

Quasi bands in Green's-function defect models

U. Lindelfelt and Alex Zunger

Solar Energy Research Institute, Golden, Colorado 80401

(Received 6 May 1981)

A basic difficulty in applying the Green's-function formalism to deep defects in solids is discussed. A cure is provided. It is shown that the conventional Green's-function formalism as applied to point-defect problems is derivable by requiring a dual representation of the defect wave functions both in terms of an expression in "local" basis functions and, *independently*, in terms of the eigenfunctions of the host-crystal Hamiltonian. It is then shown that this dual representation leads to a fundamental limitation of the method. In contrast to what may have been thought, the defect Green's-function (DGF) method does not become increasingly effective as the defect-induced potential perturbation becomes more localized in coordinate space. In fact, a consequence of this dual representation is that for impurities which are chemically mismatched to the host-crystal atoms a computationally intractable and physically undesirable enormous number of host-crystal eigenfunctions ($10^2 - 10^4$) is needed to obtain even modestly accurate defect wave functions, energies, and chemical trends. A new, formally exact approach (the "quasi-band-structure representation") is presented. This approach overcomes the difficulties underlying the dual representation in a simple way. It is based on redefining the zero-order basis set and expanding the defect wave functions in terms of such "quasi band wave functions" rather than by pure host wave functions. The former diagonalize the (finite) matrix of the host-crystal Hamiltonian and include aspects of both host and defect orbitals but need not form eigenstates to the host-crystal Hamiltonian operator. We illustrate this exact method for two analytically solvable models: a parabolic defect potential as well as a transition-atom impurity in a silicon free-electron host crystal. A comparison with the results of the conventional DGF calculation is given. Finally, the method is illustrated for a fully self-consistent calculation for substitutional Cu in silicon using nonlocal pseudopotentials.

I. INTRODUCTION

The understanding of the nature of deep impurity centers in semiconductors has long been recognized as central to the design of efficient solar cells, light-emitting and Schottky diodes, and a multitude of other semiconductor devices.¹⁻⁵ However, the chemical identity of such centers has often been impossible to determine experimentally because deep impurity centers, even when present in minute concentrations (as small as one part in 10^{10}), are capable of profoundly modifying the transport and optical properties of a semiconductor. One then frequently encounters the situation where detailed data are available on properties such as spin, g -value, thermal and optical activation energies, as well as the characteristics of the radiative decay of "ghost" defects that remain chemically and structurally unidentified. Under these conditions it is difficult, if not impossible, to design effective

chemical purifications and other defect sterilization and passivation techniques. A comprehensive theory of deep defect states seems therefore to be acutely needed. Such an approach could link the theoretically predicted properties of *specified* impurities with the observed characteristics of hitherto "ghost" states.

Among the various theoretical techniques capable of predicting deep impurity levels, the defect Green's-function (DGF) method⁶⁻⁹ often has been described as the most effective and precise (e.g., see discussions in Refs. 10-12). Unlike the crystal field,^{13,14} ligand field,^{15,16} defect molecule,¹⁷ and other cluster techniques,¹⁸⁻²¹ or the repeated-cell method,^{22,23} the DGF approach allows one to treat directly only the *defect-induced perturbations* $\Delta V(\vec{r})$ and $\Delta\rho(\vec{r})$ in the system's potential and charge density, respectively, rather than having to describe at the same time the details of the extended parts of the host (H) plus defect potential

$V_H(\vec{r}) + \Delta V(\vec{r})$.

This property has led in the past to efficient and precise descriptions of simple s,p -electron deep defect states^{8–11} in which the range of $\Delta V(\vec{r})$ is typically one or more host-crystal bond distances. However, the fact that the Green's-function approach permits one to treat directly only the physical phenomena within the range of $\Delta V(\vec{r})$ has spurred the hope that the same approach will be equally effective (if not more so) for treating localized perturbations in which $\Delta V(\vec{r})$ has dimensions typical of an atom or a core. Such perturbations frequently occur for core holes, first-row and transition-atom impurities in semiconductors, etc. It was often thought that in fact the same approach can be used to deduce chemical trends in defect energies, over a broad range of chemical coordinates characterizing $\Delta V(\vec{r})$ (Ref. 24) (e.g. electronegativity differences, impurity-host size mismatch). We show in this paper that in fact this is not the case. For sufficiently localized defect potential perturbations $\Delta V(\vec{r})$ the conventional DGF approach becomes totally impractical. If used in its present form for such localized states the results are grossly inaccurate. We will identify and discuss this limitation. A simple and efficient cure is then proposed.

The program of this paper is to show that (i) for sufficiently localized defect potential perturbations $\Delta V(\vec{r})$ the conventional DGF approach becomes totally impractical due to the need of incorporating coupling to many ($\geq 10^2 - 10^4$) host bands. (ii) Such critically localized potential perturbations are often encountered in many deep defects such as first-row as well as transition-atom impurities in conventional semiconductors, perturbations associated with extensive lattice relaxations, and core excitons in solids. In general, this difficulty with the DGF method becomes acute when the impurity atoms or structural defects are chemically and physically sufficiently different from the host crystal. (iii) This limitation of the DGF method stems fundamentally from the requirement, tacitly assumed in all of the conventional formal derivations^{6–9} and practical applications,^{10,11} that the wave function of an arbitrary defect be effectively spanned by the (frequently chemically and physically unrelated) eigenfunctions of the host-crystal Hamiltonian operator $-1/2\nabla^2 + V_H(\vec{r})$. It may not have been previously recognized that even if a defect wave function is expanded *exactly* in terms of some localized basis functions, but an additional representation in terms of host wave functions is

used,^{6–11} the conventional DGF formalism requires that the latter expansion be *independently* complete. We show, for example, that in order to determine the energy of a deep level such as that arising from a transition-atom impurity in silicon with useful precision (e.g., a tenth or two of a host energy band gap), a computationally intractable and physically undesirable large number of host eigenfunctions (hundreds to ten of thousands of host bands) is required by the DGF method. (iv) We show how this fundamental limitation of the conventional DGF method can be eliminated by redefining the zero-order (unperturbed) basis set (host wave functions) to consist of “quasi band wave functions”, which incorporate from the outset aspects of both the host and the impurity. Furthermore, such basis functions need not be eigenstates of the host-crystal Hamiltonian operator. With this representation, only a few (often as little as 5–10) quasi bands are needed to obtain the exact defect energies and wave functions, even for an exceedingly localized potential perturbation $\Delta V(\vec{r})$.

We illustrate and compare the method with the conventional DGF approach for two analytically solvable models: (1) a parabolic defect potential perturbation $\Delta V(\vec{r})$ in a silicon free-electron host, and (2) a $3d$ transition element in a silicon free-electron host. These two examples illustrate the capability of the method to treat accurately both the wave functions and the energies of arbitrarily deep defect centers. As a final illustration, we comment on the results obtained with this method for a realistic case, a substitutional Cu-atom impurity (represented by a self-consistently screened nonlocal *ab initio* pseudopotential) in a silicon crystal.

In Sec. II, we describe two different views taken up in previous work to describe the defect wave function—expanding it in terms of some localized basis $\{g_a(\vec{r})\}$, as done in cluster-type methods,^{13–21} or expanding it in terms of the host Hamiltonian eigenfunctions $\{\phi_j^0(\vec{k}, \vec{r})\}$, as done in perturbative approaches.^{9,12,25} We then show that the conventional DGF approach is derivable by requiring that both descriptions of the defect wave functions be simultaneously and independently valid (the “dual representation”). Section III discusses the consequences of the dual representation. We illustrate the underlying difficulties caused by the dual representation using a simple analytically solvable model. This illustrates the fact that a deep potential perturbation $\Delta V(\vec{r})$ requires an intractable number of host eigenfunctions in order to obtain

the defect energies and wave functions with a useful accuracy. Section IV introduces the quasi band representation that solves this difficulty in a simple and effective manner. Applications of these methods are presented in Sec. V.

II. TWO VIEWS ON DEFECT WAVE FUNCTIONS

The basic problem in the theory of the electronic structure of defects in solids is to solve the single-particle equation for the system containing a defect

$$[-\frac{1}{2}\nabla^2 + V_H(\vec{r}) + \Delta V(\vec{r})]\psi_i(\vec{r}) = \epsilon_i \psi_i(\vec{r}) \quad (1)$$

[where $V_H(\vec{r})$ and $\Delta V(\vec{r})$ are, respectively, the host (H) potential and the defect-induced potential perturbation], given the solutions ψ_i of the corresponding problem for the unperturbed host:

$$[-\frac{1}{2}\nabla^2 + V_H(\vec{r})]\phi_j^0(\vec{k}, \vec{r}) = \epsilon_j^0(\vec{k})\phi_j^0(\vec{k}, \vec{r}). \quad (2)$$

Here, $V_H(\vec{r})$ is the periodic host potential, $\{\epsilon_j^0(\vec{k})\}$ is the associated one-electron band structure, and $\phi_j^0(\vec{k}, \vec{r})$ are the Bloch eigenfunctions.

Previous theoretical models on the electronic structure of localized defect states, not using a DGF approach, have often assumed one of the two

$$[\epsilon_j^0(\vec{k}') - \epsilon_i]A_{ij}(\vec{k}') + \sum_j \sum_{\vec{k}} A_{ij}(\vec{k}) \langle \phi_j^0(\vec{k}', \vec{r}) | \Delta V(\vec{r}) | \phi_j^0(\vec{k}, \vec{r}) \rangle = 0, \quad (4)$$

given the host state $\{\epsilon_j^0(\vec{k})\}$ and $\{\phi_j^0(\vec{k}, \vec{r})\}$ over the Brillouin zone (BZ). If the perturbation $\Delta V(\vec{r})$ is delocalized to the extent that only a single host band is effectively coupled to $|i\rangle$, one can reduce the problem to the effective-mass approximation (for impurities or excitons).¹² Generally, however, the perturbation $\Delta V(\vec{r})$ may couple several host bands. The advantage of the representation of defect wave functions by host-Bloch wave functions is that the structure of Eqs. (3) and (4) permits the analysis of the evolution of defect states from the host states as $\Delta V(\vec{r})$ is turned on. At the same time, this representation relates naturally the defect energy levels ϵ_i to host-band edges $\epsilon_j^0(\vec{k})$. This approach requires, however, that the wave functions of an arbitrary defect $\psi_i(\vec{r})$ be effectively spanned by a (hopefully) manageable number M_i of (in general physically and chemically unrelated) host wave functions $\{\phi_j^0(\vec{k}, \vec{r})\}$. This will frequently lead to

fundamentally different expressions of the defect wave function $\psi_i(\vec{r})$: expansion in host-Bloch wave functions as used in perturbative methods,^{9,12,25} or expansion in some localized functions as used in cluster¹⁷⁻²¹ and ligand-field^{15,16} methods. We will first describe these approaches and then show that the standard Green's-function approach to defects as used in all recent applications is derivable by requiring that both descriptions of the defect wave functions be *simultaneously and independently* satisfied.

A. Expansion in host Bloch states

While the formation of a defect with the attendant loss of translational invariance prohibits the Brillouin-zone wave vector \vec{k} from being a good quantum number, many approaches to the problem seek a representation of the defect wave function $\psi_i(\vec{r})$ in terms of combinations of M_i perfect-host-band wave functions:

$$\psi_i(\vec{r}) = \sum_j \sum_{\vec{k}} A_{ij}(\vec{k}) \phi_j^0(\vec{k}, \vec{r}). \quad (3)$$

The problem of solving Eq. (1) is now transformed into one of finding the projections $\{A_{ij}(\vec{k})\}$ and defect energies $\{\epsilon_i\}$ in

the undesirable situation where the physics of a few defect levels, energetically closely spaced, need to be described by an intractably large number M_i of host bands spread over a wide range of one-electron energies. Various perturbative treatments of defects^{9,25-27} are rooted in the representations (3) and (4).

B. Expansion in local orbitals

An alternative view of $\psi_i(\vec{r})$ is possible if one expands it in a set $\{g_a(\vec{r})\}$ of N_i basis functions anchored to specific atomic sites (i.e., loosely speaking, "local basis functions"):

$$\psi_i(\vec{r}) = \sum_a^{N_i} C_{ia} g_a(\vec{r}). \quad (5)$$

This leads to a secular equation of the form

$$\sum_b^{N_i} C_{ib} [\langle g_a(\vec{r}) | -\frac{1}{2}\nabla^2 + V_H(\vec{r}) + \Delta V(\vec{r}) | g_b(\vec{r}) \rangle - \epsilon_i \langle g_a(\vec{r}) | g_b(\vec{r}) \rangle] = 0. \quad (6)$$

The crystal-field^{13,14} approach is based on Eq. (5) where $\{g_a(\vec{r})\}$ are limited to the impurity-atom orbitals, while the ligand-field,^{15,16} defect-molecule,¹⁷ or cluster^{18–21,28,29} approaches allow the local orbitals $g_a(\vec{r})$ to be situated also on a number of atoms (ligands) surrounding the defect site. While the description of $\psi_i(\vec{r})$ in terms of local orbitals is often intuitively appealing, such a representation describes poorly the extended host states with which the defect interacts.

C. The Green's-function way

In the standard DGF approach,^{6–11} one is tacitly requiring that both views expressed in Eqs. (3) and (5) be *simultaneously* fulfilled. In this approach it not only is required that if the expansion (3) is used enough bands M_i should be included to make this expansion valid for state i , but also that the expansion in Eq. (5) in terms of N_i localized basis functions $\{g_a(\vec{r})\}$ should be *independently* valid in the subspace of the perturbation. That is, if $\Theta(\vec{r}-\vec{R}_c)$ denotes a step function which equals unity in that part of space ($r \leq R_c$) where the defect potential $\Delta V(\vec{r})$ is nonzero and is zero (or decays to zero) otherwise, it is required that both Eqs. (3) and

$$\Theta(\vec{r}-\vec{R}_c)\psi_i(\vec{r}) = \sum_a^{N_i} C_{ia} g_a(\vec{r}) \quad (7)$$

be valid.

This is shown in Appendix A, where it is demonstrated that the basic Green's-function matrix equations for calculating one-particle energies and wave functions (e.g., those used in Refs. 6–11) are derivable from Eq. (4) under the assumption that Eqs. (3) and (7) are simultaneously satisfied. It will be illustrated in Sec. V that not only are conditions (3) and (7) sufficient to derive the conventional DGF formalism, but that also if Eq. (3) is assumed (as is the case virtually in all applications, e.g., Refs. 6–11, 24, 27, 30–33), condition (7) is *necessary*.

According to Appendix A, the one-particle energies ϵ_i and wave-function expansion coefficients C_{ia} for state i are determined by the set of simultaneous equations

$$\sum_a \left[\delta_{a'a} - \sum_{a''} \tilde{G}^0(\epsilon_i)_{a'a''} \langle g_{a''} | \Delta V | g_a \rangle \right] C_{ia} = 0, \quad (8)$$

where the nonorthogonal representation of the Green's-function matrix is

$$\tilde{G}^0(\epsilon)_{aa'} = [S^{-1}G^0(\epsilon)S^{-1}]_{aa'} \quad (9)$$

and the basis set overlap matrix is denoted as $S_{ab} = \langle g_a | g_b \rangle$. Here, the host-crystal Green's-function matrix is given in its standard form as

$$G^0(\epsilon)_{ab} = \sum_{j=1}^{M_i} \sum_{\vec{k}}^{\text{BZ}} \frac{\langle g_a | \phi_j^0(\vec{k}, \vec{r}) \rangle \langle \phi_j^0(\vec{k}, \vec{r}) | g_b \rangle}{\epsilon - \epsilon_j^0(\vec{k})}. \quad (10)$$

The overlap matrix S has been introduced for a general nonorthogonal basis. The notable advantage of the dual representations is that it leads to Eqs. (8)–(10) in which only the potential perturbation $\Delta V(\vec{r})$, rather than the full potential $V_H(\vec{r}) + \Delta V(\vec{r})$, needs to be treated. The standard Green's-function representation in Eqs. (8)–(10) forms the basis for a number of recent approaches to the defect problem; e.g., the work by Callaway and Hughes (Ref. 8) ($\{g_a\}$ represented as Wannier functions); Jaros and Brand²⁷ and Lindefelt³⁰ ($\{g_a\}$ represented as products of exponentials and Laguerre polynomials,^{30(a)} or harmonic oscillator functions^{30(b)}); Baraff and Schlüter¹⁰ ($\{g_a\}$ represented by a number of individual Gaussians on atomic shells next to the defect but not on the defect site); and Bernholc, Lipari, and Pantelides¹¹ ($\{g_a\}$ represented by a set of orthogonalized local orbitals on the defect site and few atomic shells around it). In all of these calculations, various numerical computational schemes have been devised to solve the underlying Eqs. (8)–(10) for defects whose associated perturbation $\Delta V(\vec{r})$ extends over a few host bond distances (i.e., vacancies, nontransition-atom impurities, and impurities whose core structure and size is similar to that of the host atoms). The consequences of the underlying dual representation of the defect wave functions cannot be fully appreciated for such moderately “weak” perturbations. We discuss the general implications of this requirement in the next

section and show that it leads to fundamental limitations when the defects perturbation approaches an atomic dimension.

III. CONSEQUENCES OF THE DUAL REPRESENTATION

The great strength of the Green's-function method in combining the two different views on the defect wave functions expressed in Eqs. (3) and (7) is also its fundamental weakness: One needs to make the two descriptions of $\psi_i(\vec{r})$ equally and simultaneously complete. Even if one has an exact description of a defect wave function $\psi_i(\vec{r})$ in terms of a localized basis $\{g_a(\vec{r})\}$ in Eq. (7), one still needs to describe $\psi_i(\vec{r})$ accurately via Eq. (3) by the extended periodic host wave functions, and vice versa. If the impurity atom is chemically sufficiently different from the host atoms, or if significant defect-induced lattice relaxations occur, an intractable number of host bands M_i may be required.

This feature constitutes a fundamental difference between linear direct-matrix diagonalization approaches to electronic structure [e.g., linear combination of atomic orbitals (LCAO)] and Green's-function approaches. Whereas in LCAO methods one is free to implement one's knowledge, intuition, or hypothesis on the electronic structure of the system at hand (e.g., covalent, metallic or ionic bonding, lone-pair orbitals, etc.) by directly including suitably chosen local orbitals in the wave-function's expansion [Eq. (5)], in Green's-function methods one needs to independently cope with a nonintuitive expansion in terms of host wave functions. A partial resolution of this difficulty is possible if one simply increases the number of host wave functions in the expansion (3). This often leads to intractably complicated calculations and physically intransparent results.

We will first give a simple illustration of the consequence of the requirement of dual representation using an example that can be solved analytically in closed form. We will then discuss the question of the number of bands M_i required from the structure of Eq. (4) and give a simple quantitative numerical example. We will present our solution to this problem in Sec. IV.

Consider a diamond-type host crystal with the lattice constant of bulk silicon that is described in the free-electron model, i.e., $V_H(\vec{r})=0$ and $\phi_j^0(\vec{k}, \vec{r})=1/\sqrt{\Omega} e^{i(\vec{k}+\vec{G})\cdot\vec{r}}$ in Eq. (2). Here, \vec{G} is the BZ reciprocal-lattice vector and Ω is the unit-cell volume. We place in the crystal a localized

defect perturbation potential $\Delta V(\vec{r})$ in the form of a parabolic (harmonic oscillator) potential with a spring constant K : $\Delta V(\vec{r})=K(\vec{r}-\vec{r}_0)^2$ at a lattice site $\vec{R}=\vec{r}_0$. The exact energy levels ϵ_i ($i=nlm$) and wave functions $\psi_i(\vec{r})$ of this point defect in the crystal are given by

$$\epsilon_{nlm} = \sqrt{K} (n + 3/2) \quad (11a)$$

and

$$\psi_{nlm}(r) = N_{nl} (Kr^2)^{l/2} e^{-Kr^2/2} \times L_{n-l-1}^{l+1/2}(Kr^2) Y_{lm}(\Theta, \phi), \quad (11b)$$

where N_{nl} is a normalization factor, $L_p^q(Kr^2)$ is the Laguerre polynomial, and $Y_{lm}(\Theta, \phi)$ are the spherical harmonics. We now attempt [Eq. (11a)] to find the exact energy solutions to the defect problem using the standard Green's-function method in Eqs. (8)–(10).

We will demonstrate that if one uses the standard expansion (3), then even if the defect wave function $\psi_i(\vec{r})$ is given exactly in terms of some localized orbitals, one still needs to express it exactly also in terms of host eigenfunctions to recover the correct energies in Eq. (11a). To show this, we use the *exact solution* given in Eq. (11b): $g_a(\vec{r}) = \psi_{nlm}(\vec{r})$ for the local-orbital expansion of Eq. (7). [In fact, the first 15 solutions of Eq. (11b) for $l=0, 1, 2,$ and 3 are used as a basis (cf. Sec. V); all basis functions are defect-site centered]. All integrals in Eqs. (8)–(10): $\langle g_a | \Delta V | g_b \rangle$ and $\langle g_a | \phi_j^0(\vec{k}, \vec{r}) \rangle$, are calculated by an accurate radial numerical integration. The upper part of Table I compares the exact energy eigenvalues [Eq. (11a)] for the lowest defect levels of symmetry $a_1, t_2,$ and e_2 (angular momentum of $l=0, 1,$ and $2,$ respectively), as well as the first excited a_1^* state, with those obtained by the conventional Green's-function approach when the first 10, 20, 32, and 41 host bands are used. These host bands span a very large energy range of 1.66, 2.99, 3.66, and 4.49 Ry, respectively. Yet, the errors involved in the defect energy levels ϵ_i can be seen to be enormously large on any relevant scale. Furthermore, the errors are seen to increase rapidly with angular momentum l and excitation.

Figure 1 shows the exact defect wave functions of Eq. (11b) compared with those obtained with the standard DGF technique, using in the expansion of Eq. (3), 10 and 41 host bands, respectively. It is seen that in the DGF method, an expansion of $\psi_i(\vec{r})$ in terms of as many as 41 host-band wave functions still leads to substantial errors for localized perturbations. Figure 2 shows the spectral

TABLE I: The calculated defect energy levels of a parabolic perturbation potential $\Delta V(\vec{r}) = K(\vec{r} - \vec{r}_0)^2$ in a silicon free-electron host, using the conventional defect Green's-function method and the present quasiband Green's-function method. The average number of host-band wave functions included for each \vec{k} point is M_b ; six quasi bands are used corresponding to the three lowest defect levels $a_1 + t_2 + e_2$. The potential perturbation is characterized by $K = 1 \text{ Ry/a.u.}^2$ and $|\vec{r} - \vec{r}_0| \leq 4 \text{ a.u.}$; energies are given with respect to the zero of the potential well ($1 \text{ Ry} = 13.605 \text{ eV}$). Notice the extremely slow convergence of the values obtained with the conventional DGF method relative to the far better convergence obtained with the present quasi band DGF method.

Number of bands	$n = 0, l = 0$	$n = 1, l = 1$	$n = 2, l = 2$	$n = 2, l = 0$
	a_1 (eV)	t_2 (eV)	e_2 (eV)	a_1^* (eV)
Conventional DGF method				
$M_b = 10$	69.413	130.553	186.946	200.987
$M_b = 20$	52.175	96.759	146.240	163.382
$M_b = 32$	46.257	83.507	127.139	138.349
$M_b = 41$	43.835	77.440	115.180	124.091
Exact results	40.815	68.025	95.235	95.235
Quasi band DGF method				
$M_b = 1$	40.815	68.025	95.235	184.756
$M_b = 10$	40.815	68.025	95.235	116.894
$M_b = 20$	40.815	68.025	95.235	103.765
$M_b = 32$	40.815	68.025	95.235	99.058

decomposition of the defect wave function $\psi_i(\vec{r})$ for the $|i\rangle = e$ state in terms of host wave functions $\phi_j^0(\vec{k}, \vec{r})$ for four \vec{k}_p values. Plotted are the absolute values $|A_{ij}(\vec{k}_p)|$ of the expansion coefficients in Eq. (3), normalized such that

$\sum_{j=1}^{M_i} \sum_{\vec{k}_p=1}^4 |A_{ij}(\vec{k}_p)|^2 d_p = 1$, where d_p is the number of wave vectors in the star of \vec{k}_p . It is seen that a deep defect wave function such as $\psi_e(\vec{r})$ draws its intensity from a large number of bands (their energy spread is indicated on the top abscissa); no strong attenuation of the contribution of high-energy bands is observed.

Clearly, the conventional Green's-function approach with its underlying dual description of the impurity wave function fails in reproducing effectively these correct defect states even if $\psi_i(\vec{r})$ is inputted *exactly* in terms of localized orbitals [Eq. (7)]. The lower part of Table I displays a far better convergence achieved with our "quasi band" reformulation of the Green's-function method (cf. Sec. IV): only a small number of "quasi band wave functions" (i.e., wave functions that diagonalize the host Hamiltonian matrix and include localized components) are needed.

Our second example involves an atomic impurity in a crystal. The question of how many host bands are required for an adequate description of impurity levels obviously cannot be answered in general. However, one can obtain a limit for this number for the physically interesting case of very localized impurity states, such as those produced by a transition atom in a semiconductor. When the orbital binding energy of an impurity atomic orbital is larger than the binding energy of the states at the bottom of the host's valence band (e.g., -18 eV for bulk silicon), one often obtains "hyperdeep" impurity levels below the valence-band minimum.^{28,29} These levels largely retain the characteristics of the free-atom wave function $R_{nl}(r)$, renormalized into a host Wigner-Seitz sphere (e.g., Zn 3d in silicon,²⁹ nitrogen in diamond, oxygen in GaP). While the precise location of such energy levels is rarely of direct technological interest (since they are fully occupied and sufficiently tightly bound so as not to affect the mobility of free carriers in the system), its characteristics may determine very sensitively the nature³⁴ (i.e., energy and spatial extent) of the important *gap* defect states (which need to be orthogonal to it). We will demonstrate in Sec. V

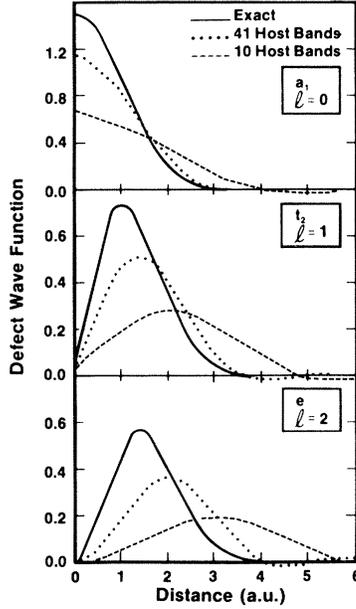


FIG. 1. Defect wave functions for the three lowest levels a_1 , t_2 , and e_2 of a parabolic defect-potential perturbation in a silicon-like free-electron host obtained in the standard DGF approach using $M_b = 10$ (dashed line) and $M_b = 41$ (dotted line) host-band wave functions per \vec{k} point. The exact solutions are given by Eq. (11b) and represented by the full line. The localized basis $\{g_a(\vec{r})\}$ [Eq. (7)] includes the exact solutions to the defect problem. Nevertheless, it is seen that the DGF approach produces poor defect wave functions if the representation in terms of host wave functions [Eq. (3)] is incomplete.

that in fact an adequate Green's-function approach to transition-atom impurities in semiconductors needs to accurately describe the atomiclike hyperdeep wave functions whose energy may lie in the lower part of the valence band, or even below it. It then becomes relevant to ask how many host bands are needed to represent the wave function $\psi_1(\vec{r})$ of such a hyperdeep defect state.

To estimate this number, let us assume that the band-structure calculation for the host crystal is performed in a plane-wave basis:

$$\phi_j^{\text{BS}}(\vec{k}, \vec{r}) = \sum_{\vec{G}} B_j(\vec{k} + \vec{G}) e^{i(\vec{k} + \vec{G}) \cdot \vec{r}}. \quad (12)$$

The question of how many host bands are needed is, of course, independent of the nature of the underlying basis set used in the band-structure calculation, provided that the calculation is *converged* with respect to the basis functions in the energy region of interest. By substituting Eq. (12) into Eq. (3), it is then clear that one has to incorporate a

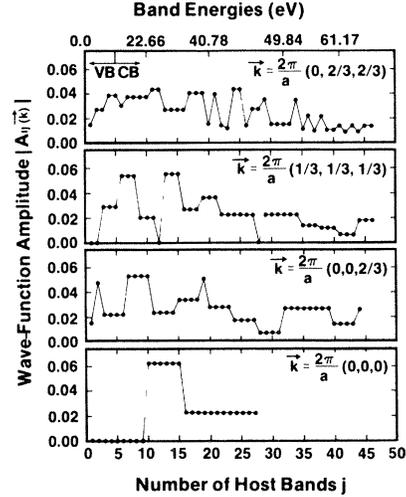


FIG. 2. Spectral decomposition of the defect wave function $\psi_i(\vec{r})$ for state e of a parabolic defect potential in a silicon free-electron host [Eq. (3)]. Shown are the absolute values of the expansion coefficients $|A_{ij}(\vec{k})|$ of the state $|i\rangle = e$ in terms of the host wave functions $|j, \vec{k}_p\rangle$ for four wave-vector values \vec{k}_p (given in Cartesian coordinates). The normalization is such that $\sum_{j=1}^{M_i} \sum_{\vec{k}_p=1}^4 |A_{ij}(\vec{k}_p)|^2 d_p = 1$, where d_p is the number of wave vectors in the star of \vec{k}_p . The bands are arranged in increasing order of $|j\rangle$; their one-electron energy is shown on the top abscissa. The discrete values of $|A_{ij}(\vec{k}_p)|$ are connected by straight lines to guide the eye. Notice that the defect wave function draws its intensity from many host wave functions, the contribution of the high-energy states being very significant.

sufficient number of plane waves in the band-structure calculation to correspond to the highest momentum component q_{max} present in the impurity wave function:

$$\begin{aligned} \psi_1(\vec{r}) &= \sum_{\vec{k}} \sum_{\vec{G}} \sum_j^{M_i} A_{ij}(\vec{k}) B_j(\vec{k} + \vec{G}) e^{i(\vec{k} + \vec{G}) \cdot \vec{r}} \\ &= \sum_{\vec{q}}^{q_{\text{max}}} A_1(\vec{q}) e^{i\vec{q} \cdot \vec{r}}, \end{aligned}$$

where

$$A_1(\vec{q}) = \int \psi_1(\vec{r}) e^{-i\vec{q} \cdot \vec{r}} d\vec{r}$$

is the Fourier transform of the impurity wave function and \vec{q}_{max} is the largest momentum $\vec{k} + \vec{G}$ appearing in the host wave functions. Furthermore, it is also clear that, in principle, one has to use as many bands as there are plane waves since only then there exists a unitary transformation between the band-structure wave functions $\{\phi_i^{\text{BS}}(\vec{k}, \vec{r})\}$ and

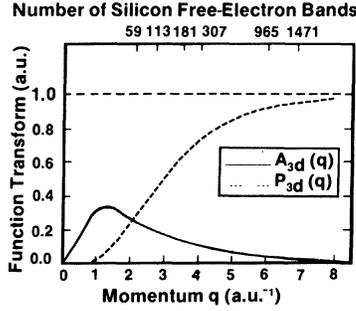


FIG. 3. The radial part of the Fourier transform $A_{nl}(q)$ [Eq. (13)] of the $3d$ orbital of Fe (full line). Note its long tail in momentum space. The dashed curve shows the function $P(q) = \int_0^q A_{nl}(k)k^2 dk$: Its departure from unity measures the incompleteness of a description of this orbital wave function in terms of silicon free-electron-band wave functions with a maximum momentum q . The number of such bands is indicated on the upper abscissa. Further, note the large number of free-electron bands required to achieve a reasonable accuracy in the description of the orbital. The consequences of small errors in such localized “defect” wave functions on the defect energy levels is demonstrated in Table II for the Cr atom.

the plane waves $\{e^{i(\vec{k} + \vec{G}) \cdot \vec{r}}\}$ to make these two sets of functions equivalent. If, however, one restricts M_i and G_{\max} so that the highest momentum q_{\max} included effectively in Eq. (12) is $q_{\max}^{\text{eff}} < q_{\max}$, then the defect wave function $\psi_1(\vec{r})$ will not equal the exact solution; instead it will be

$$\psi_1^{\text{approx}}(\vec{r}) = \sum_{\vec{q}}^{q_{\max}^{\text{eff}}} A_1(\vec{q}) e^{i\vec{q} \cdot \vec{r}}. \quad (13)$$

Similarly, the orbital energy will be $\epsilon_{nl}^{\text{approx}}$ rather

TABLE II. The orbital energy of a Cr $3d$ level as calculated by expanding its wave functions in silicon empty-lattice band wave functions [Eq. (3)]. We show separately the contributions from the orbital potential energy and kinetic energy. The exact result (“ ∞ bands”) is obtained by numerically integrating the single-particle equation without expanding the wave function in a finite basis. Note the enormous number of band wave functions needed to achieve a modest precision of 0.4–0.2 eV in the orbital energy.

Number M_i of host bands included in expansion [Eq. (3)]	q_{\max} (a.u. ⁻¹)	Orbital kinetic energy (eV)	Orbital potential energy (eV)	Orbital energy (eV)	Error in orbital energy (eV)
4 570	10	138.984	-139.887	-0.903	-4.74
15 410	15	147.396	-151.731	-4.335	-1.31
36 520	20	148.898	-154.091	-5.193	-0.45
69 630	24.8	149.226	-154.635	-5.409	-0.23
∞	∞	149.370	-155.013	-5.643	0.0

than the exact atomic value ϵ_{nl} . One can then estimate the effect of truncating q_{\max} to a smaller value q_{\max}^{eff} by making a simple approximation to $\psi_1(\vec{r})$: for hyperdeep states, one sets $\psi_1(|\vec{r}|) = R_{nl}(r)$. Figure 3 shows the Fourier transform $A_{3d}(q)$ of the Fe $3d$ orbital, as well as the quantity $P_{3d}(q) = \int_0^q A_{3d}(q')q'^2 dq'$. The approximate wave function $\psi_{nl}^{\text{approx}}(\vec{r})$ becomes exact for a momentum q_{\max} such that $P_{3d}(q_{\max}) = 1$. Not surprisingly, the long tail of $A_{3d}(q)$ in momentum space requires a high q_{\max} value (the equivalent number of host free-electron bands is indicated on the abscissa) to obtain $P_{3d}(q) \approx 1$. We can now find the error $|\epsilon_{nl} - \epsilon_{nl}^{\text{approx}}|$ associated with using various finite sizes of the host wave-function basis in Eq. (3). Table II illustrates the number of host bands M_i (and the corresponding highest momentum q_{\max} contained in them) which is required to obtain a given accuracy $|\epsilon_{nl} - \epsilon_{nl}^{\text{approx}}|$ for the $3d$ hyperdeep level of Cr: about 36 000 silicon-type free-electron bands are required for a precision of 0.45 eV, while ≈ 70 000 bands are required for a precision of 0.2 eV. While not surprisingly, the conventional Green’s-function method is inapplicable to such problems, it is of interest to notice what is the extent of the effort that would have been required to obtain a useful precision. To the extent that the hyperdeep impurity wave function $\psi_1(\vec{r})$ is similar to an atomic orbital, one would need to (i) perform the band-structure calculation for the host crystal to an accuracy corresponding to using around 70 000 plane-wave basis functions and, (ii) to use 70 000 bands in the evaluation of the host-crystal Green’s function in order to achieve an accuracy of about 0.2 eV in the defect energy. A lower number of

bands will obviously be needed if $\psi_1(\vec{r})$ becomes delocalized through hybridization. However, atomiclike hyperdeep levels are frequently found for many transition-atom and rare-earth impurities.^{28,29} Similarly, the incapability of a small number of host bands to adequately describe defect-induced perturbations is likely to show itself when lattice distortions are present.

The source of this difficulty with the standard DGF approach can be easily appreciated from the structure of Eq. (4). Even small interband mixing coefficients $A_{ij}(\vec{k})$ corresponding to high-energy conduction bands $\epsilon_j^0(\vec{k}) \gg \epsilon_i$ (and having the appropriate point-group representation) may contribute significantly to the solution due to the large size of the product $A_{ij}(\vec{k})[\epsilon_j^0(\vec{k}) - \epsilon_i]$. Further, the existence of a high density of host states at high energies in the conduction bands (where the host bands of many materials are already free-electron-like) increases their contributions to the sums in Eq. (4). Finally, as $[\epsilon_j^0(\vec{k}) - \epsilon_i]$ is either positive (j in the conduction band) or negative (j in the valence band), the determination of bandgap defect states ϵ_i involves a delicate balance of large canceling terms. Unfortunately, the small residue of these extensive cancellations contains chemically and physically important trends. This difficulty is further highlighted by the fact that while the central quantity in self-consistent defect calculations, the charge perturbation

$$\Delta\rho(\vec{r}) = \sum_i^{\text{occ}} |\psi_i|^2 - \sum_{j,\vec{k}}^{\text{occ}} |\phi_j^0(\vec{k}, \vec{r})|^2$$

[Eqs. (1) and (2)], is localized¹⁰ in the subspace of the perturbation $\Delta V(\vec{r})$, its evaluation involves the calculation of a large number of spatially extended components $\{\psi_i, \phi_j^0(\vec{k}, \vec{r})\}$ that interfere destructively to produce a localized fluctuation $\Delta\rho(\vec{r})$.

Note that while the problem of inherent slow convergence of $\psi_i(\vec{r})$ with the number of host bands exists also in the various tight-binding Green's-function approaches to defects,^{24,31-33} it remains untreated since these models are restricted by construction to have a fixed and small number of host bands (a total of 8 and 10). In such studies one uses a small and fixed number of host bands to study the variations in gap-state energies of impurities having a diverse range of central-cell potentials²⁴ (e.g., +5 to -40 eV). The assumption that the same small number of bands is sufficient to describe a large variety of impurities is likely to invalidate even the trends in the energy levels²⁴ (let

alone the absolute values) predicted by such models.³⁵

The existence of a dual representation in DGF methods, therefore, poses an obstacle to constructing computationally simple and physically transparent basis sets. This basic difficulty with the standard Green's-function approach to localized impurities may be treated with brute force by increasing the number M_i of host bands to either the convergence limit or to the computer's capability limit. (As demonstrated, it is likely that the latter limit is attained first.) However, the physical transparency of the model may be lost. In the next section, we will describe a simple method to overcome the difficulty by redefining the "unperturbed" Bloch functions to include not only host but also impurity characteristics. Only a few such "quasi band" wave functions will be required to achieve a result that is in principle exact, and in practice very precise (Table I).

IV. THE QUASI BAND STRUCTURE APPROACH TO THE DEEP DEFECT PROBLEM

In the standard derivation of the Green's-function approach to the localized defect problem (e.g., Refs. 10 and 11) one assumes that the orthogonal crystal wave functions $\phi_j^0(\vec{k}, \vec{r})$ are *eigenfunctions to the host-crystal Hamiltonian operator* $\hat{H}_0 = -\frac{1}{2}\nabla^2 + V_H$ [Eq. (2)]. By expanding the defect wave functions in this set $\{\phi_j^0(\vec{k}, \vec{r})\}$ one gets the basic Eq. (4) (cf. Appendix A). As discussed in the previous section, the requirements of Eqs. (2) and (3) will often result in a computationally intractable and physically undesirable large number M_i of host bands required to represent a localized defect state $\psi_i(\vec{r})$. The central point of our argument on this issue is that while the condition that $\phi_j^0(\vec{k}, \vec{r})$ be an eigenstate to \hat{H}_0 is sufficient for deriving the underlying formalism,⁹ it is in fact not a necessary condition. *The far weaker condition that an arbitrarily chosen orthogonal set of Bloch functions $\{\phi_j^{\text{QB}}(\vec{k}, \vec{r})\}$ diagonalizes the host-crystal Hamiltonian matrix is, in fact, a sufficient condition.* One is then free to choose the set $\{\phi_j^{\text{QB}}(\vec{k}, \vec{r})\}$ (quasi band functions) with its eigenvalues $\epsilon_j^{\text{QB}}(\vec{k})$ (quasi band structure) so that the impurity wave function $\psi_i(\vec{r})$ is expanded in principle exactly and in practice very accurately by a *small number \bar{M}_i* of such quasi bands

$$\psi_i(\vec{r}) = \sum_j^{\bar{M}_i} \sum_{\vec{k}}^{\text{BZ}} A_{ij}^{\text{QB}}(\vec{k}) \phi_j^{\text{QB}}(\vec{k}, \vec{r}) \quad (14)$$

The *conventional* choice^{6–11,31–33} where $\phi_j^{\text{QB}}(\vec{k}, \vec{r})$ is replaced by some approximation to $\phi_j^0(\vec{k}, \vec{r})$ in Eq. (14), is only a particular case. However, far better choices that incorporate in $\phi_j^{\text{QB}}(\vec{k}, \vec{r})$ the physical characteristics of the defect wave functions from the outset are possible. We will show below: (i) that in fact the weaker conditions are sufficient, and (ii) the way one can build into the set $\{\phi_j^{\text{QB}}(\vec{k}, \vec{r})\}$ the physical characteristics that make the expansion in Eq. (14) rapidly convergent. The dual description of the defect wave functions poses no difficulties in such a representation.

The minimal sufficient conditions that lead to the basic Eq. (4) can be easily derived from the variational principle. We first expand the desired solution $\psi_i(\vec{r})$ of Eq. (1) in a Bloch set $\{\phi_j^{\text{QB}}(\vec{k}, \vec{r})\}$ as indicated in Eq. (14), where $\phi_j^{\text{QB}}(\vec{k}, \vec{r})$ are at this point any set of Bloch-type functions. We will find the conditions that this set has to satisfy in order to recover the basic Eq. (4). By substituting the expansion Eq. (14) in the variational equation,

$$\delta \langle \psi_i(\vec{r}) | \hat{H}_0 + \Delta V(\vec{r}) - \epsilon_i | \psi_i(\vec{r}) \rangle = 0, \quad (15)$$

one obtains

$$\sum_{j=1}^{M_i} \sum_{\vec{k}} A_{ij}^{\text{QB}}(\vec{k}) [\langle \phi_j^{\text{QB}}(\vec{k}', \vec{r}) | \hat{H}_0 | \phi_j^{\text{QB}}(\vec{k}, \vec{r}) \rangle + \langle \phi_j^{\text{QB}}(\vec{k}', \vec{r}) | \Delta V(\vec{r}) | \phi_j^{\text{QB}}(\vec{k}, \vec{r}) \rangle - \epsilon_i \langle \phi_j^{\text{QB}}(\vec{k}', \vec{r}) | \phi_j^{\text{QB}}(\vec{k}, \vec{r}) \rangle] = 0. \quad (16)$$

It is then obvious that Eq. (16) reduces to the fundamental Eq. (4) if we require that the set $\{\phi_j^{\text{QB}}(\vec{k}, \vec{r})\}$ diagonalizes the matrix of the crystal Hamiltonian $\hat{H}_0 = -\frac{1}{2}\nabla^2 + V_H(\vec{r})$ with diagonal elements $\epsilon_j^{\text{QB}}(\vec{k})$:

$$\langle \phi_j^{\text{QB}}(\vec{k}', \vec{r}) | \hat{H}_0 | \phi_j^{\text{QB}}(\vec{k}, \vec{r}) \rangle = \epsilon_j^{\text{QB}}(\vec{k}) \delta_{jj'} \delta_{\vec{k}\vec{k}'}. \quad (17)$$

This set is then orthogonal

$$\int \phi_j^{\text{QB}*}(\vec{k}, \vec{r}) \phi_j^{\text{QB}}(\vec{k}', \vec{r}) d\vec{r} = \delta_{jj'} \delta_{\vec{k}\vec{k}'}. \quad (18)$$

Conditions (17) and (18) transform Eq. (16) into

$$[\epsilon_j^{\text{QB}}(\vec{k}') - \epsilon_i] A_{ij}^{\text{QB}}(\vec{k}') + \sum_J \sum_{\vec{k}} A_{ij}^{\text{QB}}(\vec{k}) \langle \phi_j^{\text{QB}}(\vec{k}', \vec{r}) | \Delta V(\vec{r}) | \phi_j^{\text{QB}}(\vec{k}, \vec{r}) \rangle = 0, \quad (19)$$

which is identical to Eq. (4) except that $\epsilon_j^0(\vec{k}')$ and $\phi_j^0(\vec{k}, \vec{r})$ are replaced by $\epsilon_j^{\text{QB}}(\vec{k}')$ and $\phi_j^{\text{QB}}(\vec{k}, \vec{r})$, respectively. Notice that we have not required the far stronger condition—that $\phi_j^{\text{QB}}(\vec{k}, \vec{r})$ be an eigenfunction to \hat{H}_0 [cf. Eq. (2)] or that $\epsilon_j^{\text{QB}}(\vec{k}) = \epsilon_j^0(\vec{k})$ be valid.³⁶

It is important to realize that the set $\{\phi_j^0(\vec{k}, \vec{r})\}$ which is orthogonal and an *eigenstate* to \hat{H}_0 is unique within a unitary transformation, whereas the set $\{\phi_j^{\text{QB}}(\vec{k}, \vec{r})\}$ of quasi bands which is orthogonal and merely *diagonalizes the matrix* of \hat{H}_0 is not. In fact, there is an infinite number of choices of quasi band functions $\{\phi_j^{\text{QB}}(\vec{k}, \vec{r})\}$ that fulfill Eqs. (17) and (18) but not Eq. (2).³⁷ For example, one can construct a *finite set* of \bar{M}_a Bloch functions from some arbitrary local orbitals $\{f_j(\vec{r})\}$:

$$\chi_j(\vec{k}, \vec{r}) = \frac{1}{\sqrt{N}} \sum_{\vec{R}_p} e^{i\vec{k}\cdot\vec{R}_p} f_j(\vec{r} - \vec{R}_p), \quad j = 1 \dots \bar{M}_a \quad (20)$$

then generate the Hamiltonian matrix

$\langle \chi_j(\vec{k}, \vec{r}) | \hat{H}_0 | \chi_{j'}(\vec{k}, \vec{r}) \rangle$ and diagonalize it to obtain the quasi band structure $\epsilon_j^{\text{QB}}(\vec{k})$ and the quasi-band wave functions:

$$\phi_j^{\text{QB}}(\vec{k}, \vec{r}) = \sum_{j'=1}^{\bar{M}_a} a_{jj'}(\vec{k}) \chi_{j'}(\vec{k}, \vec{r}), \quad (21)$$

where $a_{jj'}(\vec{k})$ are the elements of the diagonalizing transformation. The functions $\{\phi_j^{\text{QB}}(\vec{k}, \vec{r})\}$ fulfill Eqs. (17) and (18) by construction with some constants $\epsilon_j^{\text{QB}}(\vec{k})$. If the set $\{f_j(\vec{r})\}$ is incomplete and physically unrelated to \hat{H}_0 (e.g., if $\hat{H}_0 = -\frac{1}{2}\nabla^2$ and $\{f_j(\vec{r})\}$ is a small finite set of hydrogenic functions), then $\phi_j^{\text{QB}}(\vec{k}, \vec{r})$ and $\epsilon_j^{\text{QB}}(\vec{k})$ will be very different from $\phi_j^0(\vec{k}, \vec{r})$ and $\epsilon_j^0(\vec{k})$, respectively. We will show that a suitable choice of $\phi_j^{\text{QB}}(\vec{k}, \vec{r}) \neq \phi_j^0(\vec{k}, \vec{r})$ will make the summation over \bar{M}_i in Eq. (19) converge much faster than in the analogous Eq. (4).

Notice that in the example given in Eqs. (20) and (21) for constructing the quasi band wave functions $\phi_j^{\text{QB}}(\vec{k}, \vec{r})$ one would obtain the true eigen-

function $\phi_j^{\text{QB}}(\vec{k}, \vec{r}) \rightarrow \phi_j^0(\vec{k}, \vec{r})$ if the set $\{f_j(\vec{r})\}$ were complete. Practically all nonanalytic band-structure calculations indeed obtain wave functions $\phi_j^{\text{BS}}(\vec{k}, \vec{r})$ that satisfy orthogonality and diagonalize the matrix of \hat{H}_0 but do not satisfy Eq. (2). In this sense, the band-structure wave functions $\phi_j^{\text{BS}}(\vec{k}, \vec{r})$ are only “quasi bands.” One then often attempts in band-structure calculations to make $\phi_j^{\text{BS}}(\vec{k}, \vec{r})$ a better approximation to the true eigenstates $\phi_j^0(\vec{k}, \vec{r})$ by systematically increasing the number of basis functions $\{f_j(\vec{r})\}$. One way of doing this is to use a mixed representation³⁸:

$$\phi_j^{\text{QB}}(\vec{k}, \vec{r}) = \sum_j^{M_b} b_{jj'}(\vec{k}) \phi_j^{\text{BS}}(\vec{k}, \vec{r}) + \sum_{j''}^{M_a} a_{jj''}(\vec{k}) \chi_{j''}(\vec{k}, \vec{r}), \quad (22)$$

where the band-structure wave functions (e.g., from a plane-wave basis) $\phi_j^{\text{BS}}(\vec{k}, \vec{r})$ are augmented by another set (e.g., localized LCAO Bloch functions) $\chi_{j''}(\vec{k}, \vec{r})$ to make $\phi_j^{\text{QB}}(\vec{k}, \vec{r})$ a better approximation to the eigenstate $\phi_j^0(\vec{k}, \vec{r})$ in Eq. (2). The auxiliary basis $\{f_j(\vec{r})\}$ in Eq. (20) is then chosen in band-structure studies to have the characteristics of the *host* system [e.g., identifying $f_j(\vec{r})$ with *host* atomic orbitals in LCAO expansions]. We have shown, however, that if the defect problem is solved via a Green's-function formalism, full convergence of the band wave functions $\{\phi_j^{\text{BS}}(\vec{k}, \vec{r})\}$ to the true eigenfunctions $\{\phi_j^0(\vec{k}, \vec{r})\}$ is not even required formally; the orthogonal set $\{\phi_j^{\text{QB}}(\vec{k}, \vec{r})\}$ that spans $\psi_i(\vec{r})$ need only diagonalize the finite matrix of \hat{H}_0 . Therefore, instead of seeking convergence $\phi_j^{\text{QB}}(\vec{k}, \vec{r}) \rightarrow \phi_j^0(\vec{k}, \vec{r})$ as in conventional band-structure calculations, we take the opposite view: We will use a *manifestly incomplete* (under-converged) finite set $\{\phi_j^{\text{QB}}(\vec{k}, \vec{r})\}$ of the form in Eq. (22), but choose $\{f_j(\vec{r})\}$ of Eq. (20) [and therefore $\chi_j(\vec{k}, \vec{r})$ in Eq. (22)] to share the characteristics of the perturbation $\Delta V(\vec{r})$ rather than those of the host potential $V_H(\vec{r})$. By the variational principle, augmenting the band-structure wave functions $\{\phi_j^{\text{BS}}(\vec{k}, \vec{r})\}$ by another set $\{\chi_j(\vec{k}, \vec{r})\}$ (even if it is physically unrelated) can do no harm (in practice, cf. Fig. 6 below, it makes almost no difference). At the same time, a judicious choice of $\{f_j(\vec{r})\}$ can make the expansion of the defect wave function $\psi_i(\vec{r})$ in terms of the quasi bands $\phi_j^{\text{QB}}(\vec{k}, \vec{r})$ [Eq. (14)] converge much faster than the corresponding expansion in terms of pure host wave functions alone [Eq. (3)]. The coefficients $\{b_{jj'}(\vec{k}); a_{jj''}(\vec{k})\}$ and the quasi band energies $\{\epsilon_j^{\text{QB}}(\vec{k})\}$ in Eq. (22) are obtained directly by diagonalizing once the ma-

trix of \hat{H}_0 in the nonorthogonal basis $\{\phi_j^{\text{BS}}(\vec{k}, \vec{r}); \chi_j(\vec{k}, \vec{r})\}$ [Eq. (22)].

Next we will show that the set $\{\phi_j^{\text{QB}}(\vec{k}, \vec{r})\}$ of Eq. (22) can be used to represent arbitrary localized solutions $\psi_i(\vec{r})$. To do so we first note that the subset $\{\chi_{j''}(\vec{k}, \vec{r})\}$ of Eq. (22) alone is capable of reproducing any of the localized functions $f_i(\vec{r})$, since³⁹

$$\begin{aligned} \sum_{\vec{k}}^{\text{BZ}} \chi_j(\vec{k}, \vec{r}) &= \frac{1}{\sqrt{N}} \sum_{\vec{R}_p} \sum_{\vec{k}}^{\text{BZ}} e^{i\vec{k} \cdot \vec{R}_p} f_j(\vec{r} - \vec{R}_p) \\ &= N^{1/2} f_j(\vec{r}). \end{aligned} \quad (23)$$

Therefore, $f_j(\vec{r}) = N^{-1/2} \sum_{\vec{k}} \chi_j(\vec{k}, \vec{r})$. Secondly, we note that the set $\{\chi_j(\vec{k}, \vec{r})\}$ is a subset in Eq. (22); hence, if it can reproduce a localized function $f_j(\vec{r})$, then so can $\{\phi_j^{\text{QB}}(\vec{k}, \vec{r})\}$. This means that if $f_j(\vec{r})$ were to be the actual localized defect function $\psi_i(\vec{r})$, then the set $\{\phi_j^{\text{QB}}(\vec{k}, \vec{r})\}$ of quasi band wave functions will give us the exact solution and therefore the exact energy eigenvalue. This is demonstrated in the next section for a simple case. If, however, our choice of $f_j(\vec{r})$ is only an approximation for the true solution $\psi_i(\vec{r})$, then sufficient components of host wave functions $\{\phi_j^{\text{BS}}(\vec{k}, \vec{r})\}$ will be introduced in Eq. (22). However, these have to reproduce merely the *difference* between $\psi_i(\vec{r})$ and the trial orbitals $f_j(\vec{r})$. This is far easier to accomplish (e.g., if needed, iteratively) than expanding $\psi_i(\vec{r})$ in $\{\phi_j^{\text{BS}}(\vec{k}, \vec{r})\}$ alone, as done in previous approaches.⁶⁻¹¹

We have shown thus far how the quasi band representation overcomes the difficulties associated with representing the defect wave functions in terms of unperturbed Bloch functions [Eq. (14)]. Still, one needs to independently represent the defect wave functions by a local set [Eq. (7)]. This can be done straightforwardly by including in the local set $\{g_a(\vec{r})\}$ also the basis functions $\{f_j(\vec{r})\}$ that are likely to reproduce the prominent nonhost-like characteristics of the defect wave functions. Clearly, this choice is amenable to an iterative optimization.

To summarize the procedure, we indicate the sequence of steps involved: (i) select a localized basis of N functions $\{g_a(\vec{r})\}$ for Eq. (7) and include in it M_a basis functions $f_j(\vec{r})$ that are likely to reproduce the major features of the most localized defect wave functions. For impurities, a natural choice of $f_j(\vec{r})$ is the corresponding set of impurity atomic orbitals, compressed into a Wigner-Seitz cell. Many other choices are possible; e.g., symmetrized atomic orbitals, wave functions that result from defect cluster calculations, wave functions that result

from a previous iteration in the DGF calculation, etc. (ii) Form a set of M_a Bloch functions $\chi_j(\vec{k}, \vec{r})$ [Eq. (20)] from the set $f_j(\vec{r})$.⁴⁰ (iii) Diagonalize once the matrix of the host-crystal Hamiltonian in a basis of M_b host band-structure functions $\{\phi_j^{\text{BS}}(\vec{k}, \vec{r})\}$ and M_a local bands $\{\chi_j(\vec{k}, \vec{r})\}$ to obtain $\bar{M} = M_a + M_b$ quasi band wave functions $\{\phi_j^{\text{QB}}(\vec{k}, \vec{r})\}$ and energies $\{\epsilon_j^{\text{QB}}(\vec{k})\}$ in Eq. (22). (iv) Solve the standard DGF problem in Eqs. (8)–(10) by using $\{\phi_j^{\text{QB}}(\vec{k}, \vec{r})\}$ instead of $\{\phi_j^0(\vec{k}, \vec{r})\}$ and $\{\epsilon_j^{\text{QB}}(\vec{k})\}$ instead of $\{\epsilon_j^0(\vec{k})\}$. The N -component set of local orbitals $\{g_a(\vec{r})\}$ appearing in the DGF equations includes the basis functions $\{f_j(\vec{r})\}$. (v) Optimize the convergence of $\epsilon_i(M_a, M_b, N)$ by modifying $\{f_j(\vec{r})\}$; a suitable choice will minimize the numbers (M_b, N) of other basis functions needed. In the next section, we illustrate the method with a few detailed examples.

V. ILLUSTRATIVE EXAMPLES

The power of the quasi band approach can be simply illustrated on a case that has closed-form analytic solutions: A parabolic (harmonic oscillator) impurity potential well $\Delta V(\vec{r}) = K(\vec{r} - \vec{r}_0)^2$ in a diamond-lattice free-electron host $\hat{H}_0 = -\frac{1}{2}\nabla^2$ discussed in Sec. III. We use as before the exact solution [Eq. (11b)] as the localized basis $\{g_a(\vec{r})\}$ in Eq. (7). For the host basis, we use in Eq. (22) M_b band-structure wave functions of the form $\phi_j^{\text{BS}}(\vec{k}, \vec{r}) = 1/\sqrt{\Omega} e^{i(\vec{k} + \vec{G}) \cdot \vec{r}}$ and M_a local Bloch functions $\chi_j(\vec{k}, \vec{r})$ in which the basis $f_j(\vec{r})$ is chosen as harmonic oscillator orbitals (i.e., exact solutions to $\hat{H}_0 + \Delta V$). We diagonalize $\hat{H}_0 = -\frac{1}{2}\nabla^2$ within the basis of Eq. (22) to obtain the quasi band wave functions $\{\phi_j^{\text{QB}}(\vec{k}, \vec{r})\}$. Together with the local basis $\{g_a(\vec{r})\}$ of harmonic oscillator orbitals, we solve the DGF problem in Eqs. (8)–(10), replacing $\{\phi_j^0(\vec{k}, \vec{r})\}$ by $\{\phi_j^{\text{QB}}(\vec{k}, \vec{r})\}$, and $\{\epsilon_j^0(\vec{k})\}$ by $\{\epsilon_j^{\text{QB}}(\vec{k})\}$. We now ask how many host bands M_b are required to obtain the exact defect energy level.

The upper part of Table I, discussed in Sec. II, shows the poor convergence of the defect energy levels obtained in the standard DGF approach to the problem. The lower half of Table I shows the results obtained in the quasi band approach. We use $M_a = 6$ local functions $f_j(\vec{r})$ corresponding to the three lowest defect states $(nl) = (0,0)$, $(1,1)$, and $(2,2)$; i.e., representations a_1 , t_2 , and e_2 , respectively. It is seen that *even a single band-structure wave function ($M_b = 1$) is sufficient to recover the exact*

results. Naturally, one also obtains the exact defect wave functions for these states. To make the test more severe, we solve also for the first excited a_1^* defect level *without including a matching quasi band wave function in the basis set.* It is seen that the convergence of its energy level to the exact result is far better than that obtained in the conventional DGF method. In fact, using only $M_b = 10$ band-structure wave functions in the quasi band approach produces far better results than 41 band-structure wave functions in the conventional DGF method ($M_a = 0$). Notice, however, that this is an extremely severe test case: In practical applications for impurities in solids, one rarely finds more than a single state of a given symmetry having all of its wave function localized within an atomic sphere, such as a_1 and a_1^* here. If such a situation occurs, one can easily add a new quasi band wave function in the present formalism to represent such an excited hyperlocalized state. For example, the additional quasi band wave function can be constructed from $f_j(\vec{r}) = \psi_j^{\text{CDGF}}(\vec{r})$, where $\psi_j^{\text{CDGF}}(\vec{r})$ is an approximation to the true defect wave function obtained in the conventional defect Green's-function (CDGF) calculation.

Figure 4 compares the a_1^* defect wave function obtained in the CDGF calculation with the quasi band approach. Clearly, the conventional DGF method produces a poor approximation to $\psi_{a_1^*}(\vec{r})$, even if as many as 41 host-band wave functions are included in the representation. The quasi band approach, on the other hand, yields far smaller errors. For the lowest defect states a_1 , t_2 , and e_2 , the quasi band approach yields the exact defect wave functions, even if only a single host band (needed merely to define the zero of energy) is included in the expansion of Eq. (3).

Our second example is a Cr $3d$ impurity in a silicon free-electron host. As discussed in Sec. III, an application of the conventional DGF method to the problem requires an enormous number of host-band wave functions to recover the hyperdeep impurity energy levels (cf. Table II). We have repeated the calculation of the Cr impurity in a silicon empty lattice using the quasi band representation. Both the local-orbital set $\{g_a(\vec{r})\}$ and the quasi band set $\{\phi_j^{\text{QB}}(\vec{k}, \vec{r})\}$ include the orbital $f_j(\vec{r})$ identified here as the Cr $3d$ atomic wave function. This is readily obtained from an accurate numerical integration of the Schrödinger equation for Cr in free space using a local-density approximation for exchange (the exchange coefficient α is chosen as one). Using only two quasi bands and local orbi-

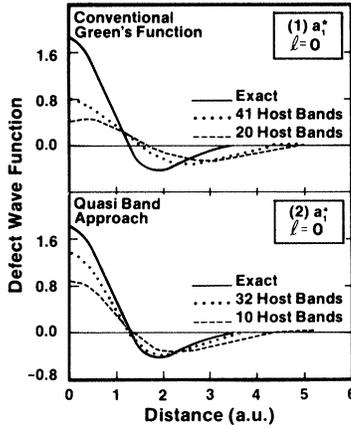


FIG. 4. The first excited a_1^* defect level of a parabolic defect potential perturbation in a silicon free-electron host: (1) as obtained in the conventional defect Green's-function method; (2) as obtained in the quasi band method but without including an a_1^* quasi band in the basis. Note that in case (1) a poor approximation is obtained to the exact result [Eq. (11b)], even when as many as 41 host-band wave functions are included in the expansion [cf. Eq. (3)].

tals, we have retrieved the Cr $3d$ defect energy level of -5.687 eV, compared to the exact result of -5.643 eV (Table II). To obtain an accuracy five times worse (0.2 eV) in the conventional DGF method, one needs to use as many as 70 000 host bands (cf. Table II). The source of the remaining small error of 0.04 eV in the defect energy level is related to the particular form of the Green's-function equation used. This is discussed in Appendix B.

Finally, we demonstrate that an adequate Green's-function approach to very deep impurity levels needs, in fact, to accurately reproduce the very localized atomiclike wave functions. We have performed a self-consistent calculation for a substitutional Cu impurity in silicon using the nonlocal first-principles atomic pseudopotentials⁴¹ and the present quasi band approach. While the detailed results will be discussed elsewhere,⁴² we comment here on the features pertinent to the present issues.

We find a partially occupied vacancylike t_2 level at 0.47 eV above valence-band maximum (E_v); the details of its wave function and energy are, however, dependent on the character of two resonances of e and t_2 symmetry (denoted e^* and t_2^*) at about 5.2 eV below E_v . As much as 90% of the charge density of these hyperdeep levels is enclosed within a sphere of radius R_0 which equals the host bond length (4.44 a.u.). Figure 5 shows the quantity $r\psi_{e^*}(r)$ for the e^* hyperdeep levels compared with

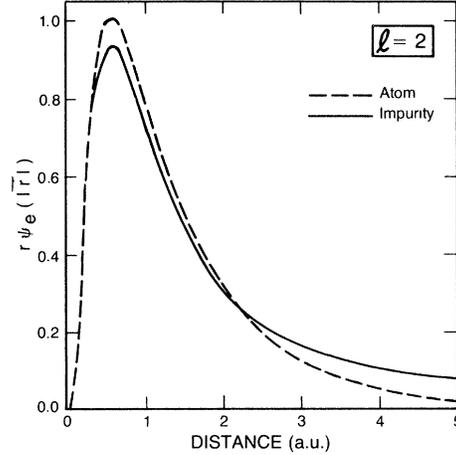


FIG. 5. r times the radial wave function $\psi_e(|\vec{r}|)$ of the hyperdeep e^* state of substitutional Cu in silicon (full line) compared with r times the radial atomic Cu $3d$ wave function (dashed line). The radial part of the t_2^* state is very similar to the e^* state.

r times the radial atomic $3d$ orbital of Cu. Clearly, these defect levels are as localized as the atomic $3d$ orbital for $r \leq 3$ a.u. Much like the case demonstrated in Tables II and III, a conventional DGF approach could not reproduce the wave function of such states without using an unreasonably high number of host wave functions. Notice further that since these hyperdeep states are within the valence bands and therefore hybridize with the host states, one could not solve the problem by simply pseudizing the Cu $3d$ orbitals, treating them as a part of an inert core. This highlights the substantial difficulties associated with treating deep impurity levels originating from localized but chemically active orbitals.

To further illustrate the nature of the quasi bands, Fig. 6 shows the quasi band structure of substitutional Cu in crystalline silicon in the Δ direction in the BZ, obtained by diagonalizing the self-consistent pseudopotential host Hamiltonian in the basis of Eq. (22). For the basis functions $\phi_j^{\text{BS}}(\vec{k}, \vec{r})$, we use the $M_b = 14$ lowest-band wave functions obtained in a standard self-consistent calculation for Si: 181 plane waves are used to expand $\phi_j^{\text{BS}}(\vec{k}, \vec{r})$ and 10 special \vec{k} points are employed in the self-consistency iterations. For the localized function $\chi_j(\vec{k}, \vec{r})$, we use $M_a = 5$ Bloch sums [Eq. (20)] (corresponding to the sums of the dimensions of the e and t_2 representations) formed from the numerical Cu $3d$ orbital obtained in a Herman-Skillman atomic calculation. The lowest M_b bands in Fig. 6 are in fact almost pure silicon bands as obtained from $\{\phi_j^{\text{BS}}(\vec{k}, \vec{r})\}$ alone: Including the local functions $\chi_j(\vec{k}, \vec{r})$ results only in an extremely small

TABLE III. The calculated 3d orbital energy of a Cr-atom impurity in an empty-lattice host using the quasi band approach. $M_a=2$ local-band wave functions and $M_b=14$ band-structure wave functions. By varying the nonlinear parameter Z^* of these Slater-type local orbitals one is able to recover the exact energy level for $Z^*=12.25$ to within a precision of better than 1%. The conventional DGF method requires $M_b \geq 7 \times 10^4$ host-band wave functions for a precision of 0.2 eV (Table II).

Exponent Z^* of local Coulomb orbitals	Defect energy level (eV)	Error relative to exact result (eV)
10.0	-5.993	0.35
11.0	-5.729	0.086
11.5	-5.707	0.064
12.0	-5.692	0.049
12.25	-5.687	0.044
12.5	-5.690	0.047
13.0	-5.731	0.088
14.0	-5.845	0.202
15.0	-5.913	0.270
Exact	-5.643	0.000
Conventional DGF, ~70000 bands	-5.409	-0.23

variational improvement (that cannot be seen on the scale of Fig. 6) over the results obtained with $\{\phi_j^{\text{BS}}(\vec{k}, \vec{r})\}$ alone. The next M_a bands correspond to a mixture of local functions and band-structure functions. All $M_a + M_b$ band wave functions are mutually orthogonal [Eq. (18)] and by construction diagonalize the Hamiltonian matrix [Eq. (17)] with eigenvalues $\epsilon_j^{\text{QB}}(\vec{k})$ shown in Fig. 6.

Expanding now the defect wave functions by the set $\{\phi_j^{\text{QB}}(\vec{k}, \vec{r})\}$,

$$\psi_i(\vec{r}) = \sum_{\vec{k}} \sum_{j=1}^{M_b} A_{ij}^{\text{QB}}(\vec{k}) \phi_j^{\text{QB}}(\vec{k}, \vec{r}) + \sum_{\vec{k}} \sum_{j=M_b+1}^{M_b+M_a} A_{ij}^{\text{QB}}(\vec{k}) \phi_j^{\text{QB}}(\vec{k}, \vec{r}), \quad (24)$$

we find that for the hyperdeep levels $|i\rangle = e^*$ and t_2^* , the quantities

$$\alpha_i = \sum_{\vec{k}} \sum_{j=M_b+1}^{M_b+M_a} |A_{ij}^{\text{QB}}(\vec{k})|^2$$

and

$$\beta_i = \sum_{\vec{k}} \sum_{j=1}^{M_b} |A_{ij}^{\text{QB}}(\vec{k})|^2, \quad (25)$$

representing contributions of local-band wave functions and host-band-structure wave functions, respectively, amount to $\alpha_{e^*} \cong \alpha_{t_2^*} = 0.72$;

$\beta_{e^*} \cong \beta_{t_2^*} = 0.28$, even if M_b is as large as 30.

Clearly, the contribution α_i missing in the conventional DGF method is overwhelmingly important; in its absence, an extremely slow convergence with the number of host-band wave functions is observed (cf. Tables I and II).

VI. SUMMARY AND CONCLUSION

The principle findings of this paper are the following:

(i) Whereas cluster and ligand-field-type models for localized defects in solids require that the defect wave function $\psi_i(\vec{r})$ be effectively spanned by some local basis $\{g_a(\vec{r})\}$, perturbative approaches require $\psi_i(\vec{r})$ to be represented by a superposition of the host-crystal Bloch functions $\{\phi_j^0(\vec{k}, \vec{r})\}$. In contrast, the conventional Green's-function approach to defects is derivable in terms of a dual representation; $\psi_i(\vec{r})$ is represented both by $\{g_a(\vec{r})\}$ and, *independently*, by $\{\phi_j^0(\vec{k}, \vec{r})\}$.

(ii) It is shown that the requirement for a dual representation in the DGF method often results in

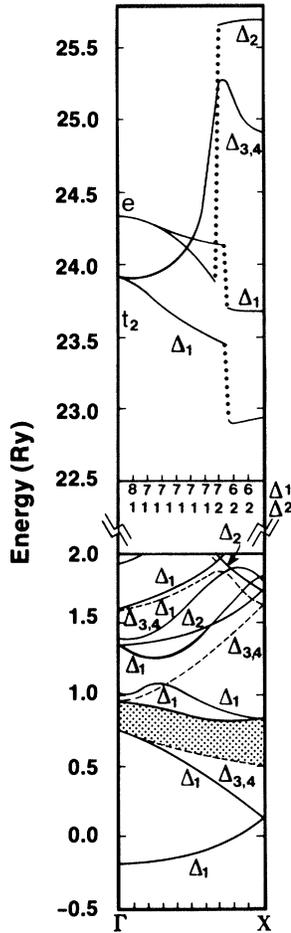


FIG. 6. The Si:Cu (substitutional) quasi band structure $\epsilon_j^{\text{QB}}(\vec{k})$ along the Δ direction in the Brillouin zone. The host bands below the energy cutoff of 2 Ry were used in this example to construct $\{\phi_j^{\text{BS}}(\vec{k}, \vec{r})\}$. The integer numbers indicate the number of such bands of symmetries Δ_1 and Δ_2 appearing below 2 Ry (the number of $\Delta_{3,4}$ levels is constant in this region). The discontinuities in the local quasi bands (upper panel) reflect the change in the number of band-structure bands of each symmetry to which the upper wave functions need to be spatially orthogonal. Note the change of energy scale above 2 Ry. The bands below 2 Ry are identical within 10^{-3} Ry to the silicon host bands calculated from plane waves alone. Combination of such bands provides a poor description of localized defect wave functions. In contrast, the quasi bands above 2 Ry are impurity related; these are very potent in accurately describing impurity wave functions.

practical application in a breakdown of the method: For sufficiently localized defect-induced perturbation potentials, an unreasonably large number of host bands is required even for a modest precision in the defect energies.

(iii) Calculation schemes that limit inherently the number of host-band wave functions available for reproducing defect levels (e.g., tight-binding models with 8 or 10 host bands^{24,31–33}) are likely to produce misleading results for deep traps. The number of host-band wave functions M_b needed to represent a defect orbital energy ϵ_i increases rapidly with the localization Δ of this state. It is therefore unlikely that the trends $\epsilon_i(\Delta, M_b)$ predicted by such methods (e.g., the trap energies ϵ_i for a series of impurities with increasing ionization energies²⁴) will be correct when ϵ_i is assumed to be M_b independent of all Δ . In the conventional DGF method, the convergence of the defect energy levels ϵ_i with the number M_i of host wave functions [Eq. (4)] can be exceedingly slow (cf. Table I). Consequently, it may be misleading to conclude that converged results have been obtained on the basis of a few calculations with small increments in the value of M_i . It is felt that $\epsilon_i(M_i)$ curves have to be published in calculations of localized defect states to demonstrate the level of convergence attained.

(iv) The fundamental limitation of the conventional DGF method is rooted in the requirement that the (orthogonal) wave functions $\{\phi_j^0(\vec{k}, \vec{r})\}$ needed to span $\psi_i(\vec{r})$ be eigenstates to the host-crystal Hamiltonian operator. While this is a sufficient condition for deriving the underlying DGF formalism, it is shown that in fact it is not a necessary condition. The far weaker condition—that an orthogonal set $\{\phi_j^{\text{QB}}(\vec{k}, \vec{r})\}$ (“quasi band wave functions”) merely diagonalizes the finite host Hamiltonian matrix—is shown to be sufficient. We show how a mixed basis set $\{\phi_j^{\text{QB}}(\vec{k}, \vec{r})\}$ of quasi band wave functions can be constructed both from pure host-band-structure functions and from localized orbitals so that the expansion of $\psi_i(\vec{r})$ in terms of $\{\phi_j^{\text{QB}}(\vec{k}, \vec{r})\}$ is very rapidly convergent. Using this approach, one is free to incorporate into the description of $\psi_i(\vec{r})$ in terms of $\{g_a(\vec{r})\}$ and $\{\phi_j^{\text{QB}}(\vec{k}, \vec{r})\}$ whatever information is available on the nature of $\psi_i(\vec{r})$ (e.g., the chemically-relevant impurity atomic orbitals, defect orbitals available from previous cluster calculations, etc.).

(v) Applications of the quasi band DGF method to two analytically solvable models: (a) a parabolic defect potential perturbation in a silicon free-electron host, and (b) a $3d$ element in the same host crystal, show excellent results relative to those obtained with the conventional DGF method. The method presented is very simple and flexible. It permits an effective application of Green’s-function techniques to arbitrarily localized defects.

ACKNOWLEDGMENTS

The authors acknowledge very helpful discussion and comments with V. Singh. Special thanks are due to P. Bendt for very helpful suggestions during the early stages of this work, as well as for providing his efficient computer subroutines for solving the generalized eigenvalue problem. Comments by J. Bernholc are gratefully acknowledged.

APPENDIX A: DERIVATION OF THE STANDARD GREEN'S-FUNCTION LOCALIZED DEFECT EQUATIONS FROM THE DUAL REPRESENTATION

In this appendix, we derive the standard defect Green's function Eqs. (8)–(10) from the dual representation in Eqs. (3) and (7). In order to solve

$$[\epsilon_j^0(\vec{k}) - \epsilon_i] A_{ij}(\vec{k}) + \sum_{j'=1}^{M_i} \sum_{\vec{k}'}^{\text{BZ}} A_{ij'}(\vec{k}') \langle \phi_j^0(\vec{k}, \vec{r}) | \Delta V | \phi_{j'}^0(\vec{k}', \vec{r}) \rangle = 0. \quad (\text{A4})$$

As will be shown in Sec. IV, the requirement that $\phi_j^0(\vec{k}, \vec{r})$ and $\epsilon_j^0(\vec{k})$ be eigenfunctions and eigenvalues, respectively, to the operator \hat{H}_0 [Eq. (A3)] is in fact unnecessarily strong; a far weaker requirement will be shown to be sufficient. However, in the derivation of this appendix we assume the former, stronger requirements [to obtain Eq. (A4)], following the standard assumptions in contemporary applications of Green's-function methods to localized defects.^{6–11}

In order to solve the basic defect problem in Eq. (A4), we can again make use of the expansion in Eq. (A2) to obtain

$$[\epsilon_j^0(\vec{k}) - \epsilon_i] A_{ij}(\vec{k}) + \langle \phi_j^0(\vec{k}, \vec{r}) | \Delta V | \psi_i \rangle = 0. \quad (\text{A5})$$

If it is assumed that the perturbation potential $\Delta V(\vec{r})$ is localized in the region $0 \leq r \leq R_c$, we can write

$$\Delta V(\vec{r}) = \Delta V(\vec{r}) \Theta(\vec{r} - \vec{R}_c), \quad (\text{A6})$$

where $\Theta(\vec{r} - \vec{R}_c)$ is a step function that projects out the subspace of the perturbation $r \leq R_c$ where $\Delta V(\vec{r})$ is nonzero.

In the subspace of the perturbation, we can make the expansion

the defect problem, i.e.,

$$[H_0 + \Delta V(\vec{r})] \psi_i(\vec{r}) = \epsilon_i \psi_i(\vec{r}), \quad (\text{A1})$$

we may expand the impurity wave function for state i in terms of the eigenfunctions $\phi_j^0(\vec{k}, \vec{r})$ of the host-crystal Hamiltonian operator \hat{H}_0 :

$$\psi_i(\vec{r}) = \sum_{j'=1}^{M_i} \sum_{\vec{k}'}^{\text{BZ}} A_{ij'}(\vec{k}') \phi_{j'}^0(\vec{k}', \vec{r}), \quad (\text{A2})$$

with

$$\hat{H}_0 \phi_j^0(\vec{k}, \vec{r}) = \epsilon_j^0(\vec{k}) \phi_j^0(\vec{k}, \vec{r}). \quad (\text{A3})$$

Substituting Eq. (A2) into (A1), multiplying to the left with $\phi_j^{0*}(\vec{k}, \vec{r})$, and integrating, we obtain the basic impurity Eq. (4):

$$\Theta(\vec{r}) \psi_i(\vec{r}) = \sum_a C_{ia} g_a(\vec{r}), \quad (\text{A7})$$

where $\{g_a(\vec{r})\}$ is some (in general nonorthogonal) basis set with an overlap matrix S with elements $S_{ab} = \langle g_a | g_b \rangle$. Substituting Eq. (A6) into Eq. (A5) gives

$$A_{ij}(\vec{k}) = \sum_a \frac{\langle \phi_j^0(\vec{k}, \vec{r}) | \Delta V | g_a \rangle}{\epsilon_i - \epsilon_j^0(\vec{k})} C_{ia}. \quad (\text{A8})$$

From Eqs. (A2) and (A7) we get

$$\sum_{j=1}^{M_i} \sum_{\vec{k}}^{\text{BZ}} A_{ij}(\vec{k}) \Theta(\vec{r}) \phi_j^0(\vec{k}, \vec{r}) = \sum_a C_{ia} g_a(\vec{r}), \quad (\text{A9})$$

so that

$$C_{ia} = \sum_{j=1}^{M_i} \sum_{\vec{k}}^{\text{BZ}} A_{ij}(\vec{k}) \times \sum_b (S^{-1})_{ab} \langle g_b | \Theta(\vec{r}) | \phi_j^0(\vec{k}, \vec{r}) \rangle. \quad (\text{A10})$$

Thus, if we multiply Eq. (A8) by $\sum_b (S^{-1})_{a'b} \langle g_b | \Theta(\vec{r}) | j\vec{k} \rangle$ and sum over all (j, \vec{k}) , $j = 1, \dots, M_i$, and $\vec{k} \in \text{BZ}$, we get after using Eq. (A10):

$$C_{ia'} = \sum_a C_{ia} \sum_b (S^{-1})_{a'b} \sum_{j=1}^{M_i} \sum_{\vec{k}}^{\text{BZ}} \frac{\langle g_b | \Theta(\vec{r}) | \phi_j^0(\vec{k}, \vec{r}) \rangle \langle \phi_j^0(\vec{k}, \vec{r}) | \Delta V | g_a \rangle}{\epsilon_i - \epsilon_j^0(\vec{k})}, \quad (\text{A11})$$

and consequently

$$\sum_a \left[\delta_{a'a} - \sum_b (S^{-1})_{a'b} \sum_{j=1}^{M_i} \sum_{\vec{k}}^{\text{BZ}} \frac{\langle g_b | \Theta(\vec{r}) | \phi_j^0(\vec{k}, \vec{r}) \rangle \langle \phi_j^0(\vec{k}, \vec{r}) | \Delta V | g_a \rangle}{\epsilon_i - \epsilon_j^0(\vec{k})} \right] C_{ia} = 0. \quad (\text{A12})$$

It should be noted that, in contrast to a similar derivation by Bassani *et al.*⁹ and Jaros and Brand,²⁷ we have not assumed formal completeness of the basis set $\{g_a(\vec{r})\}$; i.e., we have not employed so far

$$\sum_{ab} |g_a\rangle (S^{-1})_{ab} \langle g_b| = 1 \quad (\text{A13})$$

to arrive at Eq. (A12), but instead made use of the much weaker requirement in Eq. (A7). [In principle, Eq. (A7) is satisfied even by a single basis function $g_a(\vec{r}) = \psi_i(\vec{r})\Theta(\vec{r})$ which, of course, does not satisfy Eq. (A13)]. Thus, Eq. (A12) gives the exact solutions provided only that both Eqs. (A2) and (A7) are fulfilled.

It is computationally advantageous, however, to assume that we can write

$$\langle \phi_j^0(\vec{k}, \vec{r}) | \Theta(\vec{r}) \Delta V(\vec{r}) | g_a \rangle = \sum_{b''b'} \langle \phi_j^0(\vec{k}, \vec{r}) | \Theta(\vec{r}) | g_{b''} \rangle (S^{-1})_{b''b'} \langle g_{b'} | \Delta V(\vec{r}) | g_a \rangle, \quad (\text{A14})$$

where we have used the formal completeness relation and again written the impurity potential in the form $\Theta(\vec{r})\Delta V(\vec{r})$. The effect of the assumption of completeness will be discussed in Appendix B. Substitution of Eq. (A14) into (A12) gives

$$\sum_a \left[\delta_{a'a} - \sum_b \tilde{G}^0(\epsilon_i)_{a'b} V_{ba} \right] C_{ia} = 0, \quad (\text{A15})$$

where

$$\begin{aligned} \tilde{G}^0(\epsilon_i)_{a'b} &= \sum_{b''b'} (S^{-1})_{a'b''} G^0(\epsilon_i)_{b''b'} (S^{-1})_{b''b} \\ &= (S^{-1} G^0(\epsilon_i) S^{-1})_{a'b}, \end{aligned} \quad (\text{A16})$$

$$G^0(\epsilon)_{ab} = \sum_{j=1}^{M_i} \sum_{\vec{k}}^{\text{BZ}} \frac{\langle g_a | \Theta(\vec{r}) \phi_j^0(\vec{k}, \vec{r}) \rangle \langle \phi_j^0(\vec{k}, \vec{r}) | \Theta(\vec{r}) | g_b \rangle}{\epsilon - \epsilon_j^0(\vec{k})}, \quad (\text{A17})$$

and $V_{ab} = \langle g_a | \Delta V | g_b \rangle$. Equation (A15) is the standard Green's-function expression used in conventional applications for defects.

APPENDIX B: THE EFFECT OF TREATING $G^0(E)$ AND V AS TWO SEPARATE MATRICES

Appendix A indicated that it is computationally advantageous to assume formal completeness of the localized basis set $\{g_a(\vec{r})\}$ in the sense that [cf. Eq. (A14)]:

$$\langle \phi_j^0(\vec{k}, \vec{r}) | \Theta(\vec{r}) \Delta V(\vec{r}) | g_a \rangle = \sum_{b''b'} \langle \phi_j^0(\vec{k}, \vec{r}) | \Theta(\vec{r}) | g_{b''} \rangle (S^{-1})_{b''b'} \langle g_{b'} | \Delta V(\vec{r}) | g_a \rangle. \quad (\text{B1})$$

The advantage of this assumption is that the Green's-function matrix and potential matrix can be treated as two separate matrices [cf. Eqs. (A12) and (A15)], so that only the potential matrix has to be recalculated during the course of self-consistency iterations. The price for this convenience is, however, that an additional constraint is introduced: the set $\{g_a(\vec{r})\}$ needs to be complete for Eq. (B1) to be valid. This may be treated as a standard convergence problem; i.e., by either in-

creasing the size of the set $\{g_a(\vec{r})\}$ or by nonlinearly optimizing a small number of such basis functions. We have chosen the latter strategy. Table III shows the results for the Cr $3d$ level in an empty lattice host calculated with the quasi band DGF method. In this calculation the localized basis set $\{g_a(\vec{r})\}$ consisted of 19 Coulombic functions with various (nuclear charge) exponents Z^* (see Table III), augmented with the exact numerical atomic $3d$ orbitals. Furthermore, the quasi bands were

constructed from the same numerical Cr 3d wave functions. This guarantees that the dual representation of the wave function is satisfied so that, according to Appendix A and earlier discussions, the only remaining approximation is that stated in Eq. (B1). Thus, any deviation of the calculated ener-

gies from the exact numerical eigenvalues is due solely to the assumption of completeness in Eq. (B1). Table III shows that the error from this source can be made very small by optimizing Z^* : In this case, $Z^* = 12.25$ gives an error of only 0.044 eV (=0.8%).

-
- ¹H. J. Hovel, in *Semiconductors and Semimetals*, edited by R. K. Willardson and A. C. Beer (Academic, New York, 1975), Vol. II.
- ²A. G. Milnes and D. L. Feucht, *Heterojunctions and Metal-Semiconductor Junctions* (Academic, New York, 1972).
- ³A. A. Bergh and J. P. Dean, *Light Emitting Diodes* (Clarendon, Oxford, 1976).
- ⁴A. S. Grove, *Physics and Technology of Semiconductor Devices* (Wiley, New York, 1967).
- ⁵J. J. Loferski, *Solar Cells Outlook for Improved Efficiencies* (National Academy of Science, Washington, D. C., 1972).
- ⁶G. F. Koster and J. C. Slater, *Phys. Rev.* **94**, 1392 (1954); **54**, 1165 (1954).
- ⁷J. Callaway, *J. Math. Phys.* **5**, 783 (1964).
- ⁸J. Callaway and A. Hughes, *Phys. Rev.* **156**, 860 (1967); **164**, 1043 (1967).
- ⁹F. Bassani, G. Iadonisi, and B. Preziosi, *Phys. Rev.* **186**, 735 (1969).
- ¹⁰G. A. Baraff and M. Schlüter, *Phys. Rev. Lett.* **41**, 892 (1978); *Phys. Rev. B* **19**, 4965 (1979); G. A. Baraff, E. O. Kane, and M. Schlüter, *ibid.* **21**, 5662 (1980).
- ¹¹J. Bernholc, N. Lipari, and S. T. Pantelides, *Phys. Rev. Lett.* **41**, 895 (1978); *Phys. Rev. B* **21**, 3545 (1980).
- ¹²S. T. Pantelides, *Rev. Mod. Phys.* **50**, 797 (1978).
- ¹³H. Bethe, *Ann. Phys.* **3**, (1979).
- ¹⁴J. H. Van Vleck, *Phys. Rev.* **41**, 208 (1932).
- ¹⁵C. J. Ballhausen, *Ligand Field Theory* (McGraw-Hill, New York, 1962).
- ¹⁶H. Watanabe, *Operator Methods in Ligand Field Theory* (Prentice-Hall, Englewood Cliffs, 1966), p. 84.
- ¹⁷C. A. Coulson and M. J. Kearsley, *Proc. R. Soc. London Sec. A* **241**, 433 (1957).
- ¹⁸G. D. Watkins and R. P. Messmer, *Phys. Rev. Lett.* **25**, 656 (1970); *Phys. Rev. B* **7**, 2568 (1978); C. Weigel, D. Peak, J. C. Corbett, G. D. Watkins, and R. P. Messmer, *ibid.* **8**, 2906 (1973).
- ¹⁹F. P. Larkins, *J. Phys. C* **4**, 3065 (1971); **4**, 3077 (1971).
- ²⁰A. Zunger and A. Katzir, *Phys. Rev. B* **11**, 2378 (1975); A. Zunger, *J. Phys. Chem. Solids* **36**, 229 (1975).
- ²¹G. 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.
- ²²S. G. Louie, M. Schlüter, J. R. Chelikowsky, and M. L. Cohen, *Phys. Rev. B* **13**, 1654 (1976).
- ²³A. Zunger, *J. Chem. Phys.* **62**, 1861 (1975); *J. Phys. C* **7**, 96 (1974); A. Zunger and R. Englman, *Phys. Rev. B* **17**, 642 (1978).
- ²⁴H. P. Hjalmarson, P. Vogel, D. J. Wolford, and J. D. Dow, *Phys. Rev. Lett.* **44**, 810 (1980); H. P. Hjalmarson, R. E. Allen, H. Büttner, and J. D. Dow, *J. Vac. Sci. Technol.*, **17**, 993 (1980); R. E. Allen, H. P. Hjalmarson, H. Büttner, P. Vogel, D. Wolford, O. F. Sanky, and J. D. Dow, *Int. J. Quantum Chem.* **S14**, 607 (1980).
- ²⁵A. Baldereschi and N. Lipari, in *Proceedings of the Thirteenth International Conference on the Physics of Semiconductors*, edited by F. G. Fumi (Marves, Rowe, 1977), p. 595; M. Jaros, *Adv. Phys.* **29**, 409 (1980).
- ²⁶W. A. Harrison, *Pseudopotentials in the Theory of Metals* (Benjamin, New York, 1966), pp. 23, 125.
- ²⁷M. Jaros and S. Brand, *Phys. Rev. B* **14**, 4494 (1976).
- ²⁸L. Hemstreet, *Phys. Rev. B* **15**, 834 (1977); **22**, 4510 (1980).
- ²⁹A. Fazzio and J. R. Leite, *Phys. Rev. B* **21**, 4710 (1980).
- ³⁰(a) U. Lindefelt, *J. Phys. C* **11**, 3651 (1978); Corrigendum, *ibid.* **12**, 959 (1979); (b) U. Lindefelt, *J. Phys. C* **12**, L419 (1979); U. Lindefelt and S. T. Pantelides, *Solid State Commun.* **31**, 631 (1979).
- ³¹J. Bernholc and S. T. Pantelides, *Phys. Rev. B* **18**, 1780 (1978).
- ³²S. Das Sarma and A. Madhukar, *J. Vac. Sci. Technol.* **17**, 1120 (1980).
- ³³M. S. Daw and D. L. Smith, *Phys. Rev. B* **20**, 5150 (1979).
- ³⁴This is the case where the wave function of the hyperdeep level has similar dimensions as the wave function of the deep level and both have the same point-group representation (hence, orthogonality is not assured by symmetry). The latter condition is encountered very often: Donor states giving rise to an a_1 deep-gap state (e.g. N, O, F in silicon, diamond, or GaP), also have a hyperdeep state of the same symmetry; transition elements (such as Ni, Cu, Zn in silicon) often have a t_2 dangling-bond deep-gap state as well as a hyperdeep state of the same symmetry. The first condition can be visualized by considering a doughnut-shaped orbital density for the deep-gap state and a ball-shaped orbital density for the hyperdeep state.

For dangling-bond-like gap states the doughnut has dimensions characteristic of the host-bond distance whereas the hyperdeep level has the size characteristic of an impurity atomic orbital. For first-row impurity atoms in hosts made of second- or third-row elements the latter dimension is often smaller than the former dimension. On the other hand, $3d$, $4d$, and $5d$ impurity atomic orbitals have an orbital density comparable in its size to the dangling-bond radius. A correct description of the hyperdeep level is therefore essential for a valid description of the deep-gap level.

³⁵V. Singh, U. Lindefelt, and A. Zunger, Phys. Rev. B (in press).

³⁶Note that in the method of Ref. 10, the host-band structure is least-squares fitted to a limited set of localized orbitals; the fitted bands are hence not even assured to diagonalize the matrix of the host-crystal Hamiltonian [Eq. (17)]. Consequently, the fundamental Eq. (4) or Eq. (19) underlying the Green's-function formalism does not hold.

³⁷One of the common critiques of the DGF method argues that in requiring a large number M_b of host wave functions for describing localized defect states, one is in fact likely to propagate the errors associated with these numerically uncertain high conduction bands into the defect calculation. From our discussion it is clear, however, that the wave functions $\{\phi_j^{\text{BS}}(\vec{k}, \vec{r})\}$ are formally not required to converge to $\{\phi_j^0(\vec{k}, \vec{r})\}$.

³⁸R. N. Euwema, Phys. Rev. B **4**, 4332 (1971); R. A.

Deegan and W. D. Twose, Phys. Rev. **164**, 993 (1967); A. B. Kunz, J. Phys. C **3**, 1542 (1970).

³⁹Notice that if the wave vector \vec{k} in Eq. (23) is discrete, as in the case where the basic unit cell is a supercell, the sum in Eq. (23) extends only over the discrete values and equals $\sum_{\vec{R}_p} f_j(\vec{r} - \vec{R}_p)$, where \vec{R}_p are the supercell translation vectors.

⁴⁰In constructing the quasiband wave functions $\phi_j^{\text{QB}}(\vec{k}, \vec{r})$, we place the auxiliary impurity-related basis functions $f_j(\vec{r})$ in each primitive unit cell on a single impurity atomic site. The wave vector \vec{k} of $\phi_j^{\text{QB}}(\vec{k}, \vec{r})$ and the BZ sum in Eq. (10) are taken, therefore, over the regular host BZ. One could place the impurity-related basis functions $f_j(\vec{r})$ on a single site in an artificially large unit cell (e.g., super cell), or even on a site in a giant molecule with no periodic boundary conditions. Equation (23), however, indicates that irrespective of the size of the primitive lattice vector \vec{R}_p , the set $\{\chi_j(\vec{k}, \vec{r})\}$ is capable of reproducing a single local function $f_j(\vec{r})$ in each supercell. In principle, one does not have to force the localized portion $\chi_j(\vec{k}, \vec{r})$ of the quasi band $\phi_j^{\text{QB}}(\vec{k}, \vec{r})$ in Eq. (22) to be a Bloch function. In practice, however, such a choice is computationally convenient since the orthogonality condition in Eq. (17) as well as the spectral sum of Eq. (10) are more easily calculated for Bloch states.

⁴¹A Zunger and M. L. Cohen, Phys. Rev. B **18**, 5449 (1978); **20**, 4082 (1979).

⁴²U. Lindefelt and A. Zunger (unpublished).