Electronic structure of transition-atom impurities in semiconductors: Substitutional 3d impurities in silicon

Alex Zunger

Solar Energy Research Institute, Golden, Colorado 80401
and Department of Physics, University of Colorado, Boulder, Colorado 80309

U. Lindefelt

Department of Theoretical Physics, University of Lund, Sölvegatan 14A, S-223 62 Lund, Sweden
(Received 21 June 1982)

The electronic structure of neutral substitutional 3d transition-metal impurities in an infinite silicon host crystal has been calculated for the first time. The calculation is carried out self-consistently in the local-density-functional formalism to within a high precision. We use nonlocal, first-principles pseudopotentials and the recently developed quasiband crystal-field (QBCF) Green's-function method. The elements of the electronic structure of this system are discussed in detail. The calculation reveals the chemical trends in the defect energies (gap states as well as resonances) for the impurities Zn, Cu, Ni, Co, Fe, Mn, Cr, V, and Ti, as well as the regularities in the density of states, wave functions, charge distributions, and screening potentials. For charged impurities, the model explains the remarkable occurrence of many charge states in the narrow-band-gap region through a new self-regulating mechanism analogous to the homeostasis control in biological systems.

I. INTRODUCTION

A minute contamination of 1 titanium atom per 10^9 silicon atoms is sufficient to degrade substantially the efficiency of a crystalline silicon solar cell, whereas relatively huge amounts of copper contamination have almost no effect on the cell.[1] Many other transition-atom impurities have been known to considerably affect the characteristics of various optoelectric devices.[2] Among other intriguing properties of transition-atom (TA) impurities in semiconductors is the fact that as an impurity, a TA can sustain as many as five different charge states in the narrow- ($\sim$1-eV) band-gap region,[3-9] whereas the various ionized states of a free TA are spread over a range of many eV. Despite almost 30 years of research[7] on the technologically important and intellectually stimulating[8] Si:TA system, its electronic and structural properties remain unclear. In this paper we present the first theoretical study on the electronic structure of all substitutional 3d elements in an extended semiconductor host crystal. The basic elements of the electronic structure of this system as well as the various chemical trends along the 3d series are discussed.

The major obstacle in establishing a coherent experimental picture of TA impurities in silicon arises from the high diffusivity of these elements. The diffusion coefficients[5] at $\sim$1000 °C range from 10^{-4} cm²/sec for Ni (as large as typical liquid diffusivities) and decrease to 10^{-8}–10^{-10} cm²/sec for lighter impurities (Ti,V). Solutions of TA in silicon may hence be oversaturated by a few orders of magnitude and thus unstable. This has led to considerable difficulties in establishing experimentally the electronic, structural, and chemical properties of such defects. On their *electronic properties,* considerable uncertainty exists about the observed electronic activation energies of TA-defect levels in silicon. This unsettling situation is exemplified in the two most recent deep-level transient spectroscopy (DLTS) studies on 3d impurities in silicon published in 1981.[4,5] Only two out of the 35 measured transition energies agreed. A recent compilation of activation energies[6] published in 1980 contains over 20 more energy levels, almost none of which agree with the other two studies. Furthermore, a rich literature on the subject published in the Soviet literature over the past ten years (e.g., Ref. 9) reveals yet many more energy levels that are mostly unaccounted for in the studies cited previously.

On their *structural properties,* the evidence on the TA lattice site preference (substitutional versus interstitial) and aggregation state (isolated impurities versus impurity clusters) is very inconclusive. It has been known for a long term that the site preference depends on a fine control of the material preparation parameters as much as on the nature of the impurity

atom itself.[3,7] For example, a diffusion of a transition atom into *p*-type silicon near its melting temperature followed by a relatively slow quenching (a rather loosely quantified quantity[10]) often produces impurity-acceptor pairs.[3-5] A more rapid quenching, however, can also produce isolated interstitials.[3(a),3(c)] For the fast diffusers (Ni,Cu) it might be impossible to quench the sample fast enough to get stable dissolved atoms. Diffusion into *n*-type samples followed by slow quenching can produce multimers of the defect (e.g., Mn_4 clusters[11] or Fe_x clusters[12]). If the sample already contains impurities such as Cu, Ag, or Zn, a diffusion of an additional transition element can produce an impurity in a substitutional site.[3(a),3(b)] There is no evidence for any static lattice distortion or a departure from tetrahedral symmetry[3] (except, perhaps for Si:Ni [(Ref. 3(c)]). A *dynamic* Jahn-Teller distortion was, however, proposed for interstitial Cr, Mn, and Fe in silicon to explain the complete suppression of orbital contributions not accounted for by covalency effects.[3(d)] Since different measurement techniques focus on different physical properties of the impurity, it is not unusual that the same TA (e.g., Si:Co) is sometimes argued to be substitutional[13] and interstitial[3] in the same host crystal.

On their *chemical properties,* various transition-atom impurities are known to form acid-base—like complexes with dopants[3,5] and with unintentionally introduced impurities.[4] This considerably complicates the interpretation of the DLTS and EPR results. A recent systematic study[4] attributed most of the observed electron and hole traps of transition-atom—doped silicon to complexes with gold or boron in the sample. Similarly, another recent study[5] revealed defect levels that are independent of the specific transition-atom impurities diffused into the crystal. Defect levels due to transition-atom oxides (e.g., TiO_2) formed during the growth[3] and diffusion[1] of TA into semiconductors have likewise been reported.

The difficulties encountered in the experimental studies of transition-atom impurities in silicon are paralleled by the obstacles in theoretical studies on these systems.[14-22] It seems that a systematic theoretical study of a complete series of impurities (e.g., substitutional and interstitial 3*d* elements) in the same host crystal might help in organizing the large experimental data base available in terms of physically understandable regularities. It may further serve to delineate theoretically well-characterized isolated centers from defect complexes that often occur in real materials.

The existing body of experimental and theoretical results on the Si:Ta system poses a number of interesting questions. For instance, what are the basic chemical trends in the defect levels along the 3*d* series? Do all five atomic 3*d* levels appear in the semiconductor band-gap region as previously hypothesized?[3,8] Is the level ordering of *e* below t_2 indicative of substitutional defects?[3(c)] Are substitutional defects characterized by a $d \rightarrow sp$ electronic promotion?[3,4] Why are the energy levels associated with different ionization states of the impurity atom located in the narrow energy gap whereas the same ionization states in the free TA spread over a range of many eV?[4,8] Is the linear dielectric screening appropriate for shallow impurities also characteristic of localized and deep impurities? How do the states of host crystal control the position of various defect levels? How does the degree of localization of the Si:TA wave function compare with that of the more conventional defects (*s-p* defects, vacancies in semiconductors, *F* centers in ionic solids, etc.)? What is the role played by the *d* orbitals of the TA impurity in bonding to the host? Our work addresses these questions for the substitutional 3*d* impurities in silicon. It is to be complemented by a study of the interstitial 3*d* impurities.[23]

II. DEFINITION OF THE PROBLEM AND THE APPROACH

The objective of this work is to find precise self-consistent solutions to the electronic structure of isolated substitutional TA impurities in silicon within the local-density-functional formulation[24] of the interelectronic interactions. In this section we define the physical problem that we solve in terms of the effective single-particle potentials $V(\vec{r})$ and charge densities $\rho(\vec{r})$ of the unperturbed host (*H*) crystal, and the crystal containing the defect (*D*).

A. Potentials and charge densities

The single-particle potential of the pure host crystal and the impurity-containing crystal is given as a sum of the bare external potential $V^{ext}(\vec{r})$ and the screening $V^{scr}(\vec{r})$ that constitutes the electronic response to it,

$$V_H(\vec{r}) = V_H^{ext}(\vec{r}) + V^{scr}[\rho_H(\vec{r})] \;,$$
$$V_D(\vec{r}) = V_0^{ext}(\vec{r}) + V^{scr}[\rho_D(\vec{r})] \;. \tag{1}$$

The external potential is expressed as a superposition of the site-centered external potentials $v_{ps}^{(L,\beta)}(\vec{r})$, over the atomic sites at $\{\vec{R}_\beta\}$,

$$V^{ext}(\vec{r}) = \sum_{\vec{R}_\beta} \sum_L v_{ps}^{(L,\beta)}(\vec{r} - \vec{R}_\beta)\hat{P}_{L,\beta} = \sum_L V_L^{ext}(\vec{r}) \;. \tag{2}$$

In an all-electron representation, $v_{\mathrm{ps}}^{(L,\beta)}(r)$ $=-Z_\beta/r$, where Z_β is the atomic number of atom β. In a pseudopotential representation,[25] the effects of the $Z_\beta - Z_\beta^v$ core electrons are replaced by a static external potential. Hence $v_{\mathrm{ps}}^{(L,\beta)}(r)\hat{P}_{L,\beta}$ is the bare pseudopotential of atom β acting on the Lth angular momentum component of the valence wave function. The angular momentum projection operator on site β is $\hat{P}_{L,\beta}$, and Z_β^v is the number of valence (v) electrons of atom β. It includes the outer s, p, and d electrons of the atom.

The screening potential includes an interelectronic (ee) Coulomb part and an exchange-correlation (xc) contribution,

$$V_{\mathrm{scr}}[\rho(\vec{r})]=V^{ee}[\rho(\vec{r}]+V^{\mathrm{xc}}[\rho(\vec{r})] \ . \tag{3}$$

We define the effective potential perturbation $\Delta V^{\mathrm{eff}}(\vec{r})$ associated with a defect in a solid as arising from both an external perturbation and its screening perturbation

$$\Delta V^{\mathrm{eff}}(\vec{r}) \equiv \sum_L \Delta V_L^{\mathrm{ext}}(\vec{r}) + \Delta V^{\mathrm{scr}}(\vec{r}) \ . \tag{4}$$

The external potential perturbation is given as

$$\Delta V_L^{\mathrm{ext}}(r) \equiv v_{\mathrm{ps}}^{(L,D)}(\vec{r}-\vec{R}_\beta) - v_{\mathrm{ps}}^{(L,H)}(\vec{r}-\vec{R}_{\beta'}) \ , \tag{5a}$$

where it is understood that each term on the right-hand side of (5a) acts only on the L component relative to the sites β and β', respectively. For substitutional defects $\vec{R}_\beta = \vec{R}_{\beta'}$, whereas for interstitial defects at site $\vec{R}_\beta$ we have

$$\Delta V_L^{\mathrm{ext}}(\vec{r}) = v_{\mathrm{ps}}^{(L,D)}(\vec{r}-\vec{R}_\beta) \ .$$

The electronic response to this external perturbation sets up the screening fluctuation given by

$$\Delta V^{\mathrm{scr}}(\vec{r}) \equiv V_D^{\mathrm{scr}}(\vec{r}) - V_H^{\mathrm{scr}}(\vec{r}) \ . \tag{5b}$$

This screening perturbation is associated with a charge-density fluctuation $\Delta\rho(\vec{r})$ given in terms of squares of the occupied (occ) defect single-particle wave functions $\psi_i(\vec{r})$ (with occupation numbers N_i) and the occupied host-crystal Bloch wave functions $\phi_j(\vec{k},\vec{r})$ [with occupation numbers $N_j(\vec{k})$ for band j and wave vector $\vec{k}$] as

$$\Delta\rho(\vec{r}) \equiv \rho_D(\vec{r}) - \rho_H(\vec{r})$$

$$= \sum_i^{\mathrm{occ}} N_i \,|\,\psi_i(\vec{r})\,|^2$$

$$- \sum_{j,\vec{k}}^{\mathrm{occ}} N_j(\vec{k})\,|\,\phi_j(\vec{k},\vec{r})\,|^2 \ . \tag{6}$$

The basic electronic properties of the system can be obtained by self-consistently solving

$$[-\tfrac{1}{2}\nabla^2 + V_H(\vec{r}) + \Delta V^{\mathrm{eff}}(\vec{r})]\psi_i(\vec{r}) = \epsilon_i \psi_i(\vec{r}) \ . \tag{7}$$

B. Physical inputs to the problem

Other than the atomic numbers and the number of electrons in the system, there are three physical inputs to the problem:

(i) A microscopic model for describing the interelectronic exchange and correlation potential $V^{\mathrm{xc}}[\rho(\vec{r})]$ in Eq. (3). For this we use the Kohn-Sham local-density-functional formalism[24] with a Slater exchange coefficient of $\alpha_x = 1$.

(ii) The site external potentials $v_{\mathrm{ps}}^{(L,\beta)}(r)$ for the host $(\beta = H)$ and the impurity $(\beta = D)$ atoms appearing in Eq. (2). For the silicon host atom we employ the local semiempirical pseudopotential of Ref. 26. For the impurity atoms we use the first-principles nonlocal-density-functional atomic pseudopotentials of Ref. 25(a) [analytic forms are given in Ref. 25(b)]. The pseudopotential for the $L=2$ (d wave) equals, by definition,[25] the all-electron TA potential.

(iii) The atomic positions $\{\vec{R}_\beta\}$ appearing in Eq. (2) both in the host crystal and in the crystal containing a defect. For the host silicon crystal we employ its experimental crystal structure with a lattice constant $a = 10.2646$ a.u. Since experiment alone has not provided as yet any conclusive evidence on the site location of the TA in silicon, it was decided to do systematic calculations for both substitutional and interstitial[23] sites. This work reports the results for unrelaxed[3] substitutional impurities. Accordingly, $\{\vec{R}_\beta\}$ are chosen to coincide with the values of the host crystal.

Given only $v_{\mathrm{ps}}^{(\beta)}(r)$ and $\{\vec{R}_\beta\}$ for the host crystal, one can calculate the host charge density $\rho_H(r)$, the self-consistent screening potential $V_H^{\mathrm{scr}}(\vec{r})$, the band structure $\{\epsilon_j^0(\vec{k})\}$, and the host wave functions $\{\phi_j(\vec{k},\vec{r})\}$ in a standard manner. Table I depicts the calculated self-consistent silicon band structure using the exchange coefficient $\alpha_x = 1$. The calculated single-particle energies compare reasonably well with the observed excitation energies.[27-30] The same pseudopotential has been used in Ref. 31 to calculate the electronic structure of a silicon vacancy (with a similar but not identical pseudopotential used in Ref. 32); however, the exchange coefficient used in these calculations was $\alpha_x = 0.8$. A comparison of the $\alpha_x = 1$ and 0.8 results for bulk silicon (Table I) shows the two to be similar for the valence-band states; however, the $\alpha_x = 0.8$ results produce a fundamental band gap that is about 50% smaller than the observed band gap. A similar un-

TABLE I. Calculated self-consistent single-particle band structure of silicon with the semiempirical local pseudopotential of Ref. 26, employing an exchange coefficient $\alpha_x = 1$. For comparison, we give also the results for $\alpha_x = 0.8$ and experimental data (Refs. 27–30). The zero of energy is set at the valence-band maximum. The unit-cell volume used is twice the atomic volume 135.188 a.u.$^{-3}$. 181 plane waves are used at the Γ point and a comparable number at other $\vec{k}$ points. The self-consistent screening is constructed from the lowest 1471 plane waves. Four special $\vec{k}$ points are used to obtain the ground-state charge density. All components of the pseudopotential up to a momentum $q = 6$ a.u.$^{-1}$ are included.

Level	$\alpha_x = 0.8$	$\alpha_x = 1$	Experimental
$\Gamma_{1,v}$	-12.8	-12.5	-12.5 ± 0.6[a]
$\Gamma_{25',v}$	0.0	0.0	
$\Gamma_{15',c}$	2.88	3.20	3.2[b]
$\Gamma_{2'c}$	3.49	3.78	4.15 ± 0.05[c]
Γ_{12}		8.75	8.3 ± 0.1[c]
$X_{1,v}$	-8.48	-8.32	
$X_{4,v}$	-3.14	-2.98	-2.9[c]
$X_{1,c}$	0.78	1.39	
Indirect gap	0.60	1.21	1.15[d]
$L_{2',v}$	-10.39	-10.21	-9.3 ± 0.4[a]
$L_{1,v}$	-7.52	-7.24	-6.8 ± 0.2[a]
$L_{3,v}$	-1.36	-1.30	-1.2 ± 0.2[c]
$L_{1,c}$	1.60	2.02	2.02 ± 0.2[c]
$L_{3,c}$	3.69	4.11	3.9 ± 0.1[c]

[a]Reference 27.   [c]Reference 29.
[b]Reference 28.   [d]Reference 30.

derestimation of the gap occurs when first-principles nonlocal pseudopotentials are used,[25(a)] or when an all-electron local-density calculation is done. Although the discrepancy is not large on an absolute energy scale, it is felt that a 0.6-eV band gap may cloud the interpretation of the results of defect calculations in which the location of defect levels with respect to band edges is very important. We therefore use $\alpha_x = 1$ in all calculations reported here.

C. Statement of the problem

In this paper we address the problem of establishing a self-consistent solution for Eqs. (1)—(7) for $3d$ impurity atoms in silicon within the local-density formalism, given only the silicon and impurity-atom pseudopotentials $v_{ps}^{(L,\beta)}(r)$ and the atomic positions $\{\vec{R}_\beta\}$ as input. The aspects of the electronic structure in which we are most interested are the chemical trends in the single-particle energies $\{\epsilon_i\}$ and their local density of states, the wave functions $\{\psi_i(\vec{r})\}$, as well as the charge density $\rho_D(\vec{r})$, screening $V_D^{scr}(\vec{r})$, and their deformations $\Delta\rho(\vec{r})$ and $\Delta V^{scr}(\vec{r})$, respectively. Our objective is to obtain

predictions that reflect the underlying local-density theory (not the computational approximations to it, such as finite clusters, muffin-tin approximations, semiempirical tight-binding parametrizations, etc.) to a level of precision of about 0.1 eV in energies and 1—2 % in charge density and potentials. A new computational scheme[33,34] recently developed by us enables one to achieve this objective very efficiently for an entire group of impurity atoms (e.g., the $3d$ row) with only modest investment of computer time.

D. Approach

We use the quasiband crystal-field (QBCF) self-consistent Green's-function method. The method was described in detail previously.[34] For our present discussion we need only to describe the wave-function, potential, and density representations. These definitions will aid the analysis of the results (Sec. III).

1. Wave-function representations

We employ a dual representation of the defect wave functions $\{\psi_i(\vec{r})\}$. The first representation is in terms of an impurity-centered local-orbital basis $\{g_a(\vec{r})\}$. The λth partner of the αth irreducible representation of the wave function is given as

$$\psi_i^{\alpha\lambda}(\vec{r}) = \sum_a C_{ai} g_a(\vec{r}) \ . \tag{8a}$$

Since all basis functions have a unique origin, one can separate the variables in $g_a(\vec{r})$ in a crystal-field manner into a product of radial orbitals $F_{\mu l}(|\vec{r}|)$ (of principal quantum number μ and angular momentum l) and Kubic harmonics $K_l^{\alpha\lambda}(\hat{r})$, transforming according to the λth partner in the αth irreducible representation and depending on the direction of $\vec{r}$ alone:

$$\psi_i^{\alpha\lambda}(\vec{r}) = \sum_l^{l_{max}} \sum_\mu^N C_{\mu li}^\alpha F_{\mu l}(|\vec{r}|) K_l^{\alpha\lambda}(\hat{r}) \ . \tag{8b}$$

Here, $\{C_{\mu li}^\alpha\}$ are the expansion coefficients to be determined for each defect level $|i\alpha\lambda\rangle$. The local representation in Eq. (8a) can be further written as

$$\psi_i^{\alpha\lambda}(\vec{r}) = \sum_l^{l_{max}} G_{li}^\alpha(|\vec{r}|) K_l^{\alpha\lambda}(\hat{r}) \ , \tag{8c}$$

where $\{G_{li}^\alpha(|\vec{r}|)\}$, defined by Eqs. (8a) and (8b), are the independent angular degrees of freedom of the single-particle state $|i\alpha\lambda\rangle$. They reflect the $s,p,d,\ldots$ content of each state.

The second representation of the defect wave functions is given in terms of Bloch quasiband (QB) wave functions[33]:

$$\psi_i^{\alpha\lambda}(\vec{r}) = \sum_j^{\overline{M}} \sum_{\vec{k}}^{N_{\vec{k}}} A_{ij}^{\alpha\lambda}(\vec{k})\phi_j^{QB}(\vec{k},\vec{r}) \ . \tag{9}$$

Here, $\{A_{ij}^{\alpha\lambda}(\vec{k})\}$ are variational coefficients [related to $\{C_{\mu li}^{\alpha}\}$ in Eq. (8b)], whereas $\{\phi_j^{QB}(\vec{k},\vec{r})\}$ are essentially identical[33] to the host-crystal Bloch wave functions for the lowest $j \leq M_b$ bands, and are equal to the renormalized impuritylike (quasiband) wave functions for the higher bands $j = M_b + 1$ to $j = \overline{M}$, where $\overline{M} = M_a + M_b$. The coefficients $\{A_{ij}^{\alpha\lambda}(\vec{k})\}$ for the N_k wave functions used are the independent spectral degrees of freedom of the single-particle state $|i\alpha\lambda\rangle$. They reflect the contribution of each host-crystal state $|\vec{k}j\rangle$ to the defect wave function $\psi_i^{\alpha\lambda}(\vec{r})$.

The conventional expansion of the defect wave functions by pure host wave functions introduced by Koster and Slater[35] was recently shown to lead to an intolerably slow convergence in Eq. (9) (requiring $M_b = 10^2 - 10^4$ host bands) if the defect is chemically or physically sufficiently different from the host.[33] This is merely a statement of the fact that the high-energy states of the host crystal do not form a compact basis for describing a fundamentally new atomic state (3d). Our reformulation of the Koster-Slater resolvent in terms of quasiband wave functions $\{\phi_j^{QB}(\vec{k},\vec{r})\}$ instead of pure host wave functions $\{\phi_j(\vec{k},\vec{r})\}$ incorporates both hostlike and impuritylike characteristics directly in the basis $\{\phi_j^{QB}(\vec{k},\vec{r})\}$. This is achieved by renormalizing a very large number of host-crystal conduction bands, which would have been needed for an adequate expansion of localized defect states, into a much smaller number ($M_a = 5$) of local quasiband wave functions. In this representation the expansion in Eq. (9) is rapidly convergent even for arbitrary localized perturbations.[33]

2. Potential and density representations

We use a crystal-field-like factorization of the density and potential perturbations into radial and angular components:

$$\Delta\rho(\vec{r}) = \sum_{l=0}^{2l_{\max}} \Delta\rho_l(|\vec{r}|)K_l^{a_1}(\hat{r}) \ , \tag{10a}$$

$$\Delta V^{scr}(\vec{r}) = \sum_{l=0}^{2l_{\max}} \Delta V_l^{scr}(|\vec{r}|)|K_l^{a_1}(\hat{r}) \ . \tag{10b}$$

We project out the a_1 totally symmetric part of $\rho(\vec{r})$, and hence the a_1 Kubic harmonic is used. Note, however, that although these quantities are totally symmetric in the point-group representation, they are not spherically symmetric (as assumed in

the muffin-tin model[19–21]). The coefficients $\Delta\rho_l$ and ΔV_l^{scr} for $l \neq 0$ measure their degree of nonsphericity.

3. Computational input

There are four groups of computational parameters appearing in the QBCF method: (i) The number N and type of radial orbitals $\{F_{\mu l}(|\vec{r}|)\}$ with maximum angular momentum $l_{\max}$ used in the expansion (8). (ii) The number $\overline{M} = M_b + M_a$ of quasiband wave functions used in the expansion of Eq. (9). This includes M_b hostlike bands and M_a local quasibands. (iii) The number $N_{\vec{k}}$ of $\vec{k}$ points used in the calculation of the Green's function[34] (isomorphous with a supercell size). (iv) The tolerance

$$\sigma \leq \frac{1}{n}\sum_{i=1}^{n} [(\Delta V_i^{scr})_{output} - (\Delta V_i^{scr})_{input}]$$

on self-consistency,[36] where i indexes radial mesh points (extending to $R \leq 11$ a.u.).

We have searched the parameter space $\{N, l_{\max}, M_a, M_b, N_k, \sigma\}$ for values that meet our objective of $a \sim 0.1$ eV precision in energies and a few percentages in densities and potentials. Extensive convergence tests were reported previously.[34] We have found an optimal parameter set that achieves this goal for all 3d impurities in silicon using limited computational effort.

We use $l_{\max} = 4$ and $N = 47$. The local basis orbitals are distributed among the various l components as follows: 11 s orbitals, 10 p orbitals, 10 d orbitals, 8 f orbitals, and 8 g orbitals. Counting both angular and radial components, we have $11 + 30 + 50 + 56 + 72$ orbitals for $l = 0$, 1, 2, 3, and 4, respectively, or a total of 219 local basis functions. For $\{F_{\mu l}\}$ we use Coulombic orbitals with an optimized effective charge $Z^* = 10$. The local-orbital basis is augmented by the exact local-density (numerical) 3d orbital of the free-space transition-atom impurity, calculated self-consistently with the same exchange-correlation functional used for the defect calculation. All local basis functions are orthogonalized within a sphere of radius 7.5 a.u. by the Gram-Schmidt procedure. For the spectral representation we use an average of $M_b = 31$ hostlike bands per $\vec{k}$ point and $M_a = 5$ local quasibands. For $N_{\vec{k}}$ we use 10 $\vec{k}$ points (isomorphous with a supercell of 250 atoms) to obtain self-consistency; the self-consistent potential is then applied to a calculation with $N_{\vec{k}} = 20$ (isomorphous to a supercell containing 686 atoms) for testing its accuracy. The self-consistency tolerance σ is set at about 20 mRy.

To test the internal precision of the method, we apply the "atoms-in-vacuum test" which uses the Green's-function method to solve for the electronic

structure of isolated atoms. This is done by defining in the host crystal a region $\vec{r} \leq R_c$ where both $V_H^{\text{ext}}(\vec{r})$ and $V_H^{\text{scr}}(\vec{r})$ are removed (hence the potential there equals the constant host vacuum level). We have placed in this vacuum the screened pseudopotential $[v_{\text{ps}}^{L,\beta}(r)+v^{\text{scr}}(r)]$ of a free transition atom. We then solve the Green's-function problem for this system. If R_c is large enough to accommodate the atomic orbitals, the resulting QBCF energies and wave functions should equal the self-consistent atomic solutions obtained independently from a direct numerical integration of the atomic problem. For all transition atoms described in this work, we find that the QBCF method produces atomic $3d$ energies and wave functions with a precision of 0.03 eV and 1%, respectively.

Complete details of the computational scheme are described in Ref. 34. With this detailed outline, the interested reader can independently reproduce this method and its results.

E. Quantities used for the analysis of results

In this section we define some auxiliary quantities used to analyze the results of the present calculation.

1. Wave-function population analyses

The degree of spatial localization of a state $|i\alpha\lambda\rangle$ within the central-cell region $(r \leq R_{\text{cc}})$ is defined from Eq. (8c) as

$$q_i^\alpha = \sum_l^{l_{\text{max}}} \int_0^{R_{\text{cc}}} |G_{li}^\alpha(|\vec{r}|)|^2 r^2 dr \equiv \sum_l q_{li}^\alpha . \quad (11)$$

Here, the "orbital density localization parameter" q_i^α is the amount of electronic charge enclosed in a sphere of radius R_{cc} in state $|i\alpha\lambda\rangle$ for each electron occupying this state (i.e., normalized to 1). The "percentage of localization" of state $|i\alpha\lambda\rangle$ is then $100q_i^\alpha$, since $q_i^\alpha \leq 1$. The value of q_i^α is used to delineate localized defect states from extended states. Its resolution into angular components q_{li}^α in Eq. (11) provides a measure to the hybridization of this state. The "percentage of l character" in state $|i\alpha\lambda\rangle$ is given as $100q_{li}^\alpha/q_i^\alpha$. If an a_1, t_2, or e state was a pure s, p, or d state, respectively, as often assumed,[37] then the minority character in q_{li}^α would be very small. As we will see, this is often not the case. Notice that a number of q_{li}^α components are zero by symmetry, e.g. (suppressing the index i), $q_1^{a_1} = q_2^{a_1} = q_0^e = q_1^e = q_0^{t_1} = q_1^{t_1} = q_2^{t_1} = q_0^{t_2} = 0$ for T_d symmetry. Further, the orthonormality of the representation in Eq. (8c) and its single-centered character precludes ambiguity in the way one partitions a "bond charge" into its constituent atoms, as there is in a

Mulliken population analysis.[38] In all calculations reported here, we define the central-cell region in the conventional way to include the spherical space between the substitutional impurity atom and its four nearest neighbors, hence $R_{\text{cc}} = 4.44$ a.u. for silicon.

2. Local density of states

The impurity-centered local density of single-particle states belonging to the αth representation is defined as

$$D^\alpha(\epsilon) = \sum_i \delta(\epsilon - \epsilon_i^\alpha)q_i^\alpha . \quad (12)$$

This quantity describes the energy distribution of states belonging to the αth representation inside the central cell. For $R_{\text{cc}} \to \infty$, this reduces to the conventional (integrated) density of states for representation α.

3. Spectral analysis of wave functions

The resolution of the defect wave function in Eq. (9) into Bloch wave functions allows one to determine which states in the unperturbed host crystal significantly contribute to synthesizing a given defect wave function. Hence the "percentage of valence-band (VB) character" of a defect state $|i\alpha\lambda\rangle$ is given by

$$100 \sum_j^{\text{VB}} \sum_{\vec{k}}^{\text{IBZ}} |A_{ij}^{\alpha\lambda}(\vec{k})|^2 d_{\vec{k}} , \quad (13)$$

where $d_{\vec{k}}$ is the number of vectors in the star of $\vec{k}$, and IBZ indicates that the summation is carried out over the irreducible Brillouin zone. The quantity $|A_{ij}^{\alpha\lambda}(\vec{k})|^2$ measures the fraction of the norm of the defect state $|i\alpha\lambda\rangle$ contributed by the unperturbed host Bloch function $|\vec{k}j\rangle$. This will be used to identify special host-band wave functions that have dominant contributions to defect states.

Since the quasiband wave functions in Eq. (9) are constructed from M_b hostlike [band-structure (BS)] bands and M_a local quasibands (LQB), one can further decompose the norm of a defect state $|i\alpha\lambda\rangle$ into a fraction $a_{\text{BS}}^{i\alpha\lambda}$, contributed by the lowest M_b host wave functions, and a fraction $a_{\text{LQB}}^{i\alpha\lambda} = 1 - a_{\text{BS}}^{i\alpha\lambda}$ contributed by the M_a local quasibands (which effectively contain the contributions of the next-highest approximately 10^3–10^4 conduction bands):

$$\int [\psi_i^{\alpha\lambda}(\vec{r})]^2 d\vec{r} = 1 = \sum_{j=1}^{M_b} \sum_{\vec{k}}^{\text{IBZ}} [A_{ij}^{\alpha\lambda}(\vec{k})]^2 d_{\vec{k}}$$
$$+ \sum_{j=1}^{M_a} \sum_{\vec{k}}^{\text{IBZ}} [A_{ij}^{\alpha\lambda}(\vec{k})]^2 d_{\vec{k}}$$
$$\equiv a_{\text{BS}}^{i\alpha\lambda} + a_{\text{LQB}}^{i\alpha\lambda} . \quad (14)$$

If a defect state could have been constructed solely from the M_b pure-host-crystal-band wave functions, then $a_{BS}^{i\alpha\lambda} = 1$, whereas if the local quasibands are needed, then $a_{LQB}^{i\alpha\lambda}$ would be sizable. Recall that in the effective-mass approximation[39] $M_b = 1$ for donor state ($M_b = 3$ for acceptor states), $a_{BS}^{i\alpha\lambda} = 1$, and $a_{LQB}^{i\alpha\lambda} = 0$, and that in tight-binding models[37] $M_b = 8 - 10$ and $a_{LQB}^{i\alpha\lambda} = 0$.

4. Charge-density population analysis

It will be of interest to us to establish the total amount of electronic charge in the central-cell region and its distribution among various l components and representations α. We will hence define Q_l^α to be the electronic charge contribution within the central-cell region from the lth component of all occupied states belonging to the αth representation:

$$Q_l^\alpha = \sum_i^{occ} N_i \int_0^{R_{cc}} [G_{li}^\alpha(\vec{r})]^2 r^2 dr . \tag{15}$$

We refer to Q_l^α as the "orbital-representation charge." This will serve to establish "how p-like all t_2 states are," or "how much f character exists in all a_1 states," etc.

The total charge Q_l below the Fermi energy arising from the various l components in the wave functions will be defined as

$$Q_l = \sum_\alpha Q_l^\alpha . \tag{16}$$

We refer to Q_l as the "orbital occupation" and use it to judge questions such as "to what extent is a transition-atom impurity in silicon d-like or p-like?"

The distribution of the total central-cell charge among the various allowed representations will be given by the "representation occupation" Q^α,

$$Q^\alpha = \sum_l Q_l^\alpha , \tag{17}$$

which measures the contribution to the central-cell charge from occupied states belonging to the irreducible representations a_1, t_2, e, t_1. Finally, the total electronic charge enclosed in the central-cell region is given by

$$Q^{tot} = \sum_l Q_l = \sum_\alpha Q^\alpha . \tag{18}$$

5. The effective impurity atom

Q_l^α, Q_l, and Q^α will include contributions both from the impurity atom and from the host-crystal orbitals penetrating the central cell. It is instructive to define an "effective impurity atom" that best describes the properties of an "isolated" defect atom

properly renormalized by its self-consistent interactions with the host. This requires the definition of an arbitrary reference state that represents the host before a substitutional defect is introduced into it. For this reference state we select the self-consistent silicon-vacancy system calculated in Ref. 34. We conceptualize that first one prepares a vacancy and then drops into it a noninteracting effective impurity atom that produces an overall electronic distribution which is identical to that of the interacting system. The effective impurity atom may be different in its electronic structure from the isolated free-space impurity atom; that is, it may have nonspherical components in its charge density (even if the free-space atom is assumed spherical), it may be compressed, it may have a different orbital configuration, etc. We define the representation occupation of the effective impurity atom as

$$\Delta Q^\alpha = (Q^\alpha)_{impurity} - (Q^\alpha)_{vacancy} , \tag{19a}$$

its orbital occupation as

$$\Delta Q_l = (Q_l)_{impurity} - (Q_l)_{vacancy} , \tag{19b}$$

its total net electronic charge as

$$\Delta Q^{tot} = (Q^{tot})_{impurity} - (Q^{tot})_{vacancy} , \tag{19c}$$

and its charge-density components as

$$\delta_l(r) = (\Delta\rho_l)_{impurity} - (\Delta\rho_l)_{vacancy} , \tag{19d}$$

etc. Each of the two terms in Eqs. (19) is calculated separately in a self-consistent manner. Notice that since at a large distance from the impurity site the charge density of both the impurity-containing system and the vacancy are equal (to the unperturbed charge), the effective net charges in Eqs. (19) do not depend on the radius R_{cc} used in the integration [cf. Eq. (15)], provided R_{cc} is large enough. In practice, the effective charges $\Delta Q(R)$ do not change past $R \geq 7.5$ a.u. Furthermore, when a large interatomic separation is used for the host crystal, these charges reduce to the charge of a free TA. These charges are different from those used in muffin-tin calculations,[19-21] which provide instead Q^{tot} [Eq. (18)], with a cutoff radius R_{cc} [cf. Eq. (15)] equaling the muffin-tin sphere radius ($R = 2.22$ a.u.). The latter charge includes significant contributions both from the impurity atom and from the host orbitals penetrating the muffin-tin spheres, and are hence not meaningful "impurity atomic charges." It often leads to the conclusion that the impurity atom is negatively charged, whereas much of the extra
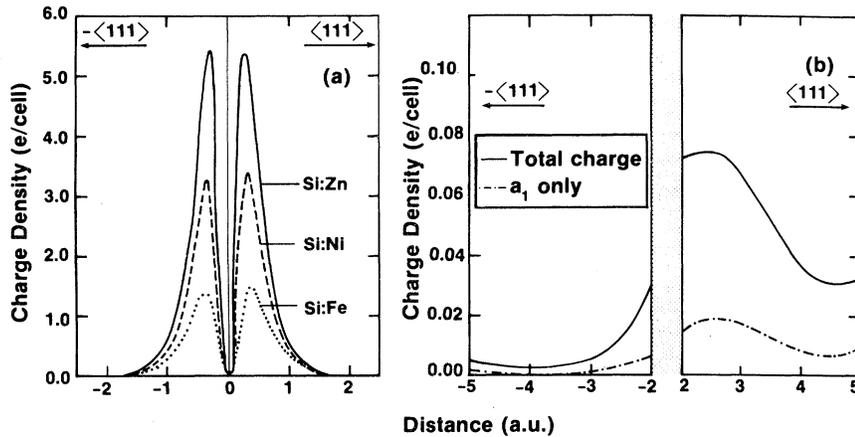
FIG. 1. Self-consistent charge density $\rho_D(\vec{r})$ for some substitutional $3d$ impurities in silicon in the $\pm\langle 111\rangle$ crystal directions. (b) shows, on an expanded scale, the total charge density (solid line) and the charge density from the a_1 states (dashed-dotted line) in the outer central-cell region. The nearest-neighbor silicon atom is located at $r=4.44$ a.u. in the $\langle 111\rangle$ direction.

charge corresponds to host states having nonzero amplitude on the impurity site. The quantities in Eq. (19) will be used to examine charge-redistribution effects. For instance, a ground-state Fe atom has a configuration $s^2p^0d^6$ [or $a_1^2(et_2)^6$ in a T_d classification]. When placed substitutionally in a crystal its effective orbital configuration will change to ΔQ_s s electrons, ΔQ_p p electrons, and ΔQ_d d electrons, its representation occupation will be $\Delta Q^{a_1}\Delta Q^e\Delta Q^{t_1}\Delta Q^{t_2}$, and it will carry a net electronic charge of ΔQ^{tot}. These are used to establish the bonding patterns, ionicity, and electronic rearrangements attendant upon impurity formation.

III. RESULTS

A. Charge densities

Figure 1(a) shows the self-consistent impurity charge density $\rho_D(\vec{r})$ [first term in Eq. (6)] of a few substitutional transition-atom impurities in silicon along the $\pm\langle 111\rangle$ crystal directions. Figure 1(b) shows, on an enlarged scale, the tails of $\rho_D(\vec{r})$ for $r\geq 2$ a.u.; the results in this range are similar for all impurities. It is seen that in the inner central-cell region [$r\leq 1.5$ a.u., Fig. 1(a)] the charge density resembles an atomic $3d$ charge density and has a very large amplitude (the maximum valence charge density of crystalline silicon is, in the same units, only $\leq 0.1e/$cell). In the outer central-cell region [$2<r<4.44$ a.u., Fig. 1(b)], on the other hand, the charge density has dropped by nearly 2 orders of magnitude and exhibits a strong asymmetric behavior, characteristic of the host-crystal charge density. It is constructed mostly from t_2+e host

states with a smaller contribution from the a_1 states. The strong valence-electron density gradient within the central cell distinguishes TA impurities from all s-p impurities that have a reasonably slowly varying central-cell charge density.[34,31,32]

Figure 2 depicts the angular momentum components $\Delta\rho_l(\,|\,\vec{r}\,|\,)$ [Eq. (10a)] of the charge-density fluctuations. First, the spherical ($l=0$) components are shown [Fig. 2(a)], indicating a monotonic decrease in the amplitude as the atomic number of the impurity decreases and a very strong localization in the inner central-cell region. Compared with the density fluctuation for a silicon vacancy[34] [Fig. 2(c)] having an overall dimension of 1–2 host bond lengths, the density fluctuation associated with a $3d$ impurity has a much smaller characteristic radius, i.e., of the order of an atomic core size. Second, we display in Fig. 2(b), on an expanded scale, the different l components of $\Delta\rho(\vec{r})$ for the Si:Cu system (typical of all other $3d$ impurities). The spherical ($l=0$) component dominates, whereas the nonspherical $l=3$ and 4 ("non-muffin-tin") terms are ~ 100 times smaller. The charge density $\rho_D(\vec{r})$ of the defect-containing system, as well as $\rho_H(\vec{r})$ of the unperturbed host crystal, is strongly nonspherical and extended in open-structure covalent systems.[40] The density fluctuation $\Delta\rho(\vec{r})$, however, is only weakly anisotropic and localized in the central-cell region, due to an effective cancellation. This highlights the reasonableness of a crystal-field expansion [viz., Eq. (10)] of the perturbation.

The spherical component $\Delta\rho_0(\,|\,\vec{r}\,|\,)$ of the density fluctuation possesses the same overall characteristic form for all substitutional $3d$ impurities: It has a positive maximum $\Delta\rho^{(1)}$ at $r=R_1$, followed by a negative minimum $\Delta\rho^{(2)}$ at R_2. The minimum is
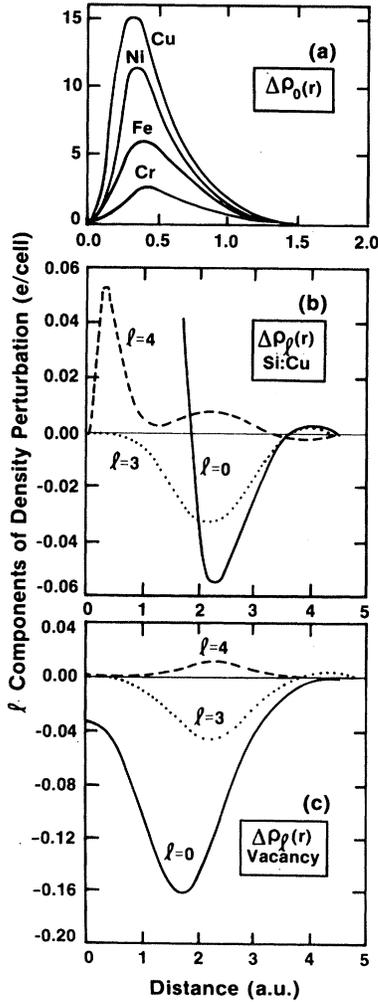
FIG. 2. Angular momentum decomposition of the charge-density fluctuation $\Delta\rho(\vec{r})$ [Eq. (10a)] for (a) substitutional impurities in silicon, and (b) for the silicon vacancy (c). Note the change in scale from (a) to (b).

bracketed by two nodes at R_{N1} and R_{N2}. A secondary, smaller positive maximum $\Delta\rho^{(3)}$ occurs at R_3. The positions R_1 and R_2 of the major extrema increase monotonically as one moves backward in the $3d$ series from Zn to Ti: R_1 varies from 0.30 to 0.45 a.u., and R_2 from 2.2 to 2.4 a.u. Following this trend, $\Delta\rho^{(1)}$ decreases rapidly from $19.3e$/cell in Si:Zn to $2.6e$/cell for Si:Cr (reflecting the delocalization of the atomiclike $3d$ orbitals with reducing atomic number). It is instructive to examine the amount of charge enclosed within a sphere of a given radius. Figure 3 depicts the total electronic charge $\Delta(R) = \int_0^R \Delta\rho(\vec{r})d\vec{r}$ in a sphere of radius R from the origin. The remarkable result of this figure is that within $R \sim 4$ a.u. the central cell is electrically neutral [i.e., $\Delta(R_{cc}) \cong \Delta Z^v$] regardless of

the number of impurity electrons, $4 \le Z_I^y \le 12$, used to populate the energy levels. A similar behavior was noted by Dimmock and Hemstreet[19(b)] for Cr in GaAs. As will be demonstrated later [Fig. 11(a)], individual wave functions associated primarily with the TA impurity extend well outside the central cell, yet by Fig. 3 the central cell preserves its electrical neutrality. This "electronic elasticity" of the central cell is facilitated by the readjustment of the hostlike state in response to the external perturbation $\Delta V_D^{ext}(\vec{r})$, and will be discussed in some detail in Sec. III F.

Figure 3 shows that the screening is effective over a large range of external perturbations and that it occurs already within a relatively short distance $r < R_{cc}$. Notice that $\Delta(R)$ is nonmonotonic: Its maximum occurs at $R \sim R_{N1}$, and its value at the maximum considerably exceeds the asymptotic limit $\Delta(\infty) = Z_I^v - Z_{Si}^v = \Delta Z^v$. Hence the position of the first node of $\Delta\rho(r)$ at R_{N1} controls the maximum available screening charge. The consequences of this "screening overshoot" will be discussed further in the next section. As a result of a negative minimum $\Delta\rho^{(2)}$, the charge $\Delta(R)$ decreases for $R_{N1} < r < R_{N2}$ and eventually saturates at the value of the point-charge limit of ΔZ^v. In contrast to the behavior of $3d$ impurities, the silicon vacancy digs a broad structureless hole in the charge (Fig. 3).
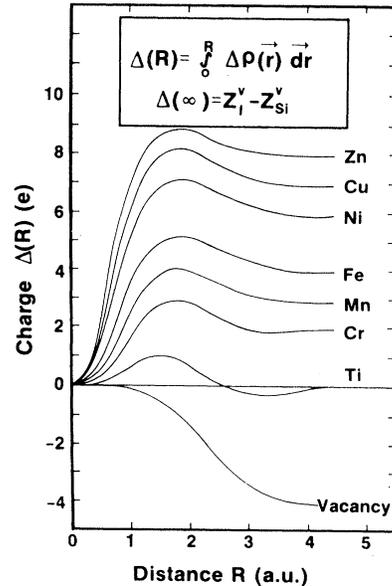
FIG. 3. Total charge $\Delta(R)$ enclosed within a sphere of radius R from the impurity for substitutional $3d$ TA in silicon and for the silicon vacancy. The nearest-neighbor Si atom is located at $R = 4.44$ a.u.
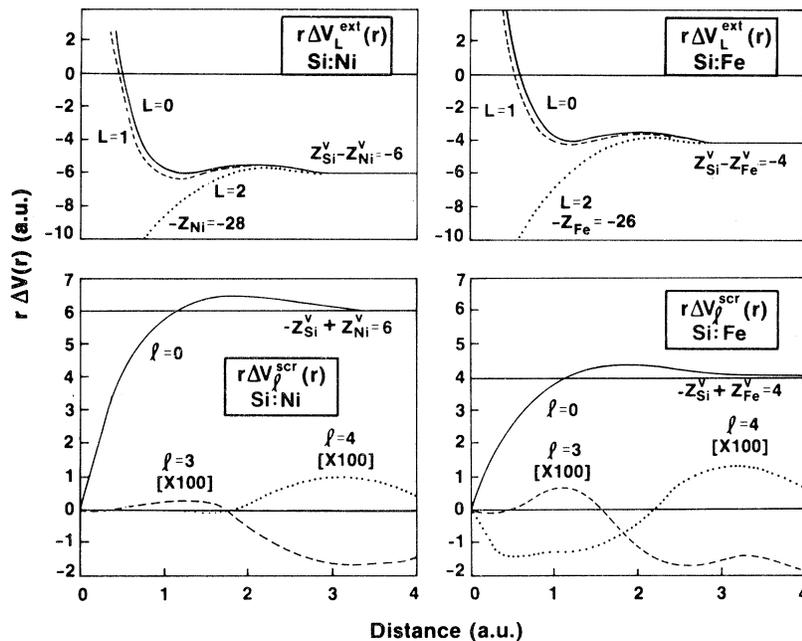
FIG. 4. External pseudopotential perturbation $r \, \Delta V_L^{\text{ext}}(r)$ [Eq. (5a)] for $L = 0$, 1, and 2 and the angular momentum components of the self-consistent screening $r \, \Delta V_l^{\text{scr}}(r)$ [Eq. (10b)] for Si:Ni and Si:Fe. The asymptotic point-ion limits are indicated at the right; the $r = 0$ limit of $r \, \Delta V_2^{\text{ext}}(r)$ is at the left.
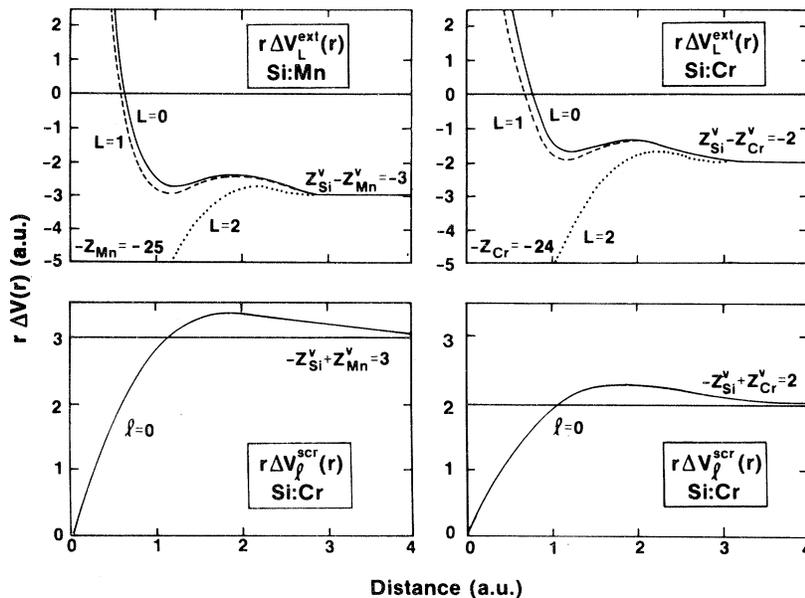
FIG. 5. Component of the external pseudopotential perturbation $r \, \Delta V_L^{\text{ext}}(r)$ and self-consistent screening $r \, \Delta V_0^{\text{scr}}(r)$ for Si:Mn and Si:Cr. The asymptotic point-ion limits are indicated at the right as $-Z_H^V + Z_D^V$; the $r = 0$ limit of $r \, \Delta V_2^{\text{ext}}(r)$ is at the left.

B. Effective potentials

Figures 4 and 5 depict the self-consistent screening perturbation $\Delta V^{\mathrm{scr}}(\vec{r})$ [Eqs. (10b)], together with the input external potential perturbation $\Delta V_L^{\mathrm{ext}}$ [Eq. (5a)] that produces it. Actually plotted are the products $r \Delta V_L^{\mathrm{ext}}(r)$ and $r \Delta V_l^{\mathrm{scr}}(r)$, which correspond to r-dependent effective charges with point-ion limits of

$$\lim_{r \to \infty} [r \Delta V_L^{\mathrm{ext}}(r)] \equiv -\Delta Z^v$$

and

$$\lim_{r \to \infty} [r \Delta V_0^{\mathrm{scr}}(r)] \equiv \Delta Z^v \; .$$

The pseudopotential perturbation (upper panels) is spherically symmetric and includes three nonlocal components: $L=0$, denoted by a solid line (experienced by s states); $L=1$, denoted by a dashed line (experienced by p states), and $L=2$, denoted by a dotted line (experienced by d, f, and g states). The difference between the three curves reflects the angular nonlocality of the impurity atomic pseudopotential. Note that since the valence $3d$ orbital has no core state of the same symmetry (i.e., $2d$), no pseudopotential cancellation occurs for $L=2$. This pseudopotential equals the (unpseudized) full all-electron Kohn-Sham potential, with

$$\lim_{r \to \infty} (r v_{\mathrm{ps}}^{(2,\beta)}) = -Z_\beta$$

(a value indicated in Figs. 4 and 5 next to the $L=2$ line). The order of the $\Delta V^{\mathrm{ext}}(r)$ curves for different elements reflects their atomic properties.[25] As the atomic number Z_β increases in going from Cr to Mn, Fe, and Ni, the minima in $\Delta V_L^{\mathrm{ext}}(r)$ deepen and the crossing points $R_{L=0}$ at which $\Delta V_L^{\mathrm{ext}}(R_L)=0$ decrease, reflecting the increase of the atomic-orbital ionization energies and atomic radius through the series. Since the atomic $4s$, $4p$, and $3d$ electrons are considered "active" valence electrons, all substitutional $3d$ elements (even Zn) constitute attractive impurities in silicon, as can be seen from the mostly negative values of $\Delta V_L^{\mathrm{ext}}(r)$ in Figs. 4 and 5.

The overall characteristics of $\Delta V_L^{\mathrm{ext}}(r)$ suggest for the high-Z_β limit the existence of deep-lying and localized, predominantly d-like states, separated from the higher-energy s-p—like states that will be considerably more delocalized. Moving towards lower Z_β values and hence to less attractive $\Delta V_2^{\mathrm{ext}}(r)$, the wave function of the deep d state will become more extended, and p-d hybridization will set in as the delocalized d state starts to overlap the domain of $\Delta V_1^{\mathrm{ext}}(r)$. These expectations are borne out by our detailed calculations.

The lower panels in Figs. 4 and 5 show the angular momentum components of the self-consistent screening perturbation $\Delta V_l^{\mathrm{scr}}(|\vec{r}|)$ [Eq. (10b)]. The spherical $l=0$ part is dominant, with the nonspherical contributions ($l=3,4$) being 2 orders of magnitude smaller (cf. Fig. 4). This behavior, together with the similar nature of the charge-density perturbation depicted in Fig. 2, highlights the physical basis for our choice of an angular momentum representation for $\Delta\rho(\vec{r})$ and $\Delta V(\vec{r})$ [Eqs. (10)]. It shows that the *angular* series converges rapidly (i.e., $\Delta\rho$ and ΔV are essentially isotropic objects) and that variational flexibility is needed predominantly for the description of the nontrivial *radial* variations in the scalar coefficients $\Delta\rho_l$ and $\Delta V_l^{\mathrm{scr}}$.

The characteristic feature of the dominant spherical $r \Delta V_0^{\mathrm{scr}}(r)$ screening component is that it always overshoots its asymptotic point-ion limit of ΔZ^v at the distance $r \approx R_{N1}$ corresponding to the first node in the density fluctuation (cf. Figs. 2 and 3). This overscreening at $r \approx R_{N1}$ is a direct consequence of the node in $\Delta\rho_0(|\vec{r}|)$. The degree of overscreening varies from one impurity to the other, scaling with the charge overshoot of Fig. 3. Since the degree of overscreening cannot be modeled intuitively, non-self-consistent calculations [which have to provide a guess for $\Delta V^{\mathrm{scr}}(\vec{r})$] are problematic.[34]

The overscreening effect causes $\Delta V^{\mathrm{scr}}(r)$ to approach its point-ion limit ΔZ^v somewhat slower than the external potential perturbation $\Delta V^{\mathrm{ext}}(r)$ ap-
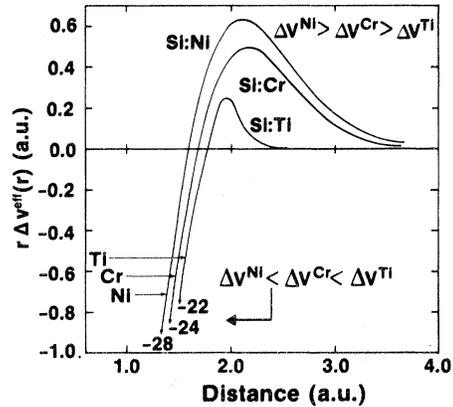
FIG. 6. r-multiplied total effective potential $r \Delta V^{\mathrm{eff}}(r)$ for Si:Ni, Si:Cr, and Si:Ti in the outer central cell. Notice the reversal of the ordering of $r \Delta V^{\mathrm{eff}}(r)$ in the outer central-cell region ($r \gtrsim 2$ a.u.) relative to their order in the inner central cell ($\lesssim 1.7$ a.u.). The $r \to 0$ limits of $r \Delta V_{L=2,l=0}^{\mathrm{eff}}(r)$ are indicated at the left. The nearest-neighbor Si atom is at $r = 4.44$ a.u.
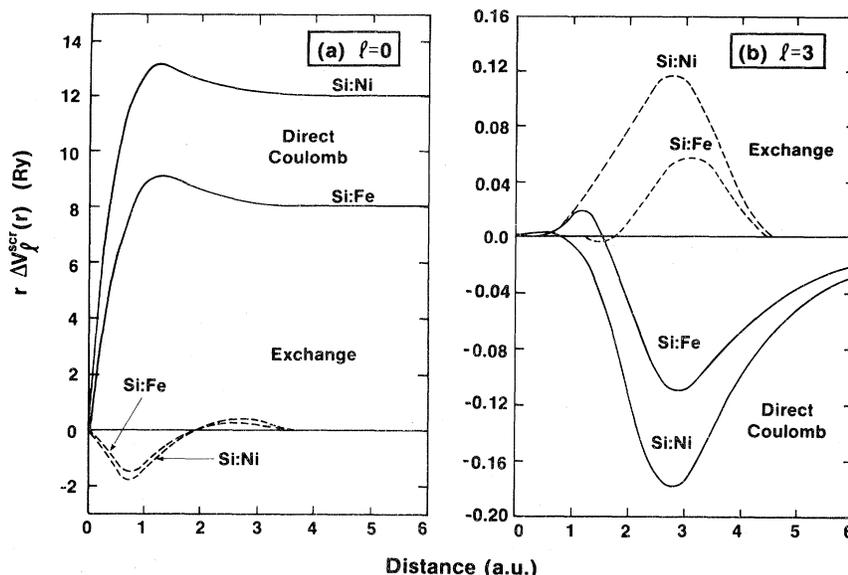
FIG. 7. Spherical (a) and lowest nonspherical (b) components of the r-multiplied direct Coulomb and exchange screening perturbations for Si:Ni and Si:Fe.

proaches its point-ion limit of $-\Delta Z^v$. Consequently, the total screened perturbation $\Delta V^{\mathrm{eff}} = \Delta V^{\mathrm{scr}}(\vec{r}) + \Delta V^{\mathrm{ext}}(\vec{r})$ of Eq. (4) is *positive* for $r \gtrsim R_N^*$. Figure 6 displays on an expanded scale $r \Delta V^{\mathrm{eff}}(r)$ in the outer central cell. For $R_0 < r \lesssim R_N^*$ (when R_0 is the classical turning point of the pseudopotential perturbation), the net effective potential is negative, due to the dominance of the attractive $\Delta V^{\mathrm{ext}}(r)$ over the repulsive $\Delta V^{\mathrm{scr}}(r)$ term (compare the corresponding panels in Figs. 4 and 5 and their sum for $r \lesssim 1.5$ a.u. in Fig. 6). Hence the effective screened perturbation $\Delta V^{\mathrm{eff}}(\vec{r})$ delineates the central cell into two distinct regoins, where R_N^* forms the demarcation point. In the inner central cell $(R_L < r < R_N^*)$, the effective potential is attractive and its depth follows the ordering of the bare pseudopotentials $\Delta V^{\mathrm{Zn}} < \Delta V^{\mathrm{Ni}} < \Delta V^{\mathrm{Cr}} < \Delta V^{\mathrm{Ti}}$, etc. The $3d$ wave functions of the free TA, as well as the wave functions of localized TA impurity states in silicon, are largely confined to this region. Hence the order of effective potentials also reflects the order of the $3d$ ionization energies I_β of free transition atoms $(I_{\mathrm{Zn}} > I_{\mathrm{Ni}} > I_{\mathrm{Cr}} > I_{\mathrm{Ti}})$, as well as the order of the impurity energy levels in the solid. In contrast, in the outer central cell the overscreening effect causes the effective perturbation to be positive, with a reversed order of effective perturbations $(\Delta V^{\mathrm{Zn}} > \Delta V^{\mathrm{Ni}} > \Delta V^{\mathrm{Cr}} > \Delta V^{\mathrm{Ti}}$, Fig. 6). This reflects the ability of the electron-rich impurities (Zn) to screen more effectively their potential relative to the electron-poor impurities (Ti) as shown in Fig. 3.

This reversal in order may lead to the interesting possibility that defect energy levels $\epsilon_i(Z_\beta)$ associated with localized wave functions (e.g., inside $r \lesssim R_N^*$) exhibit an opposite ordering, with atomic numbers Z_β, than defect levels associated with extended wave functions (e.g., excited states of the same impurities). We suspect this series reversal to also occur for impurities from other parts of the Periodic Table as discussed in a recent publication.[41] Notice that it has been often assumed both in effective-mass[39] and in tight-binding[37] calculations that the defect binding energies are monotonically decreasing as I_β decreases (or, equivalently, with the depth of the pseudopotentials[39] or with the electronegativity difference[37]). The present calculation indicates that this is the case only if all of the defect wave functions in question sample the potential in the inner central cell, but may not be so otherwise.

Figure 7 displays the decomposition of the screening perturbation $\Delta V^{\mathrm{scr}}(\vec{r})$ into its direct Coulomb [V^{ee} in Eq. (3)] and exchange-correlation [V^{xc} in Eq. (3)] components. As expected, the direct Coulomb contribution has the opposite sign to the exchange-correlation term, both for the spherical ($l=0$) and the lowest nonspherical ($l=3$) component (i.e., the screened Coulomb interactions are weaker than the bare Coulomb interactions). In the spherical screening, the direct Coulomb term is clearly dominant (ΔV_0^{xc} represents only 10–15 % of it); in the nonspherical screening the exchange-correlation screening is comparable to the direct Coulomb term.

C. Self-consistent dielectric function

One can use the calculation of the self-consistent response $\Delta V^{\text{scr}}(\vec{r})$ to the applied external perturbation $\Delta V^{\text{ext}}(\vec{r})$ to deduce an effective self-consistent dielectric function of the system:

$$\frac{\Delta V_L^{\text{ext}}(r)}{\epsilon_{l,L}(r)} = \Delta V_L^{\text{ext}}(r) + \Delta V_l^{\text{scr}}(r) \equiv \Delta V^{\text{eff}}(r) \ .$$

(20)

Here $\epsilon_{l,L}(r)$ measures the spherical $(l=0)$ and non-spherical $(l \neq 0)$ components of the effective dielectric screening responding to the Lth nonlocal component of the external perturation.

Figure 8 displays the leading spherical term $\epsilon_{0,L}^{-1}(r)$ in the inverse self-consistent dielectric function for $L=0$, 1, and 2 in Si:Fe. The results for other $3d$ impurities have similar characteristics. There are three notable features of $\epsilon_{l,L}^{-1}(r)$. First, the screening is seen to be extremely efficient: it occurs within a screening length characteristic of an *atomic dimension* (or the radius R_{N1} of the first node in the
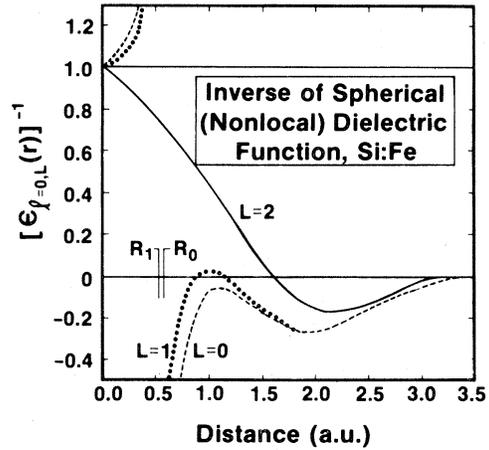
FIG. 8. Spherical inverse dielectric function for the $L=0$, 1, and 2 nonlocal pseudopotential perturbations calculated self-consistently for Si:Fe [Eq. (20)]. R_0 and R_1 denote the pseudopotential turning points at which $\Delta V_0^{\text{ext}}(R_0)=0$ and $\Delta V_1^{\text{ext}}(R_1)=0$. Notice the extremely effective screening that takes place within an atomic radius.
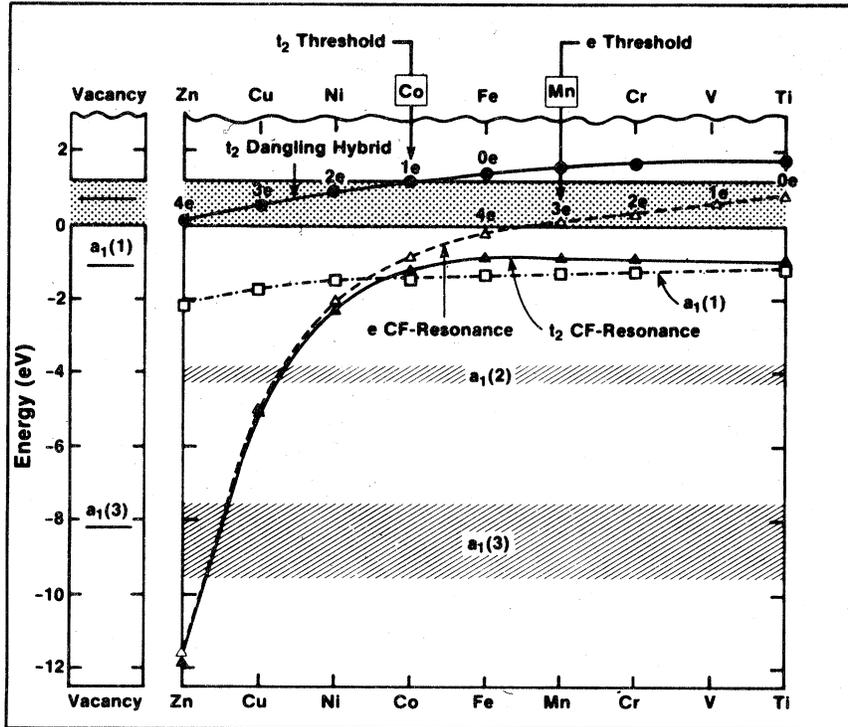
FIG. 9. Calculated defect energy levels for neutral substitutional unrelaxed transition-atom impurities in silicon. The zero of energy is set at the VB maximum. For resonances, the center-of-gravity energy is shown. The broad $a_1(2)$ and $a_1(3)$ resonances, showing little variation with the impurity, are denoted as crossed-hashed areas. For comparison, we show on the left the silicon vacancy levels (Ref. 17). Notice that the t_2^{DBH} first disappears into the conduction band at Si:Co ("t_2 threshold"), whereas the e^{CFR} first appears in the gap region in Si:Mn ("e threshold"). The numbers $4e$, $3e$ $2e$, etc. indicate the number of electrons occupying states above the VBM for the ground-state neutral impurities.

density fluctuation; cf. Fig. 2). This screening length is considerably shorter than that expected from a point-ion Thomas-Fermi model,[42] for a silicon vacancy,[34] or from the bulk linear screening used in effective-mass calculations.[39] Second, it is seen that although the screening of the s and p perturbations are similar (the small-r divergencies result from the occurrence of turning points at R_0 and R_1 for the $L=0$ and $L=1$ pseudopotentials; cf. Figs. 4 and 5), the screening of the d perturbation is both smoother and more effective. Third, a significant overscreening ($\epsilon^{-1} < 0$) corresponding to the charge overshoot (cf. Fig. 3) occurs in the bond region, with a maximum around the bond-center position ($r \sim 2.2$ a.u.).

We find that the various features of $\epsilon_{L,L}^{-1}(r)$ (e.g., position of nodes and negative minima) vary smoothly and monotonically with the atomic number of the impurity atom. This has an important practical consequence for our calculation: Having obtained $\epsilon_{L,L}^{-1}(r)$ for two impurity systems, one can use Eq. (20) to calculate a guess for a self-consistent screening $\Delta V^{\mathrm{scr}}(\vec{r})$ of other impurities from their $\Delta V_L^{\mathrm{ext}}(r)$. It turns out that this is usually an excellent guess, and we need only a few self-consistent iterations to refine it to the level of our imposed self-consistency tolerance.

D. Defect energy levels

The defect energy levels introduced by neutral substitutional $3d$ impurities in silicon are depicted in Fig. 9. For broad t_2 and e resonance states we display graphically only their centers of gravity. Energies corresponding to various impurities are connected by a line to guide the eye. The notation $4e$, $3e$, etc., denotes the gap-level occupation for the neutral state of the impurity. For comparison, we also show on the left-hand side the results for an unrelaxed silicon vacancy obtained previously.[34] The defect energies, orbital density localization parameters q_i^α and q_l^α [Eq. (11)], and percentage of valence-band character [Eq. (12)] are summarized in Tables II and III. When a defect-induced resonance has more than one peak, we give in these Tables the values corresponding to the most localized levels. We present only the results for the defect levels having q_i^α larger than $0.05e$. Next, we will discuss the defect levels in increasing order of their binding energy.

1. The t_2 dangling-bond hybrid

The highest-energy localized defect level is the t_2 dangling-bond hybrid (DBH) appearing in Si:Zn just

TABLE II. Energies, orbital-density localization parameters q^α [Eq. (11)], and their angular momentum components q_l^α, and percentage of valence-band character [Eq. (13)] for the t_2 dangling-bond hybrids and the t_2 crystal-field resonances.

Impurity atom	ϵ (eV)	t_2^{DBH} $q_p^{t_2}$ (%)	$q_d^{t_2}$ (%)	Total q^{t_2} (e)	Average VB content (%)	ϵ (eV)	t_2^{CFR} $q_p^{t_2}$ (%)	$q_d^{t_2}$ (%)	Total q^{t_2} (e)	Average VB content (%)
Zn	0.1	77	22	0.27	96	$\begin{cases} -11.9 \\ -11.8 \end{cases}$	$\begin{matrix}1.3\\0.0\end{matrix}$	$\left.\begin{matrix}98\\100\end{matrix}\right\}$	0.90	50
Cu	0.57	76	22	0.38	88.5	$\begin{cases} -5.7 \\ -5.1 \end{cases}$	$\begin{matrix}2.0\\0.6\end{matrix}$	$\left.\begin{matrix}95\\98\end{matrix}\right\}$	0.80	49
Ni	0.9	71	27	0.45	77.5	$\begin{cases} -2.4 \\ -2.2 \\ -2.0 \end{cases}$	$\begin{matrix}38\\0.5\\8\end{matrix}$	$\left.\begin{matrix}57\\97\\90\end{matrix}\right\}$	0.70	55
Co	1.16	62	37	0.49	61.5	$\begin{cases} -1.55 \\ -1.24 \end{cases}$	$\begin{matrix}11.2\\0.6\end{matrix}$	$\left.\begin{matrix}89\\97\end{matrix}\right\}$	0.62	65
Fe	1.4	50	49	0.53	55.5	$\begin{cases} -0.84 \\ -0.67 \\ -0.55 \end{cases}$	$\begin{matrix}9.0\\0.8\\10.0\end{matrix}$	$\left.\begin{matrix}90\\98\\87\end{matrix}\right\}$	0.55	74
Mn	1.52	52	49	0.50	46	$\begin{cases} -0.84 \\ -0.67 \\ -0.55 \end{cases}$	$\begin{matrix}9\\0.8\\11\end{matrix}$	$\left.\begin{matrix}90\\98\\86\end{matrix}\right\}$	0.53	75
Cr	1.58	55	45	0.42	35	$\begin{cases} -0.88 \\ -0.68 \\ -0.55 \end{cases}$	$\begin{matrix}7.8\\1.2\\12\end{matrix}$	$\left.\begin{matrix}92\\97\\87\end{matrix}\right\}$	0.43	77
Ti	1.60	57	43	0.20	14	$\begin{cases} -0.89 \\ -0.68 \\ -0.56 \end{cases}$	$\begin{matrix}10.6\\0.0\\16\end{matrix}$	$\left.\begin{matrix}88\\96\\81\end{matrix}\right\}$	0.31	78
	1.84	40	60	0.22	14					
	1.84	40	60	0.22	14					

TABLE III. Energies, orbital-density localization parameters q^α [Eq. (11)], and their angular momentum components q_l^α, and percentage of valence-band character for the e, t_2, and $a_1(1)$ crystal-field resonances.

Impurity atom	e^{CFR}				$a_1(1)$			
	ϵ (eV)	q_d^e (%)	Total q^e (e)	Average VB content (%)	ϵ (eV)	$q_s^{a}1$ (%)	Total $q^{a}1$ (e)	Average VB content (%)
Zn	−11.6	100	0.99	1	−2.1	96	0.23	93
Cu	−5.15	100	0.94	8.4	−1.7	94	0.35	85
Ni	−2.1	94	0.32	7.3	−1.5	97	0.38	85
	−1.9	100	0.32					
	−1.6	100	0.27					
Co	−1.0	98	0.88	10	−1.4	96	0.38	82
Fe	−0.28	99	0.84	14	−1.4	95	0.39	81
Mn	0.10	100	0.82	5	−1.3	97	0.41	79
Cr	0.22	99	0.78	6	−1.3	95	0.41	78
Ti	1.0	100	0.41	1	−1.4	95	0.38	82
	1.3	100	0.32					

above the valence-band maximum (VBM) and increasing in energy as one moves backwards through the $3d$ series. It first disappears into the conduction band for Si:Co ("t_2 threshold" in Fig. 9), continues to rise smoothly, and eventually saturates inside the conduction band at an energy $\epsilon \cong E_{VB} + 1.6$ eV as one reaches the Ti end of the $3d$ series. This level is not "pinned" to the silicon-vacancy gap level[37] ($\epsilon = E_{VB} + 0.7$ eV); it moves smoothly from the valence-band maximum into the conduction band. Table II shows that at the Zn end of the series the dangling-bond hybrid is made predominantly of valence-band states (96%), whereas at the Ti end the level becomes conduction-band-like (65%).

Figure 10 shows the decomposition of the t_2^{DBH} wave function into its angular-momentum components $G_l^{t_2}(|\vec{r}|)$ [Eq. (8c)] for the first two symmetry-allowed terms $l = 1$ and 2. For comparison, a partial wave-function resolution is also given for the silicon vacancy (upper left panel in Fig. 10). The significant fingerprint feature of the t_2^{DBH} state for all transition-atom impurities in silicon is a strongly *localized* atomiclike $l = 2$ component, which has a node in the *inner central-cell region*, and an *opposite* sign relative to the $l = 1$ wave-function component. This behavior is distinctly different from that observed in the silicon-vacancy dangling-bond wave function, which has a *delocal-*

ized $l = 2$ component with a node only *outside the central-cell region*, and the *same* sign as the $l = 1$ wave-function component in the inner region. As one moves backward from the Zn end to the low-Z limit of the $3d$ series, the node moves toward the central-cell boundary. Nothing special happens to the energy of this state, until the node of the $l = 2$ component extends into the central-cell boundary (Cr and Ti), at which point $\epsilon[t_2^{DBH}]$ saturates. This saturation indicates a confinement of the DBH level to a narrow energy range by the repulsive effect of two groups of levels, above and below in energy, having the same t_2 symmetry (the "avoided crossing" principle). The $l = 1$ component of the DBH is qualitatively characteristic of a hostlike dangling bond: it is extended, featureless, and has a node outside the central cell. However, a significant amount of impuritylike character is hybridized into the DBH state via the $l = 2$ components (hence the name dangling-bond hybrid).

The large $l = 2$ content of the t_2^{DBH} wave function (cf. Fig. 10 and Table II) is an important key to understanding the variation of the defect levels with atomic number (Fig. 9). Recall that the effective perturbation $\Delta V^{eff}(r)$ (cf. Fig. 6) has two distinct regions: It is attractive in the inner central cell with increased attractiveness as Z increases, and it is repulsive in the outer central cell with reduced
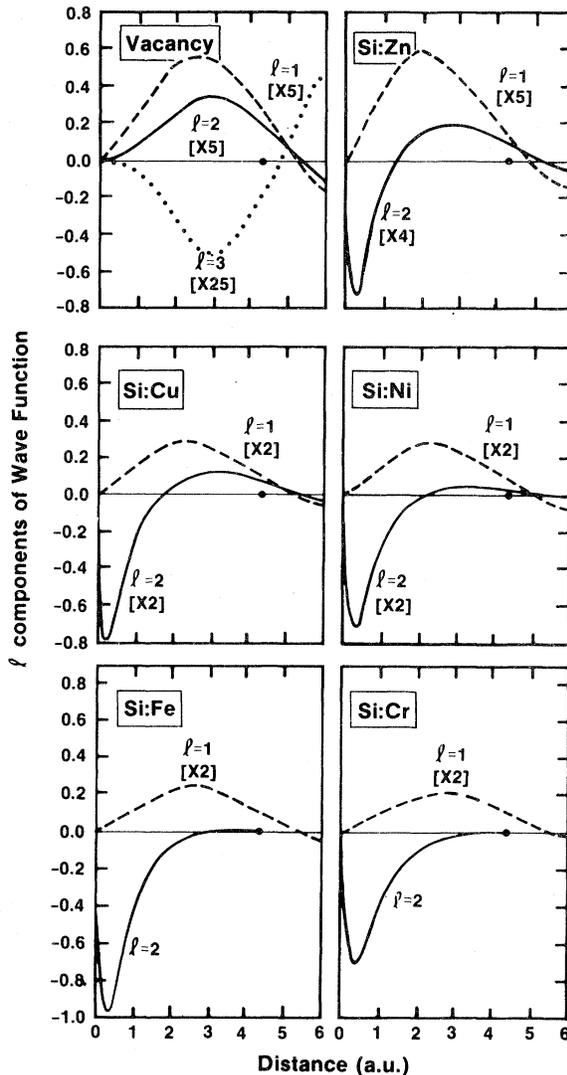
FIG. 10. Angular momentum components $G_l^{t_2}(\,|\,\vec{r}\,|\,)$ [Eq. (8c)] for the t_2 dangling-bond-hybrid wave functions of substitutional $3d$ impurities in silicon, and the silicon vacancy. Note the various scale factors for the different l components.

repulsiveness as Z decreases. The energy of the more localized defect levels (as well as the $3d$ energy of the free atoms) is controlled by the short-r behavior ($r \lesssim 2$ a.u.) simply because most of the wave-function and potential amplitudes are confined there [cf. the $\langle r \rangle_{3d}$ values for TA in Table IV (Ref. 43) and the $l = 2$ components of the impurity wave functions in Fig. 10]. As Z increases, both the atomic $3d$ orbitals and the d portion of the defect wave functions become more localized. Hence the positive kinetic-energy contribution to the orbital energy ϵ_i increases with Z. The faster decrease in

the negative and larger potential-energy contribution causes the well-known[43] increase in binding energy of atomic $3d$ orbitals with increasing Z (Table IV). The same behavior occurs for the $3d$ impurities whose binding energy (relative to the VBM) increases with increasing Z. The binding energies do not follow the (reversed) order and the magnitudes of $\Delta V^{\text{eff}}(r)$ in the outer central-cell region simply because the overlap of the wave function and $\Delta V^{\text{eff}}(r)$ is far larger in the inner central-cell region.

Notice that the energy of the vacancy t_2^{DBH} level is lower than that of the impurity atoms lighter than Co (Fig. 9) despite the fact that the vacancy potential is more repulsive than the impurity potential. The reason for the absence of pinning of the impurity levels by the vacancy level is that whereas the large and localized $l = 2$ components of the impurity wave functions have a large positive kinetic contribution to ϵ_i (including the centrifugal term), the smaller and diffused $l = 2$ components of the vacancy wave function come with a smaller kinetic term. This effect is missed by tight-binding calculations.[37]

The amount of p-d hybridization in the DBH is shown in Fig. 11, where the percentage of p and d character [q_{li}^{α} in Eq. (11)] is depicted across the $3d$ series. The DBH begins at Si:Zn as a predominantly p state (coupling with p-like valence-band states, cf. Table II). It reaches a 50%—50% p-d character near the center of the $3d$ series, after which the p character increases a little on account of the renewed availability of nearby p-like host states (this time, from the conduction bands). The large extent of p-d hybridization found near the Fermi energy is not entirely surprising; it is often found in inorganic chemistry of transition-metal tetrahedral coordination compounds [e.g., 25% p character[44] in $CuCl_4^{2-}$, 50% in CuI_4^{2-} 20—30 % in the organic complex copper $(\alpha,\alpha'\text{-Br})$ dipyrromethene[45]]. Note that a t_2 level such as the DBH can by no means be considered "p like." Figure 11(a) shows the percentage of localization of the DBH within the central cell. The level reaches its maximum localization for Fe. However, still about 50% of the wave function resides outside the central-cell region. Despite this, the central-cell region remains remarkably neutral (Fig. 3). The "electronic elasticity" effect leading to this remarkable behavior will be discussed in Sec. III F.

2. The e and t_2 crystal-field resonances

Moving in Fig. 9 to more negative energies, one observes a group of two defect levels, resonating within the host valence-band continuum, denoted as the e and t_2 crystal-field resonances (CFR). The t_2^{CFR} level is systematically below the e^{CFR} level. At

TABLE IV. Summary of calculated ground-state one-electron atomic energy levels in eV. The results of the local density (LD) and Hartree-Fock formalisms (Ref. 43) (HF) are shown. The typical size of the atomic d shell is given by the orbital moment $\langle r \rangle_{3d}$ in atomic units. Although the LD results for energies are consistently below the HF results, the two calculations produce nearly parallel trends.

| Atom | Configuration | LD $(\alpha_x = 1)$ | | | | HF | |
		$-\epsilon_s$	$-\epsilon_p$	$-\epsilon_d$	$\langle r \rangle_{3d}$	$-\epsilon_s$	$-\epsilon_d$
Zn	$s^2 d^{10}$	7.38	1.90	16.24	0.85	7.96	21.29
Cu	$s^1 d^{10}$	5.44	1.03	9.22	0.95	6.47	13.35
Ni	$s^2 d^8$	6.95	1.98	14.33	0.93	7.52	19.23
Co	$s^2 d^7$	6.73	2.01	13.34	0.98	7.27	18.38
Fe	$s^2 d^6$	6.50	2.04	12.33	1.04	7.02	17.60
Mn	$s^2 d^5$	6.26	2.05	11.27	1.10	6.74	17.38
Cr	$s^1 d^5$	4.60	1.24	5.64	1.31	6.04	10.16
V	$s^2 d^3$	5.73	2.06	9.03	1.28	6.27	13.86
Ti	$s^2 d^2$	5.43	2.04	7.82	1.40	6.00	11.99
Si	$s^2 p^2$	12.70	5.6			14.69	$-\epsilon_p = 8.1$

the Zn end of the $3d$ series, these levels occur near the bottom of the valence band and have a small crystal-field splitting. Moving toward the Ti end of the series, the levels first rise sharply toward the VBM. Then the t_2^{CFR} level saturates at $\epsilon \cong E_{VB} - (0.9-0.6)$ eV inside the valence band, whereas the energy of the e^{CFR} level continues to rise, first penetrating the band gap for Si:Mn (denoted "e threshold" in Fig. 9). It then continues to rise smoothly in the gap region as one moves toward Ti. This movement of the defect levels to higher energies directly reflects the reduced attractiveness of the effective perturbation in the inner central-cell region (cf. Figs. 4 and 5) with decreasing atomic number, as discussed previously for the t_2^{DBH} levels.

Table III indicates that the e^{CFR} is spectrally composed almost exclusively from conduction-band

states. However, the t_2^{CFR} state begins in Si:Zn with 50% valence-band character, and contrary to the t_2^{DBH} state, it acquires more valence-band character as one moves toward the Ti end of the $3d$ series (even though its energy approaches the conduction bands). Both the e and the t_2 crystal-field resonances are very much atomiclike in the central cell. This is illustrated in Fig. 12 which depicts the e^{CFR} wave function along the $\langle 110 \rangle$ crystal direction (pointing toward the next-nearest neighbors), and the t_2^{CFR} in the $\langle 111 \rangle$ crystal direction (pointing toward the nearest neighbors). The arrows in Fig. 12(a) indicate the distance R_e from the origin at which the e^{CFR} wave function attains $1/e$ of its maximum value. These R_e values are similar to the corresponding free-atom values.[43] Considering their orbital momentum content, the e^{CFR} is essentially a
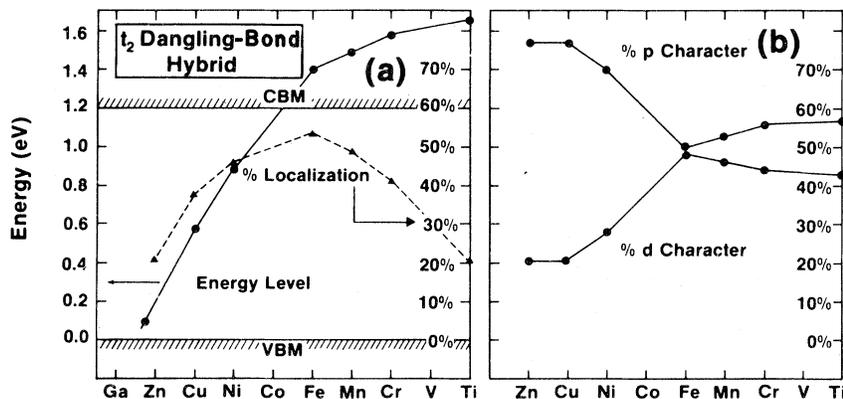
FIG. 11. (a) Variations in percentage of localization [q_i^α of Eq. (11)] and energy of the dangling-bond-hybrid level of TA in silicon (relative to the VB maximum). Notice that the DBH energy level is not pinned in the bandgap. (b) Variations in percentage of p and d character [q_{ii}^α of Eq. (11)] in the dangling-bond-hybrid wave function.
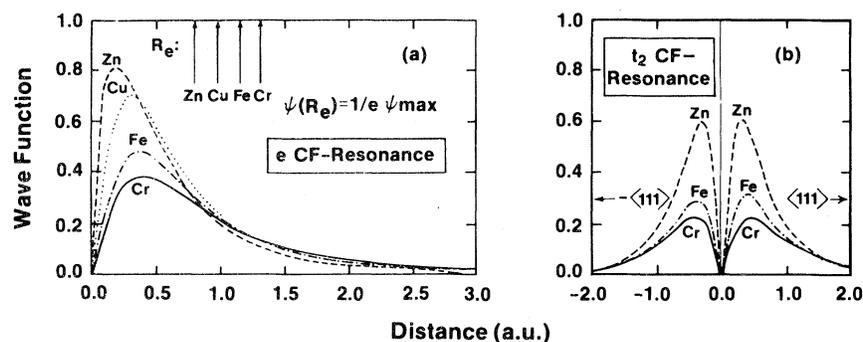
FIG. 12. Wave functions of (a) the e^{CFR} in the $\langle 110 \rangle$ crystal directions. Since the e^{CFR} wave function is symmetric, only the $+\langle 110 \rangle$ direction is shown. The arrows point to the radius R_e where this wave function attains the $1/e$ value of its maximum. The R_e values are very close to the atomic values (cf. Table IV). (b) the t_2^{CFR} wave function in the $\pm\langle 111 \rangle$ crystal directions.
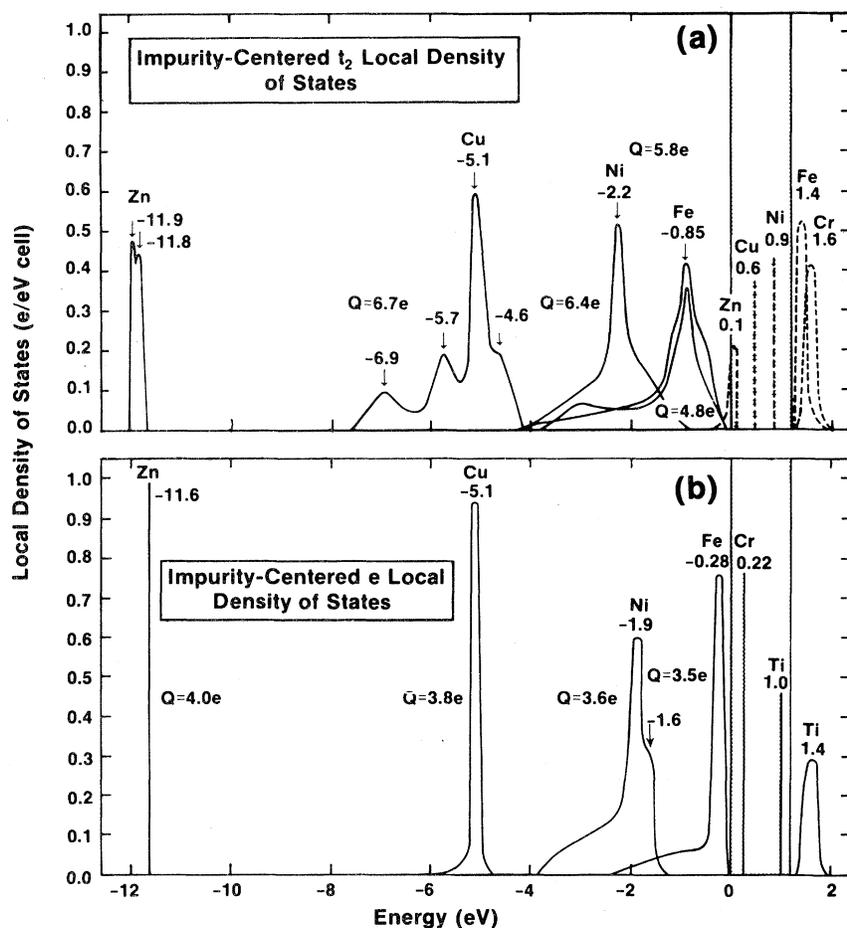
FIG. 13. Impurity-centered local density of states for (a) t_2 states and (b) e states. The zero of energy is set at the valence-band maximum. The peak energies (denoted next to the arrows) are accurate; the density of states is obtained from a coarse histogram and is less accurate. The Q values denote the total electronic charge enclosed within each resonance.

100% d state, whereas the t_2^{CFR} is 80—100 % d-like. Both crystal-field resonances commence at the Si:Zn end as very localized states (percentage of localization, 90—100%). In contrast to the DBH, they delocalize as one progresses to the Si:Ti end of the series (30% localization). We describe these states as crystal-field pairs since at the Si:Zn limit, they are essentially pure atomic d states, localized entirely within the central-cell region but with a (small) crystal-field splitting of the fivefold-degenerate $3d$ level into a threefold and twofold t_2 and e levels, respectively. As seen in Fig. 9 and Tables II and III, the e-t_2 one-electron crystal-field splitting is only ~0.1 eV for Si:Zn, and it increases as one moves backward in the $3d$ series reaching values of 1 and 1.7 eV for Si:Cr and Si:Ti, respectively.[46]

Having observed the trends in the centers of gravity of the various resonances, we now turn to examine their detailed energy distributions. Figure 13 shows the impurity-centered local density of states [Eq. (12)] for the t_2 [Fig. 13(a)] and e [Fig. 13(b)] defect levels. One can see that both e and t_2 crystal-field resonances are energetically narrow when they are close to the valence-band minimum (having a low hostlike t_2 and e density of states available for hybridization). As the resonances move to the valence-band center, they broaden as more host states of the appropriate symmetry and energy are available for hybridization. The different hybridization patterns of the e and t_2 resonances are related to their different angular symmetries. By symmetry, e-like states are composed of only $l = 2,4,\ldots$ components, whereas t_2-like states have $l = 1$, as well as $l = 2,3,4,\ldots$ components. The Kubic harmonic $K_2^e(\hat{r})$ associated with the major l component of the e states [cf. Eq. (8c)] is zero in the crystal direction $\langle 111 \rangle$ pointing toward the nearest ligand atoms. It is nonzero in the $\langle 110 \rangle$ direction, pointing to the next-nearest ligand atoms. On the other hand, the leading Kubic harmonic associated with a t_2 state ($l = 1$) is nonvanishing already for the directions pointing toward the nearest-neighboring ligand atoms. Hence, whereas the t_2 orbitals form σ bonds with the nearest neighbors, the e orbitals can form only π bonds with them (and weak σ bonds with the next-nearest neighbors). Since the e^{CFR} orbital can couple strongly only with next-nearest-neighboring host atoms, it remains only weakly hybridized (cf. the high q_d^e values in Table III) and appears as a rather narrow resonance [Fig. 13(b)] throughout the entire $3d$ series. On the other hand, the t_2^{CFR} already couples effectively with its nearest atoms, and shows pronounced p-d hybridization (cf. $q_p^{t_2}$ and $q_d^{t_2}$ values, Table II). Since the e^{CFR} does not bond strongly with the occupied host states, its wave function has only a small VB content (Table III).
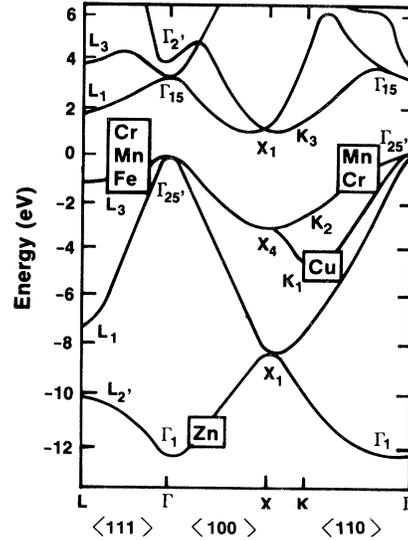
FIG. 14. Silicon band structure denoting the host states that contribute most to the t_2^{CFR} wave functions [Eq. (13)].

Many of the characteristics of the local density of states can be understood qualitatively from the nature of the host states that couple to the defect state. The increased width of the t_2^{CFR} [Fig. 13(a)], as well as its enhanced delocalization and valence-band character (Table II) as one moves backward in the $3d$ series from Zn, directly mirrors the availability of more host states of the appropriate symmetry for hybridization. For example, Fig. 14 depicts regions in the band structure of silicon that are found in our spectral analysis [Eq. (9)] to contribute significantly to the t_2^{CFR} wave function. Starting with Si:Zn, which has a very attractive pseudopotential perturbation $\Delta V_L^{ext}(\vec{r})$, its t_2^{CFR} level occurs at -11.7 eV below the VBM. The spectral analysis indicates that as much as 49.8% of this wave function is provided by the lowest silicon host-crystal band at the wave vector $\vec{k} = (2\pi/a)0.4(0,0,1)$ and energy $\epsilon_1^0(\vec{k}) = -11.83$ eV. Since the host t_2 density of states near this energy is small, the Si:Zn t_2^{CFR} has only a 50% valence-band character (Table II) and a small resonance width [Fig. 13(a)]. Moving up to Si:Cr, we see that its t_2^{CFR} level occurs at about $-(0.5-1.0)$ eV from the VBM. This region in the host band structure contains a high t_2 density of states due to the flat topmost valence band extending along the $\langle 111 \rangle$ direction from $\Gamma_{25',v}$ to $L_{3',v}$, as well as in the $\langle 110 \rangle$ direction extending from $\Gamma_{25',v}$ towards the $K_{2,v}$ point. We find that the $\vec{k}$ point $0.4(2\pi/a)(1,1,1)$ of the topmost valence band $[\epsilon_4^0(\vec{k}) = -1.22$ eV$]$ contributes 48% to the shoulder of the Si:Cr t_2^{CFR} state at $\epsilon[t_2] = -1.07$ eV, whereas the $\vec{k}$ point $0.4(2\pi/a)$ $(1,1,0)$ of the same band $[\epsilon_4^0(\vec{k}) = -0.92$ eV$]$ contributes 62% to the

TABLE V. Charge-density population analysis in electrons [Eqs. (15)—(18)] for Si:Zn, Si:Cu, Si:Ni, and Si:Fe.

Angular momentum	Si:Zn					Si:Cu				
	Q_l^{t1}	Q_l^e	Q_l^{t1}	Q_l^{t2}	Q_l	Q_l^{t1}	Q_l^e	Q_l^{t1}	Q_l^{t2}	Q_l
$l=0$	1.752	0	0	0	1.752	1.709	0	0	0	1.709
$l=1$	0	0	0	3.695	3.695	0	0	0	2.978	2.978
$l=2$	0	4.328	0	7.127	11.455	0	4.293	0	6.909	11.202
$l=3$	0.114	0	0.469	0.172	0.755	0.141	0	0.454	0.156	0.751
$l=4$	0.014	0.092	0.072	0.195	0.373	0.016	0.086	0.069	0.189	0.360
Q^α	1.880	4.420	0.541	11.189	$Q^{tot}=18.03$	1.866	4.379	0.523	10.232	$Q^{tot}=17.00$

Angular momentum	Si:Ni					Si:Fe				
	Q_l^{t1}	Q_l^e	Q_l^{t1}	Q_l^{t2}	Q_l	Q_l^{t1}	Q_l^e	Q_l^{t1}	Q_l^{t2}	Q_l
$l=0$	1.676	0	0	0	1.676	1.640	0	0	0	1.640
$l=1$	0	0	0	2.513	2.513	0	0	0	2.022	2.022
$l=2$	0	4.268	0	6.461	10.729	0	4.163	0	5.109	9.272
$l=3$	0.141	0	0.450	0.157	0.748	0.145	0	0.432	0.160	0.737
$l=4$	0.017	0.06	0.072	0.185	0.334	0.008	0.070	0.089	0.162	0.329
Q^α	1.834	4.328	0.522	9.316	$Q^{tot}=16.00$	1.793	4.322	0.521	7.453	$Q^{tot}=14.00$

t_2^{CFR} peak at $\epsilon[t_2]=-0.88$ eV. The availability of a high host t_2 density of states at this region leads to their substantial mixing into this state. This is reflected by the high percentage of valence-band character for Si:Cr (Table II). These examples illustrate how the details of the host band structure, together with the depth of the external perturbation $\Delta V_L^{ext}(\vec{r})$, control the nature of the defect levels. In sharp contrast to effective-mass impurities, the wave functions and energies of the deep Si:TA states are not controlled by host band-edge states.

There are four notable features in the variation of the main defect levels t_2^{DBH}, t_2^{CFR}, and e^{CFR} with the impurity atomic number. First, for most impurities the system sustains either a t_2^{DBH} or an e^{CFR} occupied gap state, but not both, in contrast with previous assumptions.[3,8] (Si:Fe may be a marginal case.) Second, most TA electrons appear as a valence-band resonance in t_2^{CFR} and not in the gap, in contrast to the interstitial case where all the impurity valence electrons appear in the gap levels.[21] Third, Mn forms a demarcation point in the 3d series, delineating the lighter impurities (with an impuritylike gap state e^{CFR}) from the heavier impurities (with a host-like gap state t_2^{DBH}). Fourth, the gap levels do not show pinning at the vacancy level, in contrast to the tight-binding results for s-p impurities.[37] Instead, the t_2 defect levels are confined by the host states that show a maximum in the t_2 density of states. These occur at $\sim E_{VB}-1$ eV and $\sim E_{CB}+2.7$ eV. Hence, it is the Phillips gap (~4.8 eV),[47] not the fundamental optical gap (1.1 eV) that controls these impurity levels.

3. The a_1 resonances

Figure 9 shows that there are three major a_1 resonances in the system, denoted $a_1(1)$, $a_1(2)$, and $a_1(3)$. The $a_1(1)$ resonance is relatively sharp and starts in Si:Zn at $E_{VB}-2.1$ eV, increasing in energy as we approach the Si:Ti end of the series towards $E_{VB}-1.3$ eV. It is almost a pure s state constructed predominantly from valence-band states, and has a localization parameter of $\sim40\%$ (cf. Table III). The lower resonances, $a_1(2)$ around -4 eV and $a_1(3)$ around $-(9.5-7.5)$ eV, are rather broad and weak ($q^{a_1}\approx0.05e$ per state). The lowest $a_1(3)$ resonance is an s-f hybrid with its wave function peaking past the nearest-neighbor Si atoms, whereas $a_1(2)$ has its wave function peaking within the central-cell boundaries. These resonances are very much hostlike and similar in character to the a_1 states of the silicon vacancy.[34] This can be understood qualitatively by noting that since the atomic s energy of a TA is considerably less negative than the corresponding value for Si (cf. Table IV), their bond-

TABLE VI. Charge-density population analysis in electrons [Eqs. (15)—(18)] for Si:Mn, Si:Cr, Si:Ti, and the silicon vacancy.

Angular momentum	Si:Mn $Q_l^{a_1}$	Q_l^e	$Q_l^{t_2}$	$Q_l^{t_1}$	Q_l	Si:Cr $Q_l^{a_1}$	Q_l^e	$Q_l^{t_1}$	$Q_l^{t_2}$	Q_l
$l=0$	1.617	0	0	0	1.671	1.603	0	0	0	1.603
$l=1$	0	0	1.997	0	1.997	0	0	0	2.113	2.113
$l=2$	0	3.197	5.036	0	8.233	0	2.342	0	4.922	7.264
$l=3$	0.142	0	0.202	0.452	0.796	0.146	0	0.459	0.173	0.778
$l=4$	0.018	0.09	0.177	0.072	0.357	0.018	0.089	0.076	0.06	0.243
Q^α	1.777	3.287	7.412	0.524	$Q^{tot}=13.00$	1.767	2.431	0.535	7.267	$Q^{tot}=12.00$

Angular momentum	Si:Ti $Q_l^{a_1}$	Q_l^e	$Q_l^{t_2}$	$Q_l^{t_1}$	Q_l	Vacancy $Q_l^{a_1}$	Q_l^e	$Q_l^{t_1}$	$Q_l^{t_2}$	Q_l
$l=0$	1.572	0	0	0	1.572	1.516	0	0	0	1.516
$l=1$	0	0	2.221	0	2.221	0	0	0	2.136	2.136
$l=2$	0	0.577	4.378	0	4.955	0	0.307	0	0.978	1.285
$l=3$	0.160	0	0.203	0.500	0.863	0.167	0	0.309	0.147	0.623
$l=4$	0.019	0.099	0.192	0.079	0.389	0.024	0.116	0.082	0.228	0.450
Q^α	1.751	0.676	6.994	0.579	$Q^{tot}=10.00$	1.707	0.423	0.391	3.489	$Q^{tot}=6.01$

ing combination is predominantly Si-like and resides in the valence band, whereas the antibonding TA-like states appear at much higher energies and do not contribute to the occupied states.

This phenomena is very well known in the chemistry of transition-atom compounds[44]: when a TA forms a binary compound with an element X having a lower-lying valence s orbital (e.g., $X=$Cl in CuCl), the $\Gamma_{1,v}$ bottom of the occupied valence band is composed predominantly of the X-atom s orbitals $(a\phi_s^X + b\phi_s^{TA}$ with $a \gg b$), whereas the antibonding counterpart of this state is the empty conduction-band level Γ_{1c} composed primarily of the TA s orbitals $(c\phi_s^X - d\phi_s^{TA}$, with $d \gg c$). Note, however, that although anchored at the TA site, this wave function may have most of its amplitude on the anion.[44] Hence, the TA s orbital gets effectively depopulated relative to its atomic ground state. The extent of this depopulation increases with the increased separation in the atomic s energies of the two atoms. Considering the trends in the atomic energies given in Table IV, one expects the effective s occupation of a TA in silicon to decrease as Z decreases. The a_1 valence-band resonances are accordingly mostly hostlike, whereas the atomic $4s$ state of the TA is expected to be mostly empty (diffused throughout the conduction band). This will be demonstrated in Sec. III G and Fig. 17.

E. Population analysis

Having discussed the individual defect energy levels in the system, we turn to analyzing the global charge distribution with the use of the population analysis defined in Sec. II E 4. Tables V and VI summarize the calculated orbital-representation Q_l^α, orbital Q_l, representation Q^α charges, and the total charge Q^{tot} for the Si:TA system, and the Si vacancy.

Inspection of these Tables reveals a number of chemical trends. First, the total a_1 charge is essentially s-like (93% in Zn to 90% in Cr and Ti), the e charge is essentially d-like (98% in Zn to 85% in Ti), and the t_1 charge is f-like ($\sim$85%), whereas the t_2 charge is a p-d hybrid (about 33—66%). Second, the t_1 and a_1 representation charges Q^α stay constant throughout the series, and are close to the corresponding values for the silicon vacancy. This is consistent with the fact that there are no localized t_1 defect levels for the Si:TA system, and that the a_1 resonances have their origin in the host crystal and are similar to the corresponding vacancy resonances subjected to a weakly attractive $L=0$ pseudopotential perturbation $\Delta V_{L=0}^{ext}(\vec{r})$. Third, the e representation charges stay constant from Zn to Fe, but once past Fe they drop linearly with Z. This reflects the absence of an e gap state from Zn to Fe and the gra-

dual decrease in occupation of the e gap level in going from Mn (three e^{CFR} electrons) to Ti (zero e^{CFR} electrons). On the other hand, the t_2 representation charges decrease from Zn to Co and stay constant afterwards. This reflects the emptying of the t_2^{DBH} gap states in going from Zn (four t_2^{DBH} electrons) to Co (no t_2^{DBH} electrons), and the unoccupied t_2^{DBH} levels past Co. Fourth, the total central-cell charge Q^{tot} corresponds to approximately six vacancy electrons (~ 1.5 per dangling bond, lingering into the central-cell volume) plus Z_β^v impurity valence electrons (e.g., $6 + 12$ for Si:Zn, $6 + 8$ for Si:Fe, and $6 + 4$ for Si:Ti). Inside the central-cell region, the impurity atom is hence nearly neutral, as observed in Sec. II A.

F. Stability of charged states

It has been known for a long time[3,14,39] that whereas the energy levels of different ionized states of a free ion span a large energy range, when a non-transition element exists as an impurity (Si:S, Si:P, etc.), it can sustain a number of ionized states in a very narrow energy range, often ~ 0.1 eV. Both phenomena are understandable. The weak potential perturbation of an s-p impurity in an s-p host crystal leads to defect wave functions that extend over many lattice constants [q_i of Eq. (11) is typically 0.01]. Hence, even if the intraatomic Coulomb energy per electron U is not small, the orbital energy ϵ_{gap} of this extended orbit varies only slowly with its occupation number N (proportional to the formal net charge Q_{net}). On the other hand, for free atoms the localized nature of the bound orbitals and the correspondingly large free-atom Coulomb energies U^0 lead to a rapid variation in the orbital energy with the amount of electronic charge $Q(N)$ on the atom,

$$\epsilon(N) \cong \epsilon_0 + UQ(N) . \tag{21}$$

E.g., for Mn the d^5 through d^2 ionization energies I_β (where I_β scales with Q as $I_\beta^0 - UQ$ since $I_\beta \propto -\epsilon$) are 33.7, 51.2, 72.4, and 95 eV, respectively,[48] leading to $U_d^0 \approx 20$ eV/e. (Indeed, local-density calculations for atomic ionization $d^Q \rightarrow d^{Q-1}$ show[49] that the orbital energies can be fitted by $\epsilon = \text{const} + U^0Q$, yielding $U^0 \approx 20$ eV/e for heavy $3d$ elements.) Surprisingly, however, when a TA exists as an impurity in a semiconductor, it can sustain in a narrow energy range a large number of different charge states *despite the fact that the gap level is highly levelized* (e.g., $q_i = 0.82$ for Mn; cf. Table III).

Our calculation explains this remarkable behavior in terms of two related effects: (i) The interaction with the semiconductor is shown to reduce U_{3d} by 2 orders of magnitude relative to the free-ion values U_{3d}^0, due to extremely effective *nonlinear* screening effects discussed in Sec. II C. This screening at-

tenuation is larger by an order of magnitude than that possible by a linear dielectric screening mechanism appropriate for effective-mass impurities. (ii) Despite the fact that, as in an atom, most of the impurity charge is localized in the central cell ($Q \approx Z^v$), this charge is remarkably *stable* against variations in the gap-level occupation number N (proportional to the formal net charge Q_{net}), leading to small variations in $\epsilon_{\text{gap}}(N)$. This unusual electronic elasticity of the central-cell region results from the fact that when the contribution $Q_{\text{gap}}(N)$ of the gap-level wave function to the central-cell charge is changed by altering N, the valence-band states respond *nonlinearly* by undergoing a rehybridization which changes their contribution $Q_{\text{VB}}(N)$ in the *opposite direction*. This result agrees with the prediction of Haldane and Anderson[8] based on a solution of the Anderson Hamiltonian for the semiconducting Kondo system within the unrestricted Hartree-Fock approximation.[50] Our calculation, which includes a realistic description of the host, impurity, and their interactions, provides a detailed explanation for this effect.

The total charge $Q(N)$ in the central cell due to the impurity atom alone is calculated as

$$\int_0^{R_{\text{cc}}} [\rho_D(\vec{r}) - \rho_V(\vec{r})]d\vec{r} ,$$

where we substract the vacancy (V) charge $\rho_V(\vec{r})$ from the defect charge $\rho_D(\vec{r})$ [cf. Eq. (6)] to get the effective transition-atom charge in the solid for each value of N. This central-cell charge [equal to ΔQ^{tot} of Eq. (19c) but denoted here for simplicity by $Q(N)$] represents the total number of valence electrons that an impurity atom has when it is incorporated substitutionally in Si. It depends on the number of electrons N occupying the gap level. If $N = N_0$, the system as a whole has a zero formal charge $Q_{\text{net}} = 0$ (e.g., $N_0 = 3$ for Si:Mn; cf. Fig. 9). We now separate $Q(N)$ into the contribution $Q_{\text{VB}}(N)$ from the occupied valence bands of the perturbed system, and the contribution $Q_{\text{gap}}(N)$ from the gap level alone,

$$Q(N) = Q_{\text{VB}}(N) + Q_{\text{gap}}(N) , \tag{22}$$

where

$$Q_{\text{gap}}(N) = \sum_{i,\alpha}^{\text{gap}} N_i q_i^\alpha$$

[cf. Eq. (11)]. Before discussing the self-consistently calculated results for $Q_{\text{VB}}(N)$ and $Q_{\text{gap}}(N)$, consider the prediction of simple models. If the system would have behaved as in a one-electron rigid-band model, addition of electrons to the highest occupied gap level in excess of N_0 would have kept $Q_{\text{VB}}(N)$ constant and increased $Q_{\text{gap}}(N)$ linearly by
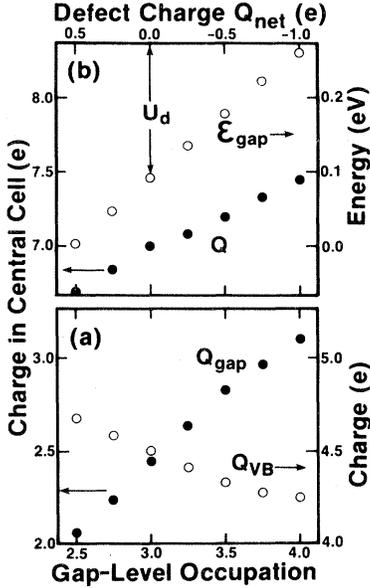
FIG. 15. (a) Contributions of the gap level $Q_{gap}(N)$ and the valence band $Q_{VB}(N)$ to the central-cell charge Q, as a function of the gap-level occupation N. (b) Variation in central-cell charge Q and gap-level energy ϵ_{gap} (relative to the VB maximum) with N.

$(N-N_0)q_i^\alpha$, where q_i^α is the gap-level charge per electron given by Eq. (11). Similarly, in an effective-mass donor model, $Q_{VB}(N)$ is conceptualized to remain almost constant as the system is charged, since Z_H^v of the impurity electrons are used to "heal" the vacancy site, and the remaining $(Z_I^v - Z_H^v + N - N_0)$ impurity electrons occupy the extended donor-gap level. Furthermore, in a non-self-consistent calculation for a deep impurity, $Q_{VB}(N)$ would likewise be independent of N, and $Q_{gap}(N)$ would increase linearly with $N - N_0$.

Figure 15(a) shows for Si:Mn the variations in $Q_{VB}(N)$ and $Q_{gap}(N)$ with N, as obtained by performing separate self-consistent calculations for each value of N. It is seen that whereas $Q_{gap}(N)$ increases nearly linearly with N, as expected, the valence-band contribution $Q_{VB}(N)$ *decreases* with N, leading to $Q(N)$ that varies with N considerably slower than $Q_{net} = N - N_0$ or $Q_{gap}(N)$ alone. Much like a pool below a waterfall, as electrons are poured into the gap level the valence-band states rearrange themselves so that they leak out of the central cell to partially compensate for the disturbance. This self-regulating behavior of the system, in response to an external perturbation, is remarkably analogous to the homeostasis control mechanisms in biology or to servomechanisms in machine operations. In such systems, despite the existence of dynamic perturbations, a stability of operation within a prescribed range of operating parameters is maintained through

a (often nonlinear) self-regulating feedback response. This is possible only in an open thermodynamic system such as Si:TA, but not in a closed system such as an isolated transition atom. In the latter case, the effect of added electrons must be absorbed by the atom itself through an increased delocalization of its wave functions and a rapid change of the ionization energy with $N - N_0 = Q_{net}$. We suspect a similar self-regulating behavior to occur in many other systems whenever localized states characterized by large correlation energies coexist in a similar energy domain with an itinerant manifold of states with which it can exchange charge. Surface states and valence-fluctuation states are probably just a few possible examples.

In the Si:TA system, almost all Z^v electrons are concentrated in the central cell much like in a free atom. For example, Table VI shows that in Si:Mn^0 the central-cell contains

$$Q(N=3) \equiv (Q^{tot})_{impurity} - (Q^{tot})_{vacancy} = 13 - 6 = 7$$

electrons, as for the free-space Mn^0 d^5s^2 atom. Notice further that $2.46e$ occupy the gap level ($3q^e$ of Table III) and the remaining $4.54e$ appear as resonances in the VB. However, in sharp contrast to the behavior of a free atom, the energy $\epsilon_{gap}(N)$ of the gap level in Si:Mn increases only slowly with N [Fig. 15(b)] due to the weak variation in $Q(N)$ facilitated by the self-regulating mechanism. This effect will allow the existence of several charged states in a narrow energy range; *it will hence lead to pinning of the Fermi energy in the band gap.*

Weak variations in $\epsilon_{gap}(N)$ are indicative, by Eq. (21), of weak variations in $U_n Q$. We find that not only does $Q(N)$ change more slowly than Q_{net}, but that U is also reduced compared to its atomic value. This is already suggested by the hyperscreening evident from Fig. 8 and discussed in Secs. II B and II C. Our model can directly provide the value of the effective intra-atomic Coulomb repulsion U_n. It is calculated as the difference between the total energy $E(-/0)$ of a neutral (0) to singly negative $(-)$ transition, and the neutral to singly positive $(+)$ transition $E(+/0)$. The use of Slater's transition-state method gives $E(-/0) \equiv \epsilon[e^{3.5}]$ and $E(+/0) = \epsilon[e^{2.5}]$, and hence $U = E(-/0) - E(+/0) = 0.2$ eV (Ref. 51) [depicted in Fig. 15(b)]. It is 2 orders of magnitude smaller than the free-atom value of $U_n^0 \cong 20$ eV. Whereas values of $U_n \sim 0.2$ eV are often observed even for s-p defects in silicon, the attenuation relative to the free-atom value is normally only $1/\epsilon_0$, where $\epsilon_0 = 11.4$ is the static dielectric constant of the host. The extra attenuation of U_d calculated here directly reflects the extremely effective nonlinear screening in the Si:TA

Defect Charge Q_{net} (e)
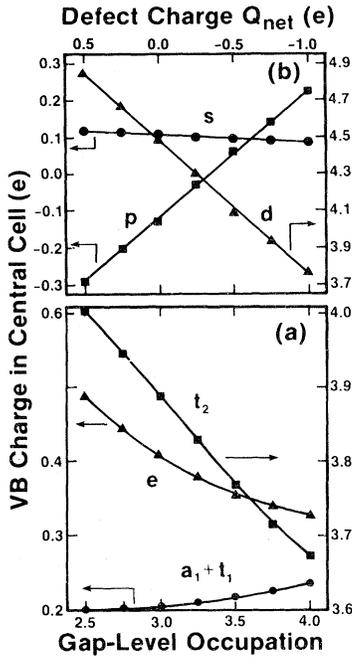
FIG. 16. Analysis of the valence-band contributions $Q_{VB}(N)$ of Si:Mn to the central-cell charge in terms of (a) wave-function representations and (b) orbital contributions.

system. The small value of U_d obtained here implies *a posteriori* that a one-electron description of this system is a reasonably good approximation.

One can further inquire to the chemical origin of the leakage of valence-band states from the central cell. Figure 16(a) analyzes $Q_{VB}(N)$ in terms of the wave-function representations [Eq. (19a)] contributing to it. It is seen that whereas the a_1 and t_1 valence states, much like in a rigid-band model, remain largely unresponsive to the change in the gap-level occupation, the e and t_2 valence-band wave functions tend to leak out of the central cell as more electrons are added to the gap level. A partial-wave analysis of $Q_{VB}(N)$ depicted in Fig. 16(b) [using Eq. (19b)] shows that the reduction in the valence-band contribution to the central-cell charge results from a strong attenuation of its d content (largely localized wave functions) accompanied by a smaller increase in the p content (largely delocalized wave functions), with almost no response from the s states. We conclude that the remarkable feature of the Si:TA system is that it *resists* an ionic charge separation in response to addition of electrons by having its valence-band states leak out from the central cell, and that the system accomplishes this leakage by undoing its $p\text{-}d$ hybridization (i.e., dehybridization).

In ionic host crystals such as alkaline-earth oxides, where $p\text{-}d$ hybridization is unimportant, one expects the self-regulating behavior to be facilitated by

a different mechanism. From the calculation of Stoneham and Sangster[52] one expects that the leakage of VB charge in response to charging the impurity is accommodated by polarization and distortion effects. This means that as the force field describing the various atom-atom interactions in the system changes with the ionization state of the impurity, the relaxed energy required to remove a host atom and replace it by a charged impurity changes as well. This polarization energy will then offset the increase in the ionization energy [second term in Eq. (21)] due to charging. Hence for TA impurities in ionic host crystals one still expects a self-regulating behavior leading to the existence of many charged states in the gap, but the role of covalent $p\text{-}d$ dehybridization is now taken up by a rearrangement in the polarization field.

G. Effective electronic configuration

The small U_d values for the Si:TA system have an interesting implication for the effective electronic configuration of the impurity atom. Whereas the large *atomic* U_d^0 values lead to the well-known preference for occupying the $4s$ subshell before the $3d$ subshell is completed, small U_d values in the *solid* may lead to a population inversion in the ground state.

Figure 17(a) depicts, with the use of Eq. 19(a), the manner in which the atomic $a_1^1(et_2)^{n+1}$ or $a_1^2(et_2)^n$ ground-state configuration is redistributed when the atom is placed in the solid. One sees that most of the atomic a_1 electrons are depleted and redistributed into the e and t_2 representations. The e representation contains nearly four electrons for Zn through Fe, then 3, 2, 1, and 0 electrons for Mn, Cr, V, and Ti, respectively, as the e^{CFR} moves into the gap and is progressively emptied (Fig. 9). The decrease in the occupation from Zn to Fe, for the t_2 representation followed by a constant occupation of approximately four electrons thereafter, reflects the evacuation of the t_2^{CFR} level as it moves through the gap.

Figure 17(b) shows the variations with atomic number of the effective orbital configuration ΔQ_l of the impurity atom [Eq. (19b)]. The total number of impurity valence electrons Z_β^v is shown for comparison (dashed line). The interesting result of this figure is that whenever the $3d$ subshell *can also accommodate* the s electrons (i.e., to the left of Ni d^8 in the periodic table), the latter are promoted into it. This trend is evident in Fig. 17(b) from the approach of the ΔQ_d line to the Z_β^v (dashed) line. Hence, whereas Fe, Mn, Cr, and Ti have 6, 5, 4, and 2 valence d electrons in the free atom, respectively, they have about 8, 7, 6, and 4 d electrons, respectively, when placed substitutionally in silicon. Transition-atom impurities in silicon tend therefore
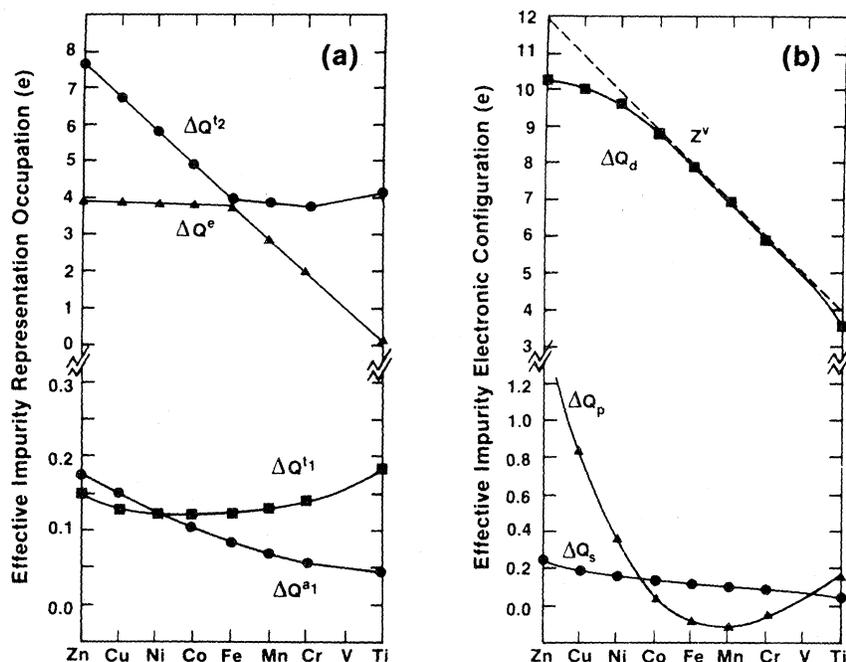
FIG. 17. Electronic configuration of an effective transition atom embedded substitutionally in silicon. (a) Representation occupations ΔQ^{α} [Eq. (19a)], and (b) orbital occupation ΔQ_l [Eq. (19b)]. For comparison the total number of valence electrons Z_{β}^{v} is shown as a dashed line.

to approach a noble-metal configuration. This is in marked contradiction to the Ludwig-Woodbury model[3] which hypothesized that for substitutional $3d$ elements, the $3d$ electrons will be promoted into the sp subshell to form a tetrahedral hybrid (i.e., $d^{n}s^{2} \rightarrow d^{n-2}s^{1}p^{3}$, or $d^{n+1}s^{1} \rightarrow d^{n-2}s^{1}p^{3}$). We find that for the completely filled d-subshell elements Zn and Cu, where the $3d$ shell cannot accommodate any s electrons, the latter are being promoted into a p state, yielding $\approx s^{0.24}p^{1.56}d^{10}$ for Zn and $s^{0.19}p^{0.84}d^{9.92}$ for Cu. The Ni impurity has the intermediate occupation of $s^{0.16}p^{0.38}d^{9.44}$; hence Ni in silicon forms a noble-metal-like compound with a nearly filled d subshell. This should be contrasted with the effective electronic configurations deduced for the bulk elemental metals[53] of $s^{1.3}d^{4.7}$, $s^{1.4}d^{6.6}$, $s^{1.3}d^{8.7}$, and $s^{1.3}d^{9.7}$ for Cr, Fe, Ni, and Cu, respectively, showing a much higher occupation of the s orbitals.

A calculation of the charge-density components $\delta_l(r)$ of the effective impurity atom [Eq. (19c)] shows them to be very similar to those of a free atom (in fact, almost indistinguishable on a regular-size plot) with somewhat compressed tails and a small nonspherical $\delta_3(r)$ term absent in the free atom. This analysis suggests a surprisingly simple chemical picture for the changes that the crystalline environment exerts on the impurity atom. Whereas the discrete atomic energy levels of the impurity

atom undergo substantial changes in the solid—they split and broaden into structured resonances covering a wide energy range—its final effective charge density simply corresponds to that of a compressed atom with its s electrons excited into the p and d shells.

H. The importance of self-consistency

In the defense of computational schemes that are inherently non-self-consistent, it is often argued that self-consistency is unimportant (or even incorrect[54]) for obtaining physically and chemically correct trends in the defect levels. We will address this issue for transition-atom impurities in silicon.

In the absence of any _a priori_ knowledge of the impure system, the perturbation potential is often modeled from the properties of the isolated host and impurity atoms (e.g., Ref. 37). Figure 18 depicts the spherical screening perturbation $\Delta V_0^{\text{scr}}(|\vec{r}|)$ for Si:Fe obtained from taking the difference between the corresponding atomic screening potentials (dashed line) together with the self-consistent screening potential from the present calculation (solid line). The two curves may appear similar, suggesting at first that the non-self-consistent atomic guess is physically reasonable. However, this is not the case. The insert to Figure 18 shows the energy-level diagram obtained with these two potentials. The atomic model leads to very large errors in the energy
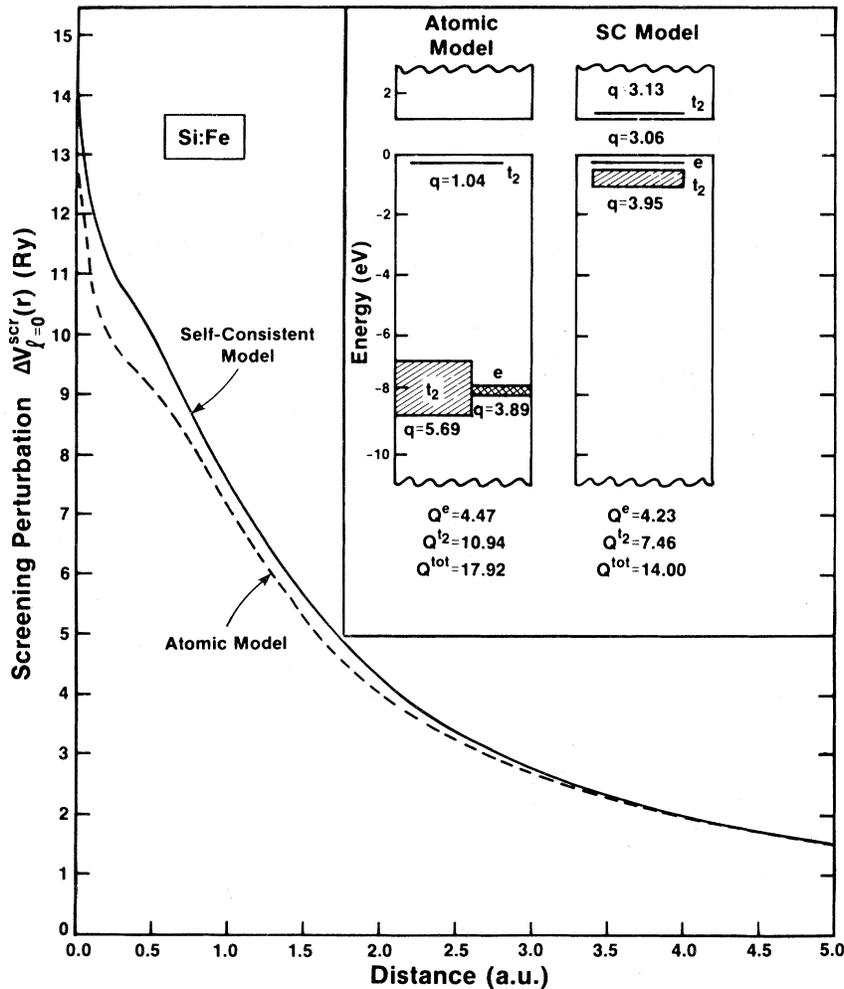
FIG. 18. Spherical $l = 0$ component of the screening perturbation obtained for Si:Fe in an atomic model (— — —), and in a self-consistent model (———). The inset compares the defect energy levels and central-cell charges produced by these two screening potentials.

levels. Compared with the self-consistent result, the atomic model predicts that the t_2^{DBH} level ($E_{VB} + 1.4$ eV in the self-consistent model) will be below the VBM at $E_{VB} - 0.3$ eV, the e^{CFR} level ($E_{VB} - 0.28$ eV) will be pulled down to $E_{VB} - 8$ eV, and the t_2^{CFR} level $\sim (E_{VB} - 0.6$ eV) will occur at $E_{VB} - (7-8.5)$eV. The overattractiveness of the atomic potential results in a far too high t_2 population in the central-cell region ($Q^{t_2} = 10.94e$, instead of $Q^{t_2} = 7.46e$). Taking into account all representations, the central-cell region in the atomic model contains almost four electrons too many.

The wave functions produced from the input atomic screening $(\Delta V_0^{scr})_{input}$ can be used to calculate an output screening $(\Delta V_0^{scr})_{output}$. Figure 19 shows their deviation

$$(\Delta V_0^{scr})_{output} - (\Delta V_0^{scr})_{input}$$

(a measure to the degree of non-self-consistency) for

both the self-consistent and the atomic input potentials. Whereas the atomic input produces an enormous error (average deviation σ of 8500 mRy), the error in the self-consistent potential is reduced by 3 orders of magnitude. There can be no doubt that a non-self-consistent atomic guess for the potential is poor.

I. Comparison with other calculations

Our results agree with the overall trend in defect energy levels found by Cartling,[19(a)] and by Hemstreet,[20] who used a small-cluster multiple-scattering $X\alpha$ (MS$X\alpha$) approach. Both cluster and Green's-function calculations show that the heavier $3d$ elements have a t_2^{DBH} gap level, whereas the lighter impurities have an e^{CFR} level in the gap. Among the differences in the results are the following (compare Fig. 9): The cluster model shows the t_2^{CFR} and e^{CFR} level for Si:Zn to be below the bottom of the valence band due to the too-narrow valence-band width.
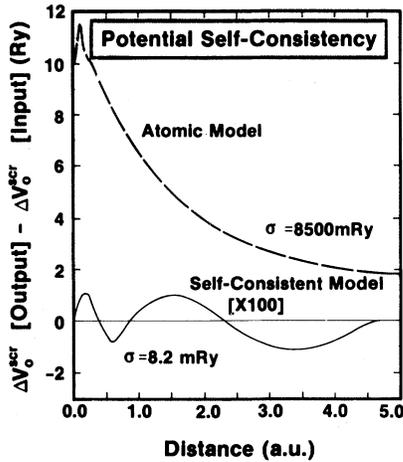
FIG. 19. Self-consistency error in potential screening produced by an atomic model (— — —), compared with that produced by a self-consistent model (———) for Si:Fe. σ indicates the energy error between output and input potentials.

For Si:Ni the t_2^{CFR} and e^{CFR} levels are about 2 eV deeper than in the present calculation; for Si:Fe the e^{CFR} is found in the gap, whereas in the present calculation it is just below the VBM, etc. If one considers the simplicity of the cluster model, these discrepancies are very reasonable. In general, the cluster results seem internally more accurate for the heavy elements Zn, Cu, Ni, and Co sustaining a t_2^{DBH} gap level (t_2^{DBH} energies of 0.26, 0.57, 0.83, and 1.03 eV, respectively, compared with our results of 0.1, 0.57, 0.9, and 1.2 eV) than for the lighter elements sustaining a e^{CFR} gap level (e^{CFR} level at 0.7 eV for Si:Cr, compared with our result of 0.2 eV). This is understandable since the Si_4X cluster (where X is a TA) used in these calculations[19(a),20] has no Si atoms as next-nearest neighbors to the impurity atom. Since the e levels can form σ bonds only with the next-nearest neighbors and none with the nearest neighbors, its stability is underestimated in a Si_4X cluster. From the results of Refs. 20 and 19(a) it appears that this causes the e gap levels for Co (Ref. 20) and Fe [Ref. 19(a)] to be 0.5 eV too high. In general, we find that the position of the e^{CFR} level is very sensitive to the variational flexibility of the basis set; incomplete sets (e.g., 135 instead of 219 local orbitals) push this level upwards by 0.5 eV. Similarly, when we zero the potential from a point between the nearest and next-nearest neighbors and beyond, the e^{CFR} level moves up by 0.55 eV. Information on wave functions and charge density cannot be compared since they are not included in the cluster work.

Somewhat more fundamental differences exist in interpretation of the origin of gap levels. Hem-

street[20] interprets his t_2^{CFR} level for the light elements (e.g., Si:Cr) to be the vacancy dangling-bond orbital moving down into the valence bands. However, if one inspects the energy levels and wave functions above the conduction-band minimum (CBM) in the present calculation, one finds that, in fact, the vacancylike dangling-bond hybrid has moved up in energy into the conduction band (cf. Fig. 9), whereas the t_2^{CFR} level inside the valence bands is not dangling-bond-like. [A similar misinterpretation occurred in Ref. 27(b) for impurities in GaAs.] Also, Hemstreet interpreted the change in the behavior of the gap levels from Fe, Cr to Co, Ni, Cu, and Zn as reflecting a crossing in the ground-state atomic $3d$ and $4s$ energy levels ($4s$ being above $3d$ in Cr, whereas the order is being reversed in Fe). In fact, both local-density and Hartree-Fock calculations[43] show the atomic $3d$ energy to be always below the $4s$ energy for all $3d$ elements in their ground states (cf. Table IV).

A further difference is that the crystal-field splitting between t_2^{CFR} and e^{CFR} tends to be somewhat smaller in the muffin-tin cluster calculation.[20] This is in line with the fact that all covalent interactions are seriously underestimated by a muffin-tin model. Recall that the band gap that measures the strength of such bonding-antibonding interactions is zero for the Si crystal when a muffin-tin potential is used in a band-structure calculation[55] and is underestimated by 30% (1.6 eV) in diamond.[40] The $\sim$1-Ry nonspherical part of the central-cell potential,[40,56] neglected in muffin-tin calculations,[19-21] clearly makes the host appear more metallic than covalent. (The only reason that a finite band gap was obtained at all in Si by Hemstreet[20] and DeLeo et al.[21] is that a finite cluster with hydrogen terminators was used, where the band edges do not correspond to the infinite crystal band edges.) The underestimation of the Si-Si covalent interactions by $\sim$1 eV shows up as a spurious reduction in the TA crystal-field splitting by $\sim$0.5–0.8 eV.

The only other self-consistent calculation for Si:TA to which we can compare our results is the Si:Zn calculation by Bernholc et al.[57] In this Green's-function calculation, the Zn atom was represented by a local empirical pseudopotential. A t_2^{DBH} state was reported at $E_{VB}+0.12$ eV, in good agreement with our results (using a nonlocal spd pseudopotential) of $E_{VB}+0.1$ eV for a 250-atom supercell, and $E_{VB}+0.08$ eV for a 1458-atom supercell.

IV. THE LOCAL-DENSITY MODEL AND BEYOND: ANALYSIS OF EFFECTS LEFT OUT

The calculation reported in this paper was performed consistently within the local-density (LD)

model[24] for interelectronic interactions to a high degree of internal numerical precision. Whereas the model describes very well the ground-state properties associated with the charge distribution, it has a number of clear deficiencies in describing ionization, excitation, and activation energies.[58] In this section we discuss these limitations and present a semiquantitative analysis of the leading corrections to it.

A. Orbital energies and removal energies

As has been pointed out previously (e.g., Refs. 58 and 49), the energy eigenvalues $-\epsilon_i^{LD}$ obtained in the LD formalism do not correspond to removal energies for three reasons. First, Koopmans theorem does not apply to the LD Hamiltonian,[49] i.e., even if the wave functions are assumed to be unaffected by the excitation process the difference in total energies E of the excited and the ground state does not equal the difference $\Delta\epsilon_{fi}^{LD}$ between the final (f) and initial (i) one-electron energies. The correction due to this effect for an $i \to f$ excitation is $\Delta\Pi_{fi}^{SIC} = \Pi_f^{SIC} - \Pi_i^{SIC}$, where Π_i^{SIC} is the self-interaction correction[58] (SIC) to orbital $|i\rangle$. (In the Hartree-Fock model, $\Pi_i^{SIC} = 0$.) Second, wave functions do relax upon excitation. The relaxation energy is denoted as $\Delta\Sigma_{fi}^R = \Sigma_f^R - \Sigma_i^R$, where Σ_i^R is the relaxation self-energy of state $|i\rangle$. It is defined here to include both the distortion of the excited orbital itself, and the rearrangement of all other occupied states, leading to an improved electrostatic screening of the excited system relative to the ground state. Third, the homogeneous-electron-gas approximation for the correlation potential is imperfect in a number of ways. When LD calculations are corrected for self-interaction and relaxation effects, a large number of its predictions become accurate.[58] However, a number of systematic discrepancies with experiment still remain, to which we refer collectively as arising from "extra correlation" (EC) corrections. The correction to the excitation energy will be denoted $\Delta\Sigma_{fi}^{EC} = \Sigma_f^{EC} - \Sigma_i^{EC}$, where Σ_i^{EC} is the extra correlation self-energy. Among the best known errors due to EC are the systematic overestimation of the inter-configurational s-d energies in transition atoms,[58] the overestimation of the spin-polarization exchange splitting in metallic Co and Ni (by factors of 1.2 and 2.2, respectively), and the deformed spectral intensities in the optical-absorption spectra of solids such as silicon and diamond.

In the absence of lattice distortions, the excitation energy between two defect states, say between the e^{CFR} with n electrons and the t_2^{DBH} with m electrons (the states are denoted hereafter for simplicity as e and t, respectively), is therefore given as

$$E[e^n t^m] - E[e^{n-1} t^{m+1}]$$
$$= \Delta\epsilon_{te}^{LD} + \Delta\Pi_{te}^{SIC} + \Delta\Sigma_{te}^R + \Delta\Sigma_{te}^{EC}, \quad (23)$$

and not by the orbital energy difference $\Delta\epsilon_{te}^{LD}$ alone. Likewise, the ionization energy I_e from state e is given as

$$I_e = -\epsilon_e^{LD} + \Pi_e^{SIC} + \Sigma_e^R + \Sigma_e^{EC}, \quad (24a)$$

whereas the electron affinity X_t of the state t is given as

$$X_t = -\epsilon_t^{LD} + \Pi_t^{SIC} + \Sigma_t^R + \Sigma_t^{EC}. \quad (24b)$$

Figure 20 depicts schematically the various terms for the e-t crystal-field excitation in the Si:TA system. It establishes the notations used in this discussion and the relevant vacuum reference energy.

Our calculation provides $\Delta\epsilon_{te}^{LD}$ (Fig. 9 and Tables II and III). The self-interaction correction Π_i^{SIC} can be calculated simply in closed form,[58] whereas the relaxation self-energy Σ_i^R can be evaluated numerically in a number of ways.[59] The extra correlation self-energy Σ_i^{EC} is extremely complicated to evaluate, and relatively little is known about its energy dependence and even its sign. We will have to neglect it in the forthcoming discussion.

Our objective here is to provide a semiquantitative analysis of the effects produced by the first three terms on the right-hand side of Eq. (23), as well as their modification due to spin polarization (Sec. IV B). A detailed derivation of these quantities was given in Ref. 58. We now summarize some of the salient features pertinent to our present problem. For simplicity of presentation, we discuss the process of ionization.

(i) The self-interaction correction Π_i^{SIC} is positive and usually larger in magnitude than the negative relaxation energy[58] Σ_i^R. Upon ionization to the vacuum, the energy of an impurity gap level will hence move to a more negative value so that its actual ionization energy will be *larger* than $-\epsilon_i^{LD}$ by the quantity $\Pi_i^{SIC} + \Sigma_i^R$. [Since $\Pi_i^{SIC} = 0$ in the Hartree-Fock (HF) model, the ionization energy to the vacuum is *smaller* in that model than the negative of the orbital energy $-\epsilon_i^{HF}$ by just Σ_i^R, in contrast to the LD result where the corrected ionization energy is *larger* than $-\epsilon_i^{LD}$.] If the excitation is into a partially occupied final state, its energy will move closer to the vacuum level by $\Pi_f^{SIC} + \Sigma_f^R$. If, however, the final state is initially empty (virtual state), an electron excited into it will experience only the reduced screening due to the hole left behind, and consequently its energy will move away from the vacuum level. This effect constitutes the polarization piece of the relaxation self-energy Σ_f^R.

(ii) Slater's transition-state (TS) concept[49] approximates the first three terms on the right-hand side of Eq. (23) by the eigenvalue difference

$$\Delta E_{te}^{TS} = \epsilon_t^{LD}[e^{n-1/2}t^{m+1/2}] - \epsilon_e^{LD}[e^{n-1/2}t^{m+1/2}] .$$

(25)

This is often a good approximation and leads, for an ionization process, to an energy ΔE_i^{TS} *larger* than the Koopmans estimate $-\epsilon_i^{LD}$. This is illustrated by the calculation of Fig. 15(b), where, relative to the vacuum level the transition-state energy for the $e^3 \rightarrow e^2$ process (given by $\epsilon^{LD}[e^{2.5}] = E_{VB} + 0.02$ eV) is seen to be farther below the vacuum level than the Koopmans estimate (given by $\epsilon^{LD}[e^3] = E_{VB} + 0.1$ eV). The self-interaction and relaxation corrections $\Pi_e^{SIC} + \Sigma_e^R$ for ionizing the e^{CFR} level of Si:Mn is hence of the order of 0.1 eV. Even if our estimate is off by 100%, the correction could not be more than 0.2 eV. Notice, however, that these corrections are expected to be substantially larger for the hyperdeep t_2^{CFR} and e^{CFR} levels of the heavy impurities.

(iii) Both the self-interaction and the relaxation corrections decrease rapidly as the wave function delocalizes spatially.[58] In particular, a diffused conduction-band state is expected to carry only small corrections. Similarly, the t_2^{DBH} state, being a dangling-bond state and hence considerably less localized than the e^{CFR} orbital, will have a smaller correction. We conclude that both donorlike transitions from e^{CFR} to the CB, and acceptorlike transitions from t_2^{DBH} to the VB, will occur at energies that exceed the LD eigenvalue differences by a rather small correction, about $\Pi_i^{SIC} + \Sigma_i^R \approx 0.1$ eV.

It is interesting to contrast the correction calculated for Si:Mn to that of atomic Mn. For the ionization of a high-spin d electron from Mn d^5s^2 we get $\Pi_d^{SIC} = 8$ eV and $\Sigma_d^R = -3.9$ eV (Ref. 59), or $\Pi_d^{SIC} + \Sigma_d^R = 4.1$ eV (Ref. 60). For the ionization of the more extended $4s$ state, we find $\Pi_s^{SIC} = 2.7$ eV and $\Sigma_s^R = -0.8$ eV, or $\Pi_s^{SIC} + \Sigma_s^R = 1.9$ eV. The significant result is that the leading corrections for the removal energies from gap levels in Si:TA are strongly reduced relative to the free-atom values (primarily due to the occurrence of a solid-state node in the wave function of the impurity state). The more delocalized nature of the defect orbitals outside the central cell (cf. Tables II and III), and the strong screening attenuation discussed in Sec. III F contribute to this reduction. Furthermore, since the corrections to the removal energies vary monotonically with the degree of orbital localization[58] and since the latter changes smoothly across the $3d$ impurity series (Tables II and III), we expect the chemical trends shown in Fig. 9 to be preserved.

B. The ground-state electronic configuration

The ground-state electronic configuration in the local-density model corresponds strictly to a population scheme obeying the *aufbau* principle, i.e., occupying levels in increasing order of one-electron energies ϵ_i^{LD} without leaving holes behind. Incomplete shells (e.g., atomic Mn d^5s^2) exist only when the one-electron energy of the incompletely occupied orbital exceeds that of the next shell upon adding an electron to the former shell. In accordance with the LD model, we have fixed the occupation numbers of the neutral TA defects according to the *aufbau* principle, as shown in Fig. 9. In the following we will denote one-electron configurations in the order of increasing one-electron energies; e.g., for Si:Mn we have e^3t^0 as the LD ground state, and for Si:Cr we have e^2t^0, etc. Notice that such LD *aufbau* configurations need not obey Hund's rule for maximum ground-state spin multiplicity. A promoted Hund-type configuration for Si:Mn, such as e^2t^1, could become a ground state only if the order of energy levels is reversed.

Obviously, the LD *aufbau* configuration need not be the ground state if a different approximation is used for the single-particle potential. An example for such an improved approximation is given by the local-spin-density (LSD) potential. In this approximation, different spin states (denoted $+$ and $-$) experience different potentials and hence lead to different spin-up (ϵ_+) and spin-down (ϵ_-) orbital energies due to exchange interactions. We will next briefly explore the condition under which such a spin splitting may alter the ground-state configuration of the defect. The effect of the spin polarization on the e-t crystal-field splitting is depicted schematically in Fig. 20.

To explore whether spin polarization can alter the LD-type e^nt^0 configuration, one needs to promote it to a $e^{n-1}t^1$ configuration. Figure 20 and the discussion of the preceding section show that this widens the e-t crystal-field splitting Δ_{CF}. Each of the one-electron levels may now exhibit an exchange splitting denoted in Fig. 20 by $\Delta_x(i)$ (depending on whether the excited electron has its spin up or down, one will get a triplet or a singlet). Since the exchange splitting reduces quadratically with decreased orbital localization, hostlike conduction-band resonances are expected to have a small exchange splitting. Clearly, the ground state will remain at its LD configuration e^nt^0 as long as

$$\Delta_{CF}[e^{n-1}t^1] \geq (\Delta_x[t^1] + \Delta_x[e^{n-1}])$$

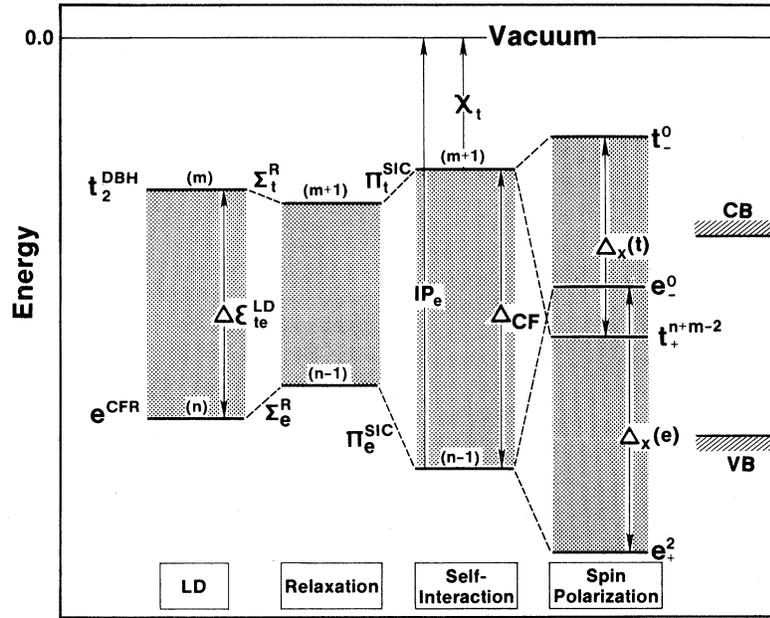(i.e., as long as the crystal-field splitting exceeds the exchange splitting). Conversely, if the exchange

FIG. 20. Schematic energy-level diagram showing the effects of relaxation, self-interaction correction, and spin polarization on the $e^{CFR} \rightarrow t_2^{DBH}$ crystal-field excitation. The zero of energy is at the vacuum level (about 5 eV above the VB maximum). An e_- vs t_+ spin-polarization level inversion is assumed. See text for explanation of symbols.

splitting exceeds the crystal-field splitting, the *aufbau* principle will produce for Si:Mn a Hund ground state $e_+^2 t_+^1 e_-^0 t_-^0$ shown in Fig. 20. Using again free transition atoms as a guide, we can calculate d-orbital exchange splittings. We find, for example, that the LSD exchange splitting[60] for Fe is 3.4, 2.5, and 1.6 eV for the $d^6 s^2$, $d^7 s^1$, and $d^8 s^0$ configurations, respectively. Since the exchange splitting is a two-electron integral, it scales with the square of the localization parameter of the orbital [Eq. (11)]. Hence with the use of atomic exchange splittings and the impurity localization parameters given in Tables II and III one can estimate the e and t exchange splittings not to exceed about 1.5 and 0.3 eV, respectively [e.g., for substitutional GaAs:Cr, Dimmock and Hemstreet[19(b)] found an exchange splitting of the e gap level of 0.6 eV, compared with the atomic exchange splitting for $d^5 s^1$ of 3.7 eV (Ref. 60)]. Clearly, a level reversal due to exchange splitting can occur only when the one-electron e-t energy gap does not exceed about 2 eV. This excludes all impurities heavier than Fe (cf. Fig. 9). From the remaining impurities, Cr and V already have a Hund ground state, and Ti has unoccupied e and t levels. The only reasonable candidates for a level reversal are hence Si:Fe and Si:Mn. Their one-electron e-t splittings are 1.7 and 1.4 eV, respectively, in the $e^4 t^0$ and $e^3 t^0$ LD configurations. Recalling that a promotion to the $e^3 t^1$ or $e^2 t^1$ configu-

rations will further widen the crystal-field splitting (cf. Fig. 20), it would appear that a spin-polarization level reversal is (marginally) possible for these systems. A complete spin-polarized calculation is needed to establish this.

Notice that the strength of the crystal-field (CF) splitting Δ_{CF} that determines the possibility of a spin-polarization configuration inversion depends on the impurity site. The crystal-field potential (i.e., the potential exerted at the impurity site by all atoms except the impurity) in silicon is far stronger at the substitutional site than in the tetrahedral interstitial site.[14] Indeed the e^{CFR}-t_2^{DBH} one-electron splitting for Si:Co was calculated in the muffin-tin cluster model to be 2.1 eV for the substitutional site[20] (between the $2T_2$ and $0E$ levels of Ref. 20), whereas the e-t splitting of the gap levels in the interstitial site was found[21] to be less than 0.1 eV. Clearly, even a very small exchange splitting will cause a level reversal for the interstitial sites,[21] whereas only a relatively large exchange splitting could reverse the level ordering for the substitutional sites (the atomic exchange splitting[60] for Co $d^8 s^1$ is only 1.8 eV). Furthermore, the spherical approximation to the potential used in the cluster calculations[20,21] is likely to severely underestimate the crystal-field potential, since the latter is determined by the $l = 3$ and $l = 4$ nonspherical components. (In the point-ion approximation the tetrahedral—crys-

tal-field splitting depends on the $l = 4$ potential component alone[14].) The nonspherical components[34,40,56] are known to be significant in an open covalent structure. This is particularly true for the tetrahedral interstitial site which in the muffin-tin approximation (with sphere radii equalling half an Si—Si bond length[20,21]) is characterized by a weak anisotropy in the potential. When the full nonspherical potential is retained the crystal field is likely to be substantially larger, inhibiting further a level inversion due to spin polarization. (Our recent QBCF calculations for interstitial Si:Fe (Ref. 23) indicate that the crystal-field splitting obtained in the cluster model[21] is underestimated by a factor of 4—5.)

From our calculated energy-level scheme (Fig. 9) and the foregoing discussion we can characterize the one-electron configuration, total spin S, and the lowest crystal-field term (in the weak-coupling limit) for the neutral defects as

$$V[e_+^1, S = \tfrac{1}{2}, {}^2E] ,$$

$$Cr[e_+^2, S = 1, {}^3A_2] ,$$

$$Mn[e_+^2 t_+^1, J = \tfrac{1}{2}, \tfrac{3}{2}, \tfrac{5}{2}, {}^4T_1] ,$$

and

$$Fe[e_+^2 t_+^2, J = 1, 2, 3, {}^5T_2] .$$

Here J is the total momentum, and a spin-polarization inversion was assumed for Mn and Fe. To lowest order, the various charged states are derived from this configuration simply by adding or subtracting electrons, e.g., V^- is like Cr^0 and Cr^+ like V^0, etc. For the heavier impurities, the e^{CFR} is too deep to be affected by spin polarization, and the

electronic configurations for the neutral defects are $Zn[t_+^3 t_-^1, S = \tfrac{1}{2}]$ (double acceptor), $Cu[t_+^3, S = \tfrac{3}{2}]$ (triple acceptor), and $Ni[t_+^2, S = 1]$ (quadruple acceptor).

V. COMPARISON WITH OBSERVED CHEMICAL TRENDS

In this section we compare qualitatively the overall chemical trends evident from our calculations with those observed experimentally. It is felt that a detailed and quantitative comparison is not warranted at this initial stage both because the interpretation of the experimental results are still clouded by a number of factors discussed in the Introduction, and because of an incomplete account of many-body effects in the present calculations (cf. Sec. IV).

A. DLTS data

Figure 21 depicts the deep-level transient spectroscopy (DLTS) activation energies for $3d$ impurities in Si, recently measured by Kimerling, Benton, and Rubin[4] and by Graff and Pieper.[5] The hole-trap (donor) energies correspond to transitions from the valence band to the defect levels, whereas the electron-trap (acceptor) energies correspond to excitations from the defect levels into the conduction band. The most recent values[5] for hole-trap ionization of a neutral TA defect X^0 to a positively charged TA defect X^+ are $E_{CB} - 0.28$ eV, $E_{CB} - 0.22$ eV, $E_{CB} - 0.42$ eV, and $E_{VB} + 0.385$ eV for Ti, V, Cr, Mn, and Fe, respectively.[61] The six electron traps at $E_{VB} + 0.6$ eV as well as the five hole traps at $E_{VB} + 0.3$ eV were assigned by Kimerling et al. as Au-related defects. Similarly, the hole
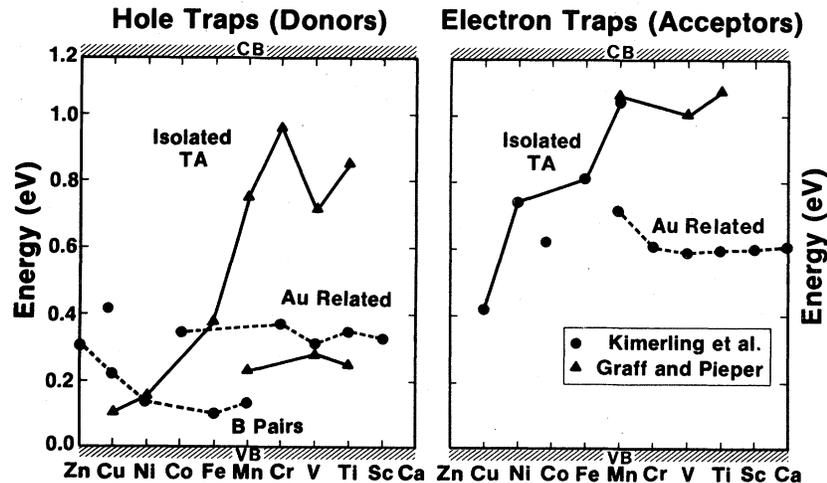

FIG. 21. Experimentally observed DLTS electron and hole traps for transition-atom-doped silicon (Refs. 4 and 5). The assignment of the various transitions and the lines connecting different points is tentative (Refs. 4 and 5).

traps near the top of the VB were shown[3,4] to be due to complexes with acceptor impurities such as boron (the most recent values are[61] $E_{VB}+0.28$ eV, $E_{CB}-0.55$ eV, and $E_{VB}+0.1$ eV for Cr, Mn, and Fe, respectively). The remaining transitions were suggested tentatively to arise from isolated TA impurities. We address only these transitions. The two sets of data[4,5] containing a total of 35 points, agree only on two points: the Mn electron trap and the Ni hole trap. The hole-trap result for Fe obtained by Graff and Pieper agrees with the results of Feichtinger.[62]

The striking result in the experimental study of Kimerling et al.[4] is the clear demarcation between the levels of elements to the right and to the left of Mn. The $3d$ elements heavier than Mn (Mn through Zn) show a single-electron trap (acceptor level) which starts up close to the CBM and drops monotonically as the impurity becomes heavier. This trend is found for our t_2^{DBH} acceptor level in Fig. 9. Recalling from the discussion of Sec. IV that the correction to the LD energies (Fig. 9) required for acceptors is a ~ 0.1-eV shift toward the CB, the calculated results are in reasonably good agreement with experiment. Graff and Pieper[5] have identified two more data points for the electron traps of Si:V and Si:Ti showing that the level has saturated, as indicated in the calculation of Fig. 9. The interpretation of the substitutional Si:Cu as a triple acceptor (viz., Fig. 9) agrees with the Hall-effect data of Hall and Racette.[63]

For the $3d$ elements lighter than Mn (Cr through Sc), Kimerling et al. found no electron trap in the upper part of the band gap, as indeed suggested by the eventual disappearance of the t_2^{DBH} level into the conduction band in Fig. 9. The donor levels associated with the e^{CFR} levels were not observed by Kimerling et al. However, Graff and Pieper did note a sequence of donor levels starting for the low-Z impurities below the CBM and dropping in energy as one goes to the heavier elements, a behavior similar to that shown by our e^{CFR} level (Fig. 9). We note that the overall agreement we find between the observed and calculated chemical trends is not based on either of the two assumptions made by Kimerling et al. to interpret the data[4]: that the elements heavier than Mn are interstitials, or that substitutional impurities are characterized by a $d \to s$ promotion. In fact, our calculation shows the opposite to be true (cf. Sec. III E). However, we must note that one cannot conclude from the agreement in the chemical trends that the observed levels are due to substitutional defects before we complete an identical calculation for all interstitial impurities to examine if the present geometry is the only one that agrees with the experimental trends.

B. EPR data

We next turn to compare our results with the chemical trends evident from the EPR studies.[3] We will first briefly discuss the line of argument used originally by Ludwig and Woodbury[3] (LW) to deduce their model from the data.

The salient features of their data are the following: First, the observed anisotropy of the g values and the hyperfine-interaction parameters clearly suggest the impurities to be in a cubic site. This means that as perceived by an EPR experiment, there is no measurable impurity-impurity, or impurity-dopant association and static lattice relaxation that alters the impurity-site symmetry (i.e., no static Jahn-Teller distortions). Second, the total spin S of the system was obtained both from counting the number $2S$ of the fine-structure lines (when resolved, e.g., Cr, but not V) and from fitting the observed frequency of the electron-nuclear double resonance (ENDOR) transitions to the transition calculated from the known impurity nuclear momentum and from an assumed value for S. This provides directly the number $2S$ of unpaired electrons.

In interpreting these data one needs to select the number and the ordering of the energy levels that will accomodate the impurity electrons. In the absence of detailed calculation, it was reasonable to assume at the time[3] that there are only two important levels t_2 and e, and that they derive from the impurity atomic $3d$ level. A simple crystal-field argument, neglecting hybridization, next-nearest-neighbor effects, charge redistribution, and incorporating only the first leading angular momentum term in the wave-function expansion, predicts for a T_d symmetry an e level below a t_2 level for substitutional defects.[14] This level ordering was assumed by LW in interpreting their results for substitutional defects. A reversed order (e above t_2) was assumed for interpreting the results that pertain to interstitial impurities. The LW level ordering for substitutional impurities implies that their e and t_2 levels correspond, in a complete calculation, to our e^{CFR} and t_2^{DBH} states. However, we find that the e and t_2 pair that actually evolves from the atomic $3d$ levels (at least at the high-Z limit) is clearly e^{CFR} and t_2^{CFR}, not e^{CFR} and t_2^{DBH}, and that their ordering is the reverse of that assumed by LW. It is hence a priori not impossible that the LW interpretation of the interstitial results (e above t_2) pertains (perhaps in part) to substitutional impurities. To make a close contact with the LW interpretation we will assume, however, in what follows the LW assignment of substitutional impurities levels, i.e., e^{CFR} and t_2^{DBH}.

The logic of the model is then as follows: If the electronic configuration of the free atom is $d^n s^m$, but only $2S < n + m$ spins are observed in the impure system, one needs to accommodate $n + m$ electrons in the t_2 and e levels, maintaining their appropriate ordering, so that $n + m - 2S$ of the spins are neutralized ("annihilated"). One can annihilate spins either by pairing electrons with opposite spins, by ionizing the electrons altogether, or by a combination of these two operations. Notice, however, that more than one solution exists to this problem. LW chose one that yields the charge state Q_{net}, consistent with the known sample doping ($Q_{net} > 0$ for p type and $Q_{net} < 0$ for n type).

Few other experimental quantities are relevant to constructing a consistent model. Simple arguments[14] show that all paramagnetic interactions that depend on the matrix elements of the orbital angular momentum, spin-orbit interactions, and hyperfine interaction among t_2 states contain opposite contributions from p- and d-orbital components. If the hybridization of p content into the TA d states is large enough, the anisotropic hyperfine interactions would even be cancelled, as observed, for instance, for tetrahedral Cu in the organic complex copper $(\alpha,\alpha'$-Br) dipyrromethene,[45] in which the p content is about 20–30%. From Fig. 11(b) and Tables II and III, we can immediately see that if the paramagnetic electrons are accomodated in the t_2^{DBH} level, large p-d "covalency" effects are expected, whereas if they are in the e^{CFR} level, such effects will be small. Further, Simanek and Müller,[64] Van Wierignen,[65] and Matumura[66] have pointed out that the hyperfine-interaction constant A for transition atoms decreases in magnitude with increasing covalency, while the g value increases. For instance, Mn^{+2} has the value $A = -97.8 \times 10^{-4}$ cm^{-1} in a strongly ionic hostlike CaF_2. The value is strongly reduced to $A = -40 \times 10^{-4}$ cm^{-1} in a covalent hostlike silicon. These trends are to be kept in mind in analyzing the data for Si:TA.

We point out that although the LW analysis is plausible, it is not unique. With the use of our calculated energy-level scheme (Fig. 9), it appears that much of the data interpreted by LW as pertinent to interstitial impurities can be equally well interpreted if one assumes substitutional impurities. This is illustrated below by a few examples. In this qualitative argument we need only the calculated level ordering for neutral defects (Fig. 9). We recall (Sec. IV) that relative to the vacuum level, addition (removal) of electrons causes a decrease (increase) in the binding energy (Fig. 21), and that spin polarization lowers the energy of the majority-spin levels. Hence, assuming an interstitial level ordering, LW interpreted the Si:Ni spectra showing $2S = 1$ to be

due to the configuration $t_+^3 t_-^3 e_+^2 e_-^1$ with $Q_{net} = 1$ (hence d^9 compared to $d^8 s^2$ in the atom). We note from Fig. 9 that a substitutional model predicts that for neutral Si:Ni^0, all but two electrons are below the VBM. The last two electrons occupy the t_2^{DBH} acceptor level. Hence, assuming also $Q_{net} = 1$, we predict a t_+^1 configuration (where t refers to the t_2^{DBH} state) consistent with $2S = 1$. However, our assignment does not require the artificial assumption of an $s \to d$ promotion. Because of its ionization, the energy of this state would be slightly *below* (cf. Fig. 20) the neutral t_2^{DBH} level for Si:Ni indicated in Fig. 9. The strong covalency effects of t_2^{DBH} [Fig. 11(b)] are expected to reduce the hyperfine constant. Indeed, LW (Ref. 3) have observed $A = 3.6 \times 10^{-4}$ cm^{-1} for this system.

If we assume, following LW, that the maximum-spin (Hund's-rule) configuration applies (i.e., the spin-polarization splitting exceeds the crystal-field splitting, cf. Fig. 20), even more data can be reconciled with a substitutional model. For example, the $2S = 3$ value observed for Si:Fe$^+$ under p-type doping, which LW interpreted to arise from an interstitial $t_+^3 t_-^2 e_+^2 e_-^0$ configuration with $Q_{net} = 1$, can be explained in a substitutional model as resulting from a $e_+^2 t_+^1 ({}^4T_1)$ configuration also with $Q_{net} = 1$. Since the highest occupied spin orbital here is the t_2^{DBH}, strong covalency effects due to p-d hybridization [cf. Fig. 11(b)] would be manifested on the spin-orbit and hyperfine interactions.[3] Indeed, $A = 3 \times 10^{-4}$ cm^{-1} was observed.[3] The same holds for the $2S = 3$ value observed for Si:Mn^0 (type of doping unspecified); the $e_+^2 t_+^1 ({}^4T_1)$ substitutional configuration with $Q_{net} = 0$ agrees with the observed data as much as the interstitial configuration $t_+^3 t_-^2 e_+^2 e_-^0$ agrees with $Q_{net} = 0$. The latter two examples are hence predicted to have an orbitally degenerate ground state and therefore a large g shift, as indeed observed[3] ($g = 3.524$ for Fe$^+$ and 3.362 for Mn^0).

For the EPR spectra, which LW assumed to be pertinent to substitutional impurities (Cr^0, Mn$^+$, and Mn^{2-}), the two interpretations are identical. For instance, for the $2S = 1$ spectra observed for Si:Mn$^+$ and Si:Cr^0 under the material preparation conditions said to favor substitutional defects, both models predict an $e_+^2 ({}^3A_2)$ configuration with $Q_{net} = 1$ for Si:Mn^0 and $Q_{net} = 0$ for Si:Cr^0. The observed[3] g values for these orbitally singlet states are indeed close to the free-electron value $g = 2$ ($g = 1.9962$ for Cr^0 and 2.0259 for Mn$^+$). Since these e states are only weakly hybridized with host states (cf. Table III), small covalency effects are expected to be manifested on the EPR parameters.[3] For the Si:Mn^{2-} spectra showing $2S = 5$, both models predict $e_+^2 t_+^3 ({}^6A_1)$ with $Q_{net} = -2$. Since, however, the highest occupied spin state is now the

strongly hybridized t_2^{DBH} state [cf. Fig. 11(b)] and not the e^{CFR} state (as in Mn^+), strong covalency effects are expected. Indeed, the hyperfine constant $|A|$ was found[3] to be dramatically smaller for Mn^{2-} $(A = -40.5 \times 10^{-4}$ cm$^{-1})$ than for Mn^+ $(A = -63 \times 10^{-4}$ cm$^{-1})$. The latter value is close to that expected for ionic hosts such as ZnS.[70]

We finally note that transition-metal impurities in silicon have recently received a renewed interest due to the possibility that they determine the interface properties (e.g., Shottky-barrier heights) in silicon-silicide interfaces.[67,68] It is hoped that more photo-emission experiments will be conducted to elucidate the energetics involved with both the deep and the hyperdeep levels predicted in this calculation (Fig. 9).

VI. SUMMARY

The electronic structure of substitutional 3d transition-atom (TA) impurities in silicon was calculated self-consistently in the local-density approximation. The salient features of the results are the following.

(i) Substitutional 3d impurities in silicon introduce two new t_2 levels, a single e level, and three a_1 resonances. The energy of these levels varies monotonically with the impurity's atomic number Z between Zn and Ti. The t_2 levels occur in pairs: a lower, bonding t_2 crystal-field-resonance (CFR) state t_2^{CFR} and a higher, antibonding t_2 dangling-bond hybrid (DBH) state t_2^{DBH}. Between them we find the nonbonding e-type crystal-field-resonance state e^{CFR}.

The fivefold-degenerate 3d levels of the impurity atom split under the T_d site symmetry into a threefold-degenerate t_2 level and a twofold-degenerate e level. The self-consistently calculated variations in the t_2 defect levels can be mapped onto a simple three-level system in which the atomic level t_2^{TA} interacts with two t_2 effective host states t_2^{VB} and t_2^{CB} from the valence band (VB) and conduction band (CB), respectively. The three-level system produces a *bonding* t_2^{CFR} which starts up in Si:Zn near the bottom of the VB as a hyperlocalized (percentage of localization in the central cell $q = 90\%$), d-like (98% d) state that is controlled by the atomic t_2^{TA} energy, and rises up and broadens in energy as Z decreases to become for Si:Ti a delocalized $(q = 31\%)$, d-p hybrid (85% d, 15% p) pinned by the energy of t_2^{VB}. The *antibonding* t_2^{DBH} is a hybrid between the host s-p dangling bonds and the TA d orbitals. It starts up in Si:Zn just above the VB maximum as a delocalized $(q = 22\%)$, predominantly p-like $(77\%, p, 22\% d)$ state, that is, controlled by the

t_2^{VB} energy, and rises up in energy as Z decreases, disappearing for the first time into the CB for Si:Co ("t_2 threshold"). As it rises up in energy, its p-to-d hybridization ratio decreases and its degree of localization increases, becoming a 50%-50% p-d hybrid with $q = 50\%$ localization for Si:Fe. At the low-Z limit it is pinned by the repulsion (avoided crossing) from t_2^{CB}. Between the bonding and antibonding t_2 states we find the nonbonding narrow e^{CFR} that is nearly a pure d state throughout, starting up as a hyperlocalized $(q = 100\%)$ state for Si:Zn near the bottom of the VB, and delocalizing as it rises up in energy with decreasing Z (e.g., $q = 78\%$ for Si:Cr). It penetrates the band-gap region for the first time for Si:Mn ("e threshold"). It shows a crystal-field splitting from the t_2^{CFR} level that increases with Z. The three defect-induced a_1 resonances are hostlike. We find that for a given impurity, the system usually sustains *either* a t_2^{DBH} *or* a e^{CFR} occupied gap state, but not both, in contrast with previous assumptions. Further, the results show that Mn forms a demarcation point in the 3d series, delineating the lighter impurities (with an impuritylike gap state e^{CFR}) from the heavier impurities (with a hostlike gap state t_2^{DBH}).

(ii) The energies of the various defect levels are not pinned either by the energy of the silicon-vacancy gap level or by the valence- or conduction-band edges. The relevant host-pinning states for the t_2 defect levels are the t_2^{VB} states ($\sim E_{VB} - 1$ eV) and the t_2^{CB} states ($\sim E_{CB} + 2.7$ eV). These constitute the major host states with a maximum in their central-cell t_2 local density of states. They are separated by an effective t_2 band gap of ~ 4.8 eV, much like the Phillips band gap. This is the physically relevant bandgap for these defect levels, not the optical gap (1.15 eV).

(iii) Whereas the charge densities $\rho(\vec{r})$ of both the host crystal and the crystal containing the defect are extended in coordinate space and strongly anisotropic, the density and potential perturbations $\Delta\rho(\vec{r})$ and $\Delta V(\vec{r})$, respectively, are both localized and nearly isotropic. Although this property is also shared by the silicon vacancy, the density fluctuation $\Delta\rho(\vec{r})$ in the Si:TA system attains its localization within a distance typical of atomic dimensions (rather than 1–2 bond lengths), and has a strong gradient in the central cell.

(iv) The screened-potential perturbation $\Delta V(\vec{r})$ exceeds its asymptotic point-ion limit of $\Delta Z/r$ inside the central cell ("screening overshoot") due to the highly localized nature of the density fluctuation. The self-consistent reciprocal dielectric function $\epsilon_{L,L}^{-1}(\vec{r})$ consequently shows that the d waves are screened extremely efficiently relative to the more conventional point-ion dielectric screening mechan-

ism appropriate for shallow effective-mass impurities.

(v) The consequence of the special screening in this system is that whereas in the inner central cell ($r \lesssim 1.5$ a.u.) the screened perturbation ΔV follows the atomically determined order of the bare pseudopotentials, $\Delta V^{Zn} < \Delta V^{Ni} < \Delta V^{Cr}$, etc., (and hence that of the atomic $3d$ ionization energies); in the outer central cell ($r > 2$ a.u.) the order is reversed. This may lead to the interesting possibility that defect energy levels $\epsilon_i(Z)$ associated with very localized wave functions will exhibit an opposite ordering with the atomic number Z to that characteristic of delocalized wave functions (e.g., excited states).

(vi) Although impurities encompassing Zn to Ti span a wide range of values for the number of valence electrons per atom (12—14), the central-cell region in the crystal is always found to be electrically neutral. This is so despite the fact that many of the individual TA-like defect wave functions extend well outside the central cell. The local neutrality is facilitated by the redistribution of hostlike valence-band states that effectively screen the external perturbation.

(vii) Transition-atom impurities in silicon are known to sustain in a narrow energy range a large number of defect states corresponding to different charge states, in sharp contrast to the behavior of free TA. This remarkable behavior is explained in our model in terms of two effects: (a) The interaction with the semiconductor is shown to reduce the intra-atomic Coulomb repulsion U_{3d} by 2 orders of magnitude relative to the free-ion value due to the extremely effective nonlinear screening effects. (b) Despite the fact that, like in an isolated TA, most of the impurity charge is localized in the central-cell

region, this charge $Q(N)$ is remarkably stable against variations in the gap-level occupation number N. This leads to a pinning of many charged states in the gap region. This unusual "electronic elasticity" of the central-cell region results from a self-regulatory behavior analogous to homeostasis in biology.

(viii) The substantial attenuation in the intra-atomic Coulomb repulsion U_d relative to the free atom suggests that a one-electron model forms a good approximation for the Si:TA system. Furthermore, the small U_d value is found to lead to an $s \to d$ population crossover in the ground state of the Si:TA system: We find that whenever the d subshell can accomodate more than its electrons, the s electrons are promoted into the d subshell. Hence Si:Ni attains a noble-metal configuration ($\approx$ Ni d^{10}), and the lighter TA attain an $\sim d^{n+2}s^0$ configuration, in contrast to the atomic $d^n s^2$ or $d^{n+1}s^1$ configurations.

(ix) The overall trends in the experimentally observed DLTS activation energies across the $3d$ series agree with the calculated defect levels.

(x) The salient features of the EPR spectra can be reconciled with an interstitial impurity configuration, as originally suggested by Ludwig and Woodbury, but are also shown to be consistent, for the most part, with a substitutional geometry.

ACKNOWLEDGMENTS

[1]J. R. Davis, A. Rohtgi, P. Rai-Choudhury, P. Blais, and R. H. Hopkins, in *Proceedings of the 13th IEEE Photovoltaic Specialists Conference 1978, Washington D.C.* (IEEE, New York, 1978), p. 490; A. M. Salama, *ibid.* p. 496; R. H. Hopkins, R. G. Seidensticker, J. R. Davis, P. R. Choudhury, P. D. Blais, and J. R. McCormik, J. Cryst. Growth **42**, 493 (1977).

[2]E.g., see *Solar Energy Conversion*, edited by B. O. Seraphin (Springer, Berlin, 1979); A. A. Bergh and J. P. Dean, *Light Emitting Diodes* (Clarendon, Oxford, 1976).

[3](a) H. H. Woodbury and G. W. Ludwig, Phys. Rev. **117**, 102 (1960); (b) Phys. Rev. Lett. **5**, 96 (1960); (c) G. W. Ludwig and H. H. Woodbury, in *Solid State Physics*, edited by F. Seitz, H. Ehrenreich, and D. Turnbull (Academic, New York, 1962), pp. 13 and 263; (d) F. S.

Ham, Phys. Rev. **138**, A1727 (1965).

[4]L. C. Kimerling, J. L. Benton, and J. J. Rubin, in *Defects and Radiation Effects in Semiconductors, Oiso, Japan, 1980*, Proceedings of the 11th Conference on Defects and Radiation Effects in Semiconductors, edited by R. R. Hasiguti (IOP, London, 1981).

[5]K. Graff and H. Pieper, in *Semiconductor Silicon, 1981*, Proceedings of the Electrochemical Society, edited by H. R. Huff, R. J. Kriegler, and Y. Takeishi (The Electrochemical Society, Pennington, NJ, 1981), Vol. 81-5, p.331.

[6]J. W. Chen and A. G. Milnes, Annu. Rev. Mater. Sci. **10**, 157 (1980).

[7]Some of the earlier works on transition-atom impurities in semiconductors include the following: G. W. Ludwig, H. H. Woodbury, and R. O. Carlson, Phys.

Rev. Lett. 1, 295 (1958); G. Finn, Phys. Rev. 91, 754 (1953); R. O. Carlson, *ibid.* 104, 937 (1954).

[8]F. D. M. Haldane and P. W. Anderson, Phys. Rev. B 13, 2553 (1976).

[9]Some examples are the following: M. M. Akhmedova, L. S. Berman, L. S. Kostina, and A. A. Lebedev, Fiz. Tekh. Poloprovdn. 9, 2351 (1975) [Sov. Phys.—Semicond. 9, 1516 (1975)]; M. A. Abdugafurova, L. M. Kapitonova, L. S. Kostina, A. A. Lebedev, and S. Makhkamov, *ibid.* 9, 450 (1975) [*ibid.* 9, 685 (1975)]; S. A. Azimov, N. A. Sultanov, L. Islamov, and R. N. Nagmatov, *ibid.* 7, 1837 (1974) [*ibid.* 7, 1227 (1974)]; M. K. Bakhadyrkhanov, S. S. Nigmankhodzhaev, and A. T. Teshabaev, *ibid.* 10, 743 (1976) [*ibid.* 10, 441 (1976)]; A. A. Lebedev and M. M. Akhmedova, *ibid.* 10, 1894 (1976) [*ibid.* 10, 1130 (1976)].

[10]In the work of Ref. 3, "slow quenching" is achieved by dropping an encapsulated sample held at 1250 °C into water, whereas "rapid quenching" is obtained by blowing the sample into ethylene glycol. In the work of Ref. 4 the sample held at 1070 °C was pulled quickly into the cool region (27 °C) of the furnace. In the work of Ref. 5 a "rapid quenching" was obtained by dropping the hot sample into liquid nitrogen. Different quenching rates control the site preference (substitutional versus interstitial) of the diffused impurity. The substrates are similarly prepared in different techniques. In Ref. 3 the silicon is grown by a pull technique, in Ref. 4 by crucible float zone, and in Ref. 5 by the Czochralski method. It is likely that different native defects exist in these samples.

[11]G. W. Ludwig, H. H. Woodbury, and R. O. Carlson, J. Phys. Chem. Solids 8, 490 (1954).

[12]S. H. Muller, G. M. Tuynman, E. G. Sierverts, and C. A. J. Ammerlaan, Phys. Rev. B 25, 25 (1982).

[13]W. Bergholtz and W. Schröter, in *International Conference on Radiation Effects in Semiconductors, Nice, 1978*, edited by J. H. Albany (IOP, London, 1979), p. 572.

[14]F. S. Ham and G. W. Ludwig, in *Paramagnetic Resonance*, Proceedings of the First International Jerusalem Conference, Israel 1962, edited by W. Low (Academic, New York, 1963), Vol. 1, p. 131.

[15]K. B. Tolpygo, Fiz. Tverd. Tela. (Leningrad) 11, 2846 (1969) [Sov. Phys.—Solid State 11, 2304 (1970)].

[16]L. S. Kogan, and K. B. Tolpygo, Fiz. Tverd. Tela. (Leningrad) 16, 3176 (1974) [Sov. Phys.—Solid State, 16, 2067 (1975)]; *ibid.* 15, 1544 (1973) [*ibid.* 15, 1034 (1973)].

[17]A. B. Roitsin and L. A. Firshtein, Fiz Tverd. Tela. (Leningrad) 13, 63 (1971) [Sov. Phys.—Solid State 13, 50 (1971)].

[18](a) L. K. Ermakov, Fiz. Tekh. Poluprovodn. 12, 1230 (1978) [Sov. Phys.—Semicond. 12, 732 (1978)]; (b) R. L. Kleinhenz, V. A. Singh, and J. W. Corbett (unpublished).

[19](a) B. G. Cartling, B. Roos, and U. Wahlgren, Chem. Phys. Lett. 21, 380 (1973); B. G. Cartling, J. Phys. C 8,

3171 (1975); 8, 3183 (1975); (b) L. A. Hemstreet and J. O. Dimmock, Phys. Rev. B 20, 1527 (1979).

[20]L. A. Hemstreet, Phys. Rev. B 15, 834 (1977).

[21](a) G. G. DeLeo, G. D. Watkins, and W. B. Fowler, Phys. Rev. B 23, 1851 (1981); (b) G. G. Deleo, G. D. Watkins, and W. B. Fowler, Phys. Rev. B 25, 4962 (1982); 25, 4972 (1982), and (unpublished).

[22]J. E. Lowther, J. Phys. C 13, 3681 (1980).

[23]A preliminary report has been given in A. Zunger and U. Lindefelt, Phys. Rev. B 26, 846 (1982).

[24]W. Kohn and L. J. Sham, Phys. Rev. 140, A1133 (1965).

[25](a) A. Zunger and M. L. Cohen, Phys. Rev. B 18, 5449 (1978); 20, 4082 (1979); (b) P. K. Lam, M. L. Cohen, and A. Zunger, *ibid.* 22, 1698 (1980).

[26]M. Schlüter, J. R. Chelikowsky, S. G. Louie, and M. L. Cohen, Phys. Rev. B 12, 4200 (1975). We use the momentum-space form

$$v_{\mathrm{ps}}^{\mathrm{Si}}(q) = \frac{8\pi Z^v}{q^2 \Omega} \frac{1}{1+b}[\cos(aq)+b]e^{cq^4} \; \mathrm{Ry} \; ,$$

with $a = 0.79065$, $b = -0.35201$, and $c = -0.01807$. The valence charge Z^v is 4.0 for silicon. The unit-cell volume per atom Ω is 135.188 a.u.$^{-3}$.

[27]L. Ley, S. Kowalcyzk, R. Pollack, and D. A. Shirley, Phys. Rev. Lett. 29, 1088 (1972).

[28]R. R. Zucca, J. P. Walter, Y. R. Shen, and M. L. Cohen, Solid State Commun. 8, 627 (1970).

[29]W. E. Spicer and R. C. Eden, in *Proceedings of the Ninth International Conference of the Physics of Semiconductors, Moscow, 1968*, edited by S. M. Ryvkin (Leningrad, 1968), Vol. 1, p. 61.

[30]G. G. MacFarlane, T. P. McLean, J. E. Quarrington, and V. Roberts, Phys. Rev. 111, 1245 (1958).

[31]G. A. Baraff and M. Schlüter, Phys. Rev. B 19, 4965 (1979).

[32]J. Bernholc, N. Lipari, and S. T. Pantelides, Phys. Rev. B 21, 3545 (1980).

[33]U. Lindefelt and A. Zunger, Phys. Rev. B 24, 5913 (1981).

[34]U. Lindefelt and A. Zunger, Phys. Rev. B 26, 846 (1982).

[35]G. F. Koster and J. C. Slater, Phys. Rev. 96, 1208 (1954).

[36]P. Bendt and A. Zunger, Phys. Rev. B 26, 3114 (1982).

[37]H. P. Hjalmarson, P. Vogel, D. J. Wolford, and J. D. Dow, Phys. Rev. Lett. 44, 810 (1980); H. P. Hjalmarson, Ph.D. thesis, University of Illinois, 1979 (unpublished).

[38]R. S. Mulliken, J. Chem. Phys. 23, 1833 (1955).

[39]See review article by F. Bassani, G. Iadonisi, and B. Preziosi, Rep. Prog. Phys. 37, 1099 (1974).

[40]G. S. Painter, D. E. Ellis, and A. R. Lubinsky, Phys. Rev. B 4, 3610 (1971).

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

[42]R. Resta, Phys. Rev. B 16, 2717 (1977).

[43]C. F. Fischer, *The Hartree-Fock Method for Atoms* (Wiley, New York, 1977).

[44]A. Goldmann, Phys. Status Solidi B 80, 425 (1977); A. Zunger and M. L. Cohen, Phys. Rev. B 20, 1189 (1979).

[45]C. A. Bates, W. S. Moore, K. J. Standley, and K. W. H. Stevens, Proc. Phys. Soc. London 79, 73 (1962).

[46]These are one-electron crystal-field splittings, not excitation energies. The latter should include both orbital relaxation and self-interaction cancellation effects and would therefore be larger than the one-electron energy differences. This is discussed in Sec. IV A.

[47]J. C. Phillips, Bonds and Bands in Semiconductors (Academic, New York, 1973), p. 36.

[48]C. Corliss and J. Sugar, J. Phys. Chem. Ref. Data 6, 1253 (1977).

[49]J. C. Slater, The Self Consistent Field for Molecules and Solids (McGraw-Hill, New York, 1974), pp. 67–74.

[50]Haldane and Anderson (Ref. 8) have assumed that all of the five TA d states occur in the gap. Our calculation shows that usually either three or two such states can exist in the gap (cf. Fig. 9). Furthermore, in their calculation the valence-band density of states was assumed to be unperturbed (rigid band) and only the extent to which valence states are admixed into the gap level is calculated self-consistently using a number-conservation constraint. The underlying difficulty is that the perturbed impurity states are expanded in terms of atomic states whereas the perturbed host states are expanded in terms of the unperturbed band states. In fact, all perturbed states need to be expressed in terms of a single complete set (e.g., the unperturbed host states, cf. Ref. 33). Only a qualitative comparison with their results is hence possible. One of us (A.Z.) is grateful to Dr. Haldane and Dr. K. Subbarao for discussions on this point.

[51]The correction to U_{3d} due to the long-range Coulomb interactions (outside the integration radius $R = 7.5$ a.u.) is estimated perturbationally to be smaller than 0.1 eV, indicating that the effective Rydberg R^* is small.

[52]A. M. Stoneham and M. J. L. Sangster, Philos. Mag. B 43, 609 (1981).

[53]R. M. Nieminen and C. H. Hodges, J. Phys. F 6, 573 (1976).

[54]See footnote 34 in R. E. Allen and J. D. Dow, J. Vac. Sci. Technol. 19, 383 (1981). These authors suggest that a self-consistent solution of the one-particle problem introduces an unbalanced "favoritism" toward the valence bands since only the occupied states are used to construct the charge density and the effective potential.

This is said to introduce a "bias" into the calculation as defect states are constructed both from valence and conduction states. We disagree with this argument. We note that the self-consistency requirement is merely a statement of the variational principle; no other choice of $V[\rho]$ is guaranteed to produce the energy-minimizing solution. Other (non-self-consistent) guesses are simply ad hoc choices.

[55]D. A. Papaconstantopoules and B. M. Klein, Solid State Commun. 34, 511 (1980).

[56]R. V. Kasowski and O. K. Anderson, Solid State Commun. 11, 799 (1972).

[57]J. Bernholc, S. T. Pantelides, N. O. Lipari, and A. Baldereschi, Solid State Commun. 37, 705 (1981).

[58]J. P. Perdew and A. Zunger, Phys. Rev. B 23, 5048 (1981).

[59]L. Hedin and A. Johansson, J. Phys. B 2, 1336 (1969); A. B. Kunz, Phys. Rev. B 12, 5890 (1975). In the Hartree-Fock (HF) approximation, the atomic relaxation energy can be calculated directly using Koopmans theorem, as

$$\sum_i^R = E_{HF}[e^{n-1}] - E_{HF}[e^n] + \epsilon_e^{HF} .$$

[60]We use a spin-polarized calculation with the Ceperley-Alder correlation [D. M. Ceperley and B. J. Alder, Phys. Rev. Lett. 45, 566 (1980)], as fitted by Perdew and Zunger in Ref. 61.

[61]We are grateful to Dr. K. Graff for communicating to us the most recent activation energies for these impurities.

[62]H. Feichtinger, in International Conference on Radiation Effects in Semiconductors, Nice, 1978, edited by J. H. Albany (IOP, London, 1979), p. 578; H. R. Szawelska and H. Feichtinger, J. Phys. C 14, 4131 (1981).

[63]R. N. Hall and J. H. Racette, J. Appl. Phys. 35, 379 (1964).

[64]E. Simanek and K. A. Müller, J. Phys. Chem. Solids 31, 1027 (1970).

[65]J. S. Van Wieringen, Discuss. Faraday Soc. 19, 11 (1955).

[66]O. Matumura, J. Phys. Soc. Jpn. 14, 108 (1959).

[67]G. Rossi, I. Abbati, L. Braicovich, I. Lindau, and W. E. Spicer, Solid State Commun. 39, 195 (1981).

[68]J. L. Freeouf, G. W. Rubloff, P. S. Ho, and T. S. Kuan, Phys. Rev. Lett. 43, 1836 (1979).