IOPscience iopscience.iop.org

Home Search Collections Journals About Contact us My IOPscience

Computational studies of conductivity in wide-band-gap semiconductors and oxides

This article has been downloaded from IOPscience. Please scroll down to see the full text article.

2008 J. Phys.: Condens. Matter 20 064230

(http://iopscience.iop.org/0953-8984/20/6/064230)

View the table of contents for this issue, or go to the journal homepage for more

Download details:
IP Address: 129.252.86.83
The article was downloaded on 29/05/2010 at 10:32

Please note that terms and conditions apply.

IOP PUBLISHING JOURNAL OF PHYSICS: CONDENSED MATTER

J. Phys.: Condens. Matter **20** (2008) 064230 (7pp) doi:10.1088/0953-8984/20/6/064230

Computational studies of conductivity in wide-band-gap semiconductors and oxides

C G Van de Walle

Materials Department, University of California, Santa Barbara, CA 93106-5050, USA

E-mail: vandewalle@mrl.ucsb.edu

Received 31 August 2007, in final form 29 October 2007
Published 24 January 2008
Online at stacks.iop.org/JPhysCM/20/064230

Abstract

The ability to control conductivity is essential for design and fabrication of (opto)electronic devices. Such conductivity control has traditionally been very difficult in wide-band-gap semiconductors, and native point defects have often been invoked to explain these problems. State-of-the-art first-principles calculations based on density functional theory have been used to elucidate these issues. Approaches for overcoming the 'band-gap problem', including the LDA + U method, allow more accurate comparisons and predictions of defect levels. The methodology is illustrated with the case of native point defects in zinc oxide. Computations reveal that the prevailing n-type conductivity cannot be attributed to native defects; it must thus be caused by impurities that are unintentionally incorporated. Hydrogen is shown to be an excellent candidate for such an impurity.

1. Introduction

Wide-band-gap semiconductors have rapidly moved to the forefront of semiconductor research and technology. The potential of materials such as SiC and GaN had been recognized for many decades, but it was only with the advent of modern epitaxial growth techniques such as molecular beam epitaxy (MBE) and metal–organic chemical vapor deposition (MOCVD) that films of sufficiently high quality could be obtained to fully exploit the materials properties. We are currently seeing similar developments in wide-band-gap oxides: materials such as ZnO, SnO_2, and In_2O_3 are being produced with higher quality than ever before, allowing the prospect of unprecedented performance and new applications.

One of the key issues that has hampered progress in applications of all of these materials is the control of their conductivity. The wider the band gap, the more difficult it becomes to achieve ambipolar conductivity (i.e., n-type conduction due to electrons generated by introducing donor impurities, and p-type conduction due to holes generated by introducing acceptor impurities). Typically, the wide-band-gap materials exhibit one type of conductivity (usually n-type), with the other type being difficult to achieve. The cause of this behavior has been debated for more than 50 years, the most widespread hypothesis being that native point defects are the cause of the unintentional conductivity. The justification was based on the following arguments: (1) The unintentional conductivity occurs even under conditions where growers think they can exclude the participation of impurities. (2) There is a perception that it is difficult to achieve stoichiometry, leading to the expectation that the deviation from stoichiometry would be accommodated through point defects. (3) Changes in partial pressures of the constituents during the growth process or during post-growth annealing have been observed to be consistent with the hypothesis that native point defects are responsible. We note that these arguments are all based on 'circumstantial evidence'; in fact, we are not aware of any case in which a direct unambiguous experimental identification of a point defect has been correlated with an increase in conductivity.

Experimental identification of native point defects and quantitative measurement of their concentration is very difficult, a problem that has impeded resolution of this important issue. The advent of first-principles calculations for native defects and impurities has made significant contributions to progress in this field. Indeed, such calculations have allowed the study of the atomic and electronic structure as well as the energetics of individual point defects at a level of detail that would be difficult to achieve in experiment, and unhampered by experimental complications such as unintentional contamination by impurities. As in so many cases where computational studies have had an impact, the advances were based on a number of parallel accomplishments: (a) development of a theoretical formalism (in this case

 © 2008 IOP Publishing Ltd Printed in the UK

based on the thermodynamics of defect formation energies) that provides a framework for use and interpretation of the computational data; (b) better algorithms and codes enabling faster computations for systems up to several hundred atoms; (c) faster computers as well as the use of parallelization; (d) verification, by comparing key results with experiment in a number of well-established cases; and (e) constant attention to the limits of validity of the underlying theory (density functional theory), and ongoing efforts to overcome those limitations to be able to address a wider range of problems and materials.

In the remainder of this paper we will have occasion to illustrate these developments with specific examples, focusing on a case study for ZnO, a material of formidable technological interest [1–3]. In section 2 we will define the key quantities that control the physical observations and describe the theoretical and computational formalism. Section 3 summarizes some major results for defects and doping in ZnO. Section 4, finally, summarizes the paper and identifies fruitful areas for continued studies.

2. Methodology

2.1. Concentration

In order to assess the impact of a defect or impurity on the conductivity of a solid, two quantities are crucial: the concentration and the ionization energy. In the following, we will use the term 'defect' to generically refer to both native point defects and impurities. Defects that occur in low concentrations have a small or negligible impact on conductivity; only those whose concentration exceeds a threshold will have observable effects. The concentration is determined by the *formation energy* through the expression:

$$c = N_{\mathrm{sites}} \exp\left(-E^{\mathrm{f}}/kT\right), \tag{1}$$

where E^{f} is the formation energy, N_{sites} is the number of sites on which the defect can be incorporated, k is the Boltzmann constant, and T the temperature.

The expression for concentration (1) assumes thermodynamic equilibrium. In principle, defects could also occur in nonequilibrium concentrations. In practice, however, the majority of bulk and epitaxial film growth techniques are close to equilibrium. Equilibration of defects is unavoidable if the diffusion barriers are low enough to allow easy diffusion at the temperatures of interest, and this is indeed the case in ZnO; a specific example will be discussed in section 3.2. Even if kinetic barriers would be present (1) is still relevant because obviously defects with a high formation energy are less likely to form.

We now focus on the definition of the formation energy [4, 5] for a native point defect, illustrating it with the specific example of an oxygen vacancy in a 2+ charge state in ZnO:

$$E^{\mathrm{f}}(V_{\mathrm{O}}^{2+}) = E_{\mathrm{tot}}(V_{\mathrm{O}}^{2+}) - E_{\mathrm{tot}}(\mathrm{ZnO}) + \mu_{\mathrm{O}} + 2E_{\mathrm{F}}. \tag{2}$$

$E_{\mathrm{tot}}(V_{\mathrm{O}}^{q})$ is the total energy of the supercell containing the (fully relaxed) defect, and $E_{\mathrm{tot}}(\mathrm{ZnO})$ is the total energy of the same supercell containing the ZnO perfect crystal. The Fermi energy E_{F} is the chemical potential of electrons, i.e., the energy of the reservoir with which electrons are exchanged. The oxygen atom that is removed is placed in a reservoir, the energy of which is the oxygen chemical potential μ_{O}. It is a variable, corresponding to the experimental notion that ZnO can be grown under oxygen-rich or oxygen-poor conditions. However, μ_{O} is subject to an upper bound equal to the energy of molecular O_2. In equilibrium, the sum of μ_{Zn} and μ_{O} corresponds to the energy of ZnO. An upper bound on μ_{Zn}, set by the energy of bulk Zn, therefore leads to a lower bound on μ_{O}. The range over which the chemical potentials can vary is thus given by the enthalpy of formation of ZnO, which we calculate to be −3.50 eV (exp.: −3.60 eV [6]).

The formation energy in (1) is, in principle, a *free energy*, i.e., it may contain vibrational entropy contributions. Such terms are usually small, and there is often a significant cancelation effect between vibrational contributions in the solid and in the reservoir [5]. Inclusion of entropy terms sometimes has a distinct impact on which configuration of a defect or impurity is most stable [7], but rarely has a significant effect on the overall concentration. Further details about the calculation of formation energies are provided in [5].

2.2. Ionization energy

A hallmark of electrically active defects is that they can occur in different charge states. For each position of the Fermi level, one particular charge state has the lowest energy. The Fermi-level positions at which the lowest-energy charge state changes are called *transition levels*. Once the formation energies are known, the transition levels immediately follow by taking energy differences:

$$\varepsilon(q/q') = [E^{\mathrm{f}}(D^{q}; E_{\mathrm{F}} = 0) - E^{\mathrm{f}}(D^{q'}; E_{\mathrm{F}} = 0)]/(q' - q), \tag{3}$$

where $E^{\mathrm{f}}(D^{q}; E_{\mathrm{F}} = 0)$ is the formation energy of the defect D in the charge state q when the Fermi level is at the valence-band maximum ($E_{\mathrm{F}} = 0$). When atomic relaxations are fully included in the calculation of the formation energies for both charge states, a thermodynamic transition level is obtained. The experimental significance of this level is that for Fermi-level positions below $\varepsilon(q/q')$ charge state q is stable, while for Fermi-level positions above $\varepsilon(q/q')$, charge state q' is stable. The transition levels should not to be confused with the Kohn–Sham states that result from band-structure calculations for a single charge state. We also note that in optical experiments (luminescence or absorption) the final state may not be completely relaxed, leading to different values for optical levels; for a discussion, see [5].

In order for a defect or impurity to make a contribution to conductivity, it has to be stable (at room temperature) in a charge state that is consistent with the presence of free carriers. For a donor-type defect, this means that the defect should be stable in a positive charge state throughout the band gap, and that the transition level from a positive to the neutral charge state should occur close to or above the conduction-band minimum (CBM). Defects with an ionization energy on the order of a few kT are easily ionized and are called 'shallow

defects' [5, 8]. Note that this usually corresponds to a situation where the +/0 transition level, as defined based on formation energies, would lie above the CBM. In practice, of course, a neutral charge state in which the electron is localized in the immediate vicinity of the defect cannot be maintained if the corresponding electronic level is resonant with the conduction band; instead, the electron will be transferred to extended states, but may still be bound to the positive core of the defect in a hydrogenic effective mass state. This is what happens, in practice, for shallow donors.

2.3. First-principles calculations

The quantities in (2) can be explicitly evaluated by performing density functional theory (DFT) [9] calculations in the local density approximation (LDA) [10] or generalized gradient approximation (GGA) [11, 12]. Introduction of a defect breaks the translational symmetry of the crystal; however, periodicity can be maintained by using a *supercell geometry*, in which the defect is placed in a cell that is a multiple of the primitive unit cell of the crystal. This allows continued use of periodicity in the calculations, which is often exploited through the use of Fast Fourier Transforms, etc. The use of supercells also has the advantage that the underlying band structure of the host remains properly described. An alternative approach would be to use a large cluster with a defect in the center. However, for computationally tractable sizes quantum confinement may significantly affect the host band structure. In addition, surface effects could interfere with the properties of the defect. In practice, the supercell method has become the most widely used approach for studying defects. Convergence as a function of supercell size should always be checked, to make sure that the quantities that are derived are representative of an isolated defect.

Supercells with 32, 64, 128, 216, and 256 atoms are typically used for materials with the zinc-blende structure. For wurtzite-structure materials, 32, 48, 72, and 96 atom cells have been used. These fairly large cell sizes call for efficient computational approaches. Pseudopotential [13, 14] and projector-augmented-wave [15, 16] approaches have proven ideal for tackling these large systems, as implemented in highly optimized codes such as the Vienna *Ab initio* Simulation Program (VASP) [17–19]

2.4. Shortcomings of DFT-LDA and how to address them

2.4.1. Semicore states. Many of the wide-band-gap materials of interest have semicore states that play an important role in the electronic structure of the material. For instance, in ZnO the Zn 3d states occur less than 8 eV below the valence-band maximum (VBM), and therefore can be expected to strongly interact with the s and p states of the valence band. Inclusion of the d states as valence states (as opposed to treating them as core states) is therefore important in order to allow hybridization. Indeed, it has been established that proper inclusion of the d states can have significant effects on band offsets, deformation potentials, and also on structure and energetics [20–22]. Unfortunately, the higher degree of localization and correlation within the d bands renders them less well described in DFT-LDA. Here we briefly mention two approaches that have been explored for correcting this problem.

- Performing *exact exchange* calculations in the optimized effective potential formalism [23, 24] leads to significant improvements. However, these calculations are at present too computationally demanding to be applied to anything but the simplest bulk systems.
- Applying the LDA + U (or GGA + U) approach, in which the electrons are separated into two subsystems: localized semicore d electrons for which the Coulomb repulsion U is taken into account via a Hubbard-like term in a model Hamiltonian, and delocalized electrons that are described by the usual orbital-independent one-electron potential. This approach had been developed and applied for materials with partially filled d shells [25, 26], but we have recently shown it greatly improves the computed properties of materials with filled d shells such as GaN and InN, as well as ZnO and CdO [21, 22]. An important issue is the choice of the parameter U. This has often been treated as a fitting parameter, with the goal of reproducing either (a) the experimental band gap or (b) the experimentally observed position of the d states in the band structure. Neither approach can be justified, because (a) LDA + U cannot be expected to correct for other shortcomings of DFT-LDA, specifically, the lack of a discontinuity in the exchange–correlation potential (see section 2.4.2), and (b) experimental observations of semicore states may include additional ('final state') effects inherent in experiments such as photoemission spectroscopy. We have therefore proposed an approximate but consistent and unbiased approach of determining U based on atomic calculations and screening this value by dividing by the optical dielectric constant of the solid [21]. Tests on a number of systems have shown that application of this formalism lowers the energy of the semicore d states, thus reducing their coupling with the p states at the VBM; simultaneously, the increased screening due to more strongly bound d states increases the energy of the s states at the CBM, leading to further opening of the band gap. Overall, the LDA + U approach provides a significantly improved description of materials with semicore states.

2.4.2. The 'band-gap problem'. The use of DFT-LDA/GGA implies that excited-state properties are not properly described. It has often been assumed that this is not a problem for defect calculations, since each individual calculation for a specific charge state could be considered to be a ground-state calculation. This notion is incorrect, in the same way that it would be incorrect to assume that LDA calculations could yield a reliable total-energy difference between an N-electron versus an $(N + 1)$-electron system [27]. Indeed, the change in the number of electrons elicits the issue of the lack of a discontinuity in the exchange–correlation potential, which is at the root of the band-gap problem [28]. Similarly, the formation energy expressed in (2) involves changes in the occupation of defect-induced states.

Phrased another way: if a specific charge state of a defect involves occupying a state in the band gap, and the band gap is incorrect in DFT-LDA, then the position of the defect state and hence the calculated total energy will suffer from this problem. Careful practitioners have always been aware of this problem and refrained from drawing conclusions that might be affected by these uncertainties. The problem is exacerbated, of course, in the case of wide-band-gap semiconductors in which the band-gap errors can be particularly severe; for example, in ZnO the LDA band gap is only 0.8 eV, compared to an experimental value of 3.4 eV.

Several approaches have been, or are being, developed to overcome these problems:

- Quasiparticle calculations in the GW approximation produce band structures in close agreement with experiment [29]. However, at present the calculation of total energies within the GW formalism [30] is still a subject of active research and currently not suitable for application to the problems we are addressing here. In addition, we note that quasiparticle calculations based on the GW approach do not correct the underbinding of semicore states [24]. As discussed in section 2.4.1, this underbinding can make a significant contribution to the band-gap error.
- Hybrid functionals are based on the inclusion of a small fraction of non-local exchange in the Hamiltonian [31–33]. The range of applicability of this approach is still being explored, and the computational demands of any method that requires evaluation of exchange will be very severe for realistic defect calculations.
- Extrapolation schemes can be based on evaluation of defect properties for two different values of the band gap followed by an extrapolation to the experimental gap. A number of empirical extrapolation approaches were described by Zhang *et al* [34], for instance based on use of different exchange and correlation potentials or different plane-wave cut-offs. Such extrapolation schemes are most likely to be successful if the calculations that produce different band gaps are physically motivated, ensuring that the shifts in defect states that give rise to changes in formation energies reflect the underlying physics of the system. A particularly attractive approach, in this respect, is an extrapolation based on LDA and LDA + U calculations, as described in [35] and [36]. As discussed in section 2.4.1, LDA + U produces genuine improvements in the electronic structure related to the energetics of the semicore states; one of these effects is an increase in the band gap. The shifts in defect-induced states between LDA and LDA + U reflect their relative valence- and conduction-band character, and hence an extrapolation to the experimental gap is expected to produce reliable results. A particular example in which this could be verified through direct comparison with experiment will be discussed in section 3.1.
- Modifications in the pseudopotentials can produce much improved band structures while maintaining reliable structure and energetics. Within the pseudopotential approach, there is no unique scheme for generating pseudopotentials. Indeed, a number of different generation schemes have been proposed over the years, often aimed at creating 'softer' potentials which can be described with a smaller plane-wave basis set. This flexibility can be exploited to generate potentials that produce a more accurate band structure, However, past attempts have not succeeded in producing such improvements while still maintaining a proper description of atomic structure and energetics [37].

A new approach has recently proven to be remarkably successful for nitride semiconductors [38, 39]. It is based on a proposal by Christensen, first implemented within the linearized muffin-tin orbital method [40], to add a highly localized (delta-function-like) repulsive potential centered on the atomic nucleus of the cation. Such a potential only affects s states, and since the CBM in compound semiconductors has largely cation s character one expects an upward shift of the corresponding eigenstates. At the same time, the highly localized character of the added potential leads one to expect only minimal changes in other aspects of the pseudopotential. These expectations were indeed borne out in the case of GaN and InN, where the modified pseudopotentials produced atomic structures and energetics that are as reliable as those obtained with standard potentials, but simultaneously producing band structures in very good agreement with experiment [39]. Even though the fitting procedure only aimed to produce the experimental value of the direct gap, the modified potentials actually produced improvements for other aspects of the band structure as well, including the position of higher-lying indirect conduction-band minima as well as the position of semicore d states [39]. This leads us to believe that the seemingly ad hoc modifications introduced by the repulsive potential are capturing some essential physics, and justifies the expectation that similarly good results can be obtained for other materials. An application of the modified pseudopotentials to the calculation of the electronic structure of nitride surfaces produced results in very good agreement with experiment [38, 41].

3. Native point defects in ZnO

We now illustrate the power of this methodology by discussing some specific results for ZnO. As mentioned in section 1, the conventional wisdom for ZnO (as well as many other oxides) is that native point defects play a dominant role in the unintentional conductivity of the material. We therefore executed a comprehensive set of investigations of point defects in the material in all possible charge states. Previous calculations were reported in [34, 42, 43] and [44]. In our recent work [35, 36] an accurate description of semicore d states was obtained by using the LDA + U approach, and the band-gap problem was addressed by extrapolating the LDA and LDA + U results to the experimental gap [35], as described in sections 2.4.1 and 2.4.2. We found that zinc and oxygen antisites, as well as oxygen interstitials, were always high in energy. We therefore focus our discussions here on those defects (vacancies and zinc interstitials) that have low enough

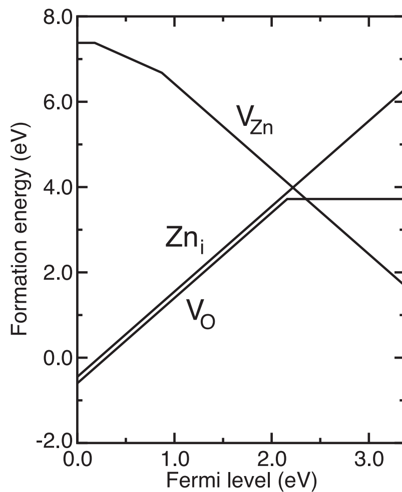

Figure 1. Calculated formation energies of most relevant native point defects in ZnO as a function of Fermi level. Zinc-rich conditions are assumed. For each Fermi-level position, only the lowest-energy charge state is shown; discontinuities in the slope of the lines therefore indicate a change in charge state, occurring at a transition level.

formation energies to form in significant concentrations. Our results for these formation energies as a function of Fermi level are shown in figure 1.

3.1. Oxygen vacancy

Figure 1 shows that oxygen vacancies have high formation energies under n-type conditions (Fermi level near the CBM). This translates into low equilibrium concentrations (much lower than the observed electron concentrations in unintentionally doped ZnO), making it unlikely that oxygen vacancies can act as sources of conductivity. Note that figure 1 assumes Zn-rich (i.e., oxygen-poor) conditions, which are most favorable for formation of oxygen vacancies; moving away from this limit would render oxygen vacancies even more unfavorable.

Figure 1 also shows that the transition level for V_O occurs at about 1 eV below the CBM. In fact, this transition is between the 2+ and the neutral charge state. The singly positive charge state is always higher in energy than either 2+ or 0. The position of the transition level indicates that oxygen vacancies are certainly *not* shallow donors; i.e., even if they were present in the material, they would not act as sources of conductivity.

Several aspects of these results can be compared with experiment. First of all, the high formation energy and resulting low concentration of oxygen vacancies in as-grown material is consistent with electron paramagnetic resonance (EPR) studies [45–47], in which a line at $g \approx 1.99$ characteristic of V_O has been observed. Indeed, this EPR signal was observed only after electron irradiation and not in the as-grown material. Second, the absence of a stable paramagnetic charge state (the +1 state) is consistent with the fact that observation of the $g \approx 1.99$ signal requires optical excitation. Third, luminescence and absorption lines associated with the defect can be calculated by mapping out a *configuration coordinate diagram* [35] and are in very good agreement with optically detected EPR experiments by Vlasenko and Watkins [47]. Cumulatively, these analyses and checks leave no doubt that oxygen vacancies are not related to unintentional n-type conductivity in ZnO.

It should be noted that the absence of oxygen vacancies in n-type ZnO does not mean they cannot play a role under p-type conditions. Acceptor doping of ZnO is a very active research area, and obtaining reproducible p-type ZnO has proven extremely challenging. One contributing factor may certainly be the tendency for self-compensation, as indicated by the low formation energy of V_O when the Fermi level is close to the VBM. A possible strategy for combatting this problem would be to move towards more oxygen-rich conditions; however, this would also decrease the likelihood of incorporation of acceptor impurities on the oxygen site, and nitrogen substituting on an oxygen site is currently considered to be the best candidate.

3.2. Zinc interstitials

Figure 1 shows that the formation energy of zinc interstitials is at least as high as that of oxygen vacancies, rendering their incorporation under equilibrium conditions equally unlikely. However, zinc interstitials do act as shallow donors (i.e., the 2+/0 transition level is near or above the CBM). A closer investigation of their possible incorporation under nonequilibrium conditions is therefore in order, and we have addressed this by studying the diffusion of zinc interstitials. In spite of the fact that zinc is a fairly large atom sitting in a confined interstitial space in ZnO, the migration barrier turns out to be remarkably low: only 0.57 eV [36, 48]. The migration process is actually quite intricate, involving the exchange of the interstitial with a substitutional Zn atom. The low value of the migration barrier indicates that diffusion of Zn interstitials is an extremely fast process that can take place even below room temperature. This implies that it would not be possible to maintain a nonequilibrium concentration of interstitials; rapid diffusion would ensure equilibration and a decline of the concentration to its (low) equilibrium level.

3.3. Zinc vacancies

Under n-type conditions (which apply to the majority of ZnO samples), zinc vacancies are the lowest-energy point defects. As pointed out in [42], V_{Zn} has a $-/2-$ transition level at about 0.9 eV above the VBM, creating the possibility of luminescence by recombination of electrons in the conduction band with holes in this defect level. The resulting emission would be around 500 nm, providing an explanation for the green luminescence that is frequently observed in ZnO samples. This explanation is consistent with a number of experimental observations [49–52]. Oxygen vacancies have also been suggested as the source of green luminescence [53]. However, this proposal was based on an incorrect assignment of the $g \approx 1.96$ line in EPR measurements; as discussed in section 3.1, oxygen vacancies are characterized by the $g \approx 1.99$, and the line at $g \approx 1.96$ is merely due to shallow donors or conduction-band electrons.

3.4. Alternative causes of n-type doping

Our investigations of native point defects lead us to conclude that they cannot be responsible for n-type conductivity. The origin of this conductivity must therefore lie in the unintentional presence of donor impurities. A number of candidates exist: Al, Ga, or In atoms substituting on a Zn site; or F or Cl on an oxygen site. However, these elements are unlikely to be systematically present in all growth environments; and many of them have actually been eliminated based on careful chemical analysis. One element, however, is virtually impossible to exclude from the growth or processing environment, and is also notoriously difficult to detect with analytical techniques such as Secondary Ion Mass Spectrometry (SIMS). That element is hydrogen.

It is, in fact, surprising that hydrogen would act as a shallow donor in ZnO. In most other semiconductors it has been established that hydrogen acts as an amphoteric impurity, i.e., as a donor in p-type material and as an acceptor in n-type [54]. In ZnO, however, interstitial hydrogen behaves exclusively as a donor and hence as a source of n-type conductivity [55]. The calculated formation energy (1.6 eV) of interstitial hydrogen in n-type ZnO is quite modest and fully consistent with the measured value of solubility for hydrogen diffused into ZnO [56], as discussed in [3]. Even more surprisingly, we have recently demonstrated based on first-principles calculations that *substitutional* hydrogen on an oxygen site (H_O) also behaves as a shallow donor [57]. Because hydrogen replaces oxygen, the partial pressure dependence of this defect is analogous to that of the oxygen vacancy, thus explaining a great deal of circumstantial evidence that in the past had been interpreted as supporting the presence of oxygen vacancies.

The bonding configuration assumed by H_O is highly unusual, since it involves hydrogen simultaneously bonding to four surrounding Zn atoms. This can be described as a 'multicenter bond' and interpreted using standard molecular orbital theory [57]. This result shows that in the process of applying first-principles theory to relevant technological issues, exciting and unexpected physics and chemistry can emerge.

4. Summary and future directions

We have described how state-of-the-art first-principles calculations are significantly contributing to solving the long-standing problem of unintentional conductivity of wide-band-gap semiconductors and oxides. Density functional theory is the cornerstone on which modern solid-state calculations are built. In section 2 we noted a number of inherent deficiencies of DFT and described various approaches to overcoming them. Because of computational demands, large-scale calculations are currently constrained to using approximate techniques such as LDA + U; even so, we demonstrated that calculations for ZnO have been able to resolve a number of long-standing puzzles. Looking to the future, some of the other approaches described in section 2 (such as exact exchange, hybrid functionals, and self-consistent GW) may become tractable and push first-principles approaches to new frontiers.

Acknowledgments

Collaborations and discussions with A Janotti, A Singh, K Delaney, P Rinke, and M Scheffler are gratefully acknowledged. The work was supported in part by the NSF MRSEC Program under award No. DMR05-20415 and by the UCSB Solid State Lighting and Energy Center. It also made use of the CNSI Computing Facility under NSF grant No. CHE-0321368.

References

[1] Reynolds D C, Look D C and Jogai B 1996 *Solid State Commun.* **99** 873
[2] Look D C 2001 *Mater. Sci. Eng.* B **80** 383
[3] Van de Walle C G 2002 *Phys. Status Solidi* b **229** 221
[4] Van de Walle C G, Laks D B, Neumark G F and Pantelides S T 1993 *Phys. Rev.* B **47** 9425
[5] Van de Walle C G and Neugebauer J 2004 *J. Appl. Phys.* **95** 3851
[6] Dean J A 1992 *Lange's Handbook of Chemistry* 14th edn (New York: McGraw-Hill)
[7] Limpijumnong S, Van de Walle C G and Northrup J E 2001 *Phys. Rev. Lett.* **87** 205505
[8] Neugebauer J and Van de Walle C G 1999 *J. Appl. Phys.* **85** 3003
[9] Hohenberg P and Kohn W 1964 *Phys. Rev.* **136** B864
[10] Kohn W and Sham L J 1965 *Phys. Rev.* **140** A1133
[11] Perdew J P and Wang Y 1991 *Phys. Rev. Lett.* **66** 508
[12] Perdew J P, Burke K and Ernzerhof M 1996 *Phys. Rev. Lett.* **77** 3865
[13] Vanderbilt D 1990 *Phys. Rev.* B **41** 7892
[14] Kresse G and Hafner J 1994 *J. Phys.: Condens. Matter* **6** 8245
[15] Blöchl P E 1994 *Phys. Rev.* B **50** 17953
[16] Kresse G and Joubert D 1999 *Phys. Rev.* B **59** 1758
[17] Kresse G and Hafner J 1993 *Phys. Rev.* B **47** 558
[18] Kresse G and Furthmüller J 1996 *Phys. Rev.* B **54** 11169
[19] Kresse G and Furthmüller J 1996 *Comput. Mater. Sci.* **6** 15
[20] Wei S H and Zunger A 1988 *Phys. Rev.* B **37** 8958
[21] Janotti A, Segev D and Van de Walle C G 2006 *Phys. Rev.* B **74** 045202
[22] Janotti A and Van de Walle C G 2007 *Phys. Rev.* B **75** 121201
[23] Städele M, Moukara M, Majewski J A, Vogl P and Görling A 1999 *Phys. Rev.* B **59** 10031
[24] Rinke P, Qteish A, Neugebauer J, Freysoldt C and Scheffler M 2005 *New J. Phys.* **7** 126
[25] Anisimov V I, Zaanen J and Andersen O K 1991 *Phys. Rev.* B **44** 943
[26] Anisimov V I, Aryasetiawan F and Liechtenstein A I 1997 *J. Phys.: Condens. Matter* **9** 767
[27] Stampfl C, Van de Walle C G, Vogel D, Krüger P and Pollmann J 2000 *Phys. Rev.* B **61** R7846
[28] Godby R W, Schlüter M and Sham L J 1988 *Phys. Rev.* B **37** 10159
[29] Hybertsen M S and Louie S G 1986 *Phys. Rev.* B **34** 5390
[30] Sánchez-Friera P and Godby R W 2000 *Phys. Rev. Lett.* **85** 5611
[31] Muscat J, Wander A and Harrison N M 2001 *Chem. Phys. Lett.* **342** 397
[32] Heyd J, Peralta J E, Scuseria G E and Martin R L 2005 *J. Chem. Phys.* **123** 174101
[33] Franchini C, Podloucky R, Paier J, Marsman M and Kresse G 2007 *Phys. Rev.* B **75** 195128
[34] Zhang S B, Wei S H and Zunger A 2001 *Phys. Rev.* B **63** 075205
[35] Janotti A and Van de Walle C G 2005 *Appl. Phys. Lett.* **87** 122102

[36] Janotti A and Van de Walle C G 2007 *Phys. Rev.* B **76** 165202
[37] Wang L W 2001 *Appl. Phys. Lett.* **78** 1565
[38] Segev D and Van de Walle C G 2006 *Europhys. Lett.* **76** 305
[39] Segev D, Janotti A and Van de Walle C G 2007 *Phys. Rev.* B **75** 035201
[40] Christensen N E 1984 *Phys. Rev.* B **30** 5753
[41] Van de Walle C G and Segev D 2007 *J. Appl. Phys.* **101** 081704
[42] Kohan A F, Ceder G, Morgan D and Van de Walle C G 2000 *Phys. Rev.* B **61** 15019
[43] Oba F, Nishitani S R, Isotani S, Adachi H and Tanaka I 2001 *J. Appl. Phys.* **90** 824
[44] Lee E C, Kim Y S, Jin Y G and Chang K J 2001 *Phys. Rev.* B **64** 085120
[45] Smith J M and Vehse W E 1970 *Phys. Lett.* A **31** 147
[46] Gonzalez C, Galland D and Herve A 1975 *Phys. Status Solidi* b **72** 309
[47] Vlasenko L S and Watkins G D 2005 *Phys. Rev.* B **71** 125210
[48] Janotti A and Van de Walle C G 2006 *J. Cryst. Growth* **287** 58
[49] Reynolds D C, Look D C, Jogai B and Morkoç H 1997 *Solid State Commun.* **101** 643
[50] Reynolds D C, Look D C, Jogai B, Van Nostrand J E, Jones R and Jenny J 1998 *Solid State Commun.* **106** 701
[51] Sekiguchi T, Ohashi N and Terada Y 1997 *Japan. J. Appl. Phys.* **36** L289
[52] Lavrov E V, Weber J, Börrnert F, Van de Walle C G and Helbig R 2002 *Phys. Rev.* B **66** 165205
[53] Vanheusden K, Seager C H, Warren W L, Tallant D R, Caruso J, Hampden-Smith M J and Kodas T T 1997 *J. Lumin.* **75** 11
[54] Van de Walle C G and Neugebauer J 2003 *Nature* **423** 626
[55] Van de Walle C G 2000 *Phys. Rev. Lett.* **85** 1012
[56] Thomas D G and Lander J J 1956 *J. Chem. Phys.* **25** 1136
[57] Janotti A and Van de Walle C G 2007 *Nat. Mater.* **6** 44