

Intrinsic n -type versus p -type doping asymmetry and the defect physics of ZnO

S. B. Zhang, S.-H. Wei, and Alex Zunger

National Renewable Energy Laboratory, Golden, Colorado 80401

(Received 9 March 2000; published 31 January 2001)

ZnO typifies a class of materials that can be doped via native defects in only one way: either n type or p type. We explain this asymmetry in ZnO via a study of its intrinsic defect physics, including Zn_O , Zn_i , V_O , O_i , and V_{Zn} and n -type impurity dopants, Al and F. We find that ZnO is n type at Zn-rich conditions. This is because (i) the Zn interstitial, Zn_i , is a *shallow* donor, supplying electrons; (ii) its formation enthalpy is *low* for both Zn-rich and O-rich conditions, so this defect is abundant; and (iii) the native defects that could compensate the n -type doping effect of Zn_i (interstitial O, O_i , and Zn vacancy, V_{Zn}), have *high* formation enthalpies for Zn-rich conditions, so these “electron killers” are not abundant. We find that ZnO cannot be doped p type via native defects (O_i , V_{Zn}) despite the fact that they are shallow (i.e., supplying holes at room temperature). This is because at both Zn-rich and O-rich conditions, the defects that could compensate p -type doping (V_O , Zn_i , Zn_O) have *low* formation enthalpies so these “hole killers” form readily. Furthermore, we identify electron-hole radiative recombination at the V_O center as the source of the green luminescence. In contrast, a large structural relaxation of the same center upon double hole capture leads to slow electron-hole recombination (either radiative or nonradiative) responsible for the slow decay of photoconductivity.

DOI: 10.1103/PhysRevB.63.075205

PACS number(s): 61.72.Bb, 61.72.Ji, 61.82.Fk

I. INTRODUCTION

Most materials exhibit an asymmetry in their ability to be doped n type or p type. For example, ZnS is a good n -type conductor but cannot be made p type,¹ and CuGaSe₂ is a good p -type conductor, but cannot be made n type through intrinsic doping.¹ A paradigm system in this respect is ZnO, which can be doped n type via intrinsic^{2–8} or extrinsic⁹ dopants to the point of becoming a metallic conductor, whereas it cannot be made p type via intrinsic doping, although recent reports suggest that p doping is possible with nitrogen.^{10,11} The defect physics of ZnO is quite complex and to a large extent unknown. For example, we know experimentally that unintentionally doped ZnO is n type but whether the donor is zinc interstitial (Zn_i) or oxygen vacancy (V_O) is still controversial. ZnO has a 3.36-eV direct band gap¹² and a green luminescence at 2.4 eV.^{12–16} Some have attributed both the subgap green luminescence, as well as the shallow dopants, to the oxygen vacancy, while others to other defects such as the F center. Adding to the controversy, positron-annihilation experiments identified¹⁷ the Zn vacancy as the dominant defect in ZnO, instead. Recently, very slow decay of photoconductivity in ZnO was also reported.¹⁸ To date, theoretical studies of the defects in ZnO are scarce. Most of the analysis still relies on the phenomenological model originated by Kroger² in 1964 (see also Mackrodt *et al.*¹⁹). Only recently, extrinsic doping of ZnO was studied²⁰ by first-principles calculation. However, a rather crude augmented spherical wave formalism was used in which the anisotropic covalent bonding is not accurately described.

In this paper we study the microscopic equilibrium mechanisms that explain the intrinsic doping asymmetry in ZnO. We consider both doping by native defects such as Zn_i , V_O , Zn_O , V_{Zn} , and O_i as well as n doping via impurities Al and F. We find the following:

(a) ZnO is n type at Zn-rich conditions. This is because (i) the zinc interstitial Zn_i is a *shallow* level, supplying elec-

trons; (ii) its formation enthalpy is *low* for both Zn-rich and O-rich conditions, so this defect is abundant; and (iii) native defects that could compensate the n -type doping effect of Zn_i (O_i or V_{Zn}) have *high* formation enthalpies at the Zn-rich conditions so these “electron killers” are rare.

(b) ZnO cannot be doped p type via native defects (O_i , V_{Zn}) despite the fact that they are shallow (i.e., supplying holes at room temperature). This is because at both Zn-rich and O-rich conditions, the defects that could compensate p -type doping (V_O , Zn_i , Zn_O) have *low* formation enthalpies, so these “hole killers” form readily.

(c) By studying the oxygen vacancy V_O , we identified the electron-hole radiative recombination at the V_O center as the source of the green luminescence in ZnO. A large structural relaxation upon capture of two holes by the same center, on the other hand, gives rise to slow electron-hole recombination (either radiative or nonradiative) responsible for the observed slow decay of photoconductivity.

II. FORMATION ENTHALPIES AND TRANSITION ENERGIES

We imagine ZnO to be in equilibrium with a reservoir of Zn and O. The formation enthalpy of a structural defect α of charge q is²¹

$$\Delta H^{(q,\alpha)} = \Delta E^{(q,\alpha)} + n_a \mu_a + q E_F, \quad (1)$$

where

$$\Delta E^{(q,\alpha)} = E^{(q,\alpha)}(\text{defect+host}) - E(\text{host only}) + n_a \mu_a(\text{solid}) + q E_V. \quad (2)$$

Here, $E^{(q,\alpha)}(\text{defect+host})$ is the total energy of a cell including the host material as well as defect α in charge state q , and $E(\text{host only})$ is the total energy of the cell containing just the host. E_F is the Fermi energy and E_V is the valence-band maximum (VBM) of the host crystal. n_a is the number of

atoms being removed during the defect formation from the host crystal to the atomic reservoir. For example, $n_a = n_{\text{Zn}} = 1$ for the Zn vacancy and $n_a = n_{\text{O}} = 1$ for the O vacancy. μ_a is the chemical potential of the reservoir, and μ_a (solid) is the energy of elemental solid a . Under thermal equilibrium $\mu_{\text{Zn}} + \mu_{\text{O}} = \Delta H_{\text{ZnO}}$, where ΔH_{ZnO} is the formation enthalpy of ZnO, so $\mu_{\text{O}} = \Delta H_{\text{ZnO}} - \mu_{\text{Zn}}$. Therefore, for ZnO, $n_{\text{O}} = 1$ and $n_{\text{Zn}} = -1$ is equivalent to $n_{\text{Zn}} = -2$. Depending on the growth condition, μ_{Zn} could vary in the energy range

$$\Delta H_{\text{ZnO}} \leq \mu_{\text{Zn}} \leq 0. \quad (3)$$

Our calculation gives $\Delta H_{\text{ZnO}} = -3.1$ eV per ZnO (with respect to hcp Zn and molecular O₂); the experimental value is -3.6 eV. The role of μ_a in Eq. (1) is to control the flow of atoms between the atomic reservoirs and the host crystal. For example, as μ_{Zn} increases, it becomes increasingly difficult to remove a Zn atom from the bulk ZnO to the Zn reservoir. The last term in Eq. (1) is the energy required to remove q electrons from the defect, placing them (in analogy with the atomic reservoir) in the electron reservoir, i.e., the Fermi energy of bulk ZnO.

While $\Delta H^{(\alpha,q)}(\mu, E_F)$ tells us how much energy it takes to form defect (α, q) , once formed, we may want to know the energy it takes to alter its charge state. This ‘‘defect transition energy’’ is defined as

$$\epsilon(q/q') = [\Delta E(q, \alpha) - \Delta E(q', \alpha)] / (q' - q), \quad (4)$$

where q and q' are two different charge states of the defect α , e.g., $q=1$ and $q'=0$ for the $(+/0)$ donor transition, $q=0$ and $q'=-1$ for the $(0/-)$ acceptor transition. If $E_F < E(q/q')$ then the defect is in its higher (i.e., more positive) charge state q , whereas if $E_F > E(q/q')$ then the defect switches to the lower (less positive) charge state q' .

III. DETAILS OF THE CALCULATION

The electronic properties of wurtzite ZnO and its defects are calculated by the plane-wave pseudopotential total-energy and force method²² using the local density approximation²³ (LDA). We use the *ab initio* pseudopotentials by Troullier and Martins²⁴ and the Ceperley-Alder exchange correlation,²⁵ as parametrized by Perdew and Zunger.²⁶ The plane-wave basis-set cutoff energy is 60 Ry. We have tested the pseudopotentials with cutoff energies up to 100 Ry. The calculated band dispersions along Γ - P , Γ - N , and Γ - H directions at 100-Ry cutoff are to within 0.1 eV of the all-electron linearized augmented plane-wave (LAPW) results, having a direct band gap of 0.8 eV. At the 60-Ry cutoff, however, the pseudopotential gap is reduced to 0.6 eV. The experimental value is 3.36 eV.¹² The large discrepancy is due to the LDA gap error. A symmetric 36-atom wurtzite supercell was constructed to model isolated point defect with four k points sampling in the irreducible Brillouin zone of the supercell. The supercell wurtzite lattice vectors are $(2\bar{1}\bar{3}1)$, $(\bar{1}101)$, and $(\bar{1}231)$, respectively, in units of the primitive cell vectors with an included angle of 120° between \mathbf{a}_1 and \mathbf{a}_2 . For charged defects, a jellium background charge was used and the background error was cor-

TABLE I. Defect formation energies (in eV) at the Zn-rich limit ($\mu_{\text{Zn}}=0$) and p -type ($\epsilon_F = \epsilon_{\text{VBM}}$) ZnO. n_a is the number of atoms exchanged with the reservoir during the formation of the defect, and q is its charge. We give both LDA and LDA-corrected formation energies. To obtain the values for the O-rich limit, add $n_a \Delta H_{\text{ZnO}}$ (where $\Delta H_{\text{ZnO}} = -3.1$ eV) to the values here. To obtain the values for $\epsilon_F > \epsilon_{\text{VBM}}$, add $q\epsilon_F$ to the energies, instead.

Defect	n_a	q	ΔE_{LDA}	$\Delta E_{\text{LDA-corr}}$
Donorlike defects				
V_{O}	-1	+2	-0.5	-3.0
		+1	0.8	1.5
		0	1.5	2.4
Zn_i	-1	+2	-0.2	-2.3
		+1	1.5	2.1
		0	3.4	6.2
Zn_{O}	-2	+2	0.2	0.4
		+1	2.0	5.2
		0	4.1	9.6
Acceptorlike defects				
V_{Zn}	+1	0	5.8	10.6
		-1	5.7	10.1
		-2	5.8	10.1
O_i	+1	0	6.2	9.7
		-1	6.4	10.4
		-2	7.4	12.1

rected to $O(L^{-3})$, where L is the dimension of the supercell.²⁷ We relax all the atoms to their equilibrium positions by force minimization. For wurtzite ZnO this gives $a = 3.25$ Å, $c/a = 1.60$, and $u = 0.375$. The error bar for the LDA calculations of ΔH is less than 0.1 eV.

IV. THE LDA RESULTS AND THE NEED FOR LDA CORRECTIONS

Table I lists the LDA defect formation enthalpies ΔH_{LDA} of Eq. (1), in terms of n_a , q , and ΔE_{LDA} , for the oxygen vacancy (V_{O}), zinc interstitial (Zn_i), Zn-on-O antisite (Zn_{O}), zinc vacancy (V_{Zn}), and oxygen interstitial (O_i). As we noted in Sec. II, the defect formation enthalpies are functions of the Zn atomic chemical potential (μ_{Zn}) and the Fermi energy (E_F). Table I gives the energies at $\mu_{\text{Zn}}=0$ and $E_F = E_V$. Table II lists the corresponding LDA defect transition energies $E(q/q')$ of Eq. (4). Figure 1 shows as lines the defect formation enthalpies as a function of the Fermi energy at the zinc- and oxygen-rich limits, respectively. Defect transition energies are shown as solid dots. One observes the following in Fig. 1:

(i) Most of the defects are shallow with transition energies either outside or near the LDA band edges $E_V=0$, $E_C = 0.6$ eV. The only exception is the oxygen interstitial where the acceptor levels are at $E_V+0.2$ and $E_V+1.0$. The oxygen vacancy is a ‘‘negative- U ’’ center: $E(2+/+) > E(+/0)$, thus under thermal equilibrium the $(2+/+)$ and $(+/0)$ transitions are experimentally inaccessible.

(ii) There is clear asymmetry in the formation enthalpies

TABLE II. Defect transition energies in eV. The donor levels are given with respect to the conduction-band minimum (E_C) while the acceptor levels are given with respect to the valence-band maximum (E_V). LDA donor levels with extrapolated band gap to the experimental value are given in parentheses.

	Donors (with respect to E_C)			Acceptors (with respect to E_V)		
	(2+/1+)	(1+/0)	(2+/0)	(0/1-)	(1-/2-)	(0/2-)
	LDA values					
V_O	0.9	0.1	0.4 (-2.3)			
Zn_i	1.1 (-1.6)	1.3(-1.4)	1.2			
Zn_O	1.2 (-1.5)	1.5(-1.2)	1.4			
V_{Zn}				-0.4	0.1	0.0
O_i				0.2	1.0	0.6
	LDA-corrected values					
V_O	1.2	-2.4	-0.6			
Zn_i	1.1	0.8	1.0			0
Zn_O	1.5	1.1	1.3			
V_{Zn}				-0.5	0.0	-0.3
O_i				0.7	1.7	1.2

for the donorlike and acceptorlike defects. Consider E_F at LDA midgap. In the zinc-rich limit, the energies of the donors Zn_i and V_O are 4 to 6 eV lower than the acceptors V_{Zn} and O_i . In the oxygen-rich limit, the energy of the low-lying

donors (Zn_i and V_O) is comparable to that of the acceptors (V_{Zn} and O_i).

Although total energies are ground-state properties and thus protected by the Hohenberg-Kohn theorem, the defect formation enthalpies $\Delta H(E_F, \mu)$ depend on the LDA band gap, since $E_V \leq E_F \leq E_C$, where the LDA gap is $E_C - E_V$. This leads to LDA errors in our enthalpies: The limits on E_F used in Fig. 1 are $E_V = 0$ and the LDA $E_C = 0.6$ eV, not the experimental value $E_C = 3.36$ eV. The reason is that extrapolation of the ΔH_{LDA} value to $E_F = E_C = 3.36$ eV would yield too low acceptor energies. For example, the formation enthalpy of the zinc vacancy for $E_F = E_C$ would be -0.8 eV at the zinc-rich or -3.9 eV at the oxygen-rich conditions. This would imply that ZnO could never be doped n type (since the Zn vacancy will act as electron killer), in sharp contradiction to experiments.²⁻⁸ Moreover from Table II, the zinc interstitial (2+/1+) donor level would be $E_C - 1.5$ eV, also in contradiction to experiment that finds this level at $E_C - 0.03$ eV.⁸ These results suggest that we need to systematically correct the LDA formation enthalpies due to the LDA band gap error. This is done in the Appendix.^{23,26,28-31}

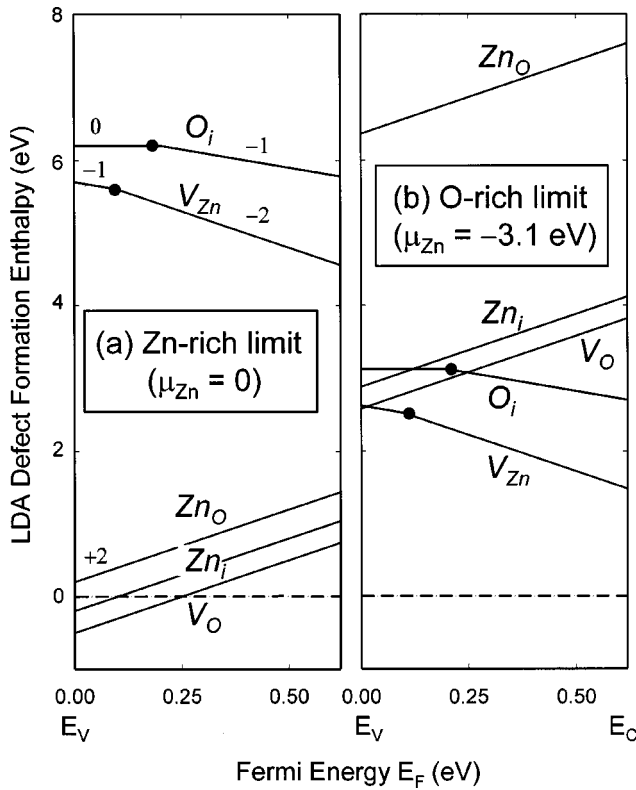


FIG. 1. LDA defect formation enthalpies are shown in lines as a function of the Fermi energy E_F at the zinc-rich ($\mu_{Zn} = 0$) and oxygen-rich ($\mu_{Zn} = -3.1$ eV) limits, respectively. The charge states of the defects are shown as -2, -1, 0, and +2. Defect transition energies are shown as solid dots.

V. DOPING ASYMMETRY IN ZnO

The conditions for n -type doping via native defects are the following.

(1) Donors (e.g., V_O , Zn_i , Zn_O) must have shallow levels, $E(+/0)$, $E(2+/+)$, or $E(2+/0)$, with respect to the conduction-band minimum (CBM), so that they readily produce electrons.

(2) Donors must have *low* formation enthalpy ΔH even if E_F is high in the gap, so that such donors become abundant.

(3) Electron-killer centers (e.g., O_i , V_{Zn}) must have *high* formation enthalpy even if E_F is high in the gap, so that they do not form.

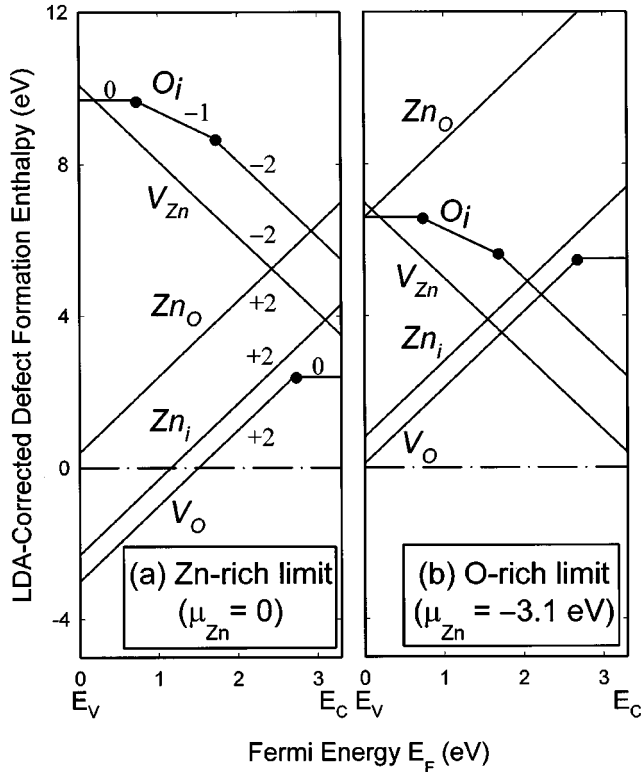


FIG. 2. Defect formation enthalpies after LDA corrections (by the $E1$ method in the Appendix) are shown as lines. The legends are the same as in Fig. 1.

The conditions for p -type doping via native defects are the following.

(a) Acceptors (e.g., O_i, V_{Zn}) must have shallow levels ($0/-$), ($-/2-$), or ($0/2-$) with respect to the VBM, so that they readily produce holes.

(b) Acceptors must have *low* formation enthalpy ΔH even if E_F is low in the gap, so that such acceptors become abundant.

(c) Hole-killer centers (e.g., V_O, Zn_i, Zn_O) must have *high* formation enthalpy even if E_F is low in the gap, so that they do not form.

Conditions (2) and (3), as well as (b) and (c) depend on the chemical potential μ so they have to be examined separately for cation-rich and anion-rich conditions. Figure 2 shows the defect formation enthalpies (solid lines) and defect transition energies (solid dots) after LDA corrections. We will discuss them in light of the above conditions.

A. Conditions for intrinsic n -type behavior

(1) We find $V_O, Zn_i,$ and Zn_O to be the negative- U defects with the ($2+/0$) donor levels at $E_C - 0.6, E_C + 1.0,$ and $E_C + 1.3$ eV, respectively. Any donor levels above E_C will ionize spontaneously, transferring the electrons to defect levels near E_C . The zinc interstitial is thus a shallow donor (regardless of the LDA corrections), in good agreement with the recent experimental results where a shallow donor level located at 30 meV below the CBM was attributed to the zinc

interstitials.⁸ We see from Fig. 2 that condition (1) is satisfied for Zn_i and Zn_O but not for V_O that is deep. This will be further discussed in Sec. VI.

(2) We see from Fig. 2 that this is true for V_O and Zn_i in Zn-rich condition.

(3) We see from Fig. 2 that this is approximately true for Zn-rich but not for O-rich conditions.

We thus conclude that intrinsic n -type doping of ZnO is possible in Zn-rich conditions via Zn interstitials.

The explanation of (2) and (3) above is as follows: The low formation enthalpy of the oxygen vacancy is in part a result of the strong binding of the O_2 molecule. The low formation enthalpy of the zinc interstitial, on the other hand, results from the relatively large ionicity of ZnO. In Si or Ge, there is no difference in the formation enthalpy between interstitials since they are identical. As the ionicity increases, it becomes easier to form cation interstitials rather than anion interstitials, and this disparity increases with ionicity. For example, it costs 3.3 eV to form a neutral Zn_i and 7.7 eV to form Se_i [in Zn-rich ZnSe (Ref. 32)]. Even in the Se-rich condition, the formation enthalpy of Zn_i (4.7 eV) is still smaller than Se_i (5.4 eV).³² In the extreme case of CuCl, it is very easy to form cation interstitials, as evidenced by the low formation enthalpy of cation interstitial-vacancy pair.³³

B. n -type doping with Al and F

It is known that ZnO can be doped n -type by extrinsic dopants⁹ such as Al on the Zn site or F on the O site. Our LDA calculations show that the single-particle energy levels at Γ are just below the CBM, i.e., $E_C - 0.05$ eV for Al_{Zn} and $E_C - 0.08$ eV for F_O . The ($+/0$) transition levels for Al and F are, on the other hand, $E_C - 0.12$ eV and $E_C - 0.08$ eV, respectively. Thus, F_O with a shallower donor level is a better n -type dopant than Al_{Zn} , in qualitative agreement with experiment.⁹ These (conduction-band-derived) donor levels remain relatively shallow upon gap corrections. The formation enthalpies of Al and F are -5.8 and -2.8 eV, respectively at the Zn-rich limit, provided that the chemical potentials of Al and F are at the respective elemental solid (gas) values.

C. Conditions for intrinsic p -type behavior

(a) We find that $E(-/2-) = E_V + 0.0$ eV for V_{Zn} and $E(0/-) = E_V + 0.7$ eV and $E(-/2-) = E_V + 1.7$ eV for O_i . The zinc vacancy is thus a shallow acceptor but not the oxygen interstitial.

(b) We see in Fig. 2 that this is true only for O-rich, but not for Zn-rich conditions.

(c) We see from Fig. 2, that this condition does not hold both in the Zn-rich and O-rich conditions.

The (b) and (c) above have the same physical origin as (2) and (3) in Sec. V A.

We conclude that ZnO cannot be doped p type under thermal equilibrium because (i) at both Zn-rich and O-rich conditions, the formation enthalpies of the hole killers $V_O, Zn_i,$ and Zn_O are low; and (ii) in Zn-rich conditions, the hole-producing acceptors are difficult to form.

TABLE III. Calculated single-particle defect energy level, $\varepsilon(V_O)$, of *neutral* oxygen vacancy with respect to the CBM. The LDA* refers to altering the band gap by nonrelativistic calculation and then extrapolating the LDA $\varepsilon(V_O)$ along with the band gap. The GW model is given in Ref. 34 and the SIC calculation follows Ref. 29.

	LDA	$E1$	LDA*	GW model	SIC
$E_{\text{CBM}} - E(V_O)$	-0.1	1.3	1.8	1.0	1.0

Recently, it has been reported that N can be used as dopant to effectively dope ZnO *p* type.^{10,11} The success of these studies demonstrated the importance of the understanding of the hole-killer defects discussed here. In one case,¹¹ Ga₂O₃ was intentionally used to reach the O-rich condition, thus suppressing the formation of oxygen vacancy and zinc interstitial. In the other case,¹⁰ the dopant N is incorporated into ZnO as a nitrogen-hydrogen complex. Because (N+H) as a whole, has valence six, identical to oxygen, a very high N level can thus be reached without the formation of hole-killer defects. Hydrogen atoms are subsequently driven out of the sample by control of the growth conditions.¹⁰ In either case, formation of the hole-killer defects are purposely avoided by carefully avoiding thermal equilibrium between the active dopant and the intrinsic defects.

VI. ENERGY LEVEL OF OXYGEN VACANCY

In the LDA calculation, the neutral oxygen vacancy produces a level of a_1 -like symmetry with occupancy of two electrons, at the energy $E_C + 0.1$ eV, i.e., inside the conduction band. Thus, in LDA the oxygen vacancy is a shallow donor. To determine if the oxygen vacancy continues to be shallow or it can become a deep donor *after* LDA correction, we calculated the position of the single-particle level of a *neutral* oxygen vacancy, $\varepsilon(V_O)$, by several methods, as listed in Table III. The $E1$ and self-interaction correction (SIC) methods are discussed in the Appendix. The model GW method³⁴ is an approximate approach to obtain the quasiparticle excitation energies. In the LDA*, we calculate the level positions without the relativistic effect.³⁵ The relativistic effect lowers the CBM by 0.2 eV, thus reducing the band gap. We then extrapolate from the results to the experimental band gap to obtain $\varepsilon(V_O)$. All the four methods in Table III indicate that the $\varepsilon(V_O)$ level is deep. Physically, the a_1 -like gap-state of the oxygen vacancy is derived from the Zn $4s$ orbitals surrounding the vacancy. This is different from the t_2 -like nitrogen vacancy in GaN where the gap state is derived from the Ga $4p$ orbital. Since the atomic s energy is about 5 to 6 eV lower than the p energy in Zn and Ga, the nitrogen vacancy state is shallow, while the oxygen vacancy state is deep. A number of recent experiments have assigned¹³⁻¹⁶ the oxygen vacancy to the green luminescence seen in ZnO. The measured 2.4 eV emission energy appears to agree with the calculated $\varepsilon(V_O)$ and E_V separation of about 2.3 eV given by the GW and SIC models. Our result for V_O is also consistent with results for the selenium vacancy in ZnSe,³⁶ where the single-particle defect level is

estimated to be only 1.0 eV above the LDA CBM, or 1.8 eV below the experimental CBM. Thus, *neutral* anion vacancy levels in II-VI compounds are most likely deep.

The situation is somewhat different for ionized vacancies. In particular, we find a large lattice distortion around the oxygen vacancy of the (2+) charge state, resulting in energy lowering of 1.3 eV. An energy lowering of 1.6 eV upon similar distortion was found for selenium vacancy V_{Se}^{2+} in ZnSe,³⁶ and for the V_{As}^{3+} vacancy in GaAs.³⁷ This energy lowering makes the oxygen vacancy a “negative- U ” center where the effective electron-vacancy Coulomb energy in the defect orbital is $U \equiv \Delta H(V^{2+}) + \Delta H(V^0) - 2\Delta H(V^+) = -0.6$ eV. The U for V_O becomes more negative upon the LDA gap correction. In comparison, $U = -1.0$ eV for V_{Se} in ZnSe.³⁶

Recently, Studenikin, Golego, and Cocivera¹⁸ reported a very slow decay (over a time interval of days) of the steady-state photoconductivity in ZnO (persistent photoconductivity or PPC). They attributed the PPC to hole-traps at the grain boundaries. PPC is common for defects in semiconductors such as for DX centers in $\text{Al}_x\text{Ga}_{1-x}\text{As}$.³⁸ Based on our results for oxygen vacancies, we suggest here an alternative explanation for the PPC in ZnO. A charge neutral V_O can capture two holes due to its negative U . Upon capturing two holes, the V_O undergoes a large lattice distortion without energy barrier. The holes thus become deeply trapped and localized. This reduces the direct recombination rate with photogenerated, delocalized electrons (responsible for conduction). The large lattice distortion and the deepness of the hole-trap level in the gap also prevent fast phonon-assisted nonradiative recombination, thus causing the PPC.

ACKNOWLEDGMENT

We thank T. Mattila for helpful discussions and S. Muller for the Zn pseudopotential without core. This work was supported by the U.S. DOE under Contract No. DE-AC36-99GO10337 and NERSC for Cray T3E time.

APPENDIX: EMPIRICAL CORRECTIONS TO THE LDA DEFECT-FORMATION ENTHALPIES

To get the correct defect formation enthalpies, we need a different approach than the LDA that does not have the band-gap error. A similar problem occurs in III-V compounds, where despite smaller LDA gap errors, we found that the calculated pinning energy, $\varepsilon_{\text{pin}}^{(n)}$, which is the upper bound for E_F in *n*-type doping, is consistently too small compared to experiments.²⁸ A general expression to correct the defect formation enthalpy is²⁸

$$\begin{aligned} \Delta E_{\text{con}}(\lambda) &= \Delta E_{\text{LDA}}(\lambda_0) + \left(\frac{\partial \Delta E}{\partial \varepsilon_g} \right)_{\lambda=\lambda_0} [\varepsilon_g(\lambda) - \varepsilon_g^{\text{LDA}}(\lambda_0)] \\ &= \Delta E_{\text{LDA}} + \delta E, \end{aligned} \quad (\text{A1})$$

where λ is a parameter that satisfies $\varepsilon_g(\lambda) = \varepsilon_g^{\text{expt}}$. Here, we test several semiempirical approaches that are known to alter the LDA band gaps and a self-interaction correction (SIC)

approach²⁶ by using SIC pseudopotentials.²⁹ All of them have the effect to push down the valence band relative to the CBM. The approaches are as follows.

(i) Use for λ the cutoff energy $E1$ in the plane-wave expansion,²⁸ i.e., $\lambda = E1 = 60, 50, \text{ and } 48 \text{ Ry}$, respectively. The shortcoming, though, is that the parameter $E1$, while restricting the short-wavelength components in the basis set, has no direct physical meaning.

(ii) Use for λ the coefficient of exchange-correlation energy functional, e.g., $\lambda = \alpha = \frac{2}{3}, 0.7 \text{ and } 0.74$ in the $X\alpha$ method.³⁰ A larger α here lowers the exchange-correlation energy. The difficulty here is, however, the strong dependence of the formation enthalpy of bulk ZnO on the parameter α . We have calculated the formation enthalpies of the oxygen and zinc vacancies at the oxygen- and zinc-rich conditions. We also calculated the zinc and oxygen interstitials at the oxygen- and zinc-rich conditions, respectively.

(iii) Use for λ the p - d repulsion³¹ in ZnO, i.e., $\lambda_0 = (d \text{ in the valence})$ and $\lambda = (d \text{ in the core})$ in Eq. (A1). The d band of Zn is too high due to the lack of the electron self-interaction energy in the LDA.²⁶ A too high d band repels unphysically the p band above, reducing the band gap.³¹ Having the d band in the core instead increases the gap from 0.6 to 1.9 eV.

(iv) Use instead the self-interaction correction (SIC).²⁶ Electron self-interaction is an unphysical effect intrinsic to the Kohn-Sham equation.²³ Namely, in the original Kohn-Sham formalism, each electron experiences an effective potential generated not only by other electrons and ions but also by itself. Recently, Vogel, Kruger, and Pollman²⁹ developed a SIC pseudopotential scheme.²⁹ They showed that the SIC pseudopotentials improve considerably the LDA single-particle band gaps. While the SIC method is considered a more rigorous method than the LDA,²⁶ the total energy for charge-neutral systems is formulated in Ref. 29 using a first-order perturbation scheme, and is not variational. As such, it cannot be extended unambiguously to charged defect calculations. To get total energy for charged defects, necessary for doping compensations, we use, instead, items (i)–(iii) above.

Figure 3 shows the various corrections δE . There is some scatter of the data: typically 1.5 eV between the $E1$ and the $X\alpha$ methods with a few exceptions, e.g., 3.2 eV for V_{O}^0 . The scattering between $E1(X\alpha)$ and d in the core is somewhat larger. Both the $X\alpha$ and d in the core methods appear to drive down the energy of the positively charged donor states more than the $E1$ method does. Compare the $E1$ and the SIC methods for the charge-neutral defects, the differences in the defect formation enthalpies are 0.8 eV for V_{O}^0 , -0.5 eV for Zn_i^0 , and 2.2 eV for Zn_{O}^0 , respectively. The relatively large magnitude of the scattering reflects the large LDA gap error, 2.7 eV, or 82% of the total gap of ZnO. There are, nevertheless, several general trends in Fig. 3, irrespective of the uncertainties in the LDA corrections.

(i) For donors [see Fig. 3(a)], the following applies.

(a) Corrections to the $2+$ states are generally large and negative. They become smaller and sometimes change sign for $q = 1+$ and 0 states.

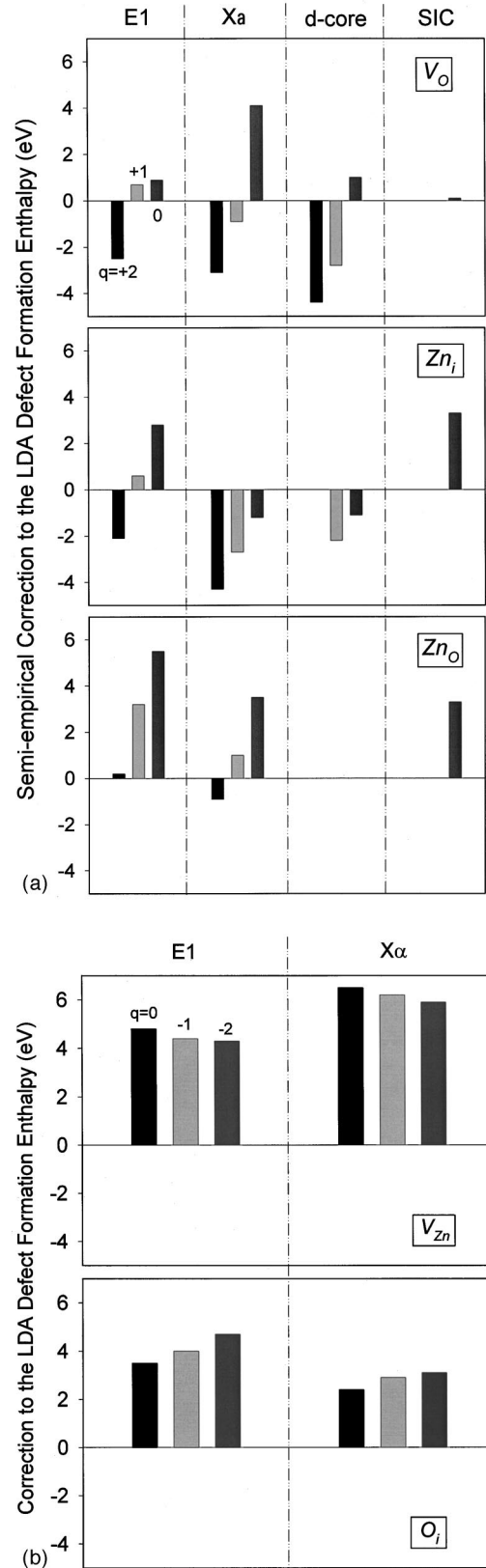


FIG. 3. Calculated LDA corrections using the various methods discussed in the Appendix. (a) For donorlike defects, the histograms from left to right in each case correspond to charge states $q = +2, +1$ and 0, respectively. (b) For acceptorlike defects, the charge states from left to right are $q = 0, -1, -2$, respectively.

(b) As a result, corrections increase the separation between different charge states.

(ii) For acceptors [see Fig. 3(b)], the following applies.

(a) Corrections are generally large and *positive*.

(b) Corrections for different charge states are similar.

(iii) Combining (i)(a) and (ii)(a) above, we see that LDA corrections for ZnO *consistently* increase the asymmetry in the formation enthalpies of the donorlike and acceptorlike defects in favor of the donorlike defects.

Of the three semiempirical methods, the $E1$ method has the smallest overall deviation from the SIC. Besides, the $E1$ method introduces only modest change in the formation enthalpy of bulk ZnO. Hence, we will use the $E1$ results (as tabulated in Table I and II along with the LDA results) for further discussions.

The corrections have the following effects:

(a) For Zn_i , Zn_O , and V_{Zn} , the positions of the defect levels relative to the band edges are unchanged after the correction, although the electron-electron correlation energy U for Zn_i and Zn_O defect orbitals become slightly negative by the order of 0.1 eV.

(b) The oxygen interstitial has two deep acceptor levels in the LDA: Both moved deeper into the gap by about 0.6 eV after the correction.

(c) The largest change occurs for V_O . While both the direct and corrected LDA calculations predict a negative U , the actual (2+/0) levels differ qualitatively. In the LDA, it is above the CBM, reflecting a too small LDA band gap. After the LDA correction, however, it is 0.6 eV below the CBM.

- ¹S. B. Zhang, S.-H. Wei, and A. Zunger, *J. Appl. Phys.* **83**, 3192 (1998).
- ²F. A. Kroger, *The Chemistry of Imperfect Crystals* (North-Holland, Amsterdam, 1964), p. 691.
- ³K. I. Hagemark, *J. Solid State Chem.* **16**, 293 (1976).
- ⁴E. Ziegler, A. Heinrich, H. Oppermann, and G. Stover, *Phys. Status Solidi A* **66**, 635 (1981).
- ⁵V. A. Nikitenko, S. A. Stenli, and N. K. Morozova, *Izv. Akad. Nauk. SSSR, Neorg. Mater.* **24**, 1830 (1988).
- ⁶J. C. Simpson and J. F. Cordaro, *J. Appl. Phys.* **67**, 6760 (1990).
- ⁷R. M. de la Cruz, R. Pareja, R. Gonzalez, L. A. Boatner, and Y. Chen, *Phys. Rev. B* **45**, 6581 (1992).
- ⁸D. C. Look, J. W. Hemsky, and J. R. Sizelove, *Phys. Rev. Lett.* **82**, 2552 (1999).
- ⁹R. G. Gordon, AIP Conference Proceedings 394, 39 (1997).
- ¹⁰K. Minegishi, Y. Koiwai, Y. Kikuchi, K. Yano, M. Kasuga, and A. Shimizu, *Jpn. J. Appl. Phys.*, **36**, L1453 (1997).
- ¹¹M. Joseph, H. Tabata, and T. Kawai, *Jpn. J. Appl. Phys.*, **38**, L1205 (1999).
- ¹²Y. Chen, D. M. Bagnall, Z. Zhu, T. Sekiuchi, K. Park, K. Hiraga, T. Tao, S. Koyama, M. Y. Shen, and T. Goto, *J. Cryst. Growth* **181**, 165 (1997).
- ¹³T. V. Butkhuzi, A. V. Bureyev, A. N. Georgobiani, N. P. Kekelidze, and T. G. Khulordava, *J. Cryst. Growth* **117**, 366 (1992).
- ¹⁴K. Vanheusden, C. H. Seager, W. L. Warren, D. R. Tallant, and J. A. Voigt, *Appl. Phys. Lett.* **68**, 403 (1996).
- ¹⁵K. Vanheusden, C. H. Seager, W. L. Warren, D. R. Tallant, J. A. Voigt, and B. E. Gnade, *J. Appl. Phys.* **79**, 7983 (1996).
- ¹⁶K. Vanheusden, C. H. Seager, W. L. Warren, D. R. Tallant, J. Caruso, M. J. Hampden-Smith, and T. T. Kodas, *J. Lumin.* **75**, 11 (1997).
- ¹⁷J. Zhong, A. H. Kitai, P. Mascher, and W. Puff, *J. Electrochem. Soc.* **140**, 3644 (1993).
- ¹⁸S. A. Studenikin, N. Golego, and M. Cocivera, *J. Appl. Phys.* **83**, 2104 (1998); **84**, 2287 (1998); **84**, 5001 (1998).
- ¹⁹W. C. Mackrodt, R. F. Stewart, J. C. Campbell, and I. H. Hillier, *J. Phys. (Paris), Colloq.* C6-64 (1980).
- ²⁰T. Yamamoto and H. Katayama-Yoshida, *Jpn. J. Appl. Phys.* **38**, L66 (1999).
- ²¹S. B. Zhang and J. E. Northrup, *Phys. Rev. Lett.* **67**, 2339 (1991).
- ²²J. Ihm, A. Zunger, and M. L. Cohen, *J. Phys. C* **12**, 4409 (1979).
- ²³W. Kohn and L. J. Sham, *Phys. Rev.* **140**, A1133 (1965).
- ²⁴N. Troullier and J. L. Martins, *Phys. Rev. B* **43**, 1993 (1991).
- ²⁵D. M. Ceperley and B. J. Alder, *Phys. Rev. Lett.* **45**, 566 (1980).
- ²⁶J. P. Perdew and A. Zunger, *Phys. Rev. B* **23**, 5048 (1981).
- ²⁷G. Makov and M. C. Payne, *Phys. Rev. B* **51**, 4014 (1995).
- ²⁸S. B. Zhang, S.-H. Wei, and A. Zunger, *Phys. Rev. Lett.* **84**, 1232 (2000).
- ²⁹D. Vogel, P. Kruger, and J. Pollmann, *Phys. Rev. B* **54**, 5495 (1996).
- ³⁰J. C. Slater, *Phys. Rev.* **81**, 385 (1951).
- ³¹S.-H. Wei and A. Zunger, *Phys. Rev. B* **37**, 8958 (1988).
- ³²D. B. Laks, C. G. Van de Walle, G. F. Neumark, P. E. Blochl, and S. T. Pantelides, *Phys. Rev. B* **45**, 10965 (1992); C. Van De Walle (private communication).
- ³³S.-H. Wei, S. B. Zhang, and A. Zunger, *Phys. Rev. Lett.* **70**, 1639 (1993).
- ³⁴P. A. Sterne and J. C. Inkson, *J. Phys. C* **17**, 1497 (1984).
- ³⁵N. E. Christensen, *Phys. Rev. B* **30**, 5753 (1984).
- ³⁶A. Garcia and J. E. Northrup, *Phys. Rev. Lett.* **74**, 1131 (1995).
- ³⁷J. E. Northrup and S. B. Zhang, *Phys. Rev. B* **50**, 4962 (1994).
- ³⁸D. V. Lang, R. A. Logan, and M. Jaros, *Phys. Rev. B* **19**, 1015 (1979).