# Quasiparticle Band Gap of ZnO: High Accuracy from the Conventional $G^0W^0$ Approach

Bi-Ching Shih, Yu Xue, and Peihong Zhang

*Department of Physics, University at Buffalo, SUNY, Buffalo, New York 14260, USA*

Marvin L. Cohen and Steven G. Louie

*Department of Physics, University of California at Berkeley, Berkeley, California 94720*
*and Materials Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA*

Contrary to previous reports, we show that the conventional $GW$ (the so-called $G^0W^0$) approximation can be used to calculate accurately the experimental band gap ($\sim 3.6$ eV) of ZnO. The widely discussed underestimate of the quasiparticle gap of ZnO within the $GW$ method is a result of an inadequate treatment of the semicore electrons and the slow and nonuniform convergence in the calculation of the Coulomb-hole self-energy in previous studies. In addition, an assumed small kinetic energy cutoff for the dielectric matrix may result in a false convergence behavior for the quasiparticle self-energy.

Quantitative understanding of electronic excitations in semiconductors has been and still is a central topic of condensed matter theory because of its fundamental and technological importance. With the advent of modern electronic structure method that includes electron-electron interactions from first principles [1,2], theory is now used to predict and accurately compare with experiment excited-state properties of semiconductors and insulators without any adjusting parameters. However, there are some perceived exceptions. One widely discussed and notable case is zinc oxide.

Zinc oxide (ZnO) has attracted considerable attention for its potential application in a variety of fields such as optoelectronics and spintronics [3,4]. Surprisingly, theoretical understanding of the electronic structure of this supposedly simple $sp$ semiconductor turns out to be rather challenging. The band gap of ZnO calculated within the local density approximation (LDA) is about 0.7 eV if the semicore 3$s$, 3$p$, and 3$d$ electrons are considered as valence electrons. Subsequent quasiparticle calculations within the *conventional* $G^0W^0$ approach [1,2] gives a band gap of 2.44 eV [5], which is more than 1 eV smaller than the experimental result of about 3.6 eV [6,7] after correcting for the lattice effects. Two recent LDA/$GW$ calculations [8,9] gave an even smaller gap of 2.12–2.14 eV, although this value is increased to about 3.2 eV [8] by subsequent self-consistent $GW$ calculations.

Aside from the well-known problem that Kohn-Sham eigenvalues are not quasiparticle energies, an additional reason for the origin of the severe underestimate of the band gap of ZnO and other related systems (e.g., InN, CdO) within the LDA has been traced to the inadequate treatment of the semicore electrons within the LDA, which underbinds the $d$ states and leads to an unphysically strong $pd$ hybridization. Since the underestimate of the binding energies of semicore $d$ states comes largely from the incomplete cancellation of the self-interaction of $d$ elec-

trons within the LDA, computational schemes (e.g., LDA + $U$ [10], self-interaction corrected (SIC) [11], or other generalized Kohn-Sham methods) in which the self-interaction is effectively removed should give a better mean-field description of the ground state properties. Subsequent quasiparticle calculations within the conventional $G^0W^0$ approach [1] may then be applied.

However, even after the removal of the spurious $pd$ hybridization using the combined LDA + $U$ and $GW$ method [12,13], it is found that the calculated quasiparticle band gap of ZnO is still well below the experimental value. This raises a serious question regarding the accuracy of the conventional $G^0W^0$ approach and prompts us to investigate more carefully another important issue that has not been addressed adequately except for a few cases [14,15], namely, the band convergence issue in the quasiparticle calculations.

Within the $GW$ method [1], the calculation of both the electron irreducible polarizability $\chi^0$ and the Coulomb-hole self-energy $\Sigma^{\text{coh}}$ involves contributions from an infinite number of unoccupied (conduction) bands. Specifically, for the Coulomb-hole self-energy we have

$$\Sigma_{n\mathbf{k}}^{\text{coh}} = \frac{1}{2} \sum_{mq\mathbf{GG}'} \frac{D_{\mathbf{GG}'}^{nm}(\mathbf{k},\mathbf{q})\Omega_{\mathbf{GG}'}^2(\mathbf{q})v(\mathbf{q}+\mathbf{G}')}{\tilde{\omega}_{\mathbf{GG}'}(\mathbf{q})[E - \epsilon_{m,\mathbf{k}-\mathbf{q}} - \tilde{\omega}_{\mathbf{GG}'}(\mathbf{q})]}, \quad (1)$$

where $D_{\mathbf{GG}'}^{nm}(\mathbf{k},\mathbf{q}) = [M_{\mathbf{G}}^{nm}(\mathbf{k},\mathbf{q})]^* M_{\mathbf{G}'}^{nm}(\mathbf{k},\mathbf{q})$, and $M_{\mathbf{G}}^{nm}(\mathbf{k},\mathbf{q}) = \langle m\mathbf{k} - \mathbf{q}|e^{-i(\mathbf{q}+\mathbf{G})\cdot\mathbf{r}}|n\mathbf{k}\rangle$. In the above expression for the Coulomb-hole energy, we adopt the generalized plasmon-pole model [1]; $\Omega_{\mathbf{GG}'}(\mathbf{q})$ and $\tilde{\omega}_{\mathbf{GG}'}(\mathbf{q})$ are the effective bare plasma and mode frequencies, and $v(\mathbf{q} + \mathbf{G}')$ is the Fourier transform of the bare Coulomb potential. In practice, the above summation is always truncated to include a finite number of empty states. However, the band convergence behavior of the $GW$ calculation in general is slow and can be very different from system to system. Even within a single system, the convergence behavior can be

very different for different states. Because of the distinct character of the valence and conduction states of ZnO, an enormous number of empty states has to be included in the summation to achieve converged quasiparticle properties.

We start our discussion by taking a close look at the electronic structure of ZnO calculated within the LDA as shown on the left panel of Fig. 1. All calculations were done using the pseudopotential plane wave approach as implemented in PARATEC [16]. All three semicore subshells of Zn, namely, $3s$, $3p$, and $3d$, are treated as valence electrons. Since these orbitals are strongly localized, a very high-energy cutoff ($E_{cut} = 350$ Ry) for the plane wave expansion of the wave functions is needed. The electronic structure of ZnO calculated within density functional theory (DFT) has been discussed by many authors. Here we summarize the main features. First, the Zn $3d$ states are located at $\sim 5$ eV below the valence band maximum (VBM). This is more than 2 eV too shallow compared with the measured value of 7.5–8.8 eV [17]. Second, these underbound $d$ states strongly hybridize with the valence $p$ states and push these states up, resulting in a greatly reduced gap ($\sim 0.7$ eV) within the LDA. We plot in Fig. 1 the decompositions of band wave functions $\psi_{n\mathbf{k}}$ onto the oxygen $2p$ (left top panel) and the Zn $3d$ (left bottom panel) orbitals as vertical bars (arbitrary scale)

superimposed on the band structure. The size of the vertical bars is proportional to $a_{n\mathbf{k}}^{p,d} = |\langle \phi^{p,d}|\psi_{n\mathbf{k}}\rangle|^2$, where $\phi^{p,d}$ is the atomic oxygen $p$ or zinc $d$ orbital. Strong $pd$ hybridization is clearly seen. Therefore, correction to this largely unphysical $pd$ hybridization is the first step towards an accurate prediction of the quasiparticle properties of ZnO and related materials.

Among various schemes that aim for a better treatment of strongly localized $d$-electron systems at a mean-field level, the LDA + U method [10,18] has been widely recognized as a simple yet powerful approach. In this work, we use a rotationally invariant LDA + U method [18,19] implemented within the pseudopotential plane wave formalism. The screened Coulomb and exchange parameters ($U = 8.0$ eV, $J = 0.9$ eV) used in this work are calculated from first-principles and will be reported separately. The right panels of Fig. 1 shows the band structure of ZnO calculated using the LDA + U method. Decompositions of the band wave functions onto oxygen $2p$ and Zn $3d$ states are shown as vertical bars similar to the case of the LDA band structure. Compared with the LDA band structure shown on the left panels, several differences are readily seen. First, the $d$-bands are located deeper in energy and are detached completely from the valence $p$ states. Second, the direct energy gap increases from 0.7 eV within the LDA to 1.28 eV in the LDA + U method. In addition, the band-width of the oxygen $p$ valence states is increased from about 4.0 eV to 4.9 eV and is in better agreement with experiment. Finally, there is now significantly less $pd$ hybridization. Therefore, it is appealing to carry out $GW$ quasiparticle calculations starting from the LDA + U solutions.

Figure 2 compares the quasiparticle gap of ZnO as a function of the number of conduction bands $N_c$ included in the Coulomb-hole summation, calculated using the LDA/$GW$ and LDA + U/$GW$ methods with a highly converged dielectric matrix which is evaluated with a cut-off energy of 80 Ry, as will be discussed later. The widely quoted value of the quasiparticle gap of ZnO (2.4–2.5 eV) within the conventional LDA/$GW$ approach results from including about 200 conduction bands in evaluating $\Sigma^{coh}$ as indicated by the vertical dashed line in the figure. We mention that this is typically the highest number of conduction bands included in previous $GW$ calculations [5,8,9,20]. As it is clearly shown in the figure, this value is far from being converged. In fact, an enormously larger number of conduction bands ($N_c \sim 3000$) is needed to obtain a converged $GW$ gap. The kinetic energy of the highest conduction band included in the calculation is also shown (see the horizontal scale at the top of the figure). For example, the energy of the $N_c = 3000$ band state is about 67 Ry. The converged gap within the LDA + U/$GW$ approach is slightly over 3.6 eV, which agrees very well with experiment. A question immediately follows: Why does the band gap of ZnO converge so slowly whereas other materials, such as Si or Ge, seem to converge much faster with respect to $N_c$?
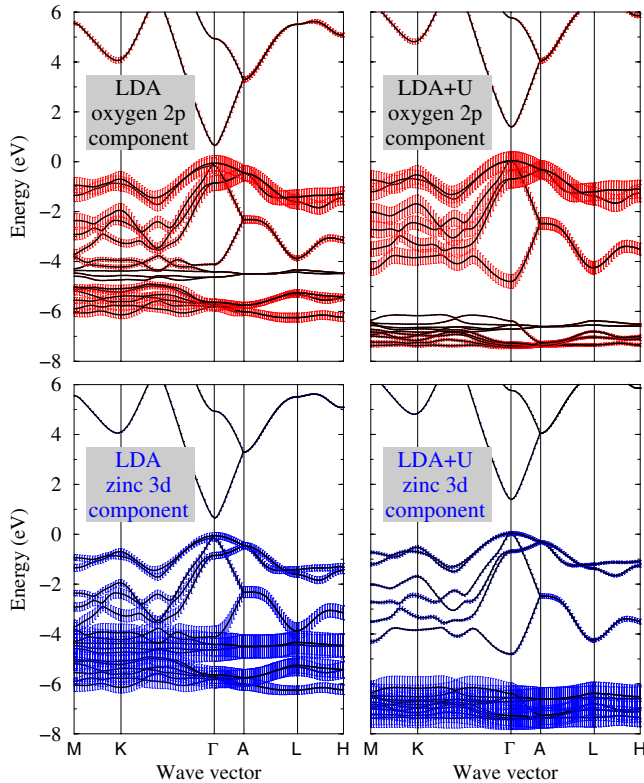


FIG. 1 (color online). The LDA (left panels) and LDA + U (right panels) band structures of ZnO. Projection of the band wave functions onto O $2p$ (top panels) and Zn $3d$ (bottom panels) orbitals are shown as vertical bars superimposed on the band structure.
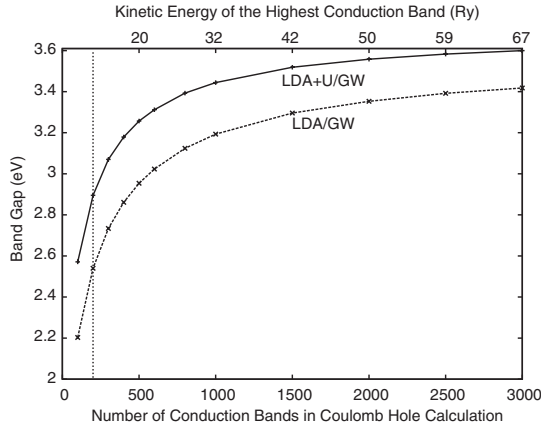
FIG. 2. Quasiparticle band gap of ZnO as a function of the number of conduction bands included in the Coulomb-hole self-energy calculation with a high cutoff energy (80 Ry) for the dielectric function as discussed in the text.

To better understand the convergence behavior of the quasiparticle band gap of ZnO, we analyze the Coulomb-hole self-energy $\Sigma^{coh}$ for the conduction band minimum (CBM) and VBM states. Figure 3 shows the dependence of $\Sigma^{coh}$ on $N_c$ for the CBM and VBM calculated within the LDA/$GW$ and LDA + $U$/$GW$ methods using a converged dielectric function. The different convergence behavior of the CBM and VBM is clearly seen. Whereas $\Sigma^{coh}$ for the CBM converges relatively quickly with respect to $N_c$, that for the VBM converges very slowly. This nonuniform convergence behavior of the VBM and CBM is a result of the dramatically different character of the CBM and VBM wave functions: The CBM is derived from the delocalized Zn 4$s$ state; the VBM, on the other hand, is derived from the fairly localized O 2$p$ state. Figure 4 shows the Fourier components of CBM, VBM, and one of the Zn 3$d$ states calculated within the LDA + $U$ method. Some artificial broadening is applied to smear out the discrete nature

of the wave functions in momentum space. For the CBM state, the contribution to the charge density from plane wave components with kinetic energy $E_G = |G|^2/2 > 10$ Ry is only 2%. In comparison, a kinetic energy of 53 Ry is needed to account for 98% of the VBM charge density. The high-$G$ plane waves components in the wave function are responsible for the strong coupling between the VBM state and high-energy conduction bands. As a result, the convergence of the $\Sigma^{coh}$ for the VBM state is much slower than that for the CBM state. In materials such as Si and Ge, both CBM and VBM are derived from the same atomic $p$ states, the quasiparticle gap converges much faster.

We now investigate in more detail the difference between the LDA/$GW$ and LDA + $U$/$GW$ results. In addition to the difference of about 5% in the converged quasiparticle gaps obtained by the two methods, there are other aspects that should be mentioned. First, the absolute value of the Coulomb-hole energy for the VBM state calculated with the LDA/$GW$ method is about 0.8 eV larger than that calculated with the LDA + $U$/$GW$ method (Fig. 3). This is because within the LDA, the valence $p$ states hybridize strongly with $d$ states as discussed earlier. As a result, the calculated VBM wave function is more "localized" within the LDA than the LDA + $U$ method. Moreover, the quasiparticle corrections to the LDA eigenvalues for the $d$ states are about $-3.8$ to $-4.1$ eV. In comparison, the corrections to the LDA + $U$ eigenvalues for the $d$ states are significantly smaller ($-1.5$ to $-1.8$ eV). This suggests that the LDA + $U$ eigenvalues are indeed closer to the quasiparticle energies.

We would like to address another important issue that relates to the convergence of the $GW$ calculation, i.e., the cutoff energy for the dielectric matrix $\epsilon_{GG'}(\mathbf{q})$. It is expected that at high $\mathbf{G}$ values, the screening effects are negligible and the interaction between electrons is reduced to the bare Coulomb interaction. Therefore, the dielectric matrix can be truncated at a high enough $\mathbf{G}$ value. However, a low cutoff energy also leads to an underestimate of the correlation effects and may give a false con-
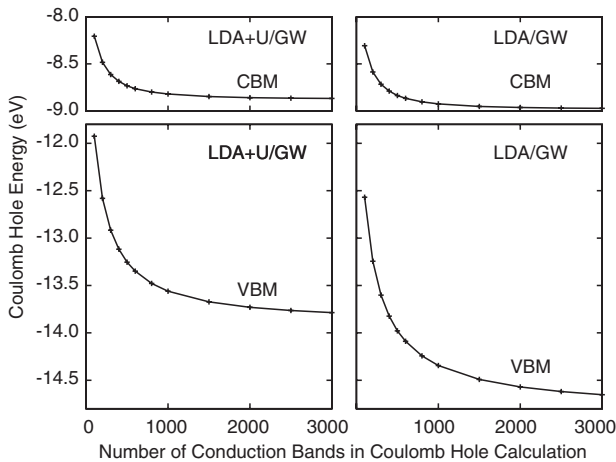


FIG. 3. Coulomb-hole self-energies of the VBM and CBM of ZnO as a function of the number of conduction bands included in the calculation, showing the different convergence behavior between the CBM and the VBM.
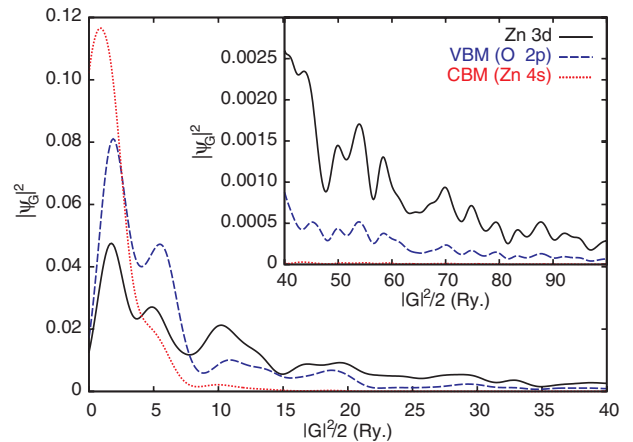


FIG. 4 (color online). Wave functions of CBM, VBM, and a Zn 3$d$ state in the momentum space.
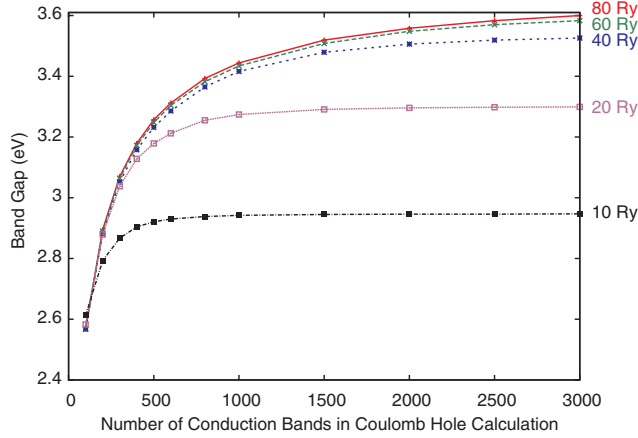
FIG. 5 (color online). Quasiparticle band gap of ZnO calculated within the LDA $+ U/GW$ method and with different dielectric matrix cutoffs, showing the false convergence behavior of the band gap when a small cutoff energy is used.

vergence behavior. Figure 5 shows the calculated quasiparticle band gap for ZnO within the LDA $+ U/GW$ method using different cutoff energies ($E = |\mathbf{G}|^2/2$ Ry) for the dielectric matrix. If a cutoff energy of 10 Ry is used, the band gap converges to about 2.95 eV with as few as 500 conduction bands in the Coulomb-hole energy calculation. This false and rapid convergence can be easily understood within the COHSEX approximation. The Coulomb-hole energy in the static approximation is

$$\Sigma_{n\mathbf{k}}^{\mathrm{coh}} = \frac{1}{2} \sum_{m\mathbf{q}\mathbf{G}\mathbf{G}'} D_{\mathbf{G}\mathbf{G}'}^{nm}(\mathbf{k}, \mathbf{q}) W_{\mathbf{G}\mathbf{G}'}^{\mathrm{pol}}(\mathbf{q}), \qquad (2)$$

where $W^{\mathrm{pol}} = (\epsilon^{-1} - 1)v$ is the polarization potential and the $D$ matrix is defined in Eq. (1). It is obvious that a low cutoff energy for the dielectric matrix will prevent contribution from high-energy conduction bands to the Coulomb-hole self-energy and hence, it leads to a false convergence behavior. The more localized the quasiparticle wave function is, the higher the cutoff energy for the dielectric matrix that is needed for an accurate account for the correlation effects. We note that the cutoff energy for the dielectric matrix used in previous calculations [5] is 10–20 Ry which is not high enough for systems such as ZnO. An absolute convergence is sometimes not necessary for systems in which all relevant electron wave functions have similar character (e.g., Si, Ge, etc) since often it is the energy differences that are of primary interest. However, for systems containing both localized and delocalized states, or in cases in which the absolute energy levels are of interest, an absolute convergence becomes very important.

Finally, we comment on the widely discussed issue about the position of the semicore $d$ states in ZnO and other related systems. The semicore $d$ states are even more localized than the oxygen $p$ states. We find that the Coulomb-hole energy for the Zn $d$ states is still not completely converged even with 3000 conduction bands included in the calculation. In addition, the cutoff energy (80 Ry) for the dielectric matrix used here does not seem to be high enough for an accurate account of the correlation effects for the $d$ states. We believe that many of the previous quasiparticle calculations (including various forms of the self-consistent GW approach) involving localized states may need to be reexamined. Our results on the semicore $d$ levels in ZnO and other related systems will be reported in a follow-up publication.

[1] M. S. Hybertsen and S. G. Louie, Phys. Rev. B **34**, 5390 (1986).
[2] R. W. Godby, M. Schlüter, and L. J. Sham, Phys. Rev. B **37**, 10 159 (1988).
[3] Ü. Özgür et al., J. Appl. Phys. **98**, 041301 (2005).
[4] C. Liu, F. Yun, and H. Morkoç, J. Mater. Sci. Mater. Electron. **16**, 555 (2005).
[5] M. Usuda et al., Phys. Rev. B **66**, 125101 (2002).
[6] S. Tsoi et al., Phys. Rev. B **74**, 165203 (2006).
[7] H. Alawadhi et al., Phys. Rev. B **75**, 205207 (2007).
[8] M. Shishkin and G. Kresse, Phys. Rev. B **75**, 235102 (2007).
[9] F. Fuchs et al., Phys. Rev. B **76**, 115109 (2007).
[10] V. I. Anisimov, J. Zaanen, and O. K. Andersen, Phys. Rev. B **44**, 943 (1991).
[11] J. P. Perdew and A. Zunger, Phys. Rev. B **23**, 5048 (1981).
[12] T. Miyake et al., Phys. Rev. B **74**, 245213 (2006).
[13] E. Kioupakis et al., Phys. Rev. B **77**, 155114 (2008).
[14] M. L. Tiago, S. Ismail-Beigi, and S. G. Louie, Phys. Rev. B **69**, 125212 (2004).
[15] F. Bruneval and X. Gonze, Phys. Rev. B **78**, 085125 (2008).
[16] http://www.nersc.gov/projects/paratec/
[17] L. Ley et al., Phys. Rev. B **9**, 600 (1974).
[18] A. I. Liechtenstein, V. I. Anisimov, and J. Zaanen, Phys. Rev. B **52**, R5467 (1995).
[19] P. Zhang et al., Phys. Rev. B **70**, 085108 (2004).
[20] M. van Schilfgaarde, T. Kotani, and S. V. Faleev, Phys. Rev. B **74**, 245125 (2006).