10
Critical Evaluation of the LDA + *U* Approach for Band Gap Corrections in Point Defect Calculations: The Oxygen Vacancy in ZnO Case Study

Adisak Boonchun and Walter R. L. Lambrecht

10.1
Introduction

The local density approximation (LDA) is well known to underestimate band gaps in semiconductors. In a recent paper, Paudel and Lambrecht [1] (PL), discussed this problem for the oxygen vacancy V_O in ZnO. The problem appears to be rather dramatic in this case as different authors do not even agree on whether the relevant defect level, the $2+/0$ transition level lies in the upper or the lower half of the band gap. Discrepancies also exist between different authors on the magnitude of the energy of formation of the defect and on the positions of the one-electron levels. At the time that paper was written, previous work had addressed the gap corrections mostly in *a posteriori* fashion [2–7]. The point of the PL paper was to use an adjusted Hamiltonian and total energy functional that gave the correct band gap for the host and then apply it to the defect transition levels. In particular, they used the LDA + U approach with U Coulomb interactions not only for the d states of Zn but also the s orbitals. The idea behind this unorthodox application of LDA + U is explained below. Since that work and even before, the problem has been investigated by several others by a variety of approaches, hybrid functionals [8–10], the GW method [9], and screened exchange (Chapter 5, [11, 12]). Here, we critically re-examine the results of PL and explore the LDA + U model further with additional U_i parameters, which overcome some of the shortcomings of the previous $U_d + U_s$ model. We compare the results from various groups for this benchmark case and look for what consensus can be reached and what remain open questions.

We first briefly remind the reader of the nature of the LDA + U method. Then we discuss some side issues, such as the potential alignment and image charge corrections. We then present the results of a new LDA + U potential and end with an overview of the various results on the oxygen vacancy in ZnO.

Advanced Calculations for Defects in Materials: Electronic Structure Methods, First Edition.
Edited by Audrius Alkauskas, Peter Deák, Jörg Neugebauer, Alfredo Pasquarello, and Chris G. Van de Walle.

10.2
LDA + *U* Basics

The LDA + *U* approach was originally introduced to deal with electrons in localized orbitals for which the standard density functional approach [13, 14] in the local (spin) density approximation (L(S)DA) is not sufficient. The main emphasis was on open-shell systems such as transition metal compounds and rare-earth metals and compounds. The original versions were strictly LDA + *U* rather than LSDA + *U* with all the magnetic effects arising from the Hubbard like on-site Coulomb terms that were added to the LDA Hamiltonian [15, 16]. A key aspect of LDA + *U* is that since LDA already contains some exchange and correlation in an orbital-independent way, a double counting correction is required when explicitly adding the Coulomb and exchange effects for these localized orbitals. While the first version [15] used an *around mean-field* approach in which the LDA is supposed to give the right answer for equal occupation of all the d orbitals, the more often used version is the so-called "fully localized limit," (FLL). In this initial discussion, we refer to the orbitals for which *U* effects are added as the d orbitals although we later will generalize this. In the FLL we assume that the LDA gives the correct total energy for the atomic limit of integer occupations of the specific d_i orbitals. They are either fully occupied or empty. What the LDA + *U* framework provides in that case, is how starting from such an atomic limit the interaction with the other bands, modifies these occupation numbers and may lead to orbital ordering. A key aspect of the LDA + *U* model is that the total energy is treated as a functional of the electron density as well as separately as a function of the occupation numbers of d orbitals, $E[n(\mathbf{r}), n_i]$. In its simplest form [16] it is given by

$$E_{\mathrm{LDA}+U} = E_{\mathrm{LDA}} - UN(N-1)/2 + U\sum_{i,j} n_i n_j. \tag{10.1}$$

The potential for the *i*-th orbital $V_i = \delta E/\delta n_i(\mathbf{r})$ with $n_i(\mathbf{r}) = |\psi_i(\mathbf{r})|^2$ then becomes

$$V_i = V_{\mathrm{LDA}} + U\left(\frac{1}{2} - n_i\right). \tag{10.2}$$

Thus the one-electron levels

$$\varepsilon_i = \partial E/\partial n_i = \varepsilon_i^{\mathrm{LDA}} + U\left(\frac{1}{2} - n_i\right) \tag{10.3}$$

are shifted from the LDA even though the LDA + *U* total energy in the limit of integer n_i remains in principle the same as the LDA total energy.

The version we presently use starts from LSDA and besides *U*, contains in general non-spherical Coulomb terms written in terms of the Slater F_k integrals and Clebsch–Gordan coefficients, specific to which spherical harmonic d_i orbitals are occupied. Also since those orbitals depend on the coordinate system, the method is formulated in a rotationally invariant form in terms of density matrices n_{ij}. This form of the LSDA + *U* method is described by Liechtenstein *et al.* [17]. It comes down to a

Hartree–Fock like treatment of the localized orbitals with a parametrized treatment of their Coulomb and exchange interactions, in particular using a screened Coulomb interaction U. Hence, one can expect some similarities between the results of this approach and hybrid functionals.

Our calculations are all carried out within a full-potential linearized muffin–tin orbital (FP-LMTO) method [18, 19]. This method uses smoothed Hankel functions as envelope functions, which are augmented inside muffin–tin spheres (in terms of ϕ and $\dot{\phi} = \mathrm{d}\phi/\mathrm{d}E$ functions, with ϕ solutions of the radial Schrödinger equation), as usual in linear methods. We use a double-κ basis set with optimized smoothing radii and radial decay constants κ including spdf functions in the basis set for the first κ and sp for the second one. In addition, we treat Zn-3d orbitals as valence bands and add 4d-local orbitals for a better description of the conduction bands. For the augmentation we use an angular momentum cut-off of $l_{\max} = 4$. We use a 127 atom supercell for which Γ-point sampling of the Brillouin zone is adequate. While the basis set corresponds to a muffin–tin potential, it should be emphasized that the full non-spherical potential inside the spheres and non-constant potential in the interstitial region is treated. Forces are calculated analytically and allow us to optimize the structure.

For the present purpose, dealing with ZnO, the semicore 3d states are completely filled, so the LDA + U treatment simply results in a downward shift of those orbitals by $U_\mathrm{d}/2$. At the same time, we note that if U_s is applied to the s orbitals of Zn, which primarily constitute the conduction band minimum and are thus almost empty, they will shift up approximately by $U_\mathrm{s}/2$. So, in a strictly empirical manner, we can adjust U_d so as to shift the d-levels down to where they are found by photoemission and U_s so as to open the correct gap at Γ.

It should be kept in mind that the justification for adding U_s is different than for the d orbitals. The physics of the d-states, is indeed that they are strongly localized and have strong Coulomb interactions. This in part opens the gap because of the resulting reduced p–d hybridization with the O-2p orbitals constituting the valence band maximum (VBM). On the other hand, the remainder of the gap is not due to localized atomic effects. Quite to the contrary, analysis of the GW approximation (see below) shows that the long-range or at least medium range $1/r$ behavior of the (dynamically screened) exchange is crucial [20]. The same could also be concluded from hybrid functionals (chapter 6, [21]). Nonetheless, in a GW approach, the difference between quasiparticle and Kohn–Sham eigenvalues is the expectation value of $\langle\psi_c|\Sigma_\mathrm{xc}^\mathrm{GW} - v_\mathrm{xc}^\mathrm{LDA}|\psi_c\rangle$ with $\Sigma_\mathrm{xc}^\mathrm{GW}$ the self-energy operator and ψ_c the conduction band minimum wave function. To the extent that the conduction band wave function ψ_c is dominated by the Zn-s orbitals, it amounts to a shift of the latter in the Hamiltonian. Thus in a strictly pragmatic sense, the U_s shift mimics this effect. Now, in practice, one should realize that a rather large and seemingly unphysical U_s is required, both because the occupation of the Zn-s orbitals is not zero and because the conduction band minimum is not purely Zn-s like. Finally, we note that the two shifts are not independent of each other. A shift in Zn-s through the self-consistently leads to a more ionic bond and this in turn affects the d-states, but ultimately, the PL model leads to a band gap and a d-band position that agree with experiment.

We also note that while the original idea behind LDA + U is that the added U terms should not destroy the already good agreement of LDA for total energies but merely adjust the one-electron levels, this does not imply that the LDA part of the energy would not change. The LDA term in the energy is indirectly affected because changing the Hamiltonian modifies the wave functions, charge density, *etc*. through the self-consistency. In fact, we find that we cannot directly use the LDA + U total energy functional for defect formation energies when applied to non-localized orbitals. In particular, we cannot apply the same LDA + U total energy functional to the free atoms or the reference systems that enter in defining the formation energies. In fact, this would not make sense because U is supposed to contain system specific screening and is not transferable from one system to another. While the U_d may be more or less transferable between different solid state environments, U_s in PL is designed to adjust the band gap specifically of ZnO and thus of course has nothing to do with the position of the Zn-s levels in for example metallic Zn. PL thus calculated the energy formation of the neutral charge state in LDA and only used the full LDA + U functional for the difference between different charge states. In the present work with even more U parameters, we decided to use only the LDA-part of the functional, without the added Hubbard-U and double counting terms. The only way in which LDA + U then enters is through the modified one-electron potential. This may seem strange and may seem to break the consistency of our one-electron levels and the total energy functional, *e.g.*, by invalidating Janak's theorem [22]. However, it should be kept in mind that the Hubbard-U and double counting terms of the FLL LDA + U are primarily designed to deal with the open shell orbital ordering within the set of localized orbitals. For closed shell systems, these terms should vanish.

To complete our discussion of LDA + U we note that the actual operator entering the calculations is a non-local projection potential $|\phi_i\rangle V_i \langle\phi_i|$, in which $\phi_i(\mathrm{r})$ is a local partial wave in the muffin–tin sphere at the LMTO linearization energy ε_v. As such it depends on the sphere radius. This is not important for well-localized wave functions like d states, but when applied to s orbitals, as we will do here, it is sphere radius dependent.

10.3 LDA + *U* Band Structures Compared to *GW*

The band structures of ZnO obtained with the LDA, LDA + U_d, and LDA + U_d + U_s potential were shown in PL. To further scrutinize them, we here compare the band structure of that model with a *GW* calculation [23, 24] in Figure 10.1.

More precisely, because the latter usually overestimates the band gap slightly we compare with a 20% LDA, 80% quasiparticle self-consistent *GW* (QSGW) band structure, which almost exactly reproduces the band gap of 3.4 eV at room temperature. Strictly, speaking we should use a zero temperature gap corrected for spin–orbit splitting of the VBM, zero-point motion corrections, and exciton effects [23] of 3.6 eV but for easy comparison to experiment, we here prefer a gap

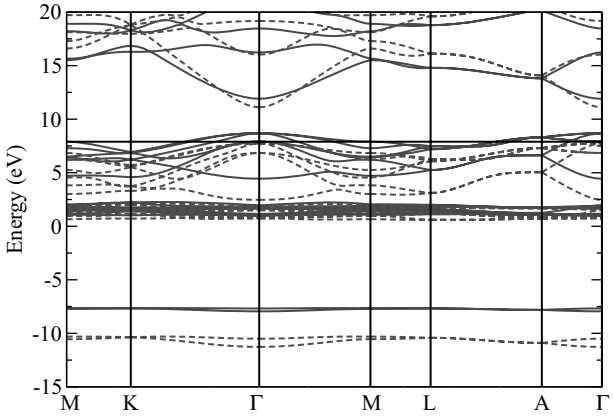

Figure 10.1 (online color at: www.pss-b.com) Band structure of ZnO in "0.8 QSGW" approximation (red dashed line) compared to the PL LDA + U_d + U_s model (blue solid line).

of 3.4 eV. In the QSGW approach, the independent particle Kohn–Sham equations from which the GW self-energy, schematically $\Sigma = iGW$, with G the one-electron Green's function and W the screened Coulomb interaction, is calculated, contains a non-local exchange correlation potential

$$V_{xc}^{\mathrm{QSGW}} = \frac{1}{2}\sum_{nm} |\psi_m\rangle \Re\{\Sigma_{mn}(\varepsilon_m) + \Sigma_{mn}(\varepsilon_n)\}\langle\psi_n|, \tag{10.4}$$

adjusted self-consistently in terms of the self-energy Σ. Here, $\Re$ means taking the Hermitean part, and ψ_n are the eigenstates of the Hamiltonian with this potential. In other words, it forms the best starting point for a single-shot perturbation theory calculation of Σ and leads to Kohn–Sham equations that equal the quasiparticle excitation energies.

In retrospect, one might raise three criticisms of the PL LDA + U_d + U_s model. First, the band gap shift induced by the U_s occurs mainly at the Γ-point only. Thus, instead of a rigid shift, the gaps at K, M, and L are not raised as much and this leads to an overall wrong curvature of the lowest conduction band, with an overestimated effective mass (EM). In particular, since the V_O defect levels are deep, one might expect that their wave function contains contributions from several host band conduction band states at different k-points when a decomposition of the defect state in host states is attempted. One might expect that thus the defect level is not sufficiently raised along with the gap correction.

Second, Alkauskas *et al.* [25, 26] recently made the observation that localized defect levels with respect to the average electrostatic potential are much less sensitive to computational model than with respect to the band-edges. Thus, to place the defect levels correctly with respect to the band-edges requires additional care in calculating the proper band edges relative to the electrostatic potential. It thus appears important that not only the band gap but also the individual band-edges agree between GW and the LDA + U model. Thus, unlike the usual practice of plotting the bands relative

Table 10.1 Structural and total energy properties of wurtzite in various LDA + U models, in-plane lattice constant a (Å), c/a ratio and internal parameter u, energy difference between rocksalt and wurtzite structure in eV/pair. PL [1], BL (this work) defined in Table 10.2.

	LDA	PL	BL	expt.
a (Å)	3.20	3.31	3.30	3.25
c/a	1.603	1.598	1.57	1.6018
u	0.3811	0.3811	0.3872	0.382
$\Delta E_{\mathrm{RS-WZ}}$	0.223	0.062	0.026	> 0

to the VBM, we here plot the bands relative to the average electrostatic potential as zero. We can see that the PL LDA + U_{d} + U_{s} significantly overestimates both the VBM and CBM.

Third, one might worry whether the LDA + U_{d} + U_{s} model provides correct total energies. This point for example was raised by Lany and Zunger [27] who found that LDA + U potentials including U_{s} may lead to too ionic bonding which leads to the h-MgO structure becoming more stable. In that structure, c/a is notably reduced and $u \approx 0.5$ and the system becomes effectively five-fold instead of four-fold coordinated [28]. The latter is closely related to the rocksalt structure. Thus, we here also examine the rocksalt to wurtzite energy difference. From the results in Table 10.1 we can see that indeed the PL model leads to an increase in lattice constant and a reduction of c/a compared to the LDA results. However, the deviation from the experimental wurtzite structure is only minor. Wurtzite stays lower in energy than rocksalt although their energy difference is reduced significantly.

10.4 Improved LDA + *U* Model

In order to remedy the first two of these problems, we constructed a new LDA + U potential, including additional parameters U on Zn-p, O-s, and O-p. Our goal is to obtain as close agreement as possible with the above GW band structure and then to explore how these potentials behave for the defect states. The role of the U_{Op}, U_{Os} which are mostly occupied is to shift the corresponding upper and lower VBM down. The U_{Znp} further allows us to adjust the lowest conduction band at K, M and L. The old and new LDA + U model parameters are summarized in Table 10.2. We note that compared to the previous model U_{Zns} is significantly reduced. We also show the actual shift potentials V_i [according to Eq. (10.2)] that result from them for bulk ZnO with the self-consistently determined density matrices as well as the average occupation numbers (averaged over the different p and d orbitals). These correspond to more reasonable shift potentials than the U_i values, perhaps with the exception of Zn-s in the PL model. For example, we see that because O-p orbitals have an occupation close to 0.5 one needs a large U to achieve a reasonable shift. One still needs to keep in mind that all actual bands have mixed atomic orbital character by

Table 10.2 Parameters U_i of the LDA + U models, resulting self-consistent occupation numbers n_i and shift potential V_i in eV. PL [1], BL (present work).

i	U_i		n_i		V_i	
	PL	**BL**	**PL**	**BL**	**PL**	**BL**
Znd	3.4	4.9	0.947	0.960	−1.56	−2.18
Zns	43.5	13.60	0.039	0.097	19.75	5.48
Znp	0	27.21		0.023		12.98
Os	0	21.77		0.795		−6.47
Op	0	39.45		0.642		−5.65

forming bonding and antibonding states and that the above occupation numbers are sphere size dependent. Note also that the shifts may vary near the defects when their occupation numbers change. This is in fact what distinguishes LDA + U from non-local external potentials.

The band structure of our new LDA + U model compared to GW is shown in Figure 10.2. The new model can be seen to adjust the conduction bands not only at Γ but also its EM at Γ and dispersion all the way to K and M. The position of the VBM relative to the average electrostatic potential is also improved. The model still somewhat underestimates the band width of the O-2p like valence bands and overestimates the valence band EM. Higher conduction bands at Γ are still slightly off and the low lying O-2s like valence band has too low binding energy.

Unfortunately, the newer model gives slightly worse wurtzite structural properties. Still the lattice constant is only 1.5% overestimated, c/a only 2% underestimated and u stays far from 0.5. The total energy difference between rocksalt and wurtzite obtained in both LDA + U models is significantly lower than in LDA. Our LDA result is close to the results by Schleife *et al.* [29] of 0.29 eV. At least, we can be reassured that

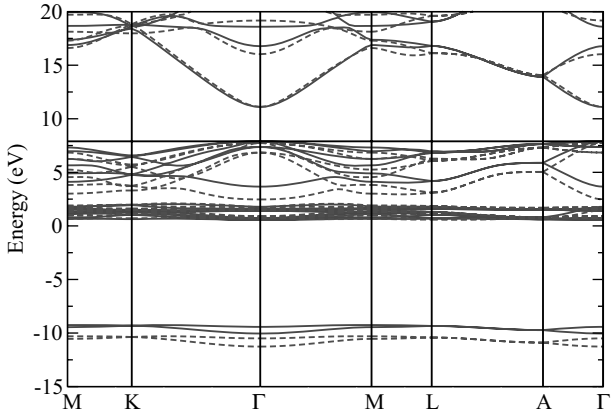

Figure 10.2 (online color at: www.pss-b.com) Band structure of ZnO in "0.8 QSGW" approximation (red dashed line) compared to the present LDA + U model (blue solid line).

the wurtzite structure remains the lower energy structure and we are nowhere near h-MgO-like c/a and u.

We note that this adjustment of the LDA + U potentials is by no means unique. One might also contemplate adding empty sphere shift potentials to adjust the potential in the interstitial region. This might actually more readily mimic a shift of the delocalized conduction band states than Zn-p but has not yet been attempted here.

10.5 Finite Size Corrections

Before proceeding to the results of the LDA + U models for the defect, we need to address two side issues which influence the results. The first of those is the finite size correction. PL examined the size convergence in different size supercells but concluded that no clear $1/L$ behavior was seen. Furthermore, the expected behavior of the image charge correction [30, 31] $q^2/L\varepsilon$ to be proportional to q^2 was not observed. They therefore did not include the image charge corrections, but instead used scaling versus $1/V$ which led to a rather small extrapolation from the largest supercells used (192 atoms). In retrospect, this failure to obtain the $1/L$ behavior is probably in part due to the use of relaxed structures, mixing up the $1/V$ elastic effects as well as possibly problems in sufficiently accurately determining the alignment potential and mostly the limited range of supercell sizes investigated which makes it difficult distinguishing between $1/L$ and $1/L^3$ behavior. Such effects can dominate the result especially for relatively small cells. Only for cells larger than say 200 atoms, the purely electrostatic terms q^2/L become dominant. This can for instance be seen in Figure 7 of Lany and Zunger [27].

We now believe that even for relatively delocalized defect electron densities, the image point charge correction is important because the latter contains always point like contributions from the ionic charge change introduced by the defect. When we add the image point charge correction, or rather 2/3 of it as recommended by Lany and Zunger [27] to mimic the additional quadrupole term, to the results of PL for the largest cell, we find that the $\varepsilon(2+/0)$ transition level shifts becomes 0.64 eV above the VBM the $\varepsilon(+/0)$ level 0.72 eV, and the $\varepsilon(2+/+)$ level 0.57 eV. The defect formation energies as function of Fermi level are shown in Figure 10.3.

The quadrupole background interaction term $\propto Qq/\varepsilon L^3$ identified by Makov and Payne [31] should strictly speaking not include the screening charge density. It is the latter that leads to an effective $Q \propto L^2$ behavior which turns the $1/L^3$ into $1/L$ behavior and allows one to combine it with the point charge term [27]. But since the screening charge is due at least in part to the background density itself, it is not clear one should include this correction. Without the factor 2/3, the levels would shift down even further to 0.53, 0.66, and 0.40 eV, respectively.

The defect then becomes a positive U_{eff} defect rather than negative U type. We note that PL already found much less negative U_{eff} behavior in other words a smaller $|U_{\text{eff}}|$ than the LDA calculations or LDA + U_{d} only. Here, U_{eff} should not be confused with

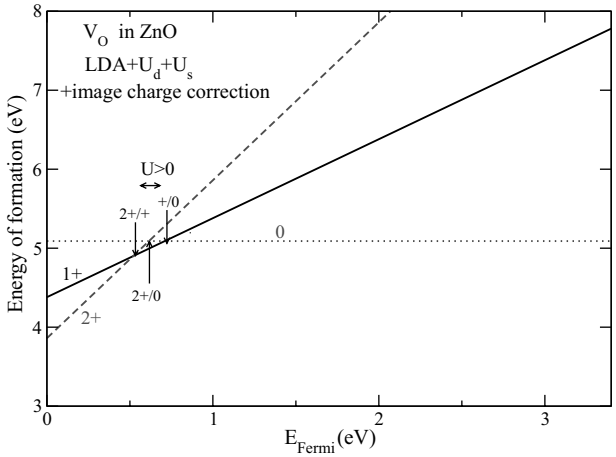

Figure 10.3 (online color at: www.pss-b.com) Formation energies of the oxygen vacancy in ZnO with the PL LDA + U_d + U_s model for 192 atoms including image charge correction as function of Fermi level.

the LDA + U parameters but is defined as $U_{\text{eff}} = \varepsilon(+/0) - \varepsilon(2+/+)$ This result in fact is consistent with Lany and Zunger's slightly different LDA + U_d + U_s model [27]. To be sure, these authors did not propose this to be their favorite approach for dealing with the gap correction. They used it for illustrative purposes and compared a U_s-only, $U_s + U_d$, and U_d-only model. In their $U_d + U_s$ model they obtain $2+/0$ at 0.61 and $+/0$ at 0.79 eV and $2+/+$ at 0.43 eV, very close to ours when including the factor 2/3 in the image point charge correction.

10.6 The Alignment Issue

Another important issue is how to align the VBM of the perfect crystal with that of the defect cell. The defect formation Gibbs free energy (at zero temperature) is defined as

$$\Delta G_f(D,q) = E(D,q) - E(X) - \sum_i \mu_i \Delta n_i + q\mu_e, \tag{10.5}$$

where $E(D,q)$ is the total energy of the supercell with the defect in the charge state q, compensated by a neutralizing homogeneous background charge density, $E(X)$ is the total energy of the perfect crystal calculated in the same size supercell to avoid k-point convergence issues, μ_i is the chemical potential of the elements whose occupation changes by the defect, and μ_e is the chemical potential of the electrons. The latter represents the energy of the electrons in the perfect crystal reservoir $\mu_e = \varepsilon_{\text{vbm}} + \varepsilon_F$ with the Fermi level ε_F measured relative to the VBM and ε_{vbm} determined relative to the average electrostatic potential. To determine ε_{vbm} we use a local reference

potential V_{loc} mark on an atom far away from the defect where the potential presumably becomes bulk like, and add $\varepsilon_{vbm}^{bulk} - V_{loc}^{bulk}$ where the latter is calculated in the perfect crystal primitive cell.

To do this accurately, one needs to average over a few atoms far away from the defect and make sure that the cell is large enough that the local potential marker indeed becomes constant over the region far away from the defect. In PL this alignment potential was determined for the neutral charge state only and then used for the other charge states. This avoids possible long-range contributions of the defect potential for the charged states.

In Figure 10.4 we plot the potential at the muffin–tin radii as function of distance from the oxygen vacancy for all three charge states. We see that sufficiently far from the defect, this potential becomes indeed constant apart from some small oscillations and is nearly the same for the three charge states.

10.7 Results for New LDA + *U*

Here, we used supercells of 128 atoms in the wurzite structure. We summarize the results of our new LDA + U potentials compared with those of PL for the same size cell in Table 10.3. The image point charge correction added was taken as $(9/10)q^2/\varepsilon R$ with R the radius of a sphere with the same volume as the supercell. This image charge correction amounts to 0.18 eV for $q = 1$ using a dielectric constant $\varepsilon = 10$ and a four times larger value for $q = 2$. This does not include the correction factor of 2/3 for the quadrupole term for the reasons explained earlier. Note that for the 1+ charge state, the one-electron levels are spin-polarized. The lower one is occupied, the higher one of minority spin is

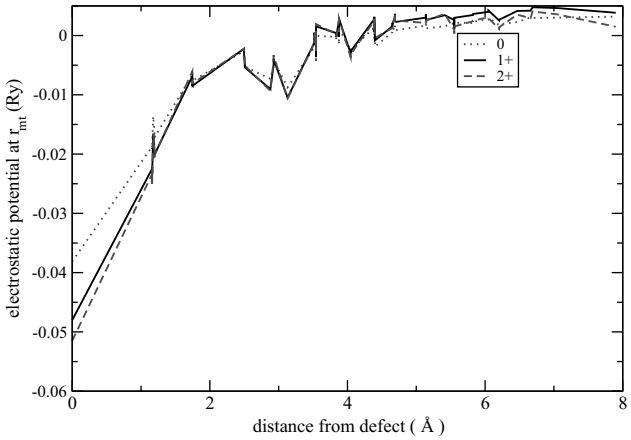

Figure 10.4 (online color at: www.pss-b.com) Potential at the muffin–tin radius as function of distance from the defect for different charge states.

Table 10.3 One-electron levels relative the VBM for different charge states (and spin σ in the $q = +1$ case) $\varepsilon_{q\sigma}$, Gibbs free energies of formation in different charge states $\Delta G_f(q)$ for $\varepsilon_F = 0$, transition level $\varepsilon(q, q')$, and U_{eff}, all in eV.

	ε_0	$\varepsilon_{+\uparrow}$	$\varepsilon_{+\downarrow}$	ε_{2+}	$\Delta G_f(0)$	$\Delta G_f(1+)$	$\Delta G_f(2+)$	$\varepsilon(2+/0)$	$\varepsilon(+/0)$	$\varepsilon(2+/+)$	U_{eff}
PL[a)]	1.0	0.6	1.1	1.8	5.06	4.41	3.92	0.57	0.65	0.49	0.16
PL[b)]					4.70	4.48	3.62	0.54	0.22	0.86	−0.64
BL[c)]	1.50	1.22	1.86	2.46	3.24	0.93	−2.41	2.82	2.31	3.34	−1.0
JV[d)]					6.7	4.6	1.8	2.42	1.94	2.90	−0.96
LZ[e)]					4.2	3.3	1.0	1.60	0.94	2.24	−1.32
Oba[f)]	1.0			$> \varepsilon_{\text{cbm}}$	4.1			2.20			< 0
SX[g)]					3.95			2.20			< 0
GGA + U + GW[h)]	0.8	0.51	2.5	$> \varepsilon_{\text{cbm}}$	4.55	3.3	1.8	1.36	1.26	1.46	−0.2
HSE + GW[h)]								1.66			
P[i)]	0.34	0.64	2.24	2.74				0.34	0.04	0.64	−0.6

a) Based on PL [1]: 128 atom cell but adding image charge corrections.
b) Using LDA + U_d + U_s as in PL but using LDA part of functional only.
c) This work's new LDA + U model.
d) Janotti and Van de Walle [2]: extrapolated LDA + U_d.
e) Lany and Zunger [3]: LDA + U_d correction of VBM.
f) HSE [32] hybrid functional with Hartree-Fock mixing fraction $\alpha = 0.375$ (Oba *et al.* [8]).
g) Screened exchange (Clark *et al.* [12]).
h) GGA + U + GW and HSE($\alpha = 0.25$) + GW (Lany and Zunger [9]).
i) B3LYP functional (Patterson [10]).

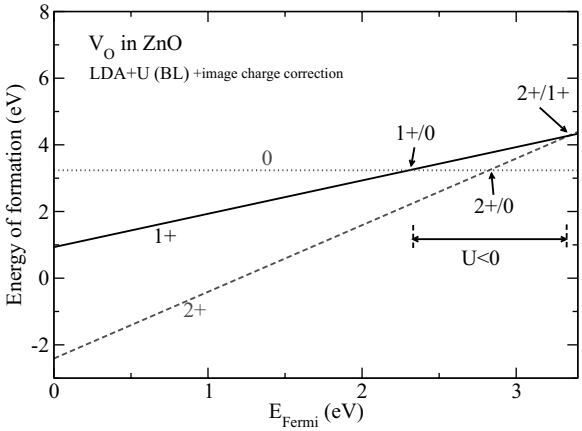

Figure 10.5 (online color at: www.pss-b.com) Energies of formation of V_O in different charge states as function of Fermi level in the oxygen rich limit calculated in present LDA + U model.

empty. The results for the energies of formation as function of Fermi level position are shown in Figure 10.5 In Table 10.3 the second row uses the same LDA + U_d + U_s model but instead of using the full LDA + U functional as PL did, we use our present approach of only using the LDA part of the total energy, using effectively only the V_i shifts. Image point charge corrections are the same as before. Note that this leads to about the same $2+/0$ level but changes the U_{eff} value to become negative.

10.8 Comparison with Other Results

In Table 10.3 we have added selected information from the literature, focusing on the latest results. Unfortunately, not all authors give values for all quantities shown here. Some are estimated from figures to the best of our ability or energies of formation were converted from Zn-rich to O-rich using an energy of formation of ZnO of -3.1 eV.

We can see that in the present LDA + U model, the one-electron level in the neutral charge state lies somewhat higher above the VBM than in PL. The one electron levels in the $1+$ and $2+$ charge states also lie significantly higher in the gap but still below the CBM even for the $2+$ charge state. Clearly, this is as expected by the fact that in our new LDA + U model the VBM drops down relative to the electrostatic potential and the gap is opened not only at Γ but throughout the lowest conduction band at other k-points.

Lany and Zunger [9] recently pointed out that one-electron levels require a finite-size correction for the effects of image point charges and the background charge density. We estimate this effect as follows. If we approximate the cells by spheres, neighboring cells do not give any contribution and the correction

amounts to the potential due the constant background charge density. This is readily calculated to be

$$\phi(r)-\phi(R) = \frac{q}{2\varepsilon}\left(\frac{1}{R}-\frac{r^2}{R^3}\right). \tag{10.6}$$

The question is now how localized the defect wavefunction is. If it is a δ-function at the origin, the upward shift is $q/(2\varepsilon R)$. If it is spread uniformly over the whole cell, we need to average the above potential over the sphere, which gives $(1/5)q/(\varepsilon R)$. For a 128 atom cell ZnO cell, these amount to 0.1 and 0.04 eV providing upper and lower bounds. Averaging, the correction is estimated to be a downward shift of order 0.07 eV for $q = 1$ and proportional to q. Our results in Table 10.3 do not include this negligible correction.

The defect formation energies here correspond to the oxygen rich limit. Our value of about 3.24 eV for the neutral charge state is somewhat lower than Paudel's result and Lany and Zunger [3], Oba *et al.* [8], and Clark *et al.* [12] but significantly smaller than that of Janotti and Van de Walle [2]. In the Zn-rich limit, the energies are lowered by 3.1 eV and the $\Delta G_f(0)$ becomes less than 1 eV, supporting the idea that this could be an abundant defect. It will, however, not be a large source of free electrons because it is a deep donor.

The transition levels move to significantly higher values, in fact even higher than Janotti and Van de Walle's and we find back a relatively strongly negative $U_{\text{eff}} = -1.0$ eV. This is mainly a result of the different relaxations in the $2+$ (outward by 15.6%), $1+$ (inward by 0.3%) and neutral (inward by 4.4%) charge states.

We note that the results by Janotti and Van de Walle [2] using an extrapolated U_d correction as well as the Lany and Zunger results from 2005 [3] without extrapolation are both based on LDA results with *a posteriori* corrections only. The VBM was simply shifted down by the LDA + U_d shift in pure ZnO. In that case, the position of the $1+$ one-electron level is above the CBM and this leads to erroneous occupation of the CBM instead of the defect level. In that sense the position of the $1+$ level is not well defined and that makes the U_{eff} untrustworthy. While most of the discrepancy between Janotti and Van de Walle [2] and Lany and Zunger [3] arises from their choice of extrapolating or not extrapolating the shifts induced by LDA + U_d, part also arises from the fact that Janotti and Van de Walle did not apply the image charge correction. GGA and LDA also slightly differ in lattice constants and this also contributes to the confusion [6, 7]. Actual GGA + U_d or LDA + U_d calculations applied to the defect place the transition levels closer to the VBM than simply adding a LDA + U_d induced downward shift of the VBM *a posteriori* to the LDA results. For example, Lany and Zunger [9] show that GGA + U gives $\varepsilon(2+/0) = 0.98$ eV, while their 2005 result [3] just shifting the VBM down *a posteriori* gave 1.60 eV. Erhart *et al.* [6, 7] used a different way of extrapolating to infinite cells and used GGA + U instead of LDA + U and reported the results either with or without the Janotti–Van de Walle type of extrapolation [6, 7] leading to somewhat intermediate results. The work by Zhang *et al.* [4] used smaller cells and other ways to estimate the gap correction effects.

Recently, several hybrid functional calculations have been carried out. The oldest one is the B3LYP calculation of Patterson [10]. It finds transition levels even closer to the VBM but also finds defect levels in the gap for all charge states, including the 2+ charge state. Oba *et al.* [8] used the HSE [32] hybrid functional but with the fraction of Hartree–Fock adjusted so as to get the correct band gap. That gives a $\varepsilon(2+/0)$ at 2.2 eV similar to Janotti and Van de Walle [2]. Unfortunately, Oba *et al.* [8] do not mention the $+/0$ or U_{eff} values nor do they show the band structure in the 1+ charge state. Their band structure in the neutral charge state finds a defect level in the usual place, about 1 eV above the VBM, while the 2+ charge state shows no level in the gap.

A similar high energy location of the $\varepsilon(2+/0)$ is obtained using screened exchange by Clark *et al.* [12] at 2.2 eV. Again, they do not mention the position of the single plus charge state levels. Lany and Zunger [9, 12] recently also applied the HSE functional but with the standard Hartree–Fock mixing fraction of 1/4 and that calculation places the defect level at 1.67 eV. They also used HSE and GGA + U as starting point for a GW calculation. They use an approach originally introduced by Rinke *et al.* [33] in which the vertical transition energies between charge states in a Franck–Condon coordination diagram are calculated as quasiparticle excitations. In other words, the transition from the neutral to the 1+ charge state is considered to be a transition of an electron from the defect level to the CBM at frozen geometry of the neutral charge state and this energy difference is first calculated as a difference between quasiparticle levels in GW. Afterward, the relaxation energy in each charge state is added as calculated either in HSE or in GGA + U. The comparison is a bit complicated because these authors use a smaller zincblende cell and hence find it necessary to correct the one-electron levels for finite size corrections as already mentioned above. They find a $\varepsilon(2+/0)$ at 1.36 eV in the GGA + U + GW approach and at 1.66 eV in the HSE + GW approach, still significantly lower than the other results. Their $|U_{\text{eff}}|$ is also significantly smaller than the LDA values or our present value.

10.9 Discussion of Experimental Results

Finally, we should discuss the connection of all these results to experiment. Two experiments are particularly relevant to the present discussion. The first is the electron paramagnetic resonance (EPR) experiment by Evans *et al.* [34]. They observe the appearance of the V_O^+ charge state from the neutral start state under excitation of light with $h\nu > 2.1$ eV. Interpreting this as an optical transition from the neutral V_O^0 one-electron level to the conduction band, and neglecting excitonic effects, this places the neutral defect's one-electron level at 1.3 eV above the VBM. In fact, this could be viewed as an upper limit taking in to account the exciton binding energy. This appears consistent with most calculations placing this level in the lower half of the band gap given the uncertainties on the experimental determination. Our present value for this one-electron level at 1.5 eV seems a bit high.

The second experiment is by Vlasenko and Watkins [35]. Using optically detected EPR (ODEPR) they provide evidence for a process of the type

$$V_O^+ + EM^0 \rightarrow V_O^0 + EM^+, \tag{10.7}$$

in which an electron is transferred from an EM type donor to the oxygen vacancy in the single positive charge state, thereby quenching its EPR signal. Since the EM and ODEPR signal (L3) associated with the V_O^+ are positive in a photoluminescence band with estimated zero phonon line at 2.48 eV, this could indicate that the empty (minority spin) one-electron eigenvalue of the V_O^+ charge state lies 2.48 eV below the EM level or roughly the same amount below the CBM or at about 1 eV above the VBM if this photoluminescence is directly a result of the electron capture. There is still considerable uncertainty on this experimental value. The peak of this photoluminescence occurs at 600 nm or closer to 2.0 eV. On the other hand, an alternative explanation of this process is that the positive photoluminescence results from subsequent recombining of the (radiatonlessly) captured electron with a hole from the VBM. If we assume that this happens before the defect has time to relax to the neutral ground state and still reflects the geometry of the single positive charge state, then it means that the empty one-electron level lies at about 2.0–2.5 eV above the VBM. This result is qualitatively consistent with Lany and Zunger's *GW* calculation [9] who find that the occupied level of the V_O^+ charge state shifts along with the VBM while the empty state shifts along with the CBM when the *GW* self-energy shifts are applied. The values for these levels in Table 10.3 are estimated from their figure. We find similarly that the $1+$ charge state minority spin level lies in the upper part of the gap at 1.86 eV, supporting the second interpretation of the Vlasenko Watkins experiment. Our value is a bit on the low side. This was the interpretation proposed by Janotti and Van de Walle [2] although they based it on the position of the $+/0$ transition state in the gap.

We conclude that neither of the two experiments provides direct evidence for the location of the thermodynamic transition levels. Instead they provide information on the one-electron eigenvalues in the neutral and $1+$ charge state, respectively.

10.10 Conclusions

In summary, the LDA $+$ *U* approach as applied to band gap correction for defects was reviewed. Shortcomings in an earlier application of the approach by PL [1] were identified and corrected. The LDA $+$ *U* approach was applied to Zn-s, p, d and O-s, p orbitals with U_i values adjusted so as to reproduce as closely as possible the band structure of ZnO in the QSGW approach. Not only band gaps but also dispersions and the position of the levels relative to the electrostatic reference potential were adjusted. This new LDA $+$ *U* model leads to transition levels close to (but slightly higher than) the recent hybrid functional and screened exchange calculations. In addition, the position of the empty minority spin level in the

1 + charge state is argued to support an explanation of the ODEPR data by Vlasenko and Watkins in terms of a two step capture + recombination with valence band hole model. The position of the one-electron level in the neutral charge state is consistent with EPR optical activation of the V_O^+ signal. The energy of formation of the oxygen vacancy is found to be relatively low, supporting the notion that this could be an abundant defect in Zn-rich material.

Acknowledgements

This work was supported by the Army Research Office under grant no. W911NF-06-1-0476 and the National Science Foundation under grant No. DMR 0710485. A. Boonchun also thanks the Commission on Higher education of Thailand for support. Computations were carried out at the Ohio Supercomputercenter under project number PDS-0145 and the HPC center at CWRU. We thank Tula Paudel for useful discussions and Mark Van Schilfgaarde for the FP-LMTO and QSGW programs.

References

1 Paudel, T.R. and Lambrecht, W.R.L. (2008) *Phys. Rev. B*, **77** (20), 205202.
2 Janotti, A. and Van de Walle, C.G. (2005) *Appl. Phys. Lett.*, **87** (12), 122102.
3 Lany, S. and Zunger, A. (2005) *Phys. Rev. B*, **72** (3), 035215.
4 Zhang, S.B., Wei, S.H., and Zunger, A. (2001) *Phys. Rev. B*, **63** (7), 075205.
5 Kohan, A.F., Ceder, G., Morgan, D., and Van de Walle, C.G. (2000) *Phys. Rev. B*, **61** (22), 15019–15027.
6 Erhart, P., Klein, A., and Albe, K. (2005) *Phys. Rev. B*, **72** (8), 085213.
7 Erhart, P., Albe, K., and Klein, A. (2006) *Phys. Rev. B*, **73** (20), 205203.
8 Oba, F., Togo, A., Tanaka, I., Paier, J., and Kresse, G. (2008) *Phys. Rev. B*, **77** (24), 245202.
9 Lany, S. and Zunger, A. (2010) *Phys. Rev. B*, **81** (11), 113201.
10 Patterson, C.H. (2006) *Phys. Rev. B*, **74** (14), 144432.
11 Clark, S.J. and Robertson, J. (2010) *Phys. Status Solidi B*, doi: 10.1002/pssb.201046110 (chapter 5 in this book).
12 Clark, S.J., Robertson, J., Lany, S., and Zunger, A. (2010) *Phys. Rev. B*, **81** (11), 115311.
13 Hohenberg, P. and Kohn, W. (1964) *Phys. Rev.*, **136** (3B), B864–B871.
14 Kohn, W. and Sham, L.J. (1965) *Phys. Rev.*, **140** (4A), A1133–A1138.
15 Anisimov, V.I., Zaanen, J., and Andersen, O.K. (1991) *Phys. Rev. B*, **44**, (3) 943–954.
16 Anisimov, V.I., Solovyev, I.V., Korotin, M.A., M.T. Czyżyk, and Sawatzky, G.A. (1993) *Phys. Rev. B*, **48** (23), 16929–16934.
17 Liechtenstein, A.I., Anisimov, V.I., and Zaanen, J. (1995) *Phys. Rev. B*, **52** (8), R5467–R5470.
18 Methfessel, M., van Schilfgaarde, M., Casali, R.A. 2000) A full-potential LMTO method based on smooth Hankel functions, in: *Electronic Structure and Physical Properties of Solids. The Use of the LMTO Method* (ed. H. Dreyssé), Lecture Notes in Physics, vol. 535 Springer Verlag, Berlin, p. 114.
19 Kotani, T. and van Schilfgaarde, M. (2010) *Phys. Rev. B*, **81** (12), 125117.
20 Maksimov, E.G., Mazin, I.I., Savrasov, S.Y., and Uspenski, Y.A. (1993) *J. Phys.: Condens. Matter*, **1**, 2493.
21 Henderson, T.M., Paier, J., and Scuseria, G.E. (2011) *Phys. Status Solidi B*, doi: 10.1002/pssb.201046303

22 Janak, J.F. (1978) *Phys. Rev. B*, **18** (12), 7165–7168.

23 Kotani, T., van Schilfgaarde, M., and Faleev, S.V. (2007) *Phys. Rev. B*, **76** (16), 165106.

24 van Schilfgaarde M., Kotani, T., and Faleev, S.V. (2006) *Phys. Rev. B*, **74** (24), 245125.

25 Broqvist, P., Alkauskas, A., and Pasquarello, A. (2009) *Phys. Rev. B*, **80** (8), 085114.

26 Alkauskas, A., Broqvist, P., and Pasquarello, A. (2010) *Phys. Status Solidi B*, doi: 10.1002/pssb.201046195 (chapter 7 in this book).

27 Lany, S. and Zunger, A. (2008) *Phys. Rev. B*, **78** (23), 235104.

28 Limpijumnong, S. and Lambrecht, W.R.L. (2001) *Phys. Rev. B*, **63** (10), 104103.

29 Schleife, A., Fuchs, F., Furthmüller, J., and Bechstedt, F. (2006) *Phys. Rev. B*, **73** (24), 245212.

30 Leslie, M. and Gillan, M.J. (1985) *J. Phys. C, Solid State Phys.*, **18**, 973.

31 Makov, G. and Payne, M.C. (1995) *Phys. Rev. B*, **51** (7), 4014–4022.

32 Heyd, J., Scuseria, G.E., and Ernzerhof, M. (2006) *J. Chem. Phys.*, **124** (21), 219906.

33 Rinke, P., Janotti, A., Scheffler, M., and C.G. Van de Walle (2009) *Phys. Rev. Lett.*, **102** (2), 026402.

34 Evans, S.M., Giles, N.C., Halliburton, L.E., and Kappers, L.A. (2008) *J. Appl. Phys.*, **103** (4), 043710.

35 Vlasenko, L.S. and Watkins, G.D. (2005) *Phys. Rev. B*, **71** (12), 125210.