PHYSICAL REVIEW B **78**, 245207 (2008)

Semiconductor thermochemistry in density functional calculations

Stephan Lany

National Renewable Energy Laboratory, Golden, Colorado 80401, USA

(Received 8 May 2008; revised manuscript received 30 September 2008; published 30 December 2008)

The local-density and generalized gradient approximations (LDA and GGA) to density functional theory (DFT) exhibit incomplete error cancellation when energy differences are taken between chemically dissimilar systems. This energy inconsistency is manifested, e.g., in the tendency to underestimate the heat (enthalpy) of formation of semiconducting and insulating compounds in LDA and, even more so, in GGA. Considering a set of 61 compounds that can be formed from 14 elements (cations: Cu, Mg, Ca, Zn, Cd, Al, Ga, and In; anions: N, P, As, O, S, and Se), optimized elemental reference energies are determined by least-squares error minimization of an overdetermined set of linear equations. These elemental energies are "optimally consistent" with the DFT energies of the semiconductor compounds and imply corrections of up to 1 eV compared to the respective LDA or GGA energies. While these "corrections" are not to be understood to yield the correct absolute total energies of the elements, they are proposed to give appropriate bounds for the chemical potentials for thermodynamic processes in semiconductors and insulators, such as, e.g., defect formation, surface reconstruction, or catalytic processes. The present model allows to evaluate thermodynamic processes using DFT energy differences taken only between systems that are expected to show good error cancellation.

DOI: 10.1103/PhysRevB.78.245207 PACS number(s): 71.15.Nc, 65.40.−b, 71.20.Nr

I. INTRODUCTION

The modeling of thermodynamic processes in the bulk or on the surface of semiconducting or insulating compounds, such as defect formation and doping,[1–3] redox reactions,[4,5] (nonstoichiometric) surface reconstruction,[6] and heterogeneous catalytic processes,[7] requires that the chemical potentials of the involved atomic species be defined. Most generally, these chemical potentials are bounded by the formation of the elemental phases, i.e., the metallic form of the cations (e.g., Al, Ga, Zn, or Cd) under the "metal-rich/anion-poor" condition, and the presence of the elemental form of the anions (e.g., elemental As, P, or gaseous N_2 or O_2) under the "metal-poor/anion-rich" condition. While the theoretical description of such processes relies to a large extent on common approximations to density functional theory (DFT), i.e., the local-density or generalized gradient approximations (LDA or GGA), these approximations yield very accurate results in particular for *energy differences* between chemically similar systems, whereas error cancellation is expected to be less perfect between chemically dissimilar systems.[4,8] For example, the difference between the *atomic* and *molecular* energies of O and N, i.e., the molecular binding energy, deviates significantly (on the order of 1 eV) from experiment in LDA and GGA.[8,9] Also, as demonstrated here for an extensive list of binary and ternary semiconducting or insulating compounds, both LDA and GGA tend to systematically underestimate the heat of formation of such compounds. Thus, while LDA and GGA are expected to describe rather consistently the relative energies within the materials class of compound semiconductors, the energies of the elemental reference phases, including metals and molecules, are much less consistent, leading to incomplete error cancellation in the calculation of thermochemical properties such as the heat of formation.

The purpose of the present work is to determine elemental reference energies for the cations A (A=Cu, Mg, Ca, Zn, Cd, Al, Ga, and In) and the anions B (B=N, P, As, O, S, and Se) that are optimally consistent with the DFT energies of their compounds. Here, the term "consistency" is understood as the tendency towards error cancellation when differences are evaluated between energies that individually have sizable absolute errors. In order to determine such optimized elemental reference energies, the LDA and GGA energies are calculated for an extensive set of 61 binary and ternary compounds, comprising the most important semiconducting and insulating compounds that are formed from the considered elements. The elemental reference energies are then obtained by solving a set of linear equations, using as input the DFT-calculated total energies of the compounds and the experimental heats of formation.[10–23] The elemental reference energies that are found to minimize the root-mean-square (rms) deviation from the experimental heats of formation imply corrections up to 1 eV compared to the directly calculated LDA or GGA energies of the respective elements. Further, in compounds containing metals with shallow d states it is often desirable to use the LDA+U or GGA+U methodology[24] to correct for residual self-interaction effects within the cation d shell, e.g., in systems containing transition metals,[4,25–27] or when shallow d states couple strongly to the valence band in semiconductors, as is the case in the photovoltaic chalcopyrites $CuInSe_2$ and $CuGaSe_2$,[28] in Cu_2O,[29,30] or in II-VI semiconductors such as ZnO.[31,32] Since, however, the appropriate values for the Coulomb parameter U are generally different in these compounds and in the metallic phase, but in practice only energies calculated with the same value for U can be compared,[5,26,33] the metallic elemental reference energies are generally undetermined in LDA+U or GGA+U calculations. By determining the elemental energies in the present work as the solution of a system of linear equations (SLE), without the need for direct calculation of the energies of the elements, appropriate chemical potentials for the elemental reference phases are obtained also for the GGA+U functional.

II. METHOD OF TOTAL-ENERGY CALCULATION

The total energies of solids and molecules are calculated in the present work within the momentum space pseudopotential formalism[34] employing the projector augmented wave method[35] as implemented in the VASP code.[36] The parametrizations of Refs. 37 and 38 are used for the LDA and GGA approximations to density functional theory, respectively. The energies are converged to better than 5 meV/atom, using an energy cutoff of up to 480 eV, as needed for accurate volume relaxation of the solids. A k-point sampling consisting of at least 2000 k points/atom^{-1} is used in case of semiconductors and insulators and denser meshes in case of the elemental metals. Initial geometries for the unit cells of the solids are taken from Ref. 39. The elemental phase of P is calculated using the structure of "black phosphorus." The energy of the standard phase "white P" is obtained by subtracting the heat of formation $\Delta H_f(\mathrm{P_{black}})=-0.41$ eV (Ref. 12) from the calculated energy of black P. In the calculations employing the GGA+U method [in the formulation according to Ref. 24], the Coulomb parameters are chosen as $U=6$ eV for Cu and Cd and 7 eV for Zn (using $J=1$ eV for the exchange parameter), in accord with previous works.[28–32] Since, the appropriate U parameter is in general system dependent and may change, in particular, in case of large variations in the oxidation state of a transition metal,[5] the choice of a single U parameter is often a compromise. In the compounds considered in the present work, however, the elements Cu, Zn, and Cd occur almost exclusively in a single oxidation state (d^{10} configuration). The only exception is the antiferromagnetic CuO (d^9), which turns out to be consistently described with the a single U parameter.

III. DETERMINATION OF OPTIMIZED ELEMENTAL REFERENCE ENERGIES

A. Formation enthalpy and bounds of chemical potentials

In order to describe theoretically thermodynamic processes in the bulk or at the surface of a solid, one generally needs to calculate the formation enthalpy of the system of interest (e.g., a particular surface reconstruction, an impurity in the bulk, etc.). The formation enthalpy ΔH_S of a solid-state system S, consisting of a set of atom α, can be written as[40]

$$\Delta H_S = E_S - \sum_{\alpha} n_\alpha(\mu_\alpha^{\mathrm{el}} + \Delta\mu_\alpha), \tag{1}$$

where E_S is the total energy of the system S, n_α is the number of atoms of type α in the system, and $\mu_\alpha=\mu_\alpha^{\mathrm{el}}+\Delta\mu_\alpha$ is the chemical potential of α, defined relative to the elemental reference phase having the chemical potential μ_α^{el} (total energy per atom of the elemental solid or diatomic molecule). In general, there exist stoichiometric compounds which can precipitate from system S, i.e., all the compounds C which can be formed by any subset of atoms β of the full set α. The heat of formation of such a compound is defined as

$$\Delta H_f(C) = E_C - \sum_{\beta} n_\beta \mu_\beta^{\mathrm{el}}, \tag{2}$$

where n_β is the number of atoms of type β per formula unit of the compound.

Under thermodynamic equilibrium conditions,[41] the chemical potentials μ_α, which determine the formation enthalpy of the system of interest [cf. Eq. (1)], are subject to the conditions

$$\mu_\alpha \le \mu_\alpha^{\mathrm{el}} \Leftrightarrow \Delta\mu_\alpha \le 0, \tag{3}$$

$$\textstyle\sum_\beta n_\beta \mu_\beta \le E_C \Leftrightarrow \displaystyle\sum_\beta n_\beta \Delta\mu_\beta \le \Delta H_f(C), \tag{4}$$

which are imposed by the possibility to precipitate the elemental phases of α and the compounds C, respectively. In Eqs. (3) and (4), the equality sign holds if the respective elemental phase or the compound actually exists in equilibrium. For illustration, consider the problem of oxygen doping in GaN, i.e., the formation of the $\mathrm{O_N}$ defect. Here, the system of interest is $S=\mathrm{GaN:O}$ (α=Ga, N, and O), and the stoichiometric compounds that can be formed from the set α are $C_1=\mathrm{GaN}$ (β_1=Ga and N) and $C_2=\mathrm{Ga_2O_3}$ (β_2=Ga and O). In this example, the equality sign holds in Eq. (4) for GaN, i.e., it is supposed that GaN exists in equilibrium, but the existence of $\mathrm{Ga_2O_3}$ is not required although the precipitation of $\mathrm{Ga_2O_3}$ will limit the solubility of oxygen in GaN.

B. Role of the elemental reference energies

For the simple case of a binary compound AB (A=cation and B=anion), without the introduction of foreign elements such as additional dopant atoms, Fig. 1 shows schematically the dependence of ΔH_S for a nonstoichiometric system S with $n_A-n_B=1$, e.g., the formation of an anion vacancy, of a cation interstitial, or of a nonstoichiometric surface reconstruction with excess cations. Considering the definition of the formation energy ΔH_S [Eq. (1)] and the bounds of the chemical potentials $\Delta\mu_A$ and $\Delta\mu_B$ according to Eqs. (3) and (4), ΔH_S is maximal when $\Delta\mu_B=0$ and $\Delta\mu_A=\Delta H_f(AB)$ (the anion-rich/cation-poor condition), and minimal when $\Delta\mu_A=0$ and $\Delta\mu_B=\Delta H_f(AB)$ (the cation-rich/anion-poor condition). As illustrated in Fig. 1, the formation enthalpies ΔH_S will change when the elemental reference energies are shifted by an offset $\delta\mu_A$ and $\delta\mu_B$. This is true even in the case $\delta\mu_A=-\delta\mu_B$, i.e., when the resulting heat of formation $\Delta H_f(AB)$ remains unchanged [cf. Eq. (2)]. Considering the example of GaN, this means that if there is an inconsistency in the DFT energies between the elemental energies (Ga and N) and the semiconductor compound (GaN), the range of possible formation energies will be affected by this inconsistency (see Sec. III G).

C. Elemental reference energies from LDA and GGA

When, as usually done, the energies of both the elemental phases and of the semiconductor compounds are determined from the same approximate functional (usually LDA or GGA), one hopes that errors in the total energies of either

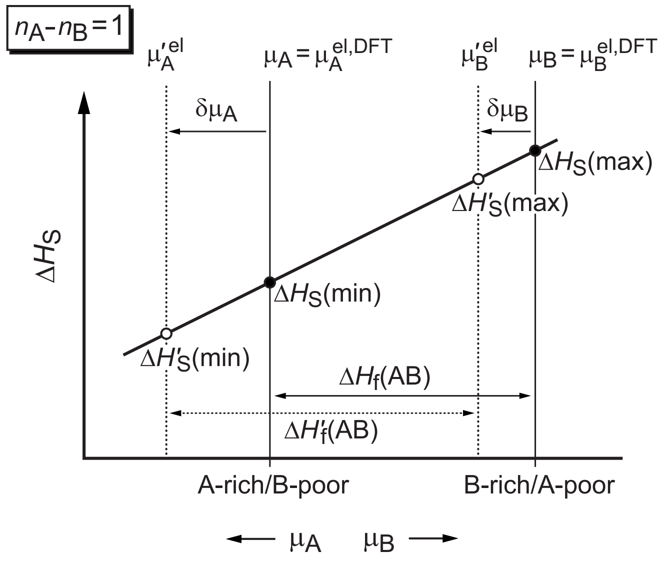

FIG. 1. Schematic illustration of the formation energy ΔH_S of a anion deficient structure of an AB compound (A=cation and B=anion), e.g., an anion vacancy, as a function of the A and B chemical potentials μ_A and μ_B, which are related to each other by Eq. (4) under equilibrium conditions. The range of possible formation energies ΔH_S is limited by the cation-rich and anion-rich conditions, i.e., $\mu_A = \mu_A^{el}$ ($\Delta\mu_A = 0$) and $\mu_B = \mu_B^{el}$ ($\Delta\mu_B = 0$), respectively. The allowed values for μ_A and μ_B span an energy corresponding to the heat of formation $\Delta H_f(AB)$. When corrections $\delta\mu_A$ and $\delta\mu_B$ are applied to the elemental reference energies μ_A^{el} and μ_B^{el}, e.g., $\mu'^{el}_A = \mu_A^{el,DFT} + \delta\mu_A$, the range of possible corrected formation energies $\Delta H'_S$ changes accordingly ($\delta\mu_A > 0$ and $\delta\mu_B < 0$ in the example shown).

calculation cancel to a large extent. In Table I, the theoretical heats of formation ΔH_f^{theor} per atom, calculated in LDA and GGA, are compared to the respective experimental values for an extensive list of 61 compounds. The rms (standard) deviation with respect to experiment is 0.18 eV/atom in LDA and 0.24 eV/atom in GGA, indicating that the error cancellation is not fully complete. Note that the numbers have to be multiplied by the number of atoms per formula unit to obtain the error in the conventional units for the heat of formation, e.g., $\Delta H_f^{theor}(\mathrm{ZnO}) = -2.91$ eV in GGA is 0.72 eV higher than the experimental value of -3.63 eV,[12] accounting for an error of $\delta H_f^{theor}/\mathrm{atom} = +0.36$ eV (see Table I). Moreover, LDA and, even more so, GGA tend to predict *systematically* too high values for ΔH_f^{theor}: the deviations of the calculated heats of formation from the respective experimental values do not statistically average out among the considered compounds, as seen in the (arithmetic) *mean deviations* of +0.08 eV/atom in LDA and +0.18 eV/atom in GGA. (See Table I; this trend has been observed before in a number of metal oxides in GGA.[4,5]) Beyond the errors that appear in the in the calculated heat of formation, less obvious but even larger errors can be present in the elemental reference energies μ_α^{el} because an offset of μ_α^{el} for cations and anions in opposite directions cancels in the calculation of ΔH_f [cf. Eq. (2)]. However, such errors in $\mu_\alpha^{el,DFT}$ would still affect in full extent the formation enthalpies ΔH_S for the system of interest [cf. Fig. 1 and Eq. (1)].

D. Elemental energies from the solution of a set of linear equations

The premise of the present work is that the error cancellation of approximate functionals is expected to be better within one class of materials (compound semiconductors) than among chemically different materials classes (molecules, metals, and semiconductors/insulators). Thus, using only the DFT energies E_C of the compounds, one can determine those energies μ_α^{el} of the elemental phases that minimize the rms error of the calculated heats of formation ΔH_f^{theor} compared to the respective experimental values ΔH_f^{expt} [cf. Eq. (2)]. These *elemental* energies μ_α^{el} can be said to be optimally consistent (in the sense defined above) with the respective DFT energies of the semiconducting or insulating *compounds*. In order to obtain such optimized elemental energies μ_α^{el} through solution of the least-squares problem, Eq. (2) is normalized to units of "energy per atom," i.e.,

$$\frac{\Delta H_f^{expt}(C)}{N} = \frac{1}{N}\left(E_C - \sum_\beta n_\beta \mu_\beta^{el}\right). \qquad (5)$$

Here, $N = \Sigma_\beta n_\beta$ is the total number of atoms per formula unit of the compound C. Having available the calculated total energies E_C and the experimental values of ΔH_f^{expt} (Ref. 42) for 61 different compounds C that can be formed from the 14 elements α considered in the present work, Eq. (5) represents an overdetermined SLE which can be solved by means of rms error minimization using standard algebraic routines. In this way, the elemental reference energies μ_α^{el} for the respective LDA, GGA, and GGA+U data sets (total energies calculated for the compounds) are determined without the need to directly calculate the total energies of the elements at all. It is convenient, however, to express the μ_α^{el} obtained from the solution of the SLE [Eq. (5)] in terms of a correction $\delta\mu_\alpha$ to the respective LDA or GGA energies of the elements, such that the corrected elemental chemical potential is obtained as, e.g., $\mu_\alpha^{el} = \mu_\alpha^{el,LDA} + \delta\mu_\alpha$ (cf. Fig. 1). The corrections $\delta\mu_\alpha$ are given in Table II. Note that in case of the data set GGA+U in Table II, the GGA+U methodology was actually applied only for the compounds containing Cu, Zn, or Cd, whereas the GGA energies are taken for the other compounds (which is equivalent to GGA+U, with $U = J = 0$). The corrections $\delta\mu_\alpha$ in GGA+U for Cu, Zn, and Cd (Table II) are given relative to the elemental energies determined with the same GGA+U parameters as used for the compounds.

E. Corrections to LDA and GGA energies of the elements

Table I gives the deviations $\delta H_f^{theor}/\mathrm{atom}$ of the theoretical compound heats of formation from experiment when evaluated with the elemental reference energies μ_α^{el} as determined from the solution of the SLE. Compared to the direct calculation of $\mu_\alpha^{el,DFT}$ in LDA or GGA, the rms deviation from the experimental ΔH_f is reduced by a factor of ~3, being only 0.07 eV/atom. A much smaller rms error cannot be expected in any event due to the uncertainty in the experimental values of ΔH_f [note, for example, that the long-time

TABLE I. The experimental heats of formation per atom, $\Delta H_f^{\mathrm{expt}}/\mathrm{atom}$, for 61 semiconducting and insulating compounds and the deviations $\delta H_f^{\mathrm{theor}}/\mathrm{atom}=(\Delta H_f^{\mathrm{theor}}-\Delta H_f^{\mathrm{expt}})/\mathrm{atom}$ of the theoretical heat of formation from experiment. The theoretical $\Delta H_f^{\mathrm{theor}}$ are determined with elemental energies μ_α^{el} that are either calculated directly in LDA and GGA (*direct*) or determined through solution of a SLE [Eq. (5)]. All numbers are in eV.

		$\delta H_f^{\mathrm{theor}}/\mathrm{atom}$				
	$\Delta H_f^{\mathrm{expt}}/\mathrm{atom}$	LDA *direct*	GGA *direct*	LDA *SLE*	GGA *SLE*	GGA+*U SLE*
Cu_3N	+0.19[a]	−0.01	0.10	0.00	−0.04	−0.01
Mg_3N_2	−0.96[b]	−0.10	0.16	−0.06	−0.10	−0.09
Ca_3N_2	−0.91[b]	−0.26	0.05	−0.02	−0.04	−0.04
Zn_3N_2	−0.05[b]	−0.08	0.16	−0.02	−0.02	−0.08
Cd_3N_2	+0.33[a]	0.05	0.21	0.10	0.08	0.05
AlN	−1.61[c]	−0.13	0.20	−0.06	−0.07	−0.06
GaN	−0.81[d]	0.00	0.32	0.09	0.09	0.11
InN	−0.10[e]	−0.25	0.18	−0.03	0.07	0.08
Cu_3P	−0.17[f]	0.01	0.03	−0.04	0.02	0.07
Mg_3P_2	−0.96[g]	0.10	0.10	0.06	0.04	0.03
Ca_3P_2	−1.22[h]	−0.17	−0.14	−0.01	−0.04	−0.06
ZnP_2	−0.21[b]	−0.31	−0.33	−0.01	−0.02	−0.01
Zn_3P_2	−0.33[b]	0.01	−0.03	−0.02	−0.02	−0.03
Cd_3P_2	−0.24[a]	0.07	−0.02	0.04	0.07	0.12
AlP	−0.85[b]	0.04	0.01	0.01	−0.01	−0.03
GaP	−0.53[b]	−0.09	−0.11	−0.11	−0.10	−0.10
InP	−0.39[b]	−0.03	−0.06	0.08	0.08	0.06
Mg_3As_2	−0.91[i]	0.26	0.27	0.12	0.11	0.08
Zn_3As_2	−0.28[b]	0.16	0.16	0.03	0.07	0.15
Cd_3As_2	−0.08[b]	0.06	−0.02	−0.08	−0.07	−0.03
AlAs	−0.61[b]	0.14	0.12	−0.03	−0.04	−0.07
GaAs	−0.37[j]	0.05	0.02	−0.10	−0.11	−0.13
InAs	−0.31[j]	0.10	0.06	0.08	0.06	0.03
CuO	−0.82[j]	−0.03	0.23	0.07	−0.01	−0.03
Cu_2O	−0.58[j]	0.04	0.17	0.00	−0.07	−0.05
MgO	−3.11[b]	−0.04	0.35	−0.06	−0.04	−0.01
CaO	−3.29[b]	−0.15	0.27	−0.01	0.01	0.04
ZnO	−1.81[b]	0.08	0.36	0.06	0.03	−0.07
CdO	−1.34[b]	0.05	0.30	0.03	0.01	−0.06
Al_2O_3	−3.47[b]	−0.03	0.43	−0.04	0.01	0.04
Ga_2O_3	−2.26[b]	0.00	0.40	0.00	0.00	0.04
In_2O_3	−1.92[b]	−0.04	0.31	0.06	0.00	0.04
$MgAl_2O_4$	−3.24[b]	−0.18	0.24	−0.20	−0.18	−0.15
$CaAl_2O_4$	−3.44[b]	0.06	0.49	0.09	0.11	0.14
$ZnAl_2O_4$	−3.07[b]	0.01	0.43	0.00	0.03	0.02
$CdGa_2O_4$	−2.01[b]	0.00	0.35	−0.01	−0.02	−0.01
CuS	−0.27[b]	0.01	0.04	−0.13	−0.11	−0.10
Cu_2S	−0.27[b]	0.29	0.32	0.10	0.14	0.09
MgS	−1.79[b]	0.24	0.32	−0.02	0.01	0.01
CaS	−2.45[b]	0.10	0.22	0.01	0.04	0.03
ZnS	−1.07[b]	0.24	0.23	−0.01	−0.02	0.01

TABLE I. (*Continued.*)

		$\delta H_f^{\text{theor}}$/atom				
	ΔH_f^{expt}/atom	LDA *direct*	GGA *direct*	LDA *SLE*	GGA *SLE*	GGA+*U* *SLE*
CdS	−0.78 [b]	0.17	0.11	−0.08	−0.10	−0.10
Al_2S_3	−1.50 [b]	0.41	0.40	0.12	0.08	0.07
Ga_2S_3	−1.07 [b]	0.37	0.37	0.09	0.07	0.07
GaS	−1.09 [b]	0.42	0.40	0.06	0.04	0.06
In_2S_3	−0.74 [b]	0.14	0.14	−0.04	−0.06	−0.06
InS	−0.70 [b]	0.13	0.16	−0.10	−0.08	−0.06
$CuSe_2$	−0.15 [a]	0.03	0.05	−0.02	0.02	0.03
CuSe	−0.21 [k]	0.11	0.07	0.00	−0.02	−0.01
Cu_3Se_2	−0.22 [k]	0.08	0.07	−0.07	−0.04	−0.02
Cu_2Se	−0.23 [k]	0.25	0.23	0.08	0.09	0.03
MgSe	−1.52 [l]	0.22	0.27	−0.01	0.03	0.02
ZnSe	−0.85 [b]	0.19	0.14	−0.04	−0.04	0.00
CdSe	−0.75 [b]	0.22	0.12	−0.01	−0.02	−0.01
Al_2Se_3	−1.18 [b]	0.36	0.33	0.10	0.09	0.07
Ga_2Se_3	−0.85 [b]	0.31	0.27	0.06	0.05	0.04
GaSe	−0.83 [b]	0.29	0.24	−0.05	−0.06	−0.05
In_2Se_3	−0.67 [b]	0.20	0.15	0.05	0.03	0.01
InSe	−0.62 [b]	0.14	0.09	−0.06	−0.08	−0.07
$CuGaSe_2$	−0.69 [m]	0.23	0.21	0.00	0.02	0.00
$CuInSe_2$	−0.53 [n]	0.12	0.08	−0.04	−0.05	−0.07
Mean deviation		0.08	0.18	0.00	0.00	0.00
rms (standard) deviation		0.18	0.24	0.07	0.07	0.07

[a] Reference 10.
[b] Reference 11.
[c] Reference 20.
[d] Reference 21.
[e] Reference 13.
[f] Reference 15.
[g] Reference 19.
[h] Reference 18.
[i] Reference 16.
[j] Reference 12.
[k] Reference 23.
[l] Reference 14.
[m] Reference 22.
[n] Reference 17.

tabulated value ΔH_f=−1.10 eV (Refs. 10–12) for GaN has recently been corrected to −1.62 eV (Ref. 21)]. Further, the systematic underestimation of the heat of formation, which is observed when the elemental reference energies $\mu_\alpha^{\text{el,DFT}}$ are directly calculated in LDA or GGA, is also removed due to the rms error minimization. It is interesting to note that after the optimization of the elemental energies, the remaining deviations of $\Delta H_f^{\text{theor}}$ from experiment are very similar in LDA

TABLE II. The optimized elemental reference energies which are obtained from solution of the system of linear equations [Eq. (5)] given as corrections $\delta\mu_\alpha$ (in eV) to the respective elemental energies per atom in LDA, GGA, or GGA+*U* (the "+*U*" methodology is applied only for the elements Cu, Zn, Cd, and their compounds).

	Cu	Mg	Ca	Zn	Cd	Al	Ga	In	N[a]	P	As	O[a]	S	Se
LDA	0.30	0.54	0.20	0.51	0.53	0.77	0.74	0.49	−0.92	−0.70	−0.44	−0.49	−0.03	−0.07
GGA	0.24	0.55	0.29	0.43	0.35	0.72	0.66	0.41	−0.20	−0.68	−0.41	0.23	0.06	−0.07
GGA+*U*	0.04	0.49	0.23	0.07	0.10	0.64	0.56	0.31	−0.13	−0.56	−0.26	0.23	0.13	0.01
Error bars	±0.09	±0.13	±0.14	±0.13	±0.13	±0.18	±0.16	±0.16	±0.18	±0.16	±0.18	±0.12	±0.13	±0.11

[a] Reference 40.

and GGA. Thus, LDA and GGA describe very similarly the thermochemical properties of the semiconductor and insulator phases, and the difference between the calculated heats of formation in LDA and GGA occur mostly due to differences in the LDA and GGA descriptions of the elemental phases. In cases where there remains a considerable deviation from experiment after the optimization, e.g., $\delta H_f^{\mathrm{theor}}/\mathrm{atom}=-0.2$ eV for $MgAl_2O_4$ (Table I), the discrepancy could also reflect an inaccuracy of the experimental data for the heat of formation.

As seen in Table II, the corrections $\delta\mu_\alpha$ obtained by solving the SLE have a magnitude up to about 1 eV, with the largest correction for molecular N_2 in LDA. The corrections are positive for cations but typically negative for anions, with the notable exception of $\delta\mu_{\mathrm{O}}=+0.2$ eV in GGA. Thus, there exists a partial cancellation of errors in the $\mu_\alpha^{\mathrm{el,DFT}}$ when ΔH_f is directly calculated from LDA or GGA. For example, $\Delta H_f(\mathrm{ZnO})=-3.48$ eV in LDA differs by only 0.08 eV/atom from experiment and gives the impression that LDA better describes the thermochemistry of ZnO than GGA which shows a 0.36 eV/atom difference. However, the better agreement with experiment in LDA results only from a fortuitous cancellation of the corrections for Zn metal and molecular oxygen, $\delta\mu_{\mathrm{Zn}}=+0.5$ eV and $\delta\mu_{\mathrm{O}}=-0.5$ eV in LDA. Indeed, the Zn correction $\delta\mu_{\mathrm{Zn}}=+0.4$ eV in GGA is very similar to that in LDA, and the difference results mostly from different corrections for molecular oxygen, $\delta\mu_{\mathrm{O}}=-0.5$ eV in LDA versus $\delta\mu_{\mathrm{O}}=+0.2$ eV in GGA. Thus, the result of the least-square solution indicates that the underestimated heat of formation of metal oxides in GGA should be corrected mostly in the energy of the metallic elements (Table II), thereby contrasting previous assumptions[4] that the error is entirely due to the O_2 molecule in case of non-transition-metal oxides. It is notable that the corrections $\delta\mu_\alpha$ for elements in the solid state are rather similar in LDA and GGA but differ strongly for the molecules N_2 and O_2. This finding bears out the general trend that GGA increasingly lowers the energy relative to LDA with increasing inhomogeneity of the electron density,[43] which is larger for molecules than for solids.

Regarding the notion of a correction for the elemental reference energies, it should be emphasized, however, that the $\delta\mu_\alpha$ do not represent corrections for the *absolute* LDA or GGA energies. For illustration, consider the binding energies of the N_2 and O_2 molecules. Since, the error in the LDA or GGA absolute energies[8,9,37] is much larger than the error in the respective binding energies, it is pointless to attribute the latter to either the atom or the molecule. Similarly, the underestimated ΔH_f of, e.g., metal oxides in GGA (cf. Table I and Refs. 4 and 5), cannot be attributed to an absolute energy error in any of the involved phases (i.e., O_2, the metal, or the oxide). In the present work, the elemental energies are treated as the variable to be optimized such that the experimental heats of formation are reproduced as well as possible when the respective compound energies are taken from the DFT calculation. Thus, the elemental reference energies are determined such to be optimally consistent with the thermochemical description of the semiconducting and insulating compounds in the respective approximation of the DFT functional. This approach requires, however, the consistency (in the above defined sense) of the calculated compound energies, a measure of which is given by the residual rms error of the least-square minimization of Eqs. (5).[44]

TABLE III. The formation energies (in eV) of the lattice vacancies in GaN, calculated in GGA, before and after applying the corrections $\delta\mu_{\mathrm{Ga}}$ and $\delta\mu_{\mathrm{N}}$ to the elemental reference energies. For V_{Ga}, the energy E_{CBM} is taken as $E_{\mathrm{VBM}}+3.55$ eV.

	$\Delta H(V_{\mathrm{N}}^{3+})$, $E_F=E_{\mathrm{VBM}}$		$\Delta H(V_{\mathrm{Ga}}^{3-})$, $E_F=E_{\mathrm{CBM}}$	
	Ga rich	N rich	Ga rich	N rich
Uncorrected	0.62	1.60	−0.47	−1.45
Corrected	−0.04	1.40	0.19	−1.25

F. GGA versus GGA+U data sets

Comparing the results in Table I obtained for the GGA and GGA+U data sets reveals no significant difference in residual rms errors between both data sets. Thus, within statistical significance (see Sec. III H), both functionals are found to describe the thermochemical properties of the compounds equally well, and the GGA and GGA+U compound energies are equally consistent. As seen in Table II, the corrections $\delta\mu_\alpha$ differ significantly between the GGA and GGA+U data sets only for Cu, Zn, and Cd, i.e., those elements on which a nonzero U parameter is applied (see also Sec. III D), whereas the $\delta\mu_\alpha$ of the other elements are hardly affected by GGA+U. Note that the observation of consistency of the GGA energies among the compound considered here contrasts the finding for transition-metal oxides with *partially filled* $3d$ shells,[4,27,45] where GGA energies exhibited inconsistency between different oxide stoichiometries, e.g., NiO and Ni_2O_3,[45] and where GGA+U dramatically improved the description of the relative stabilities between different oxidation states (despite the limitation to use a single U parameter). Thus, determining elemental reference energies for the $3d$ series with the present approach would necessitate the use of GGA+U.

G. Example: Carrier compensation in GaN due to vacancy formation

In order to illustrate an application of the corrected elemental reference energies, Table III shows the formation energies of the lattice vacancies in GaN, calculated in GGA, before and after this correction (the present defect calculations include careful treatment of supercell-size effects[46]). These defects are important for electrical doping of GaN, where the formation of the triply charged 3+ state of the N vacancy V_{N} limits p-type doping,[47] i.e., when the Fermi level E_F is close to the valence-band maximum (VBM), and where the 3− state of the Ga-vacancy limits n-type doping,[48] i.e., when E_F is close to the conduction-band minimum (CBM). As seen in Table III, the span of the formation energies $\Delta H(V_{\mathrm{N}})$ and $\Delta H(V_{\mathrm{Ga}})$ between the Ga-rich and N-rich conditions increases after the correction of the elemental energies (cf. also Fig. 1). Due to the dependence of the chemical potential of the N_2 molecule on temperature T and partial pressure P,[3,7] equilibrium conditions under typical growth conditions ($T \geq 1000$ K and $P \leq 1$ atm) are generally close to the Ga-rich limit. Before the correction, $\Delta H(V_N)$ remains well positive in this limit even when E_F lies at the VBM (Table III). After the correction, $\Delta H(V_N)$ becomes negative, indicating that hole-compensating N vacancies would spontaneously form when E_F comes close to the VBM. Thus, the correction of the elemental reference energies reconciles theory with the experimental finding[49] that successful hole doping through Mg acceptors requires growth under the addition of hydrogen—which reduces V_N formation by raising E_F—plus a postgrowth annealing (at a temperature high enough to remove H but low enough to prevent the equilibrium formation of V_N). Considering the opposite situation for n-type doping, ΔH of the electron-compensating Ga vacancy (at $E_F = E_{CBM}$) becomes positive only after the correction of the elemental energies, thus becoming consistent with the experimental fact that very high electron concentrations up to 10^{21} cm^{-3} can be achieved in GaN by Ge doping.[50]

H. Statistical analysis and error bars for the corrections $\delta\mu_\alpha$

In the limiting case of perfect accuracy of the experimental data for ΔH_f and perfect consistency of the DFT energies within the set of 61 compounds in Table I, the elemental reference energies from the solution of the SLE would reproduce exactly the experimental heats of formation.[44] This limit is not achieved, as evident from the residual rms error of 0.07 eV in the calculated $\Delta H_f^{\mathrm{theor}}$/atom (Table I). Due to the nonexact statistical nature of the data entering the SLE, the quality of the optimized reference energies μ_α^{el} increases with the degree of overdetermination of the SLE. Thus, in order to assess the statistical error bars (confidence limits) of the corrections $\delta\mu_\alpha$ for the present data sets (61 equations for 14 variables), a random error is added to the experimental values of $\Delta H_f^{\mathrm{expt}}$/atom, according to a Gaussian-type distribution. The distribution width $\sigma = 0.1$ eV is chosen such to reflect realistically a stochastic scatter in the experimental data and to be consistent with the residual rms error in $\Delta H_f^{\mathrm{theor}}$/atom after the least-squares minimization (Table I). The statistical error bars for the $\delta\mu_\alpha$, shown in Table II are then determined as the standard deviation of the $\delta\mu_\alpha$ determined from many (1000) such stochastic data sets of $\Delta H_f^{\mathrm{expt}}$. The small magnitude of 0.1–0.2 eV (Ref. 51) of the resulting error bars compared to the absolute magnitude of the $\delta\mu_\alpha$ up to 1 eV demonstrates the statistical significance of the proposed corrections for the LDA and GGA calculated elemental reference energies.

IV. CONCLUSIONS

The incomplete error cancellation between the DFT energies of compound semiconductors and their elemental constituents can be circumvented by determining the energies of the elemental phases through a rms error minimization of the heat of formation calculated from the DFT energies of the compounds. The obtained elemental energies can be said to be optimally consistent with the DFT energies of their compounds. The corresponding corrections for the elemental energies calculated in LDA and GGA are positive for metals (on the order of +0.5 eV) and typically negative for the elemental phases of the anions with the largest correction of −0.9 eV occurring for molecular nitrogen in LDA. The optimized elemental reference energies are proposed to serve as appropriate bounds for the atomic chemical potentials in the modeling of thermodynamic processes in semiconductors.

ACKNOWLEDGMENT

This work was funded by the U.S. Department of Energy, under Contract No. DE-AC36-08GO28308 through NREL's laboratory directed research and development program.

---

[1] F. A. Kröger, *The Chemistry of Imperfect Crystals* (North-Holland, Amsterdam, 1974).

[2] S. B. Zhang and J. E. Northrup, Phys. Rev. Lett. **67**, 2339 (1991).

[3] J. Osorio-Guillén, S. Lany, S. V. Barabash, and A. Zunger, Phys. Rev. Lett. **96**, 107203 (2006).

[4] L. Wang, T. Maxisch, and G. Ceder, Phys. Rev. B **73**, 195107 (2006).

[5] C. Franchini, R. Podloucky, J. Paier, M. Marsman, and G. Kresse, Phys. Rev. B **75**, 195128 (2007).

[6] J. E. Northrup and S. Froyen, Phys. Rev. Lett. **71**, 2276 (1993).

[7] K. Reuter and M. Scheffler, Phys. Rev. B **65**, 035406 (2001).

[8] R. O. Jones and O. Gunnarsson, Rev. Mod. Phys. **61**, 689 (1989).

[9] The N_2 molecule is overbound by 1.82 (0.50) eV in LDA (GGA), and the O_2 molecule is overbound by 2.28 (0.86) eV. Here, the O atom is calculated in the symmetry-broken solution; the zero-point energies of the molecules are subtracted from the experimental binding energy (Ref. 12). The error in the absolute energy of the N atom is 14.8 (4.0) eV, and that of the O atom is 15.8 (2.6) eV, as determined from a fully relativistic atomic calculation and experimental data from P. E. Cade and W. M. Huo, J. Chem. Phys. **47**, 614 (1967).

[10] D. D. Wagman, W. H. Evans, V. B. Parker, R. H. Schumm, I. Halow, S. M. Bailey, K. L. Churney, and R. L. Nutall, J. Phys. Chem. Ref. Data **11**, Suppl. 2 (1982).

[11] O. Kubaschewski, C. B. Alcock, and P. I. Spencer, *Materials Thermochemistry*, 6th ed. (Pergamon, Oxford, 1993).

[12] *CRC Handbook of Chemistry and Physics*, 60th ed. (CRC, Boca Raton, 1979); *CRC Handbook of Chemistry and Physics*, 87th ed. (Taylor & Francis, Boca Raton, 2007).

[13] H. Hahn and R. Juza, Z. Anorg. Chem. **244**, 111 (1940).

[14] L. J. Brillson, Phys. Rev. Lett. **40**, 260 (1978).

[15] S. P. Gordienko and G. Sh. Viksman, Poroshk. Metall. (Kiev) **24**, 82 (1985) [Sov. Powder Metall. Met. Ceram **24**, 573 (1985)].

[16] A. S. Alikhanyan, A. V. Steblevsky, V. A. Leontieva, M. Saidulaeva, and S. F. Marenkin, Izv. Akad. Nauk SSSR, Neorg. Mater.

21, 1271 (1985) [Inorg. Mater. **21**, 1116 (1985)].

[17] J. B. Mooney and R. H. Lamoreaux, Sol. Cells **16**, 211 (1986).

[18] A. I. Zaitsev, N. V. Korolyov, and B. M. Mogutnov, J. Chem. Thermodyn. **23**, 11 (1991); Zh. V. Dobrokhotova, A. I. Zaitsev, M. A. Zemchenko, A. D. Litvina, B. M. Mogutnov, and S. N. Yaschenko, J. Therm. Anal. Calorim. **38**, 1113 (1992).

[19] V. B. Chernogorenko and K. A. Lynchak, Fiz. Goreniya Vzryva **34**, 56 (1998) [Combust., Explos. Shock Waves **34**, 298 (1998)].

[20] J. M. McHale, A. Navrotsky, and F. J. DiSalvo, Chem. Mater. **11**, 1148 (1999).

[21] M. R. Ranade, F. Tessier, A. Navrotsky, V. J. Leppert, S. H. Risbud, F. J. DiSalvo, and C. M. Balkas, J. Phys. Chem. B **104**, 4060 (2000).

[22] N. Meyer, Ph.D. thesis, Freie Universität, Berlin, 2000.

[23] T. J. Anderson, O. D. Crisalle, S. S. Li, and P. H. Holloway, NREL Subcontractor Report No. NREL/SR-520-33997, 2001 (unpublished).

[24] (a) V. I. Anisimov, I. V. Solovyev, M. A. Korotin, M. T. Czyzyk, and G. A. Sawatzky, Phys. Rev. B **48**, 16929 (1993); (b) A. I. Liechtenstein, V. I. Anisimov, and J. Zaanen, *ibid.* **52**, R5467 (1995).

[25] O. Bengone, M. Alouani, P. Blöchl, and J. Hugel, Phys. Rev. B **62**, 16392 (2000).

[26] H. J. Kulik, M. Cococcioni, D. A. Scherlis, and N. Marzari, Phys. Rev. Lett. **97**, 103001 (2006).

[27] S. Lany, J. Osorio-Guillén, and A. Zunger, Phys. Rev. B **75**, 241203(R) (2007).

[28] S. Lany and A. Zunger, Phys. Rev. Lett. **93**, 156404 (2004).

[29] H. Raebiger, S. Lany, and A. Zunger, Phys. Rev. B **76**, 045209 (2007).

[30] A. Önsten, M. Månsson, T. Claesson, T. Muro, T. Matsushita, T. Nakamura, T. Kinoshita, U. O. Karlsson, and O. Tjernberg, Phys. Rev. B **76**, 115127 (2007).

[31] S. Lany and A. Zunger, Phys. Rev. B **72**, 035215 (2005).

[32] P. Erhart, K. Albe, and A. Klein, Phys. Rev. B **73**, 205203 (2006).

[33] G. Kresse and J. Furthmüller, Vienna *Ab initio* Simulation Package (guide) (http://cms.mpi.univie.ac.at/vasp/vasp/vasp.html).

[34] J. Ihm, A. Zunger, and M. L. Cohen, J. Phys. C **12**, 4409 (1979).

[35] P. E. Blöchl, Phys. Rev. B **50**, 17953 (1994).

[36] G. Kresse and D. Joubert, Phys. Rev. B **59**, 1758 (1999).

[37] J. P. Perdew and A. Zunger, Phys. Rev. B **23**, 5048 (1981).

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

[39] P. Villars and L. D. Calvert, *Pearson's Handbook of Crystallographic Data for Intermetallic Phases*, 2nd ed. (ASM international, Materials Park, 1991).

[40] Equations (1)–(5) are formulated for the $T=0$ K limit. Significant zero-point energies exist for the first row molecules N_2 and O_2 which are not included in the calculated total energies of the molecules. Thus, the corrections $\delta\mu_{\mathrm{N}}$ and $\delta\mu_{\mathrm{O}}$ in Table II implicitly include contributions of +0.15 and +0.10 eV, respectively, due to the zero-point energy.

[41] Note that for practical application at finite temperatures, additional enthalpy and entropy contributions may have to be considered, most importantly those associated with the ideal gas (cf. Refs. 3 and 7).

[42] For the experimental heats of formation, the values for standard conditions ($T=298$ K, $P=1$ atm) are used, which are more widely available than the $T=0$ K limits (cf. Ref. 40). Due to the small difference $\Delta H_f^{0\ \mathrm{K}} - \Delta H_f^{298\ \mathrm{K}}$ of typically less than 0.03 eV/atom, however, the resulting error is not significant.

[43] A. Zupan, K. Burke, M. Ernzerhof, and J. P. Perdew, J. Chem. Phys. **106**, 10184 (1997).

[44] Note that the rms error actually reflects both the "inconsistency" among the calculated compound energies and that among the experimental heats of formation, which is difficult to deconvolute.

[45] S. Lany, H. Raebiger, and A. Zunger, Phys. Rev. B **77**, 241201(R) (2008).

[46] S. Lany and A. Zunger, Phys. Rev. B **78**, 235104 (2008).

[47] S. Hautakangas, J. Oila, M. Alatalo, K. Saarinen, L. Liszkay, D. Seghier, and H. P. Gislason, Phys. Rev. Lett. **90**, 137402 (2003).

[48] J. Oila, J. Kivioja, V. Ranki, K. Saarinen, D. C. Look, R. J. Molnar, S. S. Park, S. K. Lee, and J. Y. Han, Appl. Phys. Lett. **82**, 3433 (2003).

[49] M. E. Zvanut, D. M. Matlock, R. L. Henry, D. Koleske, and A. Wickenden, J. Appl. Phys. **95**, 1884 (2004).

[50] P. R. Hageman, W. J. Schaff, J. Janinski, and Z. Liliental-Weber, J. Cryst. Growth **267**, 123 (2004).

[51] The different error bars for different elements reflects the fact that some elements, e.g., Cu, are more frequently represented in Eqs. (5) than others, e.g., As.