PHYSICAL REVIEW B VOLUME 25, NUMBER 4 15 FEBRUARY 1982

Evaluation of tight-binding models for deep defect levels in semiconductors

Vijay A. Singh,* U. Lindefelt, and Alex Zunger*
Solar Energy Research Institute, Golden, Colorado 80401

(Received 26 May 1981; revised manuscript received 31 August 1981)

Two of the fundamental assumptions used in contemporary tight-binding models for calculating the energies of deep defect levels in semiconductors are examined through comparison with a general model that avoids these assumptions. It is found that these assumptions are likely to invalidate the chemical trends predicted by the tight-binding models for the ordering of the impurity energy levels with atomic potentials.

In recent years, the defect Green's-function method[1] has been implemented in both a complex but precise self-consistent pseudopotential formalism,[2–5] and in a simplified tight-binding (TB) scheme,[6–9] to study the electronic structure of deep defect levels in semiconductors. By virtue of its simplicity and ease of application, the TB approach has been used to study a large number of systems, including not only defects in bulk solids[6,7] but also defects at free surfaces[9] and superlattices.[8] These applications have often produced a very interesting organization of the experimental data on defect energy levels in terms of simple semiclassical coordinates such as atomic ionization energies[7] and electronegativity differences.[8,9] Among others, one of the most intriguing regularities suggested by these models is the TB pinning phenomena: As the on-site central-cell defect potential perturbation ΔV becomes increasingly large (e.g., more positive for the repulsive defects discussed here) the gap-state defect levels $\epsilon_i(\Delta V)$ will saturate [e.g., $\partial\epsilon_i(\Delta V)/\partial\Delta V \simeq 0$] at a value $\epsilon_v(\infty)$ characteristic of the energy of the host vacancy (v); all of the repulsive ($0<\Delta V<\infty$) impurity levels ϵ_i would then lie *below* ϵ_v, i.e., be "pinned" to it. Although not supported by existing experimental data, this pinning idea has found wide applications in discussions of various interesting chemical trends.[7]

To achieve computational simplicity, a number of well-known approximations have been applied to the TB model. These were amply discussed in the literature (e.g., Refs. 10 and 11) and are often shown to lead to some uncertainties in the absolute values of the predicted energies. Here we address two different tacit assumptions which pertain to the predicted *chemical trends* (i.e., variation of calculated energies with chemical coordinates of the defect) rather than absolute energies:

(i) Irrespective of the strength of the central-cell potential ΔV, the localized defect wave functions $\psi_i(\vec{r})$ are expanded by a *small and fixed number* M of (extended) Bloch bands $\phi_j(\vec{k},\vec{r})$, i.e.,

$$\psi_i(\vec{r}) = \sum_{j=1}^{M} \sum_{\vec{k}} A_{ij}(\vec{k})\phi_j(\vec{k},\vec{r}) .$$

The number M is limited inherently by the underlying TB construction to a maximum of $M_{\rm TB}=8$ (Refs. 6, 8, and 9) or $M_{\rm TB}=10$ (Ref. 7). It is not related to the chemical identity of the impurity.

(ii) The central-cell defect potential ΔV, as sampled by the underlying basis orbitals $\{|\alpha\rangle\}$ is assumed to be "on site," i.e., it does not couple to the nearest-neighbor (or more distant) host atoms. In other words, if the basis $\{|\alpha\rangle\}$ is anchored to atomic sites, the perturbation matrix $\Delta V_{\alpha\beta}$ is assumed to be diagonal.

The conventional TB model itself does not provide any simple internal means of examing the physical validity of these two assumptions. In this report we examine the physical consequences of these two approximations using a general defect Green's-function approach recently developed.[5] Because neither of the approximations are required by this model, it provides a means of examining their effects. It is demonstrated that a model free of assumptions (i) and (ii) does not support the chemical trends predicted by the TB model.

In the TB Green's-function approach to deep defects one first obtains the band structure of the host crystal by fitting the lowest $M_{\rm TB}$ bands to a local pseudopotential band structure (e.g., that of Ref. 12). One then proceeds to express the Green's function matrix and the diagonal potential matrix in terms of these M_{TB} bands and solves for the defect energies $\epsilon_i(\Delta V, M_{\mathrm{TB}})$ and wave functions $\psi_i(\Delta V, M_{\mathrm{TB}})$ in the standard Slater-Koster fashion.[6–9] The defect wave function is hence spectrally constructed from host band wave functions extending up to the energy $\epsilon_{\mathrm{cut}}(M_{\mathrm{TB}})$ of the highest (M_{TB}th) host band included. The cutoff energy $\epsilon_{\mathrm{cut}}(M_{\mathrm{TB}})$ does not correspond to any physically special host energy. In general, however, for defect state ψ_i belonging to the ith representation of the system's point group (e.g., a t_2 representation for acceptors), all host states $\{\phi_j^0(\vec{k},\vec{r})\}$ containing the ith character (for t_2 states the $l = 1$, 2, 3, 4, . . . or p, d, f, g, . . . partial waves) could be coupled by the appropriate component of the perturbation potential ΔV. The TB model, limited inherently to small values of M_{TB}, forces the defect wave functions to be spectrally described by a narrow range of host states, irrespective of the chemical characteristics of the impurity (reflected in the radius and depth of ΔV).

Rather than limit ourselves to the lowest $M_{\mathrm{TB}}(=8-10)$ bands, we solve here for $\epsilon_i(\Delta V, M)$ as a function of $M \geq M_{\mathrm{TB}}$ using precisely the same host crystal pseudopotential band structure as inputed in most TB calculations. For each perturbation ΔV, the cutoff energy $\epsilon_{\mathrm{cut}}(M)$ is hence increased until further coupling of host states of the appropriate symmetry does not change the defect energy. The silicon host band structure is calculated self-consistently from a local pseudopotential, using a basis of 189 plane waves at Γ and an additional 200 plane waves via a second-order perturbation theory. The $i = t_2$ representation of these band wave functions is projected out and used directly to calculate the acceptor wave function and energy. To simplify our presentation we use a particularly simple form for the defect perturbation potential ΔV in the shape of a repulsive (i.e., acceptorlike) square well with radius R_0 and height V_0. This choice incorporates the two basic physical coordinates of a realistic defect potential: its "atomic size" R_0 and "atomic ionization energy" V_0. This simple choice will suffice for our illustration: We will indicate below the similarity of the results to those obtained by using for $\Delta V(\vec{r})$ a self-consistent pseudopotential. The value (and range) of V_0, hence, takes the role of "p-like" perturbation potentials in TB models.[6–9] Since the ionization energy of levels inside a square-well potential is monotonic with V_0, variations in V_0 are directly mapped into variations in the p-like impurity ionization energy used in TB models. The perturbation ΔV is not restricted to couple only impurity orbitals. The wave function of the t_2 acceptor gap state is then expanded in the standard Green's function approach by a dual representation consisting of local (impurity-anchored) orbitals and, independently, by the unperturbed host wave functions.[5] Sufficient local orbitals are included in the first expansion to make it converged for all perturbations $\Delta V(\vec{r})$ considered (15 radial Coulomb orbitals for each of the p, d, f, and g waves are used). All energy levels are converged to a precision better than 0.05 eV. One now asks how many bands M are required in the complementary expansion in terms of host wave functions such that the variation of ϵ_i with $\Delta V = \{V_0, R_0\}$ be reproduced with physically relevant precision (say, a tenth or two of the host energy band gap).

Figure 1 shows the energy of the $i = t_2$ gap level $\epsilon_{t_2}(\Delta V, M)$ in silicon as a function of the number M of host bands included in the wave-function expansion. We use a range of repulsive square-well defect potentials V_0 and a radius $R_0 = 2.2$ a.u., i.e., the Si covalent radius. The most striking result of this calculation is that the error in the defect ener-

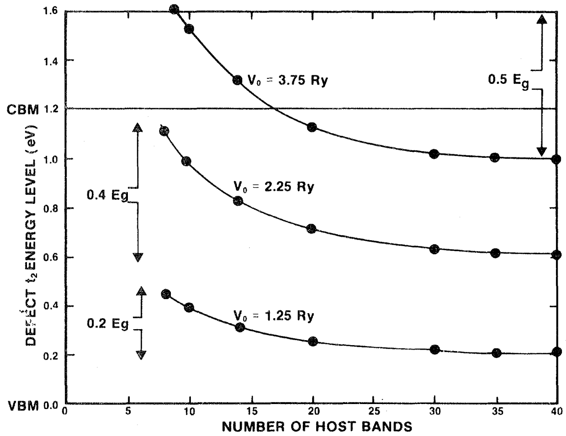

FIG. 1. Convergence of the t_2 acceptor level in silicon with respect to the number of host crystal bands used in the expansion of the defect's wave function. The substitutional acceptorlike defect potential is modeled by a repulsive square well with a radius $R_0 = 2.2$ a.u. and a height V_0. The host is modeled by a self-consistent local pseudopotential band structure (used in a fitted form in TB defect calculations). The errors made by including only 8–10 bands in the wave-function expansion (Refs. 6–9) are indicated with vertical arrows; they are indicated as a fraction of the host band gap E_g. VBM and CBM denote the valence-band maximum and conduction-band minimum, respectively.

gy level made by using only $M_{\mathrm{TB}}(=8-10)$ rather than a converged number M_{conv} of bands, is not only very large on the scale of physical interest (e.g., the band gap E_g) *but also varies strongly with the strength of the defect potential* ΔV. These errors are indicated in Fig. 1 as fractions of E_g by vertical arrows and range from 50% to 20% for different ΔV values. (For very large impurity radii, $R_0 \geq 4$ a.u., the convergence is faster. However, only Rb and Cs have such large radii. These elements were not considered in previous TB calculations). It appears, therefore, that one cannot hope to recover *physically useful trends* $\epsilon_i(\Delta V)$ in the defect energy levels (let alone absolute energy values) with the impurity's potential strength ΔV (i.e., chemical identity) for all impurities, using the same $M=M_{\mathrm{TB}}$. The "renormalization group" culation of Weigel *et al.*[13] in which host states in the conduction and valence states are discarded may also be susceptible to this criticism. The chemical trends $\epsilon_i(\Delta V)$ in such models[7–9,13] may have been interpreted over optimistically.

The physical reason behind this result is very simple and has been pointed out before[5]: the fundamental difficulty in any defect Green's function method is that one is forced to represent the localized defect wave functions $\psi_i(\vec{r})$ by a combination of often physically and chemically unrelated extended host Bloch functions $\{\phi_j^0(\vec{k},\vec{r})\}$. While this can be achieved *in principle* using a complete set $\{\phi_j^0(\vec{k},\vec{r})\}$, the number M of such wave functions needed *in practice* for a useful accuracy depends sensitively on the degree to which the properties of the host crystal differ from the properties of the impurity atom, i.e., on the nature of $\Delta V(\vec{r})$. The TB approach, on the other hand, insists that the defect states of all impurities—even if their atomic potentials differ by as much as $20-30$ eV from the host potential[7]—be represented by the same small number of host bands. (We have recently developed a new defect Green's function model—the "quasiband representation"[5]—in which this requirement is no longer needed. In this method a small number M of quasibands may be sufficient to obtain a converged representation for arbitrarily localized defect potentials.)

The results of Fig. 1 suggest that qualitative trends predicted by the TB model, like the notion of band-gap pinning, may need to be reexamined. To this end, we show in Fig. 2 the variation $\epsilon_i(R_0,V_0)$ of the t_2 gap defect level in silicon as a function of the strength V_0 of the perturbation, for various defect sizes R_0, *using a converged number M of host bands for all* $\{R_0,V_0\}$ *values.* The defect potential perturbation is allowed to couple with host atoms. Recall that the TB model which uses approximations (i) and (ii) predicts that, for a large V_0 (where V_0 is the difference between impurity and host diagonal matrix elements for representation i), the defect level $\epsilon_i(V_0)$ will show saturation at the band-gap vacancy level ϵ_v; all repulsive impurities (finite V_0) will be pinned to that saturated level at $\epsilon=\epsilon_v$. Figure 2 clearly shows that when approximations (i) and (ii) are relaxed, the energy level $\epsilon_i(V_0,R_0)$ merges into the conduction band for large enough values of V_0 (and all R_0); the level shows no "pinning" in the sense described in Ref. 7.[14] This strongly suggests that band-gap pinning is a property of the TB model rather than of nature. The behavior in Fig. 2 is suggestive of the predictions made by Krieger and Laufer[10] for the TB defect model.

Figure 2 suggests another interesting observation which focuses on the interplay between impurity size (R_0) and potential strength (V_0) in determining its energy $\epsilon_i(R_0,V_0)$: compared with smaller impurities, Fig. 2 shows that large impurities require a far smaller perturbation V_0 to split from the valence band, become deep traps, and merge eventually into the conduction band. It is hence possible that an impurity with a perturbing potential characterized by $\{V_0^{(1)},R_0^{(1)}\}$ will have its energy level *above* that of another impurity level having a stronger potential $V_0^{(2)}$, if $R_0^{(1)}$ is sufficiently larger than $R_0^{(2)}$. Hence, $\epsilon_1(V_0^{(1)},R_0^{(1)})>\epsilon_2(V_0^{(2)},R_0^{(2)})$ if $R_0^{(2)}<R_0^{(1)}$ but $V_0^{(2)}>V_0^{(1)}$. Assuming a typical radius $R_0^{(2)}=2.2$

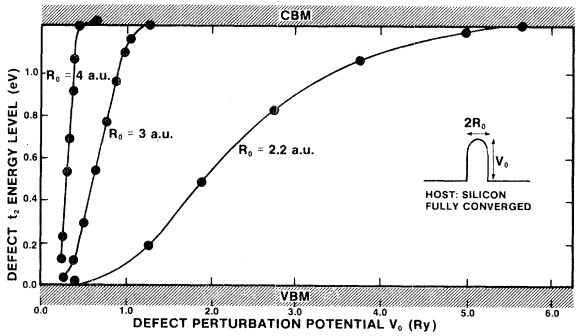

FIG. 2. Dependence of the acceptor t_2 energy level in silicon on the radius (R_0) and amplitude (V_0) of the perturbing potential. The energy levels are converged for each $\{R_0,V_0\}$ with respect to local orbitals, band wave functions, and Brillouin-zone sampling points to within better than 0.05 eV. Note that no gap pinning occurs.

a.u. (half a Si—Si bond length), one finds from Fig. 2 that the known unrelaxed Si vacancy t_2 energy $\epsilon_2=0.7$ eV (Refs. 3 and 4) is mapped[15] into the square-well model by taking $V_0^{(2)}=2.375$ Ry. One hence finds that an impurity with radius $R_0^{(1)}=3$ a.u. $>R_0^{(2)}$ will have its energy level ϵ_1 *above* ϵ_2 (e.g., $\epsilon_1=1$ eV for $V_0^{(1)}=0.875$ Ry) even if its potential amplitude $V_0^{(1)}$ is smaller than $V_0^{(2)}$. Hence, a large size impurity atom having a shallow potential well can form acceptor levels above the host vacancy level, if the host atom is sufficiently small. This clearly does not support the TB notion that "weaker" impurities will have their energy pinned below that of the "stronger" impurities.

The physical reason behind this observation is simple: in the TB model the defect is characterized by *a single variable;* the difference $\Delta V_{\alpha\beta}=V_{\alpha\alpha}-V_{\beta\beta}$ between diagonal TB matrix elements of the impurity and the host atom being replaced. Except in the work of Ref. 8, the diagonal elements $V_{\alpha\alpha}$ and $V_{\beta\beta}$ are not fixed uniquely by the fitting procedure of the band structure since the latter is invariant (up to a constant related to the host's work function[8]) under constant changes in $\{V_{ii}\}$. In a more complete model, however, the perturbing potential is characterized by a continuous function $\Delta V(\vec{r})$ and a complete basis set $\{|\alpha\rangle\}$. The physically relevant characteristics of $\Delta V(\vec{r})$ are then reflected in *two coordinates rather than one:* the radius R_0 and the potential strength V_0. [This information is contained in the matrix representation $\langle\alpha|\Delta V(\vec{r})|\beta\rangle$ if the basis $\{|\alpha\rangle\}$ is complete in the domain of $\Delta V(\vec{r})$, but not if it is grossly incomplete as in TB models.] Figure 2 shows that the predictions of a two-coordinate model[16,17] $\epsilon_i(R_0,V_0)$ (e.g., the ordering of the levels) can be the reverse of those based on a cruder one-coordinate model $\epsilon_i(V_{\alpha\alpha}-V_{\beta\beta})$. In particular, in a single-coordinate model with a diagonal perturbation matrix the vacancy is the strongest perturbation possible (i.e., $\Delta V=\infty$). Under these conditions, any repulsive impurity will have a weaker perturbing potential and consequently its energy level will be "pinned" below the energy of the vacancy level. In a more general two-coordinate model with a nondiagonal perturbation matrix, stronger perturbations are possible in principle by virtue of coupling to the host atoms as well as by steric size effects (i.e., variations in R_0). Impurity levels need not then be pinned to the vacancy level.

We conclude that while the TB defect model may be used as an expedient interpolative tool, greater caution must be exercised when applying it to deduce chemical regularities over large ranges of chemical coordinates (viz., atomic sizes and electronegativities). While extensive and careful parametrization of the model can lead to a number of valid predictions, its inherent dependence on the representation of localized states by a small number of extended host states considerably limits its applicability.

One of us (V. S.) acknowledges support by SERI through Subcontract No. HS-0-9188-4. Useful and stimulating comments by J. Joannopoulus and J. Bernholc are gratefully acknowledged.

---

[*] Also at: Department of Physics, University of Colorado, Boulder, Colorado 80309.

[1] G. F. Koster and J. C. Slater, Phys. Rev. 96, 1208 (1954); L. M. Roth, Ph.D. thesis, Harvard University, 1957 (unpublished); J. Callaway, J. Math. Phys. 5, 783 (1964).

[2] M. Jaros and S. Brand, Phys. Rev. B 14, 4494 (1976); M. Jaros, J. Phys. C 8, 2455 (1975); M. Jaros, C. O. Rodriquez, and S. Brand, Phys. Rev. B 19, 3137 (1979).

[3] J. Bernholc, N. O. Lipari, and S. T. Pantelides, Phys. Rev. Lett. 41, 895 (1978); Phys. Rev. B 21, 3545 (1980); J. Bernholc, S. T. Pantelides, N. O. Lipari, and A. Baldereschi, Solid State Commun. 37, 705 (1981).

[4] G. A. Baraff and M. Schlüter, Phys. Rev. Lett. 41, 892 (1978); Phys. Rev. B 19, 4965 (1979); G. A. Baraff, E. O. Kane, and M. Schlüter, *ibid.* 21, 5662 (1980).

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

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

[7] H. P. Hjalmarson, P. Vogl, D. J. Wolford, and J. D. Dow, Phys. Rev. Lett. 44, 810 (1980); H. P. Hjalmarson, R. E. Allen, H. Büttner, and J. D. Dow, J. Vac. Sci. Technol. 17, 993 (1980); R. E. Allen, H. P. Hjalmarson, H. Büttner, P. Vogl, D. Wolford, O. F. Sankey, and J. D. Dow, Int. J. Quantum Chem. S14, (1980).

[8] S. Das Sarma and A. Madhukar, J. Vac. Sci. Technol. 17, 1120 (1980); Solid State Commun. 38, 183 (1981).

[9] M. S. Daw and D. L. Smith, Phys. Rev. B 20, 5150 (1979); M. S. Daw and D. L. Smith, Appl. Phys. Lett.

36, 690 (1980).

[10] J. B. Krieger and P. M. Laufer, Phys. Rev. B 23, 4063 (1981).

[11] D. W. Bullet, in *Solid State Physics,* edited by H. Ehrenreich, F. Seitz, and D. Turnbull (Academic, New York, 1980), Vol. 35, p. 129.

[12] J. R. Chelikowsky and M. L. Cohen, Phys. Rev. B 14, 556 (1976).

[13] C. Weigel, S. T. Chui, and J. W. Corbett, Phys. Rev. B 18, 2377 (1978).

[14] We examined whether the results of Figs. 1 and 2 depend in any qualitative way on the fact that we model the perturbation $\Delta V(\vec{r})$ by a rounded square well rather than by a "realistic" potential. We have repeated the results of Fig. 1 by using for the perturbation $\Delta V(\vec{r})$ the self-consistently screened pseudopotential perturbation for a Si vacancy. Similar to what we show in Fig. 1, we found that by using only $M=8-10$ bands one makes errors in the t_2 energy gap level of 25% of the host band gap relative to the converged results ($M \geq 25$). Hence, the convergence properties associated with a self-consistent pseudopotential perturbation are very similar to those of curve $V_0=2.25$ Ry; $R_0=2.2$ a.u. Further, to examine the qualitative behavior shown in Fig. 2, we have calculated $\epsilon_{t_2}(f)$, where f is a constant scale factor, multiplying the self-consistent vacancy perturbation $\Delta V(\vec{r})$ of Ref. 3. For $f=1$ (no scaling), we recover the result $\epsilon_{t_2}=0.8$ eV, very close to that of Ref. 3. For $f \geq 1.5$ we find, much like in Fig. 2, that the defect level merges into the conduction band rather than saturate in the band gap. We conclude therefore that the results of the square-well potential shown in Figs. 1 and 2 correctly model the M as well as the ΔV dependence of the defect energies.

[15] Notice that the square-well defect model does not map into the TB orbital removal model (cf. Ref. 10) at the limit $V_0 \rightarrow \infty$. At this limit, not used in this work, the defect wave function is discouraged from penetrating the potential well (except through low probability tunneling) and therefore does not correspond to a dangling bond state. Tight-binding vacancy levels are obtained in the *orbital removal* scheme by taking $V_0 \rightarrow \infty$, whereas in the present more appropriate *potential removal* scheme the vacancy (as well as impurity) level correspond to a finite V_0.

[16] Although the coordinates $\{R_0, V_0\}$ are physically interrelated, one can hardly expect that this subtle correlation [provided, for instance, by careful self-consistent pseudopotential atomic calculations (Ref. 17)] will be reflected in the much cruder fitting of the host band structure achieved by varying $V_{\alpha\beta}$ (but not $V_{\alpha\alpha}$ or $V_{\beta\beta}$).

[17] A. Zunger, Phys. Rev. B 22, 5839 (1980).