Generalized Koopmans density functional calculations reveal the deep acceptor state of NO in ZnO

Stephan Lany and Alex Zunger
National Renewable Energy Laboratory, Golden, Colorado 80401, USA
(Received 27 January 2010; revised manuscript received 14 April 2010; published 19 May 2010)

Applying a generalized Koopmans condition to recover the linear behavior of the energy with respect to the fractional occupation number, we find that substitutional nitrogen (NO) in ZnO is a deep acceptor with an ionization energy of 1.6 eV, which is prohibitively large for p-type conductivity. Testing the generalized Koopmans condition in computationally more demanding hybrid-functional calculations, we obtain a very similar result for NO, but find that the simultaneous correction of defect (acceptor-level) and host (band-gap) properties remains challenging in hybrid methods. The deep character of anion-site acceptors in ZnO has important consequences for the concept of codoping, as we show that nominally charge-compensated impurity pairs such as (NO-GaZn) or (CO-TiZn) have positively charged states in the gap that act as hole traps.

DOI: 10.1103/PhysRevB.81.205209 PACS number(s): 71.55.Gs, 61.72.J−, 71.15.Mb, 71.23.An

I. INTRODUCTION

Within the active research field of p-type ZnO, nitrogen impurities substituting for oxygen (NO) are widely regarded as the most promising acceptors for successful p-type doping.1–4 Indeed, shallow acceptor levels with hole-binding energies of only 100–200 meV have been observed after N-doping.4–6 Yet, assuming that these shallow states are due to substitutional NO acceptors carries the danger of the logical post hoc ergo propter hoc (“after it, hence, because of it”) fallacy: Schirra et al.7 recently found that the shallow acceptor states observed in photoluminescence are instead related to stacking faults and cannot positively be identified with NO. It also remains an outstanding experimental contradiction that the shallow acceptor energies states are difficult to reconcile with the strongly localized, Jahn-Teller distorted hole wave function.8,9

On the theoretical side, earlier calculations for the NO acceptor, based on standard density functional theory (DFT) methods, yielded a widely quoted value of 0.4 eV for the acceptor ionization energy.10 However, the local density or generalized gradient approximations (LDA or GGA) to DFT have an intrinsic bias toward producing delocalized and shallow acceptor states,11–13 which is rooted in their failure to produce the physically correct piecewise linear behavior of the energy $E(N)$ with respect to the fractional electron number $N$.14 In case of ZnO:Li, for instance, standard DFT predicts an extended effective-mass-like state and a small acceptor ionization energy of less than 0.2 eV,10,15 whereas experimentally the acceptor state is localized on only one out of four O ligands and the ionization energy is much larger, around 0.8 eV.16 In order to resolve the bias of DFT toward shallow solutions, we recently introduced a method13 that cancels the nonlinearity of $E(N)$ by adding site- and occupation-dependent potentials and matching the potential strength parameters to satisfy a generalized Koopmans condition. This method restores the correct wave-function localization and allows to quantitatively predict the acceptor ionization energies, as demonstrated recently for metal-site acceptors in transparent conducting oxides (ZnO, In$_2$O$_3$, and SnO$_2$),13 Zn vacancies in Zn chalcogenides,17 and divalent acceptors in GaN.18 Here we revisit the oxygen-site substitutional acceptor NO in ZnO using this generalized Koopmans theory, finding an ionization energy of 1.6 eV, which is prohibitively large for p-type conductivity. (Preliminary results about the NO acceptor level and a comparison to optical transitions observed in experiment were presented in Ref. 19.) We investigate the impact of deep anion-site acceptor states on concepts of codoping,20,21 where impurity pairs involving NO single acceptors or CO double acceptors have been suggested to reduce the hole binding energy in wide-gap semiconductors.22,23 We find that, after restoring the correct Koopmans behavior, the cluster dopant (2NO-GaZn) has a slightly reduced acceptor binding energy, which however is not shallow enough for $p$-doping. Moreover, our results suggest that codoping can even impede p-type conductivity, since nominally compensated impurity pairs such as (NO-GaZn) or (CO-TiZn) exhibit a deep (+/0) transition in the gap and therefore act as hole traps.

Hybrid-functional methods,24–27 which mix Fock and LDA exchange have recently been applied to acceptors in ZnO,25–27 including also the case of NO considered here.27 Since, Hartree-Fock theory exhibits the opposite $E(N)$ nonlinearity (concave) of LDA (convex), such hybrid functionals could also be used to cancel the nonlinearity of $E(N)$. Typically, however, the parameters in hybrid functionals are either taken from predefined standards28 or are adjusted to match the experimental band gap.26,27 Neither choice guarantees that the cancellation of nonlinearity is complete. For example, Bruneval28 recently showed that hybrid functionals violate the linearity condition in small Na clusters. Moreover, some previous hybrid-functional calculations of acceptor states in oxides showed deviation from experimentally established facts, either quantitatively (ZnO:Li, Ref. 25) or even qualitatively (SiO$_2$:Al, Refs. 11 and 12). Therefore, we test here the Koopmans behavior of hybrid functionals for the nitrogen acceptor NO in ZnO using the HSE functional.29 We find that for standard HSE parameters, the nonlinearity is cancelled quite well and the NO acceptor level is close to the present prediction of $E_V+1.6$ eV. However, when the parameter controlling the mixing of Fock and GGA exchange is adjusted so that the calculated ZnO band gap agrees with...
experiment, $E(N)$ becomes concave (i.e., overcorrected) and the corresponding acceptor energy of 2.1 eV is unrealistically deep.

II. METHOD

Starting from the observation that the delocalization error in DFT originates from the residual self-interaction error that places occupied and unoccupied states too close in energy, we formulated in Ref. 13 a hole-state potential of the form

$$V_{hs} = \lambda_{hs}(1 - n_{m,\sigma} n_{rel}),$$  \hspace{1cm} (1)

which is defined for those atomic orbitals where correlation effects beyond the LDA are important, i.e., O-p and N-p in the present work. $n_{m,\sigma}$ is the fractional occupancy $0 \leq n_{m,\sigma} \leq 1$ of the $m$ sublevel of spin $\sigma$ (diagonal representation) in the respective anion-p orbital. The reference occupation $n_{rel}$ is given by the anion-p occupancies in the absence of holes, as determined from the defect-free ZnO host for $V_{hs}(O-p)$ and from the ionized $N_0^-$ acceptor for $V_{hs}(N-p)$. The second parameter $\lambda_{hs}$ is determined so to satisfy the condition

$$E(n+1) - E(n) = e_i(N),$$  \hspace{1cm} (2)

where the left-hand side is the electron addition energy $E_{add}$ for occupying the empty acceptor state $e_i(N)$. After solving for $\lambda_{hs}$ in Eq. (1) to satisfy Eq. (2), there are no remaining free parameters. While Eq. (2) resembles the Koopmans theorem, we emphasize that the original Koopmans theorem applies only to Hartree-Fock theory and only if the orbital relaxation is neglected. Instead, Eq. (2), is used here as a condition that is to be fulfilled in the presence of orbital relaxation and (general) screening effects, therefore called “generalized Koopmans condition,” and it serves to restore the physically correct piecewise linearity of the energy $E(N)$ as a function of the fractional electron number.

Since the cancellation of the nonlinearity of $E(N)$ via the Koopmans condition, Eq. (2), can be achieved both through the onsite potential $V_{hs}$, Eq. (1), and by hybrid functionals, it is interesting to compare how the two approaches achieve the increase of the energy splitting between occupied and unoccupied orbitals that is needed to compensate the convexity of $E(N)$ in LDA. Using the onsite potential $V_{hs}$, this is achieved through the dependence on the occupancies $n_{m,\sigma}$ of the $lm$ decomposed local (on site) charge symmetry, similar as in DFT+$U$.\textsuperscript{31} (Contrary to DFT+$U$, however, $V_{hs}$ is defined such that it vanishes when all symmetries are equally occupied, such as in case of the defect-free ZnO host or the ionized $N_0^-$ acceptor).\textsuperscript{35} In hybrid DFT, on the other hand, the increased energy splitting between occupied and unoccupied states is achieved through inclusion of nonlocal $V_{hs}$ (and $r'$ dependent) Fock exchange

$$V_{hs}^{nl}(r, r') = -\alpha \sum_i \frac{\psi_i^*(r') \psi_i(r)}{|r - r'|} f(|r - r'|),$$  \hspace{1cm} (3)

which is scaled by a parameter $\alpha \leq 1$ and, depending on the formulation of the hybrid functional, by an additional attenuation function $f(|r - r'|)$. Thus, by comparing the two approaches, our present work also addresses the question in how far charge-symmetry dependent potentials such as $V_{hs}$ can capture effects that otherwise would require the much more computing intensive calculation of the nonlocal Fock exchange.

The calculations employing the on-site potential $V_{hs}$, Eq. (1), are performed using the projector-augmented wave\textsuperscript{33} implementation of the VASP code\textsuperscript{34} and the GGA parameterization of Ref. 35. Since the valence-band structure of ZnO is considerably affected by the shallow Zn- inner potential, we additionally employ the GGA+$U$ method\textsuperscript{31} ($U-J=6$ eV) for Zn-$d$. The parameters $\lambda_{hs}$ are obtained as 4.5 eV for O-$p$,\textsuperscript{17} 4.3 eV for N-$p$, and 3.6 eV for C-$p$. Supercell finite-size effects were accounted for as described in Ref. 36. The Brillouin zone of the 72-atom supercell of ZnO was sampled by $2 \times 2 \times 2$ $k$ mesh and we used relatively soft pseudopotentials (300 eV cutoff) that have been tested in Ref. 36. For the hybrid-functional calculations of the N acceptor, we employed the implementation of nonlocal exchange of Ref. 37 and the range-separated HSE functional.\textsuperscript{29} Since the band gap varies between the different employed functionals (e.g., with different fractions $\alpha$ in HSE), all transition levels are calculated with respect to the valence-band maximum (VBM). In so far as the distances of defect states from the conduction-band minimum (CBM) are concerned, the experimental value of 3.44 eV for the band gap is used.

III. $N_0$ ACCEPTOR IN ZNO IN KOOPMANS-CORRECTED DFT

In Fig. 1, we compare the localization of the $N_0$ acceptor wave function in standard DFT (here, GGA+$U$ and after application of the hole-state potential $V_{hs}$, which was constructed so to satisfy the generalized Koopmans condition, Eq. (2). Whereas in standard DFT, the acceptor state has a $p_{3\pi}$ character at the N site ($e_g$ symmetry representation of the C$_{3\pi}$ point group) and considerable contributions from next-nearest-neighbor O sites, it becomes largely localized within a single N-$p$, orbital ($a_1$ symmetry) when applying the onsite potential $V_{hs}$, which is in accord with the electron paramagnetic resonance (EPR) characterization of $N_0$.\textsuperscript{8,9} The strongly directional hole localization of the neutral acceptor $N_0^0$ is accompanied by a pronounced asymmetry of the nearest-neighbor (NN) distances, i.e., a Jahn-Teller relaxation, which is not present in the ionized $N_0^-$ acceptor state.
TABLE I. Calculated properties of the NO acceptor in ZnO. The N-Zn nearest-neighbor distances $d_i$ and $d_o$, parallel and perpendicular to the wurtzite $c$ axis (cf. Fig. 1), the acceptor ionization energy $e_i$, and the optical (vertical) transition energies $e_{opt}$ (relative to the VBM).

<table>
<thead>
<tr>
<th>Property</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d_i/d_o$ (NO)</td>
<td>2.18/1.94 Å</td>
</tr>
<tr>
<td>$e_i(0/1\to)$</td>
<td>1.62 eV</td>
</tr>
<tr>
<td>$e_{opt}$:NO$^0\to$NO$^+$</td>
<td>2.47 eV</td>
</tr>
<tr>
<td>$e_{opt}$:NO$^0\to$NO$^-1$</td>
<td>0.85 eV</td>
</tr>
</tbody>
</table>

(see Table I) or in the standard DFT calculation (cf. Fig. 1). Apart from the axial configuration shown in Fig. 1(a), where the N-p orbital hosting the hole is aligned parallel to the crystal $c$ axis, there exist also nonaxial configurations of NO, in which the spin density is oriented along one of the three bonds with neighbors in the basal plane. We find that the nonaxial configuration lies 0.03 eV higher in energy than the axial geometry.

Figure 2(a) shows the calculated vertical transition energies between the neutral ($q=0$) and ionized ($q=-1$) states of NO, e.g.,

$$N_O^0 = N_O^0 + e,$$

for the transitions into or from the free-electron ($e$) state at the CBM. It was found experimentally in Ref. 38 that the EPR active neutral NO acceptor can be photoexcited by light below the band gap with energies down to 2.4 eV (514 nm). Considering that the N acceptor existed in the EPR-inactive ionized NO state before illumination, the most straightforward interpretation is that the neutral N acceptor is created due to the optical-absorption process, Eq. (4) (forward direction), i.e., the excitation from the occupied acceptor state into the CBM. Our calculated energy of 2.6 eV for this vertical transition (Fig. 2(a)) agrees well with the experimental threshold energy of the N-related absorption band observed in Ref. 38. On the other hand, for the luminescence process, Eq. (4) (backward direction), i.e., the recombination of a free electron into the empty state of NO$^0$, we obtain an energy of only 1.0 eV (Fig. 2), much smaller than the experimentally observed free-to-bound ($e$,$A^0$) transition at 3.31 eV. Thus, we cannot identify this luminescence with substitutional NO acceptors.

The large atomic relaxations involved with a change between the neutral and ionized acceptor states (Table I) lead to a large difference between the (optical) vertical and thermal acceptor transition energies. Specifically, the vertical ($0/1\to$) transitions (Fig. 2) straddle the acceptor level $e_i(0/1\to)=E_V+1.62$ eV (Table I). The Stokes shift of 1.6 eV is composed of approximately equal contributions $E_{rel}=0.8$ eV from the relaxation energy (Jahn-Teller energy) after the vertical $0\to1$ and $1\to0$ transitions (Table I). We note that relaxation energies around 0.8 eV are rather typical for deep hole states in ZnO. Similar values are calculated also for, e.g., Li$^+_2Z$o (Ref. 13) or Zn vacancies. The thermal ionization energy of 1.6 eV is much larger than common perceptions about NO (Ref. 3) and precludes the thermal activation of free holes. We conclude, therefore, that substitutional NO dopants will not lead to $p$-type conductivity in ZnO.

IV. HYBRID-FUNCTIONAL RESULTS FOR NO

In order to compare the theoretical description of the NO acceptor in ZnO by the symmetry dependent onsite potential $V_{hs}$, Eq. (1), to the description obtained by including the explicit nonlocality of Fock exchange, Eq. (3), we further performed hybrid-functional calculations for NO. Similar to the on-site potential $V_{hs}$, also hybrid functionals can qualitatively correct the localization of oxygen hole states. Typical hybrid-functional parameters are, however, motivated by molecular cohesive energies and by considerations of computational efficiency, and do not necessarily recover the linearity of $E(N)$, raising the question which parameters are appropriate for the prediction of acceptor binding energies. Using the HSE hybrid functional (with a range separation parameter $\alpha=0.25$, for which, however, the band gap of ZnO is still underestimated by about 1 eV, and for an increased fraction $\alpha=0.38$ which restores the experimental band gap). Regarding structural properties and wavefunction localization, we find that the NN distances agree within 0.02 Å with the values in Table I, irrespective of the choice of the parameter $\alpha$, and the localization of the NO acceptor state is essentially identical to that shown in Fig. 1(b). Thus, remarkably, we find that the electronic wave functions and the resulting structural properties obtained by means of the onsite potential $V_{hs}$ and by inclusion of nonlocal Fock exchange are practically identical.

Considering the NO acceptor level, however, we observe a considerable dependence on the parameter $\alpha$, obtaining $E_i(0/1\to)=E_V+1.44$ and 2.10 eV with $\alpha=0.25$ and 0.38, respectively. In order to evaluate the linearity of $E(N)$ in the HSE functional, we now determine the non-Koopmans energy $\Delta_{\alpha}=E_{\text{rel}}-e_i(N)$ [see Eq. (2)], obtaining $\Delta_{\alpha}=-0.05$ and $-0.40$ eV for $\alpha=0.25$ and 0.38, respectively. For
comparison, the non-Koopmans behavior of standard DFT is 
\[ \Delta_{\text{OK}} = +0.62 \text{ eV}. \]
Thus, we find that the generalized Koopmans condition is quite well fulfilled in HSE with the standard parameter \( \alpha = 0.25 \). The respective \( \text{N}_0 \) acceptor ionization energy of 1.44 eV is consistent with the result obtained with the \( V_{\text{hs}} \) on-site potential (see Table I). The slightly smaller value can be attributed to the fact that in HSE (with \( \alpha = 0.25 \)), the Zn-\( d \) orbitals lie 0.8 eV higher in energy compared to GGA+\( U \), thereby pushing the VBM closer to the \( \text{N}_0 \) acceptor level. The relaxation energy, e.g., following the vertical (0/1−) transition of 0.84 eV is practically identical to that obtained with \( V_{\text{hs}} \) (cf. Table I). On the other hand, the band-gap corrected HSE calculation with \( \alpha = 0.38 \) causes overcorrection (negative sign of \( \Delta_{\text{OK}} \)) and the respective acceptor level at 2.10 eV is probably unrealistically deep. Quite generally, one should not expect that increasing \( \alpha \) to match the experimental band gap will yield automatically accurate results also for defect levels.

Since the linearity condition can be matched both on-site potentials and by hybrid functionals, we now briefly discuss the implications of either choice. The addition of a fraction of nonlocal Fock exchange in hybrid-functional calculations treats all atoms uniformly and generally improves the band gap and the overall band structure. However, as shown above, one cannot expect in general that the linearity condition and the experimental gap are matched simultaneously for a single choice of the hybrid parameters. In particular, the homogeneous screening of the Fock exchange, common to all hybrid functionals (cf. Eq. (3)), has been identified as a serious simplification,\(^{42} \) which may impede the simultaneously accurate description of different properties, such as the band structure and defect states. By restoring the Koopmans condition for the respective atomic orbitals individually (here, the \( p \) orbitals of O, N, and C), the onsite potential \( V_{\text{hs}} \) method can better account for differences in the local screening behavior. On the other hand, \( V_{\text{hs}} \) is constructed [see Eq. (1)] such to vanish in the unperturbed bulk material and therefore does not improve the bulk band structure.

**V. IMPACT OF DEEP N-ACCEPTOR STATES ON CODOPING CONCEPTS**

Codoping of \( \text{N}_0 \) with group III donors (e.g., Al, Ga, In) has been proposed as a promising route to reduce the acceptor ionization energy and to increase the solubility of \( \text{N}_0 \) acceptors.\(^{20,21} \) Since such proposals were evaluated on the basis of standard LDA or GGA calculations, we now assess the ramifications of the deep \( \text{N}_0 \) state on the concept of codoping. In line with the expectation\(^{20} \) that the formation of an acceptorlike (2\( \text{N}_0 \)-\( \text{Ga}_{2\text{z}} \)) impurity cluster would have a shallower state than the isolated \( \text{N}_0 \) acceptor, we find here a hole-binding energy of 1.34 eV for this cluster in the present approach, which is indeed significantly smaller than that of isolated \( \text{N}_0 \) (cf. Table I). However, even the acceptor level of the cluster is still much too deep to release free holes by thermal excitation.

Again based on standard DFT functionals, it has been suggested\(^{22,23} \) that the acceptor ionization energy can be further lowered through formation of an impurity band (IB) due to charge-compensated donor-acceptor (D-A) pairs containing \( \text{N}_0 \) (Ref. 22) or substitutional carbon (\( \text{C}_0 \)).\(^{22,23} \) Such an IB effectively shifts the VBM closer to the acceptor level, thereby reducing the ionization energy. This concept requires that the impurity pairs must be inert and not introduce gap states themselves. Using the present approach, we find, however, that the pairs (\( \text{N}_0 \)-\( \text{Ga}_{2\text{z}} \), (\( \text{C}_0 \)-\( \text{Ti}_{2\text{z}} \)), and (\( \text{C}_0 \)-\( \text{Zr}_{2\text{z}} \)) do not only exist in their compensated, charge-neutral state, but have also a stable positively charged states, e.g., (\( \text{N}_0 \)-\( \text{Ga}_{2\text{z}} \))\(^{+} \), formed by trapping of a hole in a localized state similar to that of the isolated neutral \( \text{N}_0 \) acceptor (see Fig. 1). These positively charged states are a consequence of the strong tendency of nitrogen and, even more so, of carbon to localize holes in deep states. As shown in Fig. 3, the (0/+0) transition levels of the impurity pairs lie well above the respective IB formed by the charge-neutral pairs. This indicates that holes localize in the deep level [cf. Fig. 1(b)] and become immobile, rather than providing \( p \)-type conduction through the IB. Thus, our results obtained by generalized Koopmans DFT calculations suggest that nominally charge-compensated impurity pairs in wide-gap semiconductors have a much larger tendency to create deep gap states than it is generally recognized.

**VI. CONCLUSIONS**

The physical condition of the piecewise linearity of the total energy \( E(N) \) with respect to the occupation number plays a decisive role in determining the correct acceptor ionization energies in wide-gap oxide semiconductors. Matching this condition by means of an on-site potential \( V_{\text{hs}} \) that depends on the local charge symmetry, the predicted acceptor level of substitutional \( \text{N}_0 \) dopants in \( \text{ZnO} \) lies as deep as 1.62 eV above the valence-band edge. Testing the Koopmans behavior of \( \text{N}_0 \) in HSE hybrid-functional calculations, we find good cancellation of the nonlinearity of \( E(N) \) with the standard mixing parameter \( \alpha = 0.25 \) for the Fock exchange but increasing the parameter \( \alpha \) in HSE calculation to match the experimental band gap of \( \text{ZnO} \) overcorrects the linearity condition. The hybrid-functional description of the acceptor wave functions, the structural Jahn-Teller distortion, and the acceptor level (for \( \alpha = 0.25 \)) agree well with the results ob-
tained with $V_{\text{hs}}$ indicating that on-site potentials can largely account for the effects of nonlocal Fock exchange in localized defect states. The deep levels of the anion-site acceptors NO and CO have important implications on codoping concepts, as we find that nominally charge-compensated donor-acceptor pairs can form positively charged states and thereby trap holes into immobile deep states.

30. T. C. Koopmans, Physica 1, 104 (1934).
32. The asymmetry between the occupancies of $p_{\alpha z}$ and $p_{\alpha y}$ orbitals due to the trigonal symmetry of wurtzite ZnO is practically negligible.
41. An independent HSE study (Ref. 27) recently found an acceptor level of only 1.3 eV for N$_{\alpha}$ in ZnO, which, considering the value $\alpha=0.36$ used in that work, should be compared to our present result of $2.1$ eV for $\alpha=0.38$. The reason for this discrepancy is yet unclear.
42. Note that a critical value of $V_{\text{hs}}$ is usually needed to restore the correct wave-function symmetry (Ref. 13), so $\Delta_{\text{hs}}$ is determined as the linearly extrapolated $V_{\text{hs}} \rightarrow 0$ limit rather than the actual value for $V_{\text{hs}}=0$.