PHYSICAL REVIEW B VOLUME 10, NUMBER 2 15 JULY 1974

Theory of localized states in semiconductors. I. New results using an old method*†

Sokrates T. Pantelides‡§ and C. T. Sah

Department of Physics and Materials Research Laboratory, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801

(Received 4 June 1973)

In this paper the Kohn-Luttinger effective-mass theory is reevaluated using first-principles impurity potentials. In the past, it has generally been believed that the theory is valid for shallow levels and inapplicable to deep levels. Here, a new result is obtained, namely that the theory is valid for both shallow and *deep* levels but only for the special cases when the impurity atom is substitutional and is from the same row of the Periodic Table as the host atom. Such atoms have the same number of core electrons and have been called isocoric for convenience. On the other hand, the theory is found to be invalid in the cases of the general nonisocoric impurities, whether shallow or deep. These results are confirmed by calculations of donors in silicon. Excellent agreement with experiment is found for the isocoric phosphorus and sulfur impurities (both one- and two-electron sulfur levels are deep). No adjustable parameters are employed. These results are understood in physical terms as the one-band approximation can yield correct nodal structure for the impurity-electron wave function only in the case of isocoric impurities.

I. INTRODUCTION

Point imperfections in semiconductors, such as substitutional or interstitial chemical impurities, and lattice defects, such as vacancies and interstitials, introduce localized energy levels in the otherwise forbidden energy gaps of semiconductors. Accurate first-principles theoretical description of these bound states has been lacking.

For shallow levels, namely, levels lying close to the conduction- or valence-band edge, a simple picture gives a good description. In these cases the bound-electron (or hole) orbits are sizable and the energy levels are largely determined by the Coulomb tail of the perturbation potential. The result is a hydrogenic spectrum of levels given by

$$E_n = \frac{1}{n^2}\frac{m^*}{2\hbar^2}\left(\frac{e^2}{\epsilon}\right)^2 , \tag{1.1}$$

where m^* is an effective mass for the electron or hole and ϵ is the static dielectric constant of the semiconductor.[1]

This effective-mass-theory approach was put on a rigorous foundation by Slater[2] through a theorem by Wannier.[3] Application to real materials in a quantitative way was made by Kittel and Mitchell[4] and by Kohn and Luttinger.[5-7] The latter authors carried out more detailed work on the subject and the method has been referred to as the Kohn-Luttinger effective-mass theory (KL-EMT).

In the original KL-EMT, the perturbation potential was taken to be

$$U(\vec{r}) = -e^2/\epsilon r \tag{1.2}$$

(hence the levels given by Eq. (1.1) in the case of a simple band) and excellent agreement with experiment was obtained for the excited states.[8] For these states, the bound-electron wave function has vanishing amplitude at small r. The potential (1.2) had limited success for the ground state, whose wave function has large amplitude in the region of small r, where Eq. (1.2) is a poor approximation. Many papers have been published attempting to improve the ground-state result by including additional contributions to the perturbation potential.[9-14] The prevalent attitude is that the KL-EMT is a valid method for shallow levels. On the contrary, the method is generally believed to be inapplicable for deep levels, which lie at considerable distance from a band edge. The simple KL-EMT estimate for, say, sulfur in silicon is off by a factor of 5.

For deep levels, the bound electrons or holes have orbits of small radius and theoretical models have been developed based on the "defect molecule" concept.[15] Molecular-orbital and Hückel-theory techniques have been used in this context with largely qualitative results.[15,16] The band-theory approach has also been unable to produce reliable results.[17]

In this paper the applicability of the KL-EMT is examined when the perturbation potential $U(\vec{r})$ is constructed from first principles for individual impurities. In Sec. II the main aspects of the KL-EMT are presented and a quantitative criterion for its applicability is obtained without using an explicit expression for $U(\vec{r})$. In Sec. III the multivalley effective-mass equation (MV-EME) appropriate for donors in silicon is presented. In Sec. IV, the perturbation potential $U(\vec{r})$ is constructed from first principles for various donor impurities in silicon and the applicability of the KL-EMT is checked in each case. A striking result is obtained: from among the shallow donors, the KL-EMT is found to be applicable only in the case of phosphorus. Furthermore, contrary to traditional understanding, the KL-EMT is shown to be applicable for the deep sulfur levels. Again, the method does not apply for other deep donors. Numerical calculations are presented which verify these predictions. Excellent agreement with experiment is obtained for both phosphorus and sulfur. New calculations for the point-charge model, on which the original KL-EMT was based, are also presented, suggesting a new definition of "chemical shifts." Explanation of the above results in terms of physical reasoning is presented in Sec. V. Section VI contains a brief discussion of excited states and Sec. VI summarizes the results on the power and shortcomings of the KL-EMT. In the following paper,[18] a new method is developed, based on pseudopotential theory, which remedies the shortcomings of the KL-EMT and is applicable to a wider range of impurities.[19]

II. KOHN-LUTTINGER EFFECTIVE-MASS THEORY (KL-EMT)

This section contains a presentation of the effective-mass theory as developed by Kohn and Luttinger, but without using an explicit form for $U(\vec{r})$. A quantitative criterion for its applicability will be obtained.

The problem is to solve the perturbed crystal eigenvalue problem

$$H\psi = [-(\hbar^2/2m_0)\nabla^2 + V]\psi = E\psi \ , \tag{2.1}$$

where V is the total crystal potential. The method assumes that the perfect crystal eigenvalue problem, namely

$$H^0\psi^0_{n\vec{k}} = [-(\hbar^2/2m_0)\nabla^2 + V^0]\psi^0_{n\vec{k}} = E^0_n(\mathrm{k})\psi^0_{n\vec{k}} \tag{2.2}$$

(superscript 0 for all perfect-crystal quantities, no superscript for perturbed-crystal quantities) has been solved. Here V^0 is the perfect-crystal periodic potential, $\psi^0_{n\vec{k}}$ are the Bloch functions, and $E^0_n(\vec{k})$ is the energy of the nth band as a function of the wave vector $\vec{k}$ in the first Brillouin Zone. The potential V is then written as

$$V = V^0 + U \tag{2.3}$$

and U is treated as a perturbation to the system (2.2).

The most rigorous way to proceed is as in Ref. 5, but the method of Ref. 7 reproduces the same results. Briefly, the impurity-electron wave function ψ is expanded in terms of the complete, orthonormal set of Bloch functions as

$$\psi(\vec{r}) = \sum_n \sum_{\vec{k}} F_n(\vec{k})\psi^0_{n\vec{k}}(\vec{r}) \ . \tag{2.4}$$

Inserting this in Eq. (2.1) using (2.2) and (2.3), multiplying on the left by $\psi^{0*}_{n'\vec{k}'}$, and integrating, we get

$$E^0_n(\vec{k})F_n(\vec{k}) + \sum_{n'}\sum_{\vec{k}'} \langle\psi^0_{n\vec{k}}|U|\psi^0_{n'\vec{k}'}\rangle F_{n'}(\vec{k}') = EF_n(\vec{k}) \tag{2.5}$$

(primed and unprimed quantities have been interchanged). Now, the matrix element of the perturbation potential may be written by Fourier-analyzing $U(\vec{r})$ in the form

$$U(\vec{r}) = \sum_{\vec{q}} U(\vec{q})\, e^{i\vec{q}\cdot\vec{r}} \ , \tag{2.6}$$

whereby

$$\langle\psi^0_{n\vec{k}}|U|\psi^0_{n'\vec{k}'}\rangle = \sum_{\vec{q}} U(\vec{q})\langle\psi^0_{n\vec{k}}|e^{i\vec{q}\cdot\vec{r}}|\psi^0_{n'\vec{k}'}\rangle \ . \tag{2.7}$$

In order to simplify further we use the Bloch form of $\psi^0_{n\vec{k}}$, namely

$$\psi^0_{n\vec{k}}(\vec{r}) = (1/N\Omega)^{1/2}\, u^0_{n\vec{k}}(\vec{r})\, e^{i\vec{k}\cdot\vec{r}} \ , \tag{2.8}$$

where $N\Omega$ is the volume of the crystal, and expand the periodic function:

$$u^{0*}_{n\vec{k}}(\vec{r})u^0_{n'\vec{k}'}(\vec{r}) = \sum_h C^{nn'}_{\vec{k}\vec{k}'}(\vec{K}_h)\, e^{-i\vec{K}_h\cdot\vec{r}} \ . \tag{2.9}$$

Here $\vec{K}_h$ are the reciprocal-lattice vectors of the perfect crystal. We get

$$\langle\psi^0_{n\vec{k}}|e^{i\vec{q}\cdot\vec{r}}|\psi^0_{n'\vec{k}'}\rangle = \sum_h C^{nn'}_{\vec{k}\vec{k}'}(\vec{K}_h)\,\delta_{\vec{q},\,\vec{k}-\vec{k}'+\vec{K}_h} \ . \tag{2.10}$$

Using this result and[20]

$$C^{nn'}_{\vec{k}\vec{k}'}(0) \simeq \delta_{nn'} \ , \tag{2.11}$$

Eq. (2.7) becomes

$$\langle\psi^0_{n\vec{k}}|U|\psi^0_{n'\vec{k}'}\rangle$$

$$= U(|\vec{k}-\vec{k}'|) + \sum_{h\neq 0} U(|\vec{k}-\vec{k}'+\vec{K}_h|)\, C^{nn'}_{\vec{k}\vec{k}'}(\vec{K}_h) \ . \tag{2.12}$$

In the KL-EMT,[2-7] one drops the $h \neq 0$ terms in Eq. (2.11) so that (2.5) becomes

$$E^0_n(\vec{k})F_n(\vec{k}) + \sum_{\vec{k}'} U(|\vec{k}-\vec{k}'|)F_n(\vec{k}') = EF_n(\vec{k}) \ . \tag{2.13}$$

Clearly, from Eq. (2.12), this approximation is valid if the quantity

$$R = \frac{\sum_{h\neq 0} U(|\vec{k}-\vec{k}'+\vec{K}_h|)\, C^{nn'}_{\vec{k}\vec{k}'}(\vec{K}_h)}{U(|\vec{k}-\vec{k}'|)} \tag{2.14}$$

is small compared with 1, i.e., if $U(\vec{r})$ has no strong high Fourier components. This is equivalent to requiring that $U(\vec{r})$ must be a smooth potential that varies slowly with position.[2-7] The requirement that $R \ll 1$ gives a *quantitative* criterion for "smoothness." A consequence of this smoothness is that the perturbation U is so weak that it does not couple Bloch functions from different bands, giving the one-band result (2.13). This criterion

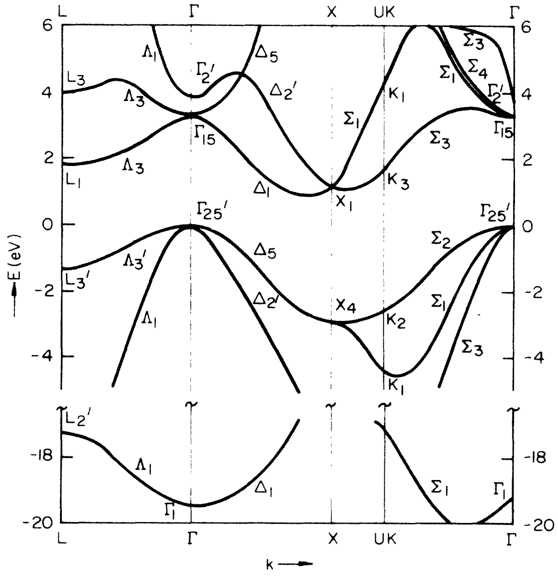

FIG. 1. Energy-band structure of crystalline silicon (from various sources).

will be developed further in Sec. IV B and a more direct physical interpretation will be arrived at in Sec. V.

The usual one-band effective-mass equation (EME) is obtained from Eq. (2.13) by further expanding $E_n^0(\vec{k})$ to powers up to k^2 about the band extremum (or extrema) of the nth band. In the simplest case of a spherical minimum at $\vec{k}=0$ (as in GaAs), this is just

$$E_n^0(\vec{k})=(\hbar^2/2m^*)k^2\,, \qquad (2.15)$$

where m^* measures the curvature of the band minimum and is the effective mass for the electron. Then, multiplying Eq. (2.13) by $e^{i\vec{k}\cdot\vec{r}}$, summing over $\vec{k}$ and using

$$F(\vec{r})=\sum_{\vec{k}}F(\vec{k})\,e^{i\vec{k}\cdot\vec{r}} \qquad (2.16)$$

(where we dropped the band index n for clarity since it is a one-band model anyway), one gets the conventional one-band EME

$$[-(\hbar^2/2m^*)\nabla^2+U(\vec{r})]F(\vec{r})=EF(\vec{r})\,, \qquad (2.17)$$

from which the bound-state energy levels E may be determined. In obtaining Eq. (2.17), the sum over $\vec{k}$ in the potential matrix element was extended to cover all of $\vec{k}$ space. This is consistent with the one-band approximation and is valid again when $R \ll 1$, in which cases $F(\vec{k})$ vanishes rapidly outside the first Brillouin zone. This approximation is actually counterbalanced by dropping the k^4 and higher-order terms in $E^0(k)$.[21] Numerical estimates for these approximations will be given in Sec. V.

III. KL-EMT FOR DONORS IN SILICON

The complete band structure of silicon is shown in Fig. 1. For donors, the perturbation is negative, whereby the relevant band is the lowest conduction band. This band is known to have six equivalent minima of ellipsoidal shape along the [100] direction (Fig. 2). For example, around the minimum along the $+x$ direction, located at $\vec{k}_1$, the energy may be expanded as

$$E^0(\vec{k})\simeq\frac{\hbar^2}{2m_l}(k_x-k_1)^2+\frac{\hbar^2}{2m_t}(k_y^2+k_z^2)\,. \qquad (3.1)$$

One then must write

$$F(\vec{k})=\sum_{i=1}^{6}\alpha_i F_i(\vec{k})\,, \qquad (3.2)$$

where $F_i(\vec{k})$ is centered about the ith minimum. The constants α_i are determined from group theory[6,7] by requiring that the total impurity-electron wave function $\psi(\vec{r})$ belongs to an irreducible representation of the point group of the Hamiltonian H. For substitutional donors, this is the tetrahedral group T_d.

The effective-mass equation for a many-valley band was first obtained by Twose[22] as reported by Fritzsche.[23] It is derived by substituting (3.2) in (2.13), multiplying by $e^{i(\vec{k}-\vec{k}_j)\cdot\vec{r}}$, summing over $\vec{k}$, and using

$$F_i(\vec{r})=\sum_{\vec{k}}F_i(\vec{k})\,e^{i(\vec{k}-\vec{k}_i)\cdot\vec{r}}\,. \qquad (3.3)$$

The result is

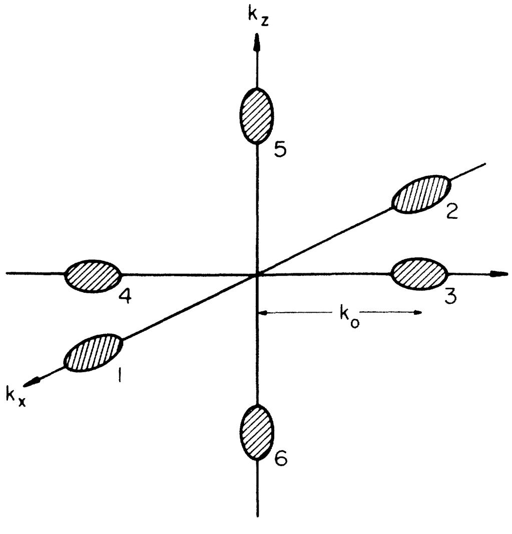

FIG. 2. Constant-energy ellipsoids around the conduction-band minima along the six [100] directions of silicon.

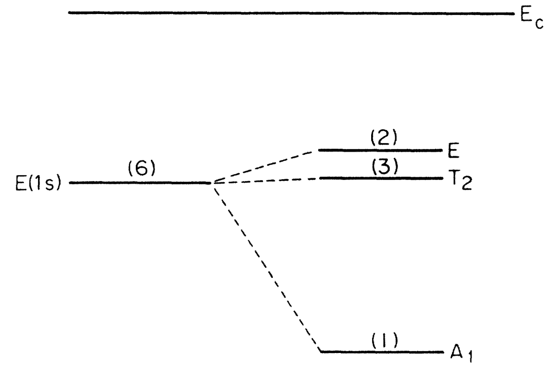

FIG. 3. Splitting of the ground-state energy level of donors in silicon. Numbers in parentheses indicate the degeneracy (not including spin degeneracy).

$$\sum_{i=1}^{6} \alpha_i e^{i(\vec{k}_i-\vec{k}_j)\cdot\vec{r}}[T_i(-i\nabla)+U(\vec{r})-E]F_i(\vec{r})=0 \, , \tag{3.4}$$

where $T_j(\vec{k})$ is the expansion of $E^0(\vec{k})$ around the jth minimum as in Eq. (3.1). It should be noted here that in the original work[4–7] and most subsequent papers, intervalley mixing, also called valley-orbit interaction, was considered negligible, and the energy was obtained from a one-valley EME. The approximation implicit in such a one-valley approach may be seen rigorously from Eq. (3.4) by rewriting it as

$$\alpha_j[T_j(-i\nabla)+U(\vec{r})-E]F_j(\vec{r})+\sum_{i=1,\, i\neq j}^{6}\alpha_i e^{i(\vec{k}_i-\vec{k}_j)\cdot\vec{r}} \times[T_i(-i\nabla)+U(\vec{r})-E]F_i(\vec{r})=0 \, ; \tag{3.5}$$

the sum represents the intervalley mixing ($i\neq j$ terms). When it is negligible, as indeed it is for excited states, the bound-state energy can be obtained from the simple one-valley EME

$$[T_j(-i\nabla)+U(\vec{r})-E]F_j(\vec{r})=0 \, . \tag{3.6}$$

This equation has proved very successful for the excited states.[8] The importance of the intervalley terms for the ground state was first recognized by Baldereschi.[24] He calculated them by perturbation theory and by omitting the exponential factor appearing in Eq. (3.5). He was able to show that intervalley mixing was largely responsible for the splitting of the ground state into a singlet of A_1 symmetry, a triplet of T_2 symmetry,[25] and a doublet of E symmetry. This splitting had been observed experimentally[26,27] and is not accounted for by the original one-valley KL-EMT according to which the ground state is sixfold degenerate (Fig. 3). Subsequent work by Ning and Sah[28] substantiated the importance of intervalley mixing. They used multivalley expressions, made a spherical-valley approximation, and fitted an empirical impurity potential to the split ground state.

Finally, when $U(r)$ binds two electrons, a helium-like model has been developed in the multivalley approximation.[21,28] In the ground state, both electrons occupy the state of A_1 symmetry of lowest energy and interact via a screened Coulomb potential. No correlation is included. The details are given in Appendix B.

IV. KL-EMT WITH FIRST-PRINCIPLES POTENTIALS

In this section the applicability of the KL-EMT, as described in Secs. I–III in its extended multivalley form, will be tested by using first-principles potentials in the cases of both shallow and deep donors. This is meant to check KL's speculation that the true impurity potential $U(r)$ would vary too rapidly in the central cell for the EMA to be valid. Another check is to be made on the prevalent understanding that a one-band approximation breaks down for deep levels situated around the middle of the energy gap. An indication that this intuitive understanding is not necessarily true comes from symmetry considerations[29] and from a comparison of emission rates from the localized center to the conduction and valence bands. For example, the deep sulfur donor has been found to have negligible hole emission to the valence band.[30]

A. Impurity potentials

First we describe how the impurity potential $U(r)$ for each impurity is calculated. We saw already that U is given by

$$U=V-V^0 \, . \tag{4.1}$$

V^0 may be written as

$$V^0(\vec{r})=\sum_j v_0^{+4}(\vec{r}-\vec{R}_j)+V_e^0(\vec{r}) \tag{4.2}$$

where $v_0^{+4}(\vec{r}-\vec{R}_j)$ is the potential of the Si^{+4}-like ion at the site $\vec{R}_j$ and $V_e^0(\vec{r})$ is the potential arising from the valence electrons.[31] In principle, these potentials are nonlocal because they include exchange, but they are denoted as functions of $\vec{r}$ for simplicity. The local Slater approximation for exchange will ultimately be used.

Similarly V, even though it is not periodic, may be written as

$$V(\vec{r})=\sum_j v^{+z_j}(\vec{r}-\vec{R}_j')+V_e(\vec{r}) \, , \tag{4.3}$$

where now $v^{+z_j}(\vec{r}-\vec{R}_j')$ is the potential of the ion at the site $\vec{R}_j'$, whose ionicity is $+z_j$. The new sites are denoted by $\vec{R}_j'$ and are, in principle, different from $\vec{R}_j$ because of lattice relaxation. Furthermore, in the case of an interstitial impurity, the sum over j in (4.3) must include the occupied interstitial site as well. Finally, as before, $V_e(\vec{r})$ is the potential of the valence electrons.

Combining Eqs. (4.1)–(4.3), we get for U

$$U = U_b + U_s \,, \tag{4.4}$$

where

$$U_b(\vec{r}) = \sum_j [v^{+z_j}(\vec{r} - \vec{R}'_j) - v_0^{+4}(\vec{r} - \vec{R}_j)] \tag{4.5}$$

and

$$U_s(\vec{r}) = V_e(\vec{r}) - V_e^0(\vec{r}) \,. \tag{4.6}$$

In order to be able to compute $U(\vec{r})$, two approximations are first made: (a) Lattice relaxation is neglected, whereby $\vec{R}'_j \simeq \vec{R}_j$; (b) the core electrons at all sites except the foreign-atom site remain unperturbed (independent-core assumption). Denoting the imperfection site by $j = 0$, this means

$$v^{+z_j}(\vec{r} - \vec{R}'_j) \simeq v_0^{+4}(\vec{r} - \vec{R}_j) \,, \quad j \neq 0 \,. \tag{4.7}$$

Estimates for the error introduced by these approximations will be discussed in Sec. V.

With these approximations, and setting $\vec{R}_0 = 0$, we get

$$U_b(\vec{r}) = v^{+z}(\vec{r}) - v_0^{+4}(\vec{r}) \tag{4.8}$$

for a substitutional impurity, and

$$U_b(\vec{r}) = v^{+z}(\vec{r}) \tag{4.9}$$

for an interstitial impurity, where, in both cases, we recall that z indicates the number of valence electrons of the impurity atom. For group-V substitutional donors $z = 5$, for interstitial magnesium $z = 2$, and so on.

We thus see that for substitutional donors, $U_b(\vec{r})$ is essentially the difference between the potential of an impurity ion and a host ion; and for interstitial donors, $U_b(\vec{r})$ is simply the potential of an impurity ion. These results could indeed have been written down from the beginning by intuition; we chose to derive them starting with Eq. (4.1) in order to identify the approximations involved in a systematic way. Finally, $U_s(\vec{r})$ arises from the redistribution of valence electrons in the crystal, caused by the presence of $U_b(\vec{r})$. We will return to this shortly.

Construction of the potentials $v_0^{+4}(\vec{r})$ and $v^{+z}(\vec{r})$ is straightforward in terms of the orbitals $\chi_\lambda^0(\vec{r})$ of the core electrons at a silicon site in the crystal and the orbitals $\chi_{\lambda'}(\vec{r})$ of the core electrons at the impurity site. Using the local Slater approximation for inclusion of exchange, we have

$$v_0^{+4}(\vec{r}) = \frac{-Z^0 e^2}{r} + \int d^3r' \frac{\rho^0(\vec{r}')}{|\vec{r} - \vec{r}'|} - 3\alpha \left(\frac{\rho^0(\vec{r})}{2\pi} \right)^{1/3} , \tag{4.10}$$

where

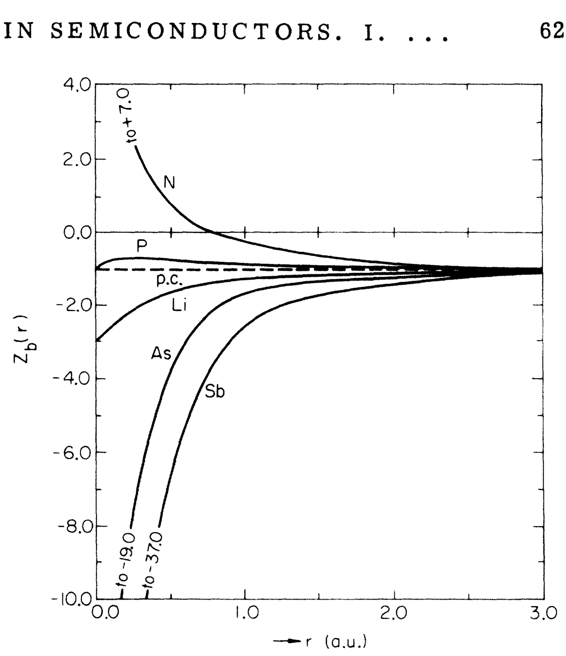

FIG. 4. Bare impurity potentials [effective charge $Z_b(r) = rU_b(r)/e^2$] for single donors in silicon, compared with (the Coulomb potential of) a point charge (p. c.).

$$\rho^0(\vec{r}) = e^2 \sum_\lambda |\chi_\lambda^0(\vec{r})|^2 \,. \tag{4.11}$$

Z^0 is the atomic number of the host and $\alpha = 1$. Similarly

$$v^{+z}(\vec{r}) = \frac{-Ze^2}{r} + \int d^3r' \frac{\rho(\vec{r}')}{|\vec{r} - \vec{r}'|} - 3\alpha \left(\frac{\rho(\vec{r})}{2\pi} \right)^{1/3} , \tag{4.12}$$

where

$$\rho(\vec{r}) = e^2 \sum_{\lambda'} |\chi_{\lambda'}(\vec{r})|^2 \,, \tag{4.13}$$

with Z the atomic number of the impurity atom. Presumably, the $\chi_\lambda^0(r)$ in the crystal must be intermediate between the core orbitals of a free ion and a free atom.[32] The two extreme choices are essentially indistinguishable and either may be used as a very good approximation. In general, one expects the same assumption to be good for the $\chi_{\lambda'}$ as well. For heavy impurity atoms, such as antimony and bismuth, it is likely that the $\chi_{\lambda'}$ are distorted from the free-atom or free-ion forms and one must exercise caution.

Using a version of the original computer program of Herman and Skillman,[33] we can generate the wave functions χ_λ^0 and χ_λ for atoms or ions. The potentials $v_0^{+4}(r)$ and $v^{+z}(r)$ are then constructed using Eqs. (4.10)–(4.13) and averaging over the angles. The resultant potentials $U_b(r)$ for a number of donors, calculated from free-atom orbitals, are shown in Figs. 4 and 5. The quantity plotted is actually the effective charge $Z_b(r)$ defined by

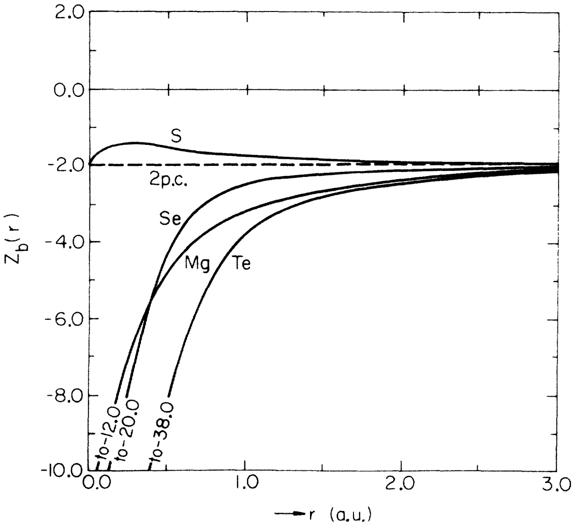

FIG. 5. Bare impurity potentials [effective charge $Z_b(r) = rU_b(r)/e^2$] for double donors in silicon, compared with (the Coulomb potential of) two point charges (2 p.c.).

$$U_b(\vec{r}) = e^2 Z_b(r)/r\ . \tag{4.14}$$

It is seen that Z_b has the correct asymptotic form

$$Z_b(r) \xrightarrow[r\to 0]{} -(Z-Z^0) \tag{4.15a}$$

for substitutional donors, or

$$Z_b(r) \xrightarrow[r\to 0]{} -Z \tag{4.15b}$$

for interstitial donors, and

$$Z_b(r) \xrightarrow[r\to \infty]{} -n \tag{4.16}$$

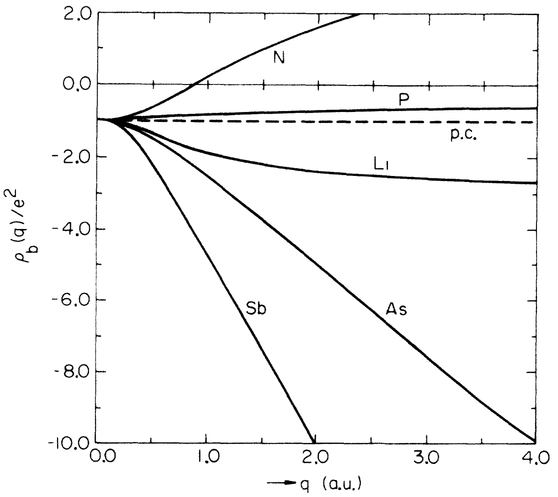

FIG. 6. Fourier transforms of the potentials of Fig. 4 [quantity plotted is $\rho_b(q) = q^2 U_b(q)/4\pi$].

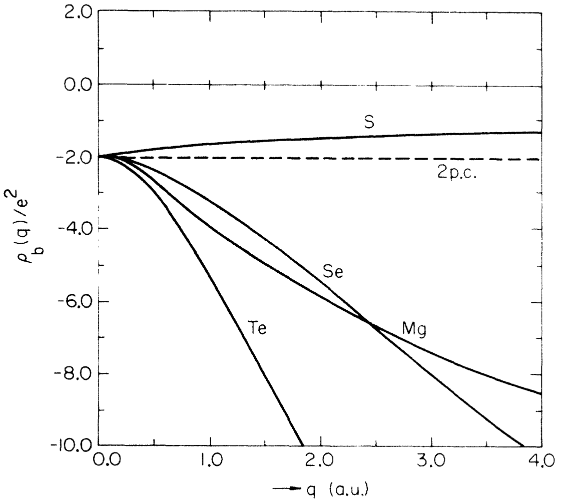

FIG. 7. Fourier transforms of the potentials of Fig. 5 [quantity plotted is $\rho_b(q) = q^2 U_b(q)/4\pi$].

for both kinds of donors. Actually $Z_b(r)$ attains the constant value n at a distance of a few atomic units. The value of n is $z-4$ for substitutional donors and z for interstitial donors. This allows us to write

$$U_b(\vec{r}) = -ne^2/r + W_b(\vec{r})\ , \tag{4.17}$$

where the first term is simply the Coulomb potential of n point charges, and $W_b(\vec{r})$ is localized within a few a.u. It is clear from Figs. 4 and 5 that only in the cases of phosphorus and sulfur in silicon, $U_b(\vec{r})$ is very nearly equal to the bare Coulomb potential of one or two point charges, respectively—in other words, W_b is very small—whereas in all other cases $U_b(\vec{r})$ deviates considerably from the point-charge model. Similar conclusions are also drawn from the Fourier transforms of these potentials, shown in Figs. 6 and 7. Again, for convenience, we plot $\rho_b(\vec{q})/e^2$ where $\rho_b(\vec{q})$ is defined by

$$U_b(\vec{q}) = (4\pi/q^2)\rho_b(\vec{q})\ . \tag{4.18}$$

Finally, before we check the applicability of the EMA for the various potentials, we see how $U_s(\vec{r})$ is calculated. As noted already, it may be viewed as the response of the valence electrons to the presence of $U_b(\vec{r})$, but linear-response theory (dielectric screening) is not necessarily valid in all cases. Indeed, as it is shown in Appendix A it is valid only in the cases of a point-charge model, phosphorus and sulfur in silicon. In these cases U_s is given Fourier-transformed by

$$U_s(\vec{q}) = [1/\epsilon(\vec{q}) - 1]U_b(\vec{q})\ . \tag{4.19}$$

The net result is that $U(\vec{q})$ is obtained from

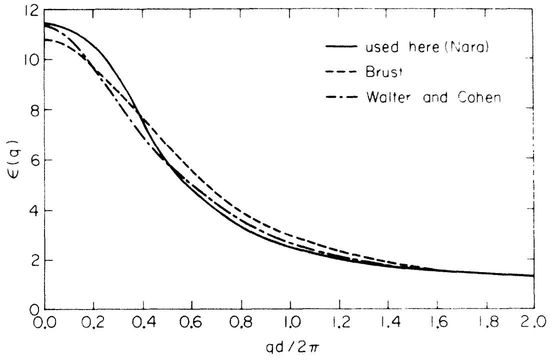

FIG. 8. Dielectric function $\epsilon(q)$ of silicon used here compared with other calculations (Brust, Ref. 36; and Walter and Cohen, Ref. 37). d is the lattice constant (5.431 Å).

$$U(\vec{q}) = U_b(\vec{q})/\epsilon(\vec{q}) \qquad (4.20)$$

and

$$U(\vec{r}) = \sum_{\vec{q}} U(\vec{q})\, e^{i\vec{q}\cdot\vec{r}} . \qquad (4.21)$$

The function $\epsilon(q)$ for silicon has been calculated by several authors. Here we used the function obtained by Nara[34] normalized to $\epsilon(0) = 11.4$, which is the low-temperature static dielectric constant for silicon.[8] The numerical results were fitted[35] to the analytic expression

$$\frac{1}{\epsilon(\vec{q})} = \frac{Aq^2}{q^2+\alpha^2} + \frac{(1-A)q^2}{q^2+\beta^2} + \frac{[1/\epsilon(0)]\gamma^2}{q^2+\gamma^2} , \qquad (4.22)$$

with $A = 1.175$, $\alpha = 0.7572$, $\beta = 0.3123$, and $\gamma = 2.044$ a.u. This function is shown in Fig. 8, where it is compared with the corresponding functions calculated by Brust[36] and by Walter and Cohen.[37] The consequences of uncertainties in the form of $\epsilon(\vec{q})$ are discussed briefly in Sec. V and more thoroughly in the following paper.[18]

In performing the calculations, the convenient separation (4.17) can be maintained by screening each piece separately, so we can write

$$U(\vec{r}) = U_{\text{pc}}(\vec{r}) + W(\vec{r}) . \qquad (4.23)$$

It should be noted that this separation is simply a mathematical convenience and does not imply that linear screening theory would apply to each piece separately. $U_{\text{pc}}(\vec{r})$ is then the screened Coulomb potential of n point charges. With $\epsilon(q)$ of Eq. (4.22), it is given by

$$U_{\text{pc}}(\vec{r}) = (-ne^2/\epsilon r) \times [1 + A\epsilon e^{-\alpha r} + (1-A)\epsilon e^{-\beta r} - e^{-\gamma r}] , \qquad (4.24)$$

where ϵ stands for $\epsilon(0)$. The short-range part $W(\vec{r})$ is obtained in numerical form and no explicit form for it may be given. It should be noted, however, that, for a potential given in numerical form, it is impractical to calculate the Fourier transform, divide by $\epsilon(q)$, and integrate back to $\vec{r}$ space. Instead, the general form (4.21) with (4.20) is modified to

$$W(r) = W_b(r) + \int dr' E(r, r') W_b(r') r'/r . \qquad (4.25)$$

The function $E(r, r')$ is a generalized dielectric function and is given in Appendix A. Equation (4.25) is then integrated numerically.

For the other donors, the dielectric-screening procedure is of course inapplicable. The calculation of $U_s(\vec{r})$ in these cases would be a forbiddable task. As we shortly shall see, however, the EMA is not applicable for these donors, so, at least for the time being, we do not have to worry about calculating $U_s(\vec{r})$ for them.

B. Checking the applicability of the EMA

Having constructed the impurity potentials, we are ready to check the criterion

$$|R| \ll 1 , \qquad (4.26)$$

where R is given by Eq. (2.14). First we look at the special case of the point-charge model for which R, denoted R_{pc}, is given by

$$R_{\text{pc}} = \sum_{h\neq 0} \frac{U_{\text{pc}}(|\vec{k}-\vec{k}'+\vec{K}_h|)}{U_{\text{pc}}(\vec{k}-\vec{k}'|)} C^{nn'}_{\vec{k}\vec{k}'}(\vec{K}_h) . \qquad (4.27)$$

Thus, if we assume that $|R_{\text{pc}}| \ll 1$, i.e., the approximations are valid for the screened-point-charge model, we can replace Eq. (4.26) by

$$|R/R_{\text{pc}}| \leqslant 1 . \qquad (4.28)$$

Then, using Eqs. (4.18) and (4.20) and the definition of U_{pc} as the screened-point-charge potential, R can be written as

$$R = \sum_{h\neq 0} \frac{U_{\text{pc}}(|\vec{k}-\vec{k}'+\vec{K}_h|)\,\rho_b(|\vec{k}-\vec{k}'+\vec{K}_h|)}{U_{\text{pc}}(|\vec{k}-\vec{k}'|)\,\rho_b(|\vec{k}-\vec{k}'|)} \times C^{nn'}_{\vec{k}\vec{k}'}(\vec{K}_h) . \qquad (4.29)$$

By comparing Eq. (4.29) with (4.27), we see that (4.28) is satisfied if

$$\frac{\rho_b(|\vec{q}+\vec{K}_h|)}{\rho_b(q)} \leqslant 1 , \quad \vec{K}_h \neq 0 . \qquad (4.30)$$

At the same time, we note that Eq. (4.30) is required only for the first few nonzero values of $\vec{K}_h$. This is because the quantities $C^{nn'}_{\vec{k}\vec{k}'}(\vec{K}_h)$, as obtained from a band-structure calculation, are nonzero only for a finite number of $\vec{K}_h$. Clearly, if $C^{nn'}_{\vec{k}\vec{k}'}(\vec{K}_h) = 0$ for a particular $\vec{K}_h$, the contribution to R is zero

TABLE I. Ionization energies for donors in silicon for which the KL-EMT in the multivalley approximation is applicable.

Impurity	Theory a (a.u.)	 c_1	 E (meV)	Experiment E (meV)
point charge	17.80	0.924	48.8	...
phosphorus	23.25	0.985	42.4	45.5[a]
two point charges (+)[b]	5.46	1.0	1085.3	...
two point charges (0)[c]	6.43	1.0	489.0	...
sulfur (S^+)[b]	6.80	1.0	659.3	613.6[d]
sulfur (S^0)[c]	7.60	1.0	297.1	302.0[e]
all other donors	theory inapplicable (see text)			

[a] Reference 27, corrected with theoretical result of Ref. 8 for $3p_{\pm}$ state.

[b] One electron is bound, so the center is positively charged.

[c] Two electrons are bound so the center is neutral.

[d] Reference 38.

[e] Reference 39. The value 302.0 is the thermal activation energy. The ionization energy should be at least 15 meV higher.

whether (4.30) is satisfied or not. In a final analysis, what we do amounts to looking at the nonzero $\vec{K}_h \neq 0$ terms for a given potential U, and if *each* of them is smaller than or equal to the corresponding term of the point-charge potential U_{pc}, then we conclude that the EMT must be as good for U as it is for U_{pc}.

The validity of the EMA for U_{pc} itself has been assumed in the literature,[11,12,24] even though it was originally established for a point-charge potential screened by the dielectric constant ϵ. R_{pc} still remains smaller than unity [despite the factor $\epsilon(\vec{q})/\epsilon(\vec{q}+\vec{K}_h)$, about 10] because of the smallness and signs of the constants $C^{nn'}_{\vec{k}\vec{k}'}(K_h)$. This will become clearer in the following paper.[18] Here our intention is to compare realistic potentials with U_{pc}. More general questions of accuracy will be discussed in Sec. V.

With the validity criterion in the form (4.30) we can use Figs. 6 and 7, and, in a straightforward manner, predict that when realistic impurity potentials are used, the EMA would be valid only in the cases of phosphorus and sulfur in silicon. Next, we shall see how the numerical calculations substantiate these predictions. A more comprehensive analysis of the results is given in the discussion in Sec. V.

C. Calculations of energy levels

For the calculations for one-electron centers, Eq. (3.4) is used. The calculation is done variationally. As in Ref. 28, we are forced to use the spherical-band approximation for $T_j(\vec{k})$,

$$T_j(\vec{k}) = \hbar^2 \kappa^2 / 2m^* \quad . \tag{4.31}$$

The value of m^* is chosen to reproduce Faulkner's result in the one-valley calculation with a Coulomb potential $-e^2/\epsilon r$. This value is $m^* = 0.2987 m_0$. The trial function was taken to be a sum of $1s$- and $2s$-like hydrogenic functions

$$F_i(\vec{r}) = c_1 F_i^{1s}(\vec{r}) + c_2 F_i^{2s}(\vec{r}) \ , \tag{4.32}$$

where

$$F_i^{1s}(\vec{r}) = (\pi a^3)^{-1/2} e^{-r/a} \ , \quad \text{for all } i \tag{4.33}$$

and

$$F_i^{2s}(\vec{r}) = (32\pi a^3)^{-1/2} (2 - r/a) e^{-r/2a} \ , \quad \text{for all } i \ ; \tag{4.34}$$

c_1, c_2, and a are variational parameters. With such a trial function, a 2×2 secular matrix is obtained, which must be diagonalized. The secular matrix and the explicit form of the matrix elements in terms of the variation parameter a are given in Appendix B.

For the two-electron centers the simpler trial function

$$F_i = F_i^{1s}(\vec{r}) \tag{4.35}$$

is used for both electrons. The total energy is then minimized. The first ionization energy is then obtained by subtracting the energy of the one-electron system. For consistency, then, the latter was also calculated using Eq. (4.35).

Calculations were performed for phosphorus and sulfur in silicon using the impurity potentials constructed earlier in this section. For comparison, calculations were also carried out using the screened Coulomb potentials $U_{pc}(\vec{r})$ of one and two point charges, which were seen to approximate the two impurity potentials, respectively. The results are presented and compared with experiment in Table I and will be discussed in Sec. V.

For the other donors, it has been seen that the EMA would not be applicable. In addition, linear screening would not be justified in calculating $U_s(\vec{r})$. Nevertheless, the latter gives the correct asymptotic form for $U(\vec{r})$ for both $r \to 0$ and r greater than a few a.u. and was used by Morita and Nara.[40] As an additional check, therefore, linear screening was used for other donors and calculations of energy levels were attempted. No bound state in the region of interest was found, confirming the failure of the KL-EMT in these cases.

V. ANALYSIS OF RESULTS AND DISCUSSION

Agreement between theory and experiment is seen from Table I to be very good. We will discuss the accuracy of the calculations at the end of this section. First we discuss the physics behind the results just presented and then make a connection with the conventional point of view.

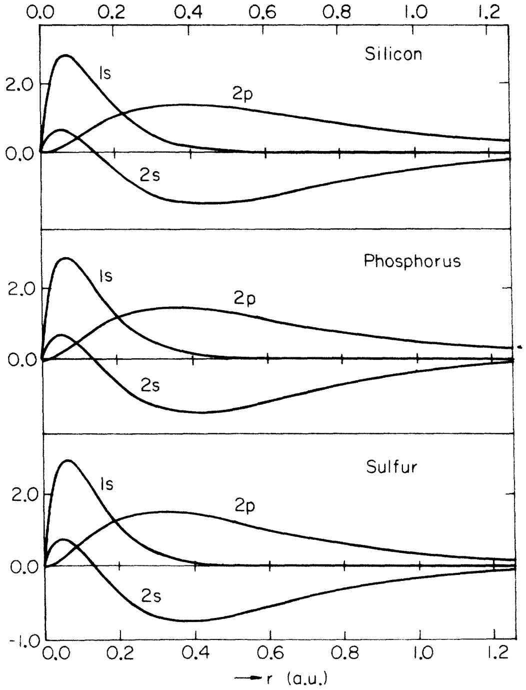

FIG. 9. Core wave functions of silicon, phosphorus, and sulfur.

A. New results

The result that effective-mass theory has been shown to work extremely well for deep levels (sulfur) while failing badly for some shallow levels (As, Sb, etc.) contradicts directly the traditional conception of the capabilities of this theory. In mathematical terms this came about because of the presence or absence of strong high Fourier components in the potential U and the satisfaction of criterion (4.30). A look at the Periodic Table, however, reveals very directly in physical terms when this should be the case: phosphorus and sulfur, for which the EMT proved to be successful, are in the same row as the host, silicon; hence, all three have the same core electronic structure, in this case $1s^2\,2s^2\,2p^6$ (Fig. 9). For convenience in subsequent discussions we call such impurities *isocoric*.[41] For such impurities the perturbation potential $U(\vec{r})$ is not responsible for the binding of additional core electrons in the impurity cell; hence, it has no strong short-range core part (Figs. 4 and 5) or, equivalently, no strong high Fourier components, whereby criterion (4.30) is satisfied and the EMA is applicable. The opposite happens in the cases of nonisocoric impurities, e.g., As in Si, which has 18 electrons more than the host core (Fig. 10). This implies a strong short-range core part in U (Fig. 4) and strong high Fourier components (Fig. 6), which render criterion (4.30) unsatisfied and the EMT inapplicable.

From this discussion, one can conclude that, as it has always been known, the EMT is good when the impurity potential is weak. Unlike the traditional understanding, however, which associated weak with shallow levels and strong with deep levels, the present work establishes that *weak means that $U(r)$ does not bind any core electrons.*

The above conclusion can be corroborated by another simple observation: In the KL-EMT, the total impurity wave function is given by[6,7,21]

$$\psi(\vec{r}) = \sum_{i=1}^{6} \alpha_i F_i(\vec{r}) \psi^0_{\vec{k}_i}(\vec{r}) \,. \tag{5.1}$$

Each term in Eq. (5.1) is a Bloch function of the bottom of the conduction band multiplied by a smooth envelope. The Bloch function $\psi^0_{\vec{k}_i}(\vec{r})$ is known to have a predominantly $3s$-$3p$-$3d$-like nodal structure at each atomic site in order to be orthogonal to the $1s$, $2s$, and $2p$ core states. It is obvious then that for an isocoric impurity the approximate form (5.1) for ψ has the correct nodal structure in both the impurity cell and the rest of the crystal. On the other hand, in the case of the nonisocoric arsenic impurity, $\psi(\vec{r})$ should have predominantly $4s$-$4p$-$4d$-like nodal structure in the impurity cell, whereby (5.1) and hence the one-band EMT are inapplicable in that region.

One might still legitimately wonder, however, how good a one-band approximation can be, particularly for deep levels, like those of sulfur. The first point to note is that most of the other bands cannot be expected to contribute significantly, because in that case, they would introduce undesirable nodal structure in ψ. The valence band is an exception, however, since its Bloch functions also have a $3s$-$3p$-like nodal structure and thus would not damage the structure of ψ. One might then argue that the two bands ought to be included

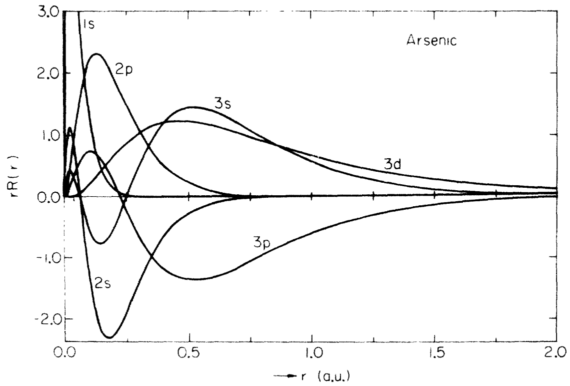

FIG. 10. Core wave functions of arsenic.

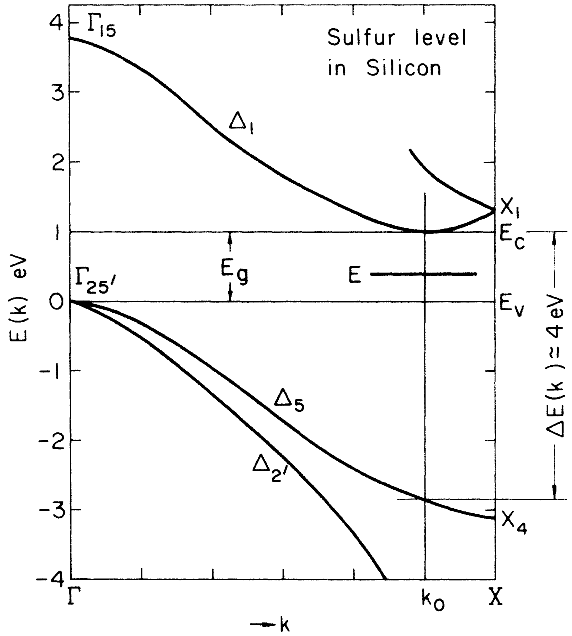

FIG. 11. Illustrating the localized sulfur level (marked E) in $\vec{k}$ space. Due to the indirect gap, the localized level is not very close to the valence band (only one of the six equivalent minima is shown).

on an equal footing, especially for the sulfur levels which lie midway in the energy gap. This is not so, however—as is illustrated in Fig. 11—because of the indirect nature of the gap. For one thing, the top of the valence band ($k=0$) cannot contribute for symmetry reasons,[29] whereas in the region of k_0 the valence band lies considerably away from the impurity level. In this region, symmetry or any other considerations do not prohibit valence-band contributions, but a perturbation-theory calculation can show that, through the energy denominator, the contribution is negligibly small.[7,28] This is also suggested by experiment, as mentioned in the preamble of Sec. IV.

Finally, another question might be raised, based on intuition. It has been argued at times[42-45] that, especially for deep levels, the effective-mass m^* should be replaced by the mass m_0 of the free electron in the impurity cell region. Ansätze expressions have been used in this context.[42-45] Examining this question from a rigorous point of view, however, we see that the only allowed correction to m^* is the inclusion of k^4 terms in the expansion of $E^0(\vec{k})$. This would result in a $\vec{k}$-dependent effective mass. Such a correction can be estimated directly (see Sec. V C and Ref. 21) and shown to be small and counterbalanced by a respective correction in the perturbation-potential-energy matrix element. From another point of view, a position-dependent m^* can be thought to include contributions from other bands in an average way. These have been seen to be negligibly small for *P* and *S*. In the following paper,[18] where a many-band method is developed which is applicable to the other donors, we will see how those results can be interpreted in terms of such a position-dependent effective mass for the impurity electron. Otherwise, any *ad hoc* replacement of m^* with some function $m^*(r)$ is not rigorously justified.

B. Connection with the conventional point of view

The conventional point of view as regards the EMA, is to solve the effective-mass equation (EME) using the point-charge potential Eq. (1.2), obtain the effective Rydberg E_0, and then look at the discrepancies between this value and the experimental ionization energies E_{exp} as "*chemical shifts*" caused by "*central cell corrections*" to the potential Eq. (1.2). The chemical shift is then defined as

$$\Delta = E_{exp} - E_0 \qquad (5.2)$$

and measures the deviation of the impurity potential from that of a point charge embedded in a dielectric medium. In the case of donors in silicon, using Eq. (1.2), one finds that a one-valley EME[4-8] gives $E_0 = 31.3$ meV.[8] Other authors[9-14] attempted to improve this number for the point charge within the one-valley model. Baldereschi[24] first calculated the intervalley mixing by perturbation theory by replacing ϵ_0 by $\epsilon(\Delta k)$ where Δk is the intervalley separation in $\vec{k}$ space. He obtained some splitting and 40.5 meV for the ground (A_1) state. The present calculation is the first in which the bare Coulomb potential is screened with $\epsilon(q)$ and the correct (though spherical) multivalley EME is used for the energy levels. Table II shows the status of the point-charge model through the years.

The present value of 48.8 meV for the point-charge ground state or its improved value using the correct ellipsoidal MV-EME is the appropriate value to use as the reference point for defining *chemical* shifts. In this framework, the chemical shift for phosphorus is very small and negative (− 3.3 meV). It should indeed be small, since the point charge should model accurately substitutional phosphorus. Furthermore, an inspection of Fig. 4 shows that $U(\vec{r})$ for phosphorus is slightly weaker than $U_{pc}(\vec{r})$, which accounts for the negative chemical shift of − 3.3 meV (it makes the phosphorus level *shallower* than that of the point charge). A similar pattern is seen to exist in the case of sulfur and the two-point-charge model (Table I and Fig. 5). As for the chemical shifts of the other donors (arsenic, interstitial lithium, etc.), they cannot be explained using the present KL-EMT. Indeed, in view of the large deviations of $U(\vec{r})$ for these donors from $U_{pc}(\vec{r})$ in the central cell shown

TABLE II. Donor ground state in silicon using a point-charge model for the perturbation, as calculated by various authors.

Authors	Year	One-valley or multivalley	Screening	Obtain splitting	$E(A_1)$ (meV)
KM[a]	1954	one-valley	$\epsilon(0)=12$	No	29.8
KL[b]	1955	one-valley	$\epsilon(0)=12$	No	29.0
BST[c]	1964	one-valley	$\epsilon(r)$	No	50.0
Müller[d]	1965	one-valley	$\epsilon(r)$	No	40.0
Faulkner[e]	1969	one-valley	$\epsilon(0)=11.4$	No	31.3
Baldereschi[f]	1970	multivalley	$\epsilon(0)$, $\epsilon(\Delta k)$	Yes	40.5
Present work[g]	1972	multivalley	$\epsilon(q)$	Yes	48.8
Experiment[h] for phosphorus				Yes	45.5

[a] Kittel and Mitchell, Ref. 4.

[b] Kohn and Luttinger, Ref. 6.

[c] Breitenecker, Sexl, and Thirring, Ref. 12; $\epsilon(r)$ is cut off at arbitrary radius $r_0 = 4.44$ a.u., result highly sensitive to this cut-off radius, see footnote d.

[d] Reference 11; same $\epsilon(r)$ as c with $r_0 = 3.55$ a.u.

[e] Reference 8.

[f] Reference 24. Intervalley terms are calculated by perturbation theory.

[g] Proper screening with $\epsilon(q)$. Direct calculation.

[h] Reference 27.

in Fig. 4, the KL-EMT is not even capable of indicating qualitatively why their ionization energies are close to that of the point charge and phosphorus. This fact will be explained using a new method in the following paper.[18]

Finally, even though in the present work ground-state energies were calculated directly by minimizing the total energy, it is instructive to examine the various contributions to the total energy by writing, symbolically,

$$E = \langle T\rangle_{\text{intra}} + \langle T\rangle_{\text{inter}} + \langle U\rangle_{\text{intra}} + \langle U\rangle_{\text{inter}} , \qquad (5.3)$$

where the subscripts denote intravalley terms [$i = j$ in the sums of Eq. (3.4)] and intervalley terms ($i \neq j$). The intervalley terms were omitted in previous one-valley calculations. For further illustration, we split U in U_{pc} and W as in Eq. (4.23) and we list in Table III the various contributions from variation calculations. It is clear that the intervalley terms are very important and cannot be neglected nor calculated accurately by perturbation theory as previously done.[24]

Equation (5.3) serves to illustrate another point. Instead of Eq. (4.23), if we write

$$U(\vec{r}) = -e^2/\epsilon r + W'(r) , \qquad (5.4)$$

where $W'(r)$ is the so-called "central-cell correction" and can be written as

$$W'(\vec{r}) = W(\vec{r}) + [U_{\text{pc}}(\vec{r}) - e^2/\epsilon r] , \qquad (5.5)$$

we get

$$E = \langle T\rangle_{\text{intra}} + \langle -e^2/\epsilon r\rangle_{\text{intra}} + [\langle T\rangle_{\text{inter}} + \langle -e^2/\epsilon r\rangle_{\text{inter}} + \langle W'\rangle_{\text{intra}} + \langle W'\rangle_{\text{inter}}] . \qquad (5.6)$$

In the original one-valley KL-EMT,[4-8] one ignores the square-bracketed term in Eq. (5.6). The first two terms give a variational value of[8] $E = E_0 = -31.3$ meV, measured from the bottom of the conduction band. The point to be made is that the quantity Δ defined by (5.2), and improperly called chemical shift, is *not* generally equal to the contribution of the terms in the square brackets in (5.6), even if W' is known exactly[46]; nor is Δ equal to $\langle W'\rangle_{\text{intra}}$, as assumed to be the case in Ref. 47 and, presumably, in Ref. 48 as well. The reason for this is that when the complete expression (5.6) or, equivalently, (5.3) is minimized, the minimum E is obtained at values of the variation parameters different from those obtained by omitting the term in square brackets in Eq. (5.6). Consequently the value of the first two terms in Eq. (5.7) is *no longer* -31.3 meV, and hence the value of the term in square brackets does not correspond to Δ. For example, for the phosphorus level calculated here, the value of the first two terms is only -23.7 meV. In the following paper,[18] we will see that this number may even be positive.

C. Accuracy of the calculations

The numbers appearing in Table I seem to agree very well with experiment. It is desirable to obtain an estimate of their uncertainty. This task can be divided in two parts—namely, investigating the accuracy of the multivalley EME and then the accuracy of our calculation of $U(\vec{r})$. For the first part, there are two major approximations: (a) the contribution from the high Fourier components of U is considered small, so that criterion (2.13) is satisfied—the one-band approximation and extending the sum over all $\vec{k}$-space are consequences of this—and (b) the k^4 and higher-order terms in the expansion of $E^0(k)$ may be dropped. Both of these contributions, which are not included in the final equations, have been estimated[21] and are discussed below. Each is found to be of order $(\lambda/a)^2$ smaller than the included contributions where $\lambda \sim 1.4$ a.u. and a is the variation parameter (orbit radius) of the state being calculated. For a shallow level ($a \sim 20$ a.u.) this correction is less than 1% and for a deeper level ($a \sim 5$ to 10 a.u.) it can be as high as 8%. Fortunately, however, the two omissions are of opposite sign, making the EMA quite accurate even for deep levels. Minor uncertainties are introduced through the value of m^* and k_0 (the distance of each minimum in $\vec{k}$ space from $k=0$, taken here to be 0.82 of the distance of each X point from $k=0$)[24]; these have also been studied.[21] The results for variations of m^* are displayed in Fig. 12 and show that a 1% uncertainty in m^* translates into 2% uncertainty in the energy level for phosphorus (shallow level) and 3% for sulfur (deep level). In Fig. 13 the energy for the point-charge model is plotted over a wider range of m^* to show that the dependence is quite complicated. The variation with k_0 is plotted in Fig. 14 showing that a 1% uncertainty in its value translates into 1% uncertainty in the value of E. Finally, the variational method employed in the calculations only gives an upper limit to the true eigenvalue. More flexible trial functions were used, but no appreciable lowering of the energy was obtained.

The uncertainties in the calculated impurity potentials arise from (a) neglect of lattice relaxation, (b) rigid undistorted neighboring ionic cores, (c) the use of free-atom core orbitals for the core-charge density in Eqs. (4.10)–(4.13), (d) the use of the Slater approximation for the exchange terms in the same equations, (e) the use of dielectric screening for $U_s(\vec{r})$, and (f) the choice of $\epsilon(q)$ from the literature. Estimates of the corresponding uncertainties have been made[21] and were found to be not very large. Some of the detailed results are presented in the following paper.[18] The net result is that over all, including all approximations for the EMA and the potential, the final numerical values are accurate to better than 5% for the shallow levels and to about 10% for the deep levels.

VI. EXCITED STATES

In the previous sections, we calculated and discussed the A_1 singlet ground states of donors in silicon using the KL-EMT with first-principles potentials. The excited-state wave functions either have zero or very small amplitudes at the

TABLE III. Various contributions to the donor ground-state energies. Units are in meV. E is given relative to the bottom of the conduction band. All symbols are defined in the text.

Impurity	$\langle T\rangle_{\text{intra}}$	$\langle T\rangle_{\text{inter}}$	$\langle T\rangle_{\text{total}}$	$\langle U_{\text{pc}}\rangle_{\text{intra}}$	$\langle U_{\text{pc}}\rangle_{\text{inter}}$	$\langle U_{\text{pc}}\rangle_{\text{total}}$	$\langle W\rangle_{\text{intra}}$	$\langle W\rangle_{\text{inter}}$	$\langle W\rangle_{\text{total}}$	$\langle U\rangle_{\text{total}}$	E
point charge	100.1	15.2	115.3	−118.0	−44.6	−162.6	...	...	...	−162.6	−47.3[a]
phosphorus	69.7	7.5	77.2	−96.8	−26.4	−123.3	0.3	3.9	4.2	−119.1	−41.9[b]
two point charges	1527.8	1971.1	3498.9	−1590.8	−2993.4	−4584.2	...	...	...	−4584.2	−1085.3
sulfur (S^+)	985.0	1033.8	2018.8	−1097.4	−1903.5	−3000.9	47.0	275.8	322.8	−2678.2	−659.3

[a] The contributions listed here are from a calculation using only $F_i^{1s}(r)$ as trial function. The energy is lowered to −48.8 meV by the inclusion of $F_i^{2s}(r)$.

[b] Same as above. The energy is lowered to −42.3 meV by the inclusion of $F_i^{2s}(r)$.

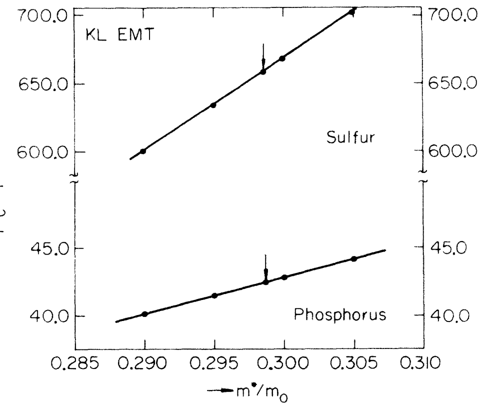

FIG. 12. Effect of uncertainty in m^* for KL-EMT energy levels.

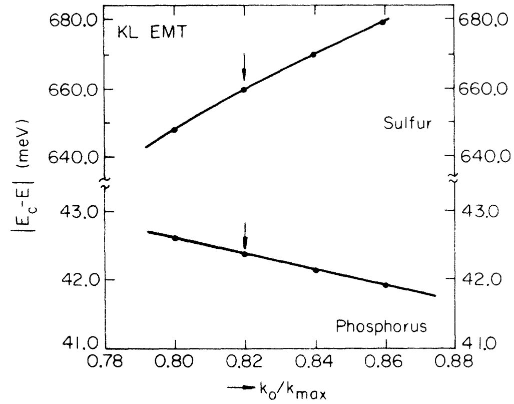

FIG. 14. Effect of uncertainty in k_0 for KL-EMT energy levels.

origin and the energies are largely determined by the Coulomb tail of $U(\vec{r})$. There are two categories of excited states, however. Directly above the $1s(A_1)$ ground state lie the $1s(T_2)$ triplet and the $1s(E)$ doublet which, in the one-valley approximation[4-8] are degenerate with the $1s(A_1)$ singlet. Above these $1s$ excited states are those corresponding to the hydrogenic spectrum ($2s$, $2p$, etc.). The latter are very accurately accounted for by a one-valley calculation using the potential $-e^2/\epsilon r$, as is shown in Ref. 8. The $1s(T_2)$ and $1s(E)$ states, however, have a basically $1s$ envelope and in the one-valley approximation, they both lie at the $1s(A_1)$ value of 31.3 meV. In the multivalley calculation, each is shifted slightly upward (Fig. 3) from this value because, for them, the intervalley terms almost cancel out, owing to the values of the constants α_i. The calculations are straightforward, as that of the A_1 level and the results are listed in Table IV. The simple trial function (4.35) was used, since the inclusion of a $2s$ contribution as in Eq. (4.32) was found to lower the energies by a negligible amount. The uncertainty is about the same as that of the A_1 levels, except that one thing is worse. The use of a spherical-band approximation is justified for the fully symmetric A_1 state where all six valleys come in with a coefficient $\alpha_i = 1/\sqrt{6}$. For the T_2 state, however, only two of the valleys come in with $\alpha_i = \pm 1/\sqrt{2}$, in which case the anisotropy of each valley should be important in calculating the intervalley mixing. This may account for the small discrepancy between theory and experiment.

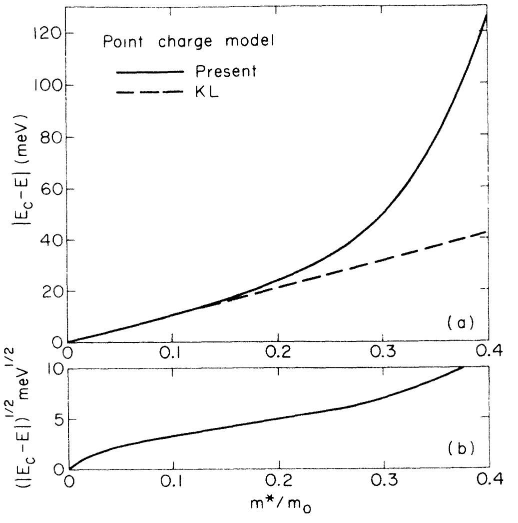

FIG. 13. Variation of the ground-state energy level of a point-charge potential with m^*: (a) illustrating the deviation from straight line, and (b) illustrating the deviation from a parabola.

VII. SUMMARY AND CONCLUSIONS

In this paper the Kohn-Luttinger effective-mass theory was reevaluated using impurity potentials constructed from first-principles crystal and atomic properties. In the past, it has generally been believed that the theory is valid for shallow levels and inapplicable to deep levels. In this paper a new conclusion was obtained which shows that the theory is indeed valid for both shallow and deep levels but only for the special cases when the impurity atom is substitutional and is from the same row of the Periodic Table as the host atom. Such atoms have the same number of core electrons and have been called isocoric for convenience. On the other hand, the theory was found to be invalid in the cases of the general nonisocoric impurities whether the levels are shallow or deep. These results have been confirmed by calculations of donors

TABLE IV. Energies of the other members of the split ground state. States of T_2 and E symmetry and s-like envelope. Reference point is the bottom of the conduction band.

Impurity	Symmetry	Theory a (a.u.)	Theory E (meV)	Experiment E (meV)
point charge	E	39.2	−30.4	···
	T_2	38.2	−31.1	···
phosphorus	E	39.0	−30.5	−32.6[a]
	T_2	37.8	−31.3	−33.9[a]
two point charges	E	20.0	−116.8	···
	T_2	18.0	−127.0	···
sulfur	E	19.6	−118.2	···
	T_2	17.4	−129.2	−188.3[b]

[a] Reference 27, corrected with theoretical result of Ref. 8 for $3p_{\pm}$ state.

[b] See Ref. 49. The identification of this level is somewhat uncertain due to the weakness of the transition.

in silicon. Excellent agreement with experiment was found for the isocoric phosphorus and sulfur energy levels (both the one- and two-electron sulfur energy levels are deep). The theory employs no adjustable parameters. The experimentally observed splitting of the ground state into a singlet, a doublet, and a triplet has been reproduced theoretically. The results were compared with those for a point-charge model and the concept of chemical shifts was redefined.

For the general nonisocoric substitutional and interstitial impurities it has been shown that a many-band expansion of the impurity-electron wave function is necessary. This will be accomplished by employing the pseudopotential transformation in the following paper.[18]

ACKNOWLEDGMENTS

We would like to express our sincere thanks to Professor John Bardeen for his many critical comments and advice, Dr. T. H. Ning for numerous discussions, Dr. N. O. Lipari for supplying a version of the Herman-Skillman program, Ray Marchand for discussions on mathematical questions, and the staff of the Sigma-5 computer at the Materials Research Laboratory for their patience and help.

APPENDIX A: USE OF DIELECTRIC SCREENING

In this appendix, we examine the conditions under which $U_s(\vec{r})$ may be calculated as linear response of the valence electrons to the introduction of $U_b(\vec{r})$. We have seen that

$$U_s(\vec{r}) = V_e(\vec{r}) - V_e^0(\vec{r}) , \tag{A1}$$

where $V_e^0(\vec{r})$ results from the charge density $\rho(\vec{r})$ of the electrons in the valence band of the perfect crystal, given by

$$\rho^0(\vec{r}) = \sum_{\vec{k}} \psi_{v\vec{k}}^{0*}(\vec{r}) \psi_{v\vec{k}}^0(\vec{r}) , \tag{A2}$$

where $\psi_{v\vec{k}}^0$ are the valence-band Bloch functions. Similarly $V_e(\vec{r})$ results from the charge density $\rho(\vec{r})$ of the electrons in the valence band of the perturbed crystal. Writing the perturbed band functions as $\psi_{v\vec{k}}(\vec{r})$ (no superscript), where $\vec{k}$ denotes only their asymptotic form away from the impurity, we have

$$\rho(\vec{r}) = \sum_{\vec{k}} \psi_{v\vec{k}}^{*}(\vec{r}) \psi_{vk}(\vec{r}) . \tag{A3}$$

In linear-response theory, $\psi_{v\vec{k}}$ is written as

$$\psi_{v\vec{k}} = \psi_{v\vec{k}}^0 + \sum_n \sum_{\vec{q}} a_n(\vec{k}+\vec{q}) \psi_{n,\vec{k}+\vec{q}}^0 \tag{A4}$$

and the coefficients $a_n(\vec{k}+\vec{q})$ are determined to first order in $U(\vec{r})$:

$$a_n(\vec{k}+\vec{q}) = \frac{\langle \psi_{n,\vec{k}+\vec{q}}^0 | U | \psi_{v\vec{k}}^0 \rangle}{E_v^0(\vec{k}) - E_n^0(\vec{k}+\vec{q})} . \tag{A5}$$

From Eqs. (A1)–(A5) one can then get Eq. (4.19) and an explicit expression for $\epsilon(q)$ in a straightforward way.[50] Clearly then this approach would apply when first-order perturbation theory is adequate for the $a_n(\vec{k}+\vec{q})$. This is so in the cases of phosphorus and sulfur in silicon which have almost identical cores with the host atom (Fig. 9) so that the nodal structure of $\psi_{v\vec{k}}$ is the same as that of $\psi_{v\vec{k}}^0$ over all the atomic sites, including the impurity site. It is not so, however, for other donors, such as arsenic, because $\psi_{v\vec{k}}$ has more nodes in the impurity cell than $\psi_{v\vec{k}}^0$, in order to satisfy the Pauli principle. First-order perturbation theory is not capable of introducing additional nodes.

The use of a dielectric function $\epsilon(\vec{q})$ which does not include umklapp terms, and hence ignores local-field corrections, is supported by the fact that umklapp terms would be important if $U(\vec{q})$ had strong high Fourier components $U(\vec{q}+\vec{K}_h)$. We saw that this is not true for the potentials calculated for phosphorus and sulfur in silicon. That this might be the case can also be demonstrated from the following physical argument: Silicon is a covalent solid because the crystal periodic potential $V^0(\vec{r})$ is strong in the tetrahedral directions. This means that the bonds are caused by strong high Fourier components, particularly a strong $V^0(\vec{K}_3)$ where $\vec{K}_3$ is the reciprocal-lattice vector in the [111] direction. If we assume that when a phosphorus or sulfur atom replaces a silicon atom, four of their valence electrons reconstruct the broken bonds with neighboring silicon atoms, this must be the result of $V(\vec{K}_3)$ being equally strong as $V^0(\vec{K}_3)$. Consequently, $U(\vec{K}_3) = V(\vec{K}_3) - V^0(\vec{K}_3)$ must be very small, or, more generally, U must not have strong high Fourier components.

Once $\epsilon(\vec{q})$ and U_b are known, therefore, $U(\vec{r})$ is calculable from

$$U(\vec{r}) = \int \frac{d^3q}{(2\pi)^3} \frac{U_b(\vec{q})}{\epsilon(\vec{q})} e^{-i\vec{q}\cdot\vec{r}} \ . \tag{A6}$$

This procedure is impractical for numerical potentials. For example, to get $W(\vec{r})$ from $W_b(\vec{r})$, instead of calculating $W_b(\vec{q})$ and using Eq. (A6) directly, we write

$$W_b(\vec{q}) = \int d^3r\, W_b(\vec{r})\, e^{i\vec{q}\cdot\vec{r}} \tag{A7}$$

and use it in Eq. (A6). It is easy to show then that $W(\vec{r})$ is given by Eq. (4.25) where

$$E(r, r') = \frac{2}{\pi} \int_0^\infty \frac{dq}{\epsilon(q)} \sin(qr)\sin(qr') \ . \tag{A8}$$

Using $\epsilon(q)$ given by Eq. (4.22), one gets

$$\begin{aligned} E(r, r') &= -A\alpha e^{-\alpha r'} \sinh(\alpha r) - (1-A)\beta e^{-\beta r'} \sinh(\beta r) \\ &\quad + \frac{1}{\epsilon(0)} \gamma e^{-\gamma r'} \sinh(\gamma r) \ , \quad r \leq r' \\ &= -A\alpha e^{-\alpha r} \sinh(\alpha r') - (1-A)\beta e^{-\beta r} \sinh(\beta r') \\ &\quad + \frac{1}{\epsilon(0)} \gamma e^{-\gamma r} \sinh(\gamma r'), \quad r \geq r' \end{aligned} \tag{A9}$$

and the integral (4.25) can be performed numerically.

APPENDIX B: ENERGY EXPRESSIONS

A. One-electron states

When the trial function (4.30) is used, a 2×2 secular matrix is diagonalized and the lower eigenvalue is minimized. Its matrix elements are written as

$$H_{mn} = (T_{mn} + U_{mn})/N_m \ . \tag{B1}$$

When the trial function (4.33) is used, the quantity

$$E(a) = H_{11}/N_1 \tag{B2}$$

is minimized. We have

$$T_{mn} = \sum_{i=1}^{6} \sum_{j=1}^{6} \alpha_j^* \alpha_i \langle F_j^{ms}(\vec{r}) | T_j(-i\nabla) | F_i^{ns}(\vec{r})\, e^{i(\vec{k}_i - \vec{k}_j)\cdot\vec{r}} \rangle , \tag{B3a}$$

$$U_{mn} = \sum_{i=1}^{6} \sum_{j=1}^{6} \alpha_j^* \alpha_i \langle F_j^{ms}(\vec{r})\, e^{-i\vec{k}_j\cdot\vec{r}} | U | F_i^{ns}(\vec{r})\, e^{i\vec{k}_i\cdot\vec{r}} \rangle , \tag{B3b}$$

$$N_m = \sum_{i=1}^{6} \sum_{j=1}^{6} \alpha_j^* \alpha_i \langle F_j^{ms}(\vec{r})\, e^{-i\vec{k}_j\cdot\vec{r}} | F_i^{ns}(\vec{r})\, e^{i\vec{k}_i\cdot\vec{r}} \rangle \ . \tag{B3c}$$

The sums and integrals can be carried out for any set of α_i's, such as sets corresponding to A_1, T_2, or E symmetry.[7] These are

$$\alpha(A_1) = \frac{1}{\sqrt{6}} (1, 1, 1, 1, 1, 1) \tag{B4a}$$

$$\begin{aligned} \alpha(E) &= \tfrac{1}{2}(1, 1, -1, -1, 0, 0) \\ &= \tfrac{1}{2}(1, 1, 0, 0, -1, -1) \end{aligned} \tag{B4b}$$

$$\begin{aligned} \alpha(T_2) &= \frac{1}{\sqrt{2}} (1, -1, 0, 0, 0, 0) \\ &= \frac{1}{\sqrt{2}} (0, 0, 1, -1, 0, 0) \\ &= \frac{1}{\sqrt{2}} (0, 0, 0, 0, 1, -1) \ . \end{aligned} \tag{B4c}$$

TABLE V. Constants $g_\lambda(\Gamma)$ resulting from the summation over the six equivalent minima of the conduction band of silicon.

g_λ \ Symmetry	A_1	E	T_2
g_1	1	1	1
g_2	1	1	−1
g_3	4	−2	0

The results can be given in compact form by making use of the constants $g_\lambda(\gamma)$ resulting from the summations (Table V). We have, in terms of the variation parameter a,

$$T_{11} = \frac{\hbar^2}{2m^*} \frac{1}{a^2} \sum_{\lambda=1}^{3} g_\lambda(\Gamma) \frac{8(2 + a^2\Delta_\lambda^2)}{(4 + a^2\Delta_\lambda^2)^2} , \tag{B5a}$$

$$T_{12} = \frac{\hbar^2}{2m^*} \frac{1}{\sqrt{2}\, a^2} \sum_{\lambda=1}^{3} g_\lambda(\Gamma) \frac{(1 + 2a^2\Delta_\lambda^2)^2 + \frac{23}{4}}{(\frac{9}{4} + a^2\Delta_\lambda^2)^3} , \tag{B5b}$$

$$T_{21} = \frac{\hbar^2}{2m^*} \frac{1}{\sqrt{2}\, a^2} \sum_{\lambda=1}^{3} g_\lambda(\Gamma) \frac{(\frac{5}{2} + 2a^2\Delta_\lambda^2)^2 + \frac{1}{2}}{(\frac{9}{4} + a^2\Delta_\lambda^2)^3} , \tag{B5c}$$

$$T_{22} = \frac{\hbar^2}{2m^*} \frac{1}{4a^2} \sum_{\lambda=1}^{3} g_\lambda(\Gamma) \times \left(\frac{1 + 4a^2\Delta_\lambda^2}{(1 + a^2\Delta_\lambda^2)^2} - \frac{3a^2\Delta_\lambda^2(1 - a^2\Delta_\lambda^2)}{(1 + a^2\Delta_\lambda^2)^4} \right) , \tag{B5d}$$

$$N_1 = \sum_{\lambda=1}^{3} g_\lambda(\Gamma) \frac{16}{(4 + a^2\Delta_\lambda^2)^2} , \tag{B6a}$$

$$N_2 = \sum_{\lambda=1}^{3} g_\lambda(\Gamma) \frac{(1 - 2a^2\Delta_\lambda^2)(1 - a^2\Delta_\lambda^2)}{(1 + a^2\Delta_\lambda^2)^4} , \tag{B6b}$$

where Δ_λ represent the distances between valleys, namely $\Delta_1 = 0$, $\Delta_2 = 2k_0$, $\Delta_3 = \sqrt{2}\, k_0$, and where k_0 is the distance of each of the six conduction-band minima from $k = 0$.

For the potential matrix elements we have two cases. First, the term $U_{\text{pc}}(\vec{r})$ is given in analytical form by Eq. (4.24) and explicit expressions are obtained. Writing Eq. (4.24) in the more compact form

$$U_{\text{pc}}(\vec{r}) = -\frac{ne^2}{\epsilon r}\sum_{\nu=1}^{4} S_\nu e^{-\sigma_\nu r} , \tag{B7}$$

where

$$S_\nu = (1, A\epsilon, (1-A)\epsilon, -1) \tag{B8a}$$

$$\sigma_\nu = (0, \alpha, \beta, \gamma) , \tag{B8b}$$

we get

$$(U_{\text{pc}})_{11} = -\frac{4ne^2}{\epsilon a}\sum_{\lambda=1}^{3} g_\lambda(\Gamma)\sum_{\nu=1}^{4}\frac{S_\nu}{(2+a\sigma_\nu)^2+a^2\Delta_\lambda^2} , \tag{B9a}$$

$$(U_{\text{pc}})_{12} = (U_{\text{pc}})_{21} = -\frac{\sqrt{2}\, ne^2}{\epsilon a}\sum_{\lambda=1}^{3} g_\lambda(\Gamma) \times \sum_{\nu=1}^{4}\frac{S_\nu}{d_{\lambda\nu}^2}(d_{\lambda\nu}-\tfrac{1}{4}) , \tag{B9b}$$

where

$$d_{\lambda\nu} = (1+a\sigma_\nu)^2 + a^2\Delta_\lambda^2 ,$$

and

$$(U_{\text{pc}})_{22} = -\frac{e^2}{2\epsilon a}\sum_{\lambda=1}^{3} g_\lambda(\Gamma) \times \sum_{\nu=1}^{4}\frac{S_\nu}{d_{\lambda\nu}^2}\left(a^2\sigma_\nu^2 + a^2\Delta_\lambda^2 - 2 + \frac{3(1+a\sigma_\nu)^2 - a^2\Delta_\lambda^2}{2d_{\lambda\nu}}\right). \tag{B9c}$$

For a short-range potential $W(\vec{r})$ given in numerical form, we get

$$W_{mn} = 4\pi\int r^2\, dr\, F_i^{ms}(\vec{r})W(r)F_i^{ns}(\vec{r}) \times \sum_{\lambda=1}^{3} g_\lambda(\Gamma)\,\frac{\sin(\Delta_\lambda r)}{\Delta_\lambda r} \tag{B10}$$

and the integrals are performed numerically. The potentials $W(r)$, are calculated as described in the text using the Herman-Skillman program for a mesh of 441 points, four times more dense than the ones published.[33]

B. Two-electron states

When two electrons are bound at a donor center, the ground state is when both electrons occupy the $1s(A_1)$ state with opposite spins, a situation analogous to the helium atom. We can easily, however, study the more general state with each electron occupying any of the three one-electron states $1s(A_1)$, $1s(T_2)$, and $1s(E)$. If we denote by Γ_1 the state of the one electron and by Γ_2 the state of the other electron (Γ_1 may or may not be equal to Γ_2), and use the simple trial function (4.33) with variation parameters a_1 and a_2, respectively, the total energy of the two electron system is given by

$$E_2(a_1, a_2) = E(\Gamma_1, a_1) + E(\Gamma_2, a_2) + \frac{\langle U_{\text{ee}}\rangle}{N_1(\Gamma_1)N_2(\Gamma_2)} , \tag{B11}$$

where $E(\Gamma, a)$ is the same as $E(a)$ defined by Eq. (B2), but with the symmetry dependence denoted explicitly. Similarly $N_1(\Gamma)$ is given by (B6a); again Γ denotes the symmetry dependence of N_1. Expressions for $E(\Gamma, a)$ and $N_1(\Gamma)$ were given in the first part of this appendix, so that now we concentrate on the electron-electron interaction term $\langle U_{\text{ee}}\rangle$. We have

$$\langle U_{\text{ee}}\rangle = \sum_{i=1}^{6}\sum_{j=1}^{6}\alpha_j^*(\Gamma_1)\alpha_i(\Gamma_1) \times \sum_{i'=1}^{6}\sum_{j'=1}^{6}\alpha_{j'}^*(\Gamma_2)\alpha_{i'}(\Gamma_2)\langle F_{j'}^{1s}(\vec{r}_1)e^{-i\vec{k}_j\cdot\vec{r}_1}F_{j'}^{1s}(\vec{r}_2)e^{-i\vec{k}_{j'}\cdot\vec{r}_2}|\,U_{\text{ee}}(\vec{r}_1,\vec{r}_2)\,|F_i^{1s}(\vec{r}_1)e^{i\vec{k}_i\cdot\vec{r}_1}F_{i'}^{1s}(\vec{r}_2)e^{i\vec{k}_{i'}\cdot\vec{r}_2}\rangle . \tag{B12}$$

Now, $U_{\text{ee}}(\vec{r}_1, \vec{r}_2)$ is a screened Coulomb interaction. The unscreened interaction is $e^2/|\vec{r}_1 - \vec{r}_2|$, whose Fourier transform is $4\pi e^2/q^2$, so that

$$U_{\text{ee}}(\vec{r}_1, \vec{r}_2) = (2\pi)^{-3}\int d^3q\,\frac{4\pi e^2}{q^2\epsilon(q)}\,e^{i\vec{q}\cdot(\vec{r}_1-\vec{r}_2)} . \tag{B13}$$

Angular integrals and sums may be performed exactly.[21] The result is

$$\langle U_{\text{ee}}\rangle = \frac{2e^2}{\pi}(16)^2\sum_{\lambda=1}^{3}\sum_{\lambda'=1}^{3} g_\lambda^*(\Gamma_2)g_{\lambda'}(\Gamma_1)\int_0^\infty\frac{dq}{\epsilon(q)}\,\frac{1}{J_{\lambda\lambda'}(q)} , \tag{B14}$$

where

$$J_{\lambda\lambda'}(q) = [4 + a_1^2(q-\Delta_\lambda)^2][4 + a_1^2(q+\Delta_\lambda)^2] \times [4 + a_2^2(q-\Delta_{\lambda'})^2][4 + a_2^2(q+\Delta_{\lambda'})^2] . \tag{B14$'$}$$

For the ground state, for which $\Gamma_1 = \Gamma_2 = A_1$ and a_1

$= a_2 = a$, this simplifies to[28]

$$\langle U_{ee} \rangle = \frac{2e^2}{\pi} (16)^2 \int_0^\infty \frac{dq}{\epsilon(q)}$$

$$\times \left(\sum_{\lambda_i}^{3} \frac{g_\lambda(A_1)}{[4 + a^2(q - \Delta_\lambda)^2]\,[4 + a^2(q + \Delta_\lambda)^2]} \right)^2 . \quad \text{(B15)}$$

The integrals over q in (B14) or (B15) are then performed numerically on a computer.

---

[*] Based on the doctoral thesis of S. T. P. submitted to the Department of Physics, University of Illinois.

[†] Supported in part by the Advanced Research Projects Agency (Contract No. NC-15-67-C-0221), the National Science Foundation (Contract No. GH-33634), and the Air Force Office of Scientific Research (Contract No. AFOSR-71-2067).

[‡] Supported in part by a University of Illinois Graduate Fellowship.

[§] Present address: W. W. Hansen Laboratories of Applied Physics, Stanford University, Stanford, Calif. 94305.

[1] For a review of the qualitative picture, see W. Shockley, *Electrons and Holes in Semiconductors* (Van Nostrand, New York, 1950), pp. 12 and 223 ff.

[2] J. C. Slater, Phys. Rev. 76, 1592 (1949).

[3] G. H. Wannier, Phys. Rev. 52, 191 (1937).

[4] C. Kittel and A. H. Mitchell, Phys. Rev. 96, 1488 (1954).

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

[6] W. Kohn and J. M. Luttinger, Phys. Rev. 98, 915 (1955).

[7] W. Kohn, Solid State Phys. 5, 257 (1957).

[8] R. A. Faulkner, Phys. Rev. 184, 713 (1969).

[9] G. Weinreich, J. Phys. Chem. Solids 8, 216 (1959); S. Shinohara, Nuovo Cimento 22, 18 (1961).

[10] P. Csavinszky, J. Phys. Chem. Solids 24, 1003 (1963).

[11] A. M. K. Müller, Solid State Commun. 2, 205 (1964); Z. Naturforsch. A 20, 1476 (1965).

[12] M. Breitenecker, R. Sexl, and W. Thirring, Z. Phys. 182, 123 (1964).

[13] J. Appel, Phys. Rev. 133, A280 (1964); L.J. Sham, Phys. Rev. 150, 720 (1966).

[14] J. C. Phillips, Phys. Rev. B 1, 1540 (1970).

[15] C. A. Coulson and M. J. Kearsley, Proc. R. Soc. A 241, 433 (1957); B. S. Gourary and A. E. Fein, J. Appl. Phys. 33, 331 (1962); T. Yamaguchi, J. Phys. Soc. Jap. 17, 1359 (1962); and the papers by G. D. Watkins and co-workers. See Ref. 16 for references.

[16] For a review of the early work, see J. W. Corbett, Solid State State Phys. Suppl. 7, 67 (1966); Solid State Phys. Suppl. 7, 155 (1966); and R. P. Messmer and G. D. Watkins, Phys. Rev. 7, 2568 (1973) for more recent works.

[17] J. Callaway and A. J. Hughes, Phys. Rev. 156, 860 (1967); K. Benneman, Phys. Rev. 137, A1497 (1965).

[18] Sokrates T. Pantelides and C. T. Sah, following paper, Phys. Rev. B 10, 638 (1974).

[19] Preliminary results were reported in a short communication; Sokrates T. Pantelides and C. T. Sah, Solid State Commun. 11, 1713 (1972).

[20] Equation (2.11) is exact for $\vec{k} = \vec{k}'$. See Ref. 5.

[21] Sokrates T. Pantelides, Doctoral thesis (University of Illinois 1973) (unpublished).

[22] W. D. Twose, in Appendix of Ref. 23.

[23] H. Fritzsche, Phys. Rev. 125, 1560 (1962).

[24] A. Baldereschi, Phys. Rev. B 1, 4673 (1970).

[25] Several earlier works have denoted this state by T_1. This is a matter of notation and the current consensus is to use T_2.

[26] R. L. Aggarwal, Solid State Commun. 2, 163 (1964).

[27] R. L. Aggarwal and A. K. Ramdas, Phys. Rev. 140, A1246 (1965).

[28] (a) T. H. Ning, Doctoral thesis, (University of Illinois, 1971) (unpublished); (b) T. H. Ning and C. T. Sah, Phys. Rev. B 4, 3468 (1971); Phys. Rev. B 4, 3482 (1971).

[29] D. L. Camphausen, J. M. James, and R. J. Sladek, Phys. Rev. B 2, 1899 (1970).

[30] L. L. Rosier and C. T. Sah, J. Solid State Electron. 13, 1547 (1970); J. Appl. Phys. 42, 4000 (1971).

[31] This separation is familiar from band-structure calculations. See, e.g., D. J. Stukel and R. N. Euwema, Phys. Rev. B 1, 1635 (1970).

[32] See, e.g., W. A. Harrison, *Solid State Theory* (McGraw-Hill, New York, 1970), p. 79.

[33] F. Herman and S. Skillman, *Atomic Structure Calculations* (Prentice-Hall, Englewood Cliffs, N. J., 1963).

[34] H. Nara, J. Phys. Soc. Jap. 20, 778 (1965).

[35] H. Nara and A. Morita, J. Phys. Soc. Jap. 21, 1852 (1966).

[36] D. Brust, Phys. Rev. B 5, 435 (1972).

[37] J. P. Walter and M. L. Cohen, Phys. Rev. B 2, 1821 (1970).

[38] W. E. Krag, W. H. Kleiner, H. J. Zeiger, and S. Fischler, J. Phys. Soc. Jap. Suppl. 21, 230 (1966); W. H. Kleiner and W. E. Krag, Phys. Rev. Lett. 25, 1490 (1970).

[39] L. L. Rosier and C. T. Sah, J. Solid State Electron. 14, 41 (1971).

[40] A. Morita and H. Nara, J. Phys. Soc. Jap. Suppl. 21, 234 (1966).

[41] It would probably be more appropriate to refer to these impurities as *isoelectronic* in parallel to the custom of calling, for example, GaAs the isoelectronic III-V compound to germanium. Unfortunately, the term isoelectronic has been used widely for impurities that have an equal number of *valence* electrons as the host atoms. Such impurities would be more appropriately called isovalent.

[42] P. Kaus, Phys. Rev. 109, 1944 (1958).

[43] A. Glodeanu, Phys. Status Solidi 19, K43 (1967).

[44] J. Hermanson, Phys. Rev. 150, 660 (1966).

[45] M. Jaros, J. Phys. C 4, 1162 (1971).

[46] In the case of interstitial lithium in silicon, $E_{\mathrm{exp}} = -31.0$ meV, so that the square-bracketed terms can be calculated by perturbation theory and, indeed, they would be equal to Δ.

[47] P. Csavinszky, J. Phys. Chem. Solids 24, 1003 (1963).

[48] J. C. Phillips, Phys. Rev. B 1, 1540 (1970).

[49] C. T. Sah, T. H. Ning, L. L. Rosier, and L. Forbes, Solid State Commun. 9, 917 (1971).

[50] This was done for a real crystal by H. Ehrenreich and M. H. Cohen, Phys. Rev. 115, 786 (1956), using the density-matrix formalism. A more elementary derivation is given in Appendix A of Ref. 21.