Convergence of density and hybrid functional defect calculations for compound semiconductors

Haowei Peng,1 David O. Scanlon,2,3 Vladan Stevanovic,1,4 Julien Vidal,5 Graeme W. Watson,6 and Stephan Lany1
1National Renewable Energy Laboratory, Golden, Colorado 80401, USA
2University College London, Kathleen Lonsdale Materials Chemistry, Department of Chemistry, 20 Gordon Street, London WC1H 0AJ, United Kingdom
3Diamond Light Source Ltd., Diamond House, Harwell Science and Innovation Campus, Didcot, Oxfordshire OX11 0DE, United Kingdom
4Colorado School of Mines, Department of Physics, Golden, Colorado 80401, USA
5Institute for Research and Development of Photovoltaic Energy (IRDEP), UMR 7174 CNRS/EDF/Chimie Paristech, 6 quai Watier, 78401 Chatou, France
6School of Chemistry and CRANN, Trinity College Dublin, Dublin 2, Ireland

(Received 31 May 2013; revised manuscript received 29 July 2013; published 3 September 2013)

Recent revisions of defect formation energy calculations based on bandgap corrected hybrid functionals have raised concerns about the validity of earlier results based on standard density functionals and about the reliability of the theoretical prediction of electrical properties in semiconductor materials in general. We show here that a close agreement between the two types of functionals can be achieved by determining appropriate values for the electronic and atomic reference energies, thereby mitigating uncertainties associated with the choice of the underlying functional.

DOI: 10.1103/PhysRevB.88.115201 PACS number(s): 71.15.—m, 71.20.Nr, 71.55.—i

I. INTRODUCTION

The theoretical prediction of the electrical properties of semiconductors relies on the ability to calculate accurate defect formation energies $\Delta H$. The methodology of supercell calculations for $\Delta H$ is a topic that continues to receive great interest.1–5 A major issue that has plagued such calculations for a long time is the “bandgap problem” of the local density and generalized gradient approximations (LDA and GGA) within density functional theory (DFT). Particularly for charged defects, such as electrically active dopants, for which $\Delta H$ depends on the Fermi level, an unambiguous prediction of the formation energy requires the correct bandgap energy, and a number of correction schemes have been discussed and applied to map the formation energies and transition levels within the underestimated DFT gap onto the full gap.6–8

A major paradigm change in the field took place when the implementation of Fock exchange into plane wave DFT codes9,10 allowed bandgap corrected hybrid functional11,12 calculations of supercells, albeit at the cost of a considerably increased computational overhead that sometimes invites compromises on convergence parameters. A considerable number of works have since been devoted to revisit previously studied cases, often finding quantitatively different results or even a qualitatively changed physical picture.13–18 Such revisions have raised concerns about the validity of previous density functional results and about the accuracy of $\Delta H$ calculations in general. Considering a set of six semiconductor materials (Cu$_2$N, Cu$_2$O, Zn$_2$N$_2$, ZnO, AlN, and Al$_2$O$_3$) and a total of 24 defects and dopants in their fully ionized charge states, we show here that the difference between standard density and hybrid functionals reflects systematically the dependence of $\Delta H$ on the reference energies for the electronic and atomic chemical potentials. The convergence of the $\Delta H$ predicted by the two types of functionals is achieved by using appropriate values for these reference energies from the one-particle Green function $G$ and the screened Coulomb interaction $W$ (GW) quasiparticle energy calculations19 and from the fitted elemental reference energies (FERE) approach,20,21 respectively. By using the $GW$ band edge shifts to predict the absolute formation energies of fully ionized dopants and defects, this paper complements previous work, where such shifts were applied to band offsets22,23 and the alignment of deep-level transition energies.24

While hybrid functionals allow one to tune the band gap via the mixing parameter $\alpha$ for the Fock exchange and the range separation parameter in the case of the Heyd, Scuseria, Ernzerhof (HSE) functional,12 there are several potential sources of uncertainty associated with the choice of these parameters: (i) The parameters needed to reproduce the experimental band gap vary significantly, e.g. from $\alpha = 0.20$ in TiO$_2$ (Ref. 17) to $\alpha = 0.38$ for ZnO (Ref. 13) (both with a HSE range separation parameter of 0.2 Å$^{-1}$), introducing some ambiguity for defect calculations, e.g. if one were to consider a Ti defect in ZnO or vice versa. (ii) The parameters chosen for the semiconductor compound are not well justified for the elemental phases (e.g. the metallic phases of Ti or Zn and the O$_2$ molecule), which could affect the thermodynamic limits that define the range of defect formation energies as a function of the chemical potentials. (iii) The parameters that correct the band gap do not necessarily correct at the same time also other features, e.g. the correct d-band position in ZnO,25 the linearity of the total energy (Koopmans behavior) needed to describe correctly localized polaronic electron or hole states,26,27 or the individual absolute band edge energies, which are needed, e.g. for band offset calculations.22,23 Thus, it is desirable to devise approaches in which the defect formation energy is not strongly dependent on the hybrid functional parameters.

II. ELECTRONIC AND ATOMIC REFERENCE ENERGIES

The formation energy $\Delta H$ of a defect $D$ in the charge state $q$ is usually expressed as a sum of three terms

\[
\Delta H(E_F,\mu_a) = [E_{tot}^{D,0} - E_{tot}^H] + q(E_{VBM}^{ref} + \Delta E_F) + \sum_a n_a (\mu_a^{ref} + \Delta \mu_a). \tag{1}
\]
i.e., it is a function of the Fermi energy $E_F = E_{\text{VBM}}^\text{ref} + \Delta E_F$ and the chemical potentials $\mu_a = \mu_a^{\text{ref}} + \Delta \mu_a$, and $n_a$ is the number of atoms of type $a$ removed from $(n > 0)$ or added to $(n < 0)$ the crystal to create the defect. Taken by itself, the first term $[E_{\text{tot}}^{D,q} - E_{\text{tot}}^{H}]$, i.e., the total energy difference between the host matrix with and without the defect, is not particularly useful, since the total energies correspond to different numbers of atoms and electrons (for $q \neq 0$). A useful form of the defect formation energy is obtained only once the references for the Fermi energy and the chemical potentials are defined by physically meaningful limits. Conventionally, these limits are taken as the energy $E_{\text{VBM}}^{\text{ref}}$ of the valence band maximum (VBM) in the second term of Eq. (1), and as the (zero-temperature) chemical potential $\mu_a^{\text{ref}}$ of the elemental phases of the atomic species $a$ in the third term. For purpose of this paper, we define the reference defect formation enthalpy $\Delta H_{\text{ref}} = [E_{\text{tot}}^{D,q} - E_{\text{tot}}^{H}] + n \mu_a^{\text{ref}}$, (2)

which does not depend on the variable ranges $\Delta E_F$ and $\Delta \mu_a$ of the Fermi level and chemical potentials.

The notorious bandgap problem in DFT leads to an ambiguity of $\Delta H$ for charged defects through the electronic reference energy $E_{\text{VBM}}^{\text{ref}}$. Its underlying principles, DFT is constructed to give the correct charge density (and, hence, the correct electrostatic potential), as well as the correct total energy. However, the single-particle (Kohn-Sham) energies do not have the meaning of quasiparticle (electron-removal or electron-addition) energies, and LDA or GGA calculations underestimate the band gap, often by more than 50%. Note that an exact DFT functional would predict the correct band gap from the total energy difference $E_g = E(N+1) + E(N-1) - 2E(N)$, irrespective of the magnitude of the single particle gap. In LDA or GGA, however, the total-energy and single-particle-energy gaps must coincide due to the absence of a discontinuity in these exchange-correlation potentials. While the closed-shell energy $E(N)$ is not directly affected by the band gap problem, the open-shell energies $E(N-1)$ and $E(N+1)$ are, which is recently being discussed in the context of the “delocalization error”. Thus, DFT approximations should describe rather accurately also the formation energies of fully ionized (closed-shell) defects, relative to the average electrostatic potential $V_{av}$, which defines the energy zero in periodic plane-wave methods. As illustrated graphically in Fig. 1, an accurate prediction of the charged-defect formation energy should therefore be accessible by correcting the charged-defect formation energy using the $GW$ quasiparticle energy shift $\delta E_{\text{VBM}} = e_{\text{VBM}}^{GW} - e_{\text{VBM}}^{\text{ref}}$, $\Delta H_{\text{ref}}^{GW} = \Delta H_{\text{ref}} + \delta E_{\text{VBM}}^{GW}$.

The need to attribute the bandgap correction to the VBM and the conduction band minimum (CBM) in a certain proportion was recognized before, but there has been so far little clarity how to appropriately determine the individual band edge shifts for charged-defect formation energies. In this paper, $\Delta H$ of dopants and defects in their fully ionized charge state is anchored at the average potential $V_{av}$, and the $GW$ quasiparticle energies are used to determine the appropriate VBM reference energy. In the fully ionized state there are no occupied defect levels (donor states populated by electrons or acceptor states populated by holes), which could require a post-DFT treatment with a contribution to the formation energy. The present approach is fully compatible with the use of the same band edge shifts for band offsets and the positioning of deep level transition energies (proposed that these levels are indeed invariant relative to $V_{av}$ as found for a variety of defects in Ref. 24). Conceptually also related is the work of Ref. 4, where the vacuum level was used as a reference for the band edge energies in the context of VBM calculations. Noting that the charged-defect formation energies do not explicitly depend on the potential step at the semiconductor-vacuum interface, we agree that integration into a common picture is possible since the abovementioned DFT principles imply a consistent description of both defect energies and the potential step.

In a similar way as $E_{\text{VBM}}^{\text{ref}}$ defines the reference energy for the electron reservoir (Fermi level), the elemental energies $\mu_a^{\text{ref}}$ define the reference energy for the atomic reservoir [cf. Eqs. (1)–(2)]. At the same time, the $\mu_a^{\text{ref}}$ enter the calculation of the compound formation enthalpy $\Delta H_f^C = E_f^{\text{ref}} + \sum a m_a \mu_a^{\text{ref}}$, (4)

which is needed to determine the thermodynamic limits of the chemical potentials $\Delta \mu_a$ [cf. Eq. (1)]. Here, $E_f^{\text{ref}}$ is the total energy of the compound consisting of $m_a$ atoms of type $a$ per formula unit. Standard LDA or GGA calculations have been found to predict $\Delta H_f^C$ with less than desirable accuracy, but sufficiently large to merit the development of approaches for correction. The origin of these inaccuracies has been traced back to incomplete error cancellation between chemically dissimilar
systems (e.g. compounds, metals, molecules).\textsuperscript{20} The fitted elemental-phase reference energies $\mu_a^\text{FERE}$ were introduced in Refs. 20 and 21 to determine atomic reference energies that are compatible with the total energies calculated for the compounds in a given functional. The shifts $\delta \mu_a = \mu_a^\text{FERE} - \mu_a^\text{DFT}$ of the FERE relative to the directly calculated elemental energies in DFT are determined by the least-square solution of the set of linear equations

$$\sum_a m_a \delta \mu_a = \Delta H_f^\text{DFT} - \Delta H_f^\text{EX},$$

(5)

for a set of compounds that can be formed from a set of elements $\{a\}$. Improved estimates for the exact formation enthalpies $\Delta H_f^\text{EX}$, needed to solve Eq. (5) for $\delta \mu_a$, could be determined from appropriate post-DFT total energy methods, such as quantum Monte Carlo,\textsuperscript{34} or the random-phase approximation.\textsuperscript{35,36} In this paper, we use instead the available tabulated experimental data\textsuperscript{37} for binary and ternary nitrides and oxides of Cu, Al, and Zn. We note that the resulting elemental reference energies $\mu_a^\text{REF} = \mu_a^\text{FERE}$ do not depend on the DFT energies of the elemental phases.\textsuperscript{20} Therefore, the abovementioned ambiguity, concerning the use of parameters ($a$ in hybrid functionals, $U$ in GGA(+$U$)) for materials other than those they have been adjusted for, is removed. The final result for the defect formation energies with correction of both the electronic and atomic reference energies reads

$$\Delta H_{\text{ref}}^\text{GW,FERE} = \Delta H_{\text{ref}}^\text{DFT} + q \delta E_{\text{VBM}} + \sum_a n_a \delta \mu_a.$$

(6)

III. GW QUASIPARTICLE ENERGY CALCULATIONS

The present calculations were performed using the projector augmented wave (PAW)\textsuperscript{38} implementation of DFT and GW in the VASP code.\textsuperscript{9,30,40} For the DFT calculations, we use the GGA(+$U$)$^{31,42}$ ($U = 5$ and 6 eV for Cu-$d$ and Zn-$d$, respectively) and the HSE$^{12}$ ($\alpha = 0.25$) functionals. The GW calculations follow the scheme applied in the previous work of Refs. 25, 27, and 43. The GGA + $U$ or HSE wave functions are maintained to keep the same average potential $V_{av}$ as reference for the single- (quasi-) particle energies. The tendency of underestimated $d$-band energies is compensated by an attractive on-site potential for Zn-$d$ ($V_d = -1.5$ eV (Ref. 25)) and Cu-$d$ ($V_d = -2.4$ eV (Ref. 43)). Further details and the computational parameters are given in the Appendix. As shown in Table I, the GW band gaps obtained with GGA(+$U$) and HSE wave functions differ by less than 0.3 eV for all cases.

IV. FITTED ELEMENTAL REFERENCE ENERGIES

Table II gives the shifts $\delta \mu_a$ of the elemental reference energies obtained from the least-squares fit of Eq. (5), and Table III shows the compound formation enthalpies $\Delta H_f$ before and after the FERE correction. The root-mean-square (RMS) deviation from the experimental data is 0.28 and 0.22 eV/atom in the direct GGA(+$U$) and HSE calculations, respectively, indicating that the thermochemistry is only slightly described better in HSE than in GGA(+$U$). With the $\mu_a^\text{FERE}$ reference energies, the deviation reduced to 0.02 and 0.01 eV/atom, respectively, which suggests that either functional describes excellently the relative energies between the different oxide and nitride compounds. The absolute $\mu_a^\text{FERE}$ energies (see Appendix) for GGA(+$U$) agree within less than 0.1 eV with previous results\textsuperscript{20,21} that were obtained by fitting of a larger set of elements and compounds. Table II further reveals that the FERE shifts in the two functionals are almost identical for Cu, N, and O, but differ significantly for Zn and Al.

V. DEFECT FORMATION ENERGIES

For the calculation of the formation energies of the intrinsic and extrinsic defects, we are focusing here only on the important case of fully ionized defects or dopants, which usually define the defect equilibrium and, therefore, determine the net doping and the carrier densities. These highest stable charge states also exhibit the strongest dependence on the band edge shift $\delta E_{\text{VBM}}$ [cf. Eq. (6) and Fig. 1]. Analogous corrections for the finite-size effects\textsuperscript{1-3,44} have been applied in the case of either functional. The formation energies of all other charge states follow from the charge transition levels inside the band gap. For shallow defects, the ionization energy depends largely on the effective mass and the dielectric constant and, therefore, is not directly affected by the DFT bandgap error. In this case, the transition level can be easily determined from the ionization energy relative to the respective band edge. When

TABLE II. The shifts $\delta \mu_a$ of the elemental reference energies $\mu_a^\text{FERE} - \mu_a^\text{DFT}$, given in electronvolts, relative to the directly calculated elemental energies in GGA(+$U$) and HSE.

<table>
<thead>
<tr>
<th></th>
<th>Cu</th>
<th>Zn</th>
<th>Al</th>
<th>N</th>
<th>O</th>
</tr>
</thead>
<tbody>
<tr>
<td>GGA(+$U$)</td>
<td>-0.02</td>
<td>-0.08</td>
<td>+0.66</td>
<td>-0.03</td>
<td>+0.34</td>
</tr>
<tr>
<td>HSE</td>
<td>±0.00</td>
<td>+0.24</td>
<td>+0.35</td>
<td>-0.04</td>
<td>+0.32</td>
</tr>
</tbody>
</table>
TABLE III. The calculated compound formation enthalpies (electronvolts per formula unit) in GGA(+ U) and HSE, compared to experiment.27

<table>
<thead>
<tr>
<th>Compound</th>
<th>ΔH_f (experiment)</th>
<th>ΔH_f (DFT)</th>
<th>ΔH_f (FERE)</th>
<th>ΔH_f (DFT)</th>
<th>ΔH_f (FERE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cu2N</td>
<td>+0.77</td>
<td>+0.80</td>
<td>+0.89</td>
<td>+0.80</td>
<td>+0.85</td>
</tr>
<tr>
<td>ZnN2</td>
<td>−0.23</td>
<td>−0.48</td>
<td>−0.18</td>
<td>+0.39</td>
<td>−0.26</td>
</tr>
<tr>
<td>AlN</td>
<td>−3.30</td>
<td>−2.73</td>
<td>−3.37</td>
<td>−3.01</td>
<td>−3.32</td>
</tr>
<tr>
<td>CuO</td>
<td>−1.63</td>
<td>−1.37</td>
<td>−1.69</td>
<td>−1.36</td>
<td>−1.68</td>
</tr>
<tr>
<td>Cu2O</td>
<td>−1.75</td>
<td>−1.49</td>
<td>−1.80</td>
<td>−1.47</td>
<td>−1.77</td>
</tr>
<tr>
<td>ZnO</td>
<td>−3.61</td>
<td>−3.38</td>
<td>−3.64</td>
<td>−3.04</td>
<td>−3.60</td>
</tr>
<tr>
<td>Al2O3</td>
<td>−17.37</td>
<td>−14.90</td>
<td>−17.26</td>
<td>−15.68</td>
<td>−17.32</td>
</tr>
<tr>
<td>Al2CuO4</td>
<td>−18.79</td>
<td>−16.10</td>
<td>−18.77</td>
<td>−16.87</td>
<td>−18.82</td>
</tr>
<tr>
<td>CuAlO2</td>
<td>−9.71</td>
<td>−8.33</td>
<td>−9.66</td>
<td>−8.71</td>
<td>−9.68</td>
</tr>
<tr>
<td>Al2ZnO4</td>
<td>−21.40</td>
<td>−18.73</td>
<td>−21.35</td>
<td>−19.19</td>
<td>−21.39</td>
</tr>
</tbody>
</table>

Figure 2 shows the differences of the defect formation energies between HSE and the GGA(+ U) predictions before and after taking into account the GW and FERE corrections for the electronic and atomic reference energies, respectively. While the uncorrected values exhibit large differences with a RMS of 1.53 eV, the GW corrections bring the results of the two functionals into much closer agreement with an RMS of 0.46 eV, and the FERE corrections afford a further reduction of the difference to 0.34 eV. Since only the FERE shifts of Zn and Al exhibit a significant difference between GGA(+ U) and HSE (cf. Table II), the most pronounced change due to the FERE is observed in Fig. 2 for the cation substitutional dopants AlZn and ZnAl (cat-S), where almost perfect agreement between the two functionals is achieved for ΔH_{ref}^{GW, FERE}. The largest remaining deviations occur for the vacancy defects, indicating that this type of defect experiences the largest total-energy contribution [first term in Eq. (2)] due to the inclusion of Fock exchange in the hybrid functional. The complete list of defect formation energies is given in Table IV. As expected, we observe that the magnitude of the GW corrections qδE_{VBM} is much larger in GGA(+ U) (RMS = 2.32 eV) than in HSE (RMS = 0.93 eV), whereas the FERE corrections ∑nδμ_a are only somewhat smaller in HSE (RMS = 0.42 and 0.25, respectively). While the adjustment of the α parameter in the HSE functional to match the correct band gap would further reduce the magnitude of the GW band edge shifts and thereby improve the predictions over the standard α = 0.25 value, the present approach removes the above noted ambiguities associated with such a materials-specific adjustment.

FIG. 2. (Color online) Difference of defect formation energies between the HSE and the GGA(+ U) functionals, based on (a) ΔH_{ref}, directly calculated in the respective DFT functional, on (b) ΔH_{ref}^{GW} taking into account GW band edge shifts, and on (c) ΔH_{ref}^{GW, FERE}, taking additionally into account the FERE shifts for the elemental reference energies [cf. Eq. (6)]. The different types of defects are anion vacancies (an-V: V_N, V_O), anion substitutional dopants (an-S: O_N, N_O), cation vacancies (cat-V: V_Cu, V_Zn, V_Al), cation substitutional dopants (cat-S: AlZn, ZnAl), and cation interstitials (cat-I: C_u).
VI. CONCLUSIONS

In conclusion, by anchoring the charged-defect formation energy at the average electrostatic potential, we used the \( GW \) quasiparticle energy shifts of the band edge energies to address the bandgap problem in supercell defect calculations. Changes in the defect formation enthalpy of smaller but still significant magnitude result from the fitted elemental reference energies, which improve the thermochemical boundary conditions. Comparing results of density and hybrid functionals, we showed that both methods give very similar results for fully ionized defects when equivalent corrections are applied in either case. This finding suggests that accurate formation energies for the important case of fully ionized defects can be obtained even with standard GGA(\(+U\)) density functionals when using appropriate electronic and atomic reference energies. Existing results without such corrections can be easily reevaluated [see Eq. (6)]. For the hybrid functionals, this approach further provides an option to avoid using different exchange mixing parameters for different materials.

ACKNOWLEDGMENTS

Financial support: For NREL and CSM (H.P., V.S., and S.L.), the U.S. Department of Energy, Office of Energy Efficiency and Renewable Energy, Next Generation Photovoltaics II (SunShot initiative), under Contract No. DE-AC36-08GO28308 to NREL. For TCD (D.O.S and G.W.W.), the SFI through the PI program (PI Grant Nos. 06/IN.1/I92 and 06/IN.1/I92/EC07). UCL (D.O.S.), Ramsay Memorial Trust and UCL Ramsay Fellowship. High performance computing resources: Red Mesa by NREL’s Computing Science Center, Kelvin by TCHPC, the Stokes cluster by ICHEC, and HECToR through membership of the HPC Materials Chemistry Consortium under
Table V. The elemental energies \( \mu^{\text{DFT}} \) calculated in DFT and given for different PAW potentials of N and O along with the respective energy cutoff \( E_{\text{cut}} \).

<table>
<thead>
<tr>
<th>PAW potential</th>
<th>( E_{\text{cut}} ) (eV)</th>
<th>( \mu^{\text{GGA+U}} ) (eV)</th>
<th>( \mu^{\text{HSE}} ) (eV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cu ((U = 5 \text{ eV in GGA}))</td>
<td>320</td>
<td>−2.03</td>
<td>−4.08</td>
</tr>
<tr>
<td>Zn ((U = 6 \text{ eV in GGA}))</td>
<td>320</td>
<td>−0.56</td>
<td>−1.98</td>
</tr>
<tr>
<td>Al</td>
<td>320</td>
<td>−3.75</td>
<td>−4.22</td>
</tr>
<tr>
<td>N (soft)</td>
<td>320</td>
<td>−8.42</td>
<td>−11.04</td>
</tr>
<tr>
<td>N (standard)</td>
<td>500</td>
<td>−8.41</td>
<td>−11.55</td>
</tr>
<tr>
<td>O (soft)</td>
<td>320</td>
<td>−5.03</td>
<td>−7.95</td>
</tr>
<tr>
<td>O (standard)</td>
<td>500</td>
<td>−5.01</td>
<td>−8.63</td>
</tr>
</tbody>
</table>

TABLE V. The elemental energies \( \mu^{\text{DFT}} \) calculated in DFT and given for different PAW potentials of N and O along with the respective energy cutoff \( E_{\text{cut}} \).