New Perspective on Formation Energies and Energy Levels of Point Defects in Nonmetals

R. Ramprasad,[1] H. Zhu,[1] Patrick Rinke,[2] and Matthias Scheffler[2]

[1]*Chemical, Materials and Biomolecular Engineering and Institute of Materials Science, University of Connecticut,*
97 North Eagleville Road, Storrs, Connecticut 06269, USA
[2]*Fritz-Haber-Institut der Max-Planck-Gesellschaft, Faradayweg 4-6, D-14195 Berlin, Germany*

We propose a powerful scheme to accurately determine the formation energy and thermodynamic charge transition levels of point defects in nonmetals. Previously unknown correlations between defect properties and the valence-band width of the defect-free host material are identified allowing for a determination of the former via an accurate knowledge of the latter. These correlations are identified through a series of hybrid density-functional theory computations and an unbiased exploration of the parameter space that defines the Hyde-Scuseria-Ernzerhof family of hybrid functionals. The applicability of this paradigm is demonstrated for point defects in Si, Ge, ZnO, and ZrO_2.

Point defects critically affect many properties of nonmetals, ranging from conductivity to optical activity to phase stability. Although first-principles computations have contributed to our understanding of point defects, confusion and faith permeate such studies. Conventional implementations of density-functional theory (DFT) utilize (semi)local exchange-correlation functionals. Because of spurious electron self-interactions, these approximations lead to inaccurate predictions of band gaps and defect levels and, consequently, to uncertainties in the computed defect formation energies and charge transition levels [1,2]. More recent hybrid DFT studies that incorporate a fraction of exact exchange involve "optimizing" the hybrid-functional parameters such that the band gap of the host material best matches experiments [3–6]. Such (semi)empiricism does not satisfactorily resolve the uncertainties, nor does it guarantee a concurrent improvement in the accuracy of the predictions [4–6].

In principle, these issues may be settled once and for all by many-body methods such as self-consistent GW or quantum Monte Carlo (QMC) calculations, except that they are not practical today for the large supercells required in point-defect studies. Even non-self-consistent G_0W_0 calculations presently do not provide total energies; hence, determination of defect formation energies within such schemes involves *ad hoc* assumptions and (semi)local DFT starting points [7–9]. Several severe assumptions also underlie QMC methods [10]. These include the usage of local-density approximation (LDA) or Hartree-Fock pseudopotentials, small supercells, and the fixed node assumption [11]. Given this scenario, it is apparent that computations of point defect properties with high fidelity are still not at hand.

Here, we adopt a different perspective, viz., we seek to discover correlations between point defect properties and electronic structure features of the defect-free parent material. In order to intentionally vary the underlying electronic structure and defect properties, we exploit the availability of the (α, ω) parameter space of the Hyde-Scuseria-Ernzerhof (HSE) "family" of hybrid functionals. The HSE hybrid exchange-correlation functionals are built from a semilocal Perdew-Burke-Ernzerhof (PBE) functional; α represents the fraction of the semilocal PBE exchange interaction replaced by a screened nonlocal functional, and ω is the inverse screening length [12]. We find that modulations of the neutral point defect formation energies (due to variations of α and ω) are strongly correlated with concommittant variations in the band gap and valence-band width (VBW) of the defect-free host material. Most notably (cf. Fig. 1), it is the apparently linear correlation of the formation energies with the VBWs that provides the best reconciliation with experimental data (for vacancies in Si and Ge) and diffusion Monte Carlo (DMC) studies (for the self-interstitial in Si). This behavior, i.e., the importance of the VBW, immediately calls into question the practice of optimizing hybrid-functional parameters to match experimental band gap values alone [3,4] and underlines the importance of other general features of the electronic structure, as has been recently alluded to [2].

A second important upshot of this study relates to the thermodynamic defect charge transition levels, $\tilde{\mu}(q/q')$, and band edge positions. $\tilde{\mu}(q/q')$ is defined as the electronic chemical potential μ at which point defects in two different charge states q and q' are at thermodynamic equilibrium with each other. When referenced to the average local electrostatic potential (V_{av}), we find that $\tilde{\mu}(q/q')$ becomes very insensitive to α and ω, consistent with prior observations [4]. This implies that the *difference* in the formation energies of defects in different charge states varies little with treatment, provided μ is referred to V_{av} (Fig. 2). On the other hand, the band gap and band edge positions—the valence-band minimum (VBm), valence-band maximum (VBM), and conduction-band minimum (CBm)—scale nearly linearly with VBW (Fig. 3). When
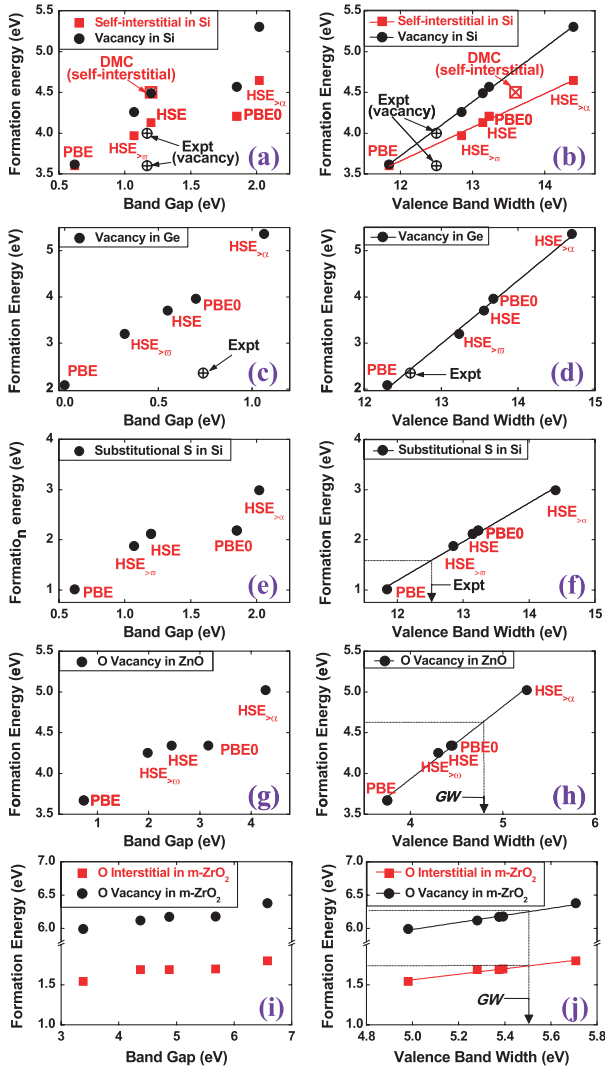
FIG. 1 (color online). Correlations between the formation energies and the band gap (left) or valence-band width (right) for point defects in bulk Si (a),(b),(e),(f), bulk Ge (c),(d), bulk wurtzite ZnO (g),(h), and bulk monoclinic ZrO_2 (i),(j). The PBE semilocal functional and four hybrid functionals within the HSE "family" (as defined in the text) were considered. Experimental (Expt), GW, and diffusion Monte Carlo (DMC) simulation results are also shown.

referenced to the VBM estimated for trustworthy VBW values, $\tilde{\mu}(q/q')$ for point defects in Si, Ge, and ZnO agrees favorably with experiments. Thus, the VBW of the defect-free parent material may be viewed as a convenient *descriptor* to assess the quality of the treatment.

For definiteness, we consider five specific functionals in the (α, ω) parameter space: PBE, PBE0, HSE, $HSE_{>\alpha}$, and $HSE_{>\omega}$. PBE0 refers to the $(0.25, 0)$ hybrid-functional, while HSE refers to the $(0.25, 0.207 \, \text{Å}^{-1})$ HSE06 functional [12]. The last two functionals correspond to those with larger values of α and ω, namely, $(0.5, 0.2 \, \text{Å}^{-1})$ and $(0.25, 0.4 \, \text{Å}^{-1})$, respectively.
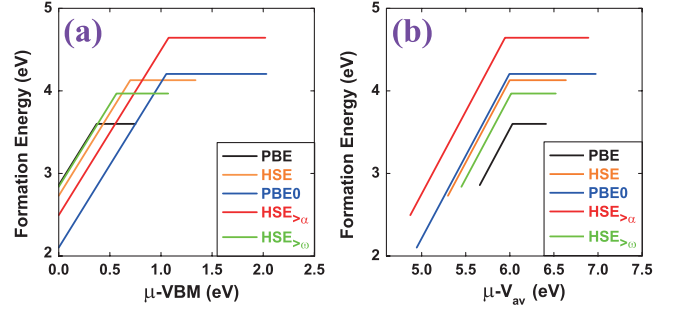

FIG. 2 (color online). The formation energy for the neutral and +2 charged Si self-interstitial as a function of the electronic chemical potential (μ). In (a), μ is referenced to the valence-band maximum (VBM), and in (b), it is referenced to the average electrostatic potential (V_{av}). For each of the five functionals considered, the curves extend from the respective VBM to the conduction-band minimum (CBm).

Specifically, we focus on point defects in Si, Ge, ZnO, and ZrO_2 (Si: $0/+2$ charged vacancy, $0/+2$ charged self-interstitial, $0/+1$ charged S substitutional; Ge: $0/-1/-2$ charged vacancy; ZnO: $0/+2$ charged O vacancy; ZrO_2: $0/-2$ charged O interstitial, $0/+2$ charged O vacancy). These material choices are intended to include well-studied and technologically important systems spanning small to large band gaps. The Si and Ge systems involved 216-atom supercells, while the ZnO and ZrO_2 systems utilized 192-atom wurtzite and 96-atom monoclinic supercells. All calculations were performed using the VASP code within the projector augmented wave methodology [13]. A Γ-centered $2 \times 2 \times 2$ k mesh and a plane wave energy cutoff of 280, 174, 400, and 400 eV for Si, Ge, ZnO, and ZrO_2, respectively, were used. Geometry optimizations were performed at the PBE level of theory, and the PBE optimized geometry was used in calculations involving all the hybrid functionals. This allowed us to directly probe the effect of the treatment of the exchange interaction. As a side remark, we note that all calculations
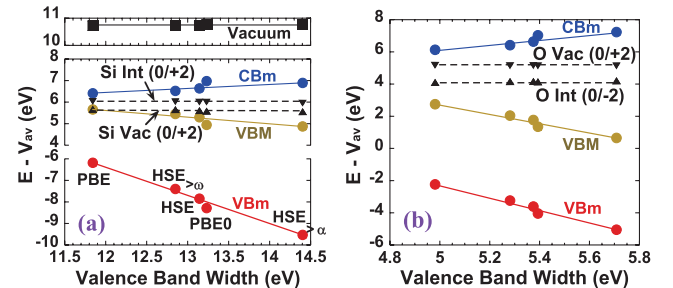

FIG. 3 (color online). Correlations between the VBW and the point defect charge transition levels [$\tilde{\mu}(q/q')$], valence-band minimum (VBm), valence-band maximum (VBM), and conduction-band minimum (CBm) for Si (a) and ZrO_2 (b). All energies are referred to V_{av} corresponding to each treatment. In (a), the vacuum level is also shown.

were performed self-consistently. However, *this was un-necessary* for calculations involving hybrid functionals (at least for the point defects considered here). The formation energy and $\tilde{\mu}(q/q')$ obtained from non-self-consistent calculations using PBE wave functions for all defects considered differed by less than 0.03 eV from the self-consistent results.

The formation energy $E_f^q(\mu)$ of a point defect in charge state q is given by

$$E_f^q(\mu) = E_{\text{def}}^q - E_{\text{bulk}} \pm \eta + q(\mu + E_{\text{ref}} + \Delta V) + E_{\text{corr}}^q,$$
(1)

where E_{def}^q and E_{bulk} are the calculated total energies of the supercells containing the point defect and the perfect bulk host materials, respectively. η, the chemical potential of the atomic species constituting the defect, is taken to be the bulk Si (Ge) energy per atom in the case of Si (Ge) defects, half the energy per O_2 molecule in the case of O defects, and the chemical potential of S corresponding to the equilibrium between Si and SiS_2 in the case of S substitutional defect. The $\pm$ sign indicates whether the defect atom was removed or added. E_{ref} corresponds to a suitable reference energy, generally taken to be the VBM (other more appropriate choices for E_{ref} are discussed below). ΔV is a correction to appropriately line up the energy zero of the supercells with and without the defect. E_{corr}^q represents spurious electrostatic interactions of charged defects due to periodicity and finite supercell sizes. Here, we have included only the first-order monopole correction [14]. Since this correction is the same for all functionals, it does not affect the comparison between the functionals.

We first consider neutral point defects, for which $E_f^{q=0}$ is independent of μ. E_f^0 varies with the type of functional used (Fig. 1), with the variation most pronounced for systems with the smallest band gaps. For instance, E_f^0 for the Si self-interstitial, Si vacancy, and S substitute in bulk Si varies by as much as 1.1, 1.8, and 2 eV, respectively, and for the Ge vacancy in bulk Ge, the variation is more than 3 eV. For O vacancy in ZnO, this variation is 1.4 eV. In contrast, the corresponding variation is 0.4 and 0.3 eV,

respectively, for the O vacancy and O interstitial in ZrO_2—the largest band gap system studied here. In all cases considered, E_f^0 increases in the order: PBE, $HSE_{>\omega}$, HSE, PBE0, and $HSE_{>\alpha}$, with the HSE and PBE0 values being within 0.3, 0.1, 0.01, and 0.01 eV of each other in the case of point defects in Ge, Si, ZnO, and ZrO_2, respectively. Table I collects the PBE and HSE results for all systems studied.

In order to comprehend the variation of E_f^0 with α and ω, we attempted to draw correlations between such variation and features of the electronic structure of the respective defect-free host material. Among all features considered, E_f^0 displays the most meaningful variation with the band gap and VBW (Fig. 1), with the practically linear correlation of E_f^0 with VBW most striking. This correlation may be understood as follows. The valence band is composed of the Si (Ge) sp^3 hybridized states in the case of bulk Si (Ge), and is dominated by O $2p$ states (with some admixture of Zn $3d$ or Zr $4d$ states) in the case of bulk ZnO and ZrO_2. The width of these valence bands is a measure of the cohesive interaction between the respective atoms within the parent material. The greater this interaction, the larger the band width. Consequently, the resistance to form defects will be greater (and hence the formation energy higher). Correlations between E_f^0 and the band gap may also be understood qualitatively by similar arguments. However, for the simple bonding or antibonding picture to be valid (and for the band gap to become a good descriptor), the band gap may have to be appropriately defined, especially for indirect band gap systems.

The identified correlation between E_f^0 and the host VBW (rather than the host band gap) provides a favorable reconciliation with available experimental results. This is immediately evident from Figs. 1(a)–1(d), which include the experimental results for E_f^0 of a vacancy in bulk Si [18,19] and bulk Ge [20], as well as the band gap and VBW of Si and Ge [15]. Table I contains our predictions of E_f^0 for all defects considered based on well-documented VBWs of the corresponding hosts. The experimental Si and Ge VBW values of 12.5 ± 0.6 eV and 12.6 eV lead to Si and Ge

TABLE I. The formation energies (in eV) of point defects in Si, Ge, ZnO, and ZrO_2. The predicted values (Pred) are determined using the correlation between the formation energies and the VBWs identified in Fig. 1 and the experimental (Si and Ge [15]) or the GW (ZnO and ZrO_2 [16,17]) VBWs (experimental data for the oxides include significant band tails, most likely due to imperfections). The experimental results for the Si and Ge vacancy formation energies are from Ref. [18–20], respectively.

System	Defect	PBE	HSE	Pred	Expt
Si	Si vacancy	3.61	4.49	4.05 ± 0.4	4.0, 3.6 ± 0.2
Si	Si interstitial	3.60	4.13	3.86 ± 0.3	$\cdots$
Si	S substitutional	1.01	2.11	1.6 ± 0.4	
Ge	Ge vacancy	2.09	3.70	2.43	2.35 ± 0.11
ZnO	O vacancy	3.67	4.32	4.62	
ZrO_2	O vacancy	5.99	6.17	6.25	$\cdots$
ZrO_2	O interstitial	1.54	1.69	1.73	$\cdots$

vacancy E_f^0 values in excellent agreement with available experimental results for these defects (Table I). Such a favorable agreement with experiment was not obtained if one attempted extrapolations based on the band gap as the descriptor. This implies that a treatment that provides a better description of electronic structure features such as the VBW may also provide a better estimate of the defect formation energies, as alluded to recently [2].

The DMC result for E_f^0 of the Si self-interstitial [10] (determined for a 16-atom supercell) is higher than the corresponding PBE value by about 1 eV—a discrepancy that has remained a puzzle for some time [11]. We argue that this large E_f^0 may be traced to the rather large VBW values underlying the DMC treatment. Indeed, fixed node DMC calculations to determine the band structure of Si yield a VBW of 13.6 ± 0.2 eV [21], compared to the PBE VBW of 11.8 eV. The correlation between the large DMC VBW value for Si and the large self-interstitial formation energy is manifest in Fig. 1(b).

$G_0 W_0$ calculations place the neutral Si self-interstitial E_f^0 0.4 eV below the DMC result [7] (i.e., ~0.3 eV larger than our predicted value). The $G_0 W_0$ correction scheme for computing E_f^0 relies on two critical ingredients: (1) the LDA E_f^{+2} value (with the assumption that the +2 charge state, owing to the empty gap levels, is least prone to self-interaction errors), and (2) the $G_0 W_0$ energies for the successive addition of two electrons to the +2 charged defect (with the assumption that the $G_0 W_0$ and LDA VBMs are aligned). As will be pointed out below, both assumptions may be questioned (the latter may be eliminated by suitable band edge realignments).

We next move on to charged point defects. E_f^q determinations for $q \neq 0$ require the specification of E_{ref}, the conventional choice being the VBM, as this allows for a comparison with experiments. However, this choice of E_{ref} leads to an issue when comparing E_f^q across treatments. The reason is that one has made the implicit questionable assumption that the VBM positions provided by each functional are all aligned. Two other natural choices for E_{ref} are V_{av} and the vacuum level. For the case of the Si self-interstitial, Fig. 2 shows the formation energies of the neutral and +2 charged defect as a function of μ defined with respect to VBM (a) and V_{av} (b). Both choices of E_{ref} lead to large variations in the E_f^{+2} values, indicating that the +2 charge state of the Si self-interstitial is not immune to self-interaction errors (see above discussion on $G_0 W_0$ calculations). A similar warning was also made in the past [2]. Also, the choice of $E_{ref} = V_{av}$ leads to two immediate observations: (1) The $\tilde{\mu}(0/+2)$ values [the "knees" in Fig. 2(b)] are aligned for all functionals considered, and (2) for any given value of μ, $(E_f^0 - E_f^{+2})$ is a constant across treatments.

The assumption of the alignment of V_{av} across functionals has thus revealed important invariant features,

making this choice of E_{ref} not only attractive but also necessary especially when comparing μ-dependent quantities across treatments. With this assumption, we explore in Fig. 3 the evolution of $\tilde{\mu}$ for point defects in Si and ZrO_2 and the band edges of the respective defect-free parent materials with respect to treatment (or VBW, our descriptor). Results for Ge and ZnO (not shown) follow the same trend. The valence (conduction) band edges move downward (upward) linearly with VBW, consistent with the varying levels of self-interaction error correction.

However, $\tilde{\mu}$ (when referenced to V_{av}) is relatively invariant to treatment (or VBW) in all cases, and the degree of variation with treatment decreases with increasing band gap of the host. In the worst case (Si vacancy), the rms variation of $\tilde{\mu}$ across treatments is ~0.08 eV. These trends, at least in the case of the localized defects considered here, are consistent with recent PBE, HSE, and PBE0 studies of a variety of point defects and insulators [4–6]. Properties related to total energy differences—such as $\tilde{\mu}$—for localized (or "atomiclike") defects appear to be well represented by (semi)local functionals just like the energy difference between highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO) or ionization potentials of atoms or molecules computed using the ΔSCF method [4–6] (a method similar to standard DFT but with one or more electrons placed in higher lying Kohn-Sham orbitals).

One question still remains, related to the choice of E_{ref}: How do our results stack up if we choose to set E_{ref} to be the vacuum level instead of V_{av} (a natural choice for the definition of ionization potentials of atoms or molecules). This point was addressed by a series of 13-layer (001) Si slab calculations. The energy difference between V_{av} in the bulk and the vacuum parts of the slab was determined for the five functionals, and corrected for the (slight) differences in the surface dipole moment in the different treatments. The vacuum level thus obtained is portrayed in Fig. 3(a). As can be seen, the vacuum levels are aligned to an acceptable level of accuracy. One may thus use either V_{av} or the vacuum as proper E_{ref} choices, with the former being a practical option [4,22].

We now attempt to make a connection between our predictions for $\tilde{\mu}$ and experiments. Once again using well-documented values for the VBW of the respective parent materials, we estimate the most likely $\tilde{\mu}$ values (defined with respect to the VBM). These predictions, along with the corresponding PBE, HSE, and available reliable experimental values for point defects in Si, Ge, and ZnO [23–25] are listed in Table II. As can be seen, the predictions are within acceptable levels of accuracy, accounting for the uncertainties in the experimental VBW value for Si, the slight fluctuations in the $\tilde{\mu}$ value across treatments, and the omission of finite size effects such as the quadrupole contributions in charged defect calculations. Such predictions, if made based on the band gap as the descriptor, were in marked

TABLE II. $[\tilde{\mu}(q/q') - \text{VBM}]$ (in eV) for point defects in Si, Ge, ZnO, and ZrO$_2$. The predicted values (Pred) are determined using the experimental (Si and Ge [15]) or the GW (ZnO and ZrO$_2$ [16,17]) VBWs. The experimental $[\tilde{\mu}(q/q') - \text{VBM}]$ results are from Refs. [23–25].

System	Defect	(q/q')	PBE	HSE	Pred	Expt
Si	Si vacancy	$(0/+2)$	0.01	0.31	0.14	0.09
Si	Si interstitial	$(0/+2)$	0.37	0.69	0.52	0.43
Si	S substitutional	$(0/+1)$	0.40	0.90	0.69	0.85
Ge	Ge vacancy	$(-1/-2)$	0.11	0.33	0.14	0.14
ZnO	O vacancy	$(0/+2)$	0.56	1.86	2.52	2.8
ZrO$_2$	O vacancy	$(0/+2)$	2.41	3.38	3.81	$\cdots$
ZrO$_2$	O interstitial	$(0/-2)$	1.36	2.40	2.86	$\cdots$

disagreement with experiments [e.g., $[\tilde{\mu}(-1/-2) - \text{VBM}]$ for the Ge vacancy was 0.5 eV].

We are thus persuaded to conclude that the properties of intrinsic point defects in nonmetals are strongly controlled by electronic structure features such as the VBW of the defect-free parent material—an important finding, in light of the emerging practice of optimizing hybrid-functional parameters to match experimental band gap values. The present work, in terms of the importance of the VBW, also brings a measure of redemption for conventional DFT calculations. Despite the severe underestimation of band gaps, reasonable descriptions of the Si and Ge valence-band manifolds by (semi)local functionals lead to defect property predictions comparable in quality to the corresponding predictions by hybrid functionals. Moreover, this analysis also indicates that assumptions underlying higher level DMC and GW treatments of defect formation energies need to be reevaluated.

A practical strategy for accurately determining E_f^q and $\tilde{\mu}(q/q')$ could be the following: (1) determination of VBW, E_f^q, VBM and $\tilde{\mu}(q/q')$ at the PBE and HSE levels of theory, (2) determination of VBW for the defect-free parent material using a higher level (e.g., GW) calculation, if reliable experimental measurements are not available, and (3) extrapolation or interpolation to the "correct" E_f^q and $\tilde{\mu}(q/q')$ (with respect to the VBM) values using the GW or experimental VBW value. While very different materials and point defect types were indeed considered in this study, the veracity of the emerging notions for an even larger variety of point defects and insulators needs to be ascertained in the future.

[1] S. Lany and A. Zunger, Phys. Rev. B **78**, 235104 (2008).
[2] P. Deak *et al.*, Phys. Status Solidi B **248**, 790 (2011).
[3] A. Janotti *et al.*, Phys. Rev. B **81**, 085212 (2010).
[4] A. Alkauskas *et al.*, Phys. Status Solidi B **248**, 775 (2011).
[5] H. P. Komsa, P. Broqvist, and A. Pasquarello, Phys. Rev. B **81**, 205118 (2010).
[6] A. Alkauskas, P. Broqvist, and A. Pasquarello, Phys. Rev. Lett. **101**, 046405 (2008).
[7] P. Rinke *et al.*, Phys. Rev. Lett. **102**, 026402 (2009).
[8] M. Hedström *et al.*, Phys. Rev. Lett. **97**, 226401 (2006).
[9] S. Lany and A. Zunger, Phys. Rev. B **81**, 113201 (2010).
[10] E. R. Batista *et al.*, Phys. Rev. B **74**, 121102(R) (2006); W.-K. Leung *et al.*, Phys. Rev. Lett. **83**, 2351 (1999).
[11] A. E. Mattsson, R. R. Wixom, and R. Armiento, Phys. Rev. B **77**, 155211 (2008).
[12] A. V. Krukau *et al.*, J. Chem. Phys. **125**, 224106 (2006).
[13] G. Kresse and J. Furthmuller, Phys. Rev. B **54**, 11 169 (1996); G. Kresse and D. Joubert, *ibid.* **59**, 1758 (1999).
[14] G. Makov and M. C. Payne, Phys. Rev. B **51**, 4014 (1995).
[15] *Numerical Data and Functional Relationships in Science and Technology*, edited by O. Madelung, Landolt-Bornstein, New Series, Group III, Vol. 23a (Springer-Verlag, Berlin, 1989).
[16] A. R. H. Preston *et al.*, Phys. Rev. B **78**, 155114, (2008).
[17] H. Jiang *et al.*, Phys. Rev. B **81**, 085119 (2010).
[18] N. Fukata *et al.*, Physica (Amsterdam) **308–310B**, 1125 (2001).
[19] S. Dannefaer, P. Mascher, and D. Kerr, Phys. Rev. Lett. **56**, 2195 (1986).
[20] J. Vanhellemont, P. Spiewak, and K. Sueoka, J. Appl. Phys. **101**, 036103 (2007).
[21] A. J. Williamson *et al.*, Phys. Rev. B **57**, 12 140 (1998).
[22] P. A. Schultz, Phys. Rev. Lett. **96**, 246401 (2006); A. Janotti and C. G. Van de Walle, Phys. Rev. B **76**, 165202 (2007).
[23] E. G. Seebauer and M. C. Kratzer, Mater. Sci. Eng. **55**, 57 (2006).
[24] *Numerical Data and Functional Relationships in Science and Technology*, edited by O. Madelung *et al.*, Landolt-Bornstein, New Series, Group III, Vol. 41a2a (Springer-Verlag, Berlin, 2002).
[25] D. M. Hofmann *et al.*, Appl. Phys. A **88**, 147 (2007).