Origins of band-gap renormalization in degenerately doped semiconductors

Aron Walsh, Juarez L. F. Da Silva, and Su-Huai Wei

National Renewable Energy Laboratory, Golden, Colorado 80401, USA

(Received 30 June 2008; revised manuscript received 8 August 2008; published 28 August 2008)

Degenerate *n*-type doping of semiconductors results in optical band-gap widening through occupation of the conduction band, which is partially offset by the so-called band-gap renormalization. From investigation of the magnitude and origin of these shifts through density-functional band-structure theory, we demonstrate that the key contribution to renormalization arises from the nonparabolic nature of the host conduction band but not the rigid shift of the band edges, as is the current paradigm. Furthermore, the carrier dependence of the band-gap widening is highly sensitive to the electronic states of the dopant ion, which can be involved in a significant reconstruction of the lower conduction band.

DOI: 10.1103/PhysRevB.78.075211

PACS number(s): 71.20.Nr, 61.72.Bb, 78.20.-e

I. INTRODUCTION

Above the critical Mott carrier density,¹ electron doping of semiconductors often results in a widening of the intrinsic optical band gap (E_g) with increasing carrier concentration: the well documented Burstein-Moss shift.^{2,3} This effect is usually attributed to occupation of the conduction band, inducing optical transitions at energies higher than the minimum-energy fundamental electronic gap (Fig. 1). The magnitude of the shift (Δ_{BM}), under free-electron theory, is described as

$$\Delta_{\rm BM} = \frac{\hbar^2}{2m^*} (3\pi^2 n_e)^{2/3},\tag{1}$$

which indicates that it is inversely proportional to the reduced effective mass m^* . m^* is derived from the valence and conduction-band effective masses, m_v^* and m_c^* , according to $\frac{1}{m_v^*} = \frac{1}{m_v^*} + \frac{1}{m_c^*}$, and proportional to $n_e^{2/3}$, where n_e is the electron carrier concentration.

However, the carrier dependence of the measured bandgap shift is generally much less than that expected from the free-electron model [Eq. (1)].^{4–6} To account for this discrepancy, a band-gap renormalization model was proposed^{6,7} where the band-gap shrinkage ($\Delta_{\rm RN}$) is considered as a result of mutual exchange and Coulomb interactions between the added free electrons in the conduction band and electronimpurity scattering. This leads to an increase in the energy of the valence-band maximum (VBM) and decrease in the energy of the conduction-band minimum (CBM).⁷ The net change in E_g can therefore be taken as a difference of the two contributions, i.e., $\Delta E_g = \Delta_{\rm BM} - \Delta_{\rm RN}$ (Fig. 1).

Although this band-gap renormalization model is widely used in the literature, the exact origin of the renormalization effect has not been well tested or understood. This is because, experimentally, the total change (ΔE_g) can be derived from optical-absorption measurements while decomposition into Δ_{BM} and Δ_{RN} component terms is difficult due to uncertainty in the determination of the absolute Fermi level and band-edge positions. Some work has been devoted to quantifying these effects in semiconductor systems^{4–10} but the existing semiempirical understanding offers little in the way of generality. Knowledge of doping effects on the band gap, however, is crucial in determining and improving material performances for optoelectronic devices.¹¹ Of particular interest for such applications are metal oxides due to the coexistence of wide band gaps and high conductivity, e.g., ZnO,^{4,12–14} CdO,^{5,8,15} SnO₂,^{9,16} and In₂O₃.^{6,17–24}

In this paper, through the application of band-structure calculations within the density-functional theory (DFT) formalism, we demonstrate that carrier induced band-gap renormalization cannot be described as a rigid shift of the band edges. While the addition of carriers lowers the energy of both the valence and conduction bands through the effect of an enhanced electron-exchange potential, it is the change in the shape (nonparabolic nature) of the conduction band that has the strongest effect on the shift in optical transitions. Taking In₂O₃ as a prototype, we contrast the predicted $\Delta_{\rm BM}$ from the ground-state effective masses to ΔE_{σ} calculated from a first-principles method for both electron and ion doping. For the former, Δ_{RN} arises purely from the nonparabolic nature of the conduction band. We show that Sn and Ge substitutional dopings produce profoundly different effects: Ge 4s undergoes stronger hybridization with the host conduction band due to its higher binding energy compared to Sn 5s, resulting in a heavier conduction band and subsequently slower rise in the Fermi energy. The implications for optimum doping of oxide systems for optoelectronic applications are discussed.



FIG. 1. (Color online) Illustrated effect of charge carriers on the lowest energy transition from the valence to conduction bands (black vertical line) of a highly doped semiconductor where Δ_{BM} is the Burstein-Moss shift and Δ_{RN} represents band-gap renormalization. Note that a positive Δ_{RN} term refers to a band-gap reduction.



FIG. 2. (Color online) Representation of the 80 atom bixbyite cubic unit cell. Oxygen are colored red (dark gray) while indium are colored blue (light gray) (Ref. 35).

II. COMPUTATIONAL APPROACH

In₂O₃ represents an interesting case as its dispersive conduction band induces a very pronounced ΔE_g upon doping and its optical behavior is of technological importance.^{6,19–23} It has recently been shown that In₂O₃ possesses a direct fundamental band gap on the order of 2.7 eV.^{19,23} Simulations were performed on the body-centered-cubic bixbyite²⁵ (T_h^7) lattice of In₂O₃ [a=10.117 Å (Ref. 26)] (Fig. 2). The electronic structure was obtained using DFT,^{27,28} as implemented in VASP.^{29,30} The In 4*d* states were treated as valence within the projector augmented wave (PAW) framework.³¹ A planewave cutoff of 500 eV and a Γ centered 8×8×8 *k*-point mesh were employed; such dense sampling of the Brillouin zone allows for more precise calculation of the Fermi energy upon doping.

The exchange-correlation effects were treated at the gradient corrected Perdew-Burke-Ernzerhof (PBE) functional level³² with a supplemental correction to the ground-state effective mass, which is generally underestimated by standard DFT functionals due to exaggerated coupling between the valence and conduction bands at the zone center. This was applied in the form of a nonlocal external potential to the conduction-band states, in the spirit of Christensen and co-workers^{33,34} within the PAW method.¹⁴ This corresponds to a partial correction of the band gap from 1.2 to 1.9 eV and a further scissors operator is applied to uniformly shift the conduction band upward so the band gap is in agreement with experiment. While all conclusions may be obtained at the PBE level, this correction enables a more reliable and quantitative description of the effective-mass changes.

Electron doping is achieved by varying the number of electrons with the excess charge compensated through a homogeneous jellium background. The occupation of the conduction-band states is determined by a Gaussian smearing width of 0.05 eV. Extrinsic isochoric ion doping is



FIG. 3. (Color online) Relationship between the total band-gap shift $\Delta E_g, \Delta_{BM}, \Delta_{RN} = \Delta_{BM}^{Eqn.1} - \Delta E_g$ (dashed line), and carrier concentration for electron doping. Results from both the parabolic [Eq. (1)] and nonparabolic [Eq. (4)] models are shown.

achieved through explicit defect calculations with supercell sizes up to 320 atoms for the lowest doping levels. In each case there is a single substitutional dopant atom situated at the center of a bixbyite supercell.

III. RESULTS

From total energy and force minimization of the bulk In_2O_3 cell, we obtain an equilibrium lattice constant a =10.260 Å. The resulting ground-state electron effective mass of $0.24m_e$ is in very good agreement with recent hybrid-DFT (HSE03) calculations $(0.22m_e)$ (Ref. 21) and is in good agreement with the reported experimental range $[m_{\nu}^* \approx 0.28 - 0.35 \ m_{e} \ (\text{Ref. 36})]$, taking into account the difficulty in achieving low carrier concentrations. The electron mass is found to be almost isotropic; however, in order to take all effects into consideration, the effective masses are weight averaged from the dispersion along the (001) and (111) directions close to the zone center. As the valence band is derived mainly from ionic O 2p orbitals, there is minimal band dispersion. From our calculations we find a valenceband effective mass of $16.14m_e$. While the hole mass of In_2O_3 is unknown experimentally, such a heavy mass is consistent with the failure in achieving *p*-type conductivity. The resulting variation of Δ_{BM} , derived from Eq. (1) with m^* $=0.22m_e$, is plotted in Fig. 3. As we can see, due to the small effective mass at the zone center, Δ_{BM} increases strongly as a function of the carrier density if a parabolic conduction band is assumed.

A. Free-electron doping effects

In Fig. 3, the directly calculated total band-gap shift ΔE_g is plotted. We find that the parabolic $\Delta_{\rm BM}$ overestimates the calculated optical gap expansion with the difference $\Delta_{\rm RN} = \Delta_{\rm BM} - \Delta E_g$ increasing with carrier concentration from 41 to 368 meV for n_e from 1 to 9×10^{20} cm⁻³. To understand the origin of this band-gap reduction, we first look at the shrink-age of the fundamental band gap due to the electron-

exchange interactions. We find that free-electron doping results in almost negligible fundamental gap shrinkage with a 17 meV decrease in the $\Gamma - \Gamma$ separation for concentrations as high as 10^{21} cm⁻³. With reference to the O core levels, we find this is because *both* the valence and conduction bands are lowered in energy due to the effect of the more attractive electron-exchange potential; however, the conduction band drops slightly faster, resulting in the mild overall shrinkage. This is at variance with the previous understanding that the valence band rises and the conduction band falls with the addition of carriers.^{6,7}

Our calculations show that, on the addition of electrons, the position and shape of the host conduction band remains mostly unperturbed except for the negligible shrinkage effects. Further investigation reveals that $\Delta_{\rm RN}$ (i.e., $\Delta_{\rm BM} - \Delta E_g$) emerges due to the *k* dependence in the shape of the conduction band not taken into account from the simple parabolic model [Eq. (1)]. The calculated effective mass strongly increases as the Fermi level rises. To account for the nonparabolic nature of the conduction band, we follow Young *et al.*¹² by introducing a function $\gamma(E)$ to describe the effective-mass *k* dependence, where

$$\frac{\hbar^2 k^2}{2m^*} = \gamma(E) = E + \frac{E^2}{E_1}.$$
 (2)

It therefore follows^{10,12} that the resulting carrier dependent transport effective mass (m_T^*) will obey the relation

$$m_T^*(E) = m_0^* \frac{d\gamma}{dE},\tag{3}$$

where m_0^* is the ground-state mass. In this way, once the parameter E_1 is fitted, the correct Δ_{BM} energy dependence can be obtained through an augmented version of Eq. (1):

$$\Delta_{\rm BM} = \frac{\hbar^2}{2m^*} (3\pi^2 n_e)^{2/3} \frac{1}{(1+\beta)},\tag{4}$$

where,

$$\beta = \frac{m_T^*(E)}{2m_0^*} - \frac{1}{2}.$$
 (5)

A fit to the conduction-band dispersion for stoichiometric In_2O_3 indicates that $E_1=2.24$ eV can reproduce the observed k dependence (Fig. 3). This clearly demonstrates that even for the ideal case of electron doping, a parabolic model is insufficient in describing quantitative changes in the bandgap shift and the error can be corrected by taking into account the nonparabolic behavior of the conduction band.

It should be noted that our values of $\Delta_{\rm RN}$ are on the order of 50% less than prior estimations for *n*-type In₂O₃.⁶ Such exaggerated $\Delta_{\rm RN}$ in previous work is related to the fitting to an extremely light valence-band effective mass (0.6m_e), resulting in a large $\Delta_{\rm BM}$ term. Due to the cancellation of errors in $\Delta_{\rm BM} - \Delta_{\rm RN}$, a reasonable value for ΔE_g was produced, in overall agreement with experimental data.

B. Ion doping effects

In real systems, *n*-type doping is often performed by introducing extrinsic dopants such as Ge and Sn substituted on



FIG. 4. (Color online) Calculated relationship between the total band-gap change ΔE_g , and carrier concentration for Sn_{In} and Ge_{In}. The magnitude of $\Delta_{\rm RN}$ ($\Delta_{\rm BM}^{\rm Eqn.1} - \Delta E_g$) is also indicated for each case (dashed lines). Experimental data for $\Delta E_g^{\rm Sn}$ [stars (Ref. 6) and squares (Ref. 36)] are also shown.

the In lattice sites in In₂O₃. The +1 difference in effective oxidation state results in a single electron donation per substitution. Both dopants result in small shrinkage of the fundamental band gap at Γ due to their lower *s* orbital energy, which induces hybridization with the In *s* conduction band, e.g., Sn_{In} at doping levels of 1×10^{21} cm⁻³ lowers the gap by 238 meV while for Ge_{In}, this increases to 286 meV.

Interestingly, both Eq. (1) and the $\Delta_{\rm RN}$ model proposed by Berggren and Sernelius⁷ are dopant independent but, clearly from the plots of ΔE_g (Fig. 4), there are profound differences between Sn_{In} and Ge_{In}, which approach 500 meV for the highest dopant concentrations. The calculated energy shifts for the two systems are much less than those predicted by $\Delta_{\rm BM}$ from both the parabolic and nonparabolic models fitted to In₂O₃. Our calculated ΔE_g for Sn_{In} is in good agreement with experimental observations.^{6,36}

Compared to Δ_{BM} arising from the parabolic model, and the calculated total band-gap shifts, ΔE_g , for Sn and Ge, the magnitudes of Δ_{RN} reach up to 600 (Sn) and 1100 meV (Ge) for 10²¹ cm⁻³ carriers: the changes are strongly dopant dependent (Fig. 4). The influence of each dopant on the host conduction band is apparent in the calculated band structures (Fig. 5). The conduction-band dispersion is significantly altered upon substitution of Sn and Ge into the lattice. This arises from hybridization of the dopant s states with the host conduction band: on doping, the CBM state gains a significant amount of dopant s character (Fig. 5). The Sn 5s orbital lies at higher binding energy than In 5s due to the increased nuclear charge, and their coupling results in both a lowering of the bottom-most conduction band and a mild decrease in the band curvature near the band edge. Ge 4s, which lies to higher binding energy than Sn 5s, induces a greater effect with a significant flattening of the conduction band. This is the origin of the greatly increased Δ_{RN} contributions for Ge_{In}; the decreased conduction-band dispersion results in a much slower rise in the Fermi level on the addition of carriers. Our calculated increase in the effective mass of In₂O₃ on Sn doping tends toward $0.43m_e$, which agrees well with measure-



FIG. 5. (Color online) Band structure of bulk In₂O₃ along the Γ -*N* line, and Sn_{In} and Ge_{In} doped cells at an effective carrier concentration of 5×10^{20} cm⁻³. The top of the valence band is set to 0 eV. The atomic contributions to the CBM state at Γ are also listed in each case.

ments of up to $0.4m_e$.³⁶ Unfortunately, Ge doping of In₂O₃ has not been investigated in detail but we predict a dampened rise in Fermi energy and increased effective mass (up to $0.76m_e$) in comparison to Sn doping.

IV. DISCUSSION

The majority of metal oxides exhibit only *n*-type behavior due to the large electronegativity of oxygen³⁷ with the good conductivity originating from the delocalized conductionband *s*-like wave function. This results in a significant blueshift of the optical band gap upon doping. Indeed for CdO, this effect is enough to lift the absorption edge above the visible range, increasing the material transparency with carrier concentration.^{5,8} Our calculations on Sn and Ge doping of In_2O_3 highlight the different effects that formally isolvalent dopants can produce. The question therefore arises on the choice of optimal dopants for a given system, which will leave the host band structure most unperturbed except for the addition of electrons.

Analysis of the atomic energy levels³⁸ can act as a simple and effective guide for predicting such dopant effects. For group 12 oxides (ZnO and CdO), the metal valence *s* states are relatively high in energy and the best *n*-type dopants are Al>In>Ga, in order of decreasing *s* orbital energy. For group 13 oxides (Ga₂O₃, In₂O₃), the best dopants are Sn >Si>Ge while for group 14 oxides (SnO₂, PbO₂), Sb>P >As. The low 4*s* orbital energy for the row 4 elements (Ga, Ge, and As) is due to the imperfect screening of the 3*d* orbitals. For all systems, both row 2 (B, C, and N) and row 6 (Tl, Pb, and Bi) cations possess valence *s* states too low in energy to act as effective *n*-type dopants if maintenance of a large optical band gap is desired. This has recently been demonstrated in Bi doping of PbO₂ where Bi can in fact act as an acceptor because, due to the large relativistic effects, the high binding energy of the Bi 6*s* states can stabilize a lower +3 charge state.³⁹

It is quite satisfying that these predications match the dopants adopted empirically in recent years with ZnO:Al, In₂O₃:Sn, and SnO₂:Sb being the most widely employed transparent conducting metal oxide systems.

V. CONCLUSIONS

In summary, we have investigated the origins of band-gap renormalization in highly doped semiconductors, taking In_2O_3 as a prototype. By decomposing ΔE_g into Burstein-Moss (Δ_{BM}) and band-gap renormalization (Δ_{RN}) terms using first-principles calculations, we have elucidated the dependence of ΔE_g on the carrier density and the chemical effects of the dopant ion. Assumption of a parabolic effective mass, as is generally considered for ΔE_g , and rigid shifts of the band edges, as previously hypothesized for Δ_{RN} , are insufficient for even a qualitative understanding of the bandgap shift. We show that Δ_{RN} arising from electron doping is caused mainly by the nonparabolic shape of the host conduction band. For impurity doping, hybridization with the host conduction states can further enhance Δ_{RN} , which depends sensitively on the relative position of the dopant s energy levels.

Based on these results, a simple guide for determining the optimal dopants for a given system was discussed with Al, Sn, and Sb recognized as the best *n*-type dopants for group 12, 13, and 14 metal oxides, respectively.

ACKNOWLEDGMENTS

We would like to thank R. G. Egdell for sharing his knowledge of doped semiconductors. The work at NREL is supported by the U.S. Department of Energy (DOE) under Contract No. DE-AC36-99GO10337. Computing resources of the National Energy Research Scientific Computing Center were employed, which is supported by DOE under Contract No. DE-AC02-05CH11231.

- ²T. S. Moss, Proc. Phys. Soc. London, Sect. B 67, 775 (1954).
- ³E. Burstein, Phys. Rev. **93**, 632 (1954).
- ⁴J. G. Lu et al., J. Appl. Phys. **101**, 083705 (2007).
- ⁵P. H. Jefferson, S. A. Hatfield, T. D. Veal, P. D. C. King, C. F. McConville, J. Zuniga-Pérez, and V. Munoz-Sanjosé, Appl. Phys. Lett. **92**, 022101 (2008).
- ⁶I. Hamberg, C. G. Granqvist, K. F. Berggren, B. E. Sernelius, and L. Engstrom, Phys. Rev. B **30**, 3240 (1984).
- ⁷K. F. Berggren and B. E. Sernelius, Phys. Rev. B **24**, 1971 (1981).
- ⁸Y. Dou, T. Fishlock, R. G. Egdell, D. S. L. Law, and G. Beamson, Phys. Rev. B **55**, R13381 (1997).
- ⁹R. G. Egdell, J. Rebane, T. J. Walker, and D. S. L. Law, Phys.

¹N. F. Mott, Philos. Mag. 6, 287 (1961).

Rev. B 59, 1792 (1999).

- ¹⁰D. Segev and S.-H. Wei, Phys. Rev. B **71**, 125129 (2005).
- ¹¹G. Thomas, Nature (London) **389**, 907 (1997).
- ¹²D. L. Young, T. J. Coutts, V. I. Kaydanov, A. S. Gilmore, and W. P. Mulligan, J. Vac. Sci. Technol. A **18**, 2978 (2000).
- ¹³ A. Walsh, J. L. F. Da Silva, and S.-H. Wei, Phys. Rev. Lett. 100, 256401 (2008).
- ¹⁴S. Lany, H. Raebiger, and A. Zunger, Phys. Rev. B 77, 241201(R) (2008).
- ¹⁵Y. Z. Zhu, G. D. Chen, H. Ye, A. Walsh, C. Y. Moon, and S.-H. Wei, Phys. Rev. B **77**, 245209 (2008).
- ¹⁶J. Robertson, Phys. Rev. B **30**, 3520 (1984).
- ¹⁷R. L. Weiher and R. P. Ley, J. Appl. Phys. **37**, 299 (1966).
- ¹⁸J. E. Medvedeva, Phys. Rev. Lett. **97**, 086401 (2006).
- ¹⁹A. Walsh et al., Phys. Rev. Lett. **100**, 167402 (2008).
- ²⁰S. Z. Karazhanov, P. Ravindran, P. Vajeeston, A. Ulyashin, T. G. Finstad, and H. Fjellvag, Phys. Rev. B **76**, 075129 (2007).
- ²¹F. Fuchs and F. Bechstedt, Phys. Rev. B 77, 155107 (2008).
- ²²P. Erhart, A. Klein, R. G. Egdell, and K. Albe, Phys. Rev. B 75, 153205 (2007).
- ²³A. Bourlange, D. J. Payne, R. G. Egdell, J. S. Foord, P. P. Edwards, M. O. Jones, A. Schertel, P. J. Dobson, and J. L. Hutchison, Appl. Phys. Lett. **92**, 092117 (2008).

- ²⁴H. Raebiger, S. Lany, and A. Zunger, Phys. Rev. Lett. 101, 027203 (2008).
- ²⁵P.-A. Glans, T. Learmonth, K. E. Smith, J. Guo, A. Walsh, G. W. Watson, F. Terzi, and R. G. Egdell, Phys. Rev. B **71**, 235109 (2005).
- ²⁶M. Marezio, Acta Crystallogr. 20, 723 (1966).
- ²⁷W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).
- ²⁸ P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964).
- ²⁹G. Kresse and J. Furthmüller, Phys. Rev. B 54, 11169 (1996).
- ³⁰G. Kresse and J. Furthmüller, Comput. Mater. Sci. 6, 15 (1996).
- ³¹G. Kresse and D. Joubert, Phys. Rev. B **59**, 1758 (1999).
- ³²J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. 77, 3865 (1996).
- ³³N. E. Christensen, Phys. Rev. B **30**, 5753 (1984).
- ³⁴A. Kamińska et al., Phys. Rev. B **76**, 075203 (2007).
- ³⁵K. Momma and F. Izumi, J. Appl. Crystallogr. 41, 653 (2008).
- ³⁶Y. Ohhata, F. Shinoki, and S. Yoshida, Thin Solid Films **59**, 255 (1979).
- ³⁷S.-H. Wei, Comput. Mater. Sci. **30**, 337 (2004).
- ³⁸S.-H. Wei and A. Zunger, Phys. Rev. B **60**, 5404 (1999).
- ³⁹S. Rothenberg, D. J. Payne, A. Bourlange, and R. G. Egdell, J. Appl. Phys. **102**, 113717 (2007).