First-principles study of intrinsic point defects in ZnO: Role of band structure, volume relaxation, and finite-size effects

Paul Erhart, Karsten Albe, and Andreas Klein
Institut für Materialwissenschaft, Technische Universität Darmstadt, Petersenstraße 23, D-64287 Darmstadt, Germany
(Received 25 November 2005; revised manuscript received 6 February 2006; published 15 May 2006)

Density-functional theory (DFT) calculations of intrinsic point defect properties in zinc oxide were performed in order to remedy the influence of finite-size effects and the improper description of the band structure. The generalized gradient approximation (GGA) with empirical self-interaction corrections (GGA+U) was applied to correct for the overestimation of covalency intrinsic to GGA-DFT calculations. Elastic as well as electrostatic image interactions were accounted for by application of extensive finite-size scaling and compensating charge corrections. Size-corrected formation enthalpies and volumes as well as their charge state dependence have been deduced. Our results partly confirm earlier calculations but reveal a larger number of transition levels: (1) For both the zinc interstitial as well as the oxygen vacancy, transition levels are close to the conduction band minimum. (2) The zinc vacancy shows a transition rather close to the valence band maximum and another one near the middle of the calculated band gap. (3) For the oxygen interstitials, transition levels occur both near the valence band maximum and the conduction band minimum.

DOI: 10.1103/PhysRevB.73.205203 PACS number(s): 61.72.Bb, 61.72.Ji, 71.15.Mb, 71.55.Gs

I. INTRODUCTION

The current interest in zinc oxide is largely driven by potential applications in optical and optoelectronic devices.[1] Since many properties of zinc oxide are highly sensitive to point and line defects present in the material, the defect physics of ZnO has been extensively studied in the past. Theoretically, a number of density functional theory (DFT) calculations have been performed to elucidate the behavior of intrinsic[2–5] as well as extrinsic point defects.[6–9] These calculations, however, are based on the local density (LDA) or generalized-gradient approximation (GGA) which suffer from an underestimation of the band gap and an improper description of the band structure. The first shortcoming is intrinsic to the DFT method in general (see, e.g., Refs. 10 and 11). The second problem is particularly pronounced for zinc oxide because self-interactions intrinsic to the LDA and GGA exchange-correlation potentials cause an energy level shift of the Zn 3d states. As a result, the calculations not only yield a band-gap error of more than 2 eV but also overestimate the covalency of the Zn-O bond. A direct comparison between data calculated within LDA or GGA-DFT and experiment is, therefore, severely hampered. In the past, this problem has been addressed in various ways.

Zhang *et al.* proposed an empirical correction scheme based on a Taylor expansion of the formation enthalpies in the plane-wave cutoff energy.[3,12] Since a profound physical motivation for this scheme is lacking, the results can only be interpreted semiquantitatively. Kohan *et al.* discussed corrections based on the electronic structure of the defect configurations,[2] while other authors resorted to a qualitative discussion of their results.[4,5]

If no correction is applied the calculated formation enthalpies reported by different authors are comparable (see Table I below), whereas the various correction schemes lead to very different results. This can be illustrated for the case of the oxygen vacancy. According to the data of Kohan *et al.* the +2/0 transition for this defect should be located in the vicinity of the valence band maximum (VBM),[2] while the corrected data by Zhang *et al.* predict the same transition to occur just below the conduction band minimum (CBM).[3] Since it is difficult to assess the reliability of these predictions, quantitatively more reliable calculations are required.

Recently, some defect calculations were carried out using the semiempirical LDA+U scheme,[13] which allows one to adjust the position of d-electron levels by implementing self-interaction corrections into the LDA or GGA exchange-correlation potentials. Hitherto, this scheme has been employed to study point defects in $CuInSe_2$, where the Cu 3d electrons play a similar role as the Zn 3d electrons in ZnO,[14] and in calculations of optical transition levels of the oxygen vacancy in ZnO.[15,16] Therefore the method is an excellent candidate for a reassessment of the thermodynamics of point defects in zinc oxide. Another issue, which has hardly been addressed in studies of point defects in zinc oxide so far, is the role of volume relaxation and finite-size effects. It is, however, well-known, that formation enthalpies can converge slowly with supercell size,[17] especially if charged defects are considered.[18]

The purpose of the present work is twofold. First, we seek to determine formation enthalpies for the intrinsic point defects of zinc oxide by taking into account the role of Zn 3d electrons. Furthermore, we study the effect of supercell size and volume relaxation by employing finite-size scaling. Thereby, we are also able to obtain defect formation volumes for point defects in ZnO. In summary, by taking into account the band structure as well as finite-size effects, this study aims to provide a consistent set of point defect properties, which will allow for a more quantitative interpretation of experimental data.

In the following section, we summarize some observations on the band structure of zinc oxide based on experimental as well as theoretical studies. This overview allows us to motivate our computational approach which is described

 ©2006 The American Physical Society

in Sec. III. The results are compiled in Sec. IV. An interpretation and comparison with literature is given in Sec. V and the paper is concluded in Sec. VI.

II. BAND STRUCTURE OF ZINC OXIDE

Experimentally, the electronic structure of zinc oxide has been investigated in some detail (see Ref. 19 and references therein). Typically, the density of states reveals two primary bands between 0 and −10 eV (measured from the valence band maximum). The upper band is primarily derived from O $2p$ and Zn $4s$ orbitals, while the lower band arises almost solely from Zn $3d$ electrons with a maximum between −7 and −8 eV.[20,21] From x-ray photoelectron spectra the admixture of Zn $3d$ states in the O $2p$ band has been determined to be about 9% indicating a small covalent contribution to bonding.[22] The band dispersion has been investigated via angle-resolved photoelectron spectroscopy along a few high-symmetry directions.[23,24] The measurements reveal a strong dispersion of the upper valence bands and a smaller dispersion of the Zn $3d$ levels. Zinc oxide displays a direct band gap of about 3.4 eV at the Γ-point.

In general, DFT calculations yield too small band gaps compared to experiment. This effect is further enhanced in ZnO due to the underestimation of the repulsion between the Zn $3d$ and conduction band levels,[21] which leads to a significant hybridization of the O $2p$ and Zn $3d$ levels[21] and eventually to an overestimation of covalency.[25] Schröer *et al.* have performed an analysis of the wave functions obtained from self-consistent pseudopotential calculations and determined a contribution of 20%–30% of the Zn $3d$ orbitals to the levels in the upper valence band[26] (to be compared with the experimental estimate of 9% covalency cited above[22]). For zinc oxide the band gap calculated with LDA or GGA is about 0.7–0.9 eV, which is just about 25% of the experimental value (3.4 eV).[2–5,21,26]

The insufficient description of strongly localized electrons (such as those occupying the Zn $3d$ states in ZnO) and the underestimation of their binding energies is a generic problem of DFT within the LDA or GGA, and at least partially a result of unphysical self-interactions.[27] In fact, it has been found that calculations based on the Hartree-Fock or the GW approximation give much more tightly bound d-electrons and significantly larger band gaps.[21,28] An alternative approach is the explicit correction of self-interaction.[29] Vogel *et al.* have developed this idea further and devised a scheme, which allows one to incorporate self-interaction corrections (SIC) and electronic relaxation corrections (SIRC) already during the construction of pseudopotentials (PP).[27] Thereby, they were able to reproduce the experimental band gap as well as the position of the $3d$ levels in several II-VI compounds with remarkable precision. In fact, the thus obtained band structure for ZnO compares better with experiment than calculations within the Hartree-Fock and GW approximations.[21,28] SIC-PPs have also been used by Zhang *et al.* in the calculation of the formation enthalpies of a few neutral point defects in zinc oxide.[3] Unfortunately, as they point out, the SIC scheme cannot be transferred unambiguously to charged defect calculations and is therefore not applicable in the present situation.

III. METHODOLOGY

A. Calibration of GGA+U method

The problems related to tightly bound electrons within LDA and GGA-DFT have motivated the development of the so-called LDA+U (or equivalently GGA+U) method.[13,30–32] In this scheme self-interaction corrections are included heuristically by considering the (orbital dependent) on-site repulsion between electrons. The scheme has been quite successful in describing the electronic properties of several transition metal oxides for which normal LDA and GGA calculations fail to reproduce the experimentally observed ground states.[13,30–32] The method has also been employed for studying ferromagnetism in ZnO codoped with transition metals[33] and for calculating absorption spectra of nanostructured ZnO.[34] The work by Zhao *et al.*[14] is of particular interest in the present context as it illustrates the applicability of the LDA+U approach to the study of point defects. These authors employed the LDA+U method to adjust the position of the Cu $3d$ levels in $CuInSe_2$ by tuning the self-interaction parameter U. More recently, the LDA+U method has also been employed in the study of the optical transitions of the oxygen vacancy in zinc oxide.[15,16] In the present work we apply a similar scheme to obtain a more realistic representation of the Zn $3d$ electrons in zinc oxide.

We begin by calibrating and benchmarking the GGA+U method. In the present work we have adopted the formulation by Dudarev *et al.*[32] In this approach, the only free parameter is the difference $\bar{U}-\bar{J}$ between the matrix elements of the screened Coulomb electron-electron interaction. We applied the on-site interaction correction for the Zn $3d$ electrons only and tuned the position of the Zn $3d$ bands by adjusting the difference $\bar{U}-\bar{J}$. The calculations were carried out within density-functional theory (DFT) as implemented in the Vienna *ab initio* simulation package (VASP)[35] using the projector-augmented wave (PAW) method.[36,37] The parameterization by Perdew and Wang (PW91)[38] was used for the exchange-correlation functional. A nonshifted Γ-point centered $4\times4\times4$ k-point mesh was used for Brillouin zone sampling. In all calculations, the plane-wave energy cutoff was set to 500 eV giving a technical error due to discretization of the Brillouin zone integrals and incomplete basis set of less than 1 meV. The equilibrium configuration was determined by calculating the energy-volume curve. At each volume the full structure was optimized until the forces were converged to better than 5 meV/Å. As experimental data on band dispersion are sparse (see Sec. II), we have decided to benchmark our calculations using the band structure calculated by Vogel *et al.* employing their self-interaction and relaxation-corrected (SIRC) pseudopotentials, which reproduces the experimental band gap as well as the position of the Zn $3d$ levels.[27]

Values for $\bar{U}-\bar{J}$ between 0 and 10 eV were considered. As $\bar{U}-\bar{J}$ is raised the Zn $3d$ states are shifted downwards and the band gap increases. At the same time the equilibrium volume decreases while the bulk modulus varies only slightly. Eventually, we settled for $\bar{U}-\bar{J}=7.5$ eV. With this value the valence band energy levels as well as the position of the Zn $3d$

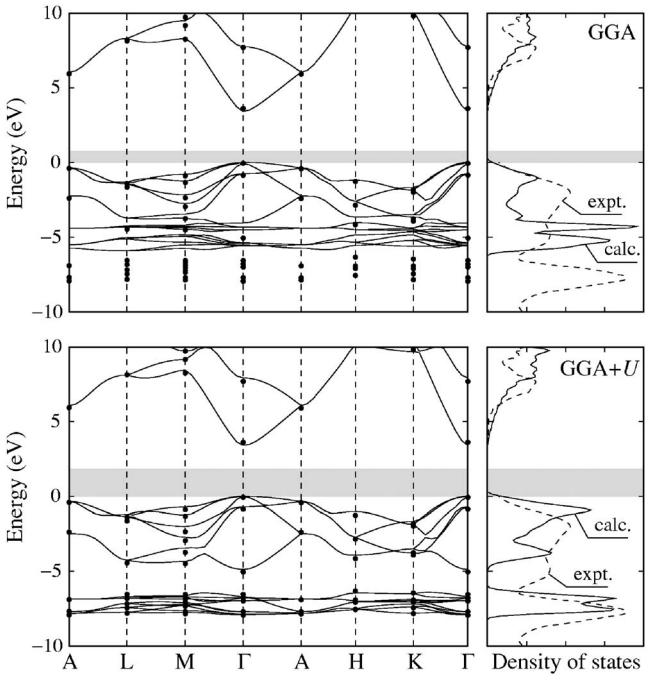

FIG. 1. Band structures obtained from density-functional theory calculations within the generalized-gradient approximation (GGA) (top) and using the GGA+U method with $\bar{U}-\bar{J}=7.5$ eV (bottom). The conduction band states have been rigidly shifted to the experimental band gap. The small black circles represent data from self-interaction and relaxation corrected (SIRC) pseudopotential calculations (Ref. 27). In the plots on the right the solid and dashed lines show the calculated and experimental (Ref. 19) density of states, respectively. The gray stripe indicates the calculated band gap.

levels are in very good agreement with the SIRC calculations and experiment. The band gap calculated using the GGA+U scheme is 1.83 eV, which constitutes a significant improvement over GGA. With otherwise identical parameters the latter yields a band gap of 0.75 eV.

The thus obtained band structure compares very well with the SIRC data from Vogel *et al.*[27] as shown in Fig. 1. (For better visualization, the conduction band has been rigidly shifted upwards by the remaining band-gap error.) The band structure obtained with $\bar{U}-\bar{J}=0$ eV, equivalent to a noncorrected GGA calculation, is included for comparison. The differences are quite striking, in particular with respect to the position of the lowest band which is predominantly derived from Zn 3d states. The right-hand panel compares the calculated density of states (DOS) with experimental data measured via photoelectron spectroscopy.[19] (Since the excitation probabilities for s, p, and d electrons change with $h\nu$ the experimental DOS varies strongly with the energy of the incoming photons.[20] In order to simplify comparison some of the bands have been rescaled). Compared to experiment, the GGA+U calculation yields a slightly smaller dispersion of the upper valence band and a yet smaller difference in the position of the Zn 3d band.[54] The calculation reproduces the double feature in the Zn 3d bands reported by Zwicker and Jacobi[23] as well as Girard *et al.*[24] Overall, the agreement of the GGA+U band structure and density of states with SIRC calculations as well as experiments is very good.

B. Defect calculations

For the defect calculations we used hexagonal supercells with 32–108 atoms equivalent to $2\times2\times2$ and $3\times3\times3$, primitive unit cells, respectively. In these calculations a nonshifted Γ-point centered $2\times2\times2$ k-point mesh was employed. All calculations were performed within GGA as well as GGA+U in order to quantify the energy corrections.

First, the atomic positions were relaxed with all cell parameters fixed at the ideal bulk values (Sec. IV A) until the forces were converged to better than 5 meV/Å. We then computed the energy-volume curves for the supercells with fixed atomic coordinates. The minimum of the curve provided the volume of the defective cell, V_D, which allowed us to calculate the formation volume according to (see, e.g., Refs. 39–41)

$$V_D^f = V_D - \frac{N_D}{N_{id}} V_{id}. \tag{1}$$

Here, V_{id} and N_{id} denote the volume and the number of atoms in the ideal reference system, and N_D denotes the number of atoms in the defective cell. The data were then subjected to finite-size scaling in a similar manner as the formation enthalpies. The formation volume determines the pressure dependence of the formation enthalpy ($V_D^f=-\partial G_D^f/\partial p$).

Earlier DFT calculations within LDA and GGA consistently gave very high formation enthalpies for antisite defects[2–4] and we have no reason to assume that their formation enthalpies would be significantly lowered within GGA+U. After all, we have considered the following point defects: the oxygen vacancy (V_O), the octahedral oxygen interstitial ($O_{i,oct}$), the dumbbell oxygen interstitial ($O_{i,db}$), the rotated dumbbell interstitial ($O_{i,db\text{-}rot}$), the zinc vacancy (V_{Zn}), and the octahedral zinc interstitial ($Zn_{i,oct}$). With the exception of the latter the geometric structures of these defects have been described in our preceding study.[5] For charged defects a homogeneous compensating background charge was added (see, e.g., Ref. 18).

The variation of the formation enthalpies with the chemical potential of the reservoir and the Fermi level was calculated using the thermodynamic formalism established in Refs. 42 and 43. The defect formation enthalpy of a defect in charge state q is calculated according to

$$H_D^f = E_D - \frac{1}{2}(n_{Zn}+n_O)\mu_{ZnO}^{bulk} - \frac{1}{2}(n_{Zn}-n_O)(\mu_{Zn}^{bulk}-\mu_O^{bulk}) - q(E_{VBM}+\mu_e) - \frac{1}{2}(n_{Zn}-n_O)\Delta\mu, \tag{2}$$

where E_D is the total energy of the system in the presence of the defect, n_i is the number of atoms of atom type i, and μ_i^{bulk} is the chemical potential of the pure constituent i in its reference state. The valence band maximum (VBM) and the (electro-)chemical potential of the electrons (Fermi energy) are denoted by E_{VBM} and μ_e, respectively. The last term describes the range within which the chemical potentials can vary. The quantity $\Delta\mu$ is restricted by the formation enthalpy ΔH_f of wurtzitic zinc oxide by $|\Delta\mu|\leq|\Delta H_f|$; zinc- and

oxygen-rich conditions correspond to $\Delta\mu=-\Delta H_f$ and $\Delta\mu=\Delta H_f$, respectively. In the zero temperature limit the chemical potentials of gaseous oxygen, solid zinc, and zinc oxide are given by the cohesive energies.

Knowledge of the formation enthalpies of the fully relaxed defects allows us to derive the thermal (equilibrium) transition level between charge states q_1 and q_2 according to

$$\varepsilon = -\frac{H_D^f(q_1) - H_D^f(q_2)}{q_1 - q_2}, \tag{3}$$

where $H_D^f(q_1)$ and $H_D^f(q_2)$ denote the formation enthalpies at the valence band maximum for charge states q_1 and q_2, respectively.

IV. RESULTS

A. Ground state properties

The crystallographic parameters for ZnO in the wurtzite structure calculated by the GGA+U method are $a=3.196$ Å, $c/a=1.606$, and $u=0.381$ which compare very well with the experimental values $a=3.24$ Å, $c/a=1.600$, and $u=0.382$ (experimental data cited in this section from Refs. 44–48). From fitting the energy-volume data to the Birch-Murnaghan equation of state[49] a bulk modulus of 136 GPa was determined in good agreement with the experimental value of 143 GPa. The calculation gave a formation enthalpy of $\Delta H_f=-3.46$ eV/f.u., which compares very well with the experimental value of $\Delta H_f=-3.58$ eV/f.u. (f.u.=formula unit). The GGA calculations gave $a=3.283$ Å, $c/a=1.611$, $u=0.378$, $B=149$ GPa, and $\Delta H_f=-3.55$ eV/f.u.

Furthermore, we calculated the cohesive energies of pure oxygen and zinc, since they enter the calculation of the formation energy of zinc oxide as well as the defect formation enthalpies. For hcp-zinc we obtained a cohesive energy of $E_c=-1.115$ eV/atom, and lattice constants of $a=2.641$ Å and $c/a=1.930$ (experimental values: $E_c=-1.359$ eV/atom, $a=2.660$ Å, $c/a=1.828$); for the oxygen dimer the calculation yielded a dimer energy of $D_0=8.80$ eV and a bond length of $r_0=1.238$ Å (experimental values: $D_0=5.166$ eV and $r_0=1.208$ Å).

B. Finite-size scaling

Strain interactions typically scale as $O(V^{-1})=O(N^{-1})=O(c)$, where V is the supercell volume, N is the number of atoms, and c is the defect concentration in the supercell. Accordingly, in order to obtain the formation enthalpies at infinite dilution, results obtained from calculations with supercells comprising 32, 48, 72, and 108 atoms were extrapolated. For neutral defects the data can well be fitted by a straight line as shown in Fig. 2. The formation enthalpy at infinite dilution is then given by the abscissa of the linear fit.

For charged defects, Makov and Payne have shown that additional energy contributions have to be included in order to correct for image charge interactions.[50] They derived the correction in the form of a multipole expansion. The first term in this expansion corresponds to monopole-monopole interactions, which scale as $O(V^{-1/3})$, and was explicitly taken into account. The magnitude of this correction is shown in Fig. 2.

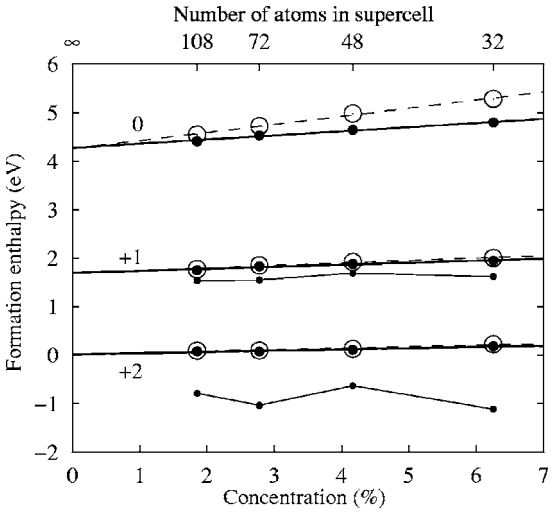

FIG. 2. Scaling behavior of formation enthalpies with concentration (equivalent to the inverse cell volume) exemplified for the case of the zinc interstitial ($Zn_{i,oct}$). For the charged defects the small circles show the data before monopole-monopole corrections are applied (Ref. 50). Large closed circles correspond to formation enthalpies after volume relaxation while large open circles show the data obtained at fixed volume. The data presented here have been computed by the GGA+U method.

Since dipole-dipole and monopole-quadrupole interactions scale as $O(V^{-1})=O(N^{-1})=O(c)$ and thus show the same scaling behavior as the interactions of the strain fields, they are implicitly considered through the finite-size scaling procedure. Figure 2 shows that indeed after correction of the monopole-monopole term the data are well-described by a linear fit.

C. Formation enthalpies

Formation enthalpies calculated by the GGA+U method are given in Table I. This compilation of data also includes the results of our GGA calculations and the data of previous DFT studies. For most defects and charge states the extrapolation errors are very small confirming the reliability of the procedure employed. In these cases the multipole-expansion-based correction scheme works excellently.[50] Larger errors arise for oxygen dumbbell interstitials in charge states ±2 and for the octahedral oxygen interstitial in charge state +2. For the dumbbell defects ($O_{i,db}$, $O_{i,db\text{-}rot}$) the defect core comprises two atoms[5] which renders them distinct from the other defects considered here. Apparently, higher order moments would be required in order to fully capture these features. For the octahedral interstitial ($O_{i,oct}$) in charge state +2 analysis of the electron density shows significant charge delocalization which cannot be described by a *finite* multipole expansion.[50]

If we compare the difference between our GGA and GGA+U results with the corrections calculated by Zhang *et al.*,[3] we find agreement of the general trends. The differences between the GGA and GGA+U are smaller than the predic-

TABLE I. Calculated formation enthalpies for point defects in bulk zinc oxide for zinc-rich and *p*-type conducting conditions (μ_e =0 eV, VBM); the second column under GGA and GGA+*U* gives the error of the extrapolation to infinite dilution. Ref. 2; DFT, LDA, ultrasoft PP; Ref. 3: DFT, LDA, norm-conserving PP; Ref. 4: DFT, GGA, ultrasoft PP; and Ref. 5: DFT, LDA, norm-conserving PP.

		This work					Ref. 3			
Defect	Charge state	GGA		GGA+*U*		Ref. 2	Uncorrected	Corrected	Ref. 4[a]	Ref. 5
$Zn_{i,oct}$	0	2.50	0.03	4.25	0.03	1.7	3.4	6.2	1.2	
	+1	0.98	0.02	1.69	0.03	1.3	1.5	2.1	⩾0.4	
	+2	0.33	0.07	0.02	0.03	0.9	−0.2	−2.3	−0.6	
V_O	0	1.00	0.06	1.71	0.04	0.0	1.5	2.4		0.9
	+1	0.26	0.03	0.71	0.03	0.2	0.8	1.5		
	+2	−0.48	0.02	−0.73	0.03	−0.3	−0.5	−3.0	−0.9	−0.5
$O_{i,db}$	0	4.61	0.05	4.70	0.07					5.1
	+1	4.76	0.07	4.59	0.02					5.1
	+2	5.36	0.33	5.08	0.27					5.2
$O_{i,db\text{-}rot}$	−2	7.70	0.40	8.79	0.17				[8.2][b]	7.2
	−1	6.51	0.28	7.08	0.04	[7.5][b]			[⩾7.1][b]	6.6
	0	4.87	0.03	4.96	0.04	[6.5][b]			[6.0][b]	5.2
	+1	5.07	0.07	4.91	0.05	[6.5][b]				5.3
	+2	5.67	0.72	5.41	0.29					5.4
$O_{i,oct}$	−2	7.84	0.13	8.97	0.07	7.8	[7.4][c]	[9.7][c]	7.8	7.4
	−1	6.65	0.05	7.33	0.03	6.8	[6.4][c]	[10.4][c]	6.9	6.7
	0	6.20	0.03	6.60	0.03	6.4	[6.2][c]	[12.1][c]	6.4	6.2
	+1	6.36	0.11	6.60	0.10	6.4				6.3
	+2	6.95	0.37	7.09	0.34					6.3
V_{Zn}	−2	6.32	0.11	7.06	0.05	6.6	5.8	10.1	5.1	5.9
	−1	5.57	0.08	5.96	0.05	5.8	5.7	10.1	5.0	5.8
	0	5.35	0.04	5.60	0.01	6.0	5.8	10.6	⩾5.1	6.0

[a]The data given here were derived from Fig. 1 in the original reference, since no explicit values are given.
[b]References 2 and 4 report formation enthalpies for a "tetrahedral interstitial" configuration but no details on the geometry of the relaxed configuration are given.
[c]The geometry of the oxygen interstitial is not specified in Ref. 3.

tions of the correction scheme of Zhang *et al.*,[3] which is, however, consistent with the band-gap underestimation still present in the GGA+*U* calculations. We also point out that for the oxygen vacancy and the zinc interstitial (where comparison is possible) the trends agree with the SIC-PP calculations.[3]

Our results yield further support for the qualitative argumentation put forth by Zhang and co-workers regarding the trends expected upon band-gap correction.[12] The formation enthalpy correction is negative for positively charged defects, positive for neutral, and yet more positive for negatively charged defects.

D. Geometries

For almost all defects the dependence of the atomic displacements on the supercell size is quite small giving evidence that the strain fields are rather short ranged. For the vacancies and the octahedral interstitials the relaxations maintain the threefold symmetry axis of the lattice. For V_{Zn}, $O_{i,oct}$, and $Zn_{i,oct}$ the displacements change continuously with the addition or subtraction of electrons. The observations regarding the charge-state dependent relaxation behavior of the oxygen vacancy described in Refs. 5, 15, and 16 are confirmed. For the neutral and positively charged oxygen dumbbell the oxygen-oxygen bond length is found to be between 15% ($q=+2$) and 23% ($q=0$) longer than the calculated dimer bond length, which is in full agreement with our previous calculation.[5]

E. Formation volumes

The defect formation volumes obtained via Eq. (1) and subjected to finite-size scaling are given in Table II and plotted as a function of charge state in Fig. 3. With the exception of the oxygen vacancy all defects display the same trend. As electrons are added to the system the formation volume rises linearly. The slope for the oxygen interstitials as well as the zinc vacancy is roughly $-0.37\ \Omega_{f.u.}/e$ while for the zinc interstitial it amounts to about $-0.56\ \Omega_{f.u.}/e$ ($\Omega_{f.u.}$ denotes the volume per formula unit). In the case of V_{Zn}, $O_{i,db}$, $O_{i,db\text{-}rot}$, and $O_{i,oct}$ electrons are added and removed predominantly

TABLE II. Point defect formation volumes calculated within GGA+U given in units of volume per formula unit ($V_D^f/\Omega_{\text{f.u.}}$). The errors of the extrapolation to infinite dilution are given in brackets.

	Charge state				
Defect	−2	−1	0	+1	+2
$Zn_{i,oct}$			0.81	0.28	−0.30
			(0.05)	(0.03)	(0.04)
V_O			−0.26	−0.18	−0.32
			(0.01)	(0.01)	(0.01)
$O_{i,db}$			0.47	0.13	−0.26
			(0.03)	(<0.01)	(<0.01)
$O_{i,db\text{-}rot}$	1.15	0.76	0.43	0.08	−0.32
	(0.10)	(0.05)	(0.01)	(0.03)	(0.01)
$O_{i,oct}$	1.05	0.68	0.33	−0.05	−0.46
	(0.08)	(0.05)	(0.01)	(0.01)	(0.01)
V_{Zn}	0.81	0.45	0.05		
	(0.03)	(0.02)	(0.06)		

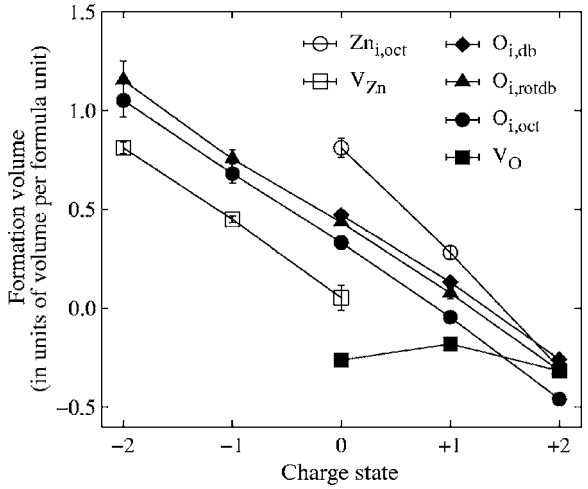

FIG. 3. Variation of defect formation volumes with charge state. Open and closed circles indicate defects on the zinc and oxygen sublattices, respectively.

from oxygenlike orbitals, while for $Zn_{i,oct}$ zinclike orbitals compensate most of the defect charge. In a simple picture the two different slopes are manifestations of the different spatial extents of the O and Zn-derived orbitals.

The oxygen vacancy behaves atypically because the local relaxation is different for the neutral and positive charge states. As explained in detail in Refs. 5, 15, and 16, Zn atoms surrounding the vacancy site in neutral charge state relax *outward*, while they show an *inward* relaxation for positive charge states. The result is a nonlinear charge state dependence of the formation volume.

V. DISCUSSION

In the following we first discuss the results of the GGA+U calculations. From the data in Table I the Fermi level dependence of the formation enthalpies (Fig. 4) and the stability map (Fig. 5) were derived. In agreement with our earlier calculations[5] three defects are found to be the most abundant: Under zinc-rich conditions the oxygen vacancy is the most likely defect for all Fermi levels; under oxygen-rich conditions the zinc vacancy and oxygen interstitials are the dominant defect types. Using Eq. (3) the defect transition levels were obtained as shown in Fig. 6(b). Earlier studies have consistently found the oxygen vacancy to be energetically slightly more favorable than the zinc interstitial. Our data imply that upon inclusion of volume relaxation this difference becomes yet larger (~0.7 eV at the VBM). For the zinc interstitial as well as for the oxygen vacancy we find transition levels close to the calculated conduction band minimum (CBM).

The *oxygen vacancy* shows a +2/0 transition 0.61 eV below the calculated CBM. This is in qualitative agreement with Zhang *et al.*[3] and Oba *et al.*[4] but contradicts the result by Kohan *et al.* who predicted the +2/0 transition of the oxygen vacancy in the vicinity of the VBM.[2] The deep level

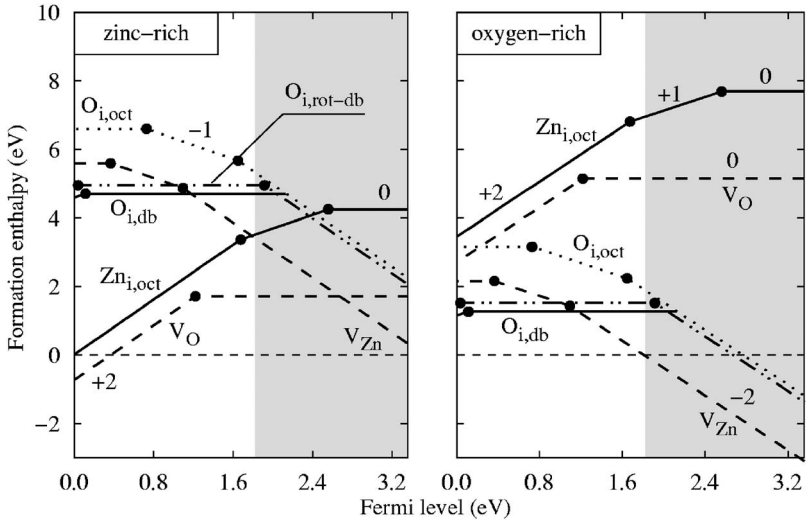

FIG. 4. Variation of defect formation enthalpies with Fermi level under zinc (left) and oxygen-rich (right) conditions as obtained from GGA+U calculations. The gray shaded area indicates the difference between the calculated and the experimental band gap. The numbers in the plot indicate the defect charge state; parallel lines imply equal charge states. Open and closed circles correspond to defects on the zinc and oxygen sublattices, respectively.

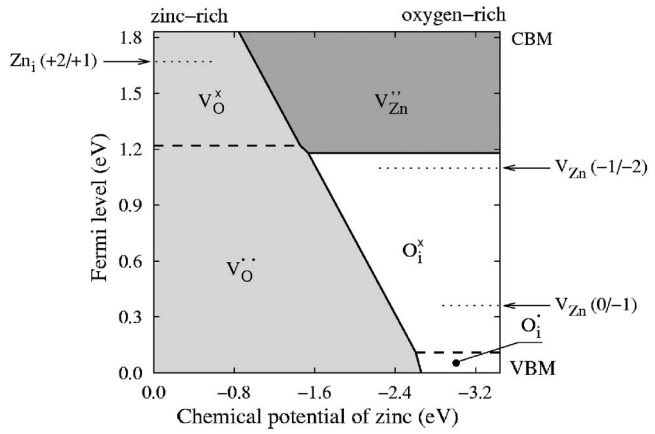

FIG. 5. Variation of dominant defect as a function of chemical potential and Fermi level as obtained from GGA+U calculations. In regions where the most stable defect is neutral, the transitions for the most stable charged defect are indicated by arrows and dotted lines.

character (strong localization of defect electrons) correlates with the significant charge dependent structural changes (see Sec. IV E and Refs. 15 and 16).

For the *zinc interstitial* our calculations locate the +2/+1 transition level 0.15 eV below the CBM. In this case, the data by Kohan *et al.* show the same charge transition, while Zhang *et al.* and Oba *et al.*[3,4] predict no transition from the +2 charge state over the entire band gap. The structure of the zinc interstitial is similar in all charge states, indicative for a shallow defect level in agreement with experimental observations.[51]

Previous studies provided evidence that dumbbell geometries ($O_{i,db}$, $O_{i,db\text{-}rot}$), in which two oxygen atoms jointly occupy a regular lattice site, are the most stable oxygen interstitial configurations.[5,6] Unlike the highly symmetric octahedral interstitial ($O_{i,oct}$) these defect configurations have comparably low formation enthalpies in positive as well as negative charge states. The present data provide a strong indication for ambipolar behavior as anticipated in Ref. 5.

For oxygen-rich conditions, oxygen dumbbell defects have the lowest formation enthalpies of all intrinsic point defects over the widest range of the (calculated) band gap confirming the results of earlier studies.[5,6] Our present data clearly prove the existence of a +1/0 transition level close to the VBM ($O_{i,db}$: 0.34 eV, $O_{i,db\text{-}rot}$: 0.02 eV above VBM). For the rotated dumbbell interstitial, the 0/−2 transition lies at the CBM indicating shallow acceptor behavior.

The *zinc vacancy* represents a good example for the importance of finite-size sampling: supercell calculations with less than 108 atoms consistently predict the neutral charge state to be unstable with respect to the negative charge states. If supercell effects are, however, properly taken into account, it turns out that the formation enthalpy of the neutral zinc vacancy at the VBM is indeed lower than for the negative charge states. This results in two transitions. The first one (0/−1) lies very close to the VBM, the second one (−1/−2) occurs near the middle of the calculated band gap. The zinc vacancy has the lowest formation enthalpy of all intrinsic point defects under oxygen-rich conditions for Fermi levels in the upper half of the band gap. Under these conditions it is therefore the dominant acceptor confirming the experimental results of Tuomisto and co-workers.[52] By using positron annihilation spectroscopy in combination with isochronal annealing cycles on electron irradiated n-type samples, they furthermore obtained evidence for a second acceptor and suggested oxygen interstitials as likely candidates. From our calculations we identify this defect to be the rotated dumbbell interstitial.

We now turn to the question to which extent the band-gap underestimation within GGA+U affects the results. Although the band gap is significantly larger with GGA+U than with GGA, it is still smaller than experiment. In the past a variety of methods has been conceived in order to correct for the band-gap error intrinsic to DFT calculations (see, e.g., Refs. 2, 3, 15, 16, and 53; also compare Secs. II and III A). As pointed out by Lany and Zunger[15] these schemes lead to somewhat different results. In the present context, we have applied the extrapolation formula applied by Janotti and Van de Walle.[16] According to this scheme the transition level, ε^{ext}, corresponding to the experimental band gap, E_G^{exp}, is obtained from the transition levels calculated within GGA ($\varepsilon^{\mathrm{GGA}}$) and GGA+$U$ ($\varepsilon^{\mathrm{GGA}+U}$) according to

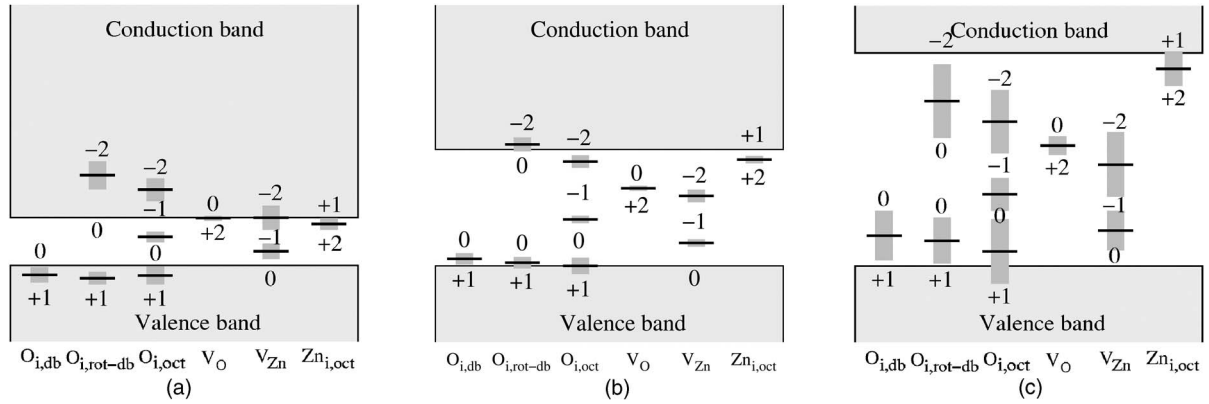

FIG. 6. Transition levels in the band gap calculated within GGA, GGA+U and using the extrapolation formula (4). The dark gray shaded areas indicate the error bars resulting from the errors given in Table I.

$$\varepsilon^{ext} = \frac{E_G^{exp} - E_G^{\mathrm{GGA}+U}}{E_G^{\mathrm{GGA}+U} - E_G^{\mathrm{GGA}}}(\varepsilon^{\mathrm{GGA}+U} - \varepsilon^{\mathrm{GGA}}) + \varepsilon^{\mathrm{GGA}+U}, \quad (4)$$

where E_G^{GGA} and $E_G^{\mathrm{GGA}+U}$ denote the GGA and GGA+U band gaps, respectively.[55] The results of this extrapolation are shown in Fig. 6 in comparison with the GGA and GGA+U data. It is apparent that the extrapolated transition levels are very sensitive to the accuracy of the GGA and GGA+U data. Nonetheless, the qualitative classification of the defect levels as well as their hierarchy, which has been established above based on the GGA+U results, is preserved in the extrapolated data.

In general, point defects will be abundant under normal conditions, since the point defect formation enthalpies in zinc oxide are very small. The resulting high defect concentrations probably do not only actively influence the electronic and optical, but to some extent also the mechanical properties. This could explain the large spread of experimental data for various materials properties.

VI. CONCLUSIONS

In the present work, we performed an extensive study of intrinsic point defects in zinc oxide in order to remedy the most significant approximations made in previous studies. To this end, we have applied the GGA+U method to correct the band structure and thereby the overestimation of covalency. As a consequence we also obtained a doubling of the band gap which is, however, still smaller than in experiments. In order to remove the effects of elastic as well as electrostatic image forces, we employed finite-size scaling which allowed us to derive more precise formation enthalpies of intrinsic point defects along with formation volumes.

Our results in some respects confirm earlier calculations but predict a larger number of transition levels for intrinsic defects. Transition levels close to the conduction band minimum are identified for both the oxygen vacancy and the zinc interstitial. The zinc vacancy possesses a transition rather close to the VBM and another one roughly in the middle of the band gap. As anticipated earlier[5] oxygen interstitials show ambipolar behavior displaying transition levels near the valence band maximum and the conduction band minimum.

ACKNOWLEDGMENTS

We acknowledge financial support through the *Sonderforschungsbereich 595* "Fatigue in functional materials" of the *Deutsche Forschungsgemeinschaft* and grants of computer time by the Center for Scientific Computing at the Johann Wolfgang Goethe-University, Frankfurt/Main.

[1] U. Özgür, Y. I. Alivov, C. Liu, A. Teke, M. Reshchikov, S. Doğan, V. Avrutin, S. J. Cho, and H. Morkoç, J. Appl. Phys. **98**, 041301 (2005).
[2] A. F. Kohan, G. Ceder, D. Morgan, and Chris G. Van de Walle, Phys. Rev. B **61**, 15019 (2000).
[3] S. B. Zhang, S. H. Wei, and A. Zunger, Phys. Rev. B **63**, 075205 (2001).
[4] F. Oba, S. R. Nishitani, S. Isotani, H. Adachi, and I. Tanaka, J. Appl. Phys. **90**, 824 (2001).
[5] P. Erhart, A. Klein, and K. Albe, Phys. Rev. B **72**, 085213 (2005).
[6] E.-C. Lee, Y.-S. Kim, Y.-G. Jin, and K. J. Chang, Phys. Rev. B **64**, 085120 (2001).
[7] Y. Yan, S. B. Zhang, S. J. Pennycook, and S. T. Pantelides, Mater. Res. Soc. Symp. Proc. **666**, F2.6 (2001).
[8] Y. Yan, S. B. Zhang, and S. T. Pantelides, Phys. Rev. Lett. **86**, 5723 (2001).
[9] S. Limpijumnong, S. B. Zhang, S.-H. Wei, and C. H. Park, Phys. Rev. Lett. **92**, 155504 (2004).
[10] M. S. Hybertsen and S. G. Louie, Phys. Rev. Lett. **55**, 1418 (1985).
[11] B. Arnaud and M. Alouani, Phys. Rev. B **62**, 4464 (2000).
[12] S. B. Zhang, S.-H. Wei, and A. Zunger, Phys. Rev. Lett. **84**, 1232 (2000).
[13] V. I. Anisimov, J. Zaanen, and O. K. Andersen, Phys. Rev. B **44**, 943 (1991).
[14] Y.-J. Zhao, C. Persson, S. Lany, and A. Zunger, Appl. Phys. Lett. **85**, 5860 (2004).
[15] S. Lany and A. Zunger, Phys. Rev. B **72**, 035215 (2005).
[16] A. Janotti and C. G. Van de Walle, Appl. Phys. Lett. **87**, 122102 (2005).
[17] M. J. Puska, S. Pöykkö, M. Pesola, and R. M. Nieminen, Phys. Rev. B **58**, 1318 (1998).
[18] J. Lento, J.-L. Mozos, and R. M. Nieminen, J. Phys.: Condens. Matter **14**, 2637 (2002).
[19] R. Blachnik, J. Chu, R. R. Galazka, J. Geurts, J. Gutowski, B. Hönerlage, D. Hofmann, J. Kossut, R. Lévy, P. Michler, U. Neukirch, T. Story, D. Strauch, and A. Waag, *Landolt-Börnstein: Numerical Data and Functional Relationships in Science and Technology, New Series*, Vol. III/41B (Springer, Heidelberg, 1999).
[20] W. Ranke, Solid State Commun. **19**, 685 (1976).
[21] M. Usuda, N. Hamada, T. Kotani, and M. van Schilfgaarde, Phys. Rev. B **66**, 125101 (2002).
[22] S. A. Leontiev, S. V. Koshtcheev, V. G. Devyatov, A. E. Cherkashin, and E. P. Mikheeva, J. Struct. Chem. **38**, 725 (1997).
[23] G. Zwicker and K. Jacobi, Solid State Commun. **54**, 701 (1985).
[24] R. T. Girard, O. Tjernberg, G. Chiaia, S. Söderholm, U. O. Karlsson, C. Wigren, H. Nylén, and I. Lindau, Surf. Sci. **373**, 409 (1997).
[25] Faraday Discuss. **124**, 275 (2003).
[26] P. Schröer, P. Krüger, and J. Pollmann, Phys. Rev. B **47**, 6971 (1993).
[27] D. Vogel, P. Krüger, and J. Pollmann, Phys. Rev. B **54**, 5495 (1996).
[28] S. Massidda, R. Resta, M. Posternak, and A. Baldereschi, Phys. Rev. B **52**, R16977 (1995).
[29] J. P. Perdew and A. Zunger, Phys. Rev. B **23**, 5048 (1981).

[30] I. V. Solovyev, P. H. Dederichs, and V. I. Anisimov, Phys. Rev. B **50**, 16861 (1994).
[31] A. I. Liechtenstein, V. I. Anisimov, and J. Zaanen, Phys. Rev. B **52**, R5467 (1995).
[32] S. L. Dudarev, G. A. Botton, S. Y. Savrasov, C. J. Humphreys, and A. P. Sutton, Phys. Rev. B **57**, 1505 (1998).
[33] M. S. Park and B. I. Min, Phys. Rev. B **68**, 224436 (2003).
[34] C. L. Dong, C. Persson, L. Vayssieres, A. Augustsson, T. Schmitt, M. Mattesini, R. Ahuja, C. L. Chang, and J.-H. Guo, Phys. Rev. B **70**, 195325 (2004).
[35] G. Kresse and J. Furthmüller, Phys. Rev. B **54**, 11169 (1996).
[36] P. E. Blöchl, Phys. Rev. B **50**, 17953 (1994).
[37] G. Kresse and D. Joubert, Phys. Rev. B **59**, 1758 (1999).
[38] J. P. Perdew, in *Electronic Structure of Solids*, edited by P. Ziesche and H. Eschrig (Akademie Verlag, Berlin, 1991).
[39] C. P. Flynn, *Point Defects and Diffusion* (Oxford University Press, Oxford, 1972).
[40] P. Ehrhart, P. Jung, H. Schultz, and H. Ullmaier, *Landolt-Börnstein: Numerical Data and Functional Relationships in Science and Technology, New Series*, Vol. III/25 (Springer, Heidelberg, 1991).
[41] S. A. Centoni, B. Sadigh, G. H. Gilmer, T. J. Lenosky, T. Diaz de la Rubia, and C. B. Musgrave, Phys. Rev. B **72**, 195206 (2005).
[42] G.-X. Qian, R. M. Martin, and D. J. Chadi, Phys. Rev. B **38**, 7649 (1988).
[43] S. B. Zhang and J. E. Northrup, Phys. Rev. Lett. **67**, 2339 (1991).
[44] J. Albertsson, S. C. Abrahams, and A. Kvick, Acta Crystallogr., Sect. B: Struct. Sci. **45**, 34 (1989).
[45] *Handbook of Chemistry and Physics*, 85th ed., edited by D. R. Lide (CRC Press, Boca Raton, FL., 2004).
[46] A. Every and A. McCurdy, *Landolt-Börnstein: Numerical Data and Functional Relationships in Science and Technology, New Series*, Vol. III/29A (Springer, Heidelberg, 1992).
[47] C. Kittel, *Introduction to Solid State Physics*, 8th ed. (Wiley, New York, 2004).
[48] K. P. Huber and G. Herzberg, *Constants of Diatomic Molecules* (Van Nostrand, New York, 1979).
[49] F. Birch, J. Geophys. Res. **83**, 1257 (1978).
[50] G. Makov and M. C. Payne, Phys. Rev. B **51**, 4014 (1995).
[51] D. C. Look, J. W. Hemsky, and J. R. Sizelove, Phys. Rev. Lett. **82**, 2552 (1999).
[52] F. Tuomisto, V. Ranki, K. Saarinen, and D. C. Look, Phys. Rev. Lett. **91**, 205502 (2003).
[53] C. Persson, Y.-J. Zhao, S. Lany, and A. Zunger, Phys. Rev. B **72**, 035211 (2005).
[54] With respect to numerical accuracy in particular regarding the Zn 3*d* levels, it must be acknowledged that the experimental data display some variation as values in the range between −7 and −8 eV have been reported (Refs. 19, 23, and 24).
[55] The underlying idea of the extrapolation scheme used by Janotti and Van de Walle resembles the strategy presented by Zhang *et al.* in Ref. 12. The calculated band gap is varied by changing some parameter of the calculation (Zhang *et al.* use the plane-wave cutoff energy, Janotti and Van de Walle use LDA and LDA+*U*). Finally, the formation enthalpies or transition energies corresponding to the full (experimental) band gap are obtained by extrapolating over the calculated band gap.