8
Accurate Gap Levels and Their Role in the Reliability of Other Calculated Defect Properties

Peter Deák, Adam Gali, Bálint Aradi, and Thomas Frauenheim

8.1
Introduction

Bulk defects and surfaces give rise to characteristic fingerprints in the electrical, optical, and magnetic spectra of non-metallic crystals, critically influencing their functionality in applications. The task of defect theory is to establish the equilibrium concentration of conceivable defects (without or in the presence of other defects) and to calculate their properties for comparison with the experimental spectra. The information gained from the joint efforts of defect theory and spectroscopy serves as database for defect engineering, which has become an integral part of technology design in electronics, optoelectronics, and photovoltaics, but also serves the understanding of surface processes like heterogeneous catalysis.

From the viewpoint of the electrical and optical properties of the bulk, as well as of the chemical reactivity of surfaces, band gap states are of paramount importance in non-metallic solids. Their position with respect to the band edges determines the electrical and optical spectra but contributes also dominantly to the formation energy. Calculating defect level positions has been the toughest challenge for defect theory, because its "work horse" in the past decades [1], density functional theory (DFT) in its standard implementations – the local density approximation (LDA) and the semi-local generalized gradient approximation (GGA) – leads to a serious underestimation of the band gap (sometimes to no gap at all), and to big uncertainties in the defect level positions in it. For a long time, this deficiency has been regarded as a relatively minor problem, hampering only the comparison of the calculated spectra with the experimental ones but, in fact, it has serious implications for the formation energy [2], and so for the relative stability of different defect configurations. In this paper we will demonstrate this and review our experience with different correction schemes for calculating defect properties free of the gap error.

Before doing so, however, let us consider the ways of calculating electrical and optical spectra. Usually, what is being measured is the energy promoting an electron from the valence band maximum (VBM) to a defect level (acceptors), or to the conduction band minimum (CBM) from a defect level (donors). Of course, internal

Advanced Calculations for Defects in Materials: Electronic Structure Methods, First Edition.
Edited by Audrius Alkauskas, Peter Deák, Jörg Neugebauer, Alfredo Pasquarello, and Chris G. Van de Walle.

excitation of the defect is also possible, and the reverse (recombination) processes can also be measured. In some experiments, the measured energy absorption or emission corresponds to the change in the electronic energy alone, while in others to the change in the total energy. With reference to a Franck–Condon diagram, the former are called vertical transitions, while the latter, where the ions have time to relax, are called adiabatic. Since (except for internal excitations) the charge state of the defect itself changes, these transition energies are often referred to as "optical" or "thermal" charge transition levels of the defect, respectively. In principle, an excitation energy should be calculated as the total energy difference between the ground and excited states, but DFT can be applied only to the former. However, excitations energies can also be deduced by comparing the ground state energies of different charge states. Using charge transition energies of the perfect and the defective crystal, as shown in Figure 8.1, one can obtain the energies of the required electronic transitions from/to the band edges to/from the defect level. For vertical (optical) transitions, the total energy differences have to be taken at the geometry of the initial state, while for adiabatic (thermal) transitions at the respective equilibria of the final and initial states. Actually, the charge transition energies with respect to the vacuum level correspond to the ionization energies and electron affinities of the systems. (*N.B.*: the electron affinity is the ionization energy of the negative state.) In DFT, if the applied functional was exact, the negative of the highest occupied Kohn–Sham (KS) level would be exactly equal with the ionization energy [3, 4]. In supercell calculations the vacuum level is not defined but the common reference level of Figure 8.1 does not appear in the required ionization energy differences. So, e.g., for the infinite system, $I_C - A_C = \varepsilon_{CB} - \varepsilon_{VB} = E_g$ [2]. The standard local and semi-local approximations of the exchange functional introduce a spurious electron self-interaction which leads to the "band gap error," irrespective of which way one calculates it (assuming that proper band-filling corrections were applied in the total energy calculated at special *k*-points) [2]. Also, while the total energy should be a linear function of the occupation number (between integer values), in the standard

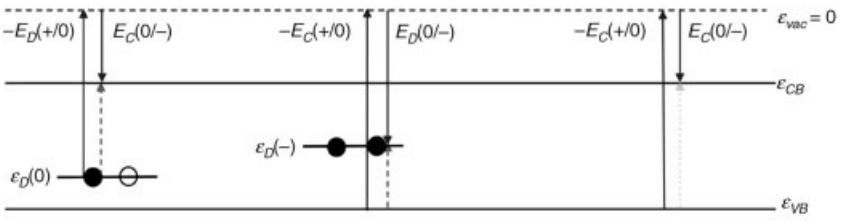

Figure 8.1 (online color at: www.pss-b.com) Transition energies as ground state energy differences. Lines denoted with the Greek letter ε represent (Kohn–Sham) one-electron levels, arrows denoted with Latin *E* correspond to transition energies between the charge states given in parentheses for the perfect (C) and the defective (D) crystal. *I* and *A* are the corresponding ionization energies and electron affinities, respectively. Arrows with dashed lines (red) are the donor and acceptor transitions, while the dotted (green) one is the fundamental absorption (E_g) of the bulk.

approximations this function has a positive curvature. This leads to the improper placement of the KS levels of the defect with respect to the band edges [4].

"Historically," the first attempts to remedy the gap problem were aimed at correcting the KS levels of the standard approximations. In Section 2 we will consider the most common methods and conclude that, save for many-body calculations on the defect containing solid, the applicability of them all are restricted to special defects even within one host material.

The "gap error," however, does not only concern the calculation of the defect level positions – it also appears in the formation energy of the defect, influencing the correct prediction of the ground state and the activation energies for diffusion or reactions. In Section 3 we will demonstrate this on a few examples. In Section 4 we will also show that total energy corrections based on the KS level corrections can only be applied in special cases.

Obviously, the whole "gap problem" could be solved by applying GW or Quantum Monte Carlo methods. The difficulty is, that – while simplified (G_0W_0) techniques can be applied to large supercells [5] to provide *a posteriori* quasiparticle (QP) corrections – self-consistent GW calculations [6] are still rather costly, restricting the size of the supercell presently, e.g., to 64–72 atoms for tetrahedral semiconductors (at the very limit), and calculation of the total energy is as yet not possible even for these. Cost factors limit Quantum Monte Carlo calculations even more, mostly forbidding even to take into account relaxation effects [7]. Therefore, in the time being, we think that the best way of dealing with the gap problem is to turn to a generalized Kohn–Sham scheme, based on approximate non-local exchange functionals [8]. There are many possibilities available (see, e.g., Refs. [9–14]). In Section 5 we will investigate the performance of the screened hybrid exchange functional of Heyd, Scuseria, and Ernzerhof [15, 16] for a set of well chosen defects. Our conclusion is that carefully tested semi-empirical, range-separated hybrids are presently the best tools for economically feasible studies of both defect spectra and energetics, with transferability within a class of hosts of similar bonding and irrespective of the nature of the defect.

8.2
Empirical Correction Schemes for the KS Levels

First attempts for correcting the consequences of the gap error for defects have concentrated on the gap levels. Since calculated effective masses were acceptably accurate, the assumption was that (semi)local approximations to DFT described both bands accurately, just the energy difference between the VBM and the CBM was too small [17]. This has led to the idea of taking the VBM as reference, open up the gap to its experimental value, and scale the energy difference of the defect level to the VBM accordingly. Needless to say that this procedure has no justification whatsoever. A more intelligent approach was to consider whether the defect state was VB or CB related and – as a first approximation – apply the same shift (w.r. to the VBM) as for the CBM in the latter case, or no shift at all in the former. Mostly, however, no clear-cut

decision is possible. This was taken into account by the "scissor operator," introduced by Baraff and Schlüter [18] for the case of a vacancy. Since the wave function of the defect can be expanded on the basis of the perfect crystalline states, they assumed that the amount of necessary shifting for a defect level can be determined by the weight of the conduction-band states in the expansion. Therefore, the shift necessary for the CBM to reproduce the experimental gap is to be scaled by the sum of overlaps between the defect wave function and all CB states of the perfect crystal. In principle, the scissor operator can be applied self-consistently, but most often it was applied *a posteriori*, which made sense only if the too low-lying CBM in the LDA or GGA calculation did not mask the localized defect level [2]. (Otherwise the calculation would lead – incorrectly – to an effective mass like state and a full shift with the CBM.)

In our experience, the scissor operator has worked reasonably well for substitutional defects and split-interstitials but for interstitial defects in the low electron density region of the crystal the results seemed to be problematic. Therefore, we carried out a case study using hydrogen as a probe in silicon and silicon carbide [19]. The hydrogen interstitial in Si has established configurations both in the high and low electron density regions of the crystal [20]. In the neutral charge state, it intercepts a Si–Si bond. At this, so-called bond-center (BC) site the hydrogen is a donor. There exists, however, a metastable site behind a Si–Si bond, in the so-called antibonding (AB) position (near to the tetrahedral interstitial site T). Here the hydrogen acts as an acceptor. This provides a unique opportunity to check how the validity of the scissor correction depends on the position of the defect. In SiC the interstitial H is always an acceptor at the AB site, behind a silicon atom. Comparing that to the case of H_{AB} in Si can show how the scissor correction works in materials with very different gaps. In our study we have used standard local (spin) density approximation, and compared the scissor correction to the QP-correction of a G_0W_0 calculation. (Details are described in Ref. [19].)

The results are shown in Table 8.1. The scissor correction is close to the G_0W_0 result for H_{BC}. This corroborates the experience obtained with substitutional defects that the scissor operator works well in the high electron density region of the crystal.

Table 8.1 Comparison of the scissor- and the G_0W_0 QP-corrections [in (eV)] to the KS levels obtained by LDA for interstitial H in Si and SiC. (Also corrections predicted by a one-parameter hybrid functional are given).

level[a)]	scissor	G_0W_0	hybrid
Si: CBM	0.61	0.66	0.61
Si : H^0_{BC}	0.53	0.44	0.47
Si : H^-_{AB}	0.44	0.17	0.12
SiC:CBM	1.08	1.17	1.12
SiC : H^-_{AB}	0.90	0.12	0.04

a) The position of the CBM has been set according to the experimental gap in the case of the scissor operator. The mixing parameter of the hybrid functional has been fitted to reproduce the lattice parameter, cohesive energy, bulk modulus, and the band gap.

In contrast the scissor yields a gross overcorrection for H_{AB} in both materials. (The error increases in 4H–SiC, relative to Si, by about the same amount as the increase in the gap correction.) The explanation lies in the different nature of the defect state at BC and AB. At BC in Si it is essentially an antibonding combination of the sp^3 hybrids on the Si neighbors, i.e., clearly conduction band derived. Therefore, it can very well be described by a linear combination of the CB states of the perfect crystal. The defect state of H at the AB site (i.e., near the tetrahedral interstitial site T) is a good example for an 1*s* effective-mass state. However, in this interstitial "hole" of the tetrahedrally bonded semiconductors, the electron density is small, so only CB states can be involved in the expansion of even such an essentially valence state [21]. As a result, the scissor gives a correction almost as big as that of the CBM, being increasingly wrong with increasing gap. Obviously, the basic assumption of the scissor operator works only in the high electron density region of the perfect crystal, and not if the defect resides in a low electron density region where the VB states of the perfect crystal do not provide an adequate basis for expanding a strongly localized defect state.

Another way of dealing with the gap problem is to correct the host band structure of a standard (semi)local DFT approximation. One possibility for that is the application of an *a posteriori* alignment scheme [22, 23]. It has been observed that the QP-corrections work like a symmetric scissor, pushing the VBM down and the CBM up, and – at least for some defects – the position of the gap levels, with respect to a suitable chosen "external" reference, hardly changes [23]. This would allow to correct the positions of the VBM and the CBM, based on – say GW – calculations done on the small unit cell of the perfect system, which is much less costly to carry out than for the defective super cell. Unfortunately, however, this idea works only for defects with well localized mid-gap levels [23], and provides no way of correcting the total energy (see next section). A self-consistent but empirical way of correcting the host band structure is offered by the use of the LDA + U (or GGA + U) methods, and/or by applying non-local empirical pseudopotentials (NLEP) for correction, as described in Ref. [4]. One problem with this procedure is that the adjustment of the KS levels is often accompanied by the deterioration in the ground state properties of the system (e.g., ionicity and lattice constant). The other problem is that a reasonably justified choice of the empirical parameters can only be made for the host system but not for an impurity. Our attempt, in the latter case, to choose parameters by enforcing the equality of the ionization energies, calculated either as total energy difference or as the negative of the highest occupied KS level, has failed.

8.3
The Role of the Gap Level Positions in the Relative Energies of Various Defect Configurations

While the methods mentioned in the previous Section can be used – at least for some defects – to obtain spectra comparable to experiment, the consequences of the gap error in the total (and formation) energy still remain. Here we would like to recall two

examples in silicon [24], which demonstrate how the error in the gap level position directly influences the relative energy of different configurations. We compare here calculations with pure GGA exchange to those with a one-parameter hybrid functional. The mixing parameter of Hartree–Fock and GGA exchange in the latter have been chosen to optimize the lattice constant, the cohesive energy, the bulk modulus, and its derivative, as well as the widths of the VB and the fundamental gap [25]. With this optimal value, also higher bulk excitations are well reproduced and, as shown in Table 8.1, the level positions of H in Si are very near to the ones obtained by GW. Later, in Section 8.5, we will show that hybrid functionals can really be used as reference, providing not only correct gap level positions but also total energy differences free of the gap error.

The first example is interstitial oxygen in silicon, in its ground state O_i, as puckered BC interstitial, and in the so called O_Y configuration, which is the saddle point along the diffusion path (Figure 8.2). The activation energy for diffusion is well established experimentally: 2.53 eV in the 270–700 °C range [26]. In a theoretical calculation at 0 K, one should expect a somewhat higher value. Based on the observed Si–O–Si stretch frequency of $\sim$1100 cm^{-1}, the zero-point energy in the ground state can be estimated to be 0.07 eV, so the 0 K theoretical barrier should be above 2.6 eV. It is known, that well converged LDA or GGA calculations underestimate this activation energy: our GGA calculation in a 64 atom supercell resulted in 2.37 eV. In contrast, the hybrid functional gave 2.69 eV, which is well in line with experiment. For both functionals, the increase of the total energy follows the emergence of a gap level from the VB, when going from the electrically inactive O_i configuration toward the saddle point at O_Y, where the central Si atom has a *p*-type dangling bond, doubly occupied due to electron transfer from the trivalent oxygen atom. Considering that the gap level is doubly occupied, the 0.14 eV difference in the level positions between the hybrid and the pure GGA calculation at the saddle point seems to explain most of the deviation of the activation energy, 0.32 eV. Although the numerical agreement is somewhat accidental (as shown in the next Section), this finding indicates that the error of GGA in predicting the activation energy is related to the gap error.

The other example is the complex of substitutional boron with a self-interstitial, $B_{Si} + Si_i$. This defect gives rise to the charge transition level shown in Table 8.2, is paramagnetic in the neutral charge state and, according to the observed hyperfine interactions, it has C_{1h} symmetry [27]. LDA and GGA studies result in a metastable configuration with such a symmetry but give also a more stable one with C_{3v}

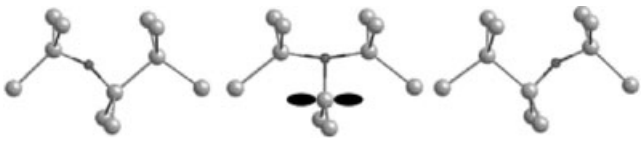

Figure 8.2 (online color at: www.pss-b.com) The diffusion path of interstitial oxygen in silicon (O_i) from ground state to ground state through the O_Y transition state. The undercoordinated Si in the latter is indicated by the lone pair *p* state. This state gives rise to a level in the gap. The oxygen atom is the dark (red in color) sphere.

Table 8.2 (+/0) charge transition levels in silicon [with respect to the perfect crystal VBM in (eV)], calculated by an LDA functional, and after correcting the total energies according to Eq. (8.1), based on a calculation by G_0W_0 or by a one-parameter hybrid functional.

E(+/0) w.r. VBM		with corrections based on level positions in		
	LDA	G_0W_0	hybrid	exptl.
H_{BC}	0.54	0.98	0.98	0.94[a)]
$B_{Si} + Si_i$	0.66		0.94	0.99[b)]

a) Ref. [29],
b) Ref. [27]

symmetry, as shown in Figure 8.3 (see Ref. [25] and references therein). This has always been suspected to be a consequence of the gap problem [28]. Calculation with the hybrid functional proves that, resulting the C_{1h} configuration lower in energy than the C_{3v} one. The reason is that in the C_{1h} configuration, which is nearly a [110] split-interstitial, a different kind of orbital is occupied than in the C_{3v} configuration, where practically the boron is an on-center substitutional and Si_i is near the tetrahedral interstitial site. The gap level position is shifted up between the GGA and the hybrid calculation much more in the C_{3v} case, than in the C_{1h}. This difference brings about a larger increase of the total energy for the C_{3v} configuration than for the C_{1h}, leading to a reversal in the stability sequence. This is clearly a case where LDA and GGA both fail to predict the correct ground state of a system because of the gap error.

These examples show, that the error in the gap level position directly influences the relative energies of defects and, without appropriate correction, this can lead to serious quantitative and qualitative errors in the predictions based on (semi)local approximations of DFT. Note, that the error of the gap level positions increases with the width of the gap (see Table 8.1), and can easily cause catastrophic problems for defects in wide band gap materials.

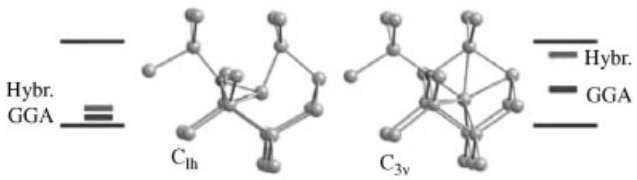

Figure 8.3 (online color at: www.pss-b.com) The C_{1h} (left) and C_{3v} (right) configurations of the $B_{Si} + Si_i$ complex in silicon. The boron atom is the dark (green in color) sphere. Also shown are the defect level positions in the gap, as obtained by a pure GGA and a hybrid functional.

8.4
Correction of the Total Energy Based on the Corrected Gap Level Positions

The examples described in the previous section suggest a possibility of correcting the total energy. The latter can always be given as the sum of the band energy and of double-counting terms, and the band energy can be split into the contribution of the occupied defect level in the gap and that of the valence band:

$$E_{\text{tot}} = E_{\text{BE}} + E_{\text{dc}} = n_{\text{D}}\varepsilon_{\text{D}} + \sum_{i}^{\text{VB}} n_i\varepsilon_i + E_{\text{dc}}(\varrho) \tag{8.1}$$

where ε_{D}, ε_{i} denote KS levels in the gap and in the VB, respectively, and n_{D}, n_{i} are the corresponding occupation numbers. Equation (8.1) clearly shows that an error in the first term of the right-hand side will influence the total energy, since the double-counting term depends only on the electron density $\varrho(r)$, which – according to GW calculations – are quite well described by the (semi)local approximations, and there is no reason to expect compensation by the second term. Based on Eq. (8.1), we have used the following correction scheme for the total energy:

$$E_{\text{tot}}^{\text{corr.}} = E_{\text{tot}} + n_{\text{D}}\Delta\varepsilon_{\text{D}}, \tag{8.2}$$

$$\Delta\varepsilon_{\text{D}} = [\varepsilon_{\text{D}}^{\text{corr.}} - \varepsilon_{\text{VBM}}^{\text{corr.}}] - [\varepsilon_{\text{D}}^{\text{uncorr.}} - \varepsilon_{\text{VBM}}^{\text{uncorr.}}] \tag{8.3}$$

Table 8.2 shows the results for the two donors discussed above in silicon, using the gap level position obtained in the hybrid calculation for correction [24]. In the first case the correction is also given based on a G_0W_0 calculation [19]. As can be seen the corrected results compare favorably to experiment. This correction scheme has recently been proposed again using GW data in Eq. (8.3) [30, 31].

Actually, we have been using this correction scheme for quite some time, initially with the scissor correction for the gap levels [32], later with the level positions obtained from one-parameter hybrid functional calculations [33]. Our experience about its success was, however, somewhat mixed, so we have examined [24] the working of Eqs. (8.2) and (8.3) on the energy difference (i) between the BC and AB sites (in the same charge state) for interstitial Si:H_i, (ii) between the ground state and saddle point configurations of Si:O_i, and (iii) between two possible configurations of a silicon vacancy in SiC, as shown in Figure 8.4. (The obvious V_{Si} configuration can transform into a $V_{\text{C}} + C_{\text{Si}}$ one [34]). Assuming that the hybrid calculation provides both gap level positions and self-consistent total energies free of the gap error, the corrections (with respect to a pure GGA calculation) to each term in Eq. (8.1) can be calculated separately. (The corrections to E_{tot} and E_{BE} can be calculated directly, while that of E_{dc} is their difference.) These are compared in Table 8.3 with the approximate correction based on Eqs. (8.2) and (8.3). As can be seen, the agreement between the first and last row is good in only one case. Looking at the contributions from E_{BE} and E_{dc}, however, it is clear that even this agreement is the result of a lucky compensation effect, which does not occur in the other two. How can this be understood in light of the success with the charge transition levels in Table 8.2?

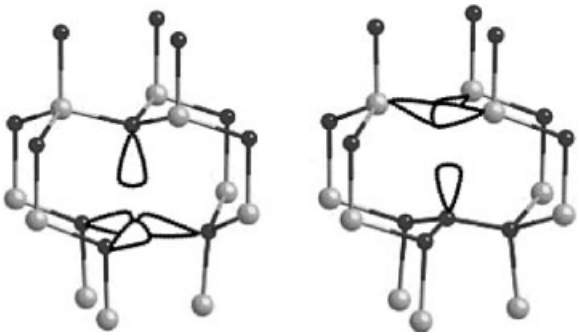

Figure 8.4 (online color at: www.pss-b.com) V_{Si} (left) and its isomer $V_C + C_{Si}$ (right) in SiC. The carbon atoms are the dark (blue) spheres. The dangling bonds are schematically indicated.

In each of the three cases examined in Table 8.3, there is considerable change in the bonding. Hydrogen at the BC site in silicon has a singly occupied gap level near the CB edge, while at the AB site it has a doubly occupied resonance just below the VB edge. No correction for the latter has been taken into account in the last row, but is included in the first. (Apparently the two corrections happen to cancel each other accidentally to a high degree.) The oxygen atom in its interstitial position in silicon is divalent, with two tetravalent Si neighbors. At the saddle point of its motion it becomes trivalent and one Si neighbor becomes undercoordinated. Obviously, very different VB resonances correspond to the two cases and the change in E_{BE} is considerably more than just the correction coming from the appearance of the gap level. (Still, the error in the latter does influence the total energy significantly.) Between the two different configurations the electron density, and so E_{dc} changes as well. For H in Si this seems to compensate the change in E_{BE}, for O only the VB part of the latter. In the case of the vacancy in SiC there is no compensation at all. The approximate correction covers the change from three dangling bonds on Si to three on C, but at the same time three Si–C bonds are replaced by three C–C bonds, so a non-self-consistent correction is obviously meaningless.

Table 8.3 The difference between a hybrid and a GGA calculation [24] in the relative energy of two defect configurations, *a* and *b*, split according to the various terms of Eq. (8.1). The last row gives the approximate total energy corrections based on Eqs. (8.2) and (8.3).

$a-b$	Si		SiC
	$H_{BC}-H_{AB}$	O_Y-O_i	$V_{Si}-V_C+C_{Si}$
$\Delta(E^a_{\mathrm{tot}}-E^b_{\mathrm{tot}})$	$+\mathbf{0.01}$	$+\mathbf{0.32}$	$-\mathbf{0.75}$
$\Delta(E^a_{\mathrm{BE}}-E^b_{\mathrm{BE}})$	-0.04	$+0.51$	$+1.08$
$\Delta(E^a_{\mathrm{dc}}-E^b_{\mathrm{dc}})$	$+0.05$	-0.19	-1.83
$n^a_{\mathrm{D}}(\varepsilon^a_{\mathrm{D}}+\Delta\varepsilon^a_{\mathrm{D}})-n^b_{\mathrm{D}}(\varepsilon^b_{\mathrm{D}}+\Delta\varepsilon^b_{\mathrm{D}})$	$+\mathbf{0.19}$	$+\mathbf{0.28}$	$-\mathbf{1.89}$

In contrast to these cases, the simple change in the occupation of a gap level may shift it a little, but VB resonances will hardly be affected and little change will occur in E_{dc} – at least in the cases mentioned here. However, in case of bistable defects, the charge transition may induce a substantial relaxation of the nuclei and a very different bonding. Therefore, the *a posteriori* corrections scheme of Eqs. (8.2) and (8.3) should be used with utter care.

8.5
Accurate Gap Levels and Total Energy Differences by Screened Hybrid Functionals

Considering all the problems with correcting the results of the (semi)local exchange functionals on the one hand, and the unfeasibility of applying many-body theories to large supercells on the other, generalized KS (*g*KS) schemes with approximate non-local exchange functionals seem to offer the solution in the time being. Such approximations are the hybrid functionals. Based on the adiabatic connection formula Becke has suggested [35, 36] an approximation to the exact DFT exchange energy by mixing GGA and Hartree–Fock (HF) exchange. The mixing parameter of these hybrid functionals were chosen semi-empirically to optimize thermochemical data of molecules. The optimal choice of 25% HF-exchange can also be justified theoretically [37]. It has been observed early on that hybrid functionals systematically improve the *g*KS gap of semiconductors [38]. Encouraged by that, we have determined materials-specific mixing parameters for crystalline Si [25], SiC [33] and SiO_2 [39], by fitting ground state properties as well as the gap to experimental values, as mentioned earlier. Table 8.4 shows the band gaps obtained. Although these hybrids have proved themselves in several applications (see, e.g., Refs. [33, 39–42]), their lack of transferability from one material to another is a severe restriction, especially for interface studies.

More recently, a new class of hybrid functionals has been introduced [15], where the mixing is done only for the short range part of the electron–electron interaction. This corresponds to screened non-local exchange (screened hybrids). The screening parameter introduces an additional degree of freedom, and optimizing these can give excellent gaps for a wide range of semiconductors (but not for all). The first version

Table 8.4 Band gaps obtained by a one-parameter hybrid exchange functional. Values in Italics in each column were obtained by the mixing parameter optimized for ground state properties and the gap of the given material. The experimental values are shown in parentheses. All values are in (eV).

HF-part	fundamental gap (eV)			
	SiO_2 (9.5)	Si (1.17)	3C-SiC (2.36)	C (5.48)
12%		*1.17*		
20%	9.0	1.44	*2.42*	5.12
28%	*9.5*			

Table 8.5 Comparison of PBE and HSE06 results [46] for the lattice constant (a, c), cohesive energy (E_{coh}), bulk modulus (B_0), fundamental (indirect) gap (E_g), first allowed direct transition at the Brillouin-zone center ($\Gamma_{25}' \rightarrow \Gamma_2'$), and valence band width (VB) with experimental data [47, 48] in case of the Group IV semiconductors.

	method	a (Å)	c (Å)	E_{coh} (eV)	B_0 (GPa)	E_g (eV)	$\Gamma_{25}' \rightarrow \Gamma_2'$ (eV)	VB (eV)
diamond	PBE	3.574		7.85	425	4.21	13.3	21.5
	exptl.	**3.567**		**7.37**	**443**	**5.48**	**15.3**	**24.2**
	HSE06	3.544		7.58	464	5.42	15.7	23.8
4H-SiC	PBE	3.091	10.116	6.51		2.22		
	exptl.	**3.073**	**10.053**			**3.23**		
	HSE06	3.069	10.045	6.37		3.21		
3C-SiC	PBE	4.375		6.51	209	1.37	6.1	15.3
	exptl.	**4.360**		**6.34**	**224**	**2.36**	**7.4**	**17.**
	HSE06	4.345		6.37	230	2.25	7.7	17.1
silicon	PBE	5.468		4.62	88	0.61	3.14	11.8
	exptl.	**5.429**		**4.63**	**99**	**1.17**	**4.15**	**12.5**
	HSE06	5.434		4.54	98	1.17	4.33	13.3
germanium	PBE					0.00		
	exptl.	**5.658**		**3.88**		**0.74**	**0.90**	**13.0**
	HSE06	5.670		3.66		0.84	0.88	13.9

(HSE03) of this screened hybrid by Heyd–Scuseria–Ernzerhof [15] has shown substantial improvement of the band gap over a one-parameter hybrid with 25% HF-part (termed PBE0 in Ref. 9) and, above all, the same quality in similarly bonded materials. It is important to emphasize that at the same time also the reproduction of the basic ground state properties (lattice constant, heat of formation, and bulk modulus) have also improved [43].

In Table 8.5 we show this for the Group IV semiconductors, diamond, SiC, Si, and Ge, using the revised HSE06 version [16]. (The screening parameter is set to 0.2 $Å^{-1}$, keeping the 25% admixture of HF-exchange to 75% GGA exchange calculated by the Perdew, Burke, and Ernzerhof – or PBE – functional [44].) We would like to point out that details of the band structure, like width of the VB or the first allowed direct transition at Γ, which have not been included into the fitting procedure, agree also well with experiment. The band gaps are reproduced in all these materials with the same high accuracy. This is even true for Ge (taking into account the lack of spin-orbit coupling in our calculation, which would lower the fundamental gap), for which the PBE calculation gives no gap at all. The transferability of HSE06 in materials with similar bonds is encouraging, provided it pertains also to defect levels. In a recent paper we have found that HSE06 works extremely well for the inner excitation of a defect in diamond [45]. We have, therefore, also investigated the transition energies between the band edges and defect states in a systematic manner [46].

From the view point of defects, the approximate non-local exchange functional in the *g*KS scheme is expected to remedy both the gap problem and the inappropriate dependence of the total energy on the occupation number, or in other words: to provide both total energy and *g*KS levels of the defect free of the gap error. To check this, we have compared the band ↔ defect transitions computed (cf. Figure 8.1) as differences of self-consistent total energies (ΔSCF method) to values obtained as differences between highest occupied *g*KS levels (ΔKS method). The average potentials between the perfect crystal and the defective supercell have been aligned using the method suggested in Ref. [1]. Charged supercells were calculated assuming a jellium charge of opposite sign. This leads to an error, dependent on the supercell size, both in the total energy and in the *g*KS levels [31]. In general, the error also depends on the nature of the defect state, requiring non-trivial correction procedures [49]. Therefore we have chosen fairly large, 512-atom supercells (in the Γ approximation) to minimize all size effects, and applied 65% of the monopole correction for charged supercells, as suggested in Ref. [2]. In case of acceptors both the ΔKS and ΔSCF transitions need correction, and we assumed them to be approximately equal for a single negative charge.

In Table 8.6 we compare the ΔKS and ΔSCF vertical transitions for a series of donors and acceptors in diamond, silicon, and germanium. For Si, the defects have been chosen to scan the whole width of the gap by their gap levels. With one exception, the agreement is within 0.1 eV, irrespective of the gap width of the host, or the shallow or deep nature of the defect.

Table 8.7 shows the comparison of the calculated adiabatic transition energies to the experimental ones. The adiabatic "ΔKS" values have been obtained by adding the relaxation energy of the charged state (with respect to the neutral one) to the vertical ΔKS transition. The values obtained this way are in stunning agreement with experiment. (*N.B.* the defects in this study have been chosen by the criterion of having accurate experimental values beyond doubt.) The same is true for the ΔSCF results, except for the case of the iron interstitial in silicon ($Si:Fe_i$), the "odd guy out" also in Table 8.6. Comparison of the two Tables show that the error is in the calculated ΔSCF

Table 8.6 Vertical transition energies [in (eV)] calculated by comparing highest occupied levels (ΔKS) or total energies (ΔSCF) according to Figure 8.1. The E(+/0) transition levels of donors are given with respect to the VBM, the E(+/0) transitions levels of acceptors with respect to the CBM.

donors			acceptors		
***E*(+/0) w.r. CBM**	**ΔKS**	**ΔSCF**	***E*(0/−) w.r. VBM**	**ΔKS**	**ΔSCF**
$C_{512}:P_C$	−0.6	−0.6	$C_{512}:B_C$	+0.3	+0.4
$Si_{512}:S_{Si}$	−0.3	−0.3	$Si_{512}:In_{Si}$	+0.2	+0.2
$Si_{512} : S_{Si}^{+}$	−0.6	−0.6	$Si_{512}:O_{Si}$	+0.9	+1.0
$Si_{512}:Fe_i$	*−1.0*	*−0.7*			
$Si_{512}:Au_{Si}$	−0.9	−0.8			
$Si_{512}:C_i$	−1.0	−0.9	$Si_{512}:C_i$	+1.0	+1.0
$Ge_{512}:S_{Ge}$	−0.4	−0.3	$Ge_{512}:O_{Ge}$	+0.4	+0.4

Table 8.7 Adiabatic ΔKS[a)] and ΔSCF transition energies compared to experiment. Experimental values are from Ref. [50].

donors				acceptors			
***E*(+/0) w.r. CBM**	**ΔKS**	**ΔSCF**	**exptl.**	***E*(0/−) w.r. VBM**	**ΔKS**	**ΔSCF**	**exptl.**
C_{512}:P_C	−0.5	−0.5	−0.6	C_{512}:B_C	+0.3	+0.4	+0.4
Si_{512}:S_{Si}	−0.3	−0.3	−0.3	Si_{512}:In_{Si}	+0.2	+0.2	+0.2
Si_{512}:S_{Si}^+	−0.5	−0.5	−0.6	Si_{512}:O_{Si}	+0.8	+0.9	+0.9
Si_{512}:Fe_i	*−0.8*	*−0.5*	*−0.8*				
Si_{512}:Au_{Si}	−0.8	−0.8	−0.8				
Si_{512}:C_i	−0.9	−0.8	−0.9	Si_{512}:C_i	+0.9	+1.0	+1.0
Ge_{512}:S_{Ge}	−0.3	−0.3	−0.3	Ge_{512}:O_{Ge}	+0.3	+0.3	+0.3

a) The relaxation energy of the charged state (with respect to the neutral one) was added to the vertical ΔKS transition.

vertical transition. This may be partly an error of the simplified charge correction, but definitely not entirely. (The error is bigger than the whole monopole correction.) Si:Fe_i has a gap state which is highly localized on a Fe 3d orbital, i.e., compared to the other defects, has the least contribution form host derived states. (In all other cases the gap state is either effective mass like or a combination of host dangling bonds. The donor state of Si:C_i is a pure p orbital on C.) Therefore, a possible explanation for the error might be that the screened hybrid can mimic the accurate non-local exchange functional in every respect only for states characteristic to the class of hosts for which the parameters have been chosen. For other states the dependence of the total energy on the occupation number is still seemingly correct, but not the absolute value.

If the above analysis is correct, the defect energetics obtained by a screened hybrid can only be fully trusted if the defect state is dominantly host-related. Still, the position of the gap level seem to be supplied by high accuracy in every case. This has three advantages: (i) the experimental spectrum can be predicted, (ii) comparison of the ΔKS and ΔSCF vertical transition energies provides a convenient way of assessing the reliability of the energetics, and (iii) the transition energies can be calculated without the need for a charge correction.

8.6 Summary

We have considered, how the "gap error" of the standard (semi)local DFT approximations influences both the spectra and relative energies of defects, and investigated several ways of correction. We conclude that from the point of view of the spectrum,

i) the scissor operator is a reasonably accurate and very convenient method of correcting the gap level position, if the defect is in the high electron density region of the perfect crystal, but fails outside of that.

ii) Methods of correcting the KS levels of the host work only for a certain class of defects even in one material.
iii) Gap level positions in semi-empirical screened hybrid calculations are just as accurate as the gap for which they have been parameterized, independent of the nature of the defect. The parameters are transferable within a class of materials with similar bonding.

From the view point of calculating the relative energy of different defect configurations (or charge states) we conclude that

iv) Different kind of gap states and VB resonances in the two configurations lead to different errors in the band energy, and so to a substantial error in the relative energy. The latter increases with the gap error (~gap width) and can lead to reversal of the stability ordering. Therefore, in such cases the LDA or GGA total energy has to be corrected for the "gap error" as well.
v) Correcting the band energy for the gap level alone is only sufficient if the relaxation upon changing the defect state is small.
vi) The total energy supplied by semi-empirical screened hybrids seems to be largely free of the consequences of the "gap error" for defect states with the character of the orbitals in those materials for which the parameters have been optimized.

It appears that, in the time being, well parameterized semi-empirical screened hybrids are the preferred method for calculating relative energies and electronic transitions for defects. Although much less expensive than GW or Quantum Monte Carlo calculations, their computational cost is still about an order of magnitude higher than for LDA or GGA. Therefore, a careful check of (iv) is recommended.

Acknowledgements

The authors are grateful for fruitful discussions with S. Lany, G. Kresse, and G. E. Scuseria. Support of the Supercomputer Center of Northern Germany (HLRN grant no. hbc00001), as well as that of the German–Hungarian bilateral research fund (436 UNG 113/167/0-1) are appreciated. AG acknowledges the support of the Hungarian grants OTKA (no. K67886) and NKTH (Nr. NKFP-07-A2-ICMET-07), as well as the János Bolyai program from the Hungarian Academy of Sciences.

References

1 Van de Walle, C.G. and Neugebauer, J. (2004) *J. Appl. Phys.*, **95**, 3851.
2 Lany, S. and Zunger, A. (2008) *Phys. Rev. B*, **78**, 235104.
3 Janak, J.F. (1978) *Phys. Rev. B*, **18**, 7165; Ambladh, C.-O. and von Barth, U. (1985) *Phys. Rev. B*, **31**, 3231.
4 Lany, S. and Zunger, A. (2009) *Phys. Rev. B*, **80**, 085202. (Chapter 11).
5 Furthmüller, J., Cappellini, G., Weissker, H.C., and Bechstedt, F. (2002) *Phys. Rev. B*, **66**, 045110.
6 Shishkin, M. and Kresse, G. (2007) *Phys. Rev B.*, **75**, 235102.

7 Batista, E.R., Heyd, J., Hennig, R.G., Uberuaga, B.P., Martin, R.L., Scuseria, G.E., Umrigar, C.J., and Wilkins, J.W. (2006) *Phys. Rev. B*, **74**, 121102(R).

8 Fuchs, F., Furthüller, J., Bechstedt, F., Shishkin, M., and Kresse, G. (2007) *Phys. Rev. B*, **76**, 115109.

9 Adamo, C. and Barone, V. (1999) *J. Chem. Phys.*, **110**, 6158.

10 Robertson, J., Xiong, K., and Clark, S.J. (2006) *Phys. Status Solidi B*, **243**, 2054. (Chapter 5).

11 Janotti, A., Segev, D., and Van deWalle, C.G. (2006) *Phys. Rev. B*, **74**, 045202.

12 Baumeier, B., Krüger, P., and Pollmann, J. (2007) *Phys. Rev. B*, **76**, 085407.

13 Broqvist, P., Alkauskas, A., and Pasquarello, A. (2009) *Phys. Rev. B*, **80**, 085114.

14 Wu, X., Selloni, A., and Car, R. (2009) *Phys. Rev. B*, **79**, 085102.

15 Heyd, J., Scuseria, G.E., and Ernzerhof, M. (2003) *J. Chem. Phys.*, **118**, 8207; Heyd, J. and Scuseria, G.E. (2004) *J. Chem. Phys.*, **121**, 1187. (Chapter 6).

16 Krukau, A.V., Vydrov, O.A., Izmaylov, A.F., and Scuseria, G.E. (2006) *J. Chem. Phys.*, **125**, 224106.

17 In fact, the total width of the bands is always underestimated. E.g., the indirect band gap in silicon is underestimated by only 0.5 eV, but the direct transition at Γ from the VBM to the second subband in the CB by 1.0 eV, indicating a "compressed" CB. The width of the VB is also too small by 6%.

18 Baraff, G.A. and Schlüter, M. (1984) *Phys. Rev. B*, **30**, 1853.

19 Deák, P., Frauenheim, T., and Gali, A. (2007) *Phys. Rev. B*, **75**, 153204.

20 Estreicher, S.K. (1995) *Mater. Sci. Eng. R*, **14**, 319; Ammerlaan, C.A.J. (2004) *Silicon. Evolution and Future of a Technology*, vol. 217 (eds P. Siffert and E. Krimmel), Springer-Verlag, Berlin, p. 261.

21 This can be seen in the almost as high a sum of overlaps with the CB states (0.72) as at the BC site (0.87).

22 Schultz, P.A. (2206) *Phys. Rev. Lett.*, **96**, 246401.

23 Alkauskas, A., Broquist, P., and Pasquarello, A. (2008) *Phys. Rev. Lett.*, **101**, 046405.

24 Deák, P., Aradi, B., Frauenheim, T., and Gali, A. (2008) *Mater. Sci. Eng. B*, **154/155**, 187.

25 Deák, P., Gali, A., Sólyom, A., Buruzs, A., and Frauenheim, T. (2005) *J. Phys.: Condens. Matter*, **17**, S2141.

26 Stavola, M., Patel, J.R., Kimerling, L.C., and Freeland, P.E. (1983) *Appl. Phys. Lett.*, **42**, 73; Takeno, H., Hayamizu, Y., and Miki, K. (1998) *J. Appl. Phys.*, **84**, 3113.

27 Watkins, G.D. (1975) *Phys. Rev. B*, **12**, 5824; Harris, R.D., Newton, J.L., and Watkins, G.D. (1987) *Phys. Rev. B*, **36**, 1094. (Chapter 7).

28 Hakala, M., Puska, M.J., and Nieminen, R.M. (2000) *Phys. Rev. B*, **61**, 8155.

29 Nielsen, K.B., Dobaczewski, S., Sogård, S., and Nielsen, B.B. (2002) *Phys. Rev. B.*, **65**, 075205.

30 Rinke, P., Janotti, A., Scheffler, M., and Van de Walle, C.G. (2009) *Phys. Rev. Lett.*, **102**, 026402.

31 Lany, S. and Zunger, A. (2010) *Phys. Rev. B*, **81**, 113201.

32 Aradi, B., Gali, A., Deák, P., Lowther, J.E., Son, N.T., Janzén, E., and Choyke, W.J. (2001) *Phys. Rev. B*, **63**, 245202; Aradi, B., Deák, P., Son, N.T., Janzén, E., Devaty, R.P., and Choyke, W.J. (2001) *Appl. Phys. Lett.*, **79**, 2746; Szűcs, B., Gali, A., Hajnal, Z., Deák, P., and Van de Walle, Ch. G. (2003) *Phys. Rev. B*, **68**, 085202.

33 Gali, A., Deák, P., Ordejón, P., Son, N.T., Janzén, E., and Choyke, J.W. (2003) *Phys. Rev. B*, **68**, 125201; Gali, A., Hornos, T., Deák, P., Son, N.T., Janzén, E., and Choyke, W.J. (2005) *Appl. Phys. Lett.*, **86**, 102108; Knaup, J.M., Deák, P., Gali, A., Hajnal, Z., Frauenheim, Th., and Choyke, W.J. (2005) *Phys. Rev. B*, **71**, 235321.

34 Bockstedte, M., Mattausch, A., and Pankratov, O. (2004) *Phys. Rev. B*, **69**, 235202.

35 Becke, A.D. (1993) *J. Chem. Phys.*, **98**, 5648.

36 Becke, A.D. (1997) *J. Chem. Phys.*, **107**, 8554.

37 Perdew, J.P., Ernzerhof, M., and Burke, K. (1996) *J. Chem. Phys.*, **105**, 9982.

38 Bredow, T. and Gerson, A.R. (2000) *Phys. Rev. B*, **61**, 5194; Muscat, J., Wander, A.,

and Harrison, N.M. (2001) *Chem. Phys. Lett.*, **342**, 397.

39 Knaup, J.M., Deák, P., Gali, A., Hajnal, Z., Frauenheim, Th., and Choyke, W.J. (2005) *Phys. Rev. B*, **71**, 235321; Knaup, J.M., Deák, P., Gali, A., Hajnal, Z., Frauenheim, Th., and Choyke, W.J. (2005) *Phys. Rev. B*, **72**, 115323.

40 Deák, P., Knaup, J.M., Hornos, T., Thill, Ch., Gali, A., and Frauenheim, Th. (2007) *J. Phys. D*, **40**, 6242–6253. Corrigendum: **41**, 049801 (2008).

41 Deák, P., Buruzs, A., Gali, A., and Frauenheim, Th. (2006) *Phys. Rev. Lett.*, **96**, 236803.

42 Aradi, B., Ramos, L.E., Deák, P., Köhler, Th., Bechstedt, F., Zhang, R.Q., and Frauenheim, Th. (2007) *Phys. Rev. B*, **76**, 035305.

43 Marsman, M., Paier, J., Stroppa, A., and Kresse, G. (2008) *J. Phys.: Condens. Matter*, **20**, 064201.

44 Perdew, J.P., Burke, K., and Ernzerhof, M. (1996) *Phys. Rev. Lett.*, **77**, 3865.

45 Gali, A., Janzén, E., Deák, P., Kresse, G., and Kaxiras, E. (2009) *Phys. Rev. Lett.*, **103**, 186404.

46 Deák, P., Aradi, B., Frauenheim, T., and Gali, A. (2010) *Phys. Rev. B*, **81**, 153203.

47 Madelung, O. (ed.) (1991) Semiconductors. Group IV Elements and II-V Compounds, in: *Data in Science and Technology*, Springer, Berlin.

48 Levinshtein, M.E., Rumyantsev, S.L., and Shur, M.S. (eds) (2001) *Properties of Advanced Semiconductor Materials: GaN, AlN, InN, BN, SiC, and SiGe*, John Wiley and Sons, New York.

49 Freysoldt, Ch., Neugebauer, J., and Van de Walle, C.G. (2009) *Phys. Rev. Lett.*, **102**, 016402. (Chapter 14).

50 Schulz, M., Dalibor, T., Martienssen, W., Landolt, H., and Börnstein, R. (2003) *Impurities and Defects in Group IV Elements, IV-IV and III-V Compounds, Landolt-Börnstein, New Series, Group III*, vol. 41, Pt. ß Springer, Berlin, Subvol. a,2.