Defect Energy Levels in Density Functional Calculations: Alignment and Band Gap Problem

Audrius Alkauskas, Peter Broqvist, and Alfredo Pasquarello

Ecole Polytechnique Fédérale de Lausanne (EPFL), Institute of Theoretical Physics, CH-1015 Lausanne, Switzerland
Institut Romand de Recherche Numérique en Physique des Matériaux (IRRMA), CH-1015 Lausanne, Switzerland

For materials of varying band gap, we compare energy levels of atomically localized defects calculated within a semilocal and a hybrid density-functional scheme. Since the latter scheme partially relieves the band gap problem, our study describes how calculated defect levels shift when the band gap approaches the experimental value. When suitably aligned, defect levels obtained from total-energy differences correspond closely, showing average shifts of at most 0.2 eV irrespective of band gap. Systematic deviations from ideal alignment increase with the extent of the defect wave function. A guideline for comparing calculated and experimental defect levels is provided.

Semilocal approximations to density functional theory, such as the local density approximation (LDA) and the generalized-gradient approximation, have proved extremely valuable to investigate energetic, atomistic, and magnetic properties of defects in solids [1]. However, these approximations have been much less successful in locating charge transition levels in the band gap because of the well known band gap problem from which they suffer. As a result, a direct comparison between calculated and experimental energy levels remains ambiguous. Furthermore, the determination of equilibrium densities of intrinsic defects and charge carriers is hindered [2]. Therefore, considerable efforts have been deployed in the study of defects to address the band gap problem going beyond semilocal approximations to density functional theory. Many-body perturbation theory in the *GW* approximation is the method of choice for calculating defect levels [3], but remains computationally demanding and therefore limited to small-size systems. Several practical routes have also been proposed, such as the scissor-operator scheme, the marker method [4], the LDA + *U* method [5], the use of adapted pseudopotentials [6], and the application of *ad hoc* extrapolation schemes [7]. However, the general applicability of these approaches is unclear. More recently, hybrid density functionals have become increasingly popular for addressing defect energy levels [8]. These functionals incorporate a fraction of Hartree-Fock exchange, leading to higher accuracy [9] and improved band gaps [10] compared to semilocal functionals.

In this Letter, we carry out a comparative study between defect energy levels calculated with semilocal and hybrid density functionals to determine their shifts as the description of the band gap improves. We aim at gaining insight into how calculated and measured defect levels should be compared when the adopted theoretical scheme is subject to the band gap problem. For this purpose, we considered materials covering a large range of band gaps and selected defect levels spanning large portions of their band gaps. Our study shows that charge transition levels obtained with

semilocal and hybrid density functionals correspond closely, provided a suitable alignment scheme is adopted. As the band gap decreases, systematic deviations from ideal alignment are found to increase with the extent of the defect wave function.

The semilocal density-functional calculations were performed within the generalized gradient approximation proposed by Perdew, Burke, and Ernzerhof (PBE) [11]. We used a hybrid density functional, denoted PBE0, which is obtained from the latter by replacing 25% of the PBE exchange energy by Hartree-Fock exchange [12]. We used a scheme based on plane-wave basis sets and norm-conserving pseudopotentials. The pseudopotentials were generated at the semilocal level and used in all calculations. The plane-wave basis set was defined by an energy cutoff of 70 Ry. The Brillouin zones of our supercells were sampled at the Γ point, but primitive cells with a converged *k*-point sampling were used for the determination of the bulk band edges. We took care of the integrable divergence of the Hartree-Fock exchange term [13]. Structural relaxations were carried out at the semilocal level [14]. Our calculations were performed with the codes QUANTUM-ESPRESSO [15] and CPMD [16].

We considered defects in four different materials of varying band gap: Si, SiC, HfO_2, and SiO_2. The band gaps of these materials calculated at the semilocal level severely underestimate the experimental values (Table I), as usual for semilocal density functional schemes. As shown in Table I, the hybrid scheme systematically gives larger band gaps, generally leading to a better agreement with experiment. For silicon, we adopted a cubic simula-

TABLE I.   Calculated and experimental band gaps (in eV).

	Si	SiC	HfO_2	SiO_2
Semilocal	0.6	2.2	4.3	5.8
Hybrid	1.8	3.9	6.7	8.3
Expt.	1.2	3.3	5.9	8.9

tion cell of 64 atoms and considered the following defects
with their relative charge states: the Si vacancy $(+2, +1, 0,$
$-1, -2)$, the Si self-interstitial $(+1, 0)$, the substitutional
O $(0, -1)$, the C interstitial $(+1, 0, -1)$, and the
P/Si-vacancy complex $(+1, 0, -1)$. For SiC, we modeled
the $4H$ polytype using an orthorhombic cell containing 96
atoms. Considered defects include the C vacancy $(+2, +1,$
$0, -1, -2)$, the Si vacancy $(+2, +1, 0, -1, -2)$, and the
complex consisting of C substitutional to Si next to a C
vacancy $(+2, +1, 0, -1, -2)$. For HfO_2, we took the
monoclinic structure and used a supercell containing 96
atoms. We considered the O vacancy $(+2, -1, 0, +1,$ and
$-2)$ and the O interstitial $(0, -1,$ and $-2)$. For SiO_2, we
modeled α-quartz with an orthorhombic cell containing 72
atoms and considered the interstitial H $(+1, 0, -1)$, the Si-
Si dimer bond $(+1, 0)$, the puckered O vacancy $(+1, 0)$,
the H bridge -Si-H-Si- $(+1, 0, -1)$, the substitutional N $(0,$
$-1)$, and the interstitial O_2 $(0, -1)$. All defect states
studied are atomically localized.

The formation energy of a defect in its charge state q can
be expressed in terms of the electron chemical potential μ
referred to the valence band maximum ε_v [1]:

$$E_f^q(\mu) = E_{tot}^q - E_{tot}^{bulk} - \sum_\alpha n_\alpha \eta_\alpha + q(\mu + \varepsilon_v), \quad (1)$$

where E_{tot}^q is the total energy of the defect system, E_{tot}^{bulk} the
total energy of the unperturbed system, n_α the number of
extra atoms of species α needed to create the defect, and
η_α the corresponding atomic chemical potential. Charge
transition levels correspond to specific values of the elec-
tron chemical potential for which two charge states have
equal formation energies. We considered both thermody-
namic and vertical charge transition levels.

To compare defect levels in semilocal and hybrid
density-functional schemes, it is necessary to use a com-
mon reference level ϕ external to the electronic system,
i.e., defined on the basis of the nuclear potentials: $\bar{\mu} =$
$\mu + \varepsilon_v - \phi$. This alignment scheme is illustrated in
Fig. 1. In our formulation, we used the same pseudopoten-
tials in the semilocal and hybrid calculations and trivially
achieved such an alignment by taking ϕ as the cell average
of the local potential originating from the ionic pseudopo-
tentials. However, we note that our formulation does not
imply any loss of generality and that a proper alignment
can also be enforced when different pseudopotentials are
used.

We calculated charge transition levels for the selected
set of defects in Si, SiC, HfO_2, and SiO_2 within both the
semilocal and hybrid schemes. Charge transition levels
obtained in either scheme were then aligned with respect
to the common reference level ϕ and reported in Fig. 2. For
each material, our results show that the defect levels calcu-
lated in the semilocal and hybrid schemes differ on average
by at most 0.2 eV when aligned in this way, despite the
significantly larger concomitant variations observed for the
band gaps. Since average shifts are similar in the four cases
studied, the identified alignment is more impressive for
large band gap materials where these shifts are small with
respect to the band gap. Indeed, the average relative shift is
only 2% for SiO_2, but increases to about 17% for Si. In
particular, these results indicate that differences between

FIG. 2 (color online).   Comparison between charge transition
levels calculated with the semilocal $(\bar{\mu}^{semiloc})$ and hybrid $(\bar{\mu}^{hyb})$
functionals for a variety of defects in Si, SiC, HfO_2, and SiO_2.
The energy levels corresponding to the valence band maximum
(VBM) and conduction band minimum (CBM) are also shown
(squares). All energies are referred to a common reference level
ϕ (see text), shifted to coincide with the VBM in the hybrid
scheme for convenience. For each material, Δ is the r.m.s. error
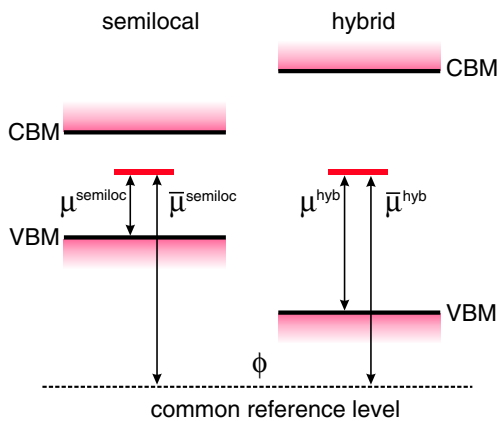with respect to the ideal alignment (dashed line).

FIG. 1 (color online).   Schematic illustration of the alignment
between energy levels obtained with a semilocal and a hybrid
density functional. The charge transition levels μ and $\bar{\mu}$ are
referred to the respective valence band maxima (VBM) and to a
common reference level, respectively. The conduction band
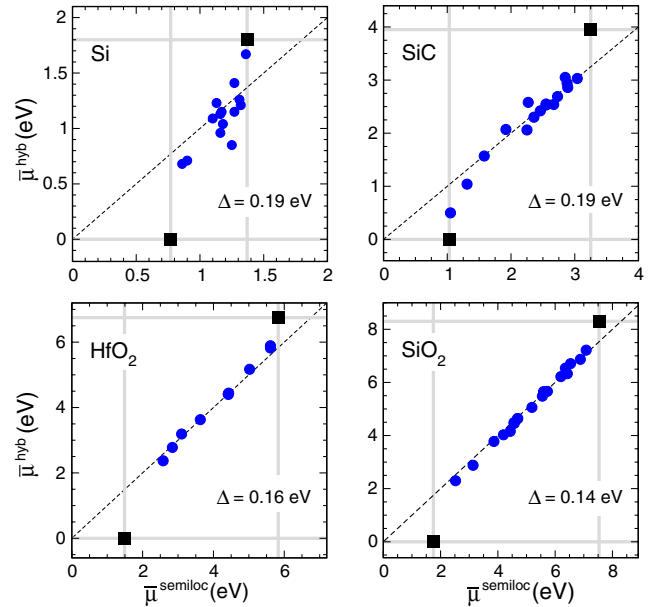minima (CBM) are also shown.

charge transition levels are already well described at the semilocal level, lending support to alignment schemes in which the defect levels are anchored to experimental marker levels given by well characterized defects [4,17]. In this respect, a key result of our work is that the defect levels are positioned on an energy scale unaffected by band gap renormalization.

To reveal systematic deviations with respect to the ideal alignment, we carried out linear regressions of the available data deriving optimal slopes [Fig. 3(a)]. Ideal alignment corresponds to unitary slope and is best illustrated for SiO_2 (slope of 1.08). When the band gap decreases, the optimal slope is found to increase indicating that defects in the upper part of the band gap tend to follow the conduction band, while defects in the lower part of the band gap tend to follow the valence band. This tendency is most pronounced for Si (slope of 1.4).

To provide a rationale for the obtained results, we first discard effects which affect the defect levels in a minor way. In all the cases studied, irrespective of band gap, the electron wave functions were found to be very similar in the semilocal and hybrid schemes. Their effect on defect levels can be quantified by calculating total energies at the hybrid level using electron wave functions optimized in the semilocal scheme without allowing for electron relaxation [18]. In this way, we inferred that the differences between the defect levels calculated in the two schemes do not arise from variations of the electron wave functions. Since the observed trends also hold for the subset of vertical transitions, we discard effects associated to differences in structural relaxation energies. The Slater transition-state approximation then allows us to focus on the single defect eigenstate rather than on the full manifold of occupied states [19]. This approximation expresses the relevant total-energy differences by the energy eigenvalue of the defect state at half filling, $\bar{\mu} \cong \langle \psi_D | \mathcal{H} | \psi_D \rangle - \phi$, and gives accurate charge transition levels in both schemes [18].

In the Slater approximation, the difference in charge transition levels can then be expressed as

$$\bar{\mu}^{\text{hyb}} - \bar{\mu}^{\text{semiloc}} \cong \langle \psi_D | \hat{V}_x^{\text{hyb}} - \hat{V}_x^{\text{semiloc}} | \psi_D \rangle, \quad (2)$$
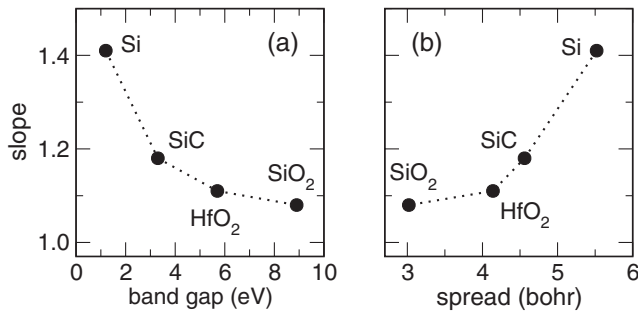

FIG. 3. Optimal slopes derived from linear regressions of the data in Fig. 2 as a function of (a) experimental band gap and (b) average spread of the defect wave functions.

where the exchange potential $\hat{V}_x$ is evaluated at half filling. Two different contributions can conceptually be distinguished in Eq. (2) referring to defect-defect and defect-bulk interactions. The separation is trivial for the Hartree-Fock exchange term, but requires some prescription for semilocal exchange. We first focus on defect-defect contributions. Since the defect wave function ψ_D is atomically localized, differences due to these interactions between charge transitions levels derived in the two schemes should be analogous to the corresponding differences for ionization potentials and electron affinities of atoms and molecules. The latter quantities can be expressed as total-energy differences and are already well described in semilocal approximations [20,21], as demonstrated by extensive quantum chemistry calculations [9]. Hence, this contribution is expected to give energy-level differences independent of the location of the defect level in the band gap. As for the defect-bulk contributions, it can be shown that they vanish in the limit of pointlike defect states. When the defect wave function has a finite extent, these contributions depend on the degree of valence-band vs conduction-band character of the defect state and can lead to a slope larger than 1 in Fig. 2. To support this picture, we calculated average spreads of the defect wave functions in each material. Figure 3(b) clearly shows that the slopes of the linear regressions increase with these spreads, as the band gap decreases.

Our results reveal a general trend which appears amenable to generalization. When the theoretical description is improved, band edges in these materials undergo significant shifts but charge transition levels of atomically localized defects remain practically unaffected. This leads us to propose the following guideline to locate charge transition levels in the experimental band gap. First, ordinary semilocal density-functional calculations are performed and charge transition levels of the targeted defect as well as band edges are determined. Second, the positions of the band edges are corrected through the use of a high-level electronic-structure theory which yields a band gap in agreement with experiment, e.g., through hybrid density-functional or GW calculations. This only requires a calculation for the bulk material, which is computationally less demanding than a defect calculation. Third, the defect level is located in the new band structure following the alignment proposed in this Letter. We demonstrate the applicability of this scheme for well-characterized defect levels: the two donor levels of substitutional Te in silicon (Te_{Si}) [22], the acceptor level of interstitial C in silicon (C_i) [23], and the optical transition between the valence band and the E_1' defect state in α-quartz [24,25]. The levels are first determined with respect to the valence band maximum within the semilocal scheme [26]. The position of the valence band is then adjusted according to recent GW calculations [27]. The resulting defect levels agree with the measured ones within the errors expected from our analysis (Table II).

TABLE II. Comparison between calculated (μ^{th}) and measured (μ^{expt}) defect levels given with respect to the valence band. Theoretical levels are obtained from semilocal levels (μ^{semiloc}) through application of GW corrections to the valence band edge $(\Delta\varepsilon_v^{GW})$ [27]. Experimental data for the defects Te_{Si}, C_i, and E_1' are from Refs. [22–24], respectively.

	Defect	q/q'	μ^{semiloc}	$\Delta\varepsilon_v^{GW}$	μ^{th}	μ^{expt}
Si	Te_{Si}	$0/+$	0.2	-0.4	0.6	0.6
Si	Te_{Si}	$+/++$	0.4	-0.4	0.8	1.0
Si	C_i	$0/-$	0.5	-0.4	0.9	1.0
SiO_2	E_1'	$+/0$	4.1	-1.9	6.0	6.0

This guideline applies to atomically localized defect states and is clearly inappropriate for effective-mass-like defect levels which are tied to band edges. Application of this procedure also requires that the defect is well described already within the semilocal density-functional scheme. For instance, the defect level should fall within the reduced band gap of the latter scheme to preserve its localized nature. An inaccurate description may also result from the occurrence of competition between defect states featuring different degrees of localization [28].

Our findings relate to other studies of defect levels and band gaps. Indeed, the band gap can also change as a result of a physical process, such as quantum confinement. For quantum dots of varying size, it has been shown that ionization potentials of deep defects remain constant as the band gap changes [29]. These potentials correspond to charge transition levels referred to the vacuum level. Another way to modify the band gap is achieved by changing the host material. It has been found that energy levels of transition-metal impurities within a set of isovalent semiconductors are aligned when referred to the vacuum level [30]. From the perspective of the present Letter, such an alignment is understood to the extent that the local chemistry of the defect is preserved and a common reference potential can be identified. Such transition-metal markers can then be used to predict band-offsets at interfaces [31].

In conclusion, calculated energy levels of atomically localized defects generally remain tied to a suitably defined reference level as the description of the band gap is improved. This leads to a guideline for comparing calculated and measured defect levels even when the adopted theoretical scheme is subject to the band gap problem.

[1] C. G. Van de Walle and J. Neugebauer, J. Appl. Phys. **95**, 3851 (2004).
[2] S. Lany and A. Zunger, Phys. Rev. Lett. **98**, 045501 (2007).
[3] S. Ismail-Beigi and S. G. Louie, Phys. Rev. Lett. **95**, 156401 (2005); M. Hedström et al., ibid. **97**, 226401 (2006).
[4] J. Coutinho et al., Phys. Rev. B **73**, 235213 (2006).
[5] A. Janotti and C. G. Van de Walle, Phys. Rev. B **76**, 165202 (2007).
[6] C. Stampfl et al., Phys. Rev. B **61**, R7846 (2000); J. Li and S.-H. Wei, Phys. Rev. B **73**, 041201(R) (2006).
[7] S. B. Zhang, S.-H. Wei, and A. Zunger, Phys. Rev. B **63**, 075205 (2001).
[8] J. M. Knaup et al., Phys. Rev. B **72**, 115323 (2005); K. Xiong et al., Appl. Phys. Lett. **87**, 183505 (2005); J. L. Gavartin et al., ibid. **89**, 082908 (2006); P. Broqvist and A. Pasquarello, ibid. **89**, 262904 (2006).
[9] L. A. Curtiss et al., J. Chem. Phys. **109**, 42 (1998).
[10] J. Muscat, A. Wander, and N. M. Harrison, Chem. Phys. Lett. **342**, 397 (2001); J. Paier et al., J. Chem. Phys. **124**, 154709 (2006); **125**, 249901 (2006).
[11] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).
[12] J. P. Perdew, M. Ernzerhof, and K. Burke, J. Chem. Phys. **105**, 9982 (1996).
[13] F. Gygi and A. Baldereschi, Phys. Rev. B **34**, 4405 (1986).
[14] Relaxations at the hybrid level for selected defects yielded negligible differences.
[15] http://www.quantum-espresso.org.
[16] CPMD, Copyright IBM Corp 1990–2006, Copyright MPI für Festkörperforsch. Stuttgart 1997–2001.
[17] P. A. Schultz, Phys. Rev. Lett. **96**, 246401 (2006).
[18] A. Alkauskas and A. Pasquarello, Physica B (Amsterdam) **401–402**, 670 (2007).
[19] J. C. Slater, Adv. Quantum Chem. **6**, 1 (1972).
[20] S. Öğüt, J. R. Chelikowsky, and S. G. Louie, Phys. Rev. Lett. **80**, 3162 (1998).
[21] J. P. Perdew and M. Levy, Phys. Rev. B **56**, 16021 (1997).
[22] V. Kalyanaraman, M. M. Chandra, and V. Kumar, J. Appl. Phys. **54**, 6417 (1983).
[23] L. W. Song and G. D. Watkins, Phys. Rev. B **42**, 5759 (1990).
[24] L. Skuja, J. Non-Cryst. Solids **239**, 16 (1998).
[25] G. Pacchioni, G. Ieranó, and A. M. Márquez, Phys. Rev. Lett. **81**, 377 (1998).
[26] For Si, the reported defect levels are extrapolated from supercell calculations with 64, 216, and 512 atoms.
[27] R. Shaltaf et al., Phys. Rev. Lett. **100**, 186401 (2008).
[28] G. Pacchioni et al., Phys. Rev. B **63**, 054102 (2000).
[29] D. V. Melnikov and J. R. Chelikowsky, Phys. Rev. Lett. **92**, 046802 (2004).
[30] M. J. Caldas, A. Fazzio, and A. Zunger, Appl. Phys. Lett. **45**, 671 (1984).
[31] J. M. Langer and H. Heinrich, Phys. Rev. Lett. **55**, 1414 (1985).