

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 intrinsic²⁻⁸ 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

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 a in charge state q is²¹

$$DH^{-q,a} = DE^{-q,a} + n_a m_a + q E_F, \quad -1!$$

where

$$DE^{-q,a} = E^{-q,a} - \text{defect} + \text{host} - E - \text{host only} + n_a m_a - \text{solid} + q E_V. \quad -2!$$

Here, $E^{(q,a)}$ -defect+host! is the total energy of a cell including the host material as well as defect a in charge state q , and E -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

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-

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. m_a is the chemical potential of the reservoir, and m_a -solid! is the energy of elemental solid a . Under thermal equilibrium $m_{\text{Zn}} + m_{\text{O}} = \Delta H_{\text{ZnO}}$, where ΔH_{ZnO} is the formation enthalpy of ZnO, so $m_{\text{O}} = \Delta H_{\text{ZnO}} - m_{\text{Zn}}$. Therefore, for ZnO, $n_{\text{O}} = 1$ and $n_{\text{Zn}} = -$

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 $DH(E_F, m)$ depend on the LDA band gap, since $E_V < E_F < E_C$, 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 DH_{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

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 !-, $\sim -/2$!-, or $\sim 0/2$!- with respect to the VBM, so that they readily produce holes.

-b! Acceptors must have *low* formation enthalpy DH 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 m 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 $\sim 2+/0$! 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-

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, $\ll(V_O)$, 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 $\ll(V_O)$. All the four methods in Table III indicate that the $\ll(V_O)$ 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 $\ll(V_O)$ and E_V separation of

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 l the cutoff energy $E1$ in the plane-wave expansion,²⁸ i.e., $l = E1 = 60, 50, \text{ and } 48 \text{ Ry}$, 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 l the coefficient of exchange-correlation energy functional, e.g., $l = a = \frac{2}{3}, 0.7 \text{ and } 0.74$ in the Xa method.³⁰ A larger a here lowers the exchange-correlation energy. The difficulty here is, however, the strong dependence of the formation enthalpy of bulk ZnO on the parameter a . 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 l the p - d repulsion³¹ in ZnO, i.e., $l_0 = (d \text{ in the valence})$ and $l = (d \text{ in the core})$ 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 Xa methods with a few exceptions, e.g., 3.2 eV for V_O^0 . The scattering between $E1(Xa)$ and d in the core is somewhat larger. Both the Xa 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+

-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 $\Delta E_{\text{F10}} = 0.610$ eV, $\Delta E_{\text{F30}} = 0.333$ eV, $\Delta E_{\text{F10}} = 0.278$ eV, $\Delta E_{\text{F10}} = 0.33$ eV. The *E1* method has