# First-principles calculation of the O vacancy in ZnO: A self-consistent gap-corrected approach

Tula R. Paudel and Walter R. L. Lambrecht
*Department of Physics, Case Western Reserve University, Cleveland, Ohio 44106-7079, USA*
(Received 14 August 2007; revised manuscript received 31 January 2008; published 8 May 2008)

The electronic structure of the oxygen vacancy in ZnO has been found to be sensitive to the corrections applied to the local (spin) density approximation (LSDA) band gap underestimate. Here, the "LSDA+$U$" approach, in which Hubbard-$U$ corrections are added to the local density approximation, is applied to both Zn $d$ and Zn $s$ orbitals. The justification of this approach is discussed. Transition state energies are calculated self-consistently instead of applying *a posteriori* corrections. The supercell size dependence and applicability of Makov–Payne corrections is investigated and an extrapolation approach inversely proportional to the cell volume is used. The 0/2+ transition level is found at 0.80 eV above the valence-band maximum and a small negative $U$ behavior is obtained with $U=-0.05$ eV. The Kohn–Sham one-electron levels in the different charge states are also presented and relevant experimental results are discussed.

DOI: 10.1103/PhysRevB.77.205202 PACS number(s): 71.55.Gs, 71.15.Mb

## I. INTRODUCTION

ZnO is an important wide-band-gap semiconductor, with many attractive features for optoelectronic applications compared to GaN: it has a higher exciton binding energy of about 50 meV (Refs. 1 and 2) and bulk single crystals can be grown, which could be useful as substrates. The defect physics has already received considerable theoretical attention of which we here quote only the most relevant papers.[3–11] Photoluminescence often shows a broad band green luminescence, which by some authors was attributed to oxygen vacancies,[5,12–15] although alternative origins, for example, a Zn vacancy[4] or Cu impurities for this luminescence have also been proposed.[16]

However, the electronic structure of the oxygen vacancy is not yet firmly established. In $p$-type material, i.e., with the Fermi level at or near the valence-band maximum, the system is in the 2+ charge state. Previous calculations[5–10] all found a negative-$U$ behavior, which means that the system has lower energy in the neutral state than in the single positive + state, owing to a much more efficient structural relaxation in the neutral state. They differ, however, in their predictions of the transition energy 2+/0, i.e., the Fermi level position where the crossover to a new ground state occurs. While Lany and Zunger[7] (LZ) proposed this transition level to lie in the lower half of the band gap, Janotti and Van de Walle[6] (JV) proposed it to lie in the upper half of the band gap. Although both used very similar pseudopotential plane wave *ab initio* calculations, they differ in the way band-gap corrections are applied. Previously, Zhang *et al.*[5] discussed various ways in which gap corrections can be applied. More recently, Erhart *et al.*[10] also found a position of the defect level in the upper half of the gap using the same extrapolation approach as JV but using the generalized gradient approximation (GGA) rather than local density approximation (LDA). Their results, however, differ in other respects, notably on the actual energies of formation. Patterson[11] found the 2+/0 state to lie close to the valence band edge at about 0.3 eV using all-electron B3LYP hybrid-functional[17,18] calculations.

On the experimental side, Vlasenko and Watkins[19] (VW) showed that either of these two results could be reconciled with the experimental data. The experimental analysis is based on optically detected electron paramagnetic resonance (ODEPR) signals in irradiated samples. The EPR signal comes from the single positive charge state, which according to the above theories is actually a metastable state. The optical luminescence associated with the ODEPR signal occurs by capture of an electron and hence turning it into the non-EPR active neutral state. If the luminescence associated with that capture involves a transition directly from a shallow donor, the level position would be deep (supporting the LZ model).[7] However, if it involves a two step process, capturing an electron from the shallow donor first, and subsequently capture of a hole from the valence band, it would indicate a much higher position of the +/0 level in the gap, in agreement with the JV model.[6] They argued in favor of the JV model on the basis of the excitation energies required for the process which did not show a sharp cutoff but stayed uniform down to 2.4 eV. A recent EPR study by Evans *et al.*,[20] however, found that the optical excitation from the $V_O^0$ state to the conduction band, required to produce the EPR signal, only occurred above 2.1 eV, thus suggesting a defect level for the neutral state in the lower half of the gap at less than 1.3 eV above the VBM. Also, Vlasenko and Watkins[19] showed that another EPR line labeled $X_2$ and potentially associated by them with a close {effective-mass-donor, $V_O$} complex would be in better agreement with the assumption of a deep +/0 level in the lower half of the gap.

Since the theoretical interpretation is based on a somewhat debatable *a posteriori* correction, it is important to further scrutinize these results. Besides the gap correction procedure, another possible source of uncertainty in the results in the finite size of the supercell. This aspect was already considered by Erhart *et al.*[10] but is also reconsidered here.

The position of the Zn $d$ levels is important for states near or in the gap.[2] The hybridization of Zn $3d$ with O $3p$ levels pushes up the valence band level. In LDA, the position of the Zn $3d$ levels is too high and thus the effect is overestimated. Both LZ and JV used the LDA+$U$ approach for the Zn $d$ orbitals. This corrects the gap partially, but not entirely. Whereas LZ simply shifted the VBM down in giving the defect level position relative to the VBM, and subsequently,

 

shift the conduction band minimum up without changes to the defect level, JV reasoned that it shows to what extent the defect level is conduction-band-like and thus they extrapolated the shift in defect level according to the full required gap correction even if the latter is not related to the Zn $d$ shift. Either of the two arguments seems to have its merits. It is thus important to obtain an independent calculation in which the full gap correction is actually included. Furthermore, LZ remark that the transition energies themselves did not differ much between LDA and LDA+$U$ and thus used the LDA transition energies but applied the LDA+$U$ correction only *a posteriori* to the host band edge. On the other hand, JV use the LDA+$U$ transition energies but *a posteriori* extrapolates them.

In this paper, we show that the LSDA+$U$ approach can be applied not only to the Zn $d$ but also to the Zn $s$ orbitals. This then allows us to fully correct the gap. Furthermore, in the LSDA+$U$ approach, we are not limited to considering one-electron levels. The method provides total energies which are consistently linked with the one-electron level shifts. Thus, we can actually obtain the defect transition levels using the usual $\Delta$SCF approach from the LSDA+$U$ Hamiltonian rather than applying *a posteriori* corrections. An alternative internally self-consistent total energy approach which also gives accurate band gaps can be found in Patterson's study.[11]

## II. COMPUTATION DETAILS

We used the full-potential linearized muffin-tin orbital (FP-LMTO) method[21] in conjunction with the density functional theory[22,23] in the local (spin) density approximation [L(S)DA] including Hubbard-$U$ corrections (LSDA+$U$) and a supercell approach. Here, we discuss the various aspects of this approach in turn.

The key feature of the present paper concerns the so-called "LSDA+$U$" approach, i.e., local spin density approximation + Hubbard-$U$ corrections. As is well known, this approach was first introduced by Anisimov *et al.*[24,25] to deal with open shell narrow $d$ bands. It essentially recognizes the fact that Coulomb interactions are orbital dependent and adds them explicitly for the narrow-band as in a Hubbard model. They are then treated in a Hartree–Fock-type mean field approximation. The fact that they are already partially accounted for in the LDA is corrected by a double-counting term. One can also apply it of course to a full shell semicore-like band such as the Zn 3$d$ band in ZnO.

In a spherically averaged version[26] of the general theory[27] there is an additional term in the total energy functional given by

$$E_{LSDA+U} = E_{LSDA} + \frac{(U-J)}{2}\sum_{\sigma}[\mathrm{Tr}\,\rho^{\sigma} - \mathrm{Tr}(\rho^{\sigma}\cdot\rho^{\sigma})], \quad (1)$$

in which $\rho^{\sigma}$ is the density matrix of the orbitals treated by LSDA+$U$. Minimization versus the density matrix, then leads to a shift in one-electron potential, given by

$$V^{\sigma}_{mm'} = (U-J)\left[\frac{1}{2}\delta_{mm'} - \rho^{\sigma}_{mm'}\right]. \quad (2)$$

In the appropriate basis, the density matrix becomes diagonal $\rho_{mm'}=n_m\delta_{mm'}$ and the corresponding elements are just the occupation numbers. Now, it is clear that if one applies this to an empty orbital $n_m=0$ for all $m$, we obtain an upward shift by $(U-J)/2$ and if we apply it to a filled orbital, we obtain a downward shift by $-(U-J)/2$. In practice, we set $J=0$ here because we have a closed shell and adjust $U$ only.

The underlying physics for justifying our $d$ and $s$ shifts is quite different. The gap underestimate in a typical semiconductor such as ZnO arises in part from the position of the Zn $d$ level but mostly from the long-range Coulomb effects in the screened exchange term. The best starting point for analyzing the gap underestimate is the $GW$ theory.[28,29] Maksimov *et al.*[30] showed that the statically screened exchange is the most nonlocal term in the $GW$ self-energy, schematically $\Sigma = iG\times W$. In the static approximation, this becomes $\Sigma = -\rho\times W$, i.e., the Green's function is replaced by the density matrix. The theory then looks exactly like Hartree–Fock but with a screened instead of bare Coulomb interaction. Now, in a semiconductor, the latter is long-ranged because even when the dielectric response becomes the macroscopic dielectric constant, the behavior is as $1/\varepsilon r$. This is what LDA is missing because it is based on an exponential metalliclike screening of the electron gas. In a $GW$ calculation, one would eventually take matrix elements of the $\Sigma - v_{xc}$ between conduction band states. Now, if the conduction band states primarily consist of cation $s$-like states, as is the case in ZnO, then the correction will manifest itself as a shift in the on-site $s$-diagonal elements of the Hamiltonian in a basis set of LMTO orbitals. Thus, we can mimic its effect by a simple shift projection operator $|\psi_s\rangle\Delta_s\langle\psi_s|$. This, as we have shown above is also what would happen in LSDA+$U$ if we apply a $U_s$ to the $s$ orbitals. Thus, although the underlying physics is different for $d$ and $s$ shifts they can in practice be treated in the same manner, as long as we are willing to include some empirical parameter like the effective shift. For $d$ shifts, the origin is intra-atomic differences in the Coulomb terms owing to the more localized nature of the $d$ orbitals compared to diffuse $sp$ orbitals. For the $s$ shift, the reason is that the conduction band minimum is primarily Zn-$s$-like and hence the gap shift owing to the quasiparticle self-energy of a $GW$ theory appears in a muffin-tin-orbital basis through a shift in the $s$-partial waves.

As far as the convergence of the FP-LMTO method is concerned, we used a double basis set of $spds_2p_2d_2$ with different $\kappa$ and smoothing radii for each orbital, a $20\times20\times32$ real space mesh for the smooth part of the wave functions, charge density and potential, and shifted $4\times4\times4$ reciprocal space mesh for the Brillouin zone integrations in the wurtzite case and correspondingly scaled versions for the supercells. The supercell size convergence is discussed in Sec. III. The charged defects are treated by including a homogeneous background charge density. The calculations are performed at the experimental lattice parameters.

The energies of formation for different charge states $q$,

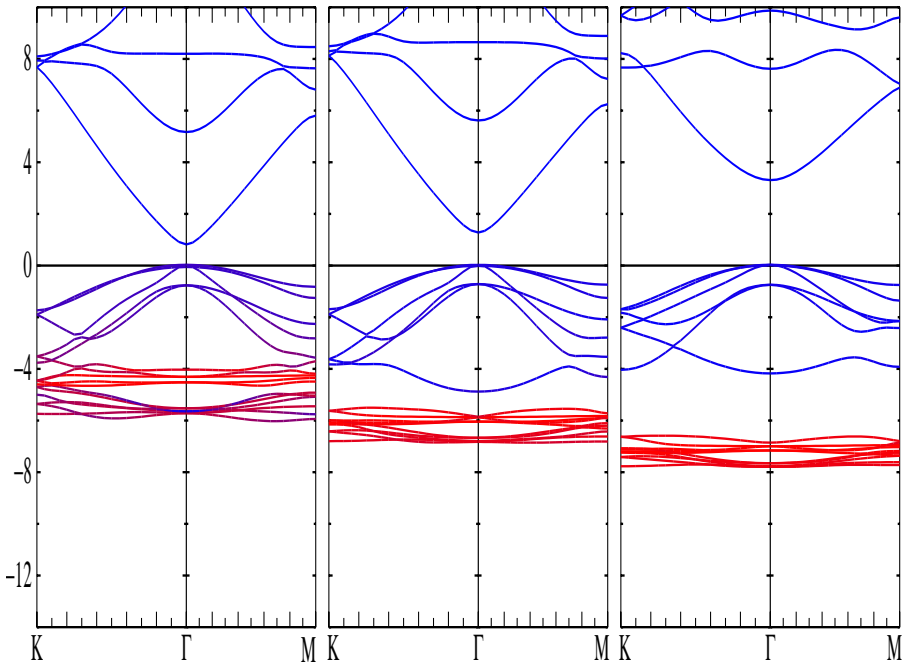


FIG. 1. (Color online) From left to right, LDA and LDA+$U_d$ and LDA+$U_d$+$U_s$ band structures of ZnO. In the color version, the contribution of the Zn $d$ orbitals to the bands is highlighted by its degree of redness against a blue counter color, with purple indicating strong hybridization. Because this information is not visible in the printed gray scale, we point out that the (red) $d$ bands occur just above −8 eV in the right panel, centered around −6 eV in the middle panel and at −4 eV in the left panel. In the left panel, strong hybridization (reddish purple) takes place with the bottom of the O $p$ bands at −6 eV and moderate hybridization (bluish purple) with the higher valence bands. In the LDA+$U$ cases, the O $p$ bands show little hybridization with the Zn $d$ and appear almost purely blue. The conduction bands contain negligible Zn $d$ and are thus purely blue.

$$\Omega_f(V_O,q,E_F) = E_{sc}(\mathrm{ZnO}{:}V_O) - E_{sc}(\mathrm{ZnO}) + \mu_O + q(E_{vbm} + E_F), \quad (3)$$

and transition energies,

$$\epsilon(q,q') = \frac{\Omega_f(V_O,q,0) - \Omega_f(V_O,q',0)}{q' - q}, \quad (4)$$

are defined as usual. Here, $E_{sc}$ are the supercell total energies (minus the energies of the free atom) for the system with and without the defect, $E_{vbm}$ is the bulk valence band maximum energy with respect to the electrostatic potential reference energy in the supercell, and $E_F$ is the Fermi level measured from the valence band maximum. To determine $E_{vbm}$, the usual procedure is to determine $E_{vbm}$ in the bulk material relative to some local potential reference, e.g., the potential at the muffin-tin radius. We can then use the same local reference at an atom far away from the defect in the defect cell where presumably the potential has become bulklike to obtain $E_{vbm}$ in the supercell. All one-electron energies, however, must be measured with respect to the volume averaged electrostatic potential, in order to be compatible with the total energy differences in Eq. (3). This is automatically the case in plane-wave pseudopotential calculations but not necessarily in all-electron methods, where special care must be taken to calculate the average electrostatic potential separately. Alternatively, we can plot the energy bands for the supercell and identify the valence band maxima and defect levels directly. Both procedures give the same value. It is important to note however that it changes from LDA to LSDA+$U_d$ to LSDA+$U_d$+$U_s$. The chemical potential for oxygen depends on whether the system is oxygen or Zn rich. In this paper, we take the chemical potential of the oxygen as the half the chemical potential of molecular oxygen (−3.76 eV) corresponding to the oxygen-rich limit. To obtain the Zn-rich limit $\mu_O = \mu_{ZnO} - \mu^0_{Zn}$, one has to add −3.1 eV, which is the energy of formation of ZnO from bulk Zn and molecular oxygen.[5]

## III. RESULTS

We first discuss the effects of the LSDA+$U$ on the band structure of ZnO. In our present calculations, we add both $U_d$ and $U_s$ within the LSDA+$U$ approach. The results are shown in Fig. 1. If we add only the $U_d$=3.40 eV, the center of the Zn $d$ orbital (−5.37 eV) with respect to the valence band maximum is shifted down by using the Hubbard parameter $U_d$ to (−6.37 eV) which is similar to its position in the *GWA* (−6.0 eV).[31] Compared to the LDA, where the gap is only 0.81 eV, the gap now becomes 1.30 eV. By adding $U_s$ =43.54 eV, the gap is adjusted to 3.3 eV, close to the experimental value of 3.4 eV. At the same time, we see that this also shifts the Zn $d$ bands further down to −7.41 eV, which is much closer to experimental position of −7.5 eV reported by Powell *et al.*[32] Such "side effects" are not so easy to predict but result from adjustments of the self-consistent potential to the added Hubbard-$U$ terms. Another effect of adding $U_d$ and $U_s$ not seen in the figure in which we aligned the VBM is

that the $E_{vbm}$ relative to the electrostatic reference potential changes. It is 8.607 eV in LDA, 8.165 eV in LDA+$U_d$, and 8.727 in LDA+$U_d$+$U_s$. As expected, $U_d$ has the effect of shifting the VBM down by about 0.4 eV. However, $U_s$ not only shifted the conduction band up but also shifted the VBM back up. It is not entirely clear if this absolute position would agree with *GW* calculations which rarely discuss the absolute position of the bands. We attribute it tentatively here to the fact that in wurtzite, the VBM has by symmetry a small *s* component allowed in it and hence may shift up slightly in view of our high $U_s$ value required to obtain the experimental gap. The value of $U_s$ needed may seem excessively high. This is because in practice the occupation of Zn *s* in the conduction band is not actually zero but of order 0.05. The O 2*s*-like band is really a Zn *s*-$O_s$ bonding state and the conduction band is the corresponding antibonding state. As long as the occupation $n_{Zns}<0.5$, we will get an upward shift of the corresponding orbital energies. Furthermore, the shift in our model is applied only to the Zn *s*-like partial wave inside the muffin-tin radius. In reality the corresponding muffin-tin orbital basis function has a significant contribution from the interstitial region. Finally, the resulting band shifts do not equal the orbital energy shifts in the Hamiltonian because of the hybridizations of the bands. Thus, we should not attach too much physical interpretation to the $U_s$ needed but only the final band structure that is produced by our effective Hamiltonian. That same Hamiltonian which gives an accurate band structure close to the experimental gap can now be applied to the defect system.

Next, we discuss the supercell size effect. The total energy calculations were performed for supercells with 72, 108, and 192 atoms in LDA and 72, 108, 128, and 192 atoms in LDA+$U$ with a regular $k$-point mesh of $2\times2\times2$ for the cell sizes of 72–128 and a $\Gamma$-point calculation 192-atom cell.

Makov and Payne[33] proposed that the total energy of the defect containing supercell should scale as $1/L$ with $L$ the linear length scale of the supercell because of the dominance of the spurious monopole electrostatic term representing the interaction of the net charge of the point defect with the neutralizing background charge density. In fact, this monopole correction was already proposed ten years earlier by Leslie and Gillan[34] and Makov and Payne[33] proposed a further quadrupole correction varying as $1/L^3$ should be added. However, this model rests on the assumption of a well-defined multipole expansion and was shown by Segev and Wei[35] to be an overestimate if the defect charge density is delocalized. This is manifestly the case here, with the charge density distributed beyond the nearest neighbor Zn atoms. They proposed instead a scaling with number of atoms or equivalently inverse scaling with volume. A more detailed approach to these corrections was proposed by Blöchl.[36] Using a simple model for the Madelung term, proposed by Carloni *et al.*,[37] one would expect the energy of formation (at $E_F=0$) to vary as

$$\Delta\Omega_f(R) = \Omega_f(R=\infty) - \frac{9q^2}{10\varepsilon R}, \tag{5}$$

where $R$ is the radius of a sphere with volume equal to that of the supercell. The dielectric constant $\varepsilon$ in this equation can in principle be obtained by fitting this equation to the supercell results of different size and $\Omega_f(R=\infty)$ is the desired limit for an infinite cell size. The sign of the correction and scaling with the charge state are important. The model predicts that the cell would have a spurious over-binding by the electrostatic interaction of the defect charges with the background and should be four times as large for a 2+ than for a + charge state. If we plot the supercell energies of formation as function of $1/R$, as shown in Fig. 2, however, we find positive slopes for the LDA case and for $q=0$ for the LDA+$U$ case. This might indicate that there is another type of $1/R$ dependence, perhaps resulting from elastic terms or other long-range effects in the total energy, as was also found by Castleton *et al.*[38] We might then attempt to fit an equation $\Omega_f(R)=\Omega_f(R=\infty)+\alpha/R-9q^2/(10\epsilon R)$ with an additional unknown coefficient $\alpha$ representing the effects independent of charge. Even this is found to be impossible because the slopes do not follow the expected dependence on $q$. We also tried fitting the initial unrelaxed formation energies versus $1/R$ and this gives indeed correct signs of the slopes, indicating indeed that other $1/R$ dependences not electrostatic in nature may occur as a result of the relaxation. This is a pronounced effect here because of the strong relaxations for this defect. One might then attempt to correct the unrelaxed results for the electrostatic $1/R$ effects and afterward include the size dependence of the relaxation energy separately. However, even the unrelaxed formation energies cannot be fit with a single effective dielectric constant, again pointing to the problems with the validity of the multipole expansion. In any case, the relaxation and electrostatic effects appear to compensate each other to some extent and the results themselves do not really support a dominant $1/R$ behavior.

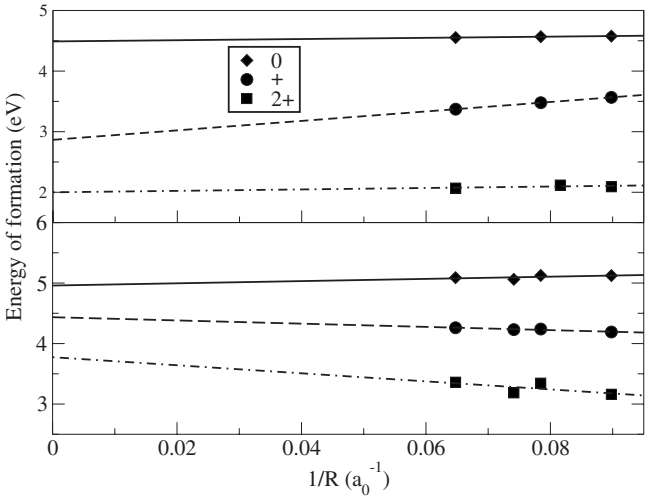


FIG. 2. Energy of formation plotted versus $1/R$ with $R$ defining the linear length scale $4\pi R^3/3=V$ in units of the Bohr radius $a_0$. Top: LDA and bottom: LDA+$U$.

From the data itself, it is not clear whether a $1/R$ or $1/R^3$ scaling or some combination of the two as suggested by Castleton *et al.*[38] provide a better fit because the range of supercell sizes treatable is too small. In the absence of a clear physical motivation for a $1/R$ behavior and considering that the latter gives an extrapolated value quite far from the actually calculated results, it appears safer to use an inverse

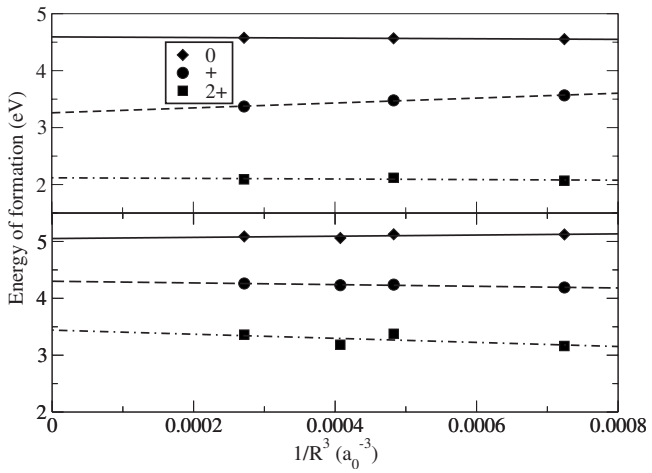


FIG. 3. Energy of formation plotted versus 1/volume (in atomic units) of the cell. Top: LDA and bottom: LDA+$U$.

volume scaling. The plot assuming an inverse volume dependence, shown in Fig. 3, appears to be better behaved in the sense that it leads to a well defined $\Delta E_f(R=\infty)$ extrapolation which is not too different from the values for the large cells. In the interest of not introducing extra fitting parameters in the extrapolation we prefer this over a fit to both types of terms.

The above defined 1/volume finite size scaling approach allows us to extract the energies of formation for the oxygen vacancy in the dilute limit (Fig. 4). In practice, they agree with those of our largest cells to within an uncertainty of about 0.1 eV. In LDA, the lowest energy of the vacancy for $E_F=0$ is for the 2+ charge state. For a slightly higher Fermi energy, it goes into the + charge state before going finally to the neutral charge state. The 2+/+ transition is lower than the +/0 transition; hence, the system does not show negative $U$ behavior. These results differ from previous results in literature, mainly as a result of our different extrapolation to the dilute limit. In fact, if we had used the finite cell results directly, we would still have found a negative $U$. Here, $U$ refers to the overall Coulomb + relaxation energy for adding

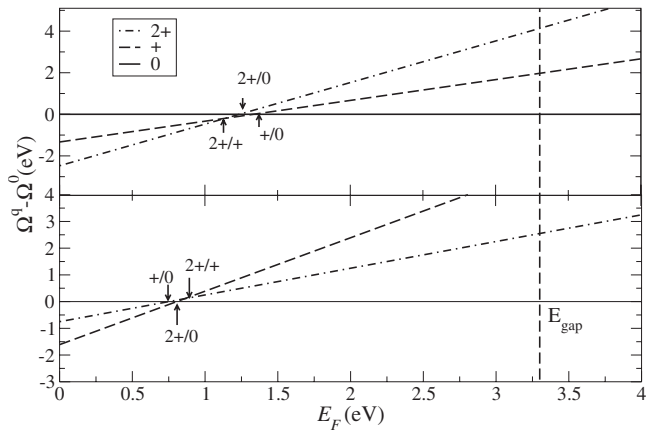


FIG. 4. Difference in energy of formation between charged and neutral defect state in the limit of infinite dilution as function of Fermi energy: Top, LDA and bottom, LDA+$U$.

a second electron to the defect level, i.e., $\epsilon(+,0)-\epsilon(2+,+)$. The latter should obviously not be confused with our $U_d$ and $U_s$ energies of the LDA+$U$ method which refer to Coulomb energies of particular orbitals.

In LDA+$U$, the system also starts out in the 2+ state but now it goes directly to the neutral state and hence shows negative $U$ behavior but $U$ is rather small. Furthermore, we see that all the energy crossings (i.e., transition states) occur somewhat closer to the valence band maximum. The LDA +$U$2+/0 level lies at 0.80 eV. This is somewhat surprising. One would have expected the direct effect of the $U_d$ to be to shift the VBM down and of the $U_s$ to shift the defect level as well as the conduction band up. However, as mentioned earlier, we find that in the LDA+$U_d$+$U_s$ calculation, the $E_{vbm}$ in Eq. (3) actually shifts up relative to the electrostatic reference potential compared to LDA and this tends to make the energies of formation of the + and 2+ states closer to the energy of the 0 state, which leads to a lowering of the transition energies.

The atomic relaxations are crucial to obtain the negative $U$ behavior. Relaxation effects are important to determine the energy of formation and hence the transition levels. Both in LDA and LDA+$U$, the Zn atoms surrounding the neutral vacancy relax inward while for the charged vacancy they relax outward. However, the extent of relaxation is quite different. In LDA the neutral vacancy relaxes inward by 3% while the + and 2+ charged vacancies relax outward on average 23% and 22%. These are percentages in distance of the Zn atom to the position of the vacancy. This indicates that for the charged vacancies, the atoms relax by forming stronger back bonds while the neutral vacancy relaxes by attempting to rebond within the vacancy. It is this strong inward relaxation of the neutral vacancy which is responsible for the negative $U$ behavior. In LDA+$U$, the neutral vacancy relaxes inward by 8%, i.e., more than in the LDA case, while the + and 2+ charged vacancies relax outward by 3% and 18%, which is smaller than in LDA, in particular, for the + charge state. Apparently, the Zn dangling bonds can more effectively rebond when filled or get less repelled from each other when empty when they include less of the Zn $d$ contribution. The amounts of relaxation found here are similar to those given by JV, which are 12% inward, 2% outward, and 23% outward for 0, +, and 2+ respectively.

Because this system has been controversial for a while, it is instructive to provide a comprehensive comparison. In Table I, we compare our results to those of other groups and also show in detail how the final results depend on the size of the superlattice and on the use of the LDA or LDA+$U$ approach. We can see that in LDA a negative $U$ would have been obtained for smaller cell sizes than our extrapolated value but the transition energies lie around 1.00–1.5 eV, which is already beyond the LDA gap.

Comparing with the other LDA+$U$ calculations, there are several important differences. First, the other calculations only included a $U_d$ which only partially corrects the gap. Secondly, LZ simply looked at how $U_d$ shifts the one-electron levels and then correspondingly introduce an *a posteriori* shift of the transition levels. Some authors, JV[6] and Erhart *et al.*[10] extrapolate the defect level shift in proportion to the gap shift to the experimental gap, while LZ[7] did not do

TABLE I. Comparison of energies of formation and transition levels between different calculations. Unless otherwise indicated, the results are obtained here. ∞ means extrapolated to the dilute limit.

| | | Energy of formation at $E_F$=0 (eV) | | | Transition energy (eV) | | | |
|---|---|---|---|---|---|---|---|---|
| Method | Supercell | 0 | + | 2+ | +/0 | 2+/0 | 2+/+ | $U$ |
| LDA | 72 | 4.55 | 3.57 | 2.07 | 0.97 | 1.24 | 1.50 | −0.53 |
| | 108 | 4.57 | 3.48 | 2.11 | 1.09 | 1.23 | 1.37 | −0.28 |
| | 192 | 4.58 | 3.38 | 2.09 | 1.20 | 1.24 | 1.29 | −0.09 |
| | ∞ | 4.59 | 3.26 | 2.11 | 1.33 | 1.24 | 1.15 | 0.18 |
| | 36[a] | 4.6 | 3.9 | 2.6 | 0.7 | 1.0 | 1.3 | −0.6 |
| | 72[b] | | | | 0.17 | 0.83 | 1.49 | −1.32 |
| | 96[c] | | | | 0.16 | 0.73 | 1.30 | −1.14 |
| GGA | ∞[d] | 4.46 | 3.72 | 2.98 | 0.74 | 0.74 | 0.74 | 0 |
| LDA+$U_d$+$U_s$ | 72 | 5.12 | 4.19 | 3.16 | 0.93 | 0.98 | 1.03 | −0.05 |
| | 108 | 5.13 | 4.24 | 3.34 | 0.89 | 0.89 | 0.89 | −0.004 |
| | 128 | 5.06 | 4.23 | 3.18 | 0.83 | 0.94 | 1.05 | −0.11 |
| | 192 | 5.09 | 4.26 | 3.36 | 0.83 | 0.86 | 0.90 | −0.04 |
| | ∞ | 5.05 | 4.30 | 3.44 | 0.75 | 0.80 | 0.85 | −0.05 |
| LDA+corr. | 36[a] | 5.5 | 4.6 | 0.1 | 0.9 | 2.7 | 4.5 | −3.6 |
| LDA+$U_d$ | 72[b] | | | | 0.94 | 1.60 | 2.24 | −1.32 |
| LDA+$U_d$ | 96[c] | | | | 0.64 | 1.18 | 1.73 | −1.09 |
| LDA+$U_d$ extrapol | 96[c] | | | | 1.94 | 2.42 | 2.90 | −0.96 |
| GGA+$U_d$ | ∞[d] | 5.17 | 4.17 | 2.73 | 1.0 | 1.22 | 1.44 | −0.44 |
| GGA+$U_d$ extrapol | ∞[d] | 5.49 | 4.37 | 1.85 | 1.12 | 1.82 | 2.52 | −1.40 |
| B3LYP | 72[e] | | | | 0.04 | 0.34 | 0.64 | −0.60 |

[a]Zhang *et al.* (Ref. 5) pseudopotential calculations with plane-wave cutoff gap correction.
[b]Lany and Zunger (Ref. 7) LDA and LDA+$U$ calculations using the projector augmented-wave method (PAW) (Ref. 39).
[c]Janotti and Van de Walle (Ref. 6) pseudopotential LDA, LDA+$U$, and LDA+$U$ extrapolated to full gap.
[d]Erhart *et al.* (Ref. 10) PAW (Ref. 39) GGA, GGA+$U$, and GGA+$U$ extrapolated to full gap correction.
[e]Patterson *et al.* (Ref. 11), B3LYP all-electron, Gaussian orbitals.

this extrapolation. In contrast, we use both $U_d$ and $U_s$ which fully corrects the gap and consistently calculate total energies and hence transition energies within this approach. Besides these differences, the size of the cells also plays a crucial role and obviously larger cells or a systematic way of extrapolating is to be preferred. Our way of extrapolating differs from that of Erhart *et al.*[10] They did include a monopole Leslie–Gillan correction first and then extrapolated versus 1/volume whereas we directly extrapolate versus 1/volume because the results were found not to scale as expected from the Leslie–Gillan electrostatic model at all. Also, while all other calculations assumed a fixed perfect crystal volume and only allowed the atomic positions to relax, Erhart *et al.*[10] allowed the volume of the cells containing the defects to relax to their optimum value. This means the results really apply to the high concentration of vacancies considered in the supercell whereas keeping the volume of the cell fixed better represents the dilute limit. Finally, Zhang *et al.*[5] discussed various ways of correcting for the band gap but used fairly small cells. Other differences in the methodology, pseudopotential versus projector-augmented wave (PAW) or all-electron FP-LMTO and $k$-point convergence may also play a minor role but are harder to trace back. The B3LYP results of Patterson[11] give a band gap of 3.34 eV close to the experimental gap and place the defect transition levels very close to the valence band edge. Interestingly, their approach which is a hybrid density-functional-theory Hartree–Fock-type approach as one could say is our LSDA+$U$ also place the defect levels closer to the VBM instead of farther away as obtained in the $U_d$ only approaches. In our case, however, this is ultimately due to the shift in $E_{vbm}$ relative to the electrostatic potential.

Our results give a much less negative $U$ than previous work. This is related to the degree of relaxation. For the + charge states, our results on the degree of relaxation agree fairly closely. However, JV for example obtain a larger difference between the 2+ and neutral states and correspondingly a more negative $U$ value.

We note that there is also some controversy on the formation energies themselves. While we as well as LZ[7] and Erhart *et al.*[10] find relatively low energies of formation of the oxygen vacancy in the neutral state, JV find a much higher

value. This has implications for the question whether the oxygen vacancy can be expected to exist in as-grown samples. We wish not to delve into this controversy here but merely note that most experimental studies on the vacancy were done with samples in which the vacancy is introduced by radiation.

Next, we turn to a comparison with the experimental studies. According to Vlasenko and Watson,[19] the optically detected EPR signal corresponds to a spin-dependent process

$$EM^0 + V_O^+ \rightarrow EM^+ + V_O^0. \quad (6)$$

The corresponding energy difference is $E_g + \Omega_f(+) - \Omega_f(0)$, which evaluates to 2.53 eV, using our calculated gap of 3.28 eV. Strictly speaking, we should here use the energy of the neutral state in the structure of the initial + state which might be somewhat higher, perhaps of order 0.1 eV. In the simplest interpretation, this electron transfer leads to the luminescence band centered at 600 nm and for which VW estimate the onset to be at 500 nm corresponding to 2.47 eV. To within 0.1 eV, there appears to be good agreement.

In the study of Evans *et al.*,[20] the reverse process is studied. Namely, the activation of the $V_O^+$ by exciting an electron from the neutral vacancy to the conduction band. They find that a minimum excitation energy of 2.1 eV is required. One might have expected on the basis of the Frank-Condon principle that this energy would have been slightly higher, rather than lower than the value given by VW. These authors notice that at the same time the $Fe^{3+}$ EPR signal present in their sample is quenched. Fe-impurity traps are thus preventing the backflow of the electron to the $V_O^+$ state resulting from the optical excitation. The possibility must be considered that the electrons in this case are directly transferred to the Fe defect without the intermediate of the conduction band The position of the defect level for $Fe^{3+}$ is not known experimentally but in a calculation by Wardle *et al.*,[40] it occurs at about 2.5 eV above the VBM or about 0.9 eV below the conduction band. This would then lead to an estimate of 1.7 eV for the activation energy of the $V_O^+$. This seems too low to be in agreement with the data and is furthermore implausible because this would require that both defects are in close proximity of each other. On the other hand, the value obtained by VW is possibly somewhat high because they base it on the onset of the luminescence band located at 500 nm instead of on its maximum occurring at 600 nm. To within the uncertainties of a few 0.1 eV, these two experiments agree on a position of the defect level in the lower half of the gap. If we then settle on a most plausible energy value of $2.2 \pm 0.1$ eV for this energy transition from the conduction band to +/0 transition level based on the experiments, it would mean that our calculation positions the defect levels a little bit too close to the VBM. As mentioned earlier, this derives in part from the upward shift of $E_{vbm}$ due to our $U_s$ parameter, which may be slightly overestimated.

It is also of interest to consider the Kohn–Sham one-electron eigenvalues of the defect in its different charge states. In Fig. 5, we show the bands of the 192 atom supercell in the different charge states in the LSDA+$U_d$+$U_s$ model, while in Fig. 6, we show them for the LDA case. In the LDA+$U$ case, defect bands can easily be recognized in

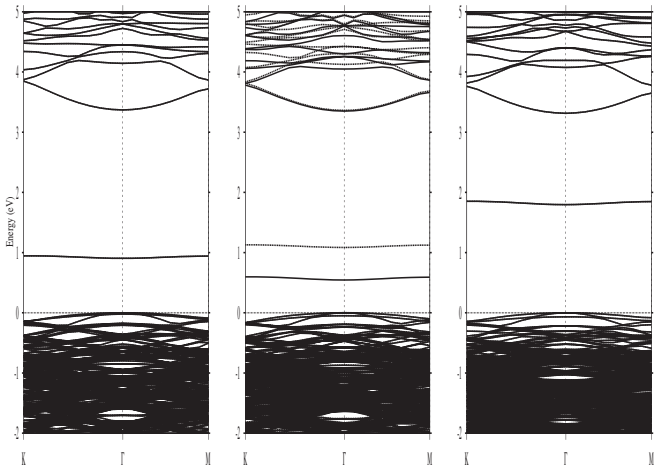


FIG. 5. Energy bands of the 192 atom supercell in LDA+$U_d$ +$U_s$ in the different charge states from left to right 0, +, and 2+.

the gap as almost dispersionless bands. We can see that the empty level of the 2+ state lies quite high in the gap at 1.79 eV. This is consistent with the fact that in this case the defect relaxes outward and thus the defect level corresponds to more or less unperturbed Zn dangling bonds. Upon catching a first electron, however, the occupied defect level moves significantly down to 0.55 eV above the VBM. This happens by rebonding the dangling bonds with the available electron leading to a significant reduction of the outward relaxation. An empty state of opposite spin occurs at 1.1 eV. Upon catching a second electron, the system again becomes nonspin-polarized and the now doubly filled level moves slightly up to 0.9 eV. Thus, the optical transition from the ground state of the $V_O^+$ to the conduction band is predicted to be about 2.4 eV, similar to the conclusions obtained from the transition energies. Again, it is somewhat larger than in the measurement of Evans *et al.*[20] We can also try to interpret the Vlasenko and Watkins[19] results using the one-electron levels. In this case, an electron is transferred from the effective mass donor to the first empty level of the 1+ charge state vacancy, which corresponds to an energy for the luminescence of 2.2 eV.

In the LDA case, the band in the gap occurs for the neutral case and it lies at 0.2 eV at the $\Gamma$ point, very close to the VBM. Even this band has significantly more dispersion than in the LDA+$U$ case. In the 2+ state, a relatively flat band different from the folded bulk bands is seen as the second conduction band. This means the empty state of the defect level is now a resonance instead of a gap level. In the 1+ state, this band is still above the conduction band minimum, and thus the extra electron was added to the conduction band minimum instead of in a localized level. This shows clearly the failure of the LDA to correctly describe the electronic structure of the defect. Because of the very strong underestimate of the gap, the filling of the levels is incorrect. This also explains why in the LDA case, the relaxation of the 1 + state is not very different from that in the 2+ state, in contrast with the LDA+$U$. It also implies that the defect level should not be moved up along with the conduction band when applying gap corrections because it is essentially

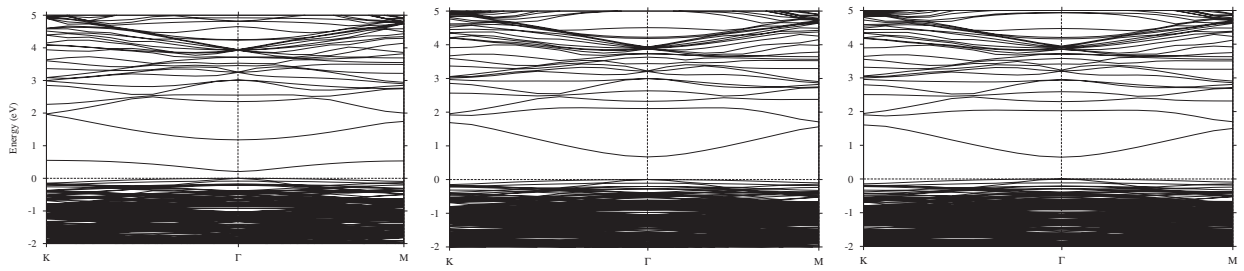


FIG. 6. Energy bands in the 192 atom supercell in LDA for the different charge states, from left to right 0, +, and 2+.

a different state. A much closer relation between the one-electron levels and the transition levels is observed in the LDA+$U$ case. The occurrence of a defect level deep in the gap even for the 2+ states in our LDA+$U$ calculation invalidates Lany and Zunger's proposed model[7] for persistent conductivity due to the oxygen vacancy because their model is based on the occurrence of the level as a resonance above the conduction band.

Our LDA+$U_d$+$U_s$ defect Kohn–Sham band structure has close similarities to that of the B3LYP calculation of Patterson,[11] except that their levels lie even closer to the valence band maximum and show a larger spin-splitting for the 1+ state. This is consistent with the fact that Hartree–Fock has stronger spin splittings. While in B3LYP, a mixture of LDA and Hartree-Fock is used, we use in some sense a screened Hartree–Fock.

## IV. CONCLUSION

All-electron first-principles total energy calculations were performed for the oxygen vacancy in ZnO using the LSDA +$U$ approach with both a shift of the Zn $d$ and Zn $s$ bands which (almost) fully corrects the band gap. The corrections were applied systematically to the total energy calculations and hence transition energies instead of applied as *a posteriori* corrections. The supercell size effects were studied and calculations for different sizes were extrapolated in a systematic way by assuming an inverse volume dependence. No justification for the Leslie–Gillan or Makov–Payne monopole correction was found since the predicted inverse linear size behavior had the wrong signs expected from electrostatics and did not scale with the square of the charge state $q^2$. We caution that this result may be specific to the type of defect and the degree of localization of the defect charge. In the present case of a vacancy, the extra charge resides on neighboring dangling bonds and further atoms and is thus not consistent with a point charge. Furthermore, relaxation is strong and leads also to competing size effects. Using an inverse volume dependence, however, we find results very close to those of our largest cells, which are significantly larger than in previous work. In the limit of infinite dilution and LDA+$U$, we still found negative $U$ behavior related to the strongly different relaxations of the neutral and 2+ charge states but our $|U|$ value was significantly smaller than found before, only −0.05 eV. We find the $2+/0$ transition level to lie at 0.80 eV, i.e., in the lower half of the gap and in fact lower than in LDA. This occurs because the $E_{vbm}$ actually moves up relative to the electrostatic reference level within our LDA+$U_d$+$U_s$ model. The correctness of this specific prediction of the model is difficult to ascertain directly from the bulk band structure, but agrees qualitatively though not quantitatively with the only other calculation that uses systematically a Hamiltonian which gives the correct band gap, namely the B3LYP calculation of Patterson,[11] which like our model builds in a certain degree of mixture of Hartree–Fock and LDA. In any case, however, the inspection of the one-electron bands shows a clear failure of LDA in populating the wrong level when considering the single + charge state. The results of our calculation are found to agree with the experimental data of VW,[19] when assuming the most straightforward interpretation of their data, their model (a) in which the luminescence observed to be correlated with the ODEPR is assumed to be directly resulting from the capture of an electron in the $V_O^+$ state from an effective mass donor, as well as with those of Evans *et al.*,[20] which measure essentially the inverse process. The quantitative agreement is to the same degree of accuracy as the agreement between those two experiments, i.e., a few 0.1 eV. When taken together, these results all point to a defect level at about $1.0 \pm 0.2$ eV.

## ACKNOWLEDGMENTS

This work was supported by the Air Force Office of Scientific Research. The calculations were done at the Ohio Supercomputer Center and the CWRU High-performance computing cluster.

[1] D. C. Reynolds, D. C. Look, B. Jogai, C. W. Litton, G. Cantwell, and W. C. Harsch, Phys. Rev. B **60**, 2340 (1999).
[2] W. R. L. Lambrecht, A. V. Rodina, S. Limpijumnong, B. Segall, and B. K. Meyer, Phys. Rev. B **65**, 075207 (2002).
[3] C. G. Van de Walle, Physica B **308**, 899 (2001).
[4] A. F. Kohan, G. Ceder, D. Morgan, and Chris G. Van de Walle, Phys. Rev. B **61**, 15019 (2000).
[5] S. B. Zhang, S.-H. Wei, and A. Zunger, Phys. Rev. B **63**, 075205 (2001).
[6] A. Janotti and C. G. Van de Walle, Appl. Phys. Lett. **87**, 122102

(2005).
[7] S. Lany and A. Zunger, Phys. Rev. B **72**, 035215 (2005).
[8] F. Oba, S. R. Nishitani, S. Isotani, H. Adachi, and I. Tanaka, J. Appl. Phys. **90**, 824 (2001).
[9] P. Erhart, A. Klein, and K. Albe, Phys. Rev. B **72**, 085213 (2005).
[10] P. Erhart, K. Albe, and A. Klein, Phys. Rev. B **73**, 205203 (2006).
[11] C. H. Patterson, Phys. Rev. B **74**, 144432 (2006).
[12] F. A. Kröger and H. J. Vink, J. Chem. Phys. **22**, 250 (1954).
[13] F. H. Leiter, H. R. Alves, A. Hofstaetter, D. M. Hoffmann, and B. K. Meyer, Phys. Status Solidi B **226**, R4 (2001).
[14] W. E. Carlos, E. R. Glaser, and D. C. Look, Physica B **308-310**, 971 (2001).
[15] K. Vanheusden, W. L. Warren, C. H. Seager, D. R. Tallant, J. A. Voigt, and B. E. Gnade, J. Appl. Phys. **79**, 7983 (1996).
[16] C. Solbrig, Z. Phys. **211**, 429 (1968).
[17] A. D. Becke, J. Chem. Phys. **98**, 5648 (1993).
[18] P. J. Stephens, F. J. Devlin, C. F. Chabolowski, and M. J. Frish, J. Phys. Chem. **98**, 11623 (1994).
[19] L. S. Vlasenko and G. D. Watkins, Phys. Rev. B **71**, 125210 (2005).
[20] S. M. Evans, N. C. Giles, L. E. Halliburton, and L. A. Kappers, in *ZnO and Related Materials*, MRS Fall Meeting Symposium L (Materials Research Society, Pittsburgh, 2007).
[21] M. Methfessel, M. van Schilfgaarde, and R. A. Casalii, in *Electronic Structure and Physical Properties of Solids: The Uses of the LMTO Method*, Lecture Notes in Physics, edited by H. Dreyssé (Springer, Berlin, 2000), pp. 114–147.
[22] P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964).
[23] W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).
[24] V. I. Anisimov, J. Zaanen, and O. K. Andersen, Phys. Rev. B **44**, 943 (1991).
[25] V. I. Anisimov, I. V. Solovyev, M. A. Korotin, M. T. Czyżyk, and G. A. Sawatzky, Phys. Rev. B **48**, 16929 (1993).
[26] S. L. Dudarev, G. A. Botton, S. Y. Savrasov, C. J. Humphreys, and A. P. Sutton, Phys. Rev. B **57**, 1505 (1998).
[27] A. I. Liechtenstein, V. I. Anisimov, and J. Zaanen, Phys. Rev. B **52**, R5467 (1995).
[28] L. Hedin and S. Lundqvist, in *Solid State Physics, Advances in Research and Applications*, edited by F. Seitz, D. Turnbull, and H. Ehrenreich (Academic, New York, 1969), pp. 1–181.
[29] L. Hedin, Phys. Rev. **139**, A796 (1965).
[30] E. G. Maksimov, I. I. Mazin, S. S. Yu, and Y. A. Uspenski, J. Phys.: Condens. Matter **1**, 2493 (1989).
[31] M. Usuda, N. Hamada, T. Kotani, and M. van Schilfgaarde, Phys. Rev. B **66**, 125101 (2002).
[32] R. A. Powell, W. E. Spicer, and J. C. McMenamin, Phys. Rev. Lett. **27**, 97 (1971).
[33] G. Makov and M. C. Payne, Phys. Rev. B **51**, 4014 (1995).
[34] M. Leslie and M. J. Gillan, J. Phys. C **18**, 973 (1985).
[35] D. Segev and S.-H. Wei, Phys. Rev. Lett. **91**, 126406 (2003).
[36] P. E. Blöchl, J. Chem. Phys. **103**, 7482 (1995).
[37] P. Carloni, P. E. Blöchl, and M. Parinello, J. Phys. Chem. **99**, 1338 (1995).
[38] C. W. M. Castleton, A. Hoglund, and S. Mirbt, Phys. Rev. B **73**, 035215 (2006).
[39] P. E. Blöchl, Phys. Rev. B **50**, 17953 (1994).
[40] M. G. Wardle, J. P. Goss, and P. R. Briddon, Phys. Rev. B **72**, 155108 (2005).