RAPID COMMUNICATIONS

PHYSICAL REVIEW B VOLUME 28, NUMBER 4 15 AUGUST 1983

Total energy of isolated point defects in solids

G. A. Baraff and M. Schluter
Bell Laboratories, Murray Hill, New Jersey 07974

(Received 19 April 1983)

Total-configurational-energy calculations have already been very successfully carried out for periodic structures and for isolated molecules, allowing one to study, e.g., crystal structures, reconstructions, and vibrational frequencies. We describe here for the first time a method for calculating the total energy of isolated, charged, point defects in an infinite crystal. It uses the self-consistent Green's-function technique. We demonstrate its quantitative utility be applying it to two defects in GaP.

Many of the interesting phenomena associated with deep-level point defects in semiconductors occur because the equilibrium positions of the constituent atoms and the forces holding them in place depend on the charge state of the defect.[1] Examples include negative-U systems,[2] radiation enhanced diffusion,[3] persistent photoconductivity,[4] and various self-trapping systems.[5] Although understanding these situations requires knowing how the total energy of the system depends on the charge and position of the nearby atoms, no quantitative scheme to solve this problem for a truly isolated point defect has yet been proposed or implemented.

Approximations to the total energy of defects have been obtained using self-consistent electronic calculations in a supercell geometry.[6,7] Aside from the well-known problems with this method, e.g., impurity banding, the supercell geometry demands either that the defect be neutral, precluding a study of the charge dependence,[6] or that there be a pair of oppositely charged defects in each neutral supercell,[7] raising questions about the long-range interactions between charged defects. Calculations also have been carried out using first-principles[8] and semiempirical methods in a finite cluster.[9] Here, in addition to the well-known problems with cluster methods, a problem arises of spurious interaction between the charged defect at the center of the cluster and the cluster terminator. Interestingly, gradients of the energy at a truly isolated defect have recently been calculated using the Hellmann-Feynman theorem.[10,11] Although this is useful for determining equilibrium configurations, it is not so for comparing situations which differ by more than a simple distortion. Clearly, a method of calculating the total energy of isolated defects will have widespread applicability. This Rapid Communication presents what is, to our knowledge, the first method for doing so and demonstrates its utility by applying it to two illustrative situations.

We work within the local-density-functional formalism,[12] using first-principles pseudopotentials[13] that are equivalent to the frozen-core approximation.[14] The total energy is given by the usual expression,[15]

$$E_T[\rho_e, V_{\mathrm{ext}}, R_A] = T[\rho_e] + V_{ee}[\rho_e] + V_{ei}[\rho_e, R_A] + V_{ii}[R_A] + E_{xc}[\rho_e] \ , \tag{1a}$$

where

$$T \equiv \sum n_i \int \Psi_i(r)^* \left(-\tfrac{1}{2}\nabla^2\right)\Psi_i(r)\,dr \ , \tag{1b}$$

$$\rho_e(r) = \sum n_i \Psi_i(r)^* \Psi_i(r) \ , \tag{1c}$$

$$V_{ee} \equiv \frac{1}{2}\int\int \frac{dr\,dr'}{|r-r'|}\rho_e(r)\rho_e(r') \ , \tag{1d}$$

$$V_{ei} \equiv \int \sum V_{\mathrm{PS}}(|r-R_A|)\rho_e(r)\,dr \ , \tag{1e}$$

$$V_{ii} \equiv \frac{1}{2}\sum_A \sum_{B\neq A} \frac{Q_A Q_B}{|R_A - R_B|} \ , \tag{1f}$$

$$E_{\mathrm{xc}} \equiv \int \rho_e(r)\epsilon_{\mathrm{xc}}(\rho_e(r))\,dr \ . \tag{1g}$$

This functional should be evaluated using the density $\rho_e(r)$ that minimizes it. Such a density can be constructed by using wave functions that satisfy a Schrödinger-type equation[12]:

$$\left[-\frac{1}{2}\nabla^2 + \phi(r) + \sum_A V_{\mathrm{PS}}(|r-R_A|) + V_{\mathrm{xc}}(r)\right]\Psi_i(r) = \epsilon_i \Psi_i(r) \ , \tag{2a}$$

where

$$\phi(r) \equiv \int \frac{dr'}{|r-r'|}\rho_e(r') \ , \tag{2b}$$

$$V_{\mathrm{xc}}(r) = \frac{d}{d\rho}[\rho\epsilon_{\mathrm{xc}}(\rho)]_{\rho=\rho_e(r)} \ . \tag{2c}$$

When the conditions (1) and (2) are satisfied, a computationally simpler form of Eq. (1a) is[15]

$$E_T = \sum n_i\epsilon_i + V_{ii} - V_{ee} + \int \rho_e(r)(\epsilon_{\mathrm{xc}} - V_{\mathrm{xc}})\,dr \ . \tag{3}$$

The defect energy we wish to calculate, ΔE, is the difference of two energy expressions similar to (3). The initial situation is the infinite, neutral, perfect crystal and the final situation is the infinite, neutral crystal containing the defect. That difference contains a large constant term, the difference between the total energy of those isolated atoms removed from the crystal and those added to it to create the defect. The large constant could be eliminated by doing total energy calculations for the isolated atoms, but we shall not do so here.

To permit charged defects in the neutral crystal, we must allow conceptually for reservoir centers at large distance from the defect. Only one property of the reservoir must be specified: its energy change μ on accepting an electron. That change depends on how one chooses the zero of potential energy within the solid, but so, too, does each eigenvalue ϵ_i. By referring μ to the same energy scale on which the eigenvalues are measured, the choice of zero for potential energy within the solid is made irrelevant.

In evaluating the change in (3), it is useful to write the electron density and ionic density as

$$\rho_e(r) = \rho_e^0(r) + \Delta\rho_e(r) \ , \tag{4a}$$

$$\rho_i(r) = \rho_i^0(r) + \Delta\rho_i(r) \ , \tag{4b}$$

where the superscript zero refers to the perfect-crystal situation. The ionic charge in (4b) is, in the usual way, related to the Laplacian of the local part of the ionic pseudopotential V_{PS}. The change in ionic density, $\Delta\rho_i(r)$, is one of the inputs to the problem. This quantity will be taken nonzero only within the very limited defect region. The change in the electronic density $\Delta\rho_e(r)$ has to be calculated and, for isolated point defects, the self-consistent Green's-function technique provides a quantitative way of doing so.[16,17] Using that technique, one also obtains the scattering phase shifts $\phi(E)$, from which the change in the eigenvalue term in (3) can be evaluated as[18]

$$\Delta\sum_i n_i\epsilon_i = -\frac{1}{\pi}\int_{-\infty}^{E_F}\epsilon\, d\epsilon\,\frac{d\phi}{d\epsilon} \quad . \tag{5}$$

For the charged defect, however, two problems arise. The first, seemingly formal, is that scattering phase shifts do not converge for a $1/r$ scattering potential. In the past, this has not been regarded as a serious problem because none of the formulations of the Green's-function method have been able to follow the Coulomb tail out to infinity, and so no numerical difficulties with (5) have been reported. The second problem is more immediate: When one evaluates the change in $V_{ii}-V_{ee}$ using (4), the first-order term is

$$\Delta^{(1)}(V_{ii}-V_{ee}) = \int\int\frac{dr\,dr'}{|r-r'|}[\rho_i^0(r)\Delta\rho_i(r') - \rho_e^0(r)\Delta\rho_e(r')] \quad . \tag{6}$$

This integral diverges at large values of r unless the defect is neutral. Actually, this problem is related to the seemingly formal one of evaluating (5). Had we studied how the phase shifts diverge, we would have found exactly the same problem with (5) as we did with (6). The two divergences would have canceled were it not for the accident that the inability to follow the Coulomb tail to infinity suppressed the infinite value concealed in (5).

To eliminate both divergences consistently, it is useful to make the cutoff of the Coulomb tail an explicit, rather than an accidental, feature of the procedure. Define the defect potential $U(r)$ in the usual way as the difference between the Schrödinger operators (2a) for the perturbed and perfect crystals. Instead of using $U(r)$ in the Green's-function equation $G=G^0+G^0UG$, we set the potential to zero identically for r larger than some cutoff radius R. That is, we replace the scattering potential $U(r)$ in the Green's-function equation by $U(r)\Theta(R-r)$, where Θ is the unit step function. This means that the perturbed wave functions no longer satisfy (2a), but instead satisfy

$$\left[-\frac{1}{2}\nabla^2+\phi(r)+\sum_A V_{PS}(r-R_A)+V_{xc}(r)\right]\psi_i(r) = [\epsilon_i+U(r)\Theta(r-R)]\psi_i(r) \quad . \tag{7}$$

We do not solve (7) directly, but instead, we solve the Green's-function equation containing the cutoff potential $U(r)\Theta(R-r)$.

The density $\rho_e(r)$, obtained from this G, does not minimize the functional (1a) because its wave functions do not satisfy (2a). However, if we evaluate (1a) using this density, the error is second-order small because the true density does minimize the functional. The steps which led from (1a) to (3) can still be carried out but now, when this is done, the second term on the right-hand side (rhs) of (7) causes an extra term, $-\int_{r>R}U(r)\rho_e(r)dr$, to appear on the rhs of (3). This term combines with Eq. (6) to give a convergent result even for charged defects. The phase shifts $\phi(\epsilon)$, evaluated for a cutoff potential, are now also well defined, and there is no ambiguity about any of the terms going into ΔE. Because of Eq. (7), Janak's theorem[19] $\partial E_{\text{tot}}/\partial n_i=\epsilon_i$ no longer holds. Instead, we now find that

$$\frac{\partial(\Delta E)}{\partial n_i} = \epsilon_i + \int_{r>R}U(r)\frac{d}{dn_i}\Delta\rho_e(r)dr \quad , \tag{8}$$

where $\Delta\rho_e(r)$ is the change in density we obtain from the Green's-function calculation using the truncated potential.

Let us now turn to results obtained by applying this formalism to calculating the total energy of two illustrative defects in GaP. The purpose of this is to demonstrate that the proposed formalism, which is so far completely untried, is a viable one. Our Green's function for GaP carries 20 orbitals per atom, and uses a 17-atom cluster for both inner- and outer-set orbitals.[16] We use nonlocal pseudopotentials[13] for Ga, for P, and O, and the Ceperly-Alder expression for correlation and exchange. The cutoff radius R is $0.70a$, where a is the lattice constant. Thus, R passes close to the centers of the second-neighbor phosphorus atoms, each of which is $a/\sqrt{2}$ distant from the central atom.

The first-model defect is created by having phosphorus as the central atom and allowing the four nearest-neighbor Ga to move, each by the same amount, radially outward ($u>0$) or inward ($u<0$). The displacement $u=0$ corresponds to a perfect crystal and should, by translational symmetry, be the minimum energy configuration. The Green's-function formalism, with the 17-atom cluster, the cutoff R, etc., does not automatically retain that symmetry. Therefore, obtaining the energy minimum at $u=0$ is an important test of the convergence and internal numerical consistency of the scheme. Strictly speaking, there is no need to use the cutoff R for this defect, which is neutral. However, we have deliberately chosen R somewhat smaller than would be needed to fully enclose the volume of nonzero defect potential in order to demonstrate that the cutoff of itself does not introduce any spurious shift in the equilibrium position. ΔE_T was calculated at four nonzero displacements (see Table I). Self-consistent iterations were carried out until the difference between input and output potential at any point was no larger than 0.004 eV for the large displacements or 0.0003 eV for the small ones. This high degree of self-consistency is needed for the convergence of the energy. It is 10 to 100 times better than what is normally needed for convergence of the electronic structure. The calculated energies satisfy

$$\Delta E = Bu+\tfrac{1}{2}Cu^2+\tfrac{1}{6}Du^3 \quad ,$$

with $B=0$, $C=70.85$ eV/Å^2, and $D=-96.77$ eV/Å^3, to better than 0.01 eV. The minimum is at $u=0$, the anhar-

TABLE I. Energy of GaP for various radial displacements of Ga centered around P.

u (Å)	+0.236	0	−0.0472	−0.0755	−0.236
ΔE (eV)	1.761	0	0.087	0.207	2.185

monicity is in the expected direction (compressing the bonds to the central phosphorus is harder than stretching them), and the spring constant C is reasonable in size. (The compressibility of the crystal, interpreted as nearest-neighbor bond stretching, would have given $C \approx 50$ eV/Å^2. Here, there are bond-bending forces of unknown amount acting to raise C.)

The second example is a model defect in the sense that we do not investigate lattice displacements. It is GaP:O$_\mathrm{P}$, where an oxygen atom replaces the central phosphorus but the nearest gallium neighbors are at $u=0$. Here, the intent is to test the operation of the cutoff scheme when there is a Coulomb tail in $U(r)$. The defect can exist in three charge states, O$^+$, O^0, and O$^-$, obtained by populating the gap state with 0, 1, or 2 electrons. We calculate the total energy $\Delta E(n)$ and the eigenvalue $\epsilon(n)$ for $n=0$, 1, and 2, where n is the occupancy of the gap state. By fitting ΔE and ϵ to quadratics in n, we can evaluate $\partial\Delta E/\partial n$ at $n=0$, 1, and 2, and ϵ at $n=\frac{1}{2}$ and $n=\frac{3}{2}$. Table II shows that $\partial\Delta E/\partial n_i - \epsilon_i$, the integral in (8), is small at the R we have used. For this reason, Janak's theorem[19] is well satisfied and the Slater transition-state eigenvalue method is justified for calculating total energy differences.

TABLE II. Total energy ΔE, eigenvalue ϵ, and $d(\Delta E)/dn$ for the three charge states of unrelaxed GaP:O$_P$ (in eV).

State	n	ΔE	$\epsilon(n)$	$\partial\Delta E/\partial n$	$\partial\Delta E/\partial n - \epsilon(n)$
0^+	0	−219.654	1.1784	1.214	+0.036
0^0	1	−218.330	1.4603	1.435	−0.025
0^-	2	−216.785	1.7188	1.656	−0.063

For a cutoff R large enough to fully enclose the defect potential of a neutral system, the integral should have been zero for $n=1$, instead of for $n \approx 0.5$ as is the case here. The discrepancy is presumably to be attributed to our deliberate choice of a rather small cutoff.

The values in these tables were calculated without taking account of the second-order difference in total energy which arose because the wave functions do not accurately minimize the functional. They were obtained with μ placed at the valence-band maximum, which we take as the zero for the eigenvalue scale. Thus, the differences

$$E(0/+) \equiv \Delta E(1) - \Delta E(0) = 1.32 \text{ eV}$$

and

$$E(-/0) \equiv \Delta E(2) - \Delta E(1) = 1.55 \text{ eV}$$

are the energies required to take an electron from the valence band to the first- and second-donor levels of GaP:O$_\mathrm{P}$ assuming no lattice relaxation. Let these differences be approximated by the transition-state eigenvalues,

$$E(0/+) \approx \epsilon(\tfrac{1}{2}) = 1.32 \text{ eV}, \quad E(-/0) \approx \epsilon(\tfrac{3}{2}) = 1.59 \text{ eV} .$$

Then the major effect of the second-order total energy change may be obtained by using perturbation theory to estimate the *first*-order change in the eigenvalues $\epsilon(\frac{1}{2})$ and $\epsilon(\frac{3}{2})$ caused by the long-range potential at these half-integer occupancies, namely, by $\mp\frac{1}{2}\Theta(r-R)/(Kr)$. ($K$ is the dielectric constant of GaP.) This lowers $\epsilon(\frac{1}{2})$ and raises $\epsilon(\frac{3}{2})$ by about 0.03 eV. Their difference, $U \equiv \epsilon(\frac{3}{2}) - \epsilon(\frac{1}{2})$ is the electron-electron repulsion energy for this defect and is calculated to be 0.33 eV. This value of U is in reasonable accord with what we expect experimentally from this system, namely, ≈ 0.4 eV. The values of $E(0/+)$ and $E(-/0)$ are 0.1 and 0.2 eV lower than the experimental values, namely, 1.45 and ~ 1.8 eV, respectively. Whether or not this discrepancy will decrease when the nearest-neighbor Ga atoms relax to their position of minimum energy is presently under investigation and will be reported elsewhere. It would also be interesting to see whether use of the local-spin-density formalism would change the results found for O^0 which, having an unpaired electron in the gap state, does have a nonzero spin. We are not, however, able to investigate this point.

In summary, we have presented a method for performing total energy calculations on isolated point defects and have validated its ability to deal with distortions and with occupancy changes. The existence of such a method allows the several groups now using self-consistent Green's functions to study problems previously inaccessible to them.

---

[1] See, for example, M. Jaros, *Deep Levels in Semiconductors* (Hilger, Bristol, 1982), or M. Lannoo and J. Bourgoin, *Point Defects in Semiconductors* (Springer, Berlin, 1981).

[2] G. D. Watkins and J. R. Troxell, Phys. Rev. Lett. 44, 593 (1980).

[3] L. C. Kimerling, in *Defects and Radiation Effects in Semiconductors,* edited by J. H. Albany, IOP Conf. Proc. No. 46 (Institute of Physics, London, 1979), p. 56.

[4] D. V. Lang and R. A. Logan, Phys. Rev. Lett. 39, 635 (1977).

[5] Y. Toyozawa, Solid State Electron. 21, 1313 (1978).

[6] G. A. Baraff, E. O. Kane, and M. Schluter, Phys. Rev. B 2, 548 (1982).

[7] D. Vanderbilt and J. D. Joannopoulos, Phys. Rev. B 27, 6311 (1983); Phys. Rev. Lett. 42, 1012 (1979).

[8] N. Sahoo *et al.*, Phys. Rev. Lett. 50, 913 (1983).

[9] P. M. Masri, A. H. Harper, and A. M. Stoneham (unpublished).

[10] M. Scheffler, J. P. Vigneron, and G. B. Bachelet, Phys. Rev. Lett. 49, 1765 (1982).

[11] U. Lindefelt (unpublished).

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

[13] D. R. Hamann, M. Schluter, and C. Chiang, Phys. Rev. Lett. 43, 1494 (1979).

[14] U. Von Barth and C. D. Gelatt, Phys. Rev. B 21, 2122 (1980).

[15] J. Ihm, A. Zunger, and M. Cohen, J. Phys. C 12, 4409 (1979).

[16] G. A. Baraff and M. Schluter, Phys. Rev. B 19, 4656 (1979).

[17] J. Bernholc, N. O. Lipari, and S. T. Pantelides, Phys. Rev. B 21, 3245 (1980).

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

[19] J. F. Janak, Phys. Rev. B 18, 7165 (1978).