Angle-resolved photoemission and quasi-particle calculation of $d$ band oxide semiconductors: The need for $d$ band shift in ZnO

Linda Y. Lim$^{1,2}$, Stephan Lany$^3$, Young Jun Chang$^4$, Eli Rotenberg$^5$, Alex Zunger$^6$, Michael F. Toney$^2$

1 Department of Materials Science and Engineering, Stanford University, Stanford, CA 94305, USA
2 Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA
3 National Renewable Energy Laboratory, Golden, CO 80401, USA
4 Department of Physics, University of Seoul, Seoul 130-743, Korea
5 Advanced Light Source, E. O. Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA
6 University of Colorado at Boulder, Boulder, CO 80309, USA

Received 19 September 2012

Abstract

ZnO is a prototypical semiconductor with occupied $d^{10}$ bands that interact with the anion $p$ states and is thus challenging for electronic structure theories. Within the context of these theories, incomplete cancellation of the self-interaction energy results in a Zn $d$ band that is too high in energy, resulting in upwards repulsion of the valence band maximum (VBM) states, and an unphysical reduction of the band gap. Methods such as GW should significantly reduce the self-interaction error, and in order to evaluate such calculations, we measured high resolution and resonant angle-resolved photoemission spectroscopy (ARPES) and compared these to several electronic structure calculations. We find that in a standard GW calculation, the $d$ bands remain too high in energy by more than 1 eV irrespective of the Hamiltonian used for generating the input wavefunctions, causing a slight underestimation of
the band gap due to the $p$-$d$ repulsion. We show that a good agreement with the ARPES data over the full valence band spectrum is obtained, when the Zn-$d$ band energy is shifted down by applying an on-site potential $V_d$ for Zn-$d$ states during the GW calculations to match the measured $d$ band position. The magnitude of the GW quasi-particle energy shift relative to the initial density functional calculation is of importance for the prediction of charged defect formation energies, band-offsets and ionization potentials.
I. INTRODUCTION

ZnO is the prototype binary semiconductor with electronically active $d^{10}$-shell and serves as an archetype for ternary materials, such as $A_2ZnO_4$ spinels (where $A=$cationic metals), or chalcopyrites such as CuInSe$_2$. In II-VI semiconductors, the Zn 3$d$ band alters the valence band structure via the $p$-$d$ band repulsion, affecting band gaps, band offsets, cohesive energies and lattice constants, as well as spin-orbit parameters. Common electronic structure methods deal well with ordinary $s$-$p$ materials, such as Si or GaAs, but active $d$-electron materials pose a problem to these methods due to the presence of stronger self-interaction energy. The trend for the position of the Zn-$d$ band relative to the anion $p$ orbitals for the series ZnTe – ZnSe – ZnS is to progressively decrease in energy. This trend poses a challenge for ZnO, because the Zn-$d$ to O-$p$ energy distance (“energy denominator” in perturbation theory) is small, and the Zn-X bond length short, and hence there is stronger coupling in perturbation theory. The incomplete cancellation of self-interaction in density functional theory places all Zn $d$ bands too high in energy. This error is particularly acute in ZnO where the Zn-$d$ to O-$p$ separation is naturally smaller than in other Zn chalcogenides. Thus, the effects of self interaction error can be significant in ZnO. Yet, metal oxides with $d$-shells are increasingly important technologically, and ZnO, in particular, is a candidate for various applications, such as in photocatalysis and photovoltaics. Thus, it is important to investigate the ZnO electronic structure with the goal of rectifying the common, theoretical “too shallow” $d$-shell, in particular in view of the more severe difficulties encountered in compounds with open and empty $d$-shells where the $d$-orbitals lie much closer to the band gap.
Previous work has been done both experimentally and theoretically to understand the
electronic structure of ZnO. The ZnO band structure has been probed experimentally by
angle-resolved photoemission spectroscopy (ARPES) using photons in both the ultraviolet
region\textsuperscript{7} and in soft x-ray region\textsuperscript{8}. However, the position and the dispersive features of the Zn-
d band were not determined precisely in these angle-resolved photoemission experiments.
Theoretical first principles band structure calculations for ZnO are available for a broad range
of electronic structure methods. Standard density functional theory (DFT) in the local density
approximation (LDA) and the generalized gradient approximation (GGA)\textsuperscript{9-12} underestimate
the band gap rather dramatically, giving values of less than 1 eV compared to the
experimental gap of 3.44 eV\textsuperscript{13}. The origin of this unusually large discrepancy is that the
typical band gap problem of LDA is compounded by the effect of the underestimated binding
energy of the Zn-\textit{d} shell, which further reduces the band gap due to \textit{p}-\textit{d} repulsion\textsuperscript{1}. In order to
predict band gaps that are comparable to experiments, it is necessary to go beyond DFT,
which, for semiconducting periodic systems, is usually through calculation of the electron
self-energy in the GW approximation\textsuperscript{14}. However, different recent GW calculations,
including quasi-particle self-consistent GW calculations (with self-consistency in the
wavefunctions), still place the Zn-\textit{d} band at too high in energy by about 1 eV\textsuperscript{15-19}. In other
words, whereas in principle GW should be self-interaction free and lower the Zn-\textit{d} energy, in
practice, current GW implementations are probably not fully self-interaction free.

In this paper, we use high resolution ARPES of ZnO (0001) together with resonant
photoemission as a test of theoretical treatments for ZnO. We show that GW calculations
based on GGA, GGA+U and HSE wavefunctions exhibit a similar too–high (overestimation)
Zn-\textit{d} band energies. The problem is resolved by applying an on-site potential for Zn-\textit{d} states
to correct the \textit{d} band energy, and we also show that this rectifies other, related errors in the \textit{p}-
When the Zn-\(d\) band energy is shifted down by applying an on-site potential \(V_d\) for Zn-\(d\) states during the GW calculations, to match the experimental \(d\) band position, an improved value for the band gap (3.30 instead of 2.94 eV) and good agreement with the ARPES data over the full valence band spectrum is obtained. The improved GW quasiparticle energies and their shift with respect to the initial DFT calculation (here in GGA+U) can be used to improve the DFT predictions for defect formation energies (charged defects), band-offsets, and ionization potentials \(^{20-25}\).

II. METHODS

A. Experiment

A single crystalline Ga-doped ZnO (0001) sample with Zn-face polished surface was purchased commercially from MTI Corporation. The dopant concentration (~ 3 \(x\) 10\(^{17}\) cm\(^{-3}\), provided by the supplier) made the sample electrically conductive to avoid charging during measurements, but was small enough so that the dopant did not change the band structure relative to pure ZnO. ARPES measurements were performed at the Electronic Structure Factory endstation at beamline 7.0.1 of Advanced Light Source at Lawrence Berkeley National Laboratory. The sample was kept at 100 K during the measurements. The combined (monochromator + electron spectrometer) energy resolution was 25 meV and the electron spectrometer had 0.1° angular resolution. Clean surfaces were prepared in a separate ultrahigh vacuum (UHV) preparation chamber directly connected to the spectrometer chamber. The surface of the sample was subjected to 500 V Ar ion sputtering at 10\(^{-5}\) Torr for 20 min and annealing to 900 K, followed by annealing in oxygen at 10\(^{-6}\) Torr and 800 K for 15 min. The above procedure was repeated three times. The position of the Fermi level (\(E_F\)) was
determined by measuring the photoemission spectra of molybdenum, which was electrically
in contact with the sample.

## B. Theory

In the present theoretical approach, we use DFT wavefunctions as input to GW calculations. The GW eigen-energies were iterated to self-consistency to remove the strong dependence of the $G_0W_0$ result on the single-particle energies of the initial DFT calculation. The input DFT wavefunctions are kept constant during the GW calculations. The results thus depend on the input wavefunctions, generated here from a few different DFT approximations. We will use the notation $GW$ (DFT) to denote the type of DFT (or hybrid-functional) wavefunctions used as input to the GW calculation. The DFT band structure calculations were performed using the projector augmented wave (PAW) implementation of DFT and GW in the VASP code. The initial DFT wavefunctions and eigenvalues were calculated for the GGA, GGA+U (U-J = 6 eV), and HSE functionals, where, for the latter, the fraction of the Fock exchange in HSE was adjusted to $\alpha = 0.38$ to match the experimental band gap of ZnO. The same lattice parameters $a = 3.24$ Å, $c = 5.22$ Å, and $u = 0.380$ were used in all calculations. The $\Gamma$-centered 16x16x4 k-mesh was chosen to increase the density of k-points in the in-plane directions that were measured experimentally, but a reduction factor of two was used in the in-plane directions for the calculation of the GW self-energy. The energy cutoff for the response functions was 150 eV, and a total number of 256 occupied and empty bands were used. DFT-derived local field effects were included in the GW calculation, which increase the dielectric constant, thereby counteracting the general tendency of the random-phase approximation (RPA) to overestimate the GW band gaps. It should be noted that there has been a recent controversy in the literature about the magnitude of the (non-self-consistent) “single-shot” $G_0W_0$ (LDA) band gap and about the convergence
behavior with respect to the number of bands used to calculate $W$. Our band $G_0W_0$(LDA) band gap is 2.45 eV in the RPA (without local field effects), thereby supporting the conclusions of Ref. [38] that the larger gap reported in Ref. [36] is due to the plasmon pole model. In addition, we tested the convergence behavior, finding that the gap changes by less than 0.1 eV for a total number of bands between 64 and 1024. Thus, our present results for 256 bands are well converged. The convergence behavior depends on both the approximation used for $W$ (as shown by in Ref. [38]) and the basis set implementation, such as norm-conserving pseudopotentials$^{36,38}$, the linearized augmented plane waves$^{37}$, or PAW (present work). Our present results indicate that the RPA implementation within PAW is a particularly robust method in regard of the convergence behavior.

Since the DFT wavefunctions are kept constant during the GW calculations, this allows the interpretation of the GW quasi-particle energies in terms of energy-shifts relative to the Kohn-Sham (KS) energies, e.g., $E_{\text{VBM}}^{\text{GW}} = E_{\text{VBM}}^{\text{DFT}} + \Delta E_{\text{VBM}}$ for the VBM. Note that under periodic boundary conditions, the band-energies relative to the $V = 0$ vacuum level are generally only defined up to a constant$^{39}$. However, by keeping the DFT wavefunctions and DFT charge density, the GW energies are defined with respect to the same reference, i.e., the average potential in DFT. By the underlying principles of DFT$^{40}$, common DFT approximations can provide good total-energies and charge-densities (and hence also electrostatic potentials), but not accurate quasi-particle energies. GW(DFT) fills this gap by providing the quasi-particle energy relative to the DFT potential. Thus, the combination of a DFT calculation of the total-energies, charge-densities and electrostatic potentials with GW(DFT) quasi-particle energy shifts allows a consistent prediction of experimentally relevant properties such as ionization potentials, band-offsets, charged defect formation energies and other properties that rely on the accurate positioning of the band-edge energies relative to the electrostatic potential. An
accurate prediction of such properties directly within a (hybrid) functional without further
corrections due to quasi-particle energy shifts can be expected only if these shifts vanish for
the respective functional.

III. RESULTS AND DISCUSSION

A comparison of theoretical and experimental band structure of ZnO is shown in Figure 1. The arrows in Figure 1(a) depict the sliced directions in the $\Gamma$-M-K plane where the measurements were taken, and the constant-energy contour plot in Figure 1(b) exhibits hexagonal symmetry, reflecting the hexagonal Brillouin zone of wurtzite ZnO. The VBM is located at the binding energy $E_B \sim -3.45$ eV, and the Fermi level is approximately located near the bottom of conduction band due to our n-type Ga-doped ZnO. The valence bands observed in the $\Gamma$-K-M and $\Gamma$-M-$\Gamma$ directions are highly dispersive, as can be seen in Figures 1(c) to 1(f). The ZnO valence band structure can be partitioned into three segments: (i) the O-$p$ bands at $-3.5$ eV < $E_B$ < $-7.5$ eV, (ii) the Zn-$s$/O-$p$ band at $-7.5$ eV < $E_B$ < $-9.5$ eV, which is formed due to hybridization between the empty $s$ states of the Zn$^{2+}$ cation and the occupied $p$ states of the O$^{2-}$ anion, and (iii) the Zn-$d$ bands at $-10$ eV < $E_B$ < $-12$ eV. The positions of O-$p$ and Zn-$d$ bands are confirmed experimentally by resonant photoemission (see supplementary material, Figure S1) of O-$p$ states.
FIG. 1. $k_{\parallel}$ scans in hexagonal plane of the ZnO Brillouin zone and comparison between experimental and theoretical band structure. A photon energy of 135 eV is chosen to study the $\Gamma$-M-K basal plane. (a) the $\Gamma$-M-K plane, and (b) its constant-energy contour plot of the photoemission intensity. The red lines denote Brillouin zone boundaries, and the dashed blue lines are connectors to the $\Gamma$ points. The arrows
indicate the sliced directions measured. (c), (d) Comparison between experimental band structure (gray scale) and theoretical band structure within the $GW$(GGA) scheme. The origin of energy is chosen at the conduction band minimum. The red lines represent O-$p$ bands, the green lines represent hybridized Zn-$s$/O-$p$ bands, and the blue line represents Zn-$d$ bands. It can be seen that the Zn-$d$ band is too high and there is overlap between Zn-$s$/O-$p$ and Zn-$d$ bands. (e), (f) Comparison between experimental band structure (gray scale) and theoretical band structure within the $GW+V_d$(GGA+U) scheme. The Zn-$d$ band is no longer too high and the Zn-$s$/O-$p$ and Zn-$d$ bands do not overlap. The horizontal dashed line indicates VBM energy. All the experimental band structures (energy distribution curves) shown here are second derivatives.

Our results are in good agreement with previously reported work$^8$ for the O-$p$ bands. However, here we use an ultraviolet photon energy which results in an improvement in the energy resolution, as well as the determination of the Zn-$d$ band position. In addition, no surface states$^7$ were observed in the present work, which shows that we probe the bulk band structure.

Figures 1(c) and 1(d) show the difficulty in correctly predicting the $d$ band position and the O-$p$ band dispersion in GW with a particular input wavefunction (in this case, GGA), as seen from the discrepancies between theoretical and experimental band structures. We observe that the “underbinding” of the Zn-$d$ states also causes a spurious mixing with the Zn-$s$/O-$p$ band, as seen in Figs.1 (c,d). Furthermore, the too small band gap of only 2.92 eV (Table 1) can
also be attributed to the too high $d$ band energy, pushing the VBM to higher energies due to the $p$-$d$ repulsion.

In order to correct the $d$ band energy, we therefore introduce the “assisted – GW” by shifting the $d$ band during the GW calculation. We call this the $GW+V_d$ approach, with an additional on-site potential$^{41}$ ($V_d = -1.5$ eV) to reproduce the experimental $d$ band position. While the functional form of the external potential $V_d$ is similar to DFT+U potential, the important distinction is that of $V_d$ is not occupation-dependent, i.e., it acts equally on occupied and unoccupied states, whereas DFT+U acts in opposite directions (attractive for occupied states, repulsive for empty states). As an initial Hamiltonian for the $GW+V_d$ approach we are using GGA+U, rather than GGA, for reasons discussed below.

| Table 1. Comparison of different computational schemes of DFT and GW for ZnO. The initial DFT wavefunctions are calculated using the GGA, GGA+U, or HSE functionals. The $GW+V_d$ with GGA+U input scheme is used as a reference (cf. Figs. 1(e) and 1(f)). $\Delta E_{\text{VBM}}$ is the GW quasi-particle energy shift with respect to the single-particle energies of the respective initial DFT functional. The experimental band gap for ZnO is 3.44 eV and $d$ band position (Fig. 1) is $-7.50$ eV.

<table>
<thead>
<tr>
<th>$GW+V_d$(GGA+U)</th>
<th>1.52</th>
<th>3.30</th>
<th>-0.99</th>
<th>-7.45</th>
</tr>
</thead>
<tbody>
<tr>
<td>$GW$(GGA)</td>
<td>0.80</td>
<td>2.92</td>
<td>-1.42</td>
<td>-6.26</td>
</tr>
<tr>
<td>$GW$(GGA+U)</td>
<td>1.52</td>
<td>2.94</td>
<td>-0.63</td>
<td>-6.33</td>
</tr>
<tr>
<td>$GW$(HSE)</td>
<td>3.46</td>
<td>3.22</td>
<td>+0.70</td>
<td>-6.21</td>
</tr>
</tbody>
</table>

From Figures 1(e), 1(f) and Table 1, we see that the addition of $V_d$ to the GW self-energy operator improves not only the $d$ band position (-7.45 eV compared to -6.33 eV), but also the band gap. The predicted band gap energy of $E_g = 3.30$ eV compares now well with the experimental value of 3.44 eV as compared to $E_g = 2.94$ eV when the on-site potential $V_d$ is not included. Furthermore, using the $GW+V_d$(GGA+U) approach, the experimental band
structure is well reproduced over the entire range of the valence band spectrum, as seen from the good agreement between theoretical and experimental in Figures 1(e) and 1(f). In particular, the Zn-s/O-p band is now well resolved as it is in the ARPES spectrum.

Next, we compare the effects of the initial Hamiltonian to calculate the wavefunctions in more detail, and illustrate the implications for band structure predictions. Figure 2 shows the band structures for GGA, GGA+U and HSE, and for each case the resulting \( \text{GW(DFT)} \) band structures. The \( \text{GW+V}_d(\text{GGA+U}) \) result is shown here only as a reference for the experimental band energies (Fig. 1(e,f)). There are several important observations: (1) The \( d \) band energies in GW are overestimated by a similar amount for all three initial Hamiltonians. Also, the trend of an underestimated GW band gap exists for all three DFT functionals, which can be interpreted as due to the \( p-d \) repulsion shifting the VBM upwards. (2) In the GGA calculation the Zn-\( d \) bands are located at a very high energy and overlap with the Zn-s/O-p band causing a spurious hybridization. As a result, the Zn-s/O-p band around 7 eV below the VBM is not resolved in the subsequent GW(GGA) calculation. In contrast, the GW(GGA+U) calculation shows the Zn-s/O-p band well resolved very close to the reference calculation, despite the proximity of the Zn-\( d \) bands. This finding indicates that GGA+U avoids the spurious hybridization and creates better wavefunctions, which is the motivation to use GGA+U wavefunctions for the GW+V\(_d\) approach, and highlights the importance of determining the initial wavefunctions with a Hamiltonian that yields the correct band ordering and hybridization. A similar improvement in the description of the wavefunction hybridization is also achieved in HSE (see Fig. 2). It can be expected that a GW+V\(_d\)(HSE) approach would yield a similar agreement with experiment as GW+V\(_d\)(GGA+U) shown in Fig. 1(e,f). Judging from the trends in Table I, the band gap would probably be slightly larger.

(3) Considering the case of the hybrid functional HSE (without the GW quasi-particle energy
shifts), we observe that the \(d\) band position is still too high (Fig. 2), even when the band gap is corrected by adjustment of the exchange parameter \(\alpha\). This finding implies that the band gap and the \(d\) band position cannot be corrected simultaneously in the hybrid functional. In fact, the unusually large value of \(\alpha = 0.38\) which yields the experimental band gap results from the need to compensate for the band gap reducing effect of the \(p-d\) repulsion. This result indicates that a physically more correct description should be achievable in a HSE+U calculation with a more moderate exchange parameter and a suitable \(U\) parameter to match the Zn-\(d\) band energies.

![FIG. 2. Comparison of ZnO band structures in different computational GW(DFT) schemes. The grey band-structure in the background shows the \(GW+V_d(GGA+U)\) result as a reference for the experimental band energies (cf. Figs.1 (e,f)). Superposed are the DFT band energies (red diamonds) and the respective GW energies (blue circles and lines).](image)

Finally, we address the quasi-particle energy shift \(\Delta E_{VBM}\) due to GW relative to the initial DFT calculation. As seen in Table 1, the VBM is lowered in GW by 1.4 eV relative to GGA, but only by 0.6 eV relative to GGA+U, since part of the VBM downshift is already included.
in GGA+U (the difference of 0.8 eV between GGA and GGA+U compares well with $\Delta E_{VBM}$ = −0.7 eV due to U found in Ref. [42]). When the $d$ band position is fully corrected in $GW+Vd$, the downshift increases to 1.0 eV relative to GGA+U. Such large values can considerably affect defect formation energies, particularly in case of high defect charges $q$ where the formation energy varies with $q \cdot \Delta E_{VBM}$, $^{20, 24, 41}$ as well as band-offsets and ionization potentials.

In the case of the HSE functional, the large magnitude of the $\alpha$ parameter needed to correct the band gap has the undesirable side effect that the VBM is lowered too much in the hybrid functional calculation, so that the subsequent GW calculation leads to an unusual upward shift of $\Delta E_{VBM} = +0.7$ eV, whereas GW quasiparticle energy corrections otherwise shift the VBM normally downwards $^{23, 25}$. Thus, additional corrections to defect energies, band offsets, and ionization potentials may be needed even when the band gap is corrected by adjusting the $\alpha$ parameter in HSE. Note that the above suggested HSE+U approach might be suitable to minimize the quasiparticle energy shifts of the band edges and thereby eliminate the need for such corrections, whereas GGA(+U) calculation will generally require such corrections.

IV. CONCLUSION

High resolution and resonant ARPES measurements of the archetype wurtzite ZnO (0001) have been used as a test of electronic structure calculations for ZnO. Comparing different GW schemes with the measured ARPES band structure of ZnO, we corroborated the general trend of GW to overestimate the $d$ band energy, and we emphasized the importance of calculating the initial wavefunctions with a Hamiltonian that yields the correct order of the energy bands and hybridization. Using GGA+U wavefunctions and applying an on-site
potential for Zn-$d$ electrons in addition to the GW self-energy operator, we obtained a consistent description of the valence band structure throughout the Brillouin zone and an improvement of the band gap energy. The predicted band edge shifts relative to the initial GGA+U calculation are expected to be useful for improving the prediction of defect formation energies, band-offsets, and ionization potentials.

**ACKNOWLEDGMENTS**

This material is based on work supported as part of the "Center for Inverse Design," an Energy Frontier Research Center funded by the U.S. Department of Energy, Office of Basic Energy Sciences, under Grant No. DE-AC36-08GO28308. The Advanced Light Source is supported by the Director, Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. Portions of this research were carried out at the Stanford Synchrotron Radiation Lightsource, a Directorate of SLAC National Accelerator Laboratory and an Office of Science User Facility operated for the U.S. Department of Energy Office of Science by Stanford University. S.L acknowledges helpful discussions with G. Kresse and V. Stevanovic.
REFERENCES