PHYSICAL REVIEW B VOLUME 53, NUMBER 15 15 APRIL 1996-I

First-principles free-energy calculations on condensed-matter systems: Lattice vacancy in silicon

Enrico Smargiassi
Centre Européen de Calcul Atomique et Moléculaire, Ecole Normale Supérieure de Lyon, 69364 Lyon CEDEX 07, France

Roberto Car
Departement de Physique, University of Geneva, Switzerland

(Received 18 October 1995; revised manuscript received 19 December 1995)

We illustrate a method for performing first-principles free-energy calculations within the Kohn-Sham scheme. We show that the method can be used in cases in which electrons have to be decoupled from the system and illustrate it with results for the formation free energy of the Si vacancy. The results agree well with the available data from experiments and *ab initio* calculations.

I. INTRODUCTION

In recent years, there has been growing interest in the calculation of thermal properties of condensed matter via computer simulation. The most general and direct way is to calculate an appropriate free energy, from which all the equilibrium thermodynamic properties can be obtained. For instance, the knowledge of the free energy of formation F_f of a defect allows one to compute its equilibrium concentration $c=e^{F_f/k_BT}$ which, in addition to being important in itself, also enters the defect-induced diffusion coefficient and other important quantities.

Unfortunately, absolute free energies are not directly measurable, neither experimentally or in a computer simulation, as they contain an entropic part which cannot be computed in any conceivable amount of time. Therefore, indirect methods must be used.[1] A popular choice is "thermodynamic integration." In this method, one takes the potential energy U of the system studied to be a smooth function of a parameter λ. It is then easy to show that $\partial F/\partial\lambda=\langle\Phi(\lambda)\rangle$ [where $\Phi(\lambda)=\partial U/\partial\lambda$ and the angled brackets indicate thermal averaging]; the free-energy difference between the states corresponding to the values $\lambda=1$ and $\lambda=0$ can then be expressed as $\int_0^1\langle\Phi\rangle d\lambda$. As F is a function of state, the states at intermediate values of λ do not need to have any physical meaning: only the smoothness of $U(\lambda)$ and its values at the end points are relevant.

This method has been used widely in the past, on simple model systems like Lennard-Jones,[2] showing that a proper account of anharmonic effects is essential for a full understanding of diffusion processes. On the other hand, similar calculations on important materials like semiconductors have been hindered to date by the lack of sufficiently reliable interatomic potentials. In fact, even well-established potentials may easily lead to unphysical results for defects, if not constructed and handled with great care.[3] And even in these cases, the results are not sufficiently accurate to allow one to draw reliable quantitative conclusions, as it can be seen by comparing the results of Ref. 4 with those obtained *ab initio*.[5] The alternative route of using *ab initio* methods has long been confined to $T=0$ or, at best, harmoniclike approximations,[6] due to the huge requirements in terms of computer resources. To overcome these problems, first-principles molecular dynamics (FPMD) schemes[7] have been developed, which make calculations of thermal properties of real materials from first principles affordable. Therefore the extension to FPMD of the above methods for calculating free energies is now an important issue.

The characteristics of the thermodynamic integration method depend in an important way on the transformation which is chosen. The parameter λ may, in principle, be anything. It can be chosen as a physical parameter of the system, in which case, the system is driven through a physically realizable transformation. In this case, the extension of the method to first-principles calculations is straightforward: all one has to do is to compute energies and forces using one of the existing *ab initio* schemes, instead of using classical interatomic potentials. The calculation of the thermal average of Φ in a first-principles scheme is in this case no more complicated than the calculation of the forces acting on the ions (actually Φ can be considered a "generalized force"): First, one computes, analytically, the derivative of the *functional* $U(\{\mathbf{r}_i\})$, with respect to λ. Second, one performs an MD simulation at the chosen λ points, taking care of keeping the system as close as possible to the instantaneous Born-Oppenheimer (BO) ground state. By invoking the Hellman-Feynman theorem and assuming ergodicity in the subspace of the ionic coordinates, one can then show that the calculated $\langle\partial U/\partial\lambda\rangle$ coincides with the desired ensemble average. In a recent work,[8] for instance, λ was taken to be the reaction coordinate of an atom jumping across a potential barrier. In this case, Φ was the component of the force acting on the jumping atom along the reaction path, thus obtaining the free energy of migration across the barrier.

Many interesting applications, however, require that the system be led through an unphysical transformation. In several important cases, one or more atoms have to be decoupled from the others, thus λ is the coupling of one or more atoms to the rest of the crystal. This is the case of the calculation of formation free energies of defects, or of free energies of extended systems. When using *ab initio* schemes, where electrons are treated explicitly, some or all of them

© 1996 The American Physical Society

must then be decoupled, together with the corresponding ions. This procedure has no counterpart in the approaches that makes use of effective interatomic potentials only (a discussion of this point in general terms is given in Ref. 9), and makes the whole exercise quite more complicated. As it will be shown in this paper in fact, the applicability of the thermodynamic integration method in the framework of the Kohn-Sham formalism to these unphysical transformations is not obvious. We had already shown that it is nevertheless feasible[10] and used the method to calculate the free energy of formation of the vacancy in Si.[11] The results have been published previously,[12] therefore, in this paper we mainly focus the discussion on the technical aspects of our work. We treat in detail the case of the vacancy, but modifications of the method for use in other cases are straightforward. To simplify matters further, we shall discuss only the case of simulations at a constant lattice parameter.

II. FORMALISM

The formation free energy F_f^N of a vacancy, as calculated by using an N atoms supercell, is

$$F_f^N(\Omega_C,T)=F_V^{N-1}(\Omega_V,T)-\frac{N-1}{N}F_C^N(\Omega_C,T), \tag{1}$$

where $F_C^N(\Omega_C,T)$ is the free energy of a crystal with N atoms, volume Ω_C, and temperature T, while $F_V^{N-1}(\Omega_V,T)$ is the free energy of a crystal with the same number of lattice sites, but one of them is vacant. After choosing a suitable single-body reference state, defined by a potential energy u, we define now the following.

(i) $U^N(\lambda=1)=U_C^N=$ potential energy of the perfect crystal.

(ii) $U^N(\lambda=0)=U_V^{N-1}+u=$ potential energy of the vacancy crystal plus the energy of the decoupled atom.

In other words, one ion is adiabatically removed from the crystal and put in the state described by u, while at the same time its valence electrons are made to disappear. The free energies of the bulk system and of the vacancy system are thus, respectively,

$$F_C^N=F^N(\lambda=1),$$

$$F_V^{N-1}=F^N(\lambda=0)-\varphi, \tag{2}$$

where φ is the free energy of an atom in a potential u, at the appropriate temperature. This is a well-defined quantity, as long as the potential u is a single-body potential. Clearly, the form of u is arbitrary: its free energy φ is implicitly added in $F^N(\lambda=0)$ and then explicitly subtracted. For calculations on solids, u is well chosen as a harmonic potential, which has the additional advantage that, in this case, φ is analytically computable. We use the following form:

$$u=\varepsilon_0+\frac{1}{2}M_I\omega^2(\mathbf{R}-\mathbf{R_0})^2, \tag{3}$$

where ε_0 is the energy of an atom in the perfect crystal at $T=0$, $\mathbf{R}_0$ is the lattice site of the atom to be made to disappear (i.e., that in which the vacancy will appear), M_I the ionic mass, and ω is a typical lattice frequency, which we chose to be the experimental Debye frequency. In this case, the decoupled atom behaves approximately as a bulk atom.

We now exploit the fact that Φ has the form of a canonical average and can, therefore, be computed by MD:

$$\frac{\partial F(\lambda)}{\partial\lambda}=-k_BT\frac{\partial}{\partial\lambda}\ln\int d^{3N}Re^{-U(\lambda)/k_BT}$$

$$=\frac{1}{\mathscr{Z}_q(\lambda)}\int d^{3N}R\Phi(\lambda)e^{-U(\lambda)/k_BT}=\langle\Phi(\lambda)\rangle \tag{4}$$

where $\mathscr{Z}_q$ is the configurational integral entering the partition function. Then,

$$F(\lambda=1)-F(\lambda=0)=\int_0^1 d\lambda\langle\Phi(\lambda)\rangle. \tag{5}$$

The formation free energy is finally obtained as

$$F_f=F(\lambda=0)-F(\lambda=1)+\frac{1}{N}F_C-\varphi=-\int_0^1 d\lambda\langle\Phi(\lambda)\rangle$$

$$+\Delta F. \tag{6}$$

The correction factor ΔF is the difference between φ, which is known independently (here analytically), and the free energy per atom in the perfect crystal. The latter is easily obtained from accurate *ab initio* calculations of phonons or directly from experimental data.

The next step is now to find a potential energy $U(\lambda)$ satisfying the appropriate conditions at $\lambda=0$ and $\lambda=1$. We chose it in the following way.

(i) We multiply by λ the occupation numbers f_i of the states occupied by Z electrons ($Z=4$ being the valence of the Si ion).

(ii) We multiply by λ the ionic charge Ze of the ion to be decoupled and its bare pseudopotential.

To compute the thermal averages needed, we use FPMD, with the standard Car-Parrinello (CP) method,[7] with the following Lagrangian:

$$\mathscr{L}=\frac{1}{2}\sum_I^N M_I\ddot{\mathbf{R}}_I+\frac{1}{2}\sum_i \mu_i\langle\dot{\psi}_i|\dot{\psi}_i\rangle-E_{\mathrm{KS}}[\{\mathbf{R}_I\},\{\psi_i\}]-E_{II}$$

$$+\ \text{constraints}, \tag{7}$$

where E_{KS} is the Kohn-Sham energy functional, E_{II} is the direct ion-ion interaction and the ‘‘constraint’’ term imposes the orthonormality condition on the wave functions.

The electron pseudomasses μ_i must be chosen in such a way that the pseudoelectronic frequencies are much higher than all ionic frequencies.[13] When Z electrons are decoupled, however, the occupation numbers of the corresponding states become smaller by a factor λ and it is easy to see[13] that, for $\lambda\rightarrow 0$, this requirement will be violated. We therefore choose to set the pseudoelectronic masses as follows: for the electrons, which stay fully coupled, $\mu_i=\mu=$ constant, for the electron states corresponding to disappearing electrons $\mu_i=\lambda\mu$ (it should be noted that this affects the form of the orthonormalization procedure: the appropriate form is given in Ref. 10). With this choice, the less coupled electrons have correspondingly low masses and thus the frequency of their motion remains high.

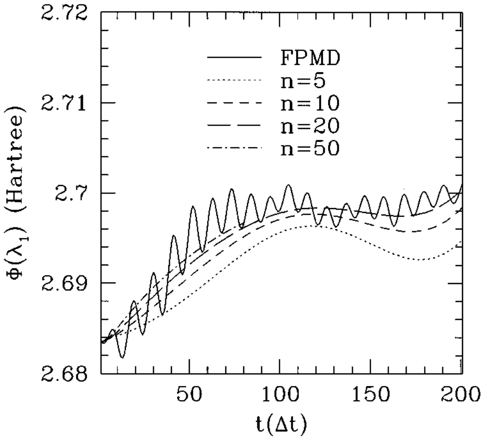

FIG. 1. Φ as calculated using the CP dynamical procedure (full line). The dashed, dotted, and dashed-dotted lines represent calculations done by minimizing the total electronic energy at every ionic step, using n steepest descent steps per each ionic step. $\Delta t=$ 10 a.u., $\mu=$ 875 a.u.

III. THE Si VACANCY

We applied the above scheme to the case of the Si vacancy. We used a 64 atom cell and the standard "local density approximation plus plane waves plus pseudopotential" scheme. We used the Γ point to sample the Brillouin zone, and a plane-wave cutoff of 8 Ry. These values are sufficient for giving phonon frequencies within 10% of the converged value. We found that using a sufficient number of λ points the results obtained from Eq. (4) at $T=0$ coincide with the ones obtained directly from Eq. (1). This is indeed the case, within 0.05 eV, if one uses a three-point Gauss-Legendre formula. It is possible to show that the states corresponding to the electrons being decoupled are necessarily the ones with the highest orbital energies. In our case these states become, at decreasing λ, the two highest states of the three introduced by the vacancy in the band gap. We verified that this indeed remains true, even during the dynamical runs.

We verified that the CP procedure gives correct results for Φ, by performing a series of calculations as follows. First, we evolved the system according to the CP Lagrangian using the same parameters that were subsequently used in the calculations $\mu=875$ a.u. and $\Delta t=10$ atu (atomic time units; 1 atu $=2.4188\ 10^{-5}$ ps). Second, we evolved the same system from the same initial conditions by minimizing explicitly the electronic energy at every ionic step, by using n steepest descent steps and computing the forces on ions via the Hellman-Feynman theorem; this must give the correct ionic evolution in the BO approximation in the limit $n\rightarrow\infty$. The results are shown in Fig. 1. The slow oscillation, which is present in all curves, comes from the ionic dynamics. The fast oscillation, present only in the CP curve, is due to the fast electronic motion around the instantaneous ground state. Apart from this oscillation, which should average to zero, it is seen that as n increases the broken curves converge to the full curve and for $n\geqslant 20$ they already lie within the oscillation band of the fast motion.

On longer runs we, however, observed that the usual criteria used to check for drifts from the BO surface are not sufficiently tight to guarantee accurate results for Φ; drifts which could be tolerated in other simulations are not acceptable here. Indeed, Φ is a differential quantity and as such it can be expected to be prone to large numerical errors. In this particular case, the problem arises because Φ mainly depends on the behavior of a small number of electrons, i.e., those that are being decoupled, which being the highest in energy are also the ones which are more likely to create problems with the adiabatic behavior.[13] A good strategy would be to force the electrons to stay close to the ground state by imposing a thermostat on them.[14] We used a simpler approach and just quenched periodically the electronic degrees of freedom on the instantaneous BO surface (one minimization every 100 steps). Of course, this implies that the system of electrons plus ions is not conservative anymore and, in fact, energy is gradually transferred from the hot ions to the cold electrons. We kept the ionic temperature constant by using a Nosé thermostat,[15] coupled to the ionic degrees of freedom only. This also ensured that under the assumption that the ionic subsystem is ergodic—for a discussion of this point see Ref. 9—the sampling of the ionic phase space is indeed canonical. We found that our method is sufficient to give accuracy on Φ better than 0.1 eV. We are then satisfied that, if the system is evolved via the CP dynamics and care is taken that the electrons do not drift away from the BO surface, the values obtained for Φ in this way are correct.

We run the program for two different temperatures, $T=500$ and $T=1000$, and the three Gauss-Legendre points of the interval $[0,1]$. We used a fixed cell, the lattice parameter a of which was chosen to be the experimental one at $T=0$. By performing a $T=0$ calculation of F_f at the value of a, corresponding to a temperature of 1400 K, we checked that thermal expansion would give a correction of the order of 0.1 eV or less. The value of F_f at $T=0$, which gives the largest contribution, was taken to be the one obtained by accurate calculation of the formation energy[5] and the dynamical runs were only used to compute the temperature-dependent part. We checked that a simulation time of 6 ps is sufficient to give reasonable thermalization and statistics, i.e., formation entropies with a statistical uncertainty of $\sim 1k_B$, this agrees with the findings of Ref. 9.

The results have been previously published[12] and here we only discuss them. The value at $T=500$ K agrees with the value obtained,[5] by using the same model of Si, but using the local harmonic approximation,[16] which shows that the anharmonic effects are still small at that temperature. Also, F_f decreases with increasing T. This is to be expected, as it is reasonable that the entropy of the defective crystal is larger than the perfect crystal and thus its free energy decreases more rapidly. It also looks like F_f is bent downward; this is an expected effect too, since the anharmonic effects, which become more and more important as T increases, are arguably larger in the defective crystal and so its entropy should increase faster than the entropy of the perfect crystal. In contrast, values calculated by using $T=0$ calculations in connection with standard statistical-mechanics methods[6] do not show any temperature dependence of the entropy.

It should be noted that the calculated values only include the ionic free energy. As the vacancy introduces closely spaced electronic levels in the band gap, only one of whom is occupied at $T=0$, there has to be a sizeable contribution to F_f coming from electronic excitations. To calculate this electronic contribution, one must have a good description of excited states, which is not the case within the local density approximation. This contribution, however, has been estimated to be approximately $1k_B$.[5] Even neglecting this, the calculated values for F_f agree well with the experimental estimates.[17] Our data, therefore, show that the observed high entropy of vacancies is a direct consequence of the vacancy-induced phonon softening. This may possibly be true for other point defects as well.

IV. CONCLUSIONS

We have shown how the thermodynamic integration method can be used within the Kohn-Sham scheme, including the cases in which the initial and final states have a different number of electrons and, thus, some electrons have to be decoupled from the rest of the crystal. The implementation of the method requires some care to avoid spoiling the adiabatic separation between the ionic and pseudoelectronic motion; we showed how these problems can be successfully overcome and how good results can be achieved.

ACKNOWLEDGMENTS

We benefited from stimulating discussions with P. E. Blöchl, S. T. Pantelides, G. De Lorenzi, A. Voter, and M. Parrinello. This work has been supported, in part, by the International School for Advanced Studies (SISSA)—Centro Interuniversitario del Nord-Est per il Calcolo Automatico (CINECA) collaborative project, under the sponsorship of the Italian Ministry for Public Education, by the European ARO Grant No. DAJA45-89-C-0025 and by the Italian National Research Council (CNR) Grant No. 89.0006.69.

---

[1] D. Frenkel, in *Molecular Dynamics Simulations of Statistical Mechanical Systems*, Proceedings of the International School of Physics ''Enrico Fermi,'' Course XCVII, Varenna, 1985, edited by G. Ciccotti and W. G. Hoover (North-Holland, Amsterdam, 1986).

[2] G. Jacucci, in *Diffusion in Crystalline Solids*, edited by G. E. Murch and A. S. Nowick (Academic Press, Orlando, 1984).

[3] S. U. Ghaisas, Phys. Rev. B **43**, 1808 (1993).

[4] D. Maroudas and R. A. Brown, Phys. Rev. B **47**, 15 562 (1993).

[5] P. E. Bloechl, E. Smargiassi, R. Car, D. B. Laks, W. Andreoni, and S. T. Pantelides, Phys. Rev. Lett. **70**, 2435 (1993).

[6] S. J. Clark and G. J. Ackland, Phys. Rev. B **48**, 10 899 (1994).

[7] R. Car and M. Parrinello, Phys. Rev. Lett. **55**, 2471 (1985).

[8] V. Milman *et al.*, Phys. Rev. Lett. **70**, 2928 (1993).

[9] E. Smargiassi and P. A. Madden, Phys. Rev. B **51**, 117 (1995).

[10] E. Smargiassi, Ph.D. thesis, Scuola Internazionale di Studi Avanzati, Trieste (I), 1990.

[11] A similar approach has also been used to calculate the melting point of Si *ab initio* [O. Sugino and R. Car, Phys. Rev. Lett. **74**, 1823 (1995)].

[12] R. Car, P. E. Blöchl, and E. Smargiassi, Mater. Sci. Forum **83-87**, 443 (1992).

[13] G. Pastore, E. Smargiassi, and F. Buda, Phys. Rev. A **44**, 6334 (1991).

[14] P. E. Blöchl and M. Parrinello, Phys. Rev. B **45**, 9413 (1992); E. S. Fois, J. I. Penman, and P. A. Madden, J. Chem. Phys. **98**, 6361 (1993).

[15] S. Nosè, J. Chem. Phys. **81**, 561 (1985).

[16] R. Najafabadi, R. LeSar, and D. J. Srolovitz, Phys. Rev. Lett. **63**, 624 (1989).

[17] S. Dannefaer, P. Mascher, and D. Kerr, Phys. Rev. Lett. **56**, 2195 (1985).