Supercell size scaling of density functional theory formation energies of charged defects

N. D. M. Hine, K. Frensch, W. M. C. Foulkes, and M. W. Finnis
Department of Physics, Imperial College, Exhibition Road, London SW7 2AZ, United Kingdom
(Received 13 August 2008; revised manuscript received 6 January 2009; published 30 January 2009)

We address the calculation within density functional theory (DFT) of defect formation energies in alumina, a ceramic oxide often considered an archetype for a wide variety of other similar oxides. We examine the conditions under which calculated defect formation energies, especially those of charged defects, are independent of the principal approximations of the plane-wave DFT formalism, most significant of which is the finite-sized supercell in which the calculation must be performed. We introduce a variation on existing methods of extrapolation to infinite system size to reduce dependence of the result on finite-size errors in the electrostatic and elastic energies of a periodic supercell containing a defect. We also show how the results can be made relatively insensitive to the choice of exchange-correlation functional and pseudopotential by a suitable treatment of the chemical potentials of the atomic species. Our results for formation energies of charged defects are less sensitive than traditional approaches to supercell size and choices of exchange-correlation functional and pseudopotential, and differ notably from previous results.

DOI: 10.1103/PhysRevB.79.024112 PACS number(s): 61.72.Bb

I. INTRODUCTION

Density functional theory (DFT) using the plane-wave pseudopotential method[1] is well established as a highly versatile tool for *ab initio* calculation of properties of point defects in metals and semiconductors. Most such calculations use the supercell approximation[2] to model a localized defect embedded in a periodic host. The success of the method in predicting the structures of defects, their formation energies and thus concentrations, and their optical, diffusional, chemical, and other properties is well proven.[3–6]

While the structures and formation energies of point defects in metals and semiconductors are obtained accurately using standard DFT methodology, defects in ionic insulators present a greater challenge. Different choices of the approximations that go into the method, namely the size and shape of the supercell, the choice of exchange-correlation functional, and the choice of pseudopotential, can make a significant difference to the predicted physical properties. Supercell calculations of materials with defects having significant localized charge give rise to strong defect-defect interactions and strongly localized defect states cause errors due to self-interaction. For some elements, the total energy of the thermodynamic reference state is poorly described by DFT, leading to uncertainties in the calculated formation energies of defects involving the addition or removal of atoms. These difficulties mean that many previously reported defect formation energies, although obtained using technically correct DFT calculations, are not suitable for use in direct comparison with and interpretation of experimental results.

In this paper we attempt to calculate as accurately as possible the defect formation energies determining defect concentrations under real experimental conditions. We study the native point defects of aluminum oxide in the corundum structure—a highly important material both in industry and as a model to test theoretical methods. While formation energies can be tentatively inferred from experiments in some cases, a great deal of uncertainty surrounds the interpretation of diffusion in alumina[7–9] and a theoretical understanding of the formation energies would be of great help. In attempting to calculate accurate formation energies for the point defects of this system, we bring together a previously reported technique[10,11] to reduce the dependence of the chemical potentials of the species present on the approximations used, and a variation of an existing method to extrapolate the results of supercell-size-dependent formation energy calculations to infinite size. In this way, we find results only weakly dependent on the approximations used.

For the calculation of defect formation energies we use the formalism of Zhang and Northrup.[12] The defect formation energies required as input to the law of mass action[13] and calculations of defect concentrations are differences between the total Gibbs free energies of the system before and after the creation of a defect at a specific site, including contributions from the chemical potentials of the atomic species and electrons added or removed. The standard form of DFT yields only zero-temperature total energies and does not provide the entropic and pV contributions to the Gibbs free energy per particle. For gases, these contributions are large and strongly dependent on temperature and pressure, but can be calculated easily and accurately using simple ideal gas physics. For solids, the pV contribution to the Gibbs free energy per unit cell is negligible at reasonable pressures, and the electronic specific heat is so small that the temperature dependence of the electronic contribution can also be neglected. Vibrational contributions can be substantial, but largely cancel when comparing defective and perfect structures. Thus, in most cases, reasonably accurate formation energies can be obtained by replacing the Gibbs free energy per unit cell of the solid by the ground-state total energy. Suitable approximations to the Gibbs free energies of the perfect and defect crystals can thus be calculated from the DFT total energies of perfect and defective supercells, while chemical potentials for the constituent atomic species in the experimentally relevant reservoirs can be determined by combining the results of DFT calculations with some well-known thermodynamics.

A number of previous attempts have been made to calculate formation energies of defects in Al_2O_3. The results of

empirical pair-potential models[14,15] vary greatly according to the potential used.[16] While methods based on empirical potentials are able to access a range of cell sizes beyond that available to DFT calculations, they experience a great deal of difficulty describing the mixed ionic-covalent bonding, the complexity and low symmetry of the crystal structure, and the major changes in nearest-neighbor (NN) coordination on creation of defects. Full electronic structure calculations have not fared very much better. Even when the defect is neutral there is significant variation between different methodologies. For example, the DFT formation energy of a neutral oxygen vacancy has been variously quoted as 12.9 eV,[17] 10.1 eV,[18] 5.8 eV,[19] 7.1 eV,[20] and 13.3 eV.[21] Most of this variation can be attributed to the use of different approaches for calculating the oxygen chemical potential and different treatments of the problems caused by the DFT underestimation of the band gap. Similar variation has been seen in the geometries accompanying these energies, with the relaxation of the nearest-neighbor Al ions away from the defect ranging from 0.4 to 12%. The variation in results with DFT methodology is often larger for defect systems than in the corresponding bulk system,[22] where, for example, crystalline lattice constants vary by little more than 2–3% between different functionals.[23]

When the defect is charged, the situation is even worse. In the Zhang-Northrup formalism, the formation energy of a charged defect depends on the chemical potential of the electron reservoir from which the charge is imagined to have come. Many calculations avoid this complication by focusing on the formation energies of charge-neutral Schottky and Frenkel combinations of defects. (We use the Kroger-Vink notation[13] for defect types, although with charge states denoted by numerical superscripts rather than dots and dashes. The charge-neutral Schottky quintet is thus $2V_{\mathrm{Al}}^{3-}+3V_{\mathrm{O}}^{2+}$ and the Frenkel pairs are $V_{\mathrm{O}}^{q+}+\mathrm{O}_i^{q-}$ and $V_{\mathrm{Al}}^{q+}+\mathrm{Al}_i^{q-}$ for some charge q.) However, to answer questions relating to defect concentrations and diffusion properties in the presence of doping and variable formation conditions, it is much more useful to consider isolated defects in all possible charge states, resulting in individual formation energies that are functions of chemical potentials μ_e and μ_{O} for the electrons and oxygen atoms, respectively.

The approach used in this paper allows one to calculate defect formation energies as a function of the position ϵ_F of the Fermi level within the gap and the oxygen chemical potential $\mu_{\mathrm{O}}(T,p_{\mathrm{O}_2})$ of the atmosphere in which the alumina is formed. The former quantity depends strongly on the level of doping by aliovalent impurities, while the latter depends strongly on temperature and oxygen partial pressure. The equilibrium defect densities are sensitive to all of these quantities. This approach is useful because it allows us to link predictions of formation energies and defect densities to experimentally controllable parameters; a study of ground-state formation energies alone would not permit direct contact with experiment. In Sec. II we introduce our methodology and discuss the techniques used to remove the dependence on the specifics of the calculation of the O_2 dimer, to reduce the effects of the band-gap errors in DFT, and the difficult issue of supercell size dependence. Section III presents our converged results for the formation energies of the intrinsic defects of alumina and discusses their implications for the concentrations of intrinsic defects in the solid.

II. METHODS

A. Bulk alumina

The DFT results presented here were calculated with the CASTEP (Ref. 24) code, using a plane-wave basis set.[1] To allow comparison of different approximations, both Vanderbilt ultrasoft[25] pseudopotentials (USPs) and norm-conserving pseudopotentials[26] (NCPPs) were used. In each case, the Al and O pseudopotentials retain three and six valence electrons, respectively, with the rest being pseudized. For each type of pseudopotential, adequate plane-wave energy cutoffs were determined by examining the convergence of a suitable supercell of bulk and then verifying that this cutoff was also sufficient for an example defect system. With the USPs, the total energy of a supercell of bulk was converged to 0.015 eV/atom at $E_{\mathrm{cut}}=550$ eV (relative to calculations at 850 eV), so this cutoff was used throughout the USP calculations. It was further verified that energy differences of the form $E_T^{\mathrm{def}}-E_T^{\mathrm{perf}}$ between typical defect and perfect cells were converged to within 0.02 eV at this cutoff. For the considerably harder NCPPs, $E_{\mathrm{cut}}=1700$ eV was required for similar convergence but 3000 eV was in fact used as it aided the generation of orbitals for further calculations within the diffusion Monte Carlo method, to be reported separately.

Exchange-correlation potentials were treated using both the local-density approximation[27,28] (LDA) and the PBE parametrization[29] of the generalized gradient approximation (GGA). For each of these approximations, the accuracy of prediction of the bulk geometry and properties was tested. α alumina belongs to the trigonal crystal system, with a rhombohedral primitive cell containing ten atoms (2 f.u.), where all atoms of each species are in equivalent positions in the perfect crystal. Each O ion is surrounded by a distorted tetrahedron of four nearest-neighbor Al ions, two of which have shorter bonds (by about 5%) than the other two. Correspondingly, each Al ion is surrounded by a distorted octahedron of O ions, three of which are 5% closer than the other three. To maintain the 2:3 coordination, in any column of atoms down the c axis, one in three aluminum ions is missing and the remaining pairs move together slightly, requiring a long unit cell corresponding to four aluminum ions in this direction.

Starting from the experimentally reported geometry,[30] optimization was carried out in the rhombohedral primitive cell of 2 f.u. of Al_2O_3 by adjusting four parameters: a_{rhom}, the rhombohedral lattice parameter; α_{rhom}, the angle between lattice vectors; and the two positions u_{Al} and v_{O} that define the distortion of the lattice resulting from the unoccupied sites in the Al lattice. In an hcp structure, u_{Al} and v_{O} would both be 1/3—their deviations from this value are the result of movement of oxygen ions toward the aluminum ions and of slight pairing of aluminum ions. Table I shows the minor variation in geometry parameters with the pseudopotential (PSP) and exchange-correlation (XC) functional used. Also shown is the mean Al-O bond length $a_{\mathrm{Al\text{-}O}}$. All of these results represent a satisfactory match to published experimental values[30] in line with expectations of the accuracy of DFT. As usual,

TABLE I. Optimized geometry parameters for bulk alumina compared to experiment. The results for different choices of exchange-correlation functional (LDA and GGA) and different types of pseudopotential (ultrasoft and norm-conserving) are all in reasonably good agreement with the experimental parameters.

Method	a_{rhom} (Å)	α_{rhom} (deg)	u_{Al}	v_{O}	$a_{\mathrm{Al\text{-}O}}$ (A)
Experiment[a]	5.136	55.28	0.352	0.306	1.915
LDA-USP	5.052	55.30	0.352	0.307	1.885
GGA-USP	5.150	55.29	0.352	0.306	1.921
LDA-NCPP	5.147	55.39	0.352	0.305	1.921
GGA-NCPP	5.169	55.29	0.352	0.305	1.928

[a]Reference 30.

the LDA underestimates the lattice constant, while the GGA mildly overestimates it. Convergence of total energies and geometry parameters to the accuracy shown in the table was achieved with a $4\times4\times4$ mesh of Monkhorst-Pack k points, with no further variation relative to those at $8\times8\times8$.

B. Defect calculations

Supercells containing defects were generated by constructing an $l\times m\times n$ array of copies of the 30-atom nonprimitive hexagonal unit cell of the bulk structure and removing or adding an atom at the center as required. The hexagonal cell corresponding to the LDA-USP values above measures 12.8 Å along the c axis by 4.7 Å along the a and b axes, so in order of increasing minimum defect-defect distance the cells used consisted of $2\times2\times1$, $2\times2\times2$, $2\times2\times3$, $3\times3\times1$, $4\times4\times1$, and $3\times3\times2$ multiples of the hexagonal cell, containing 120, 240, 360, 270, 480, and 540 atoms, respectively. Figure 1 shows the structure of the smallest of these cells, the $2\times2\times1$ supercell.

Geometry optimization for the defect structures was first performed in the smallest of these supercells ($2\times2\times1$) for each defect under consideration. The optimization was carried out in two stages to allow the effects of electrostatic interactions and elastic relaxations to be investigated separately. First, atoms within 3.5 Å of the defect were allowed to relax—this set includes the first nearest-neighbor (1NN), 2NN, and 3NN in all cases. The resulting partially relaxed structures were retained for use in the investigation of defect-defect interactions. For the final results, all the atoms within the supercell were allowed to relax. In most cases, the energy gained in the second, longer-ranged, relaxation was under 0.25 eV, but for highly charged defects such as V_{Al}^{3-} it could be over 1 eV. Energies obtained using only the shorter-ranged 3NN relaxations were therefore deemed insufficient for the determination of defect formation energies. The 3NN-relaxed positions from the $2\times2\times1$ cell, padded with perfect crystal, were, however, used as a starting point for relaxing the larger cells, to speed up the geometry optimization while avoiding a discontinuity in strain at the boundary of the embedding region. The unrelaxed energies of the larger cells were first recorded and then these structures were allowed to fully relax. The results of these investigations of finite-size errors are detailed in Sec. III. Convergence of total energies with k-point sampling was achieved in the hexagonal unit cell with a $4\times4\times2$ k-point grid centered on Γ. Supercells of size 2×2, 3×3 and 4×4 in the ab plane were sampled with 2×2, 2×2 and 1×1 k points in the reciprocal-space ab plane, respectively. Supercells containing 1 and 2 repetitions of the hexagonal cell along the c axis were sampled using 2 and 1 planes of k points along the reciprocal-space c axis. As alumina is such a good insulator, the defect states are only very weakly dispersive and the dependence of the formation energies on k-point sampling is very small.

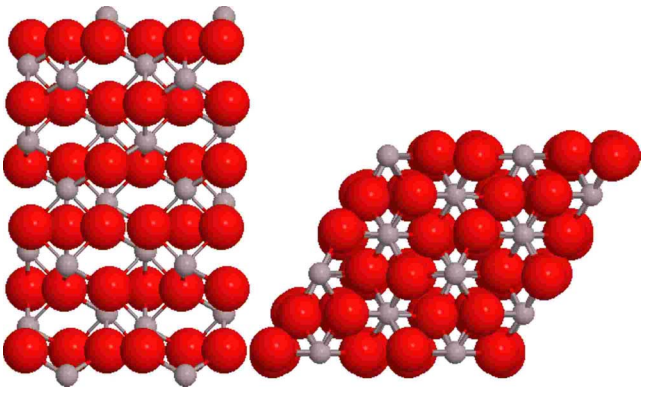

FIG. 1. (Color online) Supercell of Al_2O_3 containing $2\times2\times1$ copies of the hexagonal unit cell with 120 atoms. Left: side view along the a axis. Right: top view down the c axis.

In the Zhang-Northrup formalism, the formation energy $\Delta E_f(X^q)$ of defect species X in charge state q is

$$\Delta E_f = E_T^{\mathrm{def},q} - E_T^{\mathrm{perf}} - \sum_i \Delta n_i \mu_i + q\mu_e, \tag{1}$$

where E_T^{perf} and $E_T^{\mathrm{def},q}$ are the total energies of the perfect and defect supercells, μ_i is the chemical potential of atomic species i, Δn_i is the number of atoms of element i removed (negative Δn_i) or added (positive Δn_i) to create the defect, and μ_e is the chemical potential of an electron in the perfect crystal. Note that $q=-\Delta n_e$, where Δn_e is the number of electrons removed (negative Δn_e) or added (positive Δn_e) to create the defect. The electron chemical potential μ_e can be defined by choosing a Fermi level ε_F within the gap relative to the valence-band maximum of the perfect crystal, such that $\mu_e=E_{\mathrm{VBM}}^{\mathrm{perf}}+\varepsilon_F$. Here ε_F lies in the range 0 to E_g, where E_g is the energy gap of the perfect crystal, and $E_{\mathrm{VBM}}^{\mathrm{perf}}$ is the

position of the valence-band maximum in the perfect crystal. The use of $E_{\mathrm{VBM}}^{\mathrm{perf}}$ as a reference amounts to a choice of the zero of potential and has no effect on the accuracy of calculated defect formation energies. We shall see in Sec. III, however, that the difficulty of relating the zeros of potential in the perfect and defective supercells is an important source of finite-size error.

As explained in Sec. I, the use of total energies E_T^{perf} and E_T^{def} in place of the corresponding free energies assumes that the vibrational contributions cancel between the perfect and defect systems, and that the pV and temperature-dependent electronic terms are negligible for solids. Effects due to thermal expansivity may also play a role: the lattice constant at 1800 K is 1–2% larger than the zero-temperature lattice constant, and this produces a slight change in the band gap. Again, however, these effects can be presumed to cancel to the accuracy available in current formation energy calculations.

C. Chemical potentials

The chemical potentials of the atomic species in a nonelemental crystal are inherently troublesome quantities to define, since it is not possible without further information to assign specific fractions of the total Gibbs free energy of the crystal to the various species present. In reality, the atomic chemical potentials are set by the experimental conditions, which define the properties of the reservoirs of atoms in equilibrium with the sample. If, for example, the sample of alumina has been annealed in an oxygen atmosphere, the chemical potential of the oxygen atoms is the same as in gaseous molecular oxygen at the experimental partial pressure and temperature. The chemical potential of the oxygen atoms in well-annealed alumina is thus strongly dependent on temperature and pressure. The aluminum chemical potential μ_{Al} is tied to the corresponding μ_{O} by $2\mu_{\mathrm{Al}}+3\mu_{\mathrm{O}}=\mu_{\mathrm{Al_2O_3}}$, where $\mu_{\mathrm{Al_2O_3}}$ is the total Gibbs free energy per formula unit. If the relatively small temperature and pressure dependence of the chemical potential per unit cell of the solid is neglected, one can therefore assume that $\Delta\mu_{\mathrm{Al}}=-\frac{3}{2}\Delta\mu_{\mathrm{O}}$, where $\Delta\mu_{\mathrm{O}}$ is the change in oxygen chemical potential relative to some defined reference value. The strong dependence of μ_{O} on annealing conditions therefore implies an equally strong variation of μ_{Al}.

Although the values of the atomic chemical potentials depend on experimental conditions, one can use the formation energies of the species in their elemental states and the formation energy of the crystal to determine useful bounds. The thermodynamic stability of alumina implies that the chemical potentials of the atoms cannot be higher in the crystal than in their elemental states, gaseous O_2 and fcc aluminum metal, under the same conditions. Hence, in terms of $\mu_{\mathrm{Al}(s)}$, the total free energy per atom of fcc solid aluminum, and $\mu_{(1/2)\mathrm{O}_2(g)}$, the total free energy per atom of molecular oxygen in gaseous form, we have the inequalities

$$\mu_{\mathrm{O}} \leq \mu_{(1/2)\mathrm{O}_2(g)}(p_{\mathrm{O}_2},T), \quad \mu_{\mathrm{Al}} \leq \mu_{\mathrm{Al}(s)}. \tag{2}$$

Note that we have dropped the weak temperature and pressure dependence of the chemical potential of solid aluminum but not the much stronger dependence of the chemical potential of gaseous oxygen. The Gibbs free energy balance for the formation energy $\Delta G_f^{\mathrm{Al_2O_3}}(p_{\mathrm{O}_2},T)$ per formula unit of alumina at temperature T and oxygen partial pressure p_{O_2} gives

$$\Delta G_f^{\mathrm{Al_2O_3}}(p_{\mathrm{O}_2},T) = \mu_{\mathrm{Al_2O_3}} - 2\mu_{\mathrm{Al}(s)} - 3\mu_{(1/2)\mathrm{O}_2(g)}(p_{\mathrm{O}_2},T), \tag{3}$$

where the temperature and pressure dependence of $\mu_{\mathrm{Al_2O_3}}$ has also been dropped. Additionally, we know that under equilibrium conditions in the solid, $\mu_{\mathrm{Al_2O_3}}=2\mu_{\mathrm{Al}}+3\mu_{\mathrm{O}}$. We can now combine these four expressions to obtain

$$\tfrac{1}{2}\Delta G_f^{\mathrm{Al_2O_3}}(p_{\mathrm{O}_2},T) \leq \mu_{\mathrm{Al}} - \mu_{\mathrm{Al}(s)} \leq 0, \tag{4}$$

for the Al ions and

$$\tfrac{1}{3}\Delta G_f^{\mathrm{Al_2O_3}}(p_{\mathrm{O}_2},T) \leq \mu_{\mathrm{O}} - \mu_{(1/2)\mathrm{O}_2(g)}(p_{\mathrm{O}_2},T) \leq 0, \tag{5}$$

for the oxygen ions, giving us a range of feasible values for μ_{O} and μ_{Al}.

While this approach is often satisfactory if the formation energy is small, as in III–V semiconductors, the large value of $\Delta G_f^{\mathrm{Al_2O_3}}$ (~ -17 eV at standard temperature and pressure) makes it unsatisfactory for alumina. The allowed ranges of μ_{O} and μ_{Al} are wide and one cannot determine where specific experimental conditions lie without further calculation. Here, instead, we make direct contact with experiment by assuming that the alumina has been annealed in an oxygen atmosphere at some chosen temperature T and oxygen partial pressure p_{O_2}, implying that μ_{O} must be equal to the chemical potential $\mu_{(1/2)\mathrm{O}_2(g)}(p_{\mathrm{O}_2},T)$ per oxygen atom in gaseous molecular oxygen.

A second difficulty is that the variation in the values of $\mu_{(1/2)\mathrm{O}_2(g)}$ calculated with different functionals and pseudopotentials is of order an eV, due to the comparative failure of DFT to describe the oxygen dimer accurately. To circumvent this problem, we follow Finnis *et al.*[11] and obtain the chemical potential $\mu^0_{(1/2)\mathrm{O}_2(g)}$ of oxygen gas at standard temperature and pressure from Eq. (3) at standard temperature and pressure, with $\Delta G_f^{\mathrm{Al_2O_3}}(p^0_{\mathrm{O}_2},T^0)$ taken from experiment and $\mu_{\mathrm{Al_2O_3}}$ and $\mu_{\mathrm{Al}(s)}$ approximated by the corresponding zero-temperature total energies. The use of the experimental formation energy of the oxide renders the calculation of the DFT total energy of the oxygen dimer unnecessary and greatly reduces the sensitivity of the results to details of the DFT methodology. The chemical potential $\mu_{(1/2)\mathrm{O}_2(g)}(p_{\mathrm{O}_2},T)$ of oxygen gas at the chosen formation conditions is obtained from $\mu^0_{(1/2)\mathrm{O}_2(g)}$ using simple ideal gas relations. Comparison to thermodynamic data shows[31] that this is a good approximation. The $T=0$ total energy of fcc aluminum metal is calculated from a DFT calculation with the same cutoffs as that of the bulk alumina, with a $14\times14\times14$ Monkhorst-Pack k-point grid in a four-atom primitive cell, and a separately optimized lattice parameter for each combination of pseudopotential and functional. The Gibbs free energy of formation of alumina at standard temperature and pressure, $\Delta G^f_{\mathrm{Al_2O_3}}(p^0_{\mathrm{O}_2},T^0)=17.37$ eV, was obtained from tables,[32] as

were the values $S^0_{O_2}=0.0021$ eV/K and $C^0_P=7k_B/2=0.000302$ eV/K of the molecular entropy and constant pressure heat capacity of oxygen gas required to calculate $\mu_{(1/2)O_2(g)}(p_{O_2},T)$ from $\mu^0_{(1/2)O_2(g)}$. The full expression for $\mu_{(1/2)O_2(g)}(p_{O_2},T)$ becomes

$$\mu_{(1/2)O_2(g)}(p_{O_2},T)=\frac{1}{3}(\mu_{Al_2O_3}-2\mu_{Al(s)}-\Delta G_f^{Al_2O_3})+\Delta\mu(T)+\frac{1}{2}k_BT\log\left(\frac{p_{O_2}}{p^0_{O_2}}\right) \quad (6)$$

with the rigid-dumbbell ideal gas form for $\Delta\mu(T)$:

$$\Delta\mu(T)=-\frac{1}{2}(S^0_{O_2}-C^0_P)(T-T^0)+\frac{1}{2}C^0_PT\log\left(\frac{T}{T^0}\right). \quad (7)$$

Experimental samples of alumina are generally annealed under one set of conditions and probed under another. The conditions relevant for calculations of defect concentrations are normally the annealing conditions, which determine the defect concentrations "frozen in" after the anneal.

D. Band-gap errors

In the region surrounding a defect there may exist occupied orbitals referred to as "defect levels," that are localized around the defect site. These states, the energies of which lie in the band gap, can either be derived from former conduction-band states which have been lowered in energy by the presence of the defect, or from valence-band states which have been raised in energy. In the latter case, as the defect level is composed of states that would be occupied in the perfect crystal, there is no problem. In the former case, however, with a localized defect orbital composed of unoccupied conduction-band states, the eigenvalue of the defect state will suffer from the well-known DFT band-gap problem,[33] whereby energy gaps to conduction-band states are underestimated.

If such defect states are occupied, the eigenvalue and degree of localization of the eigenstate may be underestimated. The resulting underestimation of the formation energy is an intrinsic problem of DFT calculations on defects. A common approach to try and make the DFT defect formation energies match experimental results is to apply a rigid shift to the energy eigenvalues of the occupied defect levels. In the simplest picture, if there are m electrons occupying defect states derived from the conduction band and the energy gap is underestimated by ΔE_g, one adds $m\times\Delta E_g$ to the energy of the defect cell.[34,35] In semiconductors this adjustment appears to be vital to make the calculated defect formation energies agree with experiment. However, it represents a severely uncontrolled approximation in the case of strongly ionic insulators, where the defect states are often so localized that it is unclear whether they are derived from the conduction or valence band. In these cases, it is not obvious that the band-gap adjustment should be applied at all. Even worse, because such solids generally have large band gaps, ΔE_g itself is large.

The only approach likely to be able to fix this problem in a reliable manner would be to use a functional capable of correctly localizing the electron states and giving a correct gap. Such might be provided by a hybrid functional, although the fraction of exact exchange to be used could be regarded as an additional empirical parameter.[36] The computational demands of hybrid exchange calculations in a plane-wave basis would be very large for the supercells considered in this work, preventing us from pursuing this approach here.

Instead, we follow a much earlier suggestion of Baraff and Schlüter,[37] and calculate an effective value of m directly from the nature of the defects states themselves. A purely empirical approach to the construction of a Hamiltonian that widens the DFT band gap by ΔE_g to obtain the experimental gap is to add a so-called *scissor operator* of the form $\Delta E_g\hat{P}^{cond}$ to the DFT Hamiltonian, where $\hat{P}^{cond}$ is a projection operator onto the conduction bands of the perfect crystal. A first-order estimate of the contribution made by this additional term to the electronic energy of a system with defect levels in the gap may be obtained by summing the expectation values of the scissor operator in the occupied defect states. The effective modification to the total energy then becomes

$$m_{eff}\Delta E_g=\sum_{i\ def}\sum_{j\ cond}|\langle\psi_i^{def}|\psi_j^{cond}\rangle|^2\Delta E_g \quad (8)$$

$$=\sum_{i\ def}\left(1-\sum_{j\ val}|\langle\psi_i^{def}|\psi_j^{val}\rangle|^2\right)\Delta E_g, \quad (9)$$

where the superscripts "val," "cond," and "def" refer to the valence and conduction states of the perfect crystal and the defect states of the defective supercell, respectively. This technique was applied to the defect states calculated with norm-conserving pseudopotentials to give one value of m_{eff} which was then used for that defect across all methods.

This approach is not well justified and may not be very accurate, so it is fortunate that the band-gap error does not affect the formation energies of the most important defects in alumina. Because alumina is a very wide band-gap insulator, the most stable defects are those, such as V_O^{2+} and Al_i^{3+}, with full ionic charges. These defects do not have occupied states in the gap and are unaffected by the band-gap error. The scissor-operator technique only affects defects such as oxygen vacancies and aluminum interstitials in charge states below their ionic values, $V_O^{1+,0}$ and $Al_i^{2+,1+,0}$, which have higher formation energies than the full-charge versions over most of the available range of Fermi energy. Moreover, whenever a band-gap error reduction technique is applied to such defects, it serves only to raise their formation energies, making them even less likely to form. The band-gap error has no effect on any conclusions drawn about the concentrations of the dominant defects in alumina for any reasonable formation conditions.

E. Defect-defect interactions

The use of a plane-wave basis set in DFT has many advantages in computational cost and accuracy. However, the price is that one is restricted to periodic boundary conditions. Any defect formation energy, especially one for a defect with a nonzero charge, therefore contains spurious interactions

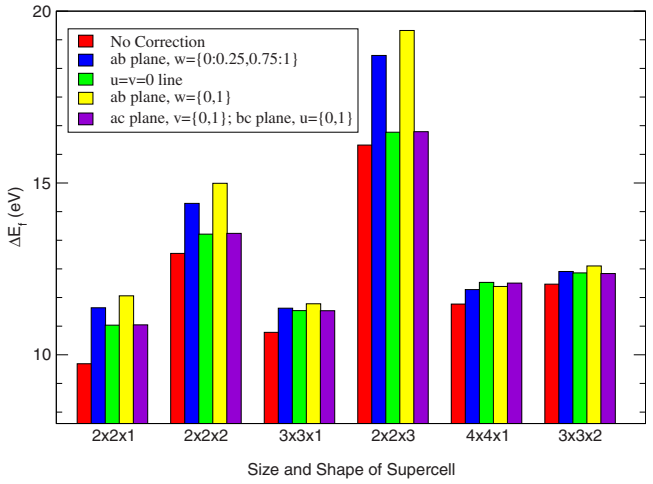

FIG. 2. (Color online) Correction of the formation energy of the aluminum vacancy V_{Al}^{3-} by "average potential" shifts. Groups of bars represent different supercell sizes indicated on the axis. Within a group, different colored bars represent different choices of region over which to average the potential. The variation with cell size and choice of region over which to average the potential remains unacceptably large even at large supercell sizes. As the supercell size increases, the variation in the shift begins to converge, but still varies by 0.5 eV depending on the choice of region at $3\times3\times2$ hexagonal cells. As the size and shape of supercell are varied, even at the larger end of the size range, calculated formation energies still vary by 1.0 eV.

from the potentials of nearby periodic copies of the defect, the magnitudes of which depend strongly on the size of the supercell.

A number of methods exist to reduce the effects of unwanted periodic boundary conditions in defect calculations. The simplest is to attempt to take account of the change in average potential caused by the periodic copies of the defect. In this approach, the "average" potentials $V_{\mathrm{av}}^{\mathrm{perf}}$ and $V_{\mathrm{av}}^{\mathrm{def}}$ are calculated by averaging over a number of points in some range as far from the defect as is possible within the supercell, and the difference is found between this part of the defect cell and the same region of perfect crystal. The resulting potential shift is then applied to the valence-band maximum, E_{VBM}, used as the zero of potential in the defect cell. The otherwise unknown monopole-background potential contribution to the defect cell total energy E_T^{def} is thus canceled from the $q\mu_e$ term in the formation energy, as μ_e contains E_{VBM}. In principle this should make the result independent of cell size. However, this approach fails in practice as the definition of "far" from the defect in a periodic supercell containing strongly localized defect charge is highly inexact. For example, in the case of the aluminum vacancy of charge −3, small movements of the line or plane along which this averaging is done can change the average potential shift applied to formation energies by 0.9 eV at smaller cell sizes. In effect, while this approach is valid for the delocalized defect states in a semiconductor, no region in the supercell of an ionic crystal is far enough from the defect for this procedure to be accurate. Figure 2 shows the wide variation in results this method can generate as the size of the supercell is varied.

A more promising route would appear to be to cancel out the potential from the periodic images of the defect. A recent paper by Dabo *et al.*[38] summarizes attempts to date to reduce finite-size errors in potentials for nonperiodic systems and classifies them according to the method used to cancel the potential from periodic images. Employing pointlike countercharges results in a theory akin to that suggested by Leslie and Gillan[39] and later expanded upon by Makov and Payne,[40] whereas a Gaussian countercharge corresponds to the local moment counter charge method of Schultz.[41,42] Dabo *et al.* contrast these approaches with the ideal solution, which would involve a density countercharge based on the full nonperiodic part of the density with all its moments and suggest a means by which this ideal solution could be implemented efficiently. However, applying such methods *post hoc* to the final total energy of a periodic supercell calculation neglects any contribution from the polarization response of the medium and any contribution to the kinetic energy resulting from the change in the wave functions caused by the change in the potential. Modifying the potential during the self-consistency cycle of the simulation would be better, but the polarization response outside the cell would still be neglected and steps would have to be taken to ensure that the potential remained continuous at the supercell boundaries.

The most commonly used finite-size error reduction method is the one due to Makov and Payne,[40] which relates the energy of a defect in a cubic supercell of size L to that of the isolated defect $(L\rightarrow\infty)$ limit via

$$E_T(L) = E_T(L\rightarrow\infty) - \frac{q^2\alpha}{2\epsilon L} - \frac{2\pi qQ}{3\epsilon L^3} + \mathcal{O}[L^{-5}], \qquad (10)$$

where q is the monopole aperiodic charge, Q is the quadrupole moment of the aperiodic charge, ϵ is the relative dielectric constant of the medium in which the charges are embedded, and α is the Madelung constant, which is a property of the shape of the supercell only and is easily calculated using the Ewald method.[43] For a simple cubic system, $\alpha=2.8373$; for our hexagonal system, with different lengths in the a and c directions, α is a function of both lengths. Since dipole-dipole terms do not necessarily cancel in noncubic systems, the higher-order Q contributions may also be important.

Castleton *et al.*[44] gave an exhaustive comparison of use of the first, and first and second terms of the Makov-Payne formula and the adjustment of the average potential. None of these methods proved very reliable, but the indications were that the form, although not the magnitude, of the terms of the Makov-Payne formula appeared accurate. Castleton *et al.* also found that if they used fitted parameters, rather than the calculated or experimental values of α, ϵ, and Q, they obtained much better results. Shim, Lee, and Nieminen[45] and Lento, Mozos, and Nieminen[46] reached similar conclusions.

In this work, we use a variant on the above fitting procedure which has several useful features. Examining Eq. (10), we see that the first-order term, which dominates for single highly localized defects where the monopole-quadrupole and dipole-dipole terms are small, contains the Madelung potential $v_M=\alpha/L$. The Madelung constant is defined such that $-q^2\alpha/2L$ is the potential energy per unit cell of an infinite periodic lattice of identical unit cells, each of which contains

a point charge q plus a uniform neutralizing background. Defect charges are very well localized in alumina, and our DFT supercell includes a canceling uniform background if the defect is charged, so it is physically reasonable that the finite-size error should scale like $-q^2\alpha/2\epsilon L$, where the division by ϵ takes account of the polarizability of the solid.

The Madelung energy $-q^2\alpha/2L$ is normally negative (and hence α is positive) because it is dominated by interactions between the point charge and the canceling background charge around it, which is on average closer than the other point charges. In some lattices with long thin unit cells, however, in which some pairs of point charges are unusually close together, the reverse applies. The Madelung energy is dominated by the contributions from nearby pairs of point charges and α is negative.

Previous methods have chosen a fixed cell shape, and hence a fixed α, and used the behavior of the total energy as a function of L for that fixed lattice geometry to estimate ϵ. However, this is unnecessarily limiting: we can change the shape of the supercell alongside the size by varying the repeat length along one axis at a time, and thus change α as well as L_i. If we then plot the total energy against v_M (a function of L_i and α) for a wide range of cell shapes and sizes, we can still extrapolate to an infinite cell by taking the limit as $v_M \to 0$. In practice this is accomplished by fitting a straight line $E(v_M)=E_\infty+bv_M$ to the calculated energies. Comparing with Eq. (10) gives $b=-q^2/2\epsilon$, so we can also find an effective dielectric constant.

This variant of Castleton's finite-size scaling approach has two advantages. The first is that, because it does not rely on uniform scaling with L, one can access a much larger range of cells within feasible system sizes. The smallest supercell that can reasonably be used to study defects in alumina contains 120 atoms. Doubling this cell in each direction would result in a cell of 960 atoms—already beyond the limit where traditional plane-wave DFT begins to hit the cubic scaling "wall." Further uniform doublings would be impossible. Therefore, with the original method, one would be fitting a line to only two points. This method makes it possible to compare six to ten different cells within the feasible range and plot a meaningful extrapolation through all of them. The second advantage of this variant of the method is that it is possible to include supercells where the Madelung constant becomes negative, so that we have points on either side of $v_M=0$. Rather than making an unreliable extrapolation to a point outside the range of the data, we can therefore interpolate (in the strict mathematical sense; we are still extrapolating the system size, of course), which is in general more reliable.

This approach to system-size extrapolation assumes that the finite-size errors depend only on v_M and neglects quadrupole and higher terms, which scale differently with system size. A further advantage of studying supercells of different shapes, as well as different sizes, is that it provides enough data points to enable the deviations from the straight line $E(v_M)=E_\infty+bv_M$ to be assessed; these deviations provide a useful measure of the importance of the neglected terms. In alumina, and presumably also in other wide band-gap insulators with highly localized defect charges, we shall see that the higher-order terms are relatively small.

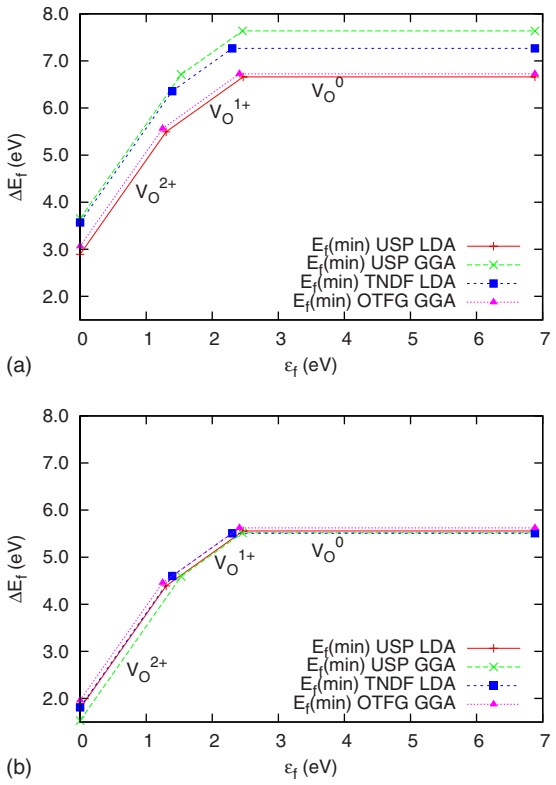

FIG. 3. (Color online) Comparison of formation energies of the oxygen vacancy (a) without and (b) with the improved chemical potential scheme. With method (a), there is a spread of values of ΔE_f of over 1 eV between different XC functionals and PSPs. With method (b), the results all fall within about 0.2 eV of each other.

III. RESULTS

A. Convergence

We first consider the calculation of the chemical potentials μ_O and μ_{Al}, comparing the standard approach, in which they are fixed at one end or the other of their range, to the method suggested by Finnis, Lozovoi, and Alavi[11] (FLA). In the former method, which we will call method (a), we obtained $\mu_{(1/2)O_2(g)}$ from the value of $E_T[O_2(g)]$ calculated with the same PSP and XC choices as for the bulk. We then set the formation conditions to represent the oxidation limit [the upper bound of the chemical potential in Eq. (5)] and simply took $\mu_O=\mu_{(1/2)O_2(g)}$. To illustrate the effect of this procedure, Fig. 3(a) shows the calculated formation energy of the oxygen vacancy V_O^q in various charge states, with no attempt made to reduce the errors due to the DFT underestimation of the band gap. The free energy of formation ΔE_f of the most stable charge state of the defect is plotted as a function of Fermi level in the range $\epsilon_F=0$ to $\epsilon_F=E_g$. The sudden slope changes occur because the most stable charge state depends on the Fermi level varying from $q=+2$ when the Fermi level is low in the gap to $q=0$ near the top of the gap. Method (a) is seen to give a result with a strong dependence on the choices of pseudopotential and exchange-correlation functional, as the total energy of O_2 represents a challenging problem for DFT for well-documented reasons. Therefore, when we compare formation energies, we find they vary by

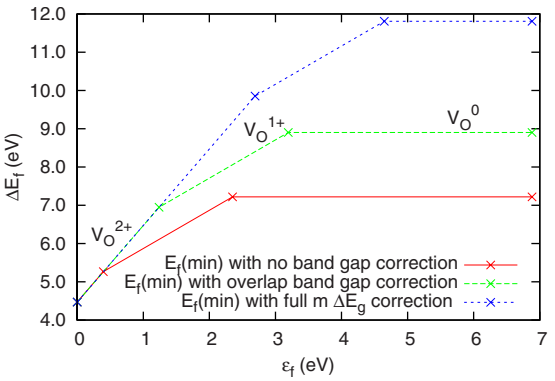

FIG. 4. (Color online) The effect of the band-gap error on the formation energy of an oxygen vacancy. The lowermost line shows the unadjusted formation energy as a function of ϵ_F; the topmost line applies the full $m \times \Delta E_g$ adjustment and the middle line applies the adjustment of Eq. (9).

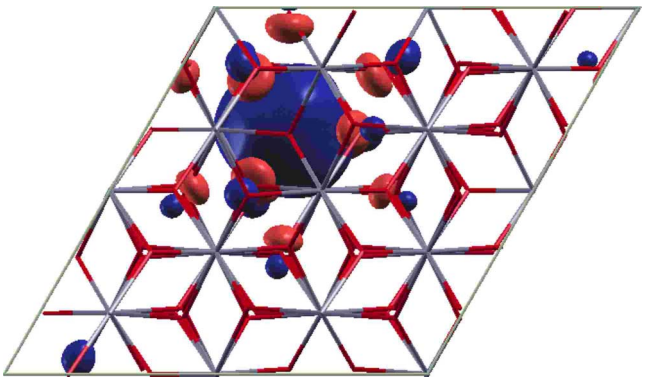

FIG. 5. (Color online) 3D isosurface projected into the *ab* plane of the defect state associated with the V_O^0 defect. Lobes associated with weight on 1NN oxygen *p* states can be seen, but most of the weight is in a diffuse *s* state on the vacancy site. The isosurface plotted corresponds to $|\psi(\mathbf{r})|^2=0.1$.

as much as 1.5 eV between different PSP/XC combinations. This situation is reproduced across other species of defect resulting in a highly unsatisfactory dependence of the formation energy on the choice of approximation scheme.

We contrast this with Fig. 3(b). Here we show analogous plots, but this time obtaining μ_O using the FLA approach, which we call method (b), bypassing the calculation of $E_T(O_2)$ by using the experimental oxide formation energy. The results obtained using different functionals and pseudopotentials now disagree by no more than about 0.2 eV. This suggests that the majority of the disparity in formation energies seen in Fig. 3(a) originates from the calculation of the ground-state energy of the O_2 molecule rather than from the solid-state calculations. This is in line with our expectations of where DFT will produce accurate answers and where it will fail. For the rest of this work we use the FLA method for chemical potentials.

Figure 4 illustrates the effect of the band-gap error in calculations of the oxygen vacancy. The neutral vacancy V_O^0 has two localized electrons on the vacancy site, which previously contained an O^{2-} ion with a filled $2p$ shell of six electrons; if the defect has charge +1, there is one localized electron; if it has charge +2, there are none. In this case, therefore, $m=2$ for V_O^0, $m=1$ for V_O^{1+}, and $m=0$ for V_O^{2+}. Note that even though there is no longer an oxygen nucleus at the defect site, it is not necessarily the case that the localized defect states must derive from the unoccupied aluminum 3*s* orbitals at the bottom of the conduction band—they may still have the symmetry of the now-missing oxygen *p* orbitals. The lowermost curve in the figure shows the formation energy of the defect as a function of ϵ_F with no attempt made to reduce the band-gap error; the topmost curve shows the formation energy with a full $m \times \Delta E_g$ adjustment applied (E_g in LDA is 6.88 eV, compared to the real gap of 9.12 eV, so ΔE_g is 2.24 eV). In the case of the neutral vacancy, the difference between the top and bottom lines is a full 4.48 eV and our uncertainty about the true value of this error brings into question the apparently high accuracy obtainable elsewhere in the calculation.

The value of m_{eff} suggested by Eq. (9) treads a middle ground between these two extremes and is better physically motivated than either. The results obtained using this projection approach are expected to be better than those obtained using the other two methods, but its success can only really be tested by going to a higher level of theory such as quantum Monte Carlo. In the case of the oxygen vacancy in LDA, the value of m_{eff} calculated via the defect-valence band overlap consistently fell around $m_{\text{eff}}=0.37$, independent of whether V_O^{1+} or V_O^0 was considered. This value corresponds closely to the position of the defect state eigenvalue within the gap ($\epsilon_{\text{def}}-E_{\text{VBM}} \simeq 2.52$ eV, $E_g=6.88$ eV in LDA), suggesting that $(\epsilon_{\text{def}}-E_{\text{VBM}})/E_g$ may provide a suitable first approximation for m_{eff}.

The defect state associated with the neutral oxygen vacancy V_O^0 has previously been identified as a "deep" defect level.[5,19] Figure 5 shows a projection in the *ab* plane of a three–dimensional (3D) isosurface of the square modulus of the Bloch state at Γ associated with this level. It can be seen that most of the weight is contained within the vacancy region in an *s*-like state, with some delocalization onto the 1NN O sites with *p*-like symmetry. None of the visible weight is distributed on the Al sites lending further support to the idea that the defect state is only minimally "conduction-band-like."

Figure 6 illustrates the procedure detailed in Sec. II E for extrapolation of formation energies to infinite separation. Results for the highest charge states of the principal intrinsic defects are shown, since these are the dominant defects under most circumstances and suffer from the largest finite-size errors. The defect formation energies were calculated using a range of different supercells but the same value of μ_e in every case. Figure 6 includes only the results of calculations at fixed geometry—that of the $2\times2\times1$ cell surrounded by a region of perfect crystal—excluding the effects of elastic relaxation far from the defect.

The unadjusted results all show a clear linear trend in $\Delta E_f(v_M)$, with finite-size errors of several electron volts for reasonably sized systems. Using a least-squares fit, we extract a "permittivity" ϵ_{fit} and a corresponding Makov-Payne finite-size extrapolation term $\Delta E_{\text{FS}}=q^2 v_M/2\epsilon_{\text{fit}}$. This also gives us a "statistical" uncertainty on the permittivity and the corresponding finite-size extrapolation term. Once the ΔE_{FS}

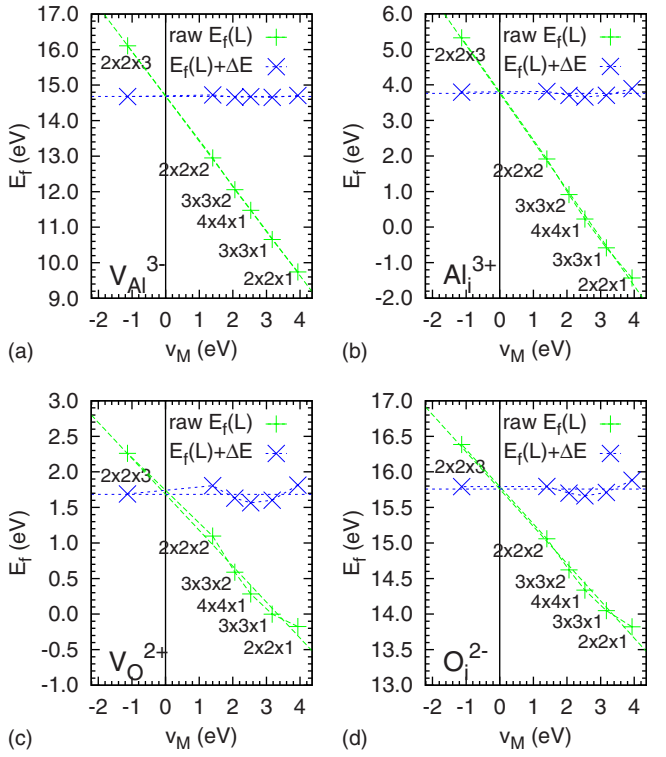

FIG. 6. (Color online) Finite-size errors in the formation energies of the principal intrinsic defects of alumina in their highest charge states. The results for larger supercells were obtained by embedding the relaxed structure of the smallest $2\times2\times1$ defect supercell in a region of perfect crystal, ignoring the effects of longer-ranged relaxations. (a) Aluminum vacancy, charge −3. (b) Aluminum interstitial, charge +3. (c) Oxygen vacancy, charge +2. (d) Oxygen interstitial, charge −2. This finite-size error reduction method works best for V_{Al}^{3-} but leaves a residual uncertainty of no more than around 0.2 eV in all cases.

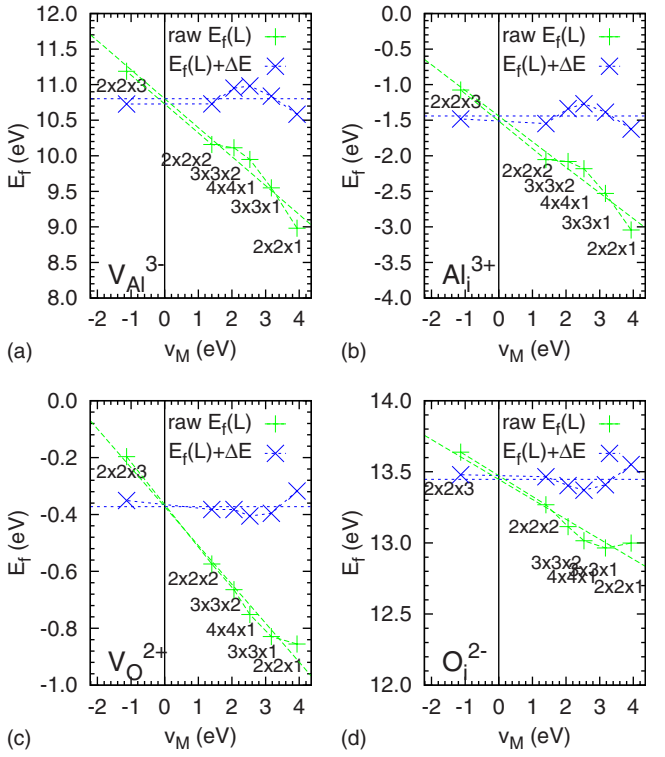

FIG. 7. (Color online) Finite-size errors in the formation energies of the principal intrinsic defects of alumina in their highest charge states including the effects of long-ranged elastic relaxation. Note the compressed scales relative to the graphs of Fig. 6. (a) Aluminum vacancy, charge −3. (b) Aluminum interstitial, charge +3. (c) Oxygen vacancy, charge +2. (d) Oxygen interstitial, charge −2. While there is more variation than for the pure electrostatic case shown in Fig. 6, it is still possible to interpolate reliably to $v_M=0$ to within around 0.2 eV.

adjustment has been added to the total energy, the results become nearly independent of size: the sloping line in each panel shows the unadjusted results, while the nearly flat line shows the adjusted results. The residual variation with supercell size and shape is less than 0.2 eV for all cell sizes considered. This residual variation is likely to be due to dipole-dipole and monopole-quadrupole terms, which cannot be expressed in terms of the Madelung potential but decay more rapidly as L increases and are negligible when L is large.

Figure 7 shows analogous results obtained allowing the whole defect supercell to relax until the residual forces are below 0.1 eV/Å. The ionic relaxation provides additional dielectric screening of the interactions between the periodic replicas and there is a marked reduction in the slopes of the lines. Since the elastic relaxations are driven primarily by the long-ranged electric field from the defect charge, the elastic contributions to the finite-size errors are expected to scale more or less as the electrostatic contributions. It is clear, however, that this is only approximately true, and that the uncertainties in the final defect formation energies are greater when elastic relaxations are included than when they are ignored. This suggests that a full anisotropic treatment of the elastic response may be required, although that would be beyond the scope of this work. However, in no case does the final energy differ from its infinite-system limit by more than 0.5 eV, which is still small enough to allow a reliable extrapolation/interpolation to infinite size.

Table II shows the effective isotropic permittivities derived from the gradients of the lines in Figs. 6 and 7, all of which were obtained using the LDA and USPs. The finite-size errors in formation energies calculated using specific defect supercells can be reduced by adding

TABLE II. Effective dielectric constants obtained by fitting the electrostatic and elastic finite-size errors of defect supercell total energies and resulting additions to the total energy required to reduce the Coulomb finite-size error of the $2\times2\times1$ cell in its LDA geometry with ultrasoft pseudopotentials. Errors due to the statistical nature of a least-squares fit are 0.1–0.2 eV.

Defect	V_O^{2+}	O_i^{2-}	V_{Al}^{3-}	Al_i^{3+}
ϵ_{fit} (unrelaxed)	3.96	3.82	3.57	3.33
ϵ_{fit} (relaxed)	14.44	14.27	11.05	12.48
$\Delta E_{FS}^{elec}(2\times2\times1)$ (eV)	1.99	2.06	4.96	5.17
$\Delta E_{FS}^{elast}(2\times2\times1)$ (eV)	−1.44	−1.51	−3.36	−3.89

TABLE III. Geometry relaxation of the region around each defect species. For each defect, the distances to and identities of the 1NN, 2NN, and 3NN atoms in the perfect crystal are shown followed by the corresponding distances after the geometry has been relaxed. The percentage changes in the 1NN, 2NN, and 3NN distances are also shown. As expected, positively charged defects attract the anions and repel the cations, while negatively charged defects do the opposite, although this rule is not obeyed strictly far from the defect.

Species	1NN	Change (to)	2NN	Change (to)	3NN	Change (to)
Al_{Al}	0.189	(6 O)	0.272	(4 Al)	0.336	(9 O)
V_{Al}^{0}	0.204	8.05%	0.261	−3.68%	0.337	0.40%
V_{Al}^{1-}	0.201	6.81%	0.250	−8.04%	0.337	0.30%
V_{Al}^{2-}	0.201	6.72%	0.257	−5.43%	0.334	−0.50%
V_{Al}^{3-}	0.204	7.96%	0.253	−6.81%	0.333	−0.70%
I site	0.189	(2 Al)	0.195	(6 O)	0.272	(6 Al)
Al_i^0	0.222	17.46%	0.213	9.23%	0.282	3.49%
Al_i^{1+}	0.228	20.63%	0.185	−5.13%	0.280	2.94%
Al_i^{2+}	0.228	20.63%	0.184	−5.64%	0.281	3.31%
Al_i^{3+}	0.228	20.63%	0.183	−6.15%	0.280	2.94%
O_O	0.189	(4 Al)	0.268	(12 O)	0.347	(8 Al)
V_O^0	0.186	−1.59%	0.268	−0.06%	0.347	0.01%
V_O^{1+}	0.199	5.57%	0.265	−0.93%	0.347	0.22%
V_O^{2+}	0.212	12.47%	0.262	−2.05%	0.348	0.36%
I site	0.189	(2 Al)	0.195	(6 O)	0.272	(6 Al)
O_i^0	0.183	−3.17%	0.213	9.06%	0.276	1.59%
O_i^{1-}	0.175	−7.41%	0.218	11.79%	0.274	0.74%
O_i^{2-}	0.166	−12.17%	0.224	14.87%	0.271	−0.37%

$$\Delta E_{\mathrm{FS}} = \frac{q^2 v_M}{2\epsilon_{\mathrm{fit}}}, \tag{11}$$

using these values of ϵ_{fit}. The resulting permittivities deviate significantly (approximately a factor of 2) from the experimentally measured zero-frequency permittivity of alumina, which is in any case anisotropic. This difference has several causes: first, the approximations of DFT lead to band-gap underestimation and thus incorrect dielectric properties; second, our periodic defect-containing supercells represent a highly concentrated solution of defects, the dielectric properties of which may differ significantly from those of the bulk; and finally, the scale of the variation of electric field considered here (nanometers) is very much smaller than that accessed by any macroscopic measurement of permittivity. The effective permittivity ϵ_{fit} can therefore be justifiably regarded as a fitting parameter. The values have associated statistical error bars from the least-squares fit of between 0.05 and 0.15, leading to "statistical" uncertainty in the energy of 0.1–0.2 eV—which for obvious reasons tallies well with the remaining variation in the results shown in Figs. 6 and 7.

In principle, if a fit of the formation energy to the first-order term was not felt to be sufficiently accurate, the variation in the second-order term in the Makov-Payne formula with both L and α could be incorporated, given careful assumptions about the dipole and quadrupole moments of the defect. Such an approach was not required here but could be an interesting topic for further work.

We can split the total finite-size error adjustment ΔE_{FS} into $\Delta E_{\mathrm{FS}}^{\mathrm{elec}}$ and $\Delta E_{\mathrm{FS}}^{\mathrm{elast}}$, these being the purely electrostatic and elastic contributions, respectively, so that $E_{\mathrm{def}}(L\to\infty) = E_{\mathrm{def}}(L) + \Delta E_{\mathrm{FS}}^{\mathrm{elec}} + \Delta E_{\mathrm{FS}}^{\mathrm{elast}}$, as shown in Table II.

The large differences between the extrapolated defect formation energies obtained with and without full lattice relaxation stress the importance of treating elastic and electrostatic effects together in a combined framework. If the atoms far from a charged defect are prevented from relaxing in response to the long-ranged electric field, the defect formation energy is substantially overestimated.

B. Defect formation energies

Having established a framework for the calculation of size-consistent defect formation energies parametrized by μ_{O} and μ_e [which are in turn known functions of (p_{O_2}, T) and ϵ_F, respectively], we now turn our attention to the formation energies and structural and electronic properties of these defects. Table III shows the relaxed and unrelaxed bond lengths around each type of defect for all accessible charge states. We note good agreement with the geometries of Matsunaga *et al.*[21] The bond-length changes clearly reflect the electronic structures of the various defects and their charge states. In forming an aluminum interstitial Al_i^q with $q=3$, we are simply inserting an Al^{3+} ion into a vacant site without affecting any existing bonds directly. The 1NN Al ions are significantly repelled and the 2NN O ions attracted. However, as q decreases to 2 and then 1, we occupy the s orbital on the Al_i

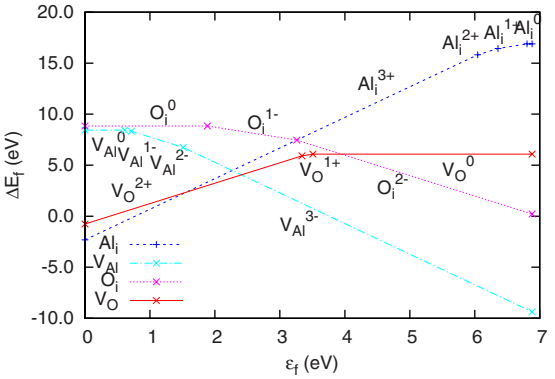

FIG. 8. (Color online) Defect formation energies of the four main types of intrinsic point defect in alumina calculated in DFT with ultrasoft pseudopotentials and the LDA. The calculation of μ_O was carried out at $T=2371$ K and $p_{O_2}=0.2$ atm.

atom and the defect charge decreases. At $q=0$, a defect state comprising s orbitals on 1NN Al sites is occupied, which causes significant repulsion of the 2NN oxygen ions: the resulting formation energy is so high that the Al_i^0 defect will be effectively absent regardless of temperature or doping level. A broadly similar effect is seen with the oxygen interstitial, although the relaxations have a smaller magnitude and the signs are reversed because the oxygen ion is negative.

The other defect that causes significant geometry rearrangement is the oxygen vacancy V_O^q. Removal of an O^{2-} ion including its electrons to create a V_O^{2+} defect causes a 12.5% relaxation of the 1NN Al ions toward the vacancy. If one or two electrons remain on the defect site ($q=1$ or $q=0$, respectively), the relaxation is greatly reduced, first to 5.6% for V_O^{1+} and then to −1.6% for V_O^0. The outward relaxation in the latter case is presumably due to the reduced localization of the electrons on the vacancy site relative to those on an O^{2-} ion. Removal of Al ions causes much smaller relaxations with less dependence on the charge state, probably because the defect states are more delocalized and associated with the 1NN oxygen atoms rather than the vacancy site itself.

From the relaxed defect supercell energies and the defect-defect interaction and band-gap error reduction formulae of Eqs. (11) and (8), respectively, we can use Eq. (1) to calculate the defect formation energies. Figure 8 collates the results of all of these calculations as functions of the Fermi energy ϵ_F. The position of the Fermi level in a real sample depends on the entire population of defects and impurities present: doping by aliovalent impurities will raise or lower ϵ_F, as will the creation of compensating native defects. If the formation energies of the native defects are sufficiently high, the intrinsic defect concentrations will be rather low. It then becomes unhelpful to talk about "undoped" alumina, since extremely small concentrations of dopants may be sufficient to swing the Fermi level one way or the other, affecting the concentrations of native defects significantly.

As a result of our methodological improvements, the formation energies of Fig. 8 differ in a number of ways from previous calculations of the same quantities. For example, Matsunaga *et al.*[21] applied the average potential shift to reduce defect-defect interactions and used full $m\times\Delta E_g$ band-gap error adjustment; they found that the four main intrinsic defects had nearly the same formation energies at $\epsilon_F=0$ and that V_{Al}^{3+} dominated for higher values of ϵ_F.

TABLE IV. Formation energies of the four simple intrinsic point defects of alumina. Values are calculated for μ_O at $T=2371$ K, $p_{O_2}=0.2$ atm, and $\epsilon_F=0$ (the valence-band maximum). The dependence on $\Delta\mu_O$ and $\Delta\epsilon_F$ is then shown explicitly. The experimental range of $\Delta\mu_O$ is approximately −1.5 to 2 eV (oxidation limit to reduction limit), while the range of $\Delta\epsilon_F$ is from 0 eV to $E_g\simeq 9.1$ eV. Errors in the formation energies due to the statistical nature of the least-squares fit used to determine the infinite size extrapolation are 0.1–0.2 eV.

Defects	Vacancy ΔE_f (eV)	Defects	Interstitial ΔE_f (eV)
V_O^0	$6.09+\Delta\mu_O$	O_i^0	$8.84-\Delta\mu_O$
V_O^{1+}	$2.58+\Delta\mu_O+\epsilon_F$	O_i^{1-}	$10.72-\Delta\mu_O-\epsilon_F$
V_O^{2+}	$-0.76+\Delta\mu_O+2\epsilon_F$	O_i^{2-}	$13.98-\Delta\mu_O-2\epsilon_F$
V_{Al}^0	$8.44-\frac{3}{2}\Delta\mu_O$	Al_i^0	$16.88+\frac{3}{2}\Delta\mu_O$
V_{Al}^{1-}	$9.05-\frac{3}{2}\Delta\mu_O-\epsilon_F$	Al_i^{1+}	$10.08+\frac{3}{2}\Delta\mu_O+\epsilon_F$
V_{Al}^{2-}	$9.77-\frac{3}{2}\Delta\mu_O-2\epsilon_F$	Al_i^{2+}	$3.72+\frac{3}{2}\Delta\mu_O+2\epsilon_F$
V_{Al}^{3-}	$11.27-\frac{3}{2}\Delta\mu_O-3\epsilon_F$	Al_i^{3+}	$-2.32+\frac{3}{2}\Delta\mu_O+3\epsilon_F$

By comparing graphs like Fig. 8 for a range of temperatures and pressures (and hence a range of oxygen chemical potentials), we find that the defect formation energy at the crossover value of ϵ_F, where the most stable defect switches from being a positively charged species to a negatively charged species, is always roughly 5 eV, regardless of μ_O. This is presumably the origin of the oft-quoted idea that all of the main defect formation energies are around this value—an idea that does indeed hold true if the Fermi energy balances out at this point in self-consistent calculations of systems containing neutral combinations of defects.

Over most of the range of ϵ_F, defects in their full formal charge states dominate. This supports the traditional picture in which intrinsic defects are created by adding or removing fully charged Al^{3+} or O^{2-} ions. It also ensures that band-gap error adjustments have no effect on the formation energies of the most important defects. However, under certain conditions of μ_O and μ_e it is possible for the formation energy of V_O^{1+} to be comparable to the other defects, which allows for the experimentally observable presence of F^+ centers in many alumina samples.[47]

As mentioned earlier, it is common practice in the study of defects to focus on μ_O- and μ_e-independent combinations: Schottky quintets, $3V_O^{2+}+2V_{Al}^{3-}$; oxygen Frenkel pairs, $V_O^{2+}+O_i^{2-}$; and aluminum Frenkel pairs, $V_{Al}^{3-}+Al_i^{3+}$. These may represent somewhat artificial quantities in alumina since the high defect formation energies render the truly intrinsic regime inaccessible. Nevertheless, our calculations support previous suggestions from experiment[48] and theory[21] that the energy ordering of these combinations is Schottky < Al Frenkel < O Frenkel.

Table IV presents parametrized expressions for the formation energies of the four main intrinsic defects in alumina including their dependence on Fermi energy and oxygen chemical potential. This table represents the main result of this work.

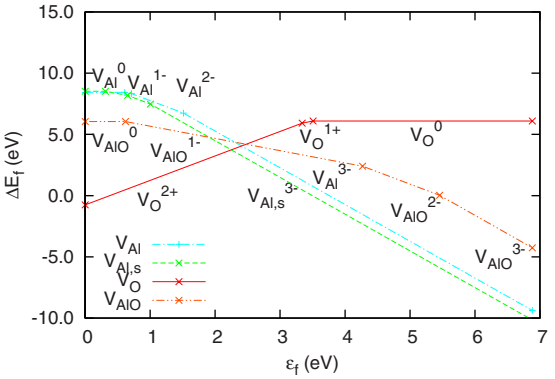

FIG. 9. (Color online) Defect formation energies of the split Al vacancy $V_{\mathrm{Al},s}=2V_{\mathrm{Al}}+\mathrm{Al}_i$ and the V_{AlO} defect alongside the V_{O} and V_{Al} single-atom vacancies. As with Fig. 8, results use ultrasoft pseudopotentials and the LDA. The calculation of μ_{O} was carried out at $T=2371$ K and $p_{\mathrm{O}_2}=0.2$ atm.

C. Alternative defects

Finally, although the main object of this study was the four simple intrinsic defects, we will address briefly two other defects that may affect diffusion in alumina. The aluminum-oxygen divacancy, V_{AlO}, created by the removal of adjacent oxygen and aluminum ions, has been suggested recently as a candidate for the mobile species in oxygen diffusion.[8,9] Because the dependence on μ_{O} is mostly canceled by the dependence on μ_{Al}, which changes as $\Delta\mu_{\mathrm{Al}}=-\frac{3}{2}\Delta\mu_{\mathrm{O}}$, the concentration of the V_{AlO} defect is only weakly affected by formation conditions. As ε_F is varied, the defect is stable in states ranging from $q=0$ to $q=-2$, and could be among the defects present in detectable quantities over some of the range, corresponding to low doping. As can be seen from its formation energy in Fig. 9, V_{AlO} could be a candidate for oxygen diffusion under certain circumstances, but only if its migration barrier was lower than that of V_{O} and concentrations under reasonable formation conditions were comparable. Such issues will be the subject of further investigation beyond the scope of this work.

The other additional defect to be considered here is an alternative structure for the aluminum vacancy that turns out to have a lower formation energy than the vacancy created by removing a single aluminum atom and relaxing the lattice. Instead, two aluminum atoms from the same column along the c axis, separated by a normally vacant site, are removed, and an aluminum interstitial is inserted into the vacant site. The split vacancy structure obtained by relaxation from this starting geometry has a formation energy lower by as much as 0.7 eV than the simple Al vacancy considered above. Its formation energy compared to the standard vacancy is shown in Fig. 9. This structure should perhaps therefore be regarded as the relaxed structure of the Al vacancy. However, its formation energy is not sufficiently much lower than the single vacancy to alter the conclusions of the preceding sections.

IV. CONCLUSIONS

This paper considered three different methods for treating some of the most challenging aspects of calculations of formation energies of charged point defects in nonelemental solids using DFT. The first, an extension of existing schemes for extrapolating calculated formation energies to the limit of isolated defects, varies both the shape and size of the supercell to achieve more accurate results with feasible supercell sizes. The second, which follows the approach of Finnis, Lozovoi, and Alavi,[11] allows chemical potentials to be calculated with greatly reduced dependence on the approximation scheme and makes it easier to relate chemical potentials to formation conditions. The third enables band-gap error adjustments to be calculated without having to identify strongly localized defect states as belonging entirely to either the valence or the conduction band, by calculating an approximation to the degree to which a state is like either band. By combining these methods, it is possible to obtain DFT estimates of defect formation energies that are both more converged with system size and choice of approximation and more easily linked to specific experimental formation conditions than previous results.

We have applied this methodology to the study of the intrinsic point defects of α-Al_2O_3 and calculated formation energies parametrized by the chemical potential of gaseous oxygen and the Fermi energy. A forthcoming paper will use these formation energies as inputs to a thermodynamic mass-action calculation, allowing realistic point defect concentrations to be obtained as functions of doping and formation conditions.[49] However, even without this thermodynamic study, several important conclusions can be drawn from the results presented here. First, the formation energies are sufficiently high that the concentrations of intrinsic defects will be rather low; the true "intrinsic" regime in which the effect of impurity doping is negligible is therefore difficult to access without exceptionally pure samples. The charge-neutral combinations of intrinsic defects such as Schottky quintets and Frenkel pairs, which must dominate in the intrinsic regime, may therefore be of little relevance in alumina: the effect of external aliovalent dopants providing charge and driving the Fermi energy up or down will almost always determine the native defect populations.

Our results also show that, while vacancies and interstitials in their highest charge states ($|q|=3$ for Al and $|q|=2$ for O) are the stablest under most conditions, there nevertheless exist regimes in which it is possible to create significant populations of defects such as V_{O}^{1+}, which are in lower charge states. Additionally, while we agree with previous work[21] showing that Schottky disorder has a lower energy per defect than either type of Frenkel disorder, it is not necessarily possible to neglect the contributions of interstitials to diffusion, especially if they have a notably lower migration barrier.

ACKNOWLEDGMENTS

N.D.M.H. and K.F. acknowledge the support from the U.K. Engineering and Physical Sciences Research Council. Computing resources were provided by the Imperial College High Performance Computing Service.

[1] M. C. Payne, M. P. Teter, D. C. Allan, T. A. Arias, and J. D. Joannopoulos, Rev. Mod. Phys. **64**, 1045 (1992).
[2] M. L. Cohen, M. Schlüter, J. R. Chelikowsky, and S. G. Louie, Phys. Rev. B **12**, 5575 (1975).
[3] A. F. Kohan, G. Ceder, D. Morgan, and Chris G. Van de Walle, Phys. Rev. B **61**, 15019 (2000).
[4] Çetin Kılıç and Alex Zunger, Phys. Rev. Lett. **88**, 095501 (2002).
[5] A. Togo, F. Oba, I. Tanaka, and K. Tatsumi, Phys. Rev. B **74**, 195128 (2006).
[6] S. B. Zhang, S.-H. Wei, and A. Zunger, Phys. Rev. B **63**, 075205 (2001).
[7] A. H. Heuer and K. P. D. Lagerlöf, Philos. Mag. Lett. **79**, 619 (1999).
[8] R. H. Doremus, J. Appl. Phys. **95**, 3217 (2004).
[9] R. H. Doremus, J. Appl. Phys. **100**, 101301 (2006).
[10] I. G. Batyrev, A. Alavi, and M. W. Finnis, Faraday Discuss. **114**, 33 (2000).
[11] M. W. Finnis, A. Y. Lozovoi, and A. Alavi, Annu. Rev. Mater. Res. **35**, 167 (2005).
[12] S. B. Zhang and J. E. Northrup, Phys. Rev. Lett. **67**, 2339 (1991).
[13] D. M. Smyth, *The Defect Chemistry of Metal Oxides* (Oxford University Press, New York, 2000).
[14] R. W. Grimes, J. Am. Ceram. Soc. **77**, 378 (1994).
[15] K. P. D. Lagerlöf and R. W. Grimes, Acta Mater. **46**, 5689 (1998).
[16] C. R. A. Catlow, R. James, W. C. Mackrodt, and R. F. Stewart, Phys. Rev. B **25**, 1006 (1982).
[17] F. Janetzko, R. A. Evarestov, T. Bredow, and K. Jug, Phys. Status Solidi B **241**, 1032 (2004) .
[18] J. Carrasco, N. Lopez, and F. Illas, Phys. Rev. Lett. **93**, 225502 (2004).
[19] Y. N. Xu, Z. Q. Gu, X. F. Zhong, and W. Y. Ching, Phys. Rev. B **56**, 7277 (1997).
[20] I. Tanaka, K. Tatsumi, M. Nakano, H. Adachi, and F. Oba, J. Am. Ceram. Soc. **85**, 68 (2002).
[21] K. Matsunaga, T. Tanaka, T. Yamamoto, and Y. Ikuhara, Phys. Rev. B **68**, 085110 (2003).
[22] A. E. Mattsson, R. R. Wixom, and R. Armiento, Phys. Rev. B **77**, 155211 (2008).
[23] Z. Wu and R. E. Cohen, Phys. Rev. B **73**, 235116 (2006).
[24] S. J. Clark, M. D. Segall, C. J. Pickard, P. J. Hasnip, M. J. Probert, K. Refson, and M. C. Payne, Z. Kristallogr. **220**, 567 (2005).
[25] D. Vanderbilt, Phys. Rev. B **41**, 7892 (1990).
[26] D. R. Hamann, M. Schlüter, and C. Chiang, Phys. Rev. Lett. **43**, 1494 (1979).
[27] D. M. Ceperley and B. J. Alder, Phys. Rev. Lett. **45**, 566 (1980).
[28] J. P. Perdew and A. Zunger, Phys. Rev. B **23**, 5048 (1981).
[29] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).
[30] H. d'Amour, D. Schiferl, W. Denner, H. Schulz, and W. B. Holzapfel, J. Appl. Phys. **49**, 4411 (1978).
[31] K. Johnston, M. R. Castell, A. T. Paxton, and M. W. Finnis, Phys. Rev. B **70**, 085415 (2004).
[32] Condensed Phase Thermochemistry Data, NIST Chemistry WebBook, NIST Standard Reference Database Number 69, 2005, (http://webbook.nist.gov).
[33] L. J. Sham and M. Schlüter, Phys. Rev. Lett. **51**, 1888 (1983); Phys. Rev. B **32**, 3883 (1985).
[34] S. B. Zhang, S.-H. Wei, A. Zunger, and H. Katayama-Yoshida, Phys. Rev. B **57**, 9642 (1998).
[35] S. B. Zhang, S.-H. Wei, and A. Zunger, Phys. Rev. Lett. **84**, 1232 (2000).
[36] J. Heyd, G. E. Scuseria, and M. Ernzerhof, J. Chem. Phys. **118**, 8207 (2003).
[37] G. A. Baraff and M. Schlüter, Phys. Rev. B **30**, 1853 (1984).
[38] I. Dabo, B. Kozinsky, N. E. Singh-Miller, and N. Marzari, Phys. Rev. B **77**, 115139 (2008).
[39] M. Leslie and M. J. Gillan, J. Phys. C **18**, 973 (1985).
[40] G. Makov and M. C. Payne, Phys. Rev. B **51**, 4014 (1995).
[41] P. A. Schultz, Phys. Rev. B **60**, 1551 (1999).
[42] P. A. Schultz, Phys. Rev. Lett. **96**, 246401 (2006).
[43] P. P. Ewald, Ann. Phys. **369**, 253 (1921).
[44] C. W. M. Castleton, A. Höglund, and S. Mirbt, Phys. Rev. B **73**, 035215 (2006).
[45] J. Shim, E.-K. Lee, Y. J. Lee, and R. M. Nieminen, Phys. Rev. B **71**, 035206 (2005).
[46] J. Lento, J.-L. Mozos, and R. M. Nieminen, J. Phys.: Condens. Matter **14**, 2637 (2002).
[47] M. Ghamnia, C. Jardin, and M. Bouslama, J. Electron Spectrosc. Relat. Phenom. **133**, 55 (2003).
[48] S. K. Mohapatra and F. A. Kröger, J. Am. Ceram. Soc. **61**, 106 (1978).
[49] K. Frensch, N. D. M. Hine, W. M. C. Foulkes, and M. W. Finnis (unpublished).