Energetics of native defects in ZnO

Fumiyasu Oba, a) Shigeto R. Nishitani, Seiji Isotani, and Hirohiko Adachi
Department of Materials Science and Engineering, Kyoto University, Sakyo, Kyoto 606-8501, Japan
Isao Tanaka
Department of Energy Science and Technology, Kyoto University, Sakyo, Kyoto 606-8501, Japan
(Received 21 March 2001; accepted for publication 26 April 2001)

We have investigated the formation energies and electronic structure of native defects in ZnO by a first-principles plane-wave pseudopotential method. When p-type conditions are assumed, the formation energies of donor-type defects can be quite low. The effect of self-compensation by the donor-type defects should be significant in p-type doping. Under n-type conditions, the oxygen vacancy exhibits the lowest formation energy among the donor-type defects. The electronic structure, however, implies that only the zinc interstitial or the zinc antisite can explain the n-type conductivity of undoped ZnO. © 2001 American Institute of Physics. [DOI: 10.1063/1.1380994]

I. INTRODUCTION

Zinc oxide (ZnO) is widely used is in our community. Polycrystalline ZnO with some additives shows highly nonlinear current–voltage characteristics, which has been utilized as a varistor.1–3 Aluminum-doped ZnO thin films have been developed as n-type transparent conductors,4,5 and extensive efforts have been paid for the search of good p-type dopants.6,7 In the applications of ZnO as electronic materials, the control of native defects is essential since the electrical properties are largely affected not only by extrinsic dopants but also by native defects. The atomic and electronic structure of the native defects in ZnO has been extensively investigated by a variety of experimental methods, e.g., electron paramagnetic resonance (EPR),8–11 cathode luminescence,12,13 deep-level transient spectroscopy,14,15 positron annihilation spectroscopy,12,16–18 and perturbed angular correlation spectroscopy.19,20 However, fundamental knowledge on the native defects is still lacking; the defect species that dominates the electrical properties of undoped ZnO has thus far been in controversy.8,9,21–23

The electrical properties of undoped systems have carefully been investigated by Hutson,24 by Hagemark and Chacka,25 and by Ziegler et al.26 They have reported that reduced samples exhibit n-type conductivity with donor energies of less than 0.05 eV,24–26 and that the donor energies decrease with the increase of carrier concentration.25,26 For an extreme case, the donor energy is reduced to 0 eV, i.e., the electrical property is metallic.26 The defect species that acts as the donor, however, has been controversial: some reports suggest the zinc interstitial,21,22 others the oxygen vacancy.8,9,23 Hagemark has discussed the defect structure on the basis of quasichemical reactions and suggested the zinc interstitial as the major donor.22 On the other hand, Mahan has reinterpreted his results and concluded that the major donor is the oxygen vacancy.23 The native defects have also been investigated using EPR, but the interpretation has been likewise controversial.8–11,21 Hagemark and Toren have reported that excessive zinc concentration shows good quantitative correlation with carrier concentration on the basis of the electrochemical and Hall effect measurements.27 This result should suggest that the origin of the n-type conductivity of ZnO is given solely by a dominant defect species. However, we do not yet have consistent answers on the defect species as introduced above.

In the present study, the formation energies and electronic structure of the native defects in ZnO are investigated through first-principles calculations. The results are discussed with a special focus on the problem of the native donor.

II. METHODOLOGY

The first-principles calculations were performed using a plane-wave pseudopotential method28 within the generalized gradient approximation (GGA).29 The self-consistent total energy in the ground state was effectively obtained by the density-mixing scheme.30 Atomic positions were optimized to minimize the total energy using the quasi-Newton method with the Broyden–Fletcher–Goldfarb–Shanno hessian update scheme.31 Ultrasoft pseudopotentials32,33 were employed, where the Zn 3d states were explicitly treated as a part of the valence.

Regarding defect species, vacancies, interstitials at the octahedral and tetrahedral sites, and antisites in the relevant charge states were considered. 72-atom supercells containing a defect were used in the calculation. For the main results, 128-atom supercells were also used to confirm the convergence with respect to cell size. Numerical integration over the Brillouin zone was carried out at the Γ point for both the 72- and 128-atom supercells. The plane-wave cutoff energy was chosen to be 380 eV. Test calculations using 16-atom supercells for two samples, the neutral oxygen vacancy and the neutral zinc interstitial at the octahedral site, showed that this cutoff energy achieves convergence of the formation energies within 0.1 eV relative to the results with the cutoff.
energies up to 800 eV. Spin polarization was not considered since it is unlikely to affect the present results significantly, e.g., an energy gain by spin polarization was calculated to be 0.1 eV for the oxygen vacancy in the charge state 1+. The atomic positions of the first and the second nearest neighbors of the defects were relaxed under the condition that the cell volume was fixed to be the value optimized for the perfect crystal; lattice constants a, c, and internal parameter u were calculated to be 100.5%, 100.3%, and 99.8% of experimental values, respectively. The optimization procedure was truncated when the residual forces for the relaxed atoms were less than 0.25 eV/Å. Further minimization of the forces was unlikely to affect the present results; total energies decreased by less than 0.001 eV/atom when the forces were reduced to 0.05 eV/Å for 16-atom supercells when the forces were reduced to 0.05 eV/Å for 16-atom supercells with a neutral oxygen vacancy or a neutral zinc interstitial at the octahedral site.

The formation energies of the defects were calculated using the total energies of the supercells. For compound systems, the formation energies depend on the atomic chemical potentials. In addition, those of charged defects also vary with the electronic chemical potential, i.e., Fermi energy. For a defect in a charge state \( q \), the formation energy is given by

\[
E_{\text{formation}}(q) = E_T(q) - n_{\text{Zn}}\mu_{\text{Zn}} - n_{\text{O}}\mu_{\text{O}} + qE_F,
\]

where \( E_T \) is the total energy of the supercell with a defect in a charge state \( q \), \( n_{\text{Zn}} \) and \( n_{\text{O}} \) are the number of zinc and oxygen atoms in the supercell. \( \mu_{\text{Zn}} \) and \( \mu_{\text{O}} \) are the atomic chemical potentials, and \( E_F \) is the Fermi energy. For charged defects, the total charge of supercells was neutralized using jellium background. Energy shifts associated with the jellium neutralization were corrected by the total energy difference between the neutral (normal) and charged systems of the perfect crystal, e.g., for the positively charged systems, electrons were removed from the valence band maximum and the jellium neutralization was included. Accordingly, the total energies of the supercells with positively charged defects were evaluated in the case that the Fermi energy is the same as the valence band maximum. Negatively charged systems were dealt with in the same manner, except that the conduction band minimum was used as the reference.

\( \mu_{\text{Zn}} \) and \( \mu_{\text{O}} \) are variables correlated as

\[
\mu_{\text{Zn}} + \mu_{\text{O}} = \mu_{\text{ZnO(bulk)}},
\]

where \( \mu_{\text{ZnO(bulk)}} \) is the chemical potential of the bulk ZnO, is a constant value calculated as the total energy per ZnO unit formula. The total energies per atom for the bulk ZnO (P6\(_3\)/mmc) and the bulk O(C2/m) were chosen as the upper limits of \( \mu_{\text{Zn}} \) and \( \mu_{\text{O}} \), respectively. In this case, ZnO may seemingly determine the upper limit of \( \mu_{\text{O}} \). Our calculation, however, showed it is unlikely: the total energy of ZnO\(_2\) was calculated to be 0.7 eV higher than the summation of the total energies of the bulk ZnO and the bulk O. \( \mu_{\text{Zn}} \) and \( \mu_{\text{O}} \) therefore vary over a range given by the heat of formation of ZnO, which was calculated to be 3.06 eV. This value is close to the experimental value reported, 3.63 eV (298.15 K).

The atomic chemical potentials describe the conditions under which materials are grown. The Fermi energy depends on the concentrations of native defects and impurities. In the following sections, we will discuss the formation energies of native defects as a function of the atomic chemical potentials and Fermi energy on the basis of Eq. (1).

### III. RESULTS AND DISCUSSION

Figure 1 shows formation energies under the two extreme conditions, the oxygen-rich limit (\( \mu_{\text{Zn}} = \mu_{\text{ZnO(bulk)}} \) and \( \mu_{\text{O}} = \mu_{\text{O(bulk)}} \)) and the zinc-rich limit (\( \mu_{\text{Zn}} = \mu_{\text{Zn(bulk)}} \) and \( \mu_{\text{O}} = \mu_{\text{O(bulk)}} \)). The slope corresponds to the charge state \( q \) as used in Eq. (1). For each defect species, only the charge state that gives the lowest formation energy with respect to the Fermi energy is shown. Change in the slope indicates transition in the charge state, which is shown with symbols. The valence band maximum is chosen as the zero of the Fermi energy; the vertical line at 0.96 eV corresponds to the conduction band minimum. The subscripts in the notation of defect species indicate defect sites, e.g., “Zn” and “oct.” denote the zinc lattice and octahedral interstitial sites, respectively.
energy of the defect electronic state, is independent of the atomic chemical potentials and hence it is common to Figs. 1(a) and 1(b). The Fermi energy is measured from the valence band maximum, and the conduction band minimum is shown at 0.96 eV. This energy, i.e., the band gap, was calculated as \( E^{(1)} - E^{(0)} - (E^{(0)} - E^{-1}) \), where \( E^{(N)} \) indicates the total energy of a perfect lattice supercell with additional \( N \) electrons. The same energy was obtained from the difference of the one-electron energies. The calculated value of 0.96 eV is considerably smaller than the experimental band gap of 3.30 eV, which is known to be due to the GGA. This may cause some systematic errors in absolute formation energy. A method to correct the formation energies of defects to overcome the underestimation has been suggested. The main points of the method are the following: The conduction band is rigidly shifted upward to match the experimental band gap. Donor-type defect states, which are expected to have characters similar to states in the conduction band, are assumed to follow the upward shift. Acceptor-type defect states are expected to have characters similar to valence states and hence left unchanged. According to the upward shift, the formation energies of donor-type defects are assumed to increase by the energy of the conduction band shift multiplied by the occupation number of the defect states. On the other hand, those of acceptor-type defects are assumed to remain unchanged.

In the present article, we show the results without corrections for simplicity. Even if the above correction is adopted, the main points of the following discussions using the relative values of the formation energies remain quantitatively identical; for all the range of the Fermi energy, the relative values are kept unchanged among the oxygen vacancy and the zinc interstitials as donor-type defects. Likewise, absolute values for all the donor-type defects are not affected when the Fermi energy is close to the valence band maximum.

The formation energies depend largely on the atomic chemical potentials, as recognized in Fig. 1. Turning from oxygen-rich to zinc-rich conditions, the defects associated with oxygen excess, i.e., the zinc vacancy, the oxygen interstitials, and the oxygen antisite are raised in formation energy, whereas those associated with zinc excess, i.e., the oxygen vacancy, the zinc interstitials, and the zinc antisite, are lowered. Among the defects associated with oxygen excess, the zinc vacancy exhibits the lowest formation energy in all the range of the Fermi energy. It is expected to act as a single or doubly ionized acceptor; the Fermi energy at which the charge state changes from 1- to 2-, i.e., the energy of the defect electronic state that will be noted \( \varepsilon(1/-2) \) hereafter, is close to the valence band maximum. The lowest formation energy is consistent with the generally accepted picture that the zinc vacancy is dominant among the defects associated with oxygen excess.

The Fermi energy is the same as the valence band maximum, and the conduction band minimum is assumed, however, the formation energies of the zinc vacancy and the zinc interstitials as donor-type defects. Consequently, the donor-type defects show lower formation energies and should therefore be dominant under most of the conditions of the chemical potentials. The extreme case is the zinc-rich limit that is shown in Fig. 1(b) and Fig. 2(a). Thus, the formation energies of the donor-type defects can be quite low when \( p \)-type conditions are assumed. The self-compensation associated with the donor-type defects may be a reason of the difficulty in the \( p \)-type doping of ZnO.

As for the defects associated with zinc excess, the oxygen vacancy and the zinc interstitial at the octahedral site, which are donor type, are as low as that of the zinc vacancy. This suggests that the effect of the self-compensation by the donor-type defects should be taken into account in \( p \)-type doping even at the oxygen-rich limit. The formation energies of the zinc vacancy and the donor-type defects are shown as a function of the atomic chemical potentials in Fig. 2. As mentioned above and can be seen in Fig. 2(a), the dependence of the formation energies on the chemical potentials is opposite in sign between the zinc vacancy and the donor-type defects. Thus, the formation energies of the donor-type defects can be quite low when \( p \)-type conditions are assumed. The self-compensation associated with the donor-type defects may be a reason of the difficulty in the \( p \)-type doping of ZnO.
4+ to 2+ in this range, is also comparable in formation energy, while it is much higher under oxygen-rich conditions as at the limit shown in Fig. 1(a). As the Fermi energy approaches the conduction band minimum, the oxygen vacancy becomes energetically preferable due to the transition in the charge state from 2+ to 0; in this case, the charge state 1+ does not appear in the diagram since the formation energy was calculated to be 0.4 eV higher than those of the oxygen vacancies in the neutral and 2+ charge states for the Fermi energy of 0.5 eV. When the Fermi energy is the same as the conduction band minimum, the formation energy of the oxygen vacancy is more than 1.2 eV lower than the other donor-type defects under any conditions of the atomic chemical potentials, as also recognized in Fig. 2(b). We believe that the energy differences do not change significantly even if supercell size is increased; using the 128-atom supercells, the difference between the oxygen vacancy and the zinc interstitial at the octahedral site was reduced only by 0.1 eV. For n-type materials where the Fermi energy is close to the conduction band minimum, the oxygen vacancy should be dominant among the donor-type defects.

We now focus on the electronic states of the donor-type defects. There has been a controversy on the native donor in ZnO; which defect plays a central role?8,9,21-23 It has been reported that undoped ZnO annealed at elevated temperatures exhibits n-type conductivity with donor energies of less than 0.05 eV.24-26 Likewise, the double ionization of the donors has been suggested by the dependence of carrier concentration on zinc partial pressure during heat treatments.25

The oxygen vacancy shows the electronic state \( \epsilon(2+/0) \) at 0.5 eV, which is quite lower than the conduction band minimum of 0.96 eV. The n-type conductivity of ZnO cannot be explained by the oxygen vacancy with such a deep defect state. On the other hand, the zinc interstitials and the zinc antisite show the electronic states \( \epsilon(2+/1+/0) \) in the conduction band; we regard them as transition levels over three charge states, 2+, 1+ and 0+ since the differences in formation energy among the charge states are less than 0.1 eV at the Fermi energies where the transition takes place. When the Fermi energy is close to the conduction band minimum, the charge states of the zinc interstitials and the zinc antisite are 2+. They are therefore expected to act as doubly ionized donors under n-type conditions, which is consistent with the observed electrical property of ZnO.25 This electronic structure is quite different from that of the oxygen vacancy with a deep electronic state.

We have thus far discussed defect electronic states on the basis of the total energy. To understand the origin of the defect states, it is useful to investigate the one-electronic band structure. Figure 3 shows band structures for the supercells with the oxygen vacancy and the zinc interstitial at the octahedral site in the neutral charge state. The conduction band for the perfect lattice supercell is superimposed by dotted curves; the valence band structure, which is not shown in the figures, is very similar to the defect models for the 72-atom supercell used in the present study.

An occupied band is recognized near the top of the valence band for the oxygen vacancy in Fig. 3(a). It was found that two electrons in this band localize mainly on the vacancy and its first nearest neighboring zinc atoms by the charge density analysis in real space. This can therefore be regarded as a defect state induced by the oxygen vacancy. The dispersion of the band with respect to wave vector is 0.57 eV, which may be caused by the interaction among the vacancies repeated under three-dimensional periodic boundary conditions. When the 128-atom supercell was used, the band became closer to the valence band maximum with the dispersion of 0.28 eV. This tendency suggests that the one-electronic state induced by the oxygen vacancy is likely to appear near the top of the valence band maximum even if supercell size is further increased. The deep one-electronic state should be the origin of the deepness of the total energy-based electronic state shown in Fig. 1. The electronic structure is quite different from nitrides such as GaN and BN for which neutral nitrogen vacancies have been reported to induce occupied defect states in the conduction band.40,41

For the zinc interstitial in Fig. 3(b), the highest occupied band is fairly similar to the lowest one in the conduction band of the perfect lattice supercell. The large dispersion implies that the defect-induced state shows resonance with the conduction band; using the 128-atom supercell, the band became closer in shape and energy to the lowest one in the conduction band of the perfect lattice supercell. This one-electronic band structure can explain why the total energy-based defect state \( \epsilon(2+/1+/0) \) appears in the conduction band. Similar features in the one-electronic structure were recognized for the zinc antisite and the zinc interstitial at the

![Image](https://example.com/image.png)
tetrahedral site, which are also consistent with the defect states based on total energy.

Thus, the electronic structure indicates that only the zinc interstitial or the zinc antisite can explain the n-type conductivity of undoped ZnO. On the other hand, the oxygen vacancy should be dominant under n-type conditions since it shows the lowest formation energy among the donor-type defects. This inconsistency may be a reason for the controversy on the native donor during past decades. Since excessive zinc concentration has been reported to show good quantitative correlation with carrier concentration, the origin of n-type conductivity is likely to be given solely by a dominant defect species. Defect complexes and/or unknown residual impurities are therefore proposed to play central roles for the n-type conductivity of undoped ZnO.

IV. SUMMARY

We have investigated the formation energies and electronic structure of native defects in ZnO through first-principles calculations. When p-type conditions are assumed, the formation energies of the donor-type defects can be quite low. The effect of self-compensation by donor-type defects should be significant in p-type doping. Regarding the native donor, the electronic structure indicates that only the zinc interstitials or the zinc antisite can explain the n-type conductivity of undoped ZnO, whereas the oxygen vacancy with the lowest formation energy should be dominant under n-type conditions. Defect complexes and/or residual impurities should be taken into account for the problem of the native donor in ZnO.

Note added in proof. After the submission of this article, a theoretical study on native defects in ZnO by Kohan et al. [Phys. Rev. B 61, 15 019 (2000)] was brought to our attention. They calculated quantities similar to those shown in the present study; some differences in the calculated formation energies and electronic states between their and our studies are presumably due to differences in computational details. However, the major points of the discussion are different; they have discussed the results with the focus on the dominant defect species and the mechanism of observed green luminescence, whereas we focus on the problem of the native donor in the present study.

ACKNOWLEDGMENTS

The authors are grateful to M. Kohyama for helpful discussions. This study was performed through the ACT program of JST.