'}{}^{\alpha\beta}k_\alpha k_\beta - (E + \lambda)\delta_{jj'}| = 0 \qquad j, j' = 5, 6. \tag{8.7b}$$

Equation (8.7a) describes the vicinity of the very top of the valence band, the point B of Fig. 8, whereas Eq. (8.7b) describes the vicinity of the split-off point B'.

The quartic equation (8.7a) can be factored and solved, giving

$$E(\mathbf{k}) = Ak^2 \pm [B^2k^4 + C^2(k_x^2k_y^2 + k_y^2k_z^2 + k_z^2k_x^2)]^{\frac{1}{2}} \tag{8.8}$$

each eigenvalue being doubly degenerate. The constants A, B, C are related to L, M, and N of Eq. (8.6) as follows:

$$\begin{aligned} A &= \tfrac{1}{3}(L + 2M) + \frac{\hbar^2}{2m} \\ B &= \tfrac{1}{3}(L - M) \\ C^2 &= \tfrac{1}{3}[N^2 - (L - M)^2]. \end{aligned} \tag{8.9}$$

We see that $E \propto k^2$ in any given direction in $\mathbf{k}$ space, but since the curva-

[57] These values were kindly communicated by Dr. H. J. Zeiger of the M.I.T. Lincoln laboratories.

[58] More precisely, this means that $E(\mathbf{k}) - E(0)$ must be small compared to λ.

ture depends on the direction of $\mathbf{k}$, the surfaces of constant energy are warped.

b. Germanium

The valence band structure of germanium is completely analogous to that of silicon. However the parameters are much more accurately known in germanium. They are:[24]

$$\begin{aligned} L &= (-32.0 \pm 0.2)\frac{\hbar^2}{2m} \\ M &= (-5.3 \pm 0.2)\frac{\hbar^2}{2m} \\ N &= (-32.2 \pm 0.2)\frac{\hbar^2}{2m}. \end{aligned} \tag{8.10}$$

The sign ambiguity which existed in the case of silicon [see remark following Eq. (8.6)] has been resolved beyond doubt by band theoretical calculations.

The spin orbit splitting at $\mathbf{k} = 0$, λ, is known from infrared absorption measurements to be

$$\lambda = 0.29 \pm 0.02 \text{ ev.}^{59} \tag{8.11}$$

9. Effective Mass Equations for Acceptor States

Because the valence bands of silicon and germanium are degenerate, the effective mass theory of acceptor states is appreciably more complicated than for donor states. Instead of a single partial differential equation, Eq. (5.4), one now finds a set of coupled partial differential equations. The reader is referred to papers by Kittel and Mitchell[35] and Luttinger and Kohn[56] for derivations of these equations. They are

$$\sum_{j'=1}^{6} D_{jj'}{}^{\alpha\beta}\left(\frac{1}{i}\frac{\partial}{\partial x^\alpha}\right)\left(\frac{1}{i}\frac{\partial}{\partial x^\beta}\right)F_{j'}(\mathbf{r}) - \lambda\epsilon_j F_j + \frac{e^2}{\kappa r}F_j(\mathbf{r}) = EF_j(\mathbf{r}),$$
$$j = 1, \ldots 6. \tag{9.1}$$

Here the parameter $D_{jj'}{}^{\alpha\beta}$, ϵ_j, and λ are the same as those occurring in the secular equation (8.2) for the determination of the energy of the valence bands, $e^2/\kappa r$ represents the Coulomb interaction of the negatively charged acceptor ion and the hole, and E is the energy measured from the top of the valence band ($E > 0$ for bound states).

The functions $F_j(\mathbf{r})$ are envelope functions in the following sense. At $\mathbf{k} = 0$ there are four degenerate Bloch waves $\varphi_1, \ldots \varphi_4$, at the top

[59] A. H. Kahn, *Phys. Rev.* **97**, 1647 (1955); W. C. Dash and R. Newman, *Phys. Rev.* **99**, 1151 (1955).

of the valence band, and two Bloch waves φ_5 and φ_6 at an energy λ below the top, the latter being split off by spin orbit coupling. We take them to be normalized according to Eq. (5.5). The total wave function of an acceptor state is then given by

$$\psi(\mathbf{r}) = \sum_{j=1}^{6} F_j(\mathbf{r})\psi_j(\mathbf{r}). \tag{9.2}$$

Its normalization is

$$\int |\psi(\mathbf{r})|^2 dr = \sum_{j=1}^{6} \int |F_j(\mathbf{r})|^2 dr. \tag{9.3}$$

The functions $F_j(\mathbf{r})$ have the same general character as in the donor states. They extend smoothly over dimensions of the order of 15 A and more and can be thought of as generalizations of hydrogenic wave functions. On the other hand, the $\varphi_j(\mathbf{r})$ are rapidly varying and completely periodic functions which have p character in the vicinity of any lattice site.

10. Approximate Solutions and Comparison with Experiment

Obtaining the Eq. (9.1) for the determination of the $F(\mathbf{r})$ is a fairly simple matter. On the other hand, it is somewhat more difficult to find reasonably good solutions of these equations. Some exact solutions have been found, but only for special values of the parameters L, M, and N of Eqs. (8.5). In general, one has to rely on the variational method coupled with group theoretical considerations. A detailed analysis of the equations and their solutions will be given in a forthcoming thesis by Schechter.[60] We shall merely sketch some of the salient features here.

a. Germanium

Let us begin with a discussion of acceptor states in germanium. These have energies of 0.013 ev or less (see Table I), which are much smaller than the spin orbit separation λ of 0.3 ev. Under these conditions the functions F_5 and F_6, associated with the split-off band are negligible compared to $F_1, \ldots F_4$ and the six coupled equations (9.1) may, to a good approximation, be replaced by the four equations

$$\sum_{j'=1}^{4} D_{jj'^{\alpha\beta}} \left(\frac{1}{i}\frac{\partial}{\partial x^\alpha}\right)\left(\frac{1}{i}\frac{\partial}{\partial x^\beta}\right) F_{j'}(\mathbf{r}) + \frac{e^2}{\kappa r} F_j(\mathbf{r}) = EF_j(\mathbf{r}) \qquad j = 1, \ldots 4. \tag{10.1}$$

The solutions of (10.1) [or (9.1)] are either fourfold or twofold degenerate,

[60] At Carnegie Institute of Technology, Pittsburgh, Pennsylvania; for a preliminary note see W. Kohn and D. Schechter, *Phys. Rev.* **99**, 1903 (1955).

corresponding to the representations Γ_8 (fourfold) and Γ_6, Γ_7 (both twofold) of the tetrahedral double group.[60a]

To obtain approximate solutions of (10.1), we expand the $F_j(\mathbf{r})$ in spherical harmonics. Since the effective Hamiltonian of (10.1) is even, every solution involves only even or only odd l values. Thus we write for even l, for example,

$$F_j(\mathbf{r}) = \sum_{l=0,2\ldots} \sum_m f_{j,l,m}(r) Y_{l,m}(\theta,\varphi). \tag{10.2}$$

In the spirit of the variational method we restrict l to values less than some l_0 and use a simple trial function for the radial parts $f_{j,l,m}(r)$. Nevertheless, the problem would still be quite unmanagable if the form of (10.2) could not be determined largely from symmetry considerations.

As a specific example let us consider the ground state. If we take $l \leq 2$, we find that (10.1) has four degenerate solutions, a typical one of which has the form

$$F(\mathbf{r}) = f_1(r)\begin{pmatrix}1\\0\\0\\0\end{pmatrix} + f_2(r)\begin{pmatrix} z^2 - \frac{1}{2}(x^2+y^2)\\ 0\\ -\frac{\sqrt{3}}{2}(x^2-y^2)\\ 0\end{pmatrix} + f_3(r)\begin{pmatrix}0\\ i(x+iy)z\\ xy\\ 0\end{pmatrix} \tag{10.3}$$

where f_1, f_2, and f_3 are radial functions. The four rows in the "spinors" represent $F_1, \ldots F_4$. An approximate eigenfunction and energy is now obtained in the usual way by varying the three functions $f_i(r)$ to maximize the expectation value of the matrix Hamiltonian appearing in (10.1). We have used the functional forms

$$f_i(r) = a_i e^{-r/r_i} \tag{10.4}$$

and found an ionization energy of 0.0089 ev, with

$$r_1 = 43.3 \text{ A}, \qquad r_2 = r_3 = 33.8 \text{ A},$$

$a_1 = 1.71 \times 10^9 \text{ cm}^{-\frac{3}{2}}$, $a_2 = -2.29 \times 10^{21} \text{ cm}^{-\frac{3}{2}}$, $a_3 = 4.97 \times 10^{21} \text{ cm}^{-\frac{3}{2}}$.[61] The inclusion of the neglected functions F_5 and F_6 by perturbation theory increases this energy by less than 1%.

The theoretical value of the ionization energy, namely 0.0089 ev, is in satisfactory agreement with the measured energies, which range be-

[60a] R. H. Parmenter, *Phys. Rev.* **100**, 573, Table VII (1955); G. Dresselhaus, *Phys. Rev.* **100**, 580, Table I (1955).

[61] W. Kohn and D. Schechter.[60]

tween 0.0102 and 0.0112 ev (Table I), particularly when one recalls that any improvement of the trial function would increase the theoretical value slightly—perhaps by 5 to 10%. The size of the remaining discrepancy, about 10–20%, is just what might have been expected. For the variation of the measured energies from acceptor to acceptor is 10%, which gives an estimate of the specific effects arising from the neighborhood of the impurity ions. Such effects are of course not included in the effective mass theory.

The excited acceptor states in germanium have not yet been observed, but the positions of four low-lying levels have been calculated by Schechter[60] and are listed—together with the ground-state energy—in Table VII. These levels are labeled $2p^{(i)}$ because their envelope functions become hydrogenic $2p$ functions in the limit where the mass parameters $(L - M)$ and N become zero. The superscript (i) merely serves to differentiate them. It may be remarked that there are 12 of these "$2p$" states in all, corresponding to the 12 linearly independent products $x_i\varphi_j$. The notation "$1s$" for the ground state has a similar significance. The fourfold degeneracy of this state corresponds to the four linearly independent functions φ_j.

TABLE VII. ACCEPTOR SPECTRA IN EFFECTIVE MASS APPROXIMATION

		Energies in units of 0.01 ev		
State	Degeneracy	Silicon[a]		Germanium
$1s$	4	3.38	(4.90)	0.89
$2p^{(1)}$	4	1.26	(1.32)	0.38
$2p^{(2)}$	4	0.86	(0.89)	0.24
$2p^{(3)}$	2	0.64	(0.65)	0.19
$2p^{(4)}$	2	0.43	(0.41)	0.06

[a] The first column under "Silicon" corresponds to a negative sign of the mass parameter $B \equiv \frac{1}{3}(L - M)$, Eq. (8.6), which is slightly favored by band theory. The second column, in parentheses, corresponds to positive B.

b. Silicon

It was not an accident that we discussed germanium first, for we encounter many troubles in silicon.

First of all the mass parameters L, M, and N in silicon are known only very roughly from cyclotron resonance experiments [see Eqs. (8.61)]. Even the sign of the combination $B \equiv \frac{1}{3}(L - M)$ is uncertain, although Kane[54] has presented an argument which favors a negative sign slightly.

Second, the spin orbit splitting λ (≈ 0.035 ev) is now of the same order of magnitude as the energies of the acceptor levels (≈ 0.05 ev), so that we

expect Bloch waves associated with all six valence bands to be involved in the acceptor wave functions. Consequently we cannot reduce our six coupled equations (9.1) to four as we did for germanium.

Finally, λ has not yet been measured in silicon. The theoretical value might be in error by as much as 20%. Fortunately this uncertainty affects the calculated energy levels by only about 5%.

All calculations have been carried out using the values

$$B \equiv (L - M)/3 = \pm 1.24\hbar^2/2m.$$

Both results are listed in Table VII.

Let us now compare the results with experiment. Beginning with the ground state, we see from Table I that, apart from In, the experimental ionization energies range from 0.045 to 0.071 ev. This alone shows that the specific effects of each type of impurity are important and that an effective mass description can have only qualitative significance. Under these circumstances the "agreement" with the theoretical values of 0.034 (or 0.049) ev is about what would be expected. The comparison between theory and experiment is similar to that for Si donors (Section 6a).

In has an ionization energy of 0.16 ev which is about four or five times the effective mass value. It is therefore questionable whether it should at all be regarded as a "shallow" acceptor. For this impurity Shulman[61a] has proposed a theory which accounts for the observed energy qualitatively (within a factor of about 2) in terms of the difference between the energies necessary to break an In-Si bond and a Si-Si bond.

Turning next to the excited acceptor states in silicon, we find a rather unclear situation. We recall first that the spacings between the observed excited *donor* levels in silicon were very nearly the same for all donors [see Eq. (6.1)]. This showed that the specific effects of the different impurity ions are very small. One would expect the same to be the case for the optically excited acceptor states, whose orbits have similar sizes and whose envelope functions also have nodes at the position of the impurity ions. In our opinion it is not clear from the available experimental data whether this is so or not.

The most accurate and complete data are those of Hrostowski and Kaiser,[13] shown in Fig. 2. They are consistent with the original work of Burstein *et al.*[3] and of Newman[12a] but show considerably more detail. The general similarity of the three spectra associated with B, Al, and Ga is striking enough. The most disturbing feature is the broad two-peaked line for Ga at 500–520 cm^{-1}. As *all* other lines have much smaller widths,

[61a] R. G. Shulman, private communication.

it appears likely that this structure has, at least in part, a different origin. The location of the two B lines at 243 and 278 cm^{-1} lines has a rather large uncertainty of about ± 10 cm^{-1} for instrumental reasons. The "best" spectrum therefore is that associated with Al. It gives the following energy differences between successive excited states:

$$\text{Al:}\quad 0.37,\ 0.56,\ 0.08,\ 0.21,\ 0.12 \times 10^{-2}\ \text{ev}. \tag{10.5a}$$

For B we have

$$\text{B:}\quad \sim 0.5,\ \sim 0.35,\ 0.12,\ 0.17,\ 0.07,\ 0.06 \times 10^{-2}\ \text{ev}. \tag{10.5b}$$

If in the interest of obtaining the best possible agreement with the other spectra we arbitrarily discard the peak at 519 cm^{-1} (0.0643 ev) for Ga, we obtain

$$\text{Ga:}\quad 0.37,\ 0.58,\ 0.15,\ 0.30 \times 10^{-2}\ \text{ev}. \tag{10.5c}$$

In view of the foregoing remarks, we can only conclude that the excited B and Ga levels may agree with those of Al within about 10%.[62] From Table VII we have the following theoretical level spacings:

$$\text{Theory:}\begin{cases} B < 0 & 0.40,\ 0.22,\ 0.21 \times 10^{-2}\ \text{ev} \\ B > 0 & 0.43,\ 0.24,\ 0.24 \times 10^{-2}\ \text{ev}. \end{cases} \tag{10.6}$$

The difference between the lowest two excited levels is in fairly good agreement with the experimental results (10.5). The other spacings do not agree well. It may be noted, however, that for all three impurities the experimental value of the lowest excited state lies 1.3 to 1.4 $\times 10^{-2}$ ev below the continuum, which is in very good agreement with the theoretical estimates of 1.3×10^{-2} ev. In judging the agreement between theory and experiment, the uncertainties in the mass parameters and in λ, as well as the difficulty of obtaining accurate solutions of the six coupled equations (9.1), should be remembered.

We should like to add a remark about the intensities of the absorption lines. In atomic hydrogen the transitions from the ground state to the $2p$, $3p$, $4p$ . . . levels have relative intensities of 1.00, 0.19, 0.07, ... No such preponderance of the lowest transition is observed in the acceptor spectra. Qualitatively, this is in agreement with the fact that in the effective mass theory the four lowest lying excited levels all have "$2p$" character; that is, all these states become hydrogenic $2p$ functions when

[62] A variation of 10% of the absolute energies of the excited states corresponds of course to a much longer fractional variation in their energy differences.

the mass parameters $(L - M)$ and N are zero.[63] Quantitative theoretical calculations of the intensities are, however, very difficult as the required matrix elements also involve the ground-state wave functions which are only imperfectly known.

V. Effects of Strains and of Static Electric and Magnetic Fields

Up to this point we have discussed the structure of donor and acceptor states only in a hypothetical perfect crystal, unperturbed by internal or external strains, by electric and magnetic fields, or by lattice vibrations. In this chapter and the next we shall discuss briefly some of the interesting effects due to such perturbations.

11. Strains

Bulk samples of silicon and germanium can be subjected to elastic strains to a maximum of about 3×10^{-3}.[64] Strains of the order of 3×10^{-4} are produced fairly easily. There are also internal strains arising from dislocation lines whose magnitude is given approximately by

$$s \approx (\sqrt{n}) \times 10^{-8} \text{ cm}. \tag{11.1}$$

Here n is the number of dislocation lines per cm^2. This gives $s \approx 10^{-7}$ for the very low dislocation line density of 100 lines/cm^2, whereas one has $s \approx 10^{-4}$ for badly deformed crystals with $n = 10^8$ lines/cm^2. The effects of small strains on the properties of an impurity state associated with a single, simple band minimum or maximum are small, of the order of magnitude of s itself. However, in silicon and germanium, much larger effects should be associated with shear strains because of the degeneracy of the conduction band minima and of the valence band maximum. Isotropic compression still has a negligible effect. So far, the effects of anisotropic strain have not been confirmed experimentally. The following discussion is intended primarily to indicate possible experiments of interest.

[63] The first experimental spectrum obtained by Burstein and co-workers[3] for B had a fortuitous resemblance to a simple hydrogenic spectrum. Therefore these authors labeled their optically excited states in order as $2p$, $3p$, and $4p$. It is now known that this labeling has no significance. First of all, Hrostowski[13] has discovered an additional lower-lying state for B which removes any similarity with the hydrogenic spectrum. Also the other acceptor spectra show no such similarity. Finally, from a theoretical point of view, no resemblance to the hydrogen spectrum is to be expected. In fact, the four lowest strong transitions are all believed to be $2p$-like states.

[64] G. L. Pearson, W. T. Read, and W. L. Feldman, *Acta Met.* **5**, 181 (1957); also W. T. Read, private communication.

a. Donor States

We mentioned in Section 4 that the conduction band has several equivalent minima in both silicon and germanium, provided the lattice is not strained. For definiteness we shall talk, in the subsequent discussion, about silicon, which possesses six equivalent minima. When the crystal is subjected to a specific shear strain of magnitude s, the equivalence of these minima is destroyed. Some will be lifted and others lowered. We can write for the shift of the jth minimum

$$\Delta E(\mathbf{k}_j) = s\epsilon_j. \tag{11.2}$$

Here the ϵ_j are proportional to the deformation potential constant Ξ_u of Herring and Vogt,[65] and are of the order of magnitude of 6 ev. For shear strains, the mean shift of the six minima vanishes:

$$\sum_{j=1}^{6} \epsilon_j = 0. \tag{11.3}$$

Let us now examine the effect of such a strain on the ground-state wave function[66] of a donor. In the absence of strains and in the effective mass approximation, the ground state is sixfold degenerate. The wave functions belonging to different irreducible representations of the tetrahedral group are of the form

$$\psi^{(i)} = \sum_{j=1}^{6} \alpha_j^{(i)} \chi_j, \qquad i = 1, \ldots 6 \tag{11.4}$$

[see Eqs. (5.48), (5.45), and (5.49)]. This degeneracy is partly split as a result of corrections to the effective mass formalism, but the wave functions are still given approximately by (11.4). The lowest, nondegenerate state is denoted by $\psi^{(0)}$. If the strain s is now produced, first-order perturbation theory shows that the ground-state wave function will be modified as follows:

$$\psi_s^{(0)} = \psi^{(0)} + s \sum_{i=2}^{6} \frac{1}{E^{(0)} - E^{(i)}} \left(\sum_{j=1}^{6} \alpha_j^{(0)} \epsilon_j \alpha_j^{(i)} \right) \psi^{(i)}. \tag{11.5}$$

Here the $E^{(i)}$ are the energies of the 6 "$1s$"-like states in the absence of strain. We see that the change of the wave function depends critically on the energy differences $E^{(0)} - E^{(i)}$. It was remarked in Section 5b that

[65] C. Herring and E. Vogt, *Phys. Rev.* **101**, 944 (1956).

[66] The importance of such effects has recently been pointed out by P. J. Price, *Phys. Rev.* **104**, 1223 (1956).

the $\psi^{(i)}$ for $i > 0$ all vanish at the donor nucleus. This leads one to believe that the corresponding $E^{(i)}$ will be given more accurately than the energy $E^{(0)}$ by the effective mass value. Unfortunately, there is no experimental evidence concerning the positions of these higher lying "$1s$"-like states. Strain-induced changes of $\psi^{(0)}$ would throw some interesting light on these levels if they could be measured.

For a rough estimate, let us replace the energy $E^{(0)} - E^{(i)}$ by Δ, where, in accordance with the above remarks,

$$\Delta \equiv E_{\text{obs}} - E_{\text{eff. mass}}. \tag{11.6}$$

Then (11.5) gives for the approximate fractional change of the wave function

$$\left| \frac{\delta\psi_s^{(0)}}{\psi^{(0)}} \right| \approx s \frac{\bar{\epsilon}}{|\Delta|}. \tag{11.7}$$

Here ϵ is of the order of magnitude of the ϵ_j, that is, about 6 ev.

For Li in silicon $\Delta = 4 \times 10^{-3}$ ev (Tables I and V). Thus a moderate strain $s = 10^{-4}$ produces a 15% change in the wave function—a very substantial effect. It must be noted, however, that since all $\psi^{(i)}$ with $i > 0$ vanish at the donor nucleus, the first-order change of the wave function at this nucleus, given by (11.5) happens to vanish. The second-order change is, of course, much smaller. This suggests that one look for changes of $\psi^{(0)}$ at other points. The double resonance technique, discussed in Section 2b, which yields the values of $|\psi^{(0)}|^2$ at the sites of the Si^{29} nuclei within the donor orbit, is particularly well suited for this purpose and should yield interesting information.

For other donors in silicon, Δ is substantially larger than in the case of Li ranging from 9×10^{-3} ev in Sb to 4×10^{-2} ev in Bi. The effect of strains on the wave function should be proportionately smaller.

Let us turn next to the change in ionization energy produced by shear strains. The first-order change of the ground-state energy is given by $\sum_j \alpha_j^{(0)} \epsilon_j \alpha_j^{(0)}$ and vanishes, according to Eq. (11.3). This conclusion does not depend on the effective mass approximation but follows from general symmetry considerations. The second-order change is given by

$$\delta E^{(0)} = -\frac{s^2\bar{\epsilon}^2}{|\Delta|} \tag{11.8}$$

where $\bar{\epsilon}$ is again of the order of magnitude of ϵ_j. At the same time, the strain depresses the lowest point of the conduction band by an amount $s|\epsilon_{\max}|$, where $|\epsilon_{\max}|$ is the magnitude of the numerically largest ϵ_j. The

energy difference E_0 (>0) between the lowest point of the conduction band and the lowest bound state is, therefore, changed by

$$\delta E_0 = -s|\epsilon_{max}| + \frac{s^2\tilde{\epsilon}^2}{|\Delta|}. \tag{11.9}$$

(See Fig. 13.) Equation (11.9) represents the change of the effective ionization energy if the temperature is sufficiently low that the occupation of all bound states other than the lowest bound state and of all conducting states other than those near the lowest conduction band minimum can be neglected. These conditions may be attainable in favorable cases (e.g., P or As in germanium, $T \approx 6°K$, $s = 3 \times 10^{-4}$). Hall measurements could then give interesting information about Δ as well as $|\epsilon_{max}|$ and hence provide values of the deformation potential constant Ξ_u.[67]

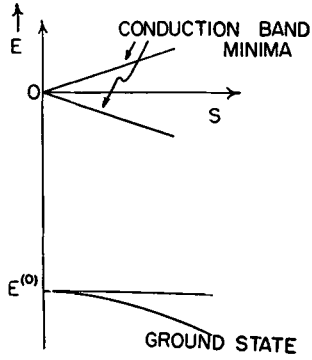

FIG. 13. Effect of a shear strain s on the donor ionization energy (schematic). Note the linear s dependence of the conduction band minima, and the quadratic s dependence of the ground state level $E^{(0)}$.

We shall now consider the excited donor levels briefly. Each optically active level belongs to the vector representation T_1 of the tetrahedral group and is threefold degenerate in the absence of strains. A strain $s = 3 \times 10^{-4}$ may be expected to split this degeneracy by roughly $s\epsilon' \approx 2 \times 10^{-3}$ ev, and should result in an appreciable broadening (or even splitting) of the absorption lines.

Finally, a remark about internal strains. In "good" crystals, having dislocation densities less than 10^5 lines/cm^2, the effects of internal strains ($s \lesssim 10^{-5}$) on donor states will be very insignificant.

b. Acceptor States

We have seen in Section 10, that in the absence of strains all acceptor levels are either twofold or fourfold degenerate. Let us consider first the ground state, which is fourfold degenerate according to theory. Under a shear strain s, this level will split into two twofold degenerate levels.[68] The magnitude of the splitting is

$$\Delta E = s\epsilon. \tag{11.10}$$

[67] F. J. Morin, T. H. Geballe, and C. Herring, *Phys. Rev.* **105**, 525 (1956).

[68] According to Kramers' theorem (time reversal), at least twofold degeneracy remains unless there is an external magnetic field.

Here ϵ is an effective deformation potential constant, which depends on the geometry of the strain and is of the order of 10 ev. The center of gravity of the four states remains fixed to first order in s. The wave functions associated with each of the two levels depend on the nature of the strain.

We see from Eq. (11.1) that internal strains arising from dislocations produce a broadening of the lowest acceptor level of the order of

$$\Delta E \approx (10^{-8} \text{ cm}) \sqrt{n}\, \epsilon. \tag{11.11}$$

This is about 3×10^{-5} ev for a typical dislocation density of $10^6/\text{cm}^2$. Such broadening has important consequences for the magnetic properties of acceptor states, which are discussed further in Sections 13b and 15c.

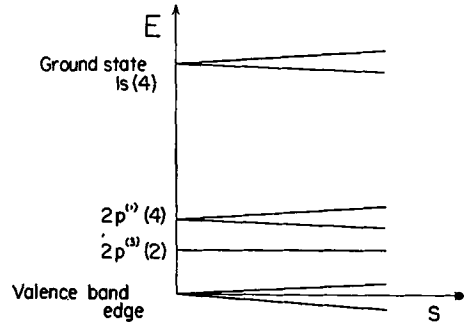

FIG. 14. Effect of a shear strain s on acceptor levels (schematic). Fourfold levels are split to first order in s, twofold levels remain unaffected.

We turn next to the excited levels. Again, the fourfold degenerate ones will be split, according to Eq. (11.10), whereas the twofold degenerate ones remain unsplit and fixed to first order in s.

Let us now see how an external shear strain $s = 10^{-4}$ affects the infrared absorption spectrum of acceptor states. At temperatures for which $kT \gg s\epsilon$ (i.e., $T \gg 10°\text{K}$) both levels which arise from the ground state will be occupied equally. Hence all absorption lines should broaden by an amount $s\epsilon \approx 10^{-3}$ ev due to this cause. Furthermore, transitions to fourfold degenerate final states will be broadened additionally as a result of the splitting of the final states. On the other hand, if $kT \ll s\epsilon$ (i.e., $T \ll 10°\text{K}$), only the lowest of the two ground-state levels will be occupied. Therefore, the transitions to the twofold excited states are not broadened by strain; however, the corresponding photon energy is increased as a consequence of the depression of the ground-state level. Transitions to fourfold excited states will be both broadened, because of the splitting of the excited level, and shifted to higher photon energies, as a result the movement of the ground state. (See Fig. 14.) Such

observations would be interesting for two reasons. They would enable one to determine the degeneracy of the various acceptor states experimentally, and they would give one interesting information about the magnitude of the deformation potential constant ϵ.

Finally, one would expect to obtain interesting effects in strained crystals with the use of *polarized* light. For, as we have just seen, only two of the four normally degenerate states are occupied when $kT \ll s\epsilon$. These have preferred "orientations" relative to the crystal axes. Consequently, the intensities of the absorption line depend strongly on the direction of polarization of the incident light.

12. Stark Effect

The Stark effect, which would be of great interest, has not yet been observed in impurity states. However, since it might be detected in silicon, we shall sketch the theory briefly.

Let $\mathcal{E}$ be the electric field at the impurity center and let us assume its direction is along the z axis. The perturbation Hamiltonian then is

$$\delta V = -\mathcal{E}ez. \tag{12.1}$$

The wave functions of the bound carrier have the form

$$\psi^{(i)}(\mathbf{r}) = \sum_j F_j^{(i)}(\mathbf{r})\varphi_j(\mathbf{r}) \tag{12.2}$$

in the effective mass theory of donor or acceptor states. [See Eqs. (5.48), (5.46), and (9.21).] The matrix element of δV between two such states is given by

$$(\psi^{(i)},\delta V\psi^{(i')}) = \sum_j (F_j^{(i)},\delta V F_j^{(i')}). \tag{12.3}$$

Since the effective mass Hamiltonians for both types of states are invariant under inversion and there are no accidental degeneracies (such as the $2s$, $2p$ degeneracy in hydrogen) the first-order Stark effect vanishes.[69]

The second-order Stark shift for a nondegenerate state $\psi^{(i)}$ is given by

$$\begin{aligned}\delta_2 E^{(i)} &= \sum_{i'} \frac{|(\psi^{(i)},\delta V\psi^{(i')})|^2}{E^{(i)} - E^{(i')}} \\ &= \frac{(\psi^{(i)},(\delta V)^2\psi^{(i)})}{\Delta E} = \frac{(F^{(i)},(\delta V)^2F^{(i)})}{\Delta E}\end{aligned} \tag{12.4}$$

[69] The *full* Hamiltonian of the impurity problem has only tetrahedral symmetry and is not invariant under inversion. As a result, if the effective mass theory is seriously in error, states belonging to the representations T_1, T_2, and Γ_8 can have an appreciable first-order Stark effect. This might possibly be of significance for the acceptor ground state (Γ_8) in silicon.

where ΔE is an average of the denominators $E^{(i)} - E^{(i')}$. The shifts will be of a similar order of magnitude for degenerate states. Let us make a rough estimate of $\delta_2 E$, by taking a simple hydrogenic wave function of "radius" R for $F^{(i)}$. This gives

$$\delta_2 E \approx 3 \frac{(e\mathcal{E}R)^2}{\Delta E}. \tag{12.5}$$

We see now that impurity states are tempting objects for Stark effect measurements because of their large orbits and small energy separations.

It may be possible to apply about 3000 volts/cm to silicon at low temperatures without causing appreciable heating.[70] The Stark shift of the ground state is, however, only about 3×10^{-5} ev for such a field, that is, is quite negligible. On the other hand, let us consider a highly excited state and take as representative values

$$\begin{aligned} R &\approx 10^{-6} \text{ cm} \\ \Delta E &\approx 5 \times 10^{-3} \text{ ev}. \end{aligned} \tag{12.6}$$

Equation (12.5) then gives

$$\delta_2 E \approx 5 \times 10^{-3} \text{ ev} \tag{12.7}$$

which is comparable to the unperturbed energies of the excited states. A shift of this magnitude could, of course, be detected easily by infrared absorption experiments, which have a precision of about 10^{-4} ev (see Section 1b).

Stark effect measurements, if feasible, would give us very useful information about the excited impurity states. For example, the fourfold degenerate acceptor levels should be split into two, whereas the twofold degenerate levels should be shifted but not split.

13. Static Magnetic Susceptibility

We mentioned in Section 3 that impurity states have been found to make significant contributions to the magnetic susceptibilities of the samples in which they are located. So far, reliable measurements of the contributions of effectively isolated impurity states are not yet available but may be expected before very long. It is interesting to see the results we would anticipate on the basis of our current models.

a. Donor States

In the case of silicon, we know from magnetic resonance measurements that the g factor of the donor electron is 1.9985,[71] which is very close to

[70] I am indebted to Dr. Burstein for a conversation on this question.
[71] G. Feher, private communication.

that of the free electron. This shows that spin-orbit coupling plays a negligible role. The corresponding spin paramagnetism is therefore, to an excellent approximation,

$$\chi_p = \frac{{\mu_0}^2}{kT} \text{ per donor} \tag{13.1}$$

where μ_0 is the Bohr magneton.

In germanium spin resonance experiments have not been successful so that we do not have an experimental source for the g factor. The spin orbit interaction in germanium may be expected to be about an order of magnitude larger than in silicon because of the higher atomic number. However, we still expect the free electron paramagnetism (13.1) for all practical purposes, provided the ground state has no orbital degeneracy.

To obtain the diamagnetic susceptibility, one must, as usual, calculate the energy of the donor state up the second order in the magnetic field H. We have seen that the ground-state wave function is a linear combination of several wave packets, each composed of Bloch waves arising from the vicinity of one of the conduction band minima. It is easily seen that these may be treated independently in the effective mass approximation, provided that we eventually take the average of their magnetic energies. Since we know from the symmetry of the donor states that the susceptibility is a scalar, we can also calculate the magnetic energy of any one wave packet and average it over all directions of the field.

The Hamiltonian for one of the packets is

$$\mathfrak{H} = \frac{1}{2m_l}\left(\frac{\hbar}{i}\frac{\partial}{\partial_z} - \frac{e}{c}A_z\right)^2 + \frac{1}{2m_t}\left[\left(\frac{\hbar}{i}\frac{\partial}{\partial_x} - \frac{e}{c}A_x\right)^2 + \left(\frac{\hbar}{i}\frac{\partial}{\partial y} - \frac{e}{c}A_y\right)^2\right] - \frac{e^2}{\kappa r} \tag{13.2}$$

where

$$\mathbf{A} = \tfrac{1}{2}(\mathbf{H} \times \mathbf{r}). \tag{13.3}$$

To second order in H, the magnetic energy consists of two contributions: δE_1, the expectation value of the second-order terms of (13.2); and δE_2, the contribution of the first-order terms of (13.2) evaluated by second-order perturbation theory.[71a] This calculation has been carried out[42] using the form

$$F = \frac{1}{(\pi a^2 b)^{\frac{1}{2}}} e^{-\sqrt{(x^2+y^2)/a^2+z^2/b^2}} \tag{13.4}$$

as an approximation for the unperturbed ground state. The result is

$$\chi = -\frac{e^2}{6c^2}\left[\frac{a^2}{m_l} + \frac{a^2+b^2}{m_t} - \frac{(a^2/m_l - b^2/m_t)^2}{(a^2/m_l + b^2/m_t)}\right] \text{ per donor.} \tag{13.5}$$

[71a] δE_2 vanishes for $m_l = m_t$.

With the effective mass values given in Eqs. (4.2) and (4.3) and the characteristic lengths a and b given in Eq. (6.5), one finds

$$\begin{aligned} \chi_{Si} &= -2.3 \times 10^{-26} \text{ cgs units/donor} \\ \chi_{Ge} &= -2.7 \times 10^{-25} \text{ cgs units/donor.} \end{aligned} \tag{13.6}$$

The fact that the observed ionization energies E_{obs} are somewhat larger than those predicted by the effective mass theory shows that the actual orbits are smaller than the theoretical ones. This may be allowed for, in a rough way, by multiplying (13.6) by the factor $(E_{eff.\ mass}/E_{obs})$, where $E_{eff.\ mass}$ is 0.029 ev in silicon and 0.0092 ev in germanium.

b. Acceptor States

The acceptor ground state is fourfold degenerate (Section 10) in the absence of any perturbations. A magnetic field will lift this degeneracy completely to first order in H and depress the center of gravity to second order. The first effect gives rise to a temperature-dependent paramagnetism, the second to a temperature-independent diamagnetism.

The paramagnetic susceptibility of germanium has been calculated by the writer.[42] The computation is quite lengthy because of the matrix nature of the effective mass Hamiltonian. One interesting feature is that the susceptibility depends not only on the usual inverse mass parameters L, M, N of Eq. (8.10), but also very critically on another parameter κ, introduced by Luttinger[72] in a study of the quantum effects of cyclotron resonance. We have estimated that κ lies in the range $\kappa = (3.5 \pm 0.2)$. This gives the paramagnetic susceptibility

$$\begin{aligned} \chi_p &= R \frac{{\mu_0}^2}{kT} \\ R &= 0.35 \pm 0.25. \end{aligned} \tag{13.7}$$

If κ is omitted from the calculation one finds $R \approx 35$. The small result contained in (13.7) is the result of seemingly fortuitous cancellations of large terms and may be altered appreciably by a slight improvement of the trial wave functions [see Eq. (10.3)].

Before applying the result (13.7) to a real sample containing acceptors, the effect of internal strains arising from dislocations must be included. We have seen in Section 11b, that a strain s splits the fourfold degenerate ground state into two twofold degenerate levels, the separation being the order of $s \times$ (10 ev). If this splitting is comparable to kT, the susceptibility is reduced below the value given by (13.7). The critical dislocation density is about 10^6 dislocation lines/cm^2 at $T = 1.4$°K.

[72] J. M. Luttinger, *Phys. Rev.* **102**, 1030 (1956); this constant κ (which occurs nowhere else in the present article) has no relation to the dielectric constant.

No calculations of the paramagnetic susceptibility of acceptor states in silicon are available. The diamagnetic susceptibility of acceptor states has not yet been calculated for either silicon or germanium.

14. Magnetic Field Dependence of the Ionization Energy

Both the energy of the impurity ground state and of the lowest point of the continuum are significantly altered in a strong magnetic field. Since the two energy changes are of different magnitudes in general, a change of the ionization energy will result.

For the case of a simple donor state, described by an isotropic mass m^*, this effect has been calculated by Yafet *et al.*[73] The strength of the magnetic field is conveniently measured by the dimensionless parameter

$$\gamma \equiv \frac{\hbar\omega/2}{E_0}. \tag{14.1}$$

Here E_0 is the ionization in the absence of a magnetic field (the "Rydberg constant" of the impurity state) and $\hbar\omega/2$ is the zero point energy of a conduction electron of mass m^* in the external magnetic field:

$$\frac{\hbar\omega}{2} = \frac{e\hbar}{2m^*c} H. \tag{14.2}$$

The interaction energy of the electron spin with the magnetic field is identical in the ground state and the conduction band and can be neglected. Instead we are concerned here with the effects of the magnetic field on the orbital energy of the electron.

The energy of the edge of the conduction band is simply given by (14.2). For weak fields, the energy of the ground state can be calculated by perturbation theory, but it must be obtained variationally for stronger fields ($\gamma \gtrsim 1$). Yafet *et al.* find that the wave function is appreciably compressed in all directions for $\gamma \gtrsim 1$, but least of all in the direction paralleled to the field. The total energy of the ground state rises, but somewhat less rapidly than that of the conduction band edge. The resulting effect is an increase of the ionization energy. The numerical results are shown in Fig. 15.

The magnetic field for which $\gamma = 1$ is given by

$$H = 2 \times 10^9 \left(\frac{m^*}{\kappa m_0}\right)^2 \text{ oersted}. \tag{14.3}$$

It is of the order of 10^5 oersted for the impurity states in germanium and is about 10^6 oersted for those in silicon.

[73] Y. Yafet, R. W. Keyes, and E. N. Adams, *J. Phys. Chem. Solids* **1**, 137 (1956).

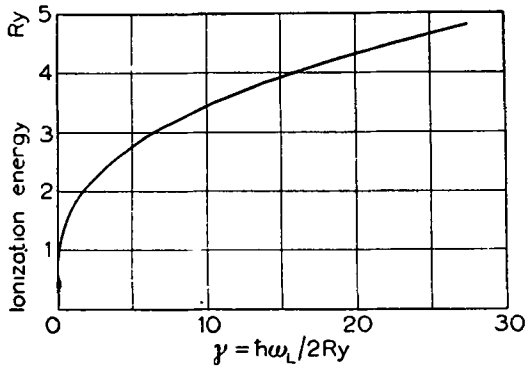

FIG. 15. Energy difference between the lowest state of the conduction band and the impurity ground state as function of the parameter γ which characterizes the strength of the magnetic field. (After Yafet *et al.*[73])

The field is only about 1300 oersted in InSb where $m^*/m_0 = 0.013$ and $\kappa = 16$. Keyes and Sladek[74] have measured the Hall coefficient in this material in fields up to 80,000 oersted at 4.2°K and have found a decrease of the Hall coefficient by a factor of 50. This large effect can be understood qualitatively in terms of the change of ionization energy. Quantitative comparison with theory was not possible because of overlap effects between neighboring impurities.

VI. Interaction with Lattice Vibrations

In this chapter we shall discuss briefly the effects of lattice vibrations on impurity states. These effects have been studied in considerable detail by Lax and Burstein[75] in order to account for the observed widths of the infrared absorption lines (Section 1b). We believe also that the failure of the attempts to observe spin resonance of acceptor states may arise from the interaction of the electrons with the lattice vibrations.

15. Absorption Line Widths

The following discussion is a brief summary of the work of Lax and Burstein.[75] Further details may be found in the original papers.

Let us focus our attention on B acceptor states in silicon for the sake of definiteness. First we want to show that it is reasonable to use the adiabatic approximation, which is familiar from the theory of molecular vibrations. If E_0 denotes the ionization energy of the impurity electron, a typical electronic frequency is

$$\omega_{el} = E_0/\hbar \approx 7 \times 10^{13}\ \text{sec}^{-1}. \tag{15.1}$$

[74] R. W. Keyes and R. J. Sladek, *J. Phys. Chem. Solids* **1**, 143 (1956).
[75] M. Lax and E. Burstein, *Phys. Rev.* **100**, 592 (1955).

Now the electron interacts appreciably only with acoustic lattice vibrations whose wavelengths are comparable to the diameter of the orbit. Vibrations of much longer wavelength describe a slow rigid displacement of the entire impurity state and the interaction with short wavelength vibrations is weakened by partial cancellation of the contributions from different parts of the impurity orbit. The frequency of the important lattice vibrations is approximately

$$\omega_{\text{vib}} \approx 1.3 \times 10^{13}\ \text{sec}^{-1} \tag{15.2}$$

which is appreciably less than the electronic frequency (15.1). This is the condition for the validity of the adiabatic approximation.

In this approximation, we can write the wave functions describing the acceptor ground state and the lattice vibrations in the form

$$\Psi_n^{(0)} = \psi^{(0)}(\mathbf{r})\Phi_n^{(0)}(a_1, a_2, \ldots). \tag{15.3}$$

Here $\psi^{(0)}(\mathbf{r})$ is the electronic wave function, a_1, a_2, . . . are a set of coordinates describing the lattice vibrations and the $\Phi_n^{(0)}$ are the wave functions describing various degrees of excitation of the normal modes of vibration. A weak dependence of $\psi^{(0)}$ on the a_i has been neglected. Similarly, the wave functions corresponding to the ith excited acceptor state have the form

$$\Psi_n^{(i)} = \psi^{(i)}(\mathbf{r})\Phi_n^{(i)}(a_1, a_2, \ldots). \tag{15.4}$$

It is an essential consequence of the electron-lattice interaction that the normal modes $\Phi_n^{(i)}$ are slightly different from $\Phi_n^{(0)}$.

a. Optical Absorption Line Widths

The matrix element for an optical transition from $\Psi_n^{(0)}$ to $\Psi_{n'}^{(i)}$ contains, apart from the electronic dipole matrix element, the scalar product

$$(\Phi_n^{(0)}, \Phi_{n'}^{(i)}). \tag{15.5}$$

Now if the sets of lattice functions $\Phi_n^{(0)}$ and $\Phi_n^{(i)}$ were identical (no electron lattice interaction), optical transitions could occur only for $n = n'$, i.e., between states which do not differ in their vibrational motion. The line would then be sharp[75a] and given by $h\nu = E^{(i)} - E^{(0)}$, the difference between the *electronic* energies. Because of the difference between the sets $\Phi_n^{(0)}$ and $\Phi_n^{(i)}$, however, transitions can take place from a given n to a variety of n', that is, transitions in which the vibrational state of motion changes. This gives rise to a line width. The shape of

[75a] The width due to the finite lifetime of the excited state is negligible.

the absorption line associated with the electronic level i clearly is given by

$$I^{(i)}(h\nu)d(h\nu) = C \sum_{n,n'} e^{-E_n^{(0)}/kT} |(\Phi_n^{(0)}, \Phi_{n'}^{(i)})|^2$$
$$h\nu \leq E_{n'}^{(i)} - E_n^{(0)} \leq h\nu + d(h\nu) \quad (15.6)$$

in which I is the intensity, C is a normalization constant, and $E_n^{(0)}$ and $E_{n'}^{(i)}$ are the *total* energies of the states $\Psi_n^{(0)}$ and $\Psi_{n'}^{(i)}$ respectively. The Boltzmann factor describes the occupation probability of the initial state.

The magnitude of the electron-lattice interaction, which is responsible for the difference between the vibrational states associated with $\psi^{(0)}$ and $\psi^{(i)}$, can be estimated from the observed lattice mobility of the electrons, since electronic wave functions of long wavelengths are involved in both cases. In this way Lax and Burstein[75] find that the root-mean-square widths of all optical absorption lines are given at $T = 0°$ by

$$\Delta(h\nu) = \left(\frac{2\hbar}{Mva^{*4}q_{\max}^3}\right)^{\frac{1}{2}} E_1 \quad (15.7)$$

to a good approximation. Here the symbols have the following meaning: M is the mass per unit cell (twice the atomic mass of silicon); v is the velocity of sound; a^* is an effective Bohr radius of the impurity *ground* state; $q_{\max}$ is the maximum propagation vector of a lattice vibration; and E_1 is an effective deformation potential constant[76] of the order of 15 ev. Substituting numbers into (15.7), Lax and Burstein obtain $\Delta(h\nu) = 3.6 \times 10^{-3}$ ev, in fairly good agreement with the observed widths which are 0.5 to 1.0×10^{-3} ev.

The reason why *all* lines are expected to have nearly the same width is as follows. The interaction between electron and lattice vibrations is so much weaker in the more extended excited states than in the ground state, that the overlap integrals $(\Phi_n^{(0)}, \Phi_n^{(i)})$ are essentially independent of i. The experimental line widths are in fact approximately equal for almost all lines (see Fig. 2).

Finally, a word about the temperature dependence of the line width. As we mentioned before, the modes of vibration which interact most strongly with the electron in its ground state have a wavelength in the vicinity of $\lambda \approx \pi a^*$. These modes become appreciably excited at a temperature T_0 given by

$$kT_0 \approx \hbar\omega(\lambda). \quad (15.8)$$

For $a^* = 13.5$ A this gives $T_0 \approx 94°$K. Near this temperature the width of the line would be expected to increase appreciably [see the Boltzmann

[76] J. Bardeen and W. Shockley, *Phys. Rev.* **80**, 72 (1950).

factor in (15.6)]. This prediction agrees with the observed dependence of the B line widths upon temperature reported in ref. 75.

Thus the main features of the widths of the optical absorption line appear to be well understood in terms of the electron lattice interaction.

b. Absence of Paramagnetic Resonance in Acceptor States

Although careful attempts have been made to find a paramagnetic resonance in p-type silicon,[64] similar to the donor state resonance described in Section 2a, the experiments have not been successful. We should like to suggest two possible reasons.

We have seen in Section 15a, that the optical absorption lines of impurity states have a width associated with electron-lattice interactions of the order of 10^{-3} ev at $T = 0°$. For quite similar reasons, we would expect the paramagnetic absorption lines corresponding to transitions between the magnetic acceptor levels to have comparable line widths.[77] Since the separation of the magnetic levels is given by $h\nu = 4 \times 10^{-5}$ ev at 9000 mc/sec, we see that such a broadening would destroy the resonance completely.

Apart from this effect, there is the broadening due to internal strains. Such strains split the ground-state level into two doublets whose mean energy difference $\overline{\Delta E}$ is given approximately by Eq. (11.11). For a typical dislocation line density of 10^5 lines/cm^2, $\overline{\Delta E} \approx 3 \times 10^{-5}$ ev, just comparable to the magnetic energy at 9000 mc.

In concluding this article, I would like to express my thanks to Dr. E. Burstein, Dr. G. Feher, Dr. H. J. Hrostowski, Dr. G. Picus, Dr. P. Price, Dr. W. T. Read, and Dr. W. W. Tyler for providing me with unpublished results and for numerous helpful comments.

Appendix: Deep-Lying Impurity Levels

The shallow impurity levels which are the main subject of this article are associated with elements of groups III and V and with Li. Besides these shallow levels a great many deeper ones have been identified which are introduced by other impurity elements.

The levels identified clearly in germanium are shown in Fig. 16. Work on these levels has been carried out largely by a group at the General Electric Research Laboratory.[78] These investigators have presented

[77] This question has not yet been fully analyzed. Preliminary studies indicate a very complicated line shape with peaks whose widths may be an order of magnitude smaller than the optical width of 10^{-3} ev.

[78] The first deep-lying level in germanium was established by W. C. Dunlap, Jr., *Phys. Rev.* **85**, 945 (1952). The review by Burton[1] and a recent article by W. W. Tyler and H. H. Woodbury, *Phys. Rev.* **102**, 647 (1955) contain references to subsequent work.

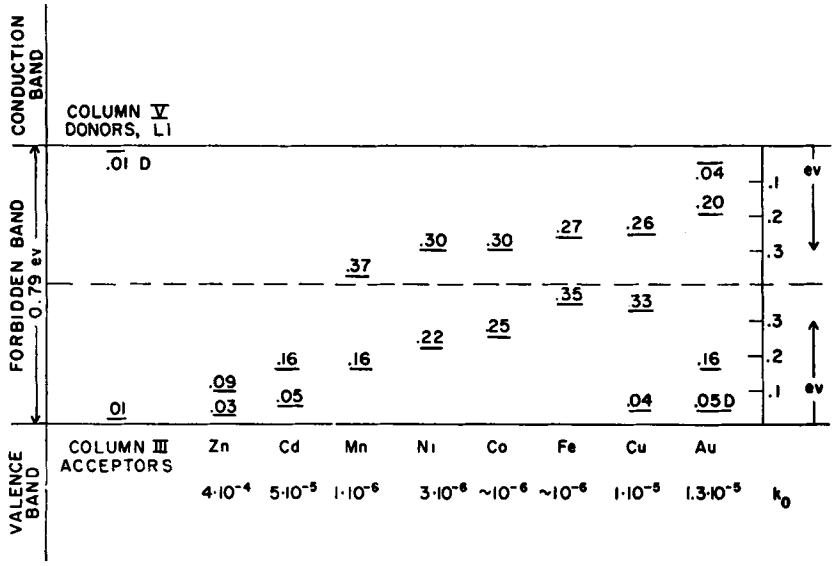

FIG. 16. Firmly established impurity levels in germanium. All levels are acceptors unless marked by D for donor. k_0 is the distribution coefficient, i.e., the ratio of the concentration in the solid phase to that in the liquid phase. (Courtesy W. W. Tyler and H. H. Woodbury.)

strong evidence that the monovalent metals Cu and Au can accept 1, 2, or 3 electrons,[79] whereas the divalent metals Zn and Cd can accept 1 or 2. This suggests that the maximum number of acceptable electrons is determined, at least in part, by the number required to complete all tetrahedral bonds. Apart from this feature, the detailed structure of the deep-lying states is not yet known.

In silicon, Au introduces a donor level 0.35 ev above the valence band[80,81] and an acceptor level 0.54 ev below the conduction band.[81] Zn introduces an acceptor level 0.31 ev above the valence band.[82] Mn produces a donor level 0.53 ev below the conduction band.[83] Moreover, Fe introduces at 0.45 ev above the valence band and one 0.55 ev below the conduction band.[84] Finally, we mention again that one of the group III elements, namely In, introduces a fairly deep acceptor level 0.16 ev above the valence band.

[79] H. H. Woodbury and W. W. Tyler, *Phys. Rev.* **105**, 84 (1957).
[80] E. A. Taft and F. H. Horn, *Phys. Rev.* **93**, 64 (1950).
[81] C. B. Collins, R. O. Carlson, and C. J. Gallagher, *Phys. Rev.* **105**, 1168 (1957).
[82] C. S. Fuller and F. J. Morin, *Phys. Rev.* **105**, 379 (1957).
[83] R. O. Carlson, *Phys. Rev.* **104**, 937 (1956).
[84] C. B. Collins, *Bull. Am. Phys. Soc.* [2] **1**, 48 (1956).