Theory of Defect Levels and the "Band Gap Problem" in Silicon

Peter A. Schultz*

Multiscale Computational Materials Methods, Sandia National Laboratories, Albuquerque, New Mexico 87185-1110, USA

(Received 10 November 2005; published 19 June 2006)

Quantitative predictions of defect properties in semiconductors using density functional theory have been crippled by two issues: the supercell approximation, which has incorrect boundary conditions for an isolated defect, and approximate functionals, that drastically underestimate the band gap. I describe modifications to the supercell method that incorporate boundary conditions appropriate to point defects, identify a common electron reservoir for net charge for all defects, deal with defect banding, and incorporate bulk polarization. The computed level spectrum for an extended set of silicon defects spans the experimental gap, i.e., exhibits no band gap problem, and agrees remarkably well with experiment.

DOI: 10.1103/PhysRevLett.96.246401 PACS numbers: 71.15.Mb, 61.72.Bb, 61.72.Ji, 71.55.−i

Formation energies of charged defects and electronic transitions between different charge states of a defect ultimately govern the electrical response of semiconductors. Density functional theory (DFT) [1,2] has proven to be an effective tool for the study of semiconductors and is the method of choice for first-principles simulations of defects. However, accurate calculations of defect properties, especially defect levels, in even the most studied semiconductor, silicon, have been problematic. First, DFT defect calculations almost invariably incorporate boundary conditions inappropriate for a dilute defect, due to use of the supercell approximation [3,4]. A second issue is the ''band gap problem'' where DFT calculations, specifically the Kohn-Sham eigenvalues, significantly underestimate the fundamental band gap [5]. The silicon DFT band gap is less than half the experimental gap and, since the band gap defines the interesting energy scale for defect levels, this appears to be a fundamental impediment to quantitative predictions.

I present a method that addresses the deficiencies of the standard supercell approximation and constitutes a more rigorous, internally consistent computational model for an isolated defect. Electrostatic boundary conditions are enforced that explicitly and properly account for net charge and bulk polarization effects. Formation energies of all charged defects are computed with a common electron reservoir, eliminating the need for *ad hoc* procedures to calibrate the electron chemical potential for each charged defect calculation. Defect dispersion is handled with a new occupation scheme for populating states. I apply this new supercell model in DFT simulations to a variety of defects in silicon to compute defect formation energies and defect energy levels. The computed defect level spectrum is not hampered by a band gap problem and shows that DFT is as successful for predicting electronic defect levels in semiconductors as it is for structural energetics, once a robust computational model of the defect and its charge is constructed.

Periodic boundary conditions (PBCs) are the theoretical framework around which solid state DFT codes are built. Although this assumption provides a rigorous computational model for perfect crystals, point defects lack three-dimensional periodicity. To apply periodic DFT codes to defects, the supercell approximation [3,4] is invoked, wherein a point defect is modeled as a periodic array of defects rather than as a single isolated defect. The Coulomb potential for a periodic array of charged defects formally diverges, however, and creates difficulties in creating a quantitative model.

A central difficulty in computing formation energies of charged defects is locating a valid chemical potential for electron charge. A variety of empirical recipes have been proposed to calibrate the electron chemical potential in formation energy calculations of charged defects [6]. The most popular aligns the valence band (VB) or conduction band (CB) edge eigenvalue with the eigenvalue of a similar state in a crystal calculation. Validity of these recipes depends on the supercell band structure not being modified by the defect, a poor assumption, and runs afoul of the band gap problem, in that the VB and CB eigenvalues of the crystal calculation are themselves dubious references. Even taking advantage of this arbitrariness in defining the chemical potential, these recipes yield no better than a few tenths of an eV accuracy [7]. In the following, I describe how to combine four modifications to the supercell approximation that provide a firmer physical foundation for supercell defect calculations.

Boundary conditions.—The local moment countercharge (LMCC) method [8] removes the divergence and installs the correct boundary conditions for net charge. The LMCC splits the supercell density $\rho(r)$ into two parts, a net charge $\rho_{\mathrm{LM}}(r)$ and a remainder defect density $\rho'_{\mathrm{def}}(r)$ with the net charge removed:

$$\rho(r) = \rho'_{\mathrm{def}}(r) + \rho_{\mathrm{LM}}(r). \tag{1}$$

The electrostatic potential $\phi_{\mathrm{LM}}(r)$ due to the net charge ρ_{LM} is solved with local boundary conditions, ($q/r \to 0$, as $r \to \infty$), while the potential $\phi'_{\mathrm{def}}(r)$ from the neutral defect density $\rho'_{\mathrm{def}}(r)$ is solved in PBC. The potential due to the net charge is truncated at the cell boundary and does not corrupt the potential of a neighboring cell. Divergence is avoided, and boundary conditions are appropriate to a local

defect. Applied in supercell calculations of charged atoms and molecules [9], the LMCC gives total energies without supercell dependence. The total potential is the sum of the separate potentials:

$$\phi(r) = \phi'_{\mathrm{def}}(r) + \phi_{\mathrm{LM}}(r). \quad (2)$$

Standard computational approaches add a flat background charge, jellium, to neutralize net charge in a supercell. A jellium-based Poisson solver does not lead to the correct local electrostatic potential [8] and hence yields an incorrect self-consistent density [10]. More critically, finding a valid electron chemical potential is difficult.

Chemical potential.—Kleinman [11] demonstrated that the Poisson Equation for the Coulomb potential in a three-dimensional periodic calculation is solvable only to within an integration constant. For uncharged supercell calculations, this is inconsequential; the total energy for a neutral defect is insensitive to a constant potential C_{def}. For a charged defect, however, the formation energy depends linearly in the charge on C_{def}. The DFT defect supercell energy $E_{\mathrm{def}}(q)$ contains qC_{def}. Conventional expressions for a charged defect formation energy [12,13] are only valid for charged defects if the integration constant C_{def} is the same for all defect calculations—which is not the case. The value of C_{def} will be defect dependent, cell size and shape dependent, and charge-state dependent; i.e., C_{def} can be different for every defect calculation.

To make explicit the indefinite nature of the defect potential, restate ϕ'_{def} in Eq. (2) as

$$\phi'_{\mathrm{def}}(r) = \phi_{\mathrm{def}}(r) + C_{\mathrm{def}}. \quad (3)$$

The problematic term qC_{def} in $E_{\mathrm{def}}(q)$ stems from the Coulomb integral of this potential with the net charge. The following expression eliminates C_{def} from $E_{\mathrm{def}}(q)$:

$$\begin{aligned} E^{\mu_e}_{\mathrm{def}}(q) = &- \int dr \eta_{\mathrm{LM}}(r)(\phi_{\mathrm{def}}(r) + C_{\mathrm{def}}) \\ &+ \int dr \phi_{\mathrm{LM}}(r)\rho'_{\mathrm{def}}(r) \\ &+ \int dr \eta_{\mathrm{LM}}(r)(\phi_p(r) + C_p) \\ &- \int dr \phi_{\mathrm{LM}}(r)\rho_p(r). \end{aligned} \quad (4)$$

The first term on the right subtracts the energy of the net charge interacting with the potential generated by the *periodic* defect density, eliminating the problematic indeterminate constant C_{def} from the energy. This Coulomb energy within the volume of the supercell is added back in the second term, computing the energy with the *local* potential due to the net charge and the defect density within the central cell. Because $\rho_{\mathrm{LM}}(r)$ has a well-defined potential, this interchange is possible. The third and fourth terms in Eq. (4) retrace the steps of the first two terms, but using the crystal density $\rho_p(r)$ and potential $\phi_p(r) + C_p$. With this correction term, the defect supercell electrostatic energy is computed in a Coulomb field defined by embedding the central cell charge in the bulk crystal. The defect-dependent constant C_{def} is replaced by a constant defined by the crystal calculation. Although unknown, C_p will be the same for every defect, in every supercell, in every charge state, so that all charged defect formation energies are grounded to the same electron reservoir.

Bulk polarization.—A net charge at a defect induces screening in the bulk crystal that is not contained within the volume of a supercell. A simple dielectric continuum approximation due to Jost [14] suffices to estimate the missing bulk polarization energy. Taking the unscreened volume as a structureless dielectric about a spherical cavity with a point charge at its center, the polarization energy (Hartree atomic units) is

$$E_{\mathrm{pol}}(q) = (1 - 1/\epsilon_0)(q^2/2R_{\mathrm{Jost}}), \quad (5)$$

where ϵ_0 is the static dielectric constant (11.8 for Si), q is the net charge, and R_{Jost} is the radius of the unscreened volume. Lento, *et al.* [10] observed that, using the LMCC, the screened volume in a supercell calculation is less than the full volume of the supercell. Noting this, I use an R_{Jost} that is smaller (by 0.8 bohr) than the radius defined by the supercell volume. The resulting shift from this offset is not large, 0.1 eV difference for doubly charged species in a 250-site supercell.

Defect level dispersion.—In the modest supercell sizes dictated by computational considerations, a defect interacts with its periodic images to form a defect band rather than a discrete state in the gap, as illustrated in the schematic band structure in Fig. 1. For a converged treatment of the bulk semiconductor, the Brillouin zone must be sampled at multiple points [15]. However, defect states at those sampling points can disperse above the CB minimum or below the VB maximum [16]. Using conventional schemes, if a defect state eigenvalue disperses above a bulk band edge, the defect state is depopulated in favor of a bulk band state and a metallic state results [Fig. 1(c)], putting charge in the delocalized band state and reducing the net localized charge on the defect. For a defect state whose energy does not cross a bulk band edge, a defect band with one electron will be populated with two electrons in the defect state at one sampling point and with zero at another [Fig. 1(b)].

To obtain a localized defect in the correct charge state, I propose a discrete defect occupation scheme, illustrated in Fig. 1(d)–1(f). States are populated at each sampling point as if that k point were the only sampling point. Metallic occupation, defined by moving electrons between sampling points, is excluded. For silicon defects, this scheme is essential to obtain meaningful results, in particular, for negatively charged defects with transitions near the CB.

Taken together, these steps define a finite defect supercell model (FDSM) that provides an internally consistent

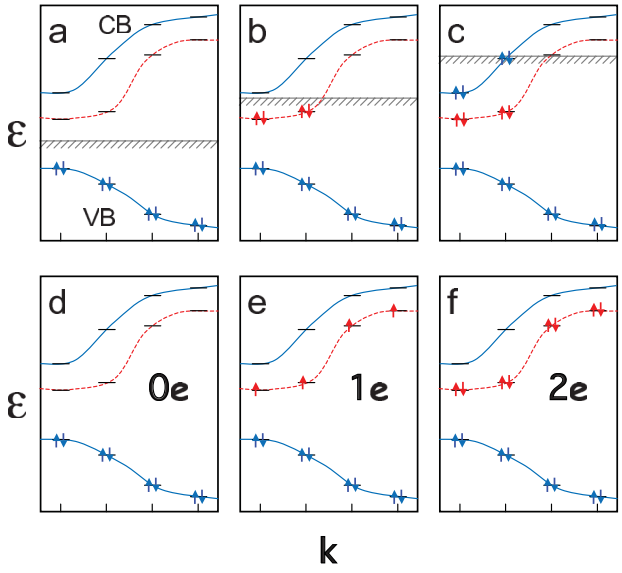

FIG. 1 (color). Schematic representation of the standard method and the discrete defect occupation method for populating states of a supercell calculation. Vertical axis corresponds to eigenvalues of the Hamiltonian as a function of the k vector along the horizontal axis. The defect band is depicted as a red dashed line. (a), (b), (c) In conventional methods, a Fermi level (gray horizontal line) determines the occupations of the states, states with eigenvalues below E_F being occupied, those above being empty. (d), (e), (f) With discrete defect occupation, states are populated within each sampling point as if the defect were isolated. Each charge state describes a valid model of an localized defect state with zero, one, or two electrons.

description of an isolated defect. The FDSM formation energy of a charge defect is written as

$$E_{\text{form}}(q) = E_{\text{def}}(q) + E_{\text{def}}^{\mu_e}(q) + E_{\text{pol}}(q) - E_p(0) - \sum_I \Delta N_I \mu_I \quad (6)$$

where $E_{\text{def}}^{\mu_e}$ [Eq. (4)] connects the formation energy to a fixed electron reservoir, E_{pol} [Eq. (5)] is the bulk polarization energy, E_p is the perfect crystal supercell energy, and the last term is the energy cost (chemical potential μ_I) associated with the difference in atom number ΔN_I between the defect and crystal cell. Charge transition energies (ionization potentials) are computed, as usual, as differences between formation energies. This determines all the charge transition energies with respect to one another, grounded to a common electron reservoir defined by an unknown constant in the crystal calculation. A single marker, such as a known defect level, then serves to align the entire computed defect level spectrum within the experimental band gap.

Calculations were performed with SEQ QUEST [17] in the local density approximation (LDA) [18]. Core electrons are replaced by pseudopotentials [19,20] and valence electron wave functions are expanded in well-converged, double-zeta plus polarization, contracted Gaussian basis sets. Results presented used 250-site supercells with a 2^3 k sample. In accord with guidelines set forth in Ref. [21], simulations were tested for sufficiency with respect to various calculational parameters, the details of which are expanded upon in the supplemental information in Ref. [22], and in subsequent publications.

The FDSM was applied to a wide variety of defects in silicon: intrinsic defects (self-interstitial i, vacancy v, divacancy vv), first row defects (boron interstitial B_i, carbon interstitial C_i, nitrogen substitutional N_s, A center O_s), and a second row defect (sulfur substitutional S_s). Figure 2 presents the computed charge transitions for this suite of defects. The transition energies range from 4.38 eV for $i(2-/-)$ to 5.45 eV for $v(+/2+)$. This 1.07 eV span in defect transition energies clearly exceeds the 0.5 eV band gap in the DFT crystal eigenvalue spectrum. The range of energies accessible to local charge transitions, the LDA effective defect gap, is nearly the width of the experimental band gap.

In Fig. 3, the computed LDA defect spectrum is compared to known experimental levels [23–29]. The defect charge transitions of Fig. 2 are aligned to set the CB edge at 4.35 eV and the VB edge at 5.52 eV. With this alignment, the largest discrepancies between experiment and theory are ~0.2 eV, and the mean average error in the defect levels is 0.10 eV. The LDA has been very successful for structural energetics in solid state systems and this broad and excellent agreement with experiment demonstrates that this success extends to charge state transitions computed as differences of ground state total energy calculations. Calculations with the generalized gradient approximation (GGA) yield a similar defect spectrum as LDA. Those results will be presented elsewhere.

The largest discrepancy with experiment is in the boron interstitial, with the consequence that LDA fails to predict the observed negative-U behavior [27]. I note here that GGA corrects this, as GGA handles more accurately than LDA the significant bonding rearrangements that occur as

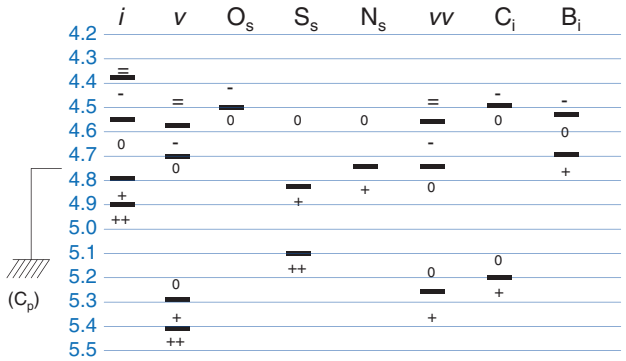

FIG. 2 (color). The computed DFT ionization potentials for eight different silicon defects, with respect to the (unknown) crystal electronic reservoir. The span of charge state transitions is more than 1 eV, demonstrating that defect charge transitions are not constrained by a band gap problem.

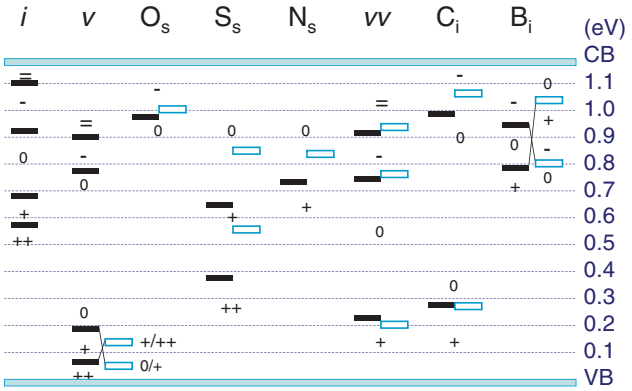

FIG. 3 (color). Computed LDA defect levels (solid blocks) are compared against experimental defect levels (open), aligned within the band gap. Agreement between LDA and experiment is good for all defects, in all parts of the band gap.

B_i changes charge states. The LDA also fails to reproduce the observed negative-U behavior in the positive charge states of the vacancy [23], but the error in the individual levels is not large, only ~0.1 eV. The negative vacancy charge levels are predicted to lie at CB-0.27 eV and CB-0.39 eV. The agreement between the LDA and experiment is remarkably good for all known levels of defects as varied as the A center, the carbon interstitial, and the divacancy. The single self-interstitial has never been observed experimentally. The LDA predicts five charge states are stable, and all the levels are predicted to lie in the upper half of the band gap.

Using a practical and robust computational model it is shown that one can compute an accurate defect level spectrum in silicon. With its small band gap and extensive experimental data, silicon serves as a particularly stringent test of the methodology developed here. The band gap problem, a problem of DFT with the eigenvalues of delocalized bulk band states, is seen not to be a problem for computing defect levels. Defect levels are computed as differences of ground state total energies of local defect states, and are fully screened with a static calculation, without reference to Kohn-Sham eigenvalues. The key aspect in computing total energies of localized charged defects in the supercell approximation is to find a chemical potential that ties charged defect formation energies to a common electron reservoir, and to provide a compatible treatment of bulk screening. The intrinsic band gap problem of computing the quasiparticle energies associated with delocalized bulk band excitations remains open.

I thank A. E. Mattsson, A. H. Edwards, R. M. Van Ginhoven, and H. P. Hjalmarson for useful discussions, and J. B. Aidun and T. R. Mattsson for critical comments on the manuscript. Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under Contract No. DE-AC04-94AL85000.

*Electronic address: paschul@sandia.gov

[1] P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964).
[2] W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).
[3] R. P. Messmer, B. McCarroll, and C. M. Singal, J. Vac. Sci. Technol. **9**, 891 (1972).
[4] M. L. Cohen, M. Schlüter, J. R. Chelikowsky, and S. G. Louie, Phys. Rev. B **12**, 575 (1975).
[5] L. J. Sham and M. Schlüter, Phys. Rev. Lett. **51**, 1888 (1983); Phys. Rev. B **32**, 3883 (1985).
[6] J. Shim, E.-K. Lee, Y. J. Lee, and R. M. Nieminen, Phys. Rev. B **71**, 035206 (2005).
[7] A. Garcia and J. E. Northrup, Phys. Rev. Lett. **74**, 1131 (1995).
[8] P. A. Schultz, Phys. Rev. Lett. **84**, 1942 (2000).
[9] P. A. Schultz, Phys. Rev. B **60**, 1551 (1999).
[10] J. Lento, J.-L. Mozos, and R. M. Nieminen, J. Phys. Condens. Matter **14**, 2637 (2002).
[11] L. Kleinman, Phys. Rev. B **24**, 7412 (1981).
[12] S. B. Zhang and J. E. Northrup, Phys. Rev. Lett. **67**, 2339 (1991); J. E. Northrup and S. B. Zhang, Phys. Rev. B **47**, R6791 (1993).
[13] G. Makov and M. C. Payne, Phys. Rev. B **51**, 4014 (1995).
[14] W. Jost, J. Chem. Phys. **1**, 466 (1933).
[15] G. Makov, R. Shah, and M. C. Payne, Phys. Rev. B **53**, 15 513 (1996).
[16] C. W. M. Castleton and S. Mirbt, Phys. Rev. B **70**, 195202 (2004).
[17] P. A. Schultz (unpublished); See http:/dft.sandia.gov/Quest.
[18] J. P. Perdew and A. Zunger, Phys. Rev. B **23**, 5048 (1981).
[19] N. Troullier and J. L. Martins, Phys. Rev. B **43**, 8861 (1991).
[20] D. R. Hamann, Phys. Rev. B **40**, 2980 (1989).
[21] A. E. Mattsson, P. A. Schultz, T. R. Mattsson, K. Leung, and M. P. Desjarlais, Model. Simul. Mater. Sci. Eng. **13**, R1 (2005).
[22] See EPAPS Document No. E-PRLTAO-96-057626 for supplemental information describing additional computational details, convergence tests, and numerical tables of the results. For more information on EPAPS, see http://www.aip.org/pubservs/epaps.html.
[23] G. D. Watkins and J. R. Troxell, Phys. Rev. Lett. **44**, 593 (1980).
[24] V. P. Markevich, A. R. Peaker, S. B. Lastovskii, L. I. Murin, and J. L. Lindström, J. Phys. Condens. Matter **15**, S2779 (2003).
[25] L. C. Kimerling, P. Blood, and W. M. Gibson, in *Defects and Radiation Effects in Semiconductors-1978*, edited by J. H. Albany (Inst. of Physics and Physical Society, London, 1979), p. 273.
[26] K. Murakami, H. Kuribayashi, and K. Masuda, Jpn. J. Appl. Phys. **27**, L1414 (1988).
[27] R. D. Harris, J. L. Newton, and G. D. Watkins, Phys. Rev. Lett. **48**, 1271 (1982); Phys. Rev. B **36**, 1094 (1987).
[28] G. D. Watkins, J. W. Corbett, and R. M. Walker, J. Appl. Phys. **30**, 1198 (1959).
[29] H. G. Grimmeiss, E. Janzén, and K. Larsson, Phys. Rev. B **25**, 2627 (1982).