PHYSICAL REVIEW B VOLUME 21, NUMBER 8 15 APRIL 1980

Scattering-theoretic method for defects in semiconductors. II. Self-consistent formulation and application to the vacancy in silicon

J. Bernholc, Nunzio O. Lipari, and Sokrates T. Pantelides
IBM Thomas J. Watson Research Center, Yorktown Heights, New York 10598
(Received 15 November 1979)

A self-consistent-field method for calculation of the electronic structure of localized defects in semiconductors is described. The method is based on Green's-function theory and follows the original idea of Koster and Slater and its developments by Callaway and coworkers. The Wannier functions of the original formulations are, however, replaced by a more flexible set of linear combination of atomic orbitals. This choice and an accurate evaluation of the perfect-crystal Green's function bring this method to the level of sophistication, accuracy, and rigor characteristic of state-of-the-art band-structure and surface calculations. The efficiency of the method stems largely from the fact that it exploits both the translational symmetry of the host crystal and the short range of the defect potential. Thus, all bulk properties (e.g., band gaps, bandwidths, etc.) are built in from the start via a band-structure calculation and are preserved. One then focuses on the changes produced by the defect potential, so that the interpretation of the results is straightforward and unambiguous. In this paper, we report an application of this method to an isolated vacancy in Si assuming no lattice relaxation. The unrelaxed vacancy introduces a bound state of T_2 symmetry at 0.7 eV above the valence-band edge and a number of resonances and antiresonances within the valence bands. A detailed analysis of these states in terms of their origin, orbital content, and of state and charge densities is presented. We find that, while many of these states are individually quite extended, they combine destructively to produce a very localized net change in the charge density. We also find that the resulting localized potential can be well approximated by a negative of an atomic silicon potential extracted from a self-consistent bulk calculation. Finally, we compare the relative merits of the three increasingly more sophisticated, but also more costly, approaches to the defect problem, namely, (1) tight-binding, (2) non-self-consistent, and (3) self-consistent calculations.

I. INTRODUCTION

The calculation of the electronic structure and properties of a perfect crystalline semiconductor is at present routinely carried out by using band-theoretic techniques, which exploit translational periodicity (Bloch theorem). Defects, on the other hand, break the translational symmetry of the host crystal, and, as a result, the calculation of their electronic properties is a considerably more complicated problem.

One class of point defects is treated very successfully by the well-known effective-mass theory (EMT).[1] The theory works best for shallow donors and acceptors whose perturbation potential is dominated by the screened Coulombic tail responsible for the hydrogenic spectra near band edges. The corresponding wave functions are very extended in real space and hence highly localized in $\vec{k}$ space. Recently, new developments in EMT suggest that some moderately deep levels can also be handled by similar techniques.[2-4] For many deep levels, however, the EMT assumptions are not suitable and alternative techniques are necessary to describe wave functions which are highly localized in real space.

Two distinct approaches have thus far been pursued for the study of deep levels in semiconductors[5]:

One of them approximates a defect in a perfect infinite solid with a finite cluster. The Schrödinger equation for the cluster is then solved directly. A variety of techniques differing in the way the cluster is terminated have evolved: free clusters,[6-8] saturated clusters,[9,10] repeated clusters,[11,12] etc. Most of the work has been carried out on clusters consisting of 4–70 atoms. Clusters of that size, however, do not contain the usually quite extended defect wave function[7,11,12] and are therefore only suitable for qualitative analysis. Recently, techniques became available for calculating local densities of states of clusters of 2000 or more atoms.[13,14] These techniques, however, have thus far been used only with semiempirical tight-binding Hamiltonians. Furthermore, none of the cluster methods has proven to be particularly efficient. The main shortcoming of all these methods is that they must produce not only the defect-induced states, but also all the bulk states of the host, without exploiting the underlying translational periodicity.

In the second approach, the main shortcoming of the cluster methods is eliminated by first calculating the host-crystal properties using band theory and then focusing on the defect-induced changes, i.e., bound states, resonances, antiresonances, etc. A variety of techniques have evolved along these lines as well.[1,15-17] With the

exception of effective-mass theory, all other techniques are variants of a method that is usually attributed to Koster and Slater.[17] These techniques are in principle especially efficient when the perturbation potential U is *localized*, even though the individual wave functions may be extended. They therefore complement EMT-type theories, which rely on U having a dominant Coulombic tail.

Since the time of the original papers, the Koster-Slater method has been generalized in the language of scattering theory and Green's functions by Callaway.[18] Both Koster and Slater[17] and Callaway[18] developed the theory by introducing the host-crystal Wannier functions to expand wave functions and represent operators in matrix form. Callaway and Hughes[19] subsequently applied the method to the single vacancy in Si, but the construction of the Wannier functions proved to be very tedious and difficult, thus severely limiting the efficiency of the method. For that reason only a few other applications[20-22] have been made. A simplified tight-binding description of the vacancy in diamond and silicon was subsequently reported by Lannoo and Lenglart[23] and by Rouhani *et al.*,[24] but the approach was not pursued further.

More recently, a variant of the Koster-Slater method introduced by Bassani, Iadonisi, and Preziosi (BIP) (Ref. 25) has been used extensively by Jaros and Brand[26] and by Lindefelt.[27] The relationship between the BIP method and the Koster-Slater method is discussed in Refs. 5 and 28. Even more recently, there has been renewed interest in the original Koster-Slater method. Bernholc and Pantelides[28] (paper I) discussed the use of an operator formulation which reveals clearly that the role of the Wannier functions is simply to represent operators in matrix form. Thus, instead of the Wannier functions, other more convenient sets can be used. A particularly convenient choice is a set of linear combinations of atomic orbitals (LCAO) which are appropriate to carry out a band-structure calculation for the host material. Using the same LCAO set for the point-defect problem is then equivalent to expanding the perturbed-crystal wave functions in terms of the host-crystal Bloch functions or Wannier functions. In paper I,[28] empirical tight-binding Hamiltonians were used to study the ideal vacancy (simple removal of an atom without lattice reconstruction) in Si, Ge, GaAs. Since then, we have extended the same basic idea to include accurate self-consistent pseudopotential Hamiltonians for the host and to allow for electronic-charge redistribution that accompanies the removal of an atom, i.e., self-consistency for the defect as well. The resulting self-consistent Green's-function method for point defects has been used to obtain a detailed description of the unrelaxed vacancy in Si. The main results have been reported in a Letter.[29] The publication of that Letter coincided with the publication of a Letter by Baraff and Schlüter[30] who reported the development of a similar LCAO self-consistent Green's-function method and an application to the unrelaxed vacancy in Si.

In this paper we give a full account of our formulation of the self-consistent Green's-function method and its application to the unrelaxed vacancy in Si. This work brings the method to the same level of sophistication, accuracy, and resolving power as achieved by the most recent band-structure and surface calculations. In particular, the iteration to self-consistency frees our results from any dependence on the assumed similarity of interactions in the perturbed system to those in the unperturbed bulk crystal.[28] In this work we prove for the first time that the vacancy potential is very *localized*, while the individual vacancy states are quite *extended*. We also provide charge-density maps of the various states and make a detailed analysis of the electron distribution in the vicinity of the vacancy. Finally, we examine the effects of self-consistency and propose the bulk silicon potential to be used in inexpensive semiquantitative studies of vacancy complexes in silicon.

This paper is organized as follows. In Sec. II, the general Green's-function formalism is described. This section is based mainly on the papers by Koster and Slater,[17] Callaway,[18] and on general results of formal scattering theory.[31,32] The representation of operators is dealt with in Sec. III and the choice of basis functions is discussed in Sec. IV. Our computational procedures are outlined in Sec. V. Section VI contains the results of the unreconstructed vacancy in silicon and is followed by a brief summary and conclusions (Sec. VII). In a series of appendices we give additional details of our calculations and compare our formulation and results with those of Ref. 30.

II. SELF-CONSISTENT GREEN'S-FUNCTION METHOD

The Green's-function method assumes the knowledge of the solutions of the perfect-crystal Hamiltonian H^0:

$$H^0\psi^0_{n\vec{k}} = E_{n\vec{k}}\psi^0_{n\vec{k}}\,, \tag{1}$$

where $E_{n\vec{k}}$ and $\psi^0_{n\vec{k}}$ denote the band energies and wave functions, respectively. The corresponding Schrödinger equation for the imperfect solid is

$$H\psi = (H^0 + U)\psi = E\psi\,. \tag{2}$$

We define the perfect- and perturbed-crystal Green's operators by the relations

$$G^0(E^+)=\lim_{\epsilon\to 0^+}(E+i\epsilon-H^0)^{-1} \tag{3}$$

and

$$G(E^+)=\lim_{\epsilon\to 0^+}(E+i\epsilon-H)^{-1}. \tag{4}$$

Combining Eqs. (2)–(4), one immediately gets Dyson's equation, i.e.,

$$G(E^+)=G^0(E^+)+G^0(E^+)UG(E^+). \tag{5}$$

The formal solution of Eq. (5) is

$$G(E^+)=[1-G^0(E^+)U]^{-1}G^0(E^+). \tag{6}$$

Using the above definitions, one can immediately obtain formal solutions of Eq. (2). For energies within the band gaps, where $G^0(E)$ is real, the solution is

$$\psi=G^0(E)U\psi. \tag{7}$$

Within the band continua, one must add a solution of the homogeneous Eq. (1) to Eq. (7). We thus obtain the Lippman-Schwinger equation[33]

$$\psi_{n\vec{k}}=\psi^0_{n\vec{k}}+G^0(E^+)U\psi_{n\vec{k}} \tag{8}$$

(n and $\vec{k}$ are not conserved in the perturbed crystal, but are convenient labels for the scattering states). The condition for the existence of bound states is, from Eq. (7)

$$D(E)=\det\|1-G^0(E)U\|=0, \tag{9}$$

when the operator $1-G^0(E)U$ is expanded in any complete set of states. This condition determines the energy of the bound state. The corresponding wave function is then obtained as the nontrivial solution to Eq. (7) at that energy. For states within the band continua, we write Eq. (8) in the form

$$[1-G^0(E^+)U]\psi_{n\vec{k}}=\psi^0_{n\vec{k}}. \tag{10}$$

The determinant of $1-G^0(E^+)U$ is now nonzero because the imaginary part of $G^0(E^+)$ is nonzero within the band continua. As a result, the operator is always invertible and solutions exist at all energies within the energy bands of the perfect crystal. Band edges, therefore, are not shifted by the perturbation. The wave function ψ corresponding to a solution at $E=E^0_{n\vec{k}}$ is not, of course, equal to $\psi^0_{n\vec{k}}$.

The density operators for the perfect and perturbed crystal are related to the respective Green's operators by[31,32]

$$\rho^0(E)=-(2/\pi)\,\mathrm{Im}G^0(E) \tag{11}$$

and

$$\rho(E)=-(2/\pi)\,\mathrm{Im}G(E). \tag{12}$$

Using Eq. (6), we now obtain the following expression[31] for the change in the charge density $\Delta\rho$ induced by the defect

$$\Delta\rho=\int_{-\infty}^{E_F}[\rho(E)-\rho^0(E)]\,dE$$
$$=\frac{2}{\pi}\,\mathrm{Im}\int_{-\infty}^{E_F}\{1-[1-G^0(E)U]^{-1}\}G^0(E)\,dE. \tag{13}$$

This expression is of central importance for self-consistent calculations. Similarly, the state density $N(E)$ is altered in the vicinity of the defect. Using the relations[31]

$$\Delta N(E)=\mathrm{Tr}[\rho(E)-\rho^0(E)], \tag{14}$$

$$[G^0(E)]^2=-(d/dE)G^0(E), \tag{15}$$

and Eqs. (6), (11), and (12), one obtains

$$\Delta N(E)=(2/\pi)\mathrm{Im}\,\mathrm{Tr}\{(d/dE)[G^0(E)]\,U[1-G^0(E)U]^{-1}\} \tag{16}$$

by cyclic properties of the trace. Since the derivative of the density of states is infinite at a critical point, one should in general expect some structure in $\Delta N(E)$ at those points (dependent on the strength and details of the potential). The potential will also introduce *additional* structure through the factor $[1-G^0(E)U]^{-1}$. It can be shown[18] that the change in the state density is also given by

$$\Delta N(E)=\frac{2}{\pi}\frac{d\delta(E)}{dE}, \tag{17}$$

(spin included) where the phase shift $\delta(E)$ is defined by

$$\delta(E)=-\tan^{-1}[\mathrm{Im}\,D(E)/\mathrm{Re}\,D(E)]. \tag{18}$$

It follows that $\delta(E)$ goes through an odd multiple of $\pi/2$ every time $\mathrm{Re}\,D(E)=0$. An expansion around such an energy E_0 gives

$$\tan\delta(E)\approx-\Gamma/[2(E-E_0)], \tag{19}$$

where

$$\Gamma=2\mathrm{Im}\,D(E_0)/\mathrm{Re}\,D'(E_0) \tag{20}$$

and the prime denotes differentiation with respect to energy. In the region close to E_0, $\Delta N(E)$ becomes

$$\Delta N(E)=\frac{\Gamma}{2\pi}\frac{1}{(E-E_0)^2+\frac{1}{4}\Gamma^2}. \tag{21}$$

This characteristic Breit-Wigner form indicates that for $\Gamma>0$ one has a *resonance* or a peak in $\Delta N(E)$ with a half width Γ. For $\Gamma<0$, one has an *antiresonance*, i.e., a negative peak with half width $|\Gamma|$. (This analysis applies only for isolated resonances or antiresonances. If the spacing between a resonance and an antiresonance is smaller than the width, a more complicated spectrum oc-

curs.) The above considerations also assume that the background state density (i.e., the state density of the perfect crystal) is smooth. The resonances and antiresonances associated with the critical points, i.e., originating from the factor $(d/dE)G^0(E)$ in (16), will be called quasiresonances and quasiantiresonances, respectively.

According to Levinson's theorem[31] the total number of states remains unchanged in the presence of a perturbation, i.e.,

$$\int_{-\infty}^{\infty} \Delta N(E)\, dE = 0 . \tag{22}$$

When the states in the gaps are counted separately from the state density changes within the bands, Eq. (22) becomes

$$\int_{\text{bands}} \Delta N(E)\, dE = -N_b , \tag{23}$$

where N_b is the total number of bound states in the gaps.

Let us consider a continuous group of bands separated by band gaps, e.g., the valence bands of a semiconductor. Denoting by E_{bot} and E_{top} the bottom and the top of this group of bands, one obtains from (17)

$$\int_{E_{\text{bot}}}^{E_{\text{top}}} \Delta N(E)\, dE = \frac{2}{\pi}[\,\delta(E_{\text{top}}) - \delta(E_{\text{bot}})\,] . \tag{24}$$

Since the Green's function and therefore also the determinant $D(E)$ are real at the band edges it follows from (18) that

$$\int_{E_{\text{bot}}}^{E_{\text{top}}} \Delta N(E)\, dE = 2m , \tag{25}$$

where m is in integer, (including spin) independently of either the strength or the details of the potential.

III. REPRESENTATION OF OPERATORS

For applications to particular problems, the operators of the preceeding section must be represented in a basis set. The operator equations (7), (9), and (13) then become matrix equations. In order to exploit the limited range of the defect potential, we will assume that the individual basis functions are localized in space (for example, Wannier functions or LCAO's centered at the atomic positions). The subject of the specific choices of the basis sets will be discussed in the next section. In such a representation $\{\varphi_\alpha\}$, the matrix elements of the potential $U_{\alpha\beta}$ will be nonzero only if both φ_α and φ_β are centered on atoms close to the defect, so that they both overlap with the potential. The space spanned by this basis set can therefore be divided into two subspaces: subspace A which overlaps with the potential and subspace B in which the potential is effectively zero. The potential matrix may then be written schematically as

$$U = \begin{pmatrix} U_{AA} & 0 \\ 0 & 0 \end{pmatrix} . \tag{26}$$

(Note that the size of the subspace A depends also on the range of energies one wishes to study. For high-energy scattering, a larger subset would be necessary.) After writing the Green's-function matrix as

$$G^0(E) = \begin{pmatrix} G^0_{AA}(E) & G^0_{AB}(E) \\ G^0_{BA}(E) & G^0_{BB}(E) \end{pmatrix} , \tag{27}$$

the matrix of the operator $1 - G^0(E)U$ becomes

$$1 - G^0(E)U = \begin{pmatrix} 1 - G^0_{AA}(E)U_{AA} & 0 \\ -G^0_{BA}(E)U_{AA} & 1 \end{pmatrix} . \tag{28}$$

From (28) it is now evident that

$$\det\|1 - G^0(E)U\| = \det\|1 - G^0_{AA}(E)U_{AA}\| , \tag{29}$$

so that the size of the determinant reduces to the size of the nonzero part of the potential matrix.[17] Let us write the bound-state wave function ψ as $\{\psi_A, \psi_B\}$, where ψ_A and ψ_B are the components of ψ in the subspace A and subspace B, respectively. If E_0 is the bound-state energy, ψ_A is the nontrivial solution to the matrix equation

$$[\,1 - G^0_{AA}(E_0)U_{AA}\,]\psi_A = 0 ; \tag{30}$$

ψ_B can then be obtained from [cf. Eq. (7)]

$$\psi_B = G_{BA}(E_0)U_{AA}\psi_A . \tag{31}$$

The correct normalization of ψ_A can be obtained without calculating ψ_B (Refs. 34, 30) as follows: From Eq. (7) we get

$$\begin{aligned} 1 &= \langle\psi|\psi\rangle = \langle\psi U G^0(E)\,|\,G^0(E)U\psi\rangle \\ &= \langle\psi\,|\,U[\,G^0(E)\,]^2 U\,|\,\psi\rangle . \end{aligned} \tag{32}$$

Using the relation

$$U\psi = \{U_{AA}\psi_A, 0\} \tag{33}$$

and Eqs. (15) and (27), one obtains the normalization condition

$$-\langle\psi_A\,|\,U_{AA}G^{0\prime}_{AA}(E)U_{AA}\,|\,\psi_A\rangle = 1 . \tag{34}$$

Since the self-consistent total change in the charge density cannot extend further than the defect potential,[35] it is sufficient to calculate $\Delta\rho$ in Eq. (13) only in the subspace A. It follows from Eqs. (26), (27), and (28)

$$\Delta\rho_{AA}^{\text{val}} = \text{Im}\,\frac{2}{\pi}\int_{\text{val bands}} dE\{1-[1-G_{AA}^{0}(E)U_{AA}]^{-1}\} \times G_{AA}^{0}(E)\,. \tag{35}$$

The relations (29), (30), (34), and (35) show explicitly that the size of the problem in the Green's-function method is determined by the range of the perturbation potential even for self-consistent calculations.

In order to complete the operator formalism for the present Green's-function method, we give some general expressions for the evaluation of G^0. By introducing the Bloch functions $\psi_{n\vec{k}}^{0} = |n\vec{k}\rangle$ as a complete set, G^0 can be written in the form

$$G^0(E^+) = \sum_{n\vec{k}} \frac{|n\vec{k}\rangle\langle n\vec{k}|}{E^+ - E_{n\vec{k}}}\,. \tag{36}$$

Defining the spectral density operator $A^0(E)$ (Ref. 31) by

$$A^0(E) = \sum_{n\vec{k}} |n\vec{k}\rangle\langle n\vec{k}|\,\delta(E - E_{n\vec{k}})\,, \tag{37}$$

and using the Dirac identity

$$\lim_{\epsilon\to 0^+}\int \frac{f(t)\,dt}{t+i\epsilon} = \text{P}\int \frac{f(t)\,dt}{t} - i\pi f(0)\,, \tag{38}$$

where P denotes the principal part of the integral, we obtain

$$G^0(E^+) = \text{P}\int \frac{A(E')}{E-E'}\,dE' - i\pi A(E)\,. \tag{39}$$

IV. THE CHOICE OF THE BASIS SET

The selection of the basis set for the expansion of the operators in actual calculations is obviously problem-dependent and must be guided by physical considerations. It also constitutes one of the most important choices that determine the efficiency of the method.

The traditional basis set, i.e., the Wannier functions, have proved very convenient for the formal development of the method.[17,18] For actual calculations, on the other hand, they pose problems because their construction is quite laborious.[19,36] In this work we chose to use a set of LCAO's. Such sets have recently proved to be quite powerful and efficient in band-structure calculations[37,38] and a great deal of experience for their use has been amassed by atomic and molecular calculations.

The first step is to determine an LCAO set that is capable of yielding an accurate charge density and band structure for the perfect crystal (see Sec. V for the actual basis set used). The same set is then used to represent the operators $G^0(E)$ and U for the description of an unrelaxed vacancy. This choice can be justified as adequate for describing the perturbed crystal by invoking the usual quantum-chemical practice of using LCAO's characteristic of particular atoms only at the atomic sites. It is also adequate to describe the perturbation, because U is expected to be essentially the negative of an atomic-like potential. In other words, since the LCAO's at atomic sites in the perfect crystal are adequate for the description of bond formation, the same orbitals at the vacant site and the neighboring sites can be expected to describe the breaking of the bonds.

Our choice of a common LCAO set to describe both the perfect and the perturbed crystal has a number of useful consequences: in particular, by using an orthonormal LCAO set (see Sec. V for details), the operator equations of Secs. II and III can be read directly as matrix equations. The matrix elements between basis orbitals and Bloch functions necessary to evaluate the Green's-function matrix [cf. Eq. (36)] become very simple. Finally, because the perfect-crystal Hamiltonian has a finite spectrum, any ambiguities[39] associated with the termination of the energy integral of Eq. (39) are completely removed.

In order to exploit the point-group symmetry of the vacancy, we form symmetrized linear combinations of LCAO's on each shell of atoms surrounding the vacancy (shell orbitals). Finally, since the LCAO's, and therefore the shell orbitals are not orthogonal, we orthogonalize each shell orbital to all shell orbitals closer to the defect site, forming a set of orthonormal shell orbitals (OSO). This procedure, in contrast to the symmetric Löwdin orthogonalization,[40] has the advantage of preserving the localization of the basis functions and allows an easy and systematic study of the contributions of various shells and of convergence with respect to the number of shells. Also, additional shells may be added to the basis without the need of recalculating any of the matrix elements between existing orbitals. The orthonormalization process by itself avoids the need for carrying the overlap matrix in Eqs. (30)–(36), leading to large savings in computer time. The details of the symmetrization and orthogonalization procedure are given in Appendix A.

V. CALCULATIONS

In this section we will describe the calculational procedures used in applying the Green's-function method to study an ideal (undistorted) vacancy in Si. The calculations consist of three major steps: (1) Solution of the bulk problem and the calculation of the perfect-crystal Green's function G^0, (2) solution of the Green's-function equations for a given potential, and (3) construction of the self-

consistent potential. The methods of calculation employed at each step are described and tested in Secs. V A–C below.

A. Perfect crystal

1. Energy bands

In the self-consistent pseudopotential formalism,[41] the interaction between the valence electrons and the ionic core is modeled by a potential whose short-range form is fitted to the bulk properties,[41] atomic term values,[42] or atomic calculations.[43] The valence charge density, and thereby the valence crystalline potential, is then determined self-consistently within the local-density theory.[44] Usually, such calculations are carried out using a plane-wave basis set. One of the advantages of this set is that convergence studies can be carried out in a straightforward and systematic manner. We have therefore used a plane-wave basis set to calculate fully convergent energy bands for Si to serve as standards against which to compare our LCAO calculations. Such a calculation is necessary because we wish to use a minimal LCAO basis, optimizing the orbital exponents to yield bands that are in good agreement with known accurate calculations. We have carried out plane-wave calculations using the ionic potentials determined in Refs. 42 and 43. Both these potentials are specified in terms of the same analytical expression, namely,

$$V_{\text{ion}}(q) = (a_1/q^2)[\cos(a_2 q) + a_3]\, e^{a_4 q^4}. \tag{40}$$

In Ref. 42, the coefficients in Eq. (40) were fitted to experimental atomic term values, whereas in Ref. 43 they were fitted to atomic calculations. Both sets of coefficients are given in Table I. The charge density for each iteration was calculated using ten special k points of Chadi and Cohen.[45] We found 65 plane waves in the wave-function expansion to be sufficient to obtain a converged expansion of the crystalline charge density. The crystalline potential was calculated using the same exchange correlation as in the atomic calculations used to fit Eq. (40), i.e., the X_α approximation with $\alpha = 0.8$. The band-structure energies for the high-symmetry points are given in the first two columns of Table II. In particular, one should notice that the fundamental band gap is 0.6 and 0.8 eV for the two potentials. The discrepancy with the experimental value of 1.17 eV is attributed to the failure of the local approximation in the Kohn-Sham[44] density-functional formalism. In the following, we will use the ionic potential from Ref. 43, which is derived from atomic calculations. This choice makes our calculations both free from any empirical adjustments, and internally consistent, i.e., every quality has been calculated entirely within the local-density theory with the same exchange. In our defect calculations we use the LCAO's of Chadi[37] which consist of s, p, and d orbitals, and one f orbital transforming as the potential ($\sim xyz$). We have carried out fully self-consistent calculations for silicon in this basis without any shape approximation for the crystalline potential.[46] The band energies of the LCAO calculation with an LCAO self-consistent potential are compared with the fully converged plane-wave calculations in Table II. The largest discrepancy between the two calculations is about 0.2 eV, in the band-gap region, i.e., the same as obtained by Chadi[37] using a different potential. In particular, the band gap is 1.0 eV.

TABLE I. The values (in atomic units) of parameters defining the ionic Si pseudopotential V_{ion} [Eq. (41)] and bulk pseudopotential V_{bulk} [Eq. (43)]. The normalization volume for V_{ion} is 135.1 (a.u.)^3.

	$V^{\text{Si}}_{\text{ion}}$ (Ref. 42)	$V^{\text{Si}}_{\text{ion}}$ (Ref. 43)	$V^{\text{Si}}_{\text{bulk}}$
a_1	−1.12507	−1.233	20.0
a_2	0.79065	0.7370	0.633
a_3	0.35201	−0.3969	−17.7
a_4	−0.01807	−0.0177	0.459

2. Green's function

Once the band-structure calculation is carried out, the matrix elements of the Green's operator $G^0(E)$ between any pair of basis orbitals (the OSO's) can be evaluated using Eqs. (37) and (39). For a point defect of T_d symmetry, as is the case of the unrelaxed vacancy, the $\vec{k}$-space summation has to be done only over one irreducible wedge W, with the summand chosen by symmetry considerations so that the result is equivalent to a full-zone summation. If we denote each OSO by $|\alpha i l\rangle$, where α is the representation, i is a running index for OSO's, excluding partners, and l is a partner index, it is straightforward to show that

$$\langle \alpha i l | A(E) | \alpha' i' l' \rangle = \frac{48}{d_\alpha} \operatorname{Re}\Bigg(\sum_n \sum_{\vec{k}\in W} \sum_{l''} \langle \alpha i l'' | n\vec{k} \rangle \langle n\vec{k} | \alpha' i l'' \rangle \times \delta(E - E_n(\vec{k}))\, \delta_{\alpha\alpha'} \delta_{ll'} \Bigg), \tag{41}$$

where d_α is the dimensionality of the αth representation. [It is $\delta_{\alpha\alpha'}$ and $\delta_{ll'}$ in Eq. (41), which arise from symmetrization of the basis orbitals, that provide the savings in labor.] The evaluation of the matrix elements of $A(E)$ as given by Eq. (41) was carried out using the Gilat-Raubenheimer[47,48]

TABLE II. Band-structure energies for the high-symmetry points using potentials from Ref. 42 and Ref. 43. The point along the Δ axis is $\vec{k}=(0.85,0,0)$. The number of plane waves corresponds to an energy cutoff $|\vec{k}+\vec{G}|<9$ Ry. The Slater orbitals are those of Chadi (Ref. 37). The band energies for those orbitals have been shifted by 0.14 eV with respect to the plane-wave results (see Ref. 37).

	Plane waves[a]	Plane waves[b]	Slater orbitals[b]	Deviation from plane-wave[b] results
Γ_{1v}	−12.79	−12.89	−12.52	−0.37
$\Gamma_{25'v}$	0.00	0.00	0.00	
$\Gamma_{15'c}$	2.88	2.98	3.10	−0.72
$\Gamma_{2'c}$	3.49	2.72	2.92	−0.2
X_{1v}	−8.48	−8.63	−8.38	−0.25
X_{4v}	−3.14	−3.06	−2.95	−0.11
X_{1c}	0.78	0.95	1.11	−0.16
X_{4c}	11.83	11.67	11.93	−0.26
$L_{2'v}$	−10.39	−10.55	−10.25	−0.3
L_{1v}	− 7.52	− 7.52	− 7.27	−0.25
L_{3v}	− 1.36	− 1.34	− 1.29	−0.21
L_{1c}	1.60	1.38	1.59	−0.21
L_{3c}	3.69	3.89	4.07	−0.13
Δ_{1v}	− 9.63	− 9.76	− 9.50	−0.26
Δ_{2v}	− 7.21	− 7.38	−7.12	−0.26
Δ_{5v}	− 3.04	− 2.97	− 2.86	−0.11
Δ_{1c}	0.62	0.78	0.97	−0.19
Δ_{2c}	1.20	1.39	1.56	−0.17

[a] Potential from Ref. 42.
[b] Potential from Ref. 43.

technique. Convergence tests were carried out and it was found that 308 $\vec{k}$ points in the irreducible wedge were adequate. Once the matrix elements of $A(E)$ were calculated, the matrix elements of $G^0(E)$ were obtained using Eq. (39). The large number of Hilbert transforms of rather strongly varying functions with sharp peaks has been handled using a fast-Hilbert-transform (FHT) algorithm developed by Cooley and Bernholc.[49] The use of the FHT resulted in a very accurate determination of the Green's-function matrix elements on an energy mesh spaced by 0.07 eV. This mesh was found dense enough to allow interpolation for intermediate energy values when needed. The Green's-function matrix elements were calculated once and stored.

B. Solutions for a given defect potential (the non-self-consistent vacancy)

For the purpose of testing our computational scheme of determining the solution for a given defect potential, we first used the negative of a bulk self-consistent "atomic" pseudopotential, obtained by decomposing the full crystalline pseudopotential into a sum of atom-centered spherical pseudopotentials, neglecting nonspherical terms. The resulting atomic pseudopotential was fitted to an analytical expression of the form

$$V_{\text{bulk}}(r)=a_1e^{-a_2r^2}+a_3e^{-a_4r^2}. \tag{42}$$

The a_i's are given in Table I. Superposition of these atom-centered spherical potentials reproduces the total bulk pseudopotential quite adequately (Fig. 1). Thus, using the negative of one of these atom-centered potentials as a model defect potential corresponds to an approximation for the vacancy that amounts to removing an atom from the

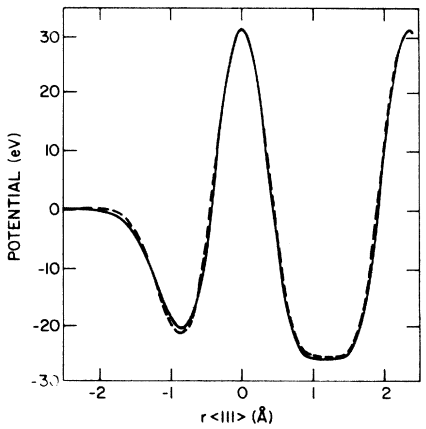

FIG. 1. Comparison of the truly self-consistent bulk silicon potential (dashed line) to the superposition of the atom-centered spherical potentials (solid line) along the bonding and antibonding directions. See text.

crystal without allowing the remaining valence electrons to redistribute (absence of self-consistency). The calculations using the non-self-consistent vacancy potential, as defined above, serve two purposes. First, they allow us to test the convergence properties of our computational procedure for a given defect potential, and, second, to assess the importance of self-consistency and the resulting screening fields once the fully self-consistent solutions are obtained (Sec. V C below).

Given the defect potential, its matrix elements between pairs of orthogonalized shell orbitals (OSO's) were calculated numerically on a cubic mesh. Symmetry was again used to reduce the integration region by a factor of 24. We found that 1000–2000 points per atom in the irreducible part of the cube were sufficient for an accuracy of 2–3 mRy in the potential matrix elements.[50] Having the matrix elements of both $G^0(E)$ and the defect potential U, the quantity $D(E)$ given by Eq. (9) is evaluated in a straightforward way. Making use of symmetry once more, one gets[18,19]

$$D(E)=\prod_{\alpha}[D_{\alpha}(E)]^{d_{\alpha}}, \tag{43}$$

where $D_{\alpha}(E)$ is defined as in Eq. (9), but evaluated in the subspace of OSO's of the αth representation. Accordingly, bound states in the gap belonging to the αth representation are obtained by locating the zeros of $D_{\alpha}(E)$. Similarly, one can show that

$$\Delta N(E)=\sum_{\alpha}\Delta N_{\alpha}(E), \tag{44}$$

where $\Delta N_{\alpha}(E)$ is again defined as $\Delta N(E)$, but within the subspace of OSO's of the αth representation (ΔN_{α} is defined so as to contain the degeneracy factor d_{α}).

The calculations were carried out using one, two, and three shells of orbitals (one shell means orbitals at the vacant site only). In all cases, we find only one bound state of T_2 symmetry. Its energy level is at 0.75, 0.75, and 0.76 eV for one, two, and three shells, respectively, indicating an adequate convergence with respect to the number of shells. The corresponding changes in state densities within the band continua, shown as dashed curves in Fig. 2, were also found to be indistinguishable for one, two, and three shells on the scale of the figure. (Only changes in densities of states of A_1 and T_2 symmetries are shown. Changes in the densities of states of the other symmetries are of order 0.1 and thus insignificant by comparison. These results are interpreted in Sec. VI, where comparison is made with fully self-consistent results.) Integrals of the state-density changes provide an additional check of the calculations. It can easily be shown that Levinson's theorem [Eqs. (22) and (23)] is valid for states of each symmetry separately. Thus, the state-density changes for each symmetry ought to integrate to the number (integer) of bound states of the same symmetry. Indeed, we found that the change in the density of T_2 states integrates to -6, compensating the sixfold degenerate T_2 bound state in the fundamental gap. The changes in the densities of states of the remaining symmetries integrate to zero, as they should. Thus, since the creation of a neutral vacancy involves the removal of four electrons, we conclude that the T_2 bound state contains two electrons.

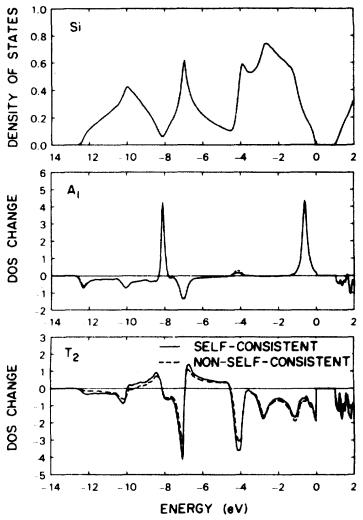

FIG. 2. The density of states of silicon (top panel) and the change in the density of states of A_1 and T_2 symmetries. The curves are broadened by 0.2 eV and the reference energy is the top of the valence bands.

C. Fully self-consistent solutions

After the calculations have been initiated with a starting potential (e.g., the non-self-consistent potential described above or some other convenient form), the Green's-function equations have to be iterated to self-consistency. At every iteration the construction of a new potential consists of the following steps: (1) calculation of the change in charge density in the valence bands and of the charge density associated with the bound states, (2) solution of the Poisson equation, and (3) calculation of the exchange-correlation term (within the local-density theory[44]).

The change in the charge density associated with the valence bands was obtained by carrying out the matrix integral (35). Since the integrand is a rather singular function (see below), the integral was calculated by the adaptive (variable-step size)

trapezoidal rule. The average step length was 0.07 eV and was monitored by examining $\mathrm{Tr}\Delta\rho(E)$, which is equal to the change in the density of states at that energy. The change in the charge density in real space is now

$$\Delta\rho^{\mathrm{val}}(\mathbf{r})=\sum_{\alpha\alpha'}\psi_{\alpha}^{*}(\mathbf{r})\Delta\rho_{\alpha\alpha'}^{\mathrm{val}}\psi_{\alpha'}(\mathbf{r})\,. \tag{45}$$

The charge density associated with the two electrons in the bound state was obtained from the bound-state wave function within the region spanned by Eq. (30), and the normalization constant was calculated from relation (34). It will be shown in the next section by explicit construction that the rather extended tail of the bound-state wave function is canceled by the charge redistribution within the valence bands localizing the total change in the charge density almost entirely within the cavity defined by the nearest neighbors. This localization is important to the efficiency of the method and has been exploited in the next stages of the calculation.

The Hartree part of the defect potential V_H can now be obtained by solving the Poisson equation

$$\nabla^2 V_H=-8\pi\Delta\rho\,, \tag{46}$$

with the boundary condition, for large r,

$$V(\mathbf{r})\sim -Z/r\,, \tag{47}$$

where

$$Z=-\int d^3r\,\Delta\rho(\mathbf{r})\,. \tag{48}$$

In order to facilitate a numerical solution, it is convenient to add and later subtract a neutralizing, positive, and spherically symmetric charge density ρ^{+}. The charge

$$\rho^{N}=\Delta\rho+\rho^{+} \tag{49}$$

is then neutral and the corresponding potential V^N behaves as $\mathrm{const}\times r^{-(l+1)}$ at large distances where l is the lowest multipole component of the defect potential allowed by symmetry ($l=3$ for the T_d group). V^N is determined by solving Poisson's equation in a large cube (cube side = 18 a.u.) around the vacancy. Because of the smallness of the r^{-4} and higher-order terms on the surface of the cube, the boundary condition is taken to be $V^N=0$. This geometry allows us to use the very efficient fast-Fourier-transform[51] techniques in order to solve the Poisson equation numerically,[51,46,52] the only constraint being that the numerical mesh must be equally spaced (see Appendix B for the discussion of the algorithm). This algorithm defines the cubic mesh for the calculation of the potential matrix elements (cf. Sec. V B and Ref. 50). The Hartree potential is obtained by subtracting the potential V^{+} corresponding to the charge density ρ^{+}. In our calculations, we choose

$$\rho^{+}(\vec{\mathbf{r}})=\frac{1}{\pi^{3/2}}\frac{Z}{r_0^3}\exp(-r^2/r_0^2)\,, \tag{50}$$

whereby

$$V^{+}(\vec{\mathbf{r}})=\frac{2Z}{r}\,\mathrm{erf}\left(\frac{|\vec{\mathbf{r}}|}{r_0}\right). \tag{51}$$

Finally, the exchange-correlation potential is calculated in the local-density approximation. Because of the $\rho^{1/3}$ dependence of this potential on the charge density, it is necessary to calculate both the perfect-crystal charge density $\rho^0(r)$ as well as $\Delta\rho(r)$ at each mesh point. The defect exchange-correlation potential is then proportional to $[\rho^0(r)+\Delta\rho(r)]^{1/3}-[\Delta\rho(r)]^{1/2}$. The calculation is done numerically[50] on the same mesh points used for the Hartree part of the potential.

VI. RESULTS AND DISCUSSION

In this section we present the results of our self-consistent calculations for the unrelaxed vacancy in Si. The calculations were brought to self-consistency using one, two, and three shells for the purpose of checking convergence of charge densities and wave functions with respect to the number of shells.

A. Energy levels and state densities

As with the non-self-consistent calculations described in the previous section, we obtain one bound state in the gap of T_2 symmetry. Its energy level was found to be at 0.63, 0.66, and 0.68 eV for one, two, and three shells, respectively. Convergence is therefore somewhat slower, but still quite adequate. We note that self-consistency has lowered the position of the bound state by approximately 0.1 eV. This lowering is the result of screening arising from the response of the valence electrons to the removal of the atom, which reduces the strength of the defect potential. The final self-consistent vacancy potential is shown in Fig. 3 and compared with the non-self-consistent vacancy potential defined in Sec. V. Note that the effect of self-consistency is very small, as reflected by the lowering of the bound-state energy level by less than 0.1 eV.

The changes in the densities of states of A_1 and T_2 symmetries are shown as solid curves in Fig. 2. On the scale of the figure, the curves obtained from the one-, two-, and three-shell calculations are indistinguishable. Changes in the densities of states of the other symmetries are again insignificant by comparison. The change in the densities

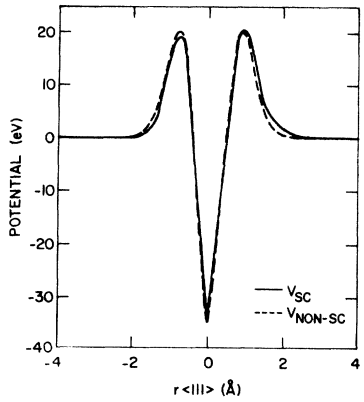

FIG. 3. Comparison of the fully self-consistent vacancy potential to the negative of the spherical atom-centered potential extracted from bulk data. See text.

of T_2 states integrates to -6, as required by Levinson's theorem, and the changes in the densities of states of the remaining symmetries integrate to zero. The self-consistent state-density changes are compared in Fig. 2 with the corresponding changes obtained without self-consistency, indicating once more that the effect of self-consistency is small.

In Fig. 2 we have included a plot of the density of states of the perfect crystal (top panel), against which the calculated changes induced by the vacancy can be compared. We note that a sharp resonance of A_1 symmetry appears at -8 eV, where the perfect-crystal state density has a minimum. The antiresonances appear at or near maxima in the perfect-crystal state densities. These results are consistent with the general analysis given in Sec. II B, since extrema in the perfect-crystal state density correspond to critical points in the band structure. The A_1 resonance at -0.7 eV, on the other hand, is caused by the particular nature of the vacancy potential and will be discussed further later on.

B. Wave functions and charge densities

For further analysis of the nature of the solutions we examine wave functions and charge densities. The most convenient and informative way to display these results is in the form of contour maps in a (110) plane. All the results that will be presented were obtained from three-shell calculations. The three-shell basis is sufficiently large to allow plotting beyond the second neighbors of the vacancy.

In Fig. 4 we show the charge density of the perfect crystal, the charge density in a crystal containing a single vacancy, and the change in the charge density produced by the introduction of the

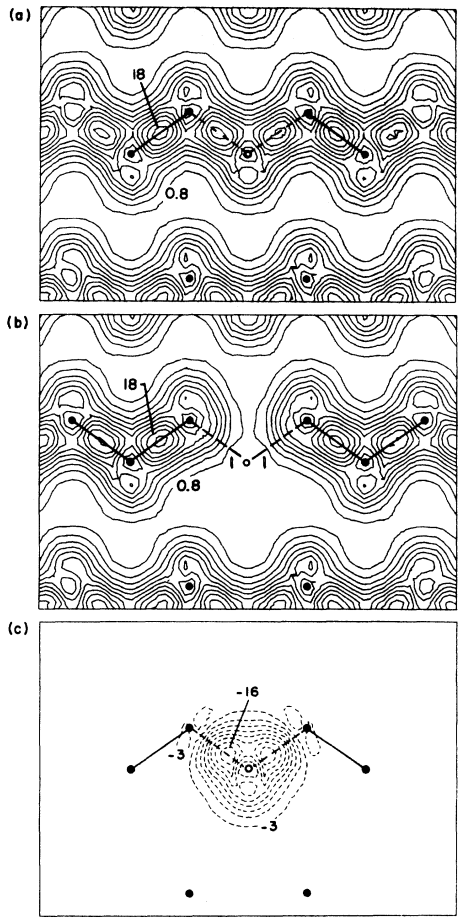

FIG. 4. Contours of constant electron density (in electrons per bulk Si unit cell) for (a) charge density of perfect silicon. (b) Charge density in the presence of the vacancy. (c) The change in the charge density i.e., the difference between (b) and (a).

vacancy. As an internal check of the calculations, the change in the charge density is found to integrate to -4, corresponding to the net number of electrons that have been removed from the crystal. We note that the change in the charge density is localized almost entirely within the cavity defined by the nearest neighbors. This charge produces a potential that becomes equal to $-4e^2/r$ beyond the nearest neighbors and cancels exactly the $4e^2/r$ tail of the ionic part of the vacancy potential. The net vacancy potential is therefore almost completely localized within a distance of a bond length, as we already saw in Fig. 3. These results are consistent with the fact that orbitals on the second and the third shells of atoms do not contribute significantly.

We turn now to examine the wave functions of individual states. In Fig. 5 we show a contour plot of the square of the wave function of the T_2 bound

TABLE III. The integrated changes in the valence-charge density for each representation.

Representation	Charge(electrons)
A_1	−0.10
A_2	0.00
E	−0.04
T_1	−0.03
T_2	−5.20

state. *In contrast to the total change in the charge density* [Fig. 4(c)], *we see that the charge density associated with the bound state is quite extended.* In fact, the basis orbitals on the three shells of atoms included in the plot of Fig. 5 contain only approximately 70% of the total bound-state charge (1.4 bound-state electrons lie inside the three-shell volume). Note that this result does not imply that additional shells of orbitals ought to have been included in the calculation. Recall that an important virtue of the Green's-function method is that it does not require a basis set capable of expanding individual wave functions. The role of the basis set is to represent the defect potential and *change* in the charge density. As we saw already, both of these quantities are quite localized.[53]

The rather delocalized bound-state charge density might at first glance appear to be inconsistent with the fact that the total change in the charge density is highly localized. Note that previous non-self-consistent calculations *assumed* a localized vacancy potential and obtained a delocalized bound state. The present self-consistent calculations confirm this assumption and provide a justification: In addition to the T_2 bound state in the gap, the defect potential induces a series of resonances and antiresonances in the valence bands, which we already saw in Fig. 2. Contour plots of the charge density associated with some of the resonances are shown in Figs. 6 and 7, revealing that the corresponding charge density is quite delocalized. It is a subtle cancellation of the tails of individual states that produces the strong localization of the *net* charge disturbance. Another illustration of this fact is provided by plots of the total change in the charge density for each type of symmetry separately (Fig. 8). Each of them is quite delocalized [the solid contours are the sum of all positive changes (resonances), whereas the dashed contours are the sum of all negative changes (antiresonances)] and not necessarily contained in the volume defined by three shells of atoms. In fact their integrals within the three-shell basis are given in Table III. All five together add up to −5.4. Recall that the bound state contributes only 1.4 electrons in the volume defined by three shells of atoms. The net change within this volume is therefore − 4, indicating that indeed the *net* change outside the box is zero and providing an additional internal check of the calculations.

The above results for the total charge densities, the defect-induced change in the total charge density and individual wave functions provide a clear illustration of the advantages of the Green's-function method over cluster methods. Cluster methods attempt to describe the perturbed crystal directly [Fig. 4(b)] and terminate it as a matter of necessity to make the calculation feasible. In contrast, the Green's-function method describes first the infinite perfect crystal [Fig. 4(a)] and then focuses on the change [Fig. 4(c)] which is naturally localized. Furthermore, whereas the Green's-function method needs a basis set capable of de-

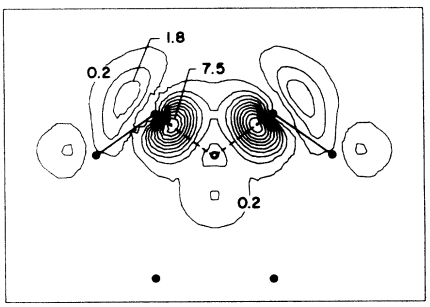

FIG. 5. Contours of constant electron density for the T_2 bound state. Units as in Fig. 4.

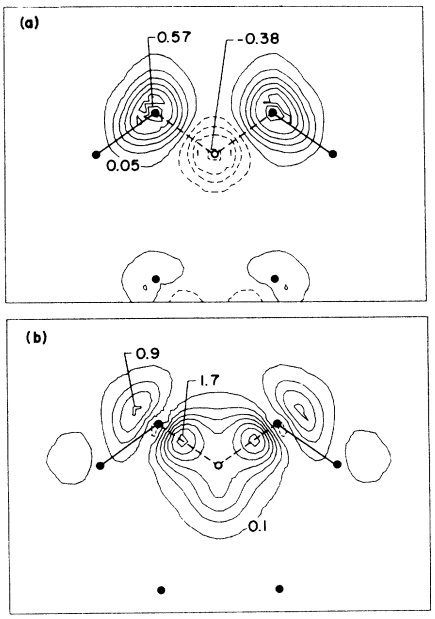

FIG. 6. Contours of constant electron density for the A_1 resonances at (a) −8 and (b) − 0.7 eV. Units as in Fig. 4.

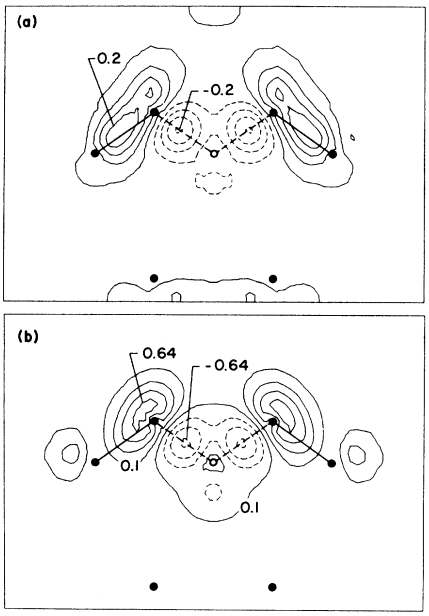

FIG. 7. Contours of constant electron density for the T_2 quasiresonances at (a) −8.4 and (b) −6.8 eV. Units as in Fig. 4.

scribing accurately only the total change in the charge density, which is highly localized [Fig. 4 (c)], the cluster method needs a basis set capable of describing accurately individual wave functions (Fig. 5) which are considerably more extended. Only when the size of the cluster is large enough to contain individual wave functions would a cluster calculation give reliable quantitative results.

C. LCAO analysis

The use of an LCAO basis for the calculations provides a convenient framework for exploring the structure and origins of the solutions. First, we observe that the valence electrons in the perfect Si crystal have mainly s (A_1) and p (T_2) character about each atom. This is a well-known result that has often been exploited in constructing simple

TABLE IV. The decomposition of the bound-state wave function of the undistorted vacancy into the bands of perfect silicon $c_n = \sum_k |\langle n\vec{k} | \psi \rangle|^2$.

Band no.	c_n (%)
1	0.4
2	4.8
3	23.7
4	51.1
5	10.9
6	4.1
7	2.2
8	1.4
9	0.6
10	0.3

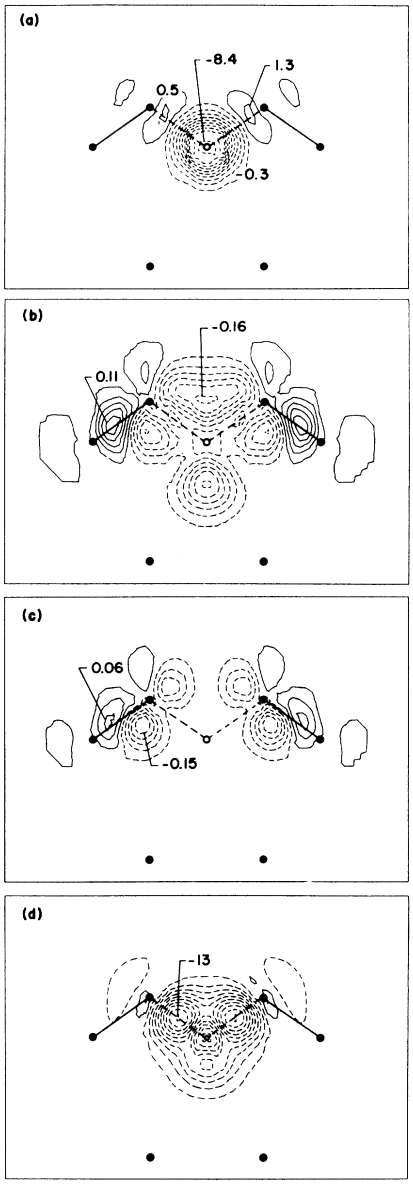

FIG. 8. Contours of constant electron density for the change in the charge density of symmetry (a) A_1, (b) E, (c) T_1, (d) T_2. Units as in Fig. 4.

semiempirical tight-binding Hamiltonians.[54] Our self-consistent calculations reveal that the valence electrons also have a small d ($E + T_2$) character.[55] Since the creation of a vacancy amounts to removing four valence electrons from the crystal, one can expect that the major changes in the charge and state densities occur in the A_1 and T_2 representations. Our results, Figs. 2 and 8, confirmed this simple LCAO prediction. In particular, note in Fig. 8 that the A_1 and T_2 changes in the charge density are large and concentrated in the cavity defined by the nearest neighbors. (The T_2 change

does not have a p-like shape because it is an average of the "x", "y", and "z" components.) Note also the small change in the E representation (d-like) in Fig. 8(c) in agreement with the LCAO analysis given above. Figure 8(d) shows the change of charge density in the T_1 states. This change, which is also quite small compared with the changes in the charge density of A_1 and T_2 states, comes from linear combinations of orbitals on the nearest neighbors and is therefore localized about those sites. Finally, the change of charge density of A_2 states is too small to be seen on the scale of Fig. 8, because it arises from linear combinations of orbitals on the second-nearest neighbors.

We now focus our attention on the A_1 and T_2 results and show that they can be understood in terms of some very elementary LCAO models. For the perfect crystal, one can use s and p orbitals on every atom and combine them into tetrahedrally directed sp^3 hybrid orbitals. These hybrids can then be combined in pairs to form bonding and antibonding orbitals, which in turn give rise to the valence and conduction bands, respectively.[56] For the crystal containing a single vacancy, one can again use s and p orbitals on every atom and form sp^3 hybrids. Bonding and antibonding orbitals can again be formed by pairing hybrids, with the exception of the four hybrids on the nearest neighbors pointing toward the vacant site. We will refer to these orbitals as "dangling hybrids" (the charge associated with such an orbital is often referred to as a "dangling bond"). As a first approximation, one would expect that the states associated with the vacancy are linear combinations of these four hybrids. This assumption forms the basis of the "defect-molecule" model, originally proposed by Coulson and Kearsley[6] for the vacancy in diamond. Symmetry requires that the four hybrids be combined into an A_1 singlet and a T_2 triplet. In fact, our self-consistent pseudopotential calculations reveal that the T_2 bound state (Fig. 5) and the A_1 resonance at -0.7 eV [Fig. 6(b)] are essentially linear combinations of the dangling-hybrid-like orbitals. The remaining resonances and antiresonances are mainly associated with critical points in the band structure (cf. Secs. II and VI A). The A_1 resonance at about -8 eV appears precisely at the point where the density of states has a cusp. (This cusp arises from the X_1 point in the valence bands and is a peculiarity of the diamond lattice. It may be viewed as a pseudogap, since it is precisely at the point where a gap opens up in the zinc-blende structure.) In fact, this resonance is a band-structure effect and appears at the same place for several substitutional acceptors.[57] The contour plot of the charge redistribution corresponding to this resonance is shown in Fig. 6(a). Note that it is quite localized since it is essentially a bound state. It consists mainly of s-like orbitals on the nearest neighbors.

In Fig. 7 we show the changes in the charge density corresponding to the two T_2 quasiresonances at (a) -8.4 eV and (b) -6.8 eV, as obtained from the self-consistent calculations. The quasiresonance at -8.4 corresponds to a strengthening of the back bonds [recall however, that, overall, the back bonds are not strengthened because the *total* change is localized entirely within the cavity defined by the nearest neighbors; cf Fig. 4(c).] The quasiresonance at -6.8 eV shifts some of the charge into the antibonding direction in the normally empty interstitial regions.

D. The effective-mass nature of the bound state

In the self-consistent Green's-function calculations, we obtain the bound-state wave functions in terms of our LCAO basis set in the form

$$\psi(\vec{r}) = \sum_{\mu} C_{\mu} \phi_{\mu}(\vec{r}) , \tag{52}$$

where ϕ_{μ} is an OSO (for simplicity, we use $\mu = \{\alpha, i, l\}$ as a composite index for an OSO). A common expansion of the ψ, e.g., in the case of effective-mass theory, is in terms of Bloch functions in the form

$$\psi(\vec{r}) = \sum_{n\vec{k}} F_n(\vec{k}) \psi_{n\vec{k}}(\vec{r}) . \tag{53}$$

Combining (52) and (53), we immediately obtain

$$F_n(\vec{k}) = \langle n\vec{k} | \psi \rangle = \sum_{\mu} C_{\mu} \langle n\vec{k} | \phi_{\mu} \rangle . \tag{54}$$

An alternative expression can be obtained by using Eqs. (7) and (36):

$$F_n(\vec{k}) = \sum_{\mu} C_{\mu} \langle nk | U | \phi_{\mu} \rangle / (E - E_{n\vec{k}}) . \tag{55}$$

This expression shows that only those OSO's for which the matrix elements of U are nonzero need be kept in calculating $F_n(k)$. We have calculated these quantities and summed them over the Brillouin zone for each band in order to establish the relative weight of each band in the expansion of ψ. The results are given in Table IV. We note that 80% of the wave function comes from the top three valence bands, with the remainder distributed over a total of seven more bands. The lowest conduction band contributes 11%, whereas the rest of the conduction bands together contribute less than 9%.

In Fig. 9 we give plots of the various $F_n(\vec{k})$ functions along two important symmetry directions in $\vec{k}$ space. Only the $F_n(\vec{k})$ for bands 2, 3, 4, and 5 can be plotted on the scale of Fig. 9. This figure

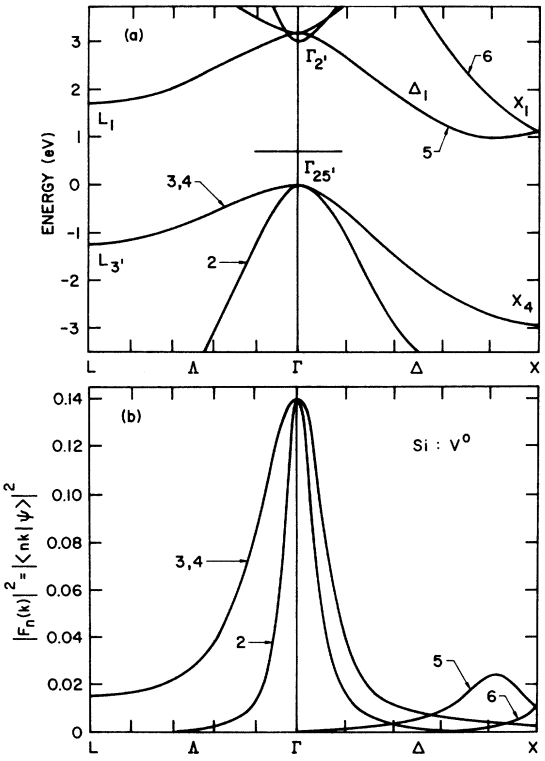

FIG. 9. The band structure of silicon of the vicinity of the band gap and the envelope functions $F_n(\vec{k})$ plotted along the Λ and Δ directions. See text.

illustrates clearly that ψ is composed mainly from Bloch functions in the vicinity of the valence-band maximum at $\vec{k}=0$, very much like effective-mass-like bound states. We may recall at this stage that effective-mass equations have recently been rederived[58] for an arbitrary impurity potential by making only two approximations: (1) restricting the number of bands that participate in the expansion of the bound-state wavefunction, and (2) evaluating all matrix elements to order k^2 (thus including the so-called umklapp terms[59]). The results shown in Fig. 9 for the vacancy bound-state wave function indicate that the new generalized effective-mass equations might be applicable even for quite deep levels. Numerical work would have to be carried out to establish their actual usefulness. It should be noted, in any case, that the effective-mass equations would be useful only if the defect or impurity potential were known (the theory cannot determine the potential self-consistently) and then only for the bound state(s) in the gap. Further discussion of the connections with EMT is given in Ref. 60.

E. Comparison with tight-binding models

So far, we have carried out Green's-function calculations of three different degrees of complexity: (1) using a semiempirical tight-binding Hamiltonian, reported in paper I,[28] (2) using a non-self-consistent vacancy pseudopotential taken to be the negative of a bulk atomic pseudopotential, and (3) using a fully self-consistent vacancy pseudopotential. (In the last two cases, the underlying band structure is also based on self-consistent pseudopotentials.) We have already seen in Sec. V B that the results of calculations (2) and (3) differ only in their details. We therefore turn now to a comparison with the tight-binding results of Ref. 28. In Fig. 10 we compare the tight-binding results for the changes in the densities of A_1 and T_2 states with the results of the fully self-consistent calculations described above. The two curves in each case are remarkably similar, confirming the usefulness of the tight-binding calculations. We may recall that in the tight-binding model[28] all Hamiltonian matrix elements are assumed to be identical to those of the unperturbed bulk, even those corresponding to the backbonds. The self-consistent results for the total change in the charge density [Fig. 4(c)] show that changes are restricted within the vacancy cavity and provide an explanation and justification for the validity of the tight-binding model.

The bound-state energy, however, obtained by the tight-binding calculations[28] is far too low (0.3 eV compared with the present 0.7 eV). We speculate that this unsatisfactory result is a consequence of the fact that the tight-binding conduction bands are not very accurate. As discussed in Ref. 28, the valence-band state-density changes are deter-

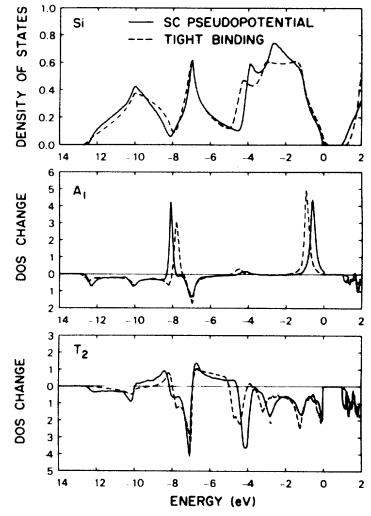

FIG. 10. Comparison of the tight-binding and self-consistent results for density of states (DOS) of perfect silicon and for the DOS changes of A_1 and T_2 symmetries. The curves are broadened by 0.2 eV and the reference energy is the top of the valence bands.

mined primarily by the valence bands, whereas the position of the level in the gap is determined by a subtle balance between valence- and conduction-band contributions. Note that this observation [Eq. (43) of Ref. 28] is not in contradiction with the fact that the bound-state wave function is composed mainly of valence-band Bloch functions (Fig. 9). The tight-binding result comes about because the bound-state wave function is expanded in terms of s and p LCAO's which give rise to both valence and conduction bands. Note that Eq. (43) of Ref. 28 is only a condition that determines the position of the bound-state energy level and contains no information about the composition of the bound-state wave function. On the other hand, this relation suggests that the bound-state energy level obtained by the tight-binding model is too low because the tight-binding conduction bands are too narrow compared with the true sp conduction bands, so that the integral on the right-hand side of Eq. (43) of Ref. 28 is larger than it should be.

VII. SUMMARY AND CONCLUSIONS

In paper I, we have shown that the Koster-Slater Green's-function technique becomes very efficient and accurate when the Wannier functions of the original formulation are replaced by an LCAO basis set. In that paper the method was applied to tight-binding Hamiltonians and its advantages over small- and large-cluster techniques were explicitly illustrated.

In this paper we have demonstrated the feasibility of the Green's-function method in conjunction with first-principles potentials and band structures and we have extended it to fully include self-consistency effects. The principal advantages of this method are the exploitation of (1) the short-range nature of the defect potential, (2) the translational symmetry of the host crystal, and (3) the analytic separation between the bulk and defect properties. Our first application of the method has been to the vacancy in Si. We have found one bound threefold degenerate state of T_2 symmetry at 0.7 eV, which is occupied by only two electrons. There are also several resonances and antiresonances within the valence bands which are associated with the vacancy. The wave functions associated with the bound state, the resonances, and the antiresonances are individually quite *extended*, while the vacancy potential is *short ranged* and extends only up to the nearest neighbors of the vacancy. The localization of the potential is caused by a detailed cancellation of the bound-state charge density by a corresponding charge redistribution within the valence bands.

ACKNOWLEDGMENTS

It is a pleasure to thank S. G. Louie and A. R. Williams for valuable discussions. We would also like to thank J. W. Cooley for advice concerning the use of fast-Fourier-transform algorithms. This work was supported in part by AFOSR under Contract No. F49620-77-C-0005.

APPENDIX A: SYMMETRIZATION AND ORTHOGONALIZATION PROCEDURE

The symmetrized combinations of LCAO's on each shell surrounding the vacancy can be generated by letting the projection operator

$$P^{(j)}_{\mu\mu} = \frac{l_j}{h} \sum_R \Gamma^{(j)}(R)^*_{\mu\mu} P_R \tag{A1}$$

for each representation j act on each of the basis functions on that shell. [R is a rotation belonging to the point group, P_R is the corresponding rotation operator, l_j is the dimension of the irreducible representation, and $\Gamma^{(j)}(R)$ is the matrix of the representation for that rotation.]

The above procedure has been automated and both the symmetrized shell orbitals and the OSO's have been generated on a computer. The result of applying $P^{(j)}_{\mu\mu}$ on an LCAO function is

$$P^{(j)}_{\mu\mu}\psi_\alpha(\vec{r}-\vec{\tau}) = \frac{l_j}{h} \sum_R \Gamma^{(j)}(R)_{\mu\mu}\psi_\beta(\vec{r}-\vec{\tau}'), \tag{A2}$$

where α denotes the transformation property of an LCAO (s, x, etc.), $\vec{\tau}$ is its position; $\beta = P_R^{-1}\alpha$ and $\tau' = P_R\tau$ denote the transformation property and the position of the rotated orbital, respectively. After discarding the linearly dependent functions, the partners to the symmetrized shell orbitals are formed with the help of the transfer operator $P^{(j)}_{\lambda\mu}$. Finally, the resulting basis functions are orthonormalized by the Schmidt process. The orthonormalized shell orbitals (OSO) can then be compactly expressed in terms of the LCAO's on the shell in question and the OSO's on the previous shells.

APPENDIX B

In this appendix we will solve the Poisson equation

$$\nabla^2 V = -8\pi\rho, \tag{B1}$$

subject to the boundary condition zero applied on the surface of a cube. Let us denote the cube side by a. Since V vanishes on the surface of the cube, V can be expanded in a three-dimensional sine series. If the charge density ρ is given on a cubic mesh of N points in each direction, the expansions

$$V(x_1,x_2,x_3)=\sum_{k_1=1}^{N}\sum_{k_2=1}^{N}\sum_{k_3=1}^{N} d(k_1,k_2,k_3)\sin\frac{\pi k_1 x_1}{Na} \times\sin\frac{\pi k_2 x_2}{Na}\sin\frac{\pi k_3 x_3}{Na} \quad \text{(B2)}$$

and

$$\rho(x_1,x_2,x_3)=\sum_{k_1=1}^{N}\sum_{k_2=1}^{N}\sum_{k_3=1}^{N} c(k_1,k_2,k_3)\sin\frac{\pi k_1 x_1}{Na} \times\sin\frac{\pi k_2 x_2}{Na}\sin\frac{\pi k_3 x_3}{Na} \quad \text{(B3)}$$

are related by

$$d(k_1,k_2,k_3)=-\frac{8N^2a^2}{\pi(k_1^2+k_2^2+k_3^2)}\,c(k_1,k_2,k_3)\,. \quad \text{(B4)}$$

The Fourier analysis to determine $c(k_1,k_2,k_3)$ and the Fourier synthesis to obtain $V(x_1,x_2,x_3)$ are carried out iteratively using the one-dimensional sine-transform algorithm.

APPENDIX C

In this appendix we discuss the similarities and differences between our implementation of the Green's-function formalism and that developed independently by Baraff and Schlüter (BS).[30] The two formulations have thus far produced virtually identical results for the unrelaxed vacancy in Si.

Our approach is based directly on operator equations which are general results of scattering theory. The basic operators are $G^0(E)$ and U. All quantities of physical interest are given as traces or determinants of operators that involve only $G^0(E)$ and U. (The trace and the determinant of an operator are invariant in the sense that they can be calculated by using any orthonormal complete set of functions to represent the operator in matrix form.) In our work, we chose a basis set which is physically transparent and, at the same time, practical for accurate self-consistent calculations.

BS's formulation, on the other hand, is based on expansions of wave functions and the Green's function in two different basis sets. Expressions for the quantities of interest are *derived* in terms of the expansion coefficients. The same expressions can, however, be obtained directly from the standard scattering-theoretic operator equations by representing the operators in the two basis sets. Differences arise in the final expressions of the two approaches largely because of the two different basis sets as opposed to our single set. In what follows we compare our choices with those of BS and identify advantages and disadvantages.

The central operator of our formulation is

$$Q_1(E)=1-G^0(E)U\,. \quad \text{(C1)}$$

For example, bound-state energies are given by the zeros of the determinant of $Q_1(E)$. In contrast, the corresponding BS result is equivalent to using the operator

$$Q_2(E)=U-UG^0(E)U\,. \quad \text{(C2)}$$

We observe immediately that

$$Q_2(E)=UQ_1(E)\,, \quad \text{(C3)}$$

so that, as long as $\det\|U\|\neq 0$, either Q_1 or Q_2 will in principle give identical results. In practice, however, differences can in fact arise.

One of the appealing properties of $Q_2(E)$ is that it satisfies a variational principle, in the sense that it yields energies that are accurate to second order in the wave function.[30] This property suggests that a desired degree of accuracy can be achieved with fewer basis orbitals. However, it is straightforward to see that, as long as one uses the same number of orbitals to represent U and G^0 in matrix form, the factorization (C3) follows. As a consequence, Q_1 and Q_2 will have precisely the same zeros even though one is variational and one is not. In order to take advantage of the variational nature of Q_2 one must use two different-size basis sets for U and G^0. The factorization (C3) is then not possible and Q_1 and Q_2 would yield different results. In particular, BS used a set of LCAO's at each atomic site, including the vacant site, for G^0 (the "inner" set), and a smaller set that did not include orbitals at the vacant site for U (the "outer" set). The variational principle was then in effect. In contrast, we included orbitals at the vacant site for both G^0 and U, so that the use of Q_1 and Q_2 would yield identical results. We believe that including orbitals at the vacant site for both U and G^0 is mandated by physical considerations, even though the size of the U matrix is increased. As we saw in Sec. IV, the role of the basis set is to represent U and the change in the charge density $\Delta\rho$, both of which are highly localized in the vacant atomic volume (Figs. 3 and 4). In fact, one can argue[53] that orbitals at the vacant site alone ought to be sufficient to expand $\Delta\rho$. We further believe that the omission of orbitals at the vacant site by BS has affected their results, albeit in a minor way. The effect is noticeable for states of A_1 symmetry which have nonzero amplitude at the origin. In particular, we find an A_1 resonance at $E_v-0.7$ eV, whereas BS find it at $E_v-1.1$ eV. We believe that the most likely source of this discrepancy is BS's omission of orbitals at the vacant site in their outer set, i.e., the set that is used to expand the wave functions and hence $\Delta\rho$ and U.

The use of Q_2 with two different-size basis sets has an additional disadvantage[61] that must be dealt with carefully: Note that $\det\|Q_2\|$ is in principle

proportional to $\det\|U\|$. Note also that the strength of the method lies in the fact that the matrix of U is nonzero in a small subset [cf. Eq. (26)]. Normally, one would like to increase this subset and check the convergence of the results. As convergence is reached, $\det\|U\|$ tends to zero and hence $\det\|Q_2\|$ tends to zero as well. Since Q_2 is inverted in the calculation of the change in the charge density, one must exercise caution in carrying out convergence tests (in the limit of full convergence, when $\det\|U\|=0$, special procedures must be used to ensure the inversion of Q_2 is effected in the subspace in which $\det\|U\|$ is nonzero). None of these complications arise when one works with Q_1.

Finally, one last technical difference between our approach and that of BS is the following: As stated earlier, we work with operator equations and represent each operator in matrix form so that the operator equations become matrix equations.[62] Thus, we in fact calculate the Green's-function matrix in the chosen basis set. BS, on the other hand, as they point out, do not calculate matrix elements of the Green's function in any particular set of states. Instead, they *expand* the Green's function in a set of states as follows:

$$G^0(E,r,r')=\sum_{mm'}\Phi_m^*(r)\tilde{G}^0_{mm'}(E)\Phi_{m'}(r')\,. \tag{C4}$$

The $\tilde{G}^0_{mm'}(E)$ are expansion coefficients which are evaluated by expanding the Bloch functions in terms of the orbitals $\Phi_m(r)$. [Note that the Bloch functions are calculated in a plane-wave basis and then *fitted* to an expansion in terms of the $\Phi_m(r)$.] In fact, one can show that the $\tilde{G}^0_{mm'}(E)$ would be Green's-function matrix elements if the Φ_m were orthogonal. If not, and S_{ml} is the overlap matrix of the Φ_m, we have

$$\tilde{G}^0_{mm'}(E)=\sum_{ll'}S^{-1}_{ml}G^0_{ll'}(E)S^{-1}_{l'm'}\,, \tag{C5}$$

where the $G^0_{ll'}(E)$ are true Green's-function matrix elements and can be evaluated in a straightforward way as described in Sec. V.

[1] W. Kohn, *Solid State Physics*, edited by F. Seitz and D. Turnbull (Academic, New York, 1957), Vol. 5.

[2] S. T. Pantelides and C. T. Sah, Phys. Rev. B 10, 621 (1974); 10, 638 (1974); J. Bernholc and S. T. Pantelides, *ibid.* 15, 4935 (1977).

[3] A. Baldereschi and N. Lipari, in *Proceedings of the Thirteenth International Conference on the Physics of Semiconductors*, edited by F. G. Fumi (Marves, Rome, 1977), p. 595.

[4] M. Altarelli, W. Y. Hsu, and R. A. Sabatini, J. Phys. C 10, L605 (1977); S. T. Pantelides, Solid State Commun. 30, 65 (1979).

[5] For a recent review, see S. T. Pantelides, Rev. Mod. Phys. 50, 797 (1978).

[6] C. A. Coulson and M. J. Kearsley, Proc. R. Soc. London A 241, 433 (1957).

[7] G. D. Watkins and R. P. Messmer, Phys. Rev. Lett. 25, 656 (1970); Phys. Rev. B 7, 2568 (1973).

[8] C. Weigel, D. Peak, J. C. Corbett, G. D. Watkins, and R. P. Messmer, Phys. Rev. B 8, 2906 (1973).

[9] F. P. Larkins, J. Phys. C 4, 3065 (1971); 4, 3077 (1971).

[10] B. Cartling, B. Roos, and U. Wahlgren, Chem Phys. Lett. 21, 380 (1973); B. Cartling, J. Phys. C 8, 3171 (1975); 8, 3183 (1975).

[11] G. D. Watkins and R. P. Messmer, in *Computational Methods for Large Molecules and Localized States in Solids*, edited by F. Herman, A. D. McLean, and R. K. Nesbet (Plenum, New York, 1972), p. 133.

[12] S. G. Louie, M. Schlüter, J. R. Chelikowsky, and M. L. Cohen, Phys. Rev. B 13, 1654, (1976).

[13] J. D. Joannopoulos and E. J. Mele, Solid State Commun. 20, 729 (1976).

[14] E. Kauffer, P. Pêcheur, and M. Gerl, J. Phys. C 9, 2913 (1976); Phys. Rev. B 15, 4107 (1977).

[15] See M. Lifshitz, Nuovo Cimento Suppl. 3, 716 (1956) for the original Russian references.

[16] M. Lax, Phys. Rev. 94, 1391 (1954).

[17] G. F. Koster and J. C. Slater, Phys. Rev. 94, 1392 (1954); 95, 1165 (1954).

[18] J. Callaway, J. Math. Phys. 5, 783 (1964); Phys. Rev. 154, 515 (1967).

[19] J. Callaway and A. Hughes, Phys. Rev. 156, 860 (1967); 164, 1043 (1967).

[20] N. J. Parada, Phys. Rev. B 3, 2042 (1971).

[21] J. Callaway, Phys. Rev. B 3, 2556 (1971).

[22] S. P. Singhal, Phys. Rev. B 4, 2497 (1971); 5, 4203 (1972).

[23] M. Lannoo and P. Lenglart, J. Phys. Chem. Solids 30, 2409 (1969).

[24] D. Rouhani, M. Lannoo, and P. Lenglart, in *Radiation Effects in Semiconductors*, edited by J. W. Corbett and G. D. Watkins (Gordon and Breach, London, 1971), p. 15.

[25] F. Bassani, G. Iadonisi, and B. Preziosi, Phys. Rev. 186, 735 (1969).

[26] M. Jaros and S. Brand, Phys. Rev. B 14, 4494 (1976).

[27] U. Lindefelt, J. Phys. C 11, 3651 (1978).

[28] J. Bernholc and S. T. Pantelides, Phys. Rev. B 18, 1780 (1978).

[29] J. Bernholc, N. O. Lipari, and S. T. Pantelides, Phys. Rev. Lett. 41, 895 (1978).

[30] G. A. Baraff and M. Schlüter, Phys. Rev. Lett. 41, 892 (1978); Phys. Rev. B 19, 4965 (1979).

[31] F. Garcia-Moliner, in *Theory of Imperfect Crystalline Solids: Trieste Lectures 1970* (IAEA, Vienna, 1971).

[32] R. G. Newton, *Scattering Theory of Waves and Particles* (McGraw-Hill, New York, 1966).

[33] B. A. Lippman and J. Schwinger, Phys. Rev. 79, 469 (1963).

[34]J. F. Janak, A. R. Williams, and V. L. Moruzzi, Phys. Rev. B 11, 1522 (1975).
[35]In general, the defect potential extends further than the change in the charge density due to multipole fields.
[36]A systematic and practical way to construct Wannier functions for the valence bands has recently been reported by E. O. Kane and A. B. Kane [Phys. Rev. B 17, 2691 (1978)].
[37]E. O. Kane, Phys. Rev. B 13, 3478 (1976); D. J. Chadi, *ibid.* 16, 3572 (1977).
[38]R. C. Chaney, C. C. Lin, and E. E. Lafon, Phys. Rev. B 3, 459 (1971); J. Langlinais and J. Callaway, *ibid.* 5, 124 (1972); S. Ciraci and I. P. Batra, *ibid.* 15, 3254 (1977).
[39]R. Zeller and P. H. Dederichs, Phys. Rev. Lett. 42, 1713 (1979).
[40]P. O. Löwdin, J. Chem. Phys. 18, 365 (1950).
[41]J. A. Appelbaum and D. R. Hamann, Phys. Rev. B 8, 1977 (1973).
[42]M. Schlüter, J. R. Chelikowsky, S. G. Louie, and M. L. Cohen, Phys. Rev. B 12, 4200(1975).
[43]S. G. Louie (private communication).
[44]P. Hohenberg and W. Kohn, Phys. Rev. 136, B864 (1964); W. Kohn and L. J. Sham, *ibid.* 140, *A*1133 (1965).
[45]D. J. Chadi and M. L. Cohen, Phys. Rev. B 8, 5747 (1973).
[46]S. G. Louie, K. M. Ho, and M. L. Cohen, Phys. Rev. B 19, 1774 (1979).
[47]G. Gilat and L. J. Raubenheimer, Phys. Rev. 144, 390 (1966).
[48]J. F. Janak, in *Computational Methods in Band Theory*, edited by P. M. Marcus, J. F. Janak, and A. R. Williams (Plenum, New York, 1971), p. 323.
[49]J. W. Cooley and J. Bernholc (unpublished).
[50]Alternatively, the basis functions and the potential could be fitted to functions for which analytical expressions for integrals entering the calculation are available. [e.g., Gaussians (Refs. 38 and 30)]. Also, the Poisson equation can be solved analytically in the Gaussian basis. The exchange-correlation potential, however, must be treated numerically and thereafter *fitted* to Gaussians. The fitting of a function in three dimensions is a rather cumbersome procedure, because of the great accuracy required. Furthermore, additional difficulties arise if the problem at hand requires a large number of Gaussians. In this case, the charge density being a bilinear function of the basis contains an impractical number of terms and requires refitting with a new set of Gaussians which places an even stronger requirement on the accuracy of the fit. A direct numerical integration scheme is much easier to program, to automate, and to test for convergence. We also believe that numerical sampling of 1000–2000 points per atom represents the potential much more accurately than the fitted function, thereby making a numerical evaluation of the matrix elements not only a simpler, but also a more accurate procedure.
[51]J. W. Cooley and J. W. Tukey, Math Comput. 19, 297 (1965).
[52]The efficiency of this procedure stems from the fact that ∇^2 is diagonal in the plane-wave representation.
[53]This fact and the results displayed in Figs. 2 and 5, which show that the total change in the charge density and vacancy potential have a predominantly single-center character, have led U. Lindefelt and S. T. Pantelides [Solid State Commun. 30, 631 (1979)] to suggest a single-center basis set consisting of harmonic-oscillator eigenfunctions as an alternative to the multicenter LCAO basis set used in the present work.
[54]For a review of tight-binding Hamiltonians, see S. T. Pantelides and J. Pollmann, J. Vac. Sci. Technol. 16, 1349 (1979).
[55]See also Ref. 37.
[56]S. T. Pantelides and W. A. Harrison, Phys. Rev. B 11, 3006 (1975).
[57]J. Bernholc, N. O. Lipari, and S. T. Pantelides (unpublished).
[58]S. T. Pantelides, Solid State Commun. 30, 65 (1979).
[59]The importance of the umklapp terms was first pointed out by K. Shindo and H. Nara, J. Phys. Soc. Jpn. 40, 1640 (1976); M. Altarelli, W. Y. Hsu, and R. A. Sabatini, J. Phys. C 10, L605 (1977).
[60]S. T. Pantelides, N. O. Lipari, and J. Bernholc, Solid State Commun. 33, 1045 (1980).
[61]See also Lindefelt and Pantelides, Ref. 53.
[62]We found it convenient to orthogonalize our basis orbitals ahead of time. One could use nonorthogonal orbitals, in which case the inverse of the overlap matrix is inserted between the matrices of operators that are to be multiplied.