System-size convergence of point defect properties: The case of the silicon vacancy

Fabiano Corsetti[*] and Arash A. Mostofi

Thomas Young Centre, Imperial College London, London SW7 2AZ, United Kingdom
(Received 19 October 2010; revised manuscript received 10 May 2011; published 29 July 2011)

We present a comprehensive study of the vacancy in bulk silicon in all its charge states from 2+ to 2−, using a supercell approach within plane-wave density-functional theory, and systematically quantify the various contributions to the well-known finite size errors associated with calculating formation energies and stable charge state transition levels of isolated defects with periodic boundary conditions. Furthermore, we find that transition levels converge faster with respect to supercell size when only the Γ-point is sampled in the Brillouin zone, as opposed to a dense k-point sampling. This arises from the fact that defect level at the Γ-point quickly converges to a fixed value which correctly describes the bonding at the defect center. Our calculated transition levels with 1000-atom supercells and Γ-point only sampling are in good agreement with available experimental results. We also demonstrate two simple and accurate approaches for calculating the valence band offsets that are required for computing formation energies of charged defects, one based on a potential averaging scheme and the other using maximally-localized Wannier functions (MLWFs). Finally, we show that MLWFs provide a clear description of the nature of the electronic bonding at the defect center that verifies the canonical Watkins model.

I. INTRODUCTION

The presence of point defects can have many consequences for the optical, electrical, and mechanical properties of materials. In particular, the behavior of defects in semiconductors has been the subject of thorough and ongoing research due to their use in devices such as transistors, solar cells, and light-emitting diodes. This is because point defects strongly influence the electrical conductivity of semiconductors by adding states in the band gap, thus changing the number of charge carriers available. Therefore, both native defects (vacancies and self-interstitials) and impurity-related defects (dopants) play a crucial role in understanding and controlling the performance of semiconducting materials such as silicon, germanium, and gallium arsenide.

To this end, the study of point defects in semiconductors using first-principles electronic structure simulations has received a considerable amount of attention in recent years.[1] Most notably, density-functional theory[2] (DFT) has proven extremely popular due to its balance of computational speed and predictive success. Nevertheless, it has been noted that even for one of the simplest cases of point defects in semiconductors, that of the neutral vacancy in silicon, theoretical studies have shown a large scatter in results for basic quantities such as the defect formation energy.[3,4]

For a point defect in an infinite crystal lattice, the relaxation of atomic positions around the defect can extend to many successive shells of atoms. In simulations with period boundary conditions, this gives rise to elastic interactions between the defect and its images in neighboring supercells; the relaxation must therefore be contained within one supercell. Additionally, there is a spurious electrostatic interaction between a defect and its periodic replicas that depends on the size of the supercell.

We focus on the isolated vacancy in bulk silicon (denoted v^q, where q is the charge state of the defect center). A number of previous studies have been undertaken on this defect center using a variety of theoretical techniques;[5] in particular, Probert

and Payne,[4] Puska *et al*.,[6] and Wright[7] have performed studies of the convergence of the defect formation energy.

In this paper, we investigate systematically and quantify the main contributions to the finite size error that lead to the well-known slow convergence with respect to system size of, for example, the vacancy formation energy, which arise from spurious electrostatic interactions, elastic interactions, and wave function orthogonality constraints between periodic images of the defect center in the supercell approach. In addition, our calculations demonstrate that the defect formation energy and equilibrium charge state transition levels exhibit different convergence behavior with respect to supercell size, depending on the Brillouin zone sampling used: the former benefits from the use of a dense k-point grid, while the latter from sampling at the Γ-point only. The reasons for this difference will be discussed in the text. Furthermore, we present two simple and accurate methods for calculating the potential alignment correction to the valence band maximum of charged defect supercells, one based on averaging the potential with a Voronoi cell construction, the other on matrix elements between maximally-localized Wannier functions[8] (MLWFs). Finally, we relate the MLWFs associated with the defect center for each charge state to the canonical model of Watkins[9–11] (described in Fig. 5, Appendix) and show that the qualitative description given by this simple model is in full agreement with parameter-free DFT calculations.

The rest of the paper is organized as follows: in Sec. II, we describe the computational techniques that we employ and give the technical details of our simulations. In Sec. III, we illustrate the two methods we use to perform the potential alignment correction and show preliminary results comparing them. In Sec. IV, we present our main results; first, we describe the convergence properties of the unrelaxed (Sec. IV A) and relaxed (Sec. IV B) neutral vacancy, and then we describe the results obtained for different charge states of the defect (Sec. IV C) in terms of the transition levels (Sec. IV D). In Sec. V, we give a brief summary of our main conclusions.

II. COMPUTATIONAL METHODS

In the supercell approximation, the formation energy of the vacancy E_f^q is defined (following Zhang and Northrup[12]) as

$$E_f^q = E_{\text{vac}}^q - \left(\frac{N-1}{N}\right) E_{\text{bulk}} + q\mu_e, \qquad (1)$$

where E_{bulk} is the total energy of the bulk (perfect crystal) supercell, E_{vac}^q is the total energy of the same supercell containing a single vacancy with charge q, N is the number of atoms in the bulk supercell, and μ_e is the electron chemical potential. This last term can be divided into $\varepsilon_v + \Delta\mu_e$, where ε_v is the value of the valence band maximum (VBM) and $\Delta\mu_e$ is the position of the Fermi level in the band gap. The determination of ε_v is discussed in more detail in Sec. III.

For supercell sizes up to and including 256 atoms, the calculations are performed using the CASTEP (Ref. 13) code (version 5.0). The Ceperley-Alder local-density approximation[14] (LDA) is used to describe exchange and correlation. For charged supercell calculations, a compensating uniform background charge is added. We use two pseudopotentials for silicon: CASTEP's "on-the-fly" Vanderbilt ultrasoft pseudopotential,[15] and a norm-conserving pseudopotential;[16] both have four valence electrons and give accurate lattice constants for bulk silicon (5.39 and 5.38 Å, respectively, compared with an experimental value of 5.43 Å). This slight underestimation of less than 1% is typical for LDA DFT. The DFT-optimized lattice parameter is used in all calculations.

For our largest calculations, on a 1000-atom simple cubic (sc) supercell, we use the linear-scaling DFT code ONETEP (Refs. 17 and 18) (version 2.4). The "crossover," namely, the system size at which it becomes computationally more efficient to use ONETEP as opposed to conventional cubic-scaling DFT, is highly system dependent and lies at around 500 atoms for silicon. For this reason, we use conventional plane-wave DFT for supercells smaller than this crossover, and linear-scaling DFT for the largest supercell. ONETEP makes use of the single-particle density matrix, expressed in separable form in terms of a localized basis of nonorthogonal generalized Wannier functions (NGWFs) and a density kernel.[17,18] We use the same norm-conserving pseudopotential as for CASTEP, and nine NGWFs on each silicon atom with a truncation radius of 3.97 Å. We do not truncate the density kernel.

We follow the methodology outlined by Probert and Payne in their study of the neutral silicon vacancy;[4] however, numerical parameters are converged independently for all system sizes. The supercells are constructed from three unit cell shapes: fcc (the primitive cell, with 2 atoms), sc (with 8 atoms), and bcc (with 32 atoms). The supercells are then made from n^3 unit cells. We perform calculations on both unrelaxed and relaxed geometries; our convergence tolerance for the defect formation energy of a given supercell is 10 meV. We use a regular Monkhorst-Pack (MP) mesh[19] of k-points for the Brillouin zone integration. In the text, the parameter k_{MP} refers to a k_{MP}^3 grid. The issue of Brillouin zone sampling shall be discussed in Sec. IV; in general, we make use of two sampling schemes: a Γ-point only sampling (i.e., $k_{\text{MP}} = 1$), and what we call a "dense" sampling. This last term we shall use to indicate that k_{MP} has been converged with respect to the formation energy for a particular supercell.

For the relaxation, a quasi-Newton BFGS scheme is used. All the atoms in the supercell are allowed to move, and their positions are slightly randomized at the start of the procedure to allow for symmetry breaking. Our convergence tolerance for the geometry optimization is 5×10^{-3} eV/Å for the root mean square force on all the atoms.

III. DETERMINING THE ELECTRON CHEMICAL POTENTIAL

As described in Sec. II, the formation energy of charged defects depends on the electron chemical potential, which is given relative to the VBM eigenvalue position ε_v. ε_v is more precisely defined as the energy difference between the pure host material and the host with a single electron hole. In the limit of a dilute hole gas, this energy difference is equivalent to the VBM eigenvalue. For a finite supercell calculation, however, the energy difference between the bulk system with and without the hole only slowly converges to the infinite limit as the system size is increased.[20] Instead, the VBM will always remain at the same position (given a sufficient Brillouin zone sampling), which is the limiting value.

Two approaches can be taken to determine ε_v: either calculating the energy difference of the bulk supercell with $q = 0$ and $q = +1$, or using the bulk VBM eigenvalue and employ a potential alignment correction.[20–22] In the rest of this section, we describe the two methods that we employ for calculating the potential alignment correction: a partitioning of the real-space potential using Voronoi cells, and an MLWF-based approach. To the best of our knowledge, this is the first time these methods have been used to calculate the correction.

A. The real-space Voronoi cell method

Our first method is based on considering the electrostatic potential directly. In such approaches, the correction is typically determined by plotting the electrostatic potential obtained from the DFT calculation, and averaging it either in the x–y plane, within atomic spheres or over primitive cells of the host material. The important point is that the averaging must be done in some localized manner, such that it is possible to measure the value of the bulk-like potential in the charged defect supercell by considering a region far away from the defect center. The correction to the defect formation energy is then $q(V_{\text{vac}} - V_{\text{bulk}})$, where V_{bulk} and V_{vac} are the average potentials in the bulk supercell and in a "bulk-like" region of the vacancy supercell.

In order for the correction to be unambiguous and accurate, the averaging volume should be as small as possible while still covering a region of space that is completely representative of the bulk material (i.e., all regions in the bulk supercell should give the macroscopic average for the material). Our approach is to use Voronoi cells: each region corresponds to the real-space volume of points closest to one particular atom in the supercell. There are, therefore, as many regions as there are atoms. The electrostatic potential, however, is given on a discrete real-space grid whose spacing depends on the basis set cut-off energy and which, in general, is not commensurate with the atomic spacing, leading to inaccuracies in the averaging procedure that may swamp the difference between the bulk
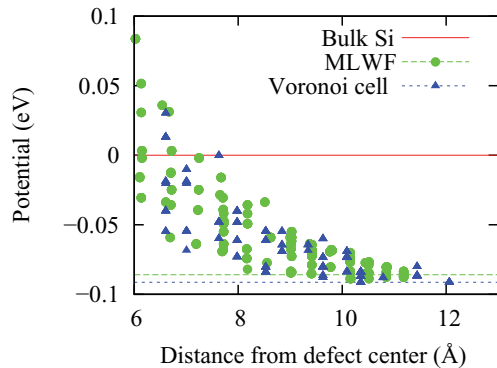
FIG. 1. (Color online) Potential alignment correction for v^{2-} using the MLWF and real-space Voronoi cell methods. The values are adjusted with respect to the bulk supercell. The dashed horizontal lines show the value for the furthest point from the defect center.

and defect supercells. In order to overcome this drawback, we Fourier interpolate the potential onto a finer grid that is commensurate with the atomic spacing, which results in each Voronoi cell containing exactly the same number of points. We note that care must be taken to give the correct fractional weighting to points which are directly between two or more atoms.

The potential alignment correction is then determined by considering Voronoi cells belonging to the atoms which are furthest from the defect center, as shown in Fig. 1.

B. The MLWF method

Our second approach to calculating the potential alignment correction is based on matrix elements of the Hamiltonian in the basis of MLWFs. The shift in the reference of potential that is sought is reflected in the position of the energy eigenvalues; however, comparing the eigenvalues of bulk-like states in the defect supercell with those of the bulk supercell is problematic since the eigenstates are delocalized. MLWFs, on the other hand, provide a probe of the localized properties of the system. Once the Hamiltonian is transformed into a basis of MLWFs, the potential alignment correction can be determined by considering on-site matrix elements $\tilde{H}_{nn} = \langle \omega_n | \hat{H} | \omega_n \rangle$ (where ω_n is a MLWF belonging to the cell at the origin) for Wannier functions whose centers, defined by their first moments, are far away from the defect center. The shift in potential is then simply $\tilde{H}_{nn,\text{vac}} - \tilde{H}_{nn,\text{bulk}}$. In practice, as shown in Fig. 1, we plot this difference for all n as a function of the distance of the Wannier function center from the defect center, and the value of the potential alignment correction is taken as the mean of the values which are furthest from the defect center.

This method is feasible since the wannierization procedure results in the same partitioning of the electronic density in the bulk and defect supercells everywhere except in the direct neighborhood of the defect; hence it is possible to identify Wannier functions that are associated with the defect center and Wannier functions that are not. In the latter case, therefore, it is straightforward to unambiguously match equivalent Wannier functions in the bulk and defect supercells.

From Fig. 1, it can be seen that the MLWF correction method is in an excellent agreement with the Voronoi cell

TABLE I. List of all supercells up to 256 atoms with their respective symmetries. Also listed are the value of k_{MP} used (converged with respect to the formation energy for each supercell), the corresponding k-point volume in reciprocal space, the unrelaxed defect formation energy, and the kinetic energy contribution to this value. Calculations were performed with an ultrasoft pseudopotential and a plane-wave energy cutoff of 400 eV.

N	Symmetry	k_{MP}	k-point volume (10^{-3}Å^{-3})	E_f^0 (eV) Total	Kinetic
2	fcc	8	12.35	2.65	−10.19
8	sc	7	4.61	3.13	−11.49
16	fcc	6	3.66	3.10	−11.68
32	bcc	6	1.83	3.83	−12.08
54	fcc	4	3.66	3.51	−12.36
64	sc	6	0.91	3.81	−12.45
128	fcc	4	1.54	3.79	−12.53
216	sc	3	2.17	4.05	−12.51
250	fcc	3	1.87	3.95	−12.60
256	bcc	3	1.83	4.13	−12.47

method. Among all our calculations for the various charge states, the maximum discrepancy was 0.01 eV. The MLWF method gives a finer representation of the system than averaging over atomic sites (as in the Voronoi cell approach), since each silicon atom has four bonding Wannier functions connecting it to its neighbors. Therefore, there are twice as many Wannier functions as atoms, and hence twice as much information to consider.[23]

MLWFs are now a standard tool for the analysis of electronic structure calculations; therefore, this approach provides a simple and accurate method for calculating the potential alignment correction, since all the necessary information is readily available once the standard wannierization procedure has been performed.[24]

IV. RESULTS

A. The neutral unrelaxed vacancy

In order to study the finite size convergence properties of the system, we first simulate the neutral vacancy in its unrelaxed state, with all the atoms in their perfect crystalline positions; our results are reported in Table I. Considering the different supercell geometries separately, it can be seen that E_f^0 increases monotonically with system size. We therefore take our highest value of 4.13 eV (for the 256-atom bcc supercell) as a lower bound to the unrelaxed defect formation energy.

The slow convergence of the formation energy as a function of system size is due to the spurious interaction between periodic images of the defect center. As the defect is both unrelaxed and uncharged, this is neither caused by elastic interactions nor monopole-monopole electrostatic interactions; rather, it is the result of higher multipole interactions and overlap between the wave functions of the periodically repeated defect centers.

The importance of wave function overlap may be gauged by considering the kinetic energy contribution to the defect formation energy, which we define analogously to Eq. (1), except using only the noninteracting kinetic energy

contribution to the total energy instead of the total energy itself. The results are listed in the right-hand column of Table I. It is immediately apparent that the kinetic energy component of the defect formation energy varies on a scale *larger* than the total formation energy as the supercell size is increased, and that, even at our largest supercell sizes, it is not converged to better than 0.1 eV. These changes are caused by subtle changes in the defect wave functions of neighboring defect centers that occur in order to maintain orthogonality as the periodic images of the defect move apart. While the electronic density is rather insensitive to them and, therefore, the electrostatic interactions are not affected, the kinetic energy is not. As a result, any correction scheme that accounts solely for the classical electrostatic interaction will not be sufficient in this case to predict the correct formation energy for an infinite system beyond this level of accuracy.

We consider one last question before moving on to ionic relaxations: is it desirable to use a dense k-point grid when simulating point defects? Although this is obviously necessary for describing the delocalized bulk states, it seems reasonable that a Γ-point sampling might give a better approximation of the localized defect levels in the dilute limit, as a way of deliberately eliminating unwanted dispersion, and by occupying the state which most closely resembles the bonding defect state for the infinite system.[25] Indeed, Fig. 2 shows that the position of the defect level at Γ rapidly converges to a fixed value, even if the total dispersion of the level throughout the Brillouin zone is large.

As far as the defect formation energy is concerned, at least, our results show that Γ-point calculations do not give a better estimate of the defect formation energy: for the 256-atom supercell, E_f^0 is underestimated (as compared to our lower bound estimate) both for the unrelaxed and relaxed cases, by 0.2 and 0.4 eV, respectively. This conclusion is in agreement with two other studies on the relaxed silicon vacancy: Puska *et al.*,[6] who report that a 2^3 MP sampling shows a faster convergence than a Γ-point sampling for the neutral charge state; and Wright,[7] whose results for the 216-atom supercell show a greater average error across all charge states for Γ-point sampling than any finer MP grid sampling. We note that, conversely, quantities of interest other than the defect formation energy might benefit from a Γ-point only sampling;

as we shall see in Sec. IV D, the stable charge state transition levels are an example of this.

B. The neutral relaxed vacancy

On relaxation of the ionic positions, the predicted symmetry for v^0 from Watkins' model (D_{2d}) is not seen in any supercell smaller than 256 atoms, as shown in Fig. 3. Calculations on the 32-atom bcc, 64-atom sc, and 128-atom fcc supercells do not exhibit a change in symmetry, only an inward relaxation. Consequently, the defect formation energy is only lowered by $\sim$0.2 eV in these cases.

The 256-atom bcc supercell undergoes the predicted change in symmetry and a reduction in defect volume by more than 40%. This results in the defect formation energy being lowered by 0.7 eV (full results are given in Table II). Therefore, even though the unrelaxed defect formation energy is highest for the largest supercell, after relaxation it becomes the lowest, demonstrating the well-known importance of Jahn-Teller distortion and the long-ranged nature of the elastic interactions which can lead to both qualitatively incorrect relaxation patterns and quantitatively inaccurate formation energies in small supercells.

These results confirm several previous DFT studies on this system.[4,6,7] In particular, Puska *et al.*[6] report that small supercells can show a range of different symmetries depending on the k-point sampling, and there is some evidence from their results that a Γ-point only sampling favors D_{2d}. It is clear from our results that a dense k-point sampling favors the lattice's unrelaxed tetrahedral point group symmetry T_d for small supercells.

Figure 3 also shows the relaxed defect formation energy in a 1000-atom sc supercell, as calculated with the linear-scaling
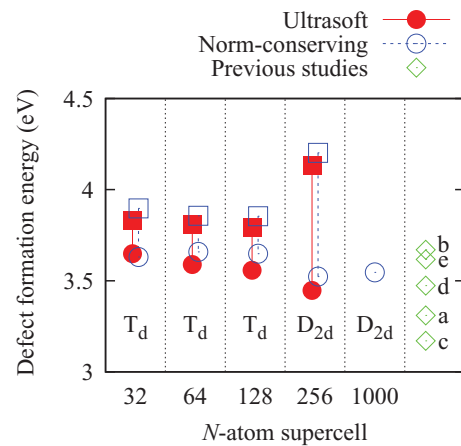
FIG. 3. (Color online) Relaxation effects for the neutral vacancy. Squares show the defect formation energy for the unrelaxed lattice and circles for the relaxed one. Labeled below is the point group symmetry of the defect center after relaxation. The last column shows the relaxed formation energy from previous studies: (a) Ref. 6 ($N = 216$, $k_{MP} = 2$, LDA); (b) Ref. 3 ($N = 216$, $k_{MP} = 2$, GGA); (c) Ref. 4 ($N = 256$, $k_{MP} = 2$, GGA); (d) Ref. 7 ($N = 1000$, $k_{MP} = 3$, LDA); (e) Ref. 7 ($N = 1000$, $k_{MP} = 3$, GGA). Our calculations were performed with an ultrasoft (norm-conserving) pseudopotential and the LDA functional, with a converged plane-wave energy cutoff of 400 eV (800 eV) and a k-point grid of $k_{MP} = 3$ ($k_{MP} = 4$).
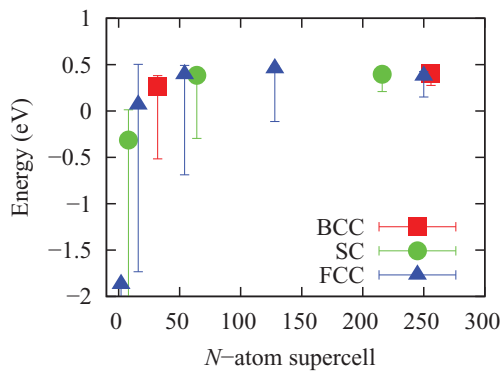
FIG. 2. (Color online) Dispersion of the unrelaxed defect level. The symbols show the position of the level at Γ, and the bars show the extent of the dispersion obtained from our dense k-point sampling. The positions are given with respect to the bulk VBM.

DFT code ONETEP.[26] The result (calculated with the same pseudopotential) is within 23 meV of the 256-atom bcc system, and the relaxed defect structure is almost identical, with only a 33 mÅ root mean square difference in the lengths of the bonds at the defect center.

C. Charged vacancies

Our results for the relaxed symmetries and defect formation energies of the various charge states of the vacancy are shown in Table II. All the charge states apart from v^{2-} are in agreement with Watkins' model. For v^{2-}, a completely different "split vacancy" configuration is found, with a point group symmetry of D_{3d} (in agreement with previous studies by Nieminen *et al.*[3,6] and Wright[7]). The symmetry predicted by the model (C_{2v}) was found as a metastable state. For all charge states, there were no qualitative differences between the Γ-point only and multiple k-point calculations, although the former consistently showed a greater inward relaxation and thus a smaller defect volume. For a detailed discussion of the electronic structure of the different charge states of the vacancy, the reader is referred to the Appendix.

We note that the relaxation of v^{1-} is the most problematic; in fact, if the initial random displacement of the atomic positions is not large enough, the system consistently converges to a metastable T_d configuration with a completely spin-polarized triply degenerate defect level (i.e., filled with three spin-aligned electrons). Furthermore, our final lowest configuration is only approximately C_{2v}, as there is a small but noticeable splitting of $\sim$0.1 Å of the bond lengths between the two pairs

of neighbors of the vacancy. Nieminen and Wright both report that the LDA does not give the expected symmetry for this charge state, although they find a D_{3d} configuration.

As explained in Sec. III, there are two possible approaches for estimating the position of ε_v in the system: we can either use a calculation of the bulk supercell with an electron hole (denoted "Hole" in Tables II–III), or the bulk VBM value with a potential alignment correction[27] ("VBM"). The agreement between the two approaches is better for the positively charged vacancies (-0.05 to 0.01 eV) than the negatively charged ones (0.05 to 0.21 eV); however, these uncertainties are small enough not to affect the level ordering.

D. Transition levels

Following the careful description of Baraff *et al.*,[10] we define the stable charge state transition level $E(m/n)$ to be the value of the Fermi level (with respect to the perfect crystal valence band edge) at which there is a crossing of the defect formation energies of two charge states m and n, leading to a change in the most stable state from v^m to v^n as the Fermi level is raised. This is given by

$$E(m/n) = \frac{E_f^n(\Delta\mu_e = 0) - E_f^m(\Delta\mu_e = 0)}{m - n}. \quad (2)$$

As first predicted by Baraff *et al.*[10] and then experimentally verified by Watkins and Troxell,[28] the isolated silicon vacancy exhibits a "negative-U" effect, by which the stable charge state changes directly from v^{2+} to v^0. The v^{1+} state is therefore only

TABLE II. Summary table of the results obtained for the 256-atom bcc supercell and the 1000-atom sc supercell. Bond lengths are given between the atoms surrounding the vacancy, as labeled in Fig. 6(a) for v^{2+}, v^{1+}, v^0, v^{1-}, and Fig. 6(g) for v^{2-}. Atoms labeled with the same letter are equivalent by symmetry. $\Delta\alpha = \alpha - 90°$ is the distortion of the regular octahedron in the split vacancy configuration. The defect volume is calculated from the tetrahedron formed by the four neighbors of the vacancy site. The defect formation energy is given for $\Delta\mu_e = 0$. Symmetry labels are given in Schönflies notation.

Charge	N	k_{MP}	Relaxed	Pseudopotential	Symmetry	Bond lengths (Å) a–a	b–b	a–b	Defect volume (Å^3)	E_f^q (eV) Hole	VBM
2+	256	1	Yes	Ultrasoft	T_d	3.22	3.22	3.22	3.93	2.86	2.85
	256	3	Yes	Ultrasoft	T_d	3.63	3.63	3.63	5.61	3.62	3.66
1+	256	1	Yes	Ultrasoft	D_{2d}	2.93	2.93	3.27	3.62	2.99	3.00
	256	3	Yes	Ultrasoft	D_{2d}	2.86	2.86	3.40	3.73	3.50	3.55
	1000	1	No[a]	Norm conserving	D_{2d}	2.87	2.87	3.37	3.70	–	3.50
0	256	1	No	Ultrasoft	T_d	3.81	3.81	3.81	6.54	3.90	
	256	3	No	Ultrasoft	T_d	3.81	3.81	3.81	6.54	4.13	
	256	1	Yes	Ultrasoft	D_{2d}	2.83	2.83	3.31	3.53	3.06	
	256	3	Yes	Ultrasoft	D_{2d}	2.86	2.86	3.40	3.73	3.45	
	256	4	No	Norm conserving	T_d	3.81	3.81	3.81	6.50	4.20	
	256	4	Yes	Norm conserving	D_{2d}	2.86	2.86	3.41	3.74	3.52	
	1000	1	Yes	Norm conserving	D_{2d}	2.87	2.87	3.37	3.70	3.55	
1−	256	1	Yes	Ultrasoft	$\sim C_{2v}$	3.10	2.70	3.24 (av.)	3.50	3.48	3.44
	256	3	Yes	Ultrasoft	$\sim C_{2v}$	2.94	2.87	3.26 (av.)	3.55	3.80	3.71
						a–a	a–c	$\Delta\alpha$			
2−	256	1	Yes	Ultrasoft	D_{3d} (split)	3.45	2.60	6.63°	2.86	3.70	3.57
	256	3	Yes	Ultrasoft	D_{3d} (split)	3.46	2.60	6.52°	2.88	3.92	3.72

[a]For this system, we use the relaxed atomic positions obtained for the 1000-atom v^0 supercell, without further relaxation.

TABLE III. Transition levels for the 256-atom bcc supercell and the 1000-atom sc supercell. Results from previous studies are also shown. Our calculations use the LDA functional. Crosses (×) are used when there is no transition in the band gap. Asterisks (*) denote thermodynamically stable transitions. All values are given in eV.

| | | $N = 256$ | | | | $N = 1000$ | Reference 6[a]
$N = 216$ | Reference 3[a]
$N = 216$ | Reference 7[a]
$N = 1000$ | |
| | | $k_{MP} = 1$ | | $k_{MP} = 3$ | | $k_{MP} = 1$ | $k_{MP} = 1$ | $k_{MP} = 2$ | $k_{MP} = 3$ | |
	Experimental (Ref. 9)	Hole	VBM	Hole	VBM	VBM	LDA	GGA	LDA	GGA
$E(2+/1+)$	0.13	0.13	0.16	×	×	–	0.19	0.13	0.27*	0.27
$E(2+/0)$	0.09*	0.10*	0.11*	×	×	–	0.15*	0.10*	0.28	0.19*
$E(1+/0)$	0.05	0.07	0.06	×	×	0.04	0.11	0.06	0.28*	0.12
$E(0/1-)$	–	0.43	0.38	0.35	0.26	–	0.57	0.37	0.76	0.63
$E(0/2-)$	–	0.32*	0.25*	0.24*	0.13*	–	0.49*	0.37*	0.66*	0.53*
$E(1-/2-)$	–	0.22	0.13	0.13	0.00	–	0.40	0.36	0.56	0.42

[a]The transition levels have been calculated from the quoted values of the formation energy of the various charge states using Eq. (2).

metastable; this is a direct consequence of the Jahn-Teller distortion, which lowers the energy of the neutral vacancy much more than the singly positive vacancy. It has also been suggested that the same effect might be observed in the sequence v^0, v^{1-}, v^{2-}.

Our transition levels are given in Table III; as can be seen, the Γ-point calculations predict the negative-U behavior both for the negatively and positively charged levels as expected, and are in a good agreement with the available experimental results (also shown in Fig. 4). Surprisingly, the calculations using a dense k-point sampling ($k_{MP} = 3$) give the opposite ordering for the sequence v^{2+}, v^{1+}, v^0; this results in no transitions in the band gap between these levels, and only v^0 being a stable charge state. The negative-U behavior is, however, still present for the negatively charged levels, although their positioning is lower than for the Γ-point calculations. The Γ-point, therefore, gives better estimates of the transition levels. However, as stated previously, the
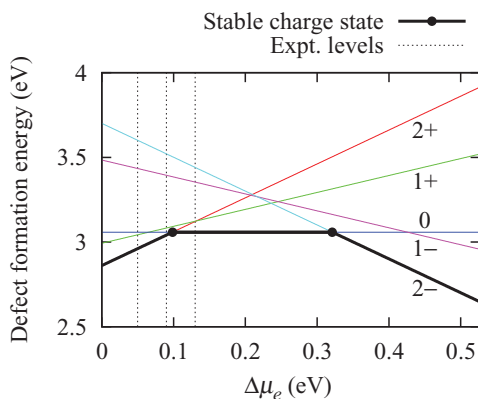

FIG. 4. (Color online) Formation energy of the different charge states of the vacancy as a function of the electron chemical potential (plotted relative to the VBM). The energy range shown covers the DFT band gap for silicon. The thermodynamically stable charge state at each point is highlighted in bold, and the circles indicate the level position for the stable transitions. The dashed vertical lines show the available experimental values for the transition levels. The results shown are for the 256-atom bcc supercell with a Γ-point sampling.

absolute value of the defect formation energy is not as well converged with respect to the dilute limit when using this Brillouin zone sampling. However, calculations of v^0 and v^{1+} in the 1000-atom supercell at Γ (also included in Tables II and III) are converged with respect to system size both in terms of the transition level $E(1+/0) = 0.04$ eV and the absolute value of the formation energy $E_f^0 = 3.55$ eV. Therefore, this suggests that at this system size, a Γ-point sampling can be employed for simultaneous convergence of both quantities of interest.

V. CONCLUSIONS

We have studied the silicon vacancy in all its charge states using the supercell approach and plane-wave pseudopotential DFT. Our calculations confirm the slow finite size convergence of defect formation energies and transition levels, due to electrostatic interactions and wave function overlap between periodic images of the defect, and long-ranged ionic relaxations. The impact of each of these has been quantified, and it has been found that all three provide non-negligible contributions to the total error. Furthermore, due to the hybridization of the defect levels with the perfect crystal band structure, we find that the choice of k-points has a noticeable impact on the results. In particular, Γ-point calculations converge faster than calculations with uniform, multiple k-point sampling when considering stable charge state transition levels (given by differences in formation energies of different charge states), and vice versa when considering the absolute value of formation energies.

Due to this slow convergence, the calculations may still benefit from the use of even larger supercells than the ones currently employed. This, however, will present additional numerical challenges: the defect formation energy is an *intensive* quantity that is obtained by taking the difference of two total energies that are *extensive* with the system size. As a result, to achieve a given accuracy for E_f as the supercell size is increased, it is necessary to proportionately increase the precision per atom in the total energy calculations in order to avoid the result being swamped by numerical noise.

We have also introduced two methods for correcting the alignment of the valence band maximum for charged defects:

one is based on averaging the electrostatic potential using a Voronoi cell construction, and the other on Hamiltonian matrix elements in a basis of maximally-localized Wannier functions. The two methods give an excellent mutual agreement and constitute simple and robust ways to calculate potential alignment corrections. However, the determination of ε_v remains a somewhat uncontrolled source of error, as demonstrated by the discrepancy between the potential alignment correction ("VBM") and total energy difference ("Hole") approaches (described in Sec. III), which can be as small as 0.01 eV (for positively charged defects), or as large as 0.2 eV (for negatively charged defects), as shown in Table II (right hand columns). We note that the problem of correcting errors in charged supercells arising from periodic boundary conditions has been addressed by many previous studies.[21,29–33]

The accuracy of the LDA functional for describing exchange and correlation should also be investigated. GGA calculations have already been shown to give very similar qualitative and quantitative results (as shown in Sec. IV); both local and semilocal functionals, however, suffer in particular with respect to the well-known effect of self-interaction error. Unfortunately, higher accuracy methods are at present confined to relatively small system sizes due their computational expense, and are therefore limited by the large finite size errors that we have discussed earlier. Nonetheless, qualitative differences have been suggested for the vacancy from screened-exchange calculations.[34]

Finally, it is interesting to note that maximally-localized Wannier functions give a chemically intuitive picture of the electronic structure at the vacancy site, and without any external input confirm Watkins' LCAO model of the defect. The only charge state which does not follow the predictions from the model (the doubly negative vacancy) is also explained in terms of its Wannier functions. Full details are given in the Appendix.

APPENDIX: VISUALIZATION USING MLWFs

Wannier functions provide a localized view of the electronic structure. This is particularly useful in the study of point defects, since the Bloch states associated with the defect levels are not completely localized due to the periodicity of the system; in fact, in small supercells, it is not possible to disentangle them from the bulk band structure.

In the case of bulk silicon, previous work has shown that it is possible to recover the typical σ bond orbitals between silicon ions by wannierization of the valence band; alternatively, the bottom conduction bands (once disentangled from higher bands) give the corresponding antibonding orbitals, while treating both these sets of bands together as a single manifold

produces four sp^3 orbitals on each ion.[8,35] The bulk silicon σ orbitals obtained from our calculations are shown in Fig. 6(b).

For the wannierization of the defect systems, we only use the occupied manifold, except in the case of the unrelaxed vacancy where we include all the defect levels. We can loosely define the concept of a defect Wannier function to be any MLWF in the defect supercell which differs qualitatively from the σ bond MLWFs of the bulk system. For all charge states, such defect Wannier functions are only present within the first ionic shell around the vacancy; between the first and second shells, we already recover σ-like bonding orbitals.

Figure 6(c) shows the defect Wannier functions for the neutral unrelaxed system; as expected from Watkins' model (described in Fig. 5), these are sp^3 orbitals pointing toward the vacancy. These orbitals cannot be obtained by wannierizing the three visible defect levels in the gap alone, as the nodeless combination which lies within the valence band is also needed; therefore, the entire valence manifold plus the defect levels in the gap must be wannierized as a whole.

The defect Wannier functions obtained for v^{2+}, v^0, and v^{1-} in their relaxed configurations, as shown in Figs. 6(d)–6(f), also
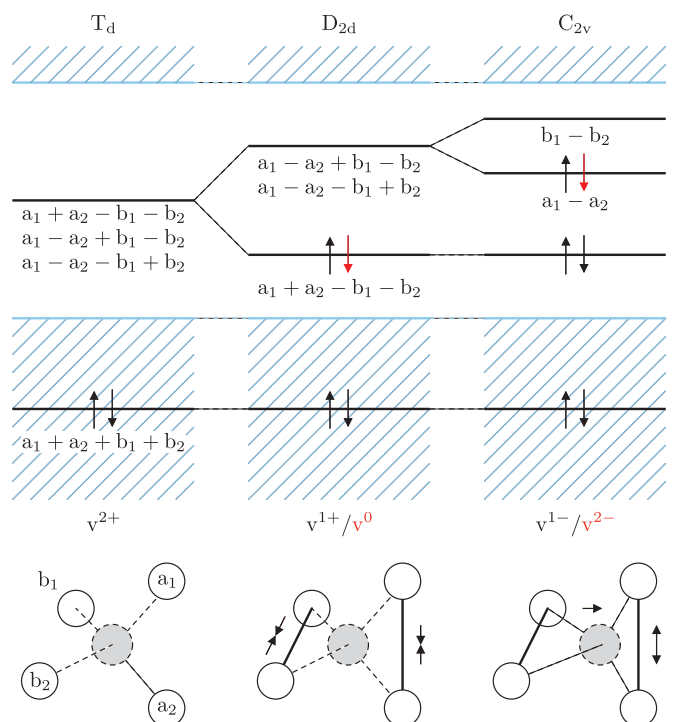

FIG. 5. (Color online) Watkins' LCAO model for the silicon vacancy, deduced from electron paramagnetic resonance (EPR) studies of the defect center (Ref. 36), and later confirmed by electron-nuclear double resonance (ENDOR) measurements (Refs. 37 and 38). The four orbitals associated with the neighboring atoms of the defect site are the dangling bonds resulting from the removal of the central silicon atom. The figure shows the predicted ionic configuration and point group symmetry for different charge states of the vacancy due to Jahn-Teller distortion (Ref. 39). For the D_{2d} and C_{2v} configurations, the first charge state listed in the figure refers to the spin-polarized case with only one electron in the highest occupied orbital, and the second charge state refers to the non-spin-polarized case with two electrons.
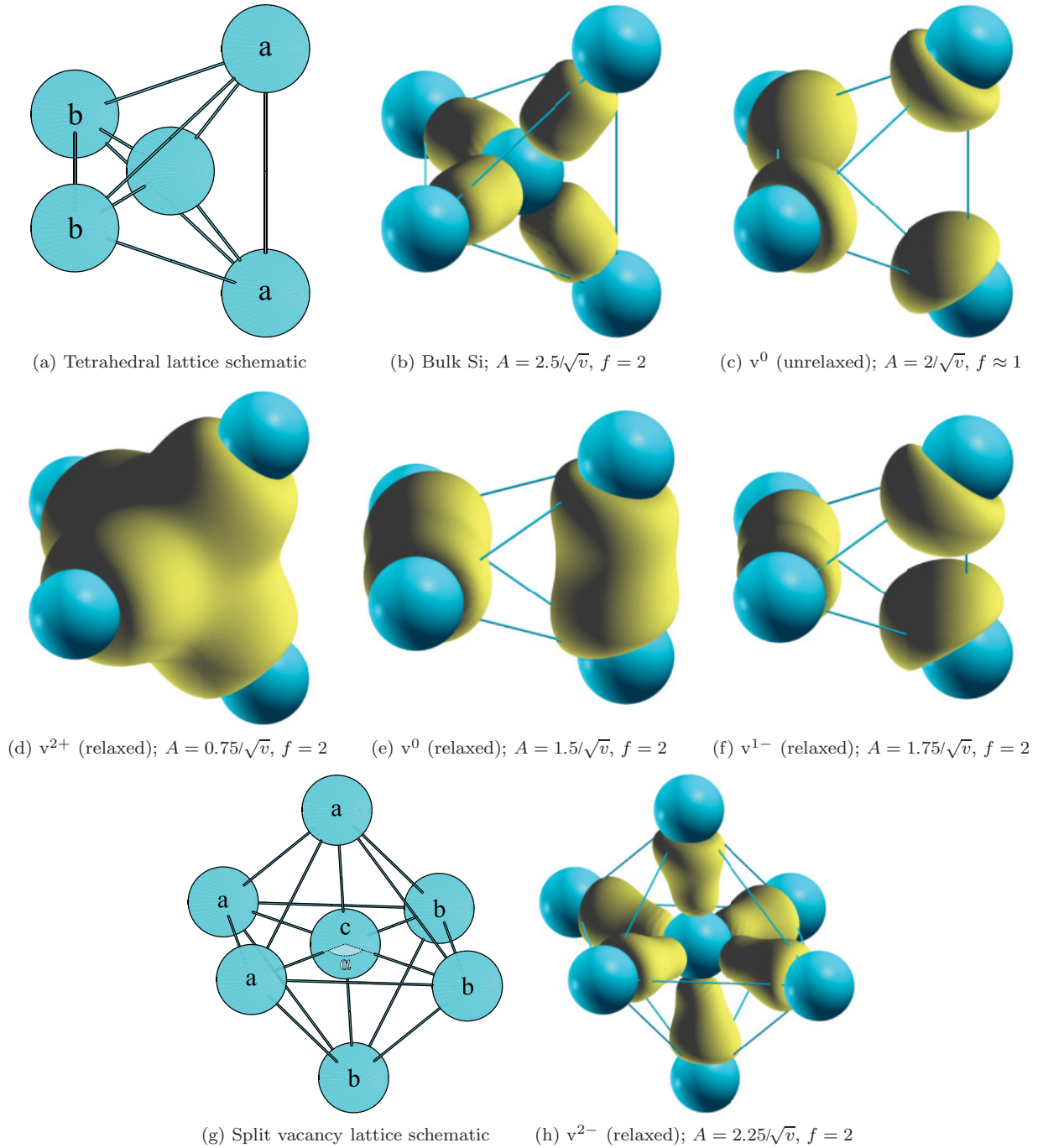
(a) Tetrahedral lattice schematic

(b) Bulk Si; $A = 2.5/\sqrt{v}$, $f = 2$

(c) v^0 (unrelaxed); $A = 2/\sqrt{v}$, $f \approx 1$

(d) v^{2+} (relaxed); $A = 0.75/\sqrt{v}$, $f = 2$

(e) v^0 (relaxed); $A = 1.5/\sqrt{v}$, $f = 2$

(f) v^{1-} (relaxed); $A = 1.75/\sqrt{v}$, $f = 2$

(g) Split vacancy lattice schematic

(h) v^{2-} (relaxed); $A = 2.25/\sqrt{v}$, $f = 2$

FIG. 6. (Color online) Contour-surface plots of the MLWFs most strongly associated with the defect center for different charge states of the vacancy, calculated using the 256-atom bcc supercell and a norm-conserving pseudopotential; for computational efficiency, we only include the Γ-point wave function in the wannierization procedure, since we expect a negligible difference in the qualitative features of the resulting Wannier functions. Each separate closed surface is an individual Wannier function. The amplitude A of the contour shown in each figure is given in the caption in terms of the primitive cell volume v. In general, the Wannier functions also have small negative components on the backbonds (not pictured). f is the electronic occupancy of a single orbital. The wannierization procedure is performed using the WANNIER90 (Ref. 24) code (version 1.2).

support Watkins' model. For the doubly positive vacancy, the defect levels in the gap are empty and therefore not included; the result is a single MLWF corresponding to the symmetric s-type nodeless combination of the four sp^3 orbitals. The neutral vacancy includes one lowered defect level only, and produces the two bonding orbitals between the pairs of nearest neighbors

of the defect center. The singly negative vacancy includes two defect levels in the gap (for the majority spin), producing one bonding orbital (between the pair of ions with the shortest bond length) and two sp^3 orbitals.

Finally, we show the split vacancy configuration obtained for v^{2-} in Figs. 6(g)–6(h). In this case, one of the neighbors

(labeled c) moves halfway along the line connecting it with the vacancy site, thus placing itself at the center of an octahedron made up of two pairs of second-nearest neighbors (labeled a and b). This octahedron is approximately regular (having as faces equilateral triangles); however, there is a small distortion due to the difference in distance between a–a/b–b pairs and a–b pairs. This distortion is quantified by the angle α between any three ions a, c, b; for a regular octahedron, α is a right angle, and so the distances a–a, b–b, a–b are identical. In general, the split configuration can be specified with only two parameters: the distortion angle α and the distance a–c between the central ion and any of the ions forming the octahedral cage.

The wannierization of the occupied manifold produces six defect Wannier functions, each one a bond between the central ion and one of its neighbors. This suggests that the c ion is forming six sp^3d^2 orbitals which then bond to the dangling sp^3 orbitals of the a, b ions. The inward relaxation of the ionic positions also shortens the bonds and reduces the distortion of the octahedron. This configuration is favored since these bonds are shorter than the ones obtained by dimerization in the C_{2v} arrangement predicted by Watkins' model, and so are closer in length to the bulk silicon bond. However, only in the case of the doubly negative vacancy are there enough electrons to fully occupy the six orbitals.

---

[*]fabiano.corsetti08@imperial.ac.uk

[1]For a review, see: C. G. Van de Walle and J. Neugebauer, J. Appl. Phys. **95**, 3851 (2004).

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

[3]M. G. Ganchenkova, L. E. Oikkonen, V. A. Borodin, S. Nicolaysen, and R. M. Nieminen, Mater. Sci. Eng. B **159-160**, 107 (2009).

[4]M. I. J. Probert and M. C. Payne, Phys. Rev. B **67**, 075204 (2003).

[5]For a review, see: Probert and Payne[4] and Puska et al.[6]

[6]M. J. Puska, S. Pöykkö, M. Pesola, and R. M. Nieminen, Phys. Rev. B **58**, 1318 (1998).

[7]A. F. Wright, Phys. Rev. B **74**, 165116 (2006).

[8]N. Marzari and D. Vanderbilt, Phys. Rev. B **56**, 12847 (1997).

[9]G. D. Watkins, in *Deep Centres in Semiconductors*, edited by S. T. Pantelides (Gordon and Breach, New York, 1986), p. 147.

[10]G. A. Baraff, E. O. Kane, and M. Schlüter, Phys. Rev. B **21**, 5662 (1980).

[11]G. D. Watkins, Physica B + C **117B-118B**, 9 (1983).

[12]S. B. Zhang and J. E. Northrup, Phys. Rev. Lett. **67**, 2339 (1991).

[13]S. J. Clark, M. D. Segall, C. J. Pickard, P. J. Hasnip, M. I. J. Probert, K. Refson, and M. C. Payne, Z. Kristallogr. **220**, 567 (2005).

[14]D. M. Ceperley and B. J. Alder, Phys. Rev. Lett. **45**, 566 (1980).

[15]D. Vanderbilt, Phys. Rev. B **41**, 7892 (1990).

[16]D. R. Hamann, M. Schlüter, and C. Chiang, Phys. Rev. Lett. **43**, 1494 (1979).

[17]C.-K. Skylaris, P. D. Haynes, A. A. Mostofi, and M. C. Payne, J. Chem. Phys. **122**, 084119 (2005).

[18]N. D. M. Hine, P. D. Haynes, A. A. Mostofi, C.-K. Skylaris, and M. C. Payne, Comput. Phys. Commun. **180**, 1041 (2009).

[19]H. J. Monkhorst and J. D. Pack, Phys. Rev. B **13**, 5188 (1976).

[20]C. Persson, Y.-J. Zhao, S. Lany, and A. Zunger, Phys. Rev. B **72**, 035211 (2005).

[21]S. Lany and A. Zunger, Phys. Rev. B **78**, 235104 (2008).

[22]S. Pöykkö, M. J. Puska, and R. M. Nieminen, Phys. Rev. B **53**, 3813 (1996).

[23]Only the valence bands have been taken into account for the wannierization.

[24]A. A. Mostofi, J. R. Yates, Y.-S. Lee, I. Souza, D. Vanderbilt, and N. Marzari, Comput. Phys. Commun. **178**, 685 (2008).

[25]J. Neugebauer, private communication..

[26]The unrelaxed defect formation energy has not been calculated with ONETEP, since the density matrix method does not allow for a degenerate ground state without an explicit (nonlinear scaling) diagonalization.

[27]The results given use the Voronoi cell method for determining the potential alignment; this is to ensure consistency among all results, since for the 1000-atom ONETEP calculation only the Voronoi cell method is appropriate. For all comparisons of CASTEP calculations between the Voronoi cell and MLWFs methods, the agreement was within 0.01 eV (calculations not shown).

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

[29]G. Makov and M. C. Payne, Phys. Rev. B **51**, 4014 (1995).

[30]C. W. M. Castleton, A. Höglund, and S. Mirbt, Phys. Rev. B **73**, 035215 (2006).

[31]A. Alkauskas, P. Broqvist, and A. Pasquarello, Phys. Rev. Lett. **101**, 046405 (2008).

[32]C. Freysoldt, J. Neugebauer, and C. G. Van de Walle, Phys. Rev. Lett. **102**, 016402 (2009).

[33]N. D. M. Hine, K. Frensch, W. M. C. Foulkes, and M. W. Finnis, Phys. Rev. B **79**, 024112 (2009).

[34]J. Lento and R. M. Nieminen, J. Phys.: Condens. Matter **15**, 4387 (2003).

[35]I. Souza, N. Marzari, and D. Vanderbilt, Phys. Rev. B **65**, 035109 (2001).

[36]G. D. Watkins, in *Defects and Their Structure in Non-metallic Solids*, edited by B. Henderson and A. E. Hughes (Plenum, New York, 1976), p. 203.

[37]M. Sprenger, S. H. Muller, and C. A. J. Ammerlaan, Physica B + C **116B**, 224 (1983).

[38]M. Sprenger, S. H. Muller, E. G. Sieverts, and C. A. J. Ammerlaan, Phys. Rev. B **35**, 1566 (1987).

[39]H. A. Jahn and E. Teller, Proc. R. Soc. Lond. A **161**, 220 (1937).