Theory of Donor States in Silicon*

W. KOHN, *Department of Physics, Carnegie Institute of Technology, Pittsburgh, Pennsylvania and Bell Telephone Laboratories, Murray Hill, New Jersey*

AND

J. M. LUTTINGER, *Department of Physics, University of Michigan, Ann Arbor, Michigan and Bell Telephone Laboratories, Murray Hill, New Jersey*

(Received February 2, 1955)

By using the recently measured effective masses for n-type Si, $m_1=0.98\ m$ and $m_2=0.19$ m, approximate solutions of the resulting effective mass Schroedinger equation are obtained. The accuracy of the solutions was tested in the limiting cases where $m_2/m_1=1$ and 0 respectively. The nature of the resulting states and their degeneracy is discussed in some detail, taking into account the fact that the conduction band of Si has six equivalent minima. Experimentally measured ionization energies show that the effective mass theory is seriously in error in the case of the ground state. This error is attributed to failure of the effective mass theory near the donor nucleus, and allowance for this failure is made in the case of higher states. This leads finally to a theoretical spectrum for the electrons bound by P, As, or Sb donors.

1. INTRODUCTION

IT is well known that P, As, or Sb—all Group V atoms—when added to Si, act as so-called electron donors. The following qualitative picture has been well confirmed by experiment[1]: The donor atom has five electrons outside of closed shells. Four of these serve to complete bonds with neighboring atoms. The fifth finds itself in the Coulomb field of the remaining charge (reduced by the dielectric constant of the medium). In this field there exist bound states, with ionization energies of a few hundredths of a volt. The purpose of this paper is to examine these states theoretically in the light of existing experimental information.

This investigation was stimulated by the spin resonance experiments of Fletcher *et al.*[2] on n-type Si, and we have examined only donor states in Si in some detail. However, most of our considerations apply *mutatis mutandis* also to n-type Ge. On the other hand, acceptor states in Si and Ge represent a rather different situation, because of the special structure of the top of the valence band, to which we hope to come back in a later publication.

2. EXPERIMENTAL SITUATION

The following experimental data are of special significance for our purposes:

(a) Ionization Energies

These have been determined from the temperature dependence of the Hall coefficient. The results are given in Table I. We have included in this table results for Li donors which very likely occupy interstitial positions in the lattice.

(b) Band Structure

Recent measurements of piezo-resistance,[3] magneto-resistance[4] and cyclotron resonances[5] lead conclusively to the following picture. The conduction band of Si has six minima located on the (1,0,0) and equivalent axes. If the minimum on the (1,0,0) axis is denoted by $\mathbf{k}^{(1)}=(k_0,0,0)$, the energy near this point is given by

$$E=E_0+(\hbar^2/2m_1)(k_x-k_0)^2+(\hbar^2/2m_2)(k_y^2+k_z^2), \quad (2.1)$$

where

$$m_1=0.98m, \quad m_2=0.19m, \quad (2.2)$$

m being the free electron mass.

3. EFFECTIVE MASS THEORY

The donor state wave functions satisfy a Schrödinger equation of the form

$$\left(-\frac{\hbar^2}{2m}\nabla^2+V(\mathbf{r})+U(\mathbf{r})\right)\psi(\mathbf{r})=E\psi(\mathbf{r}), \quad (3.1)$$

where $V(\mathbf{r})$ is the periodic potential of an electron in a perfect Si lattice, and $U(\mathbf{r})$ is the additional potential due to the replacement of a Si atom by a donor ion. We choose the donor nucleus as coordinate origin. Then, for large r,

$$U(\mathbf{r})=-e^2/\kappa r, \quad (3.2)$$

TABLE I. Experimental ionization energies of donor states.

Donor	Ionization energy (ev)
P	0.044[a]
As	0.049[a,b]
Sb	0.039[a]
Li	0.033[a,c]

[a] F. J. Morin *et al.*, Phys. Rev. **96**, 833 (A) (1954).
[b] E. Burstein *et al.*, J. Phys. Chem. **57**, 849 (1953).
[c] C. S. Fuller and J. A. Ditzenberger, Phys. Rev. **91**, 193 (1953).

* Supported in part by the Office of Naval Research. Preliminary reports have appeared in Bull. Am. Phys. Soc. **30**, No. 2, 38 (1955) and Phys. Rev. **97**, 1721 (1955).

[1] E.g., G. L. Pearson and J. Bardeen, Phys. Rev. **75**, 865 (1949).
[2] R. C. Fletcher *et al.*, Phys. Rev. **94**, 1392 (1954); **95**, 844 (1954).
[3] C. Smith, Phys. Rev. **94**, 42 (1954).
[4] G. L. Pearson and C. Herring, Physica (to be published).
[5] R. N. Dexter *et al.*, Phys. Rev. **96**, 222 (1954).

where κ is the dielectric constant of Si, having the value 12.0. Because of this large reduction of the Coulomb field of the donor ion, one expects bound states around the donor having dimensions large compared to the lattice spacing. In this limit it can be shown[6] that (3.1) has solutions of the form

$$\psi^{(i)}(\mathbf{r})=F^{(i)}(\mathbf{r})\psi(\mathbf{k}^{(i)},\mathbf{r}), \quad i=1, 2, \cdots 6, \tag{3.3}$$

where $\psi(\mathbf{k}^{(i)},\mathbf{r})$ is the Bloch wave at the minimum $\mathbf{k}^{(i)}$ assumed to be nondegenerate. The $F^{(i)}(\mathbf{r})$ satisfy the effective mass equations,

$$\left[-\frac{\hbar^2}{2m_1}\frac{\partial^2}{\partial z_i^2}-\frac{\hbar^2}{2m_2}\left(\frac{\partial^2}{\partial x_i^2}+\frac{\partial^2}{\partial y_i}\right)-\frac{e^2}{\kappa r}-\epsilon\right]F^{(i)}(\mathbf{r})=0, \tag{3.4}$$

where m_1 and m_2 are defined in (2.1) and z_i lies along the direction of $\mathbf{k}^{(i)}$. The function $\psi^{(i)}(\mathbf{r})$ is normalized, provided we set

$$\int_{\text{unit cell}}|\psi(\mathbf{k}^{(i)},\mathbf{r}|^2 d\mathbf{r}=\Omega, \tag{3.5}$$

Ω being the volume of the unit cell, and

$$\int_{\text{all space}}|F^{(i)}(\mathbf{r})|^2 d\mathbf{r}=1. \tag{3.6}$$

Finally it is convenient to choose the phases of the $\psi(\mathbf{k}^{(i)},\mathbf{r})$ such that the $\psi(\mathbf{k}^{(i)},0)$ are all real. The eigenvalue ϵ of (3.4) represents the energy of the state (3.3) relative to the bottom of the conduction band.

In the following two sections we shall study the solutions (3.3) and their energy spectrum. Actually, however, the assumption of a sufficiently large orbit—which leads to (3.3)—is not always too well satisfied and we shall come back to this difficulty in Sec. 6.

4. GROUND STATE

We begin by examining the lowest lying solution of (3.4). We introduce

$$a_B=\hbar^2\kappa/m_2e^2, \quad \epsilon_0=m_2e^4/2\hbar^2\kappa^2, \tag{4.1}$$

as units of length and energy respectively, and write

$$\gamma\equiv m_2/m_1(<1). \tag{4.2}$$

This leads to the equation

$$\left[-\left(\gamma\frac{\partial^2}{\partial z^2}+\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}\right)-\frac{2}{r}-\epsilon\right]F(\mathbf{r})=0, \tag{4.3}$$

where we have dropped the subscripts and superscripts i. For $\gamma<1$ this equation does not completely separate and no exact solutions have been found. It is therefore of interest to consider two limiting cases:

[6] J. M. Luttinger and W. Kohn, Phys. Rev. **97**, 869 (1955).

(i) $m_1=m_2$ (or $\gamma=1$)

This is the hydrogenic case, leading to the following normalized solution:

$$F(r)=\pi^{-\frac{1}{2}}e^{-r}, \tag{4.4}$$

with an energy

$$\epsilon=-1. \tag{4.5}$$

(ii) $m_1\gg m_2$ (or $\gamma\ll 1$)

This we call the adiabatic limit for the following reason. Equation (4.3) may be regarded as the Schrödinger equation describing three one-dimensional particles of masses $1/\gamma(\gg 1)$, 1, and 1, respectively, interacting through the potential $-2/(x^2+y^2+z^2)^{\frac{1}{2}}$. Since one of these masses is much larger than the other two, we follow the well known adiabatic procedure of first solving the equation

$$\left[-\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}\right)-\frac{2}{(x^2+y^2+z^2)^{\frac{1}{2}}}-\lambda(z)\right]\Phi(x,y;z)=0, \tag{4.6}$$

where z is treated as parameter (*cf.* electronic motion for fixed nuclear position in molecular problems), and then

$$[-\gamma d^2/dz^2+\lambda(z)-\epsilon]u(z)=0 \tag{4.7}$$

(*cf.* nuclear motion in molecular problems). Let us now introduce

$$\rho=(x^2+y^2)^{\frac{1}{2}}, \quad \varphi=\tan^{-1}(y/x). \tag{4.8}$$

Then (4.6) becomes

$$\left[\frac{1}{\rho}\frac{\partial}{\partial\rho}\rho\frac{\partial}{\partial\rho}+\frac{1}{\rho^2}\frac{\partial^2}{\partial\varphi^2}-\frac{2}{(\rho^2+z^2)^{\frac{1}{2}}}-\lambda(z)\right]\Phi(\rho,\varphi;z)=0. \tag{4.9}$$

The ground state is of course independent of φ.

In the extreme adiabatic limit ($\gamma\ll 1$), the lowest eigenvalue of (4.9) is

$$\lim_{\gamma\to 0}\epsilon=\lambda(0), \tag{4.10}$$

where $\lambda(0)$ is the lowest eigenvalue of Eq. (4.9) with $z=0$. The corresponding solution is

$$\Phi=e^{-2\rho}; \quad \lambda(0)=-4. \tag{4.11}$$

Thus we see that if m_2 is kept fixed while m_1 increases from m_2 to ∞, the ionization energy changes from $-\epsilon_0$ to $-4\epsilon_0$.

It is also of interest to calculate $u(z)$ in the extreme adiabatic limit. For this purpose we require $\lambda(z)$, Eq. (3.13), for small z. A simple perturbation calculation gives

$$\lambda(z)=-4+32|z|+\cdots. \tag{4.12}$$

Substituting into (4.7) leads to

$$\left[-\frac{d^2}{dz^2}+\frac{32}{\gamma}|z|-\left(\frac{4+\epsilon}{\gamma}\right)\right]u(z)=0. \tag{4.13}$$

By changing the variable to

$$\xi \equiv z(32/\gamma)^{\frac{1}{3}}, \quad u(z)=v(\xi), \tag{4.14}$$

Eq. (4.13) becomes

$$[-d^2/d\xi^2+|\xi|-\eta]v(\xi)=0, \tag{4.15}$$

where

$$\eta=\frac{1}{(32)^{\frac{2}{3}}\gamma^{\frac{1}{3}}}(4+\epsilon). \tag{4.16}$$

Equation (4.10) can be solved in terms of Bessel functions. The lowest solution is symmetric in ξ and is given by

$$\eta=1.02,$$

$$v(\xi)=C(\eta-\xi)^{\frac{1}{2}}J_{-1/3}[\tfrac{2}{3}(\eta-\xi)^{\frac{3}{2}}]+J_{1/3}[\tfrac{2}{3}(\eta-\xi)^{\frac{3}{2}}], \quad 0\leq\xi\leq\eta, \tag{4.17}$$

$$v(\xi)=C(\sqrt{3}/\pi)(\xi-\eta)^{1/2}K_{1/3}[\tfrac{2}{3}(\xi-\eta)^{\frac{3}{2}}], \quad \eta\leq\xi\leq\infty, \tag{4.18}$$

where

$$C^2\approx 5.3, \tag{4.19}$$

when v is normalized such that

$$\int_{-\infty}^{\infty}[v(\xi)]^2 d\xi=1. \tag{4.20}$$

From (4.16) and (4.17) we obtain the following refinement of (4.11):

$$\gamma\to 0: \epsilon=-4+10.3\gamma^{1/3}, \tag{4.21}$$

$$F=[(8/\pi)^{1/2}(32)^{1/6}\gamma^{-1/6}]e^{-2\rho}v[(32/\gamma)^{\frac{1}{3}}z]. \tag{4.22}$$

The qualitative aspect of this function is of interest: As $\gamma\to 0$ (i.e., $m_1\to\infty$) the dependence on y and z tends to a constant function, $e^{-2\rho}$, which extends over dimensions of the order of a_B, the Bohr radius associated with m_2. On the other hand the extent of the wavefunction in the x-direction is of the order of $(m_2/m_1)^{1/3}a_B$, i.e., the function becomes pancake-like as $m_1/m_2\to\infty$. At the origin we have from (4.22), (4.17), and (4.19):

$$\gamma\to 0: F(0)\to A\gamma^{-1/6}, \tag{4.23}$$

$$A\cong 1.98. \tag{4.24}$$

Variational Method

Neither the hydrogenic nor the adiabatic case are of much direct practical value for the actual case $m_2/m_1=0.19$, which lies in between. However a simple variational function, namely,

$$F^{(v)}=(\pi a^2 b)^{-\frac{1}{2}}\exp[-(\rho^2/a^2+z^2/b^2)^{\frac{1}{2}}],^{13} \tag{4.25}$$

represents, with a suitable choice of a and b, a very good approximation to the actual situation for all $\gamma\leq 1$. [We use again (4.3) and (4.2) as units of length and energy.] Its accuracy can be nicely tested in both limiting cases. $F^{(v)}$ is evidently exact for $\gamma=1$. For $\gamma\to 0$, one finds, by minimizing the energy,

$$a\to 4/3\pi=0.424, \quad b\to\tfrac{1}{3}(4/\pi)^{2/3}\gamma^{1/3},$$
$$\epsilon\to-\tfrac{2}{3}(3\pi/4)^2=-3.701, \tag{4.26}$$
$$F^{(v)}(0)\to 2.12\gamma^{-1/6}.$$

The agreement with the true limiting function (4.23) and (4.24) and energy (4.11) is remarkably good. Figure 1 shows a plot of the energy as a function of $\gamma=m_2/m_1$; as calculated with the variational function (4.25). The error is only 7.5 percent at $\gamma=0$ and no doubt smaller for all intermediate γ. For Si, with $\gamma=0.19$ we find

$$\epsilon=-0.029 \text{ ev}. \tag{4.27}$$

Degeneracy

We remarked in Sec. 3 that in the effective-mass approximation, the Schrödinger equation has six equivalent and evidently degenerate solutions of the form (3.3). These functions must form a basis for a representation of the symmetry group of our system, *viz.*, the tetrahedral group T_d [14] about the donor atom. The six functions (3.3) behave under operation of T_d like $x, -x$; $y, -y$; $z, -z$ respectively. T_d has five classes and we list in Table II the characters of the present representation, $\{1s\}$, as well as of the 5 irreducible representations of T_d:

TABLE II. Characters of tetrahedral point-group representations.

Representation \ Group element	E	$8C_3$	$3C_2$	$6\sigma_d$	$6S_4$
A_1	1	1	1	1	1
A_2	1	1	1	−1	−1
E	2	−1	2	0	0
T_1	3	0	−1	1	−1
T_2	3	0	−1	−1	1
$1s$	6	0	2	2	0

Comparison with the characters of the irreducible representations shows that $\{1s\}$ decomposes into

$$\{1s\}=A_1+E+T_1, \tag{4.28}$$

which are one-, two-, and three-dimensional respectively. The degeneracy of the six functions (3.3) obtains of course only in our effective mass approximation. In reality this degeneracy will be partially lifted according to (4.28) and one singly-, one doubly-, and one triply-degenerate level will be produced. If we number the minima in the (1,0,0), (−1,0,0), (0,1,0), (0,−1,0), (0,0,1), (0,0,−1) directions by 1, 2, ⋯, 6, respectively, the correct linear combinations are given by

$$\Psi_i=\sum_{j=1}^{6}\alpha_i^{(j)}\psi^{(j)}, \quad i=1, 2, \cdots, 6, \tag{4.29}$$

[13] This function was independently introduced by M. Lampert, and by G. Mitchell and C. Kittel (to be published).

[14] See, e.g., Eyring, Walter, and Kimball, *Quantum Chemistry* (Jonn Wiley and Sons, Inc., New York, 1944), p. 388.

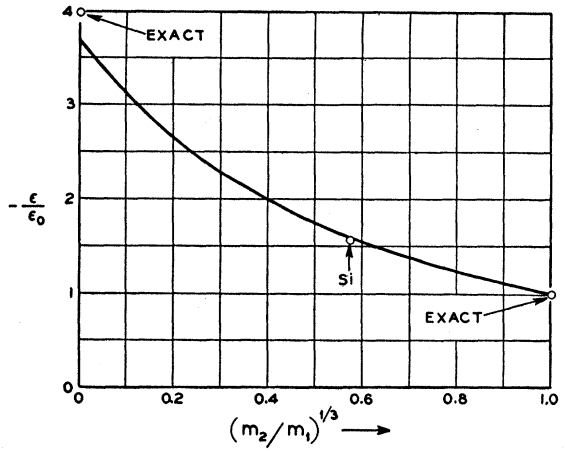

FIG. 1. Energy of ground state as function of mass ratio, $\gamma=m_2/m_1$, calculated with the trial function (4.25); $\epsilon_0=m_2e^4/2\hbar^2\kappa^2$.

where

$$\begin{aligned}
&\alpha_1{}^{(j)}=(1/\sqrt{6})(1,1,1,1,1,1), \quad (A_1)\\
&\left.\begin{aligned}\alpha_2{}^{(j)}&=\tfrac{1}{2}(1,1,-1,-1,0,0),\\ \alpha_3{}^{(j)}&=\tfrac{1}{2}(1,1,0,0,-1,-1),\end{aligned}\right\}(E)\\
&\left.\begin{aligned}\alpha_4{}^{(j)}&=(1/\sqrt{2})(1,-1,0,0,0,0),\\ \alpha_5{}^{(j)}&=(1/\sqrt{2})(0,0,1,-1,0,0),\\ \alpha_6{}^{(j)}&=(1/\sqrt{2})(0,0,0,0,1,-1).\end{aligned}\right\}(T_1)
\end{aligned} \tag{4.30}$$

Since all $\psi^{(j)}(0)$ are equal (see Sec. 3), we notice that only the nondegenerate function Ψ_1 has a finite value at the donor nucleus. It therefore will have a much larger hyperfine structure than the other states. The experiments of Fletcher *et al.*,[2] who have measured the hyperfine splitting of electron spin resonances in donor-doped Si at liquid helium temperatures, show that in fact Ψ_1 must be the ground state. Other considerations which will be discussed in Sec. 6 support the same conclusion.

5. EXCITED STATES

We now wish to obtain some orientation concerning the excited solutions of Eq. (4.3). This equation is invariant under rotations about the z axis and under inversion through the origin. Hence the solutions of (4.3) may be labelled by an integral magnetic quantum number m [$F(\mathbf{r})\sim e^{im\varphi}$] and a parity $P=\pm1$.

It is of interest to trace, qualitatively, the spectrum of (3.4) as we pass from the hydrogenic case to the adiabatic limit.

(a) Hydrogenic Limit

For $\gamma=1$, the spectrum is given:

$$\epsilon=-1/n^2, \quad n=1,2,\cdots, \tag{5.1}$$

where n is the principal quantum number. The states of even and odd l have $P=\pm1$, respectively.

The case where γ is just less than one, $\gamma=1-\delta$, must be treated by degenerate perturbation theory, the perturbation being $\delta\partial^2/\partial z^2$ [see Eq. (4.3)]. This mixes states of given n and m and all l of a given parity, e.g., the states $n=2$, $m=0$, $l=0$ and 2, and the accidental degeneracy due to different l and m will be lifted. (The double degeneracy corresponding to $\pm m$ remains of course.)

(b) Adiabatic Limit

Proceeding as in the last section, we first solve (4.9), obtaining, for $z=0$,

$$\begin{aligned}
&\Phi_{m,t}(\rho,\varphi)=\rho^{|m|}L_{|2m|+t}{}^{|2m|}\left(\frac{4}{|2m|+2t+1}\rho\right)e^{im\varphi},\\
&\lambda_{m,t}(0)=-4/(|2m|+2t+1)^2,\\
&\qquad m=0,\pm1,\pm2,\cdots, \quad t=0,1,2,\cdots,
\end{aligned} \tag{5.2}$$

where $L_\alpha{}^\beta(x)$ denotes an associated Laguerre polynomial.

In principle, $\lambda(z)$ for $z\neq0$ can be similarly found and used to solve (4.7). Since $\lambda(z)$ is a symmetric function of z the solutions of (4.7) will be either even or odd. They may be labeled by a quantum number q representing the number of nodes. Thus the total wave function becomes, in the adiabatic limit,

$$F_{m,t;q}(\mathbf{r})=\Phi_{m,t}(\rho,\varphi)u_{m,t;q}(z), \tag{5.3}$$

and

$$\epsilon_{m,t;q}=\lambda_{m,t}(0)+\mu_{m,t;q}, \tag{5.4}$$

where $\mu_{m,t;q}$ is defined by the above equation. Evidently

$$\lim_{\gamma\to0}\mu_{m,t;q}=0. \tag{5.5}$$

The parity of $F_{m,t;q}$ is given by

$$P=(-1)^{m+q}. \tag{5.6}$$

(c) Connection Between Hydrogenic and Adiabatic Limit

We now raise the following question: What happens to the hydrogenic level scheme as γ decreases from 1 to 0. This change to γ may be considered as the gradual appearance of a perturbation, $\partial^2/\partial z^2$. Since this perturbation commutes with rotations about the z-axis and inversion, m and P are conserved. However, all states with the same m and P are mixed. In Table III we list, for $\gamma=1$, the states corresponding to given m and P in order of increasing energy. We shall use the symbol R_i to denote a set of states with given m and P. The R_i

TABLE III. Hydrogenic levels grouped according to values of m and P.

R_1 $m=0, P=1$	R_2 $m=0, P=-1$	R_3 $m=1, P=-1$	R_4 $m=1, P=1$
1*s*	2*p*	2*p*	3*d*
2*s*	3*p*	3*p*	4*d*
3*s*, 3*d*	4*p*, 4*f*	4*p*, 4*f*	5*d*, 5*g*
4*s*, 4*d*	5*p*, 5*f*	5*p*, 5*f*	6*d*, 6*g*

belong to different representations of the symmetry group of the Hamiltonian in (4.3). As the perturbation $\partial^2/\partial z^2$ comes into play, two levels belonging to different R_i may cross freely but those in the same R_i "interact" and hence cannot cross.

Let us begin by considering R_1. For $\gamma=1-\delta$ (δ being a small positive number), its levels are all discreet and in a definite order

$$R_1: \quad 1s,\ 2s,\ (3s,d)^{(1)},\ (3s,d)^{(2)},\ \cdots, \tag{5.7}$$

where the superscripts (1) and (2) indicate certain linear combinations. These levels represent the lowest, second lowest, etc., level corresponding to $m=0$, $P=1$. As γ approaches zero, they remain the lowest, second etc. level corresponding to $m=0$, $P=1$, as a result of the noncrossing principle; i.e., they go over into the adiabatic levels, $m=0$, $t=0$ and $q=0, 2, 4, \cdots$. [In the adiabatic limit these all lie lower than the levels $m=0$, $t>0$, $q=0, 2, 4, \cdots$, which belong to the same representation, see (5.2) and (5.5)]. Thus we have the following correspondences:

$$\begin{array}{lllll} & & m & t & q \\ R_1: & 1s,\ m=0 & \rightarrow 0 & 0 & 0 \\ & 2s,\ m=0 & \rightarrow 0 & 0 & 2 \\ & (3s,d)^{(1)},\ m=0 & \rightarrow 0 & 0 & 4 \\ & (3s,d)^{(2)},\ m=0 & \rightarrow 0 & 0 & 6. \end{array} \tag{5.8}$$

Similarly for the other groups:

$$\begin{array}{lllll} & & m & t & q \\ R_2: & 2p,\ m=0 & \rightarrow 0 & 0 & 1 \\ & 3p,\ m=0 & \rightarrow 0 & 0 & 3 \\ R_3: & 2p,\ m=1 & \rightarrow 1 & 0 & 0 \\ & 3p,\ m=1 & \rightarrow 1 & 0 & 2 \\ R_4: & 3d,\ m=1 & \rightarrow 1 & 0 & 1 \\ & 4d,\ m=1 & \rightarrow 1 & 0 & 3. \end{array} \tag{5.9}$$

We see then that all hydrogenic levels go over, in the adiabatic limit, into adiabatic levels with the same m, and $t=0$.[15] Therefore, as γ diminishes from 1 to 0, the energy of a hydrogenic level n, l, m changes from $\epsilon=-1/n^2$ to $\epsilon=-4/(2|m|+1)^2$ [see Eq. (5.2)]. Figure 2 illustrates the correspondence for the first few hydrogenic levels.

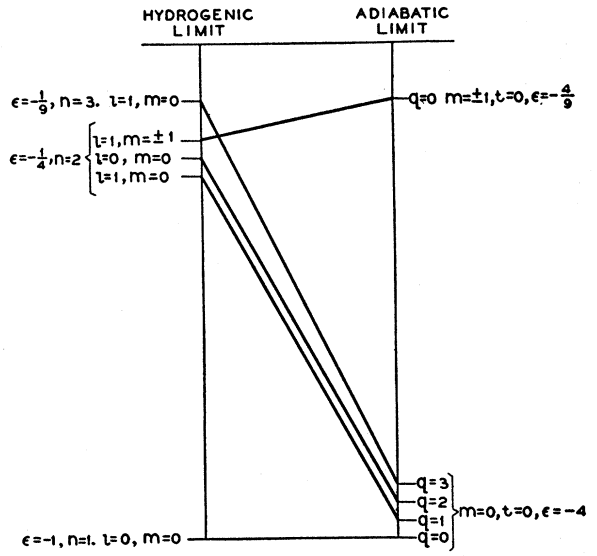

FIG. 2. Correspondence between energy levels in the hydrogenic limit, $m_2/m_1=1$, and the adiabatic limit, $m_2/m_1=0$ (schematic). ϵ is in units of $\epsilon_0=m_2e^4/2\hbar^2\kappa^2$. Note that ϵ is drawn on different scales in the two limits.

(d) Variational Calculations

To determine the energies of the first few excited states of donors in Si, we have made variational calculations which we have tested in the limits $\gamma=1$ and 0. This comparison enabled us to make slight corrections and to estimate the error of our calculation.

TABLE IV. Energies in the effective mass approximations.[a]

State	Trial function	γ	E_{var}	E_{ex}	E_I
$1s$ (0,0,0)	$C\exp[-(\rho^2/a^2+z^2/b^2)^{\frac{1}{2}}]$	0	−3.70	−4.00	
		0.19	−1.60		−1.63±0.03
		1.00	−1.00	−1.00	
$2p$, $m=0$ (0,0,1)	$Cz\exp[-(\rho^2/a^2+z^2/b^2)^{\frac{1}{2}}]$	0	−3.47	−4.00	
		0.19	−0.597		−0.63±0.03
		1.00	−0.250	−0.250	
$2s$, $m=0$ (0,0,2)	$(C_1+C_2\rho^2+C_3z^2)\exp[-(\rho^2/a^2+z^2/b^2)^{\frac{1}{2}}]$	0	−3.34	−4.00	
		0.19	−0.456		−0.49±0.03
		1.00	−0.245	−0.250	
$3p$, $m=0$ (0,0,3)	$(C_1+C_2\rho^2+C_3z^2)z\exp[-(\rho^2/a^2+z^2/b^2)^{\frac{1}{2}}]$	0	−3.22	−4.00	
		0.19	−0.288		−0.32±0.03
		1.00	−0.111	−0.111	
$2p$, $m=1$ (1,0,0)	$Cx\exp[-(\rho^2/a^2+z^2/b^2)^{\frac{1}{2}}]$	0	−0.434	−0.444	
		0.19	−0.324		−0.33±0.005
		1.00	−0.250	−0.250	

[a] All energies are expressed in units of ϵ_0, Eq. (4.1), which for Si has the value 0.0179 ev.

[15] The adiabatic levels with $t>0$ have no counterpart in the hydrogenic spectrum.

The results are listed in Table IV, together with the exact results in the two limits. The description of a state by, say, $2p$, $m=1$ means that as $\gamma\to 1$, the state becomes a hydrogenic $2p$ state, with $m=1$ along the direction of $\mathbf{k}^{(i)}$. The numbers in parentheses are the quantum numbers m, t, q in the adiabatic limit. The following notation is used:

$$\begin{aligned} E_{\text{var}} &= \text{energy calculated variationally with the indicated trial function;} \\ E_{\text{ex}} &= \text{exact energy;} \\ E_f &= \text{final energy, i.e., variational energy, corrected by comparison with exact energies, including an error estimate;} \\ \gamma &= m_2/m_1. \end{aligned} \tag{5.10}$$

(e) Degeneracy

Considerations similar to those of Sec. 4 show that each effective mass wave function $F^{(i)}(\mathbf{r})\psi(\mathbf{k}^{(i)},\mathbf{r})$ is one of a 12-fold degenerate set for $m\neq 0$, and a 6-fold degenerate set for $m=0$. It is of interest to find the linear combinations of these functions which transform according to the irreducible representations of the tetrahedral group. For this purpose we given in Table V the characters of the tetrahedral point group for the representations corresponding to different m values.

Comparison with Table II gives the following decomposition:

$$\begin{aligned} |m|=0: &\quad A_1+E+T_1, \\ |m| \text{ even } (\neq 0): &\quad A_1+A_2+2E+T_1+T_2, \\ |m| \text{ odd}: &\quad 2T_1+2T_2. \end{aligned} \tag{5.11}$$

Thus deviations from the effective mass model will split each degenerate set into groups according to the following scheme:

$$\begin{aligned} m=0: &\quad 6=1+2+3, \\ m \text{ even } (\neq 0): &\quad 12=1+1+2+2+3+3, \\ m \text{ odd}: &\quad 12=3+3+3+3. \end{aligned} \tag{5.12}$$

6. DEVIATIONS FROM THE EFFECTIVE-MASS FORMALISM

The shortcomings of the effective-mass formalism are best seen from the following comparisons with experiment. The effective-mass theory predicts an ionization energy of 0.029 ev for all donors, whereas the experimental values for P, As, and Sb range from 0.039 to 0.049 ev (see Table I). The discrepancies undoubtedly arise largely due to the breakdown of our assumptions in the immediate vicinity of the donor atom, for the following two reasons: (1) The perturbing potential $U(\mathbf{r})$ changes from $-e^2/\kappa r$ to $-\Delta Ze^2/\kappa r$ near $r=0$, where ΔZ is the excess charge of the donor nucleus over that of Si. (2) The effective-mass formalism does not apply in this region since the fractional change of $U(\mathbf{r})$ in a typical lattice distance, a, is large.

TABLE V. Characters of tetrahedral point group representations.

$\lvert m\rvert$ \ Group element	E	$8C_3$	$3C_2$	$6\sigma_d$	$6S_4$
0	6	0	2	2	0
even, $\neq 0$	12	0	4	0	0
odd	12	0	−4	0	0

We do not have a good quantitative theory to take these effects into account. But it is clear that they will be largest for the states of greatest amplitude near the donor nucleus. In the remainder of this section we shall discuss the effect of deviations from the effective mass formalism on the most important low-lying levels.

"1s"-states

We begin by considering the six states (4.29), (4.30), whose modulating functions $F^{(i)}(\mathbf{r})$ becomes "1s"-like when $m_2/m_1\to 1$. The lowest of these is almost certainly Ψ_1 corresponding to the identity representation. Since all $\Psi^{(i)}(0)$ are equal, the 6 terms in the sum (4.29) are all in phase at $r=0$ so that $\Psi_1(0)$ would be expected to be quite large. This is borne out by the observed large hyperfine splitting of the lowest donor state.[2] It is therefore not surprising that the energy of this state, as calculated by the effective mass approximation is substantially in error. As was noted previously, the observed ionization energies for group V donors are from 35 to 70 percent greater than the effective mass energy. We have discussed this state in some detail elsewhere.[16] The other five "1s"-states evidently vanish at $r=0$ [see Eq. (4.30)].

For orientation let us study these six 1s-states in the tight binding approximation. In this case, in the vicinity of the donor nucleus,

$$\psi^{(i)}=\alpha R_s(r)+\beta\frac{\mathbf{k}^{(i)}\cdot\mathbf{r}}{k^{(i)}r}R_p(r), \tag{6.1}$$

where $R_s(r)$ and $R_p(r)$ are the normalized radial atomic $3s$ and $3p$ functions, and α and β are numerical constants of comparable order of magnitude. The p-part of this function contains the factor $\mathbf{k}^{(i)}\cdot\mathbf{r}$, because it must be invariant under rotations which leave $\mathbf{k}^{(i)}$ invariant.

With the help of (6.1), we can make a plausible guess as to the positions of the 1s-levels belonging to the representations A_1, E, and T_1. By (4.29) and (4.30), we see that in the vicinity of the donor nucleus (and extending over about one atomic volume), we have:

$$A_1:\quad \Psi_1(\mathbf{r})=\sqrt{6}\alpha R_s(r), \tag{6.2}$$

$$E:\begin{cases}\Psi_2(\mathbf{r})=0,\\ \Psi_3(\mathbf{r})=0,\end{cases} \tag{6.3}$$

$$T_1:\begin{cases}\Psi_4(\mathbf{r})=\sqrt{2}\beta(x/r)R_p(r),\\ \Psi_5(\mathbf{r})=\sqrt{2}\beta(y/r)R_p(r),\\ \Psi_6(\mathbf{r})=\sqrt{2}\beta(z/r)R_p(r).\end{cases} \tag{6.4}$$

[16] W. Kohn and J. M. Luttinger, Phys. Rev. **97**, 883 (1955).

Therefore we would expect Ψ_1 to be lowest because the electron spends a relatively larger time near the nucleus, where the attraction is greater. Similarly we would expect the states belonging to T_1 to have somewhat lower energies than the effective-mass energy, but because of their p-nature to lie substantially above Ψ_1. Finally the energy of the two states belonging to E should be rather accurately described by the effective-mass energy.

"2s"-states

Among the excited levels, we expect the greatest deviation from the effective-mass energy in the case of that $2s$-state which belongs to the identical representation A_1 of T_d (see Table II). Since the energy difference ΔE between this state and the corresponding $1s$-state is small ($\sim$0.03 ev), their wave functions will be practically identical in the vicinity of the donor atom.[17] It is therefore possible to estimate the shift of this level relative to its effective mass value from the experimentally observed shift of the corresponding $1s$-level.

To make the analysis manageable let us replace the two different masses m_1 and m_2 by a single mean mass m^*. A reasonable choice is to take

$$m^*=0.31\, m, \tag{6.5}$$

which in the effective-mass approximation gives the same ground-state energy as the measured masses. Now for $r \geq a$, where a is a typical lattice distance, the wave functions $\Psi_{n,s}$ for both the $1s$- and $2s$-states, belonging to the identity representation, will be given by the effective-mass expressions

$$\Psi_{ns}=F_{ns}(r)\sum_{i=1}^{6}\psi(\mathbf{k}^{(i)},\mathbf{r}); \quad n=1, 2, \quad r\geq a. \tag{6.6}$$

where because of the single mass m^* all $F^{(i)}(\mathbf{r})$ are identical for $i=1, \cdots 6$ [see Eqs. (3.3), (3.4), and (4.29)]. The functions $F_{ns}(\mathbf{r})$ satisfy equations of the hydrogenic form

$$\left(-\frac{\hbar^2}{2m^*}\nabla^2-\frac{e^2}{\kappa r}-\epsilon_{ns}\right)F_{ns}(\mathbf{r})=0, \tag{6.7}$$

where ϵ_{ns} is the actual energy of the ns-state. Since the ϵ_{ns} do not exactly coincide with the hydrogenic levels, $-m^*e^4/2\hbar^2\kappa^2n^2$, the F_{ns} which vanish at $r=\infty$ would become infinite if continued to $r=0$. However, for $r<a$ the effective-mass function (6.6) no longer is reliable. On the other hand, in view of the negligible value of $\Delta\epsilon(\equiv\epsilon_{2s}-\epsilon_{1s})$ we have in this region

$$\Psi_{1s}=\Psi_{2s}, \quad r\leq a. \tag{6.8}$$

[17] It can be easily seen that the relevant criterion is $\Delta E \ll(\hbar^2/2m)a_s^2$, where a_s is the radius of the Wigner-Seitz sphere. In the present case $\Delta E\approx 0.02(\hbar^2/2m)a_s^2$.

TABLE VI. Correction of $2s$-energy.

Element	$1s$-energy[a] Eff. mass	$1s$-energy[a] Observed	δ	$\frac{1}{(1-\delta/2)^2}$	$2s$-energy[a] Eff. mass	$2s$-energy[a] Corrected
P	−1.0	−1.52	0.18	1.21	−0.30	−0.36
As	−1.0	−1.69	0.23	1.27	−0.30	−0.38
Sb	−1.0	−1.35	0.14	1.16	−0.30	−0.35
Li	−1.0	−1.14	0.07	1.07	−0.30	−0.32

[a] All energies are in units of $\bar{\epsilon}=-0.029$ ev.

By combining (6.6) and (6.8), we find

$$\left[\frac{\partial F_{2s}/\partial r}{F_{2s}}\right]_{r=a}=\left[\frac{\partial F_{1s}/\partial r}{F_{1s}}\right]_{r=a}. \tag{6.9}$$

Since a is much smaller than the "Bohr" radius a^* of this problem ($a^*=\kappa\hbar^2/m^*e^2\approx$20A), the situation is entirely analogous to that encountered in the optical spectrum of the alkalies. One finds therefore in the same way that

$$\epsilon_{ns}=-\bar{\epsilon}/(n-\delta)^2, \tag{6.8}$$

where $\bar{\epsilon}$ is the "Rydberg" ($=m^*e^4/2\hbar^2\kappa^2$) and δ is the quantum defect which is independent of n. (6.8) may also be written as

$$\epsilon_{ns}=-\frac{\bar{\epsilon}}{n^2}\frac{1}{(1-\delta/n)^2}, \tag{6.7}$$

the last factor being the correction factor due to departures from a simple Coulomb law in the interior region. Thus δ may be determined from a comparison of $\bar{\epsilon}$ ($=-0.029$ ev) and the observed ground-state energy, and the correction factor for the $2s$-state can then be calculated from (6.7), and applied to the effective-mass value of the $2s$-energy, -0.0088 ev (see Table IV). The results are listed in Table VI.

We see that in all cases the correction of the $2s$-level is fairly small.

It can be shown that the replacement of m_1 and m_2 by m^* does not materially affect the correction we have just calculated.

The remaining five $2s$-states are expected to be much less shifted (see the discussion of the $1s$-levels), presumably slightly downward.

"p"-states

These states have the property that the modulating functions $F^{(i)}(\mathbf{r})$ vanish at the donor nucleus. Since the relevant length in these functions is of the order a_B (which is about 20 A), their amplitude will be completely negligible in that region near the donor nucleus where the effective mass theory breaks down. Therefore we would expect the effective mass theory to be excellent for "p"-states.†

† See note added in proof at the end of the paper.

TABLE VII. Level scheme of donor states in silicon.

State	Representations of T_d [a]	Number of degenerate [b] states	Eff. mass theory	(Energy in ev) $\times 10^2$ [c] P	As	Sb
$1s, m=0$	A_1	1	-2.9 ± 0.1	-4.4 [d]	-4.9 [d]	-3.9 [d]
$1s, m=0$	$E+T_1$	5	-2.9 ± 0.1	-3.2 ± 0.3	-3.3 ± 0.4	-3.1 ± 0.2
$2p, m=0$	A_1+E+T_1	6	-1.13 ± 0.06	-1.13 ± 0.06	-1.13 ± 0.06	-1.13 ± 0.06
$2s, m=0$	A_1	1	-0.88 ± 0.06	-1.06 ± 0.10	-1.11 ± 0.10	-0.94 ± 0.08
$2s, m=0$	$E+T_1$	5	-0.88 ± 0.06	-0.93 ± 0.11	-0.95 ± 0.13	-0.90 ± 0.08
$2p, m=+1$	$2T_1+2T_2$	12	-0.59 ± 0.01	-0.59 ± 0.01	-0.59 ± 0.01	-0.59 ± 0.01
$3p, m=0$	A_1+E+T_1	6	-0.57 ± 0.06	-0.57 ± 0.06	-0.57 ± 0.06	-0.57 ± 0.06

[a] See reference 14.
[b] These states are only approximately degenerate, consisting in general of several strictly degenerate sets, as appears from the second column. Spin degeneracy is not included.
[c] The indicated errors represent estimated uncertainties within the framework of the present model.
[d] Experimental.

7. SUMMARY AND DISCUSSION

On the basis of the previous sections, our picture of the donor levels in Si is represented in Table VII and Fig. 3, which latter shows graphically the level scheme for one of the donors. In summary we may recall that this picture is based on the following model:

(1) The conduction band of Si has six minima in the (1,0,0) and equivalent directions. At each minimum the band is nondegenerate. The minima are not too close to $k=0$.

(2) The effective masses are $m_1=0.98\ m$, $m_2=0.19\ m$ (twice).

(3) Except in the immediate vicinity of the donor atom, the donor states are described by functions of the form

$$\Psi=\sum_{j=1}^{6}\alpha^{(j)}F^{(j)}(\mathbf{r})\psi(\mathbf{k}^{(j)},\mathbf{r}), \tag{7.1}$$

where the $F^{(j)}(\mathbf{r})$ are modulating functions satisfying the effective-mass equations (3.4), the $\psi(\mathbf{k}^{(j)},\mathbf{r})$ are the Bloch functions at the minima $\mathbf{k}^{(j)}$ of the conduction band, and the $\alpha^{(j)}$ are constants obeying the requirements of tetrahedral symmetry.

(4) Shifts of the energy levels relative to their values in the effective-mass theory are attributed to failure of the effective-mass formalism in the vicinity of the donor atom. From the known shift of the ground state, the shifts of other levels are estimated.

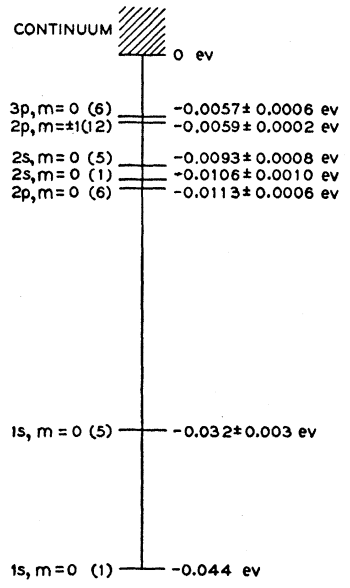

FIG. 3. Spectrum of bound states around a phosphorus-donor. Numbers in parentheses indicate number of approximately degenerate states; spin degeneracy is not included. See Table VII.

We should like to mention explicitly that the effects of lattice vibrations on the position and width of the levels has not been considered.

All the states can be excited thermally. However, only the p-states are expected to play a major role in the optical transitions from the ground state. Oscillator strengths for these transitions will be published shortly.

There exist as yet no experimental data with which our results can be quantitatively compared. It is hoped that when such data become available their agreement or disagreement with our results will throw some light on the nature of donor states.

This work was begun at the Bell Telephone Laboratories in the summer of 1954. It is a great pleasure to thank their staff for the hospitality they extended to us. We are grateful to Dr. R. C. Fletcher, Dr. C. Herring, and Dr. G. Wannier for many stimulating and helpful conversations.

Note added in proof.—Since this paper was submitted we have learned of experimental work by Burstein, Picus, and Henvis (to be published) who have observed optical transitions from the ground state to various excited states of donor electrons. These measurements are in excellent agreement with the theoretical positions of the excited "p"-states reported in the present paper. A discussion of the experimental results in the light of the theory is in press.