Error estimates for density-functional theory predictions of surface energy and work function

Sam De Waele,[1,2] Kurt Lejaeghere,[1] Michael Sluydts,[1,2] and Stefaan Cottenier[1,2]
[1]*Center for Molecular Modeling, Ghent University, 9052 Zwijnaarde, Belgium*
[2]*Department of Electrical Energy, Metals, Mechanical Constructions and Systems, Ghent University, 9052 Zwijnaarde, Belgium*
(Received 10 July 2016; revised manuscript received 19 September 2016; published 14 December 2016)

Density-functional theory (DFT) predictions of materials properties are becoming ever more widespread. With increased use comes the demand for estimates of the accuracy of DFT results. In view of the importance of reliable surface properties, this work calculates surface energies and work functions for a large and diverse test set of crystalline solids. They are compared to experimental values by performing a linear regression, which results in a measure of the predictable and material-specific error of the theoretical result. Two of the most prevalent functionals, the local density approximation (LDA) and the Perdew-Burke-Ernzerhof parametrization of the generalized gradient approximation (PBE-GGA), are evaluated and compared. Both LDA and GGA-PBE are found to yield accurate work functions with error bars below 0.3 eV, rivaling the experimental precision. LDA also provides satisfactory estimates for the surface energy with error bars smaller than 10%, but GGA-PBE significantly underestimates the surface energy for materials with a large correlation energy.

DOI: 10.1103/PhysRevB.94.235418

I. INTRODUCTION

Due to ever increasing hardware performance and algorithm efficiency, the capabilities of quantum-based computational materials physics continue to grow rapidly. Methods that solve the Schrödinger equation to improve the understanding of materials consequently gain ever more prominence as their reach in system size expands. Density-functional theory (DFT) [1,2] has become one of the most prevalent of these methods. One of the areas in which DFT is being applied, is the modeling of surfaces and interfaces of crystalline solids [3], with an increased emphasis on a computational screening to identify interesting materials without experimental input [4–6]. The usefulness of DFT simulations therefore depends on their *accuracy*, i.e., the deviation with respect to experimental results. A number of studies have attempted to assess the accuracy of DFT-predicted surface properties [7–18]. Most of these studies, however, only consider a limited set of test surfaces, which brings into question the transferability and statistical significance of the accuracy estimate. The present work uses a test set of elemental materials, spanning most of the periodic table, for a statistical analysis of the agreement between DFT-calculated and experimentally measured surface properties. It makes use of the framework established by Lejaeghere *et al.*, who estimated the accuracy of DFT predictions for structural, elastic, and thermal properties of crystalline solids [19,20]. In the same spirit, the objective is to characterize the DFT value as a predictor for the experimental result. This leads to an estimate of the accuracy of DFT-predicted surface properties, the identification of experimental outliers and a protocol to correct calculated quantities *a posteriori* for predictable deviations.

The surface properties considered in this benchmark study are the surface energy and work function of crystalline solids. The surface energy is the energy required to create a surface from the bulk crystal. It is an important thermodynamic quantity governing the equilibrium shape of monocrystals [21–23], brittle fracture [24], or the rate of sintering [25]. The work function is the minimum energy required to extract an electron from the bulk of a material to the surrounding vacuum. It is one of the principal quantities governing thermionic emission [26] and band bending at semiconductor contacts [27]. Because of their fundamental importance, many experimental data are available for both quantities.

Both the work function and surface energy can be derived directly from periodic DFT calculations. This is in contrast to the experimental difficulties encountered when measuring these properties (see Sec. II A). Theoretical predictions of work functions and surface energies therefore offer a convenient and complementary method to cumbersome experiments. Nevertheless, these predictions are not perfect, due to approximations that are meant to make the theory tractable. This work aims to provide a confidence interval for DFT-based surface properties, in order to correctly interpret them with respect to their experimental counterparts.

DFT surface calculations were performed for all group 1 through group 15 elemental materials lighter than Po, except for B, Mn, N, P, and the lanthanides. For N, P, the lanthanides, and materials in groups 16–18, there is very little to no experimental surface information available. A comparison between DFT and experiment is therefore not possible. B and Mn were not considered either, because they both have a very complex unit cell from which surface construction is computationally very cumbersome. For all surfaces both the LDA (in the Perdew-Zunger parametrization) [28] and PBE [29] exchange-correlation functionals were evaluated, as these are two of the most prevalent functionals in the field of solid-state DFT. Additionally, a comparison between both functionals provides insight into how DFT-calculated surface properties depend on the level of theory. In Sec. II the experimental data, details of the calculations and statistical concepts necessary for comparing theory and experiment, are presented. Section III discusses the result of this comparison by evaluating the predictable errors, analyzing the residual errors, and validating the derived correction protocol from theory to experiment.

II. METHODOLOGY

A. Experimental data

The definition of the surface energy is thermodynamically straightforward, but measuring the surface energy of a solid

is extremely difficult and prone to a large imprecision. The experimental challenge is underlined by the myriad of different methods that have been proposed to determine the surface energy. These vary from cleavage experiments [24] for brittle materials, the zero-creep method for more ductile materials [30,31], the analysis of the equilibrium shape of crystallites [22,23], or contact angle measurements of liquid droplets [32]; to less direct solutions such as measuring the elastic modulus [33], the electrical conductivity [34], or even the speed of acoustic waves [35] in submicrometer powder assemblies. These experiments often suffer from a low precision and inaccuracies due to surface contaminants, but more importantly from a benchmark point of view, each method is biased towards a certain class of materials. This inconsistency severely hampers a uniform quantitative comparison across the entire test set of materials evaluated in this work.

Because of the above-described difficulties involved in obtaining the surface free energy of a solid, Tyson and Miller [36] derived a semiempirical approach that uses the liquid surface tension, which can be measured much more accurately. Because the derivation applies to all elemental materials, a consistent database is obtained. Moreover, the transparency of the method by Tyson and Miller makes it possible to estimate a consistent experimental error bar for the surface energies they present, which is necessary for the comparison with DFT predictions. To motivate how the experimental error bars are determined, a short overview of the considerations of Tyson and Miller is necessary.

Tyson and Miller first derived the equality

$$\gamma_{SV} = \gamma_{SL} + \gamma_{LV}, \tag{1}$$

which is valid at melting temperature. γ_{SV} is the solid-vapor interface energy, which is a synonym for the surface free energy being sought for, γ_{SL} is the solid-liquid interface energy, and γ_{LV} the liquid-vapor interface energy (i.e., the surface tension). Although it ignores anisotropic effects of γ_{SL} or γ_{SV}, Eq. (1) adequately represents the average energetic tradeoffs involved in surface formation.

To obtain the relation between the known liquid surface tension γ_{LV} and the desired solid surface energy γ_{SV}, Tyson and Miller then introduced two new ratios, the product of which is the ratio of γ_{SL} and γ_{SV}:

$$\frac{\gamma_{SL}}{\gamma_{gb}}\frac{\gamma_{gb}}{\gamma_{SV}} = \frac{\gamma_{SL}}{\gamma_{SV}} = \alpha, \tag{2}$$

where γ_{gb} is the grain boundary energy for an average high-angle grain boundary. The ratio γ_{SL}/γ_{gb} was determined to be 0.45 ± 0.05 by extrapolation of dihedral angle data of liquid precipitates at grain boundaries in binary alloys [37]. The second ratio γ_{gb}/γ_{LV} was derived from measurements of dihedral angles at thermally etched grain boundaries and estimated to be 0.33 ± 0.04 (see Tyson and Miller). Combining Eqs. (1) and (2), one obtains

$$\frac{\gamma_{SV}}{\gamma_{LV}} = \frac{1}{1-\alpha}. \tag{3}$$

This results in $\gamma_{SV}/\gamma_{LV} = 1.18 \pm 0.03$ as an estimate for a general material. In principle, however, the ratio should be determined for every individual material. This suggests the error margin of ± 0.03 given by Tyson and Miller is somewhat optimistic. Other authors [36] proposed alternative values for the ratio γ_{SV}/γ_{LV} in the range 1.09–1.33, which means a somewhat larger uncertainty of about $\pm 10\%$ should be taken into account for the derived values of γ_{SV} at melting temperature.

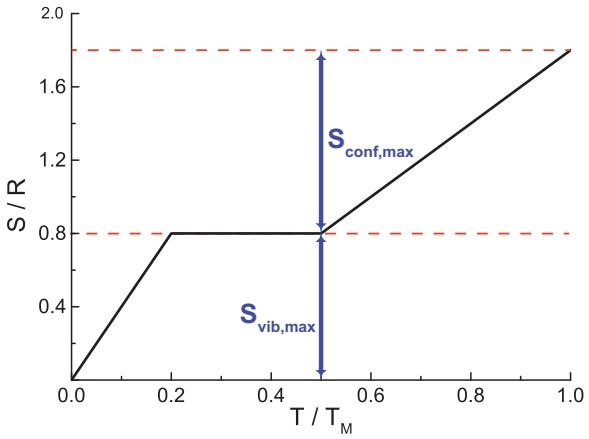

FIG. 1. The surface entropy S as a function of temperature in the simplified model of Tyson and Miller [36]. The first linear increase is due to the activation of surface-related vibrational contributions. The second linear increase is associated with surface roughening.

The conversion from the liquid-vapor to the solid-vapor interface energy is only applicable at the melting temperature T_m. To compare with DFT data, γ_{SV} must be transformed to a surface energy at 0 K. Two entropic contributions were singled out by Tyson and Miller as the most important factors for the temperature dependence of γ_{SV}: the vibrational entropy S_{vib}, and the configurational entropy S_{conf}. The former includes all vibrational modes associated with the surface, whereas the latter aims to describe the surface roughening observed when the material is heated. For both contributions a simplified approximation was proposed [38]. S_{vib} was assumed to increase linearly from 0 to 0.8 R (where R is the universal gas constant) between 0 K and the Debye temperature $T_D \approx 0.2T_m$. Surface roughening, on the other hand, becomes significant from $T = 0.5T_m$ to $T = T_m$ and causes a linear increase in S_{conf} from 0 to R in that temperature interval. The entire temperature-dependent evolution of the surface entropy is depicted in Fig. 1. If the effective [39] surface area A for a surface of general orientation is approximated as $A \approx 1.612N^{1/3}V_m^{2/3}$ [38], with N as Avogadro's number and V_m the molar volume, the temperature-dependent transformation is

$$\begin{aligned}\gamma_{SV}(0) &= \gamma_{SV}(T_m) - \int_{T_m}^{0} \frac{S(T)}{A} dT \\ &\approx \gamma_{SV}(T_m) + \frac{RT_m}{A}.\end{aligned} \tag{4}$$

The transformation of γ_{SV} from T_m to 0 K utilizes a rough approximation for the surface entropy. Other estimates of S_{vib} and S_{conf} suggest a large uncertainty of $\pm 50\%$ should be attached to the RT_m/A value [38]. However, as this term only contributes about 10% of the total surface energy ($RT_m/A \approx 0.1\gamma_{SV}$) [36], the dominant part of the uncertainty remains the conversion from γ_{LV} to γ_{SV}. The total error bar on

the experimental surface energy of a material i is then

$$\sigma_{i,\mathrm{expt}} = \sqrt{(0.1\gamma_{\mathrm{SV}})^2 + (0.5\,\mathrm{R}T_{\mathrm{m}}/A)^2} \approx 0.112\gamma_{\mathrm{SV}}. \quad (5)$$

It must be emphasized that the entire derivation of Tyson and Miller revolves around *averages*. A material-independent conversion from γ_{LV} to γ_{SV} is applied and the temperature effect is estimated by employing a simplified model for the entropy. The only material-specific properties used are surface tension at the melting temperature $\gamma_{\mathrm{LV}}(T_{\mathrm{m}})$ and molar volume V_{m}. This means no crystallographic or anisotropic information is present. Moreover, many materials undergo a phase transformation between 0 K and T_{m}. This begs the question which phase should be associated with the data by Tyson and Miller. A decisive argument to identify the most appropriate surface is the ratio $\gamma_{\mathrm{gb}}/\gamma_{\mathrm{SV}}$. As measurements of this ratio are performed at room temperature, the preferred crystallographic structures are those stable at room temperature. Moreover, work functions are typically measured for those room temperature allotropes as well. Using these when comparing both experimental surface energies and experimental work functions to the DFT-predicted values (see Sec. III), serves to maintain consistency in this accuracy assessment.

Work function measurements are more precise than surface energy measurements. The main source of inaccuracies lies in the presence of surface defects. Adsorbed impurities or structural surface defects can strongly alter the measured work function. For the elemental materials, the work function data compiled by Michaelson [40] forms the largest reference set. However, a rigorous discussion of experimental precision is lacking. A review encompassing fewer elemental materials but a larger amount of data was more recently presented by Kawano [41]. Because more experimental results are available for each material, it is possible to determine reliable experimental error bars for the work functions reviewed by Kawano. The standard deviations on sets of experiments for the same surface are at most 0.18 and 0.25 eV for anisotropic and polycrystalline samples, respectively. For materials with limited experimental data, a safe value of 0.32 eV was applied. There are a number of materials (As, Ba, Bi, Cd, Hg, Mg, Sn, Tl, Zn, and Zr) for which a work function was given by Michaelson, but the data were judged to be of unknown reliability. The accuracy with which the work functions of these materials is predicted by DFT is evaluated, but they are not incorporated in the statistical analysis.

B. DFT calculations for surfaces

A crystalline solid is usually modeled computationally by applying three-dimensional (3D) periodic boundary conditions. A surface of the material is created by inserting a vacuum region within a periodic cell. This essentially mimics a cleavage procedure and creates periodically repeated slabs of N atomic layers, which have two surfaces of an imposed crystallographic orientation (Fig. 2). The surface energy $E_{\mathrm{surf}} = \gamma A$ corresponds to the energy required to create a slab surface of area A. It is obtained from the DFT calculation by subtracting the bulk energy per layer (E_{bulk}) N times from the total slab energy (E_{slab}) and dividing by two to account for

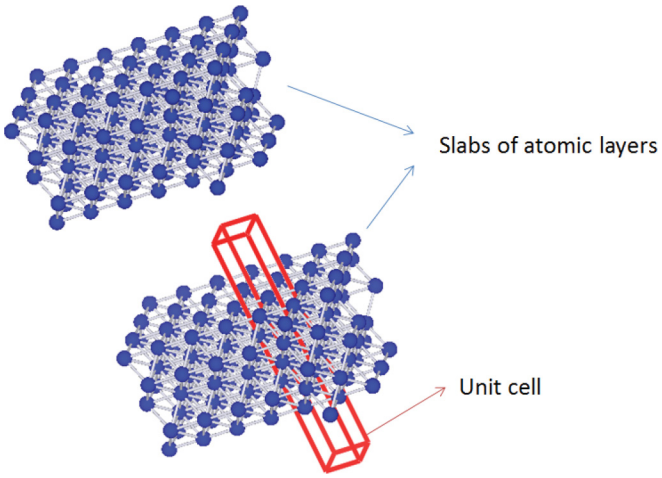

FIG. 2. To model a surface on the atomic scale in a 3D periodic code, slabs of atomic layers are constructed. As an example the {100} surface of a face-centered cubic material is shown. The slab unit cell is indicated by the red box.

the presence of two identical surfaces at both sides of the slab:

$$E_{\mathrm{surf}} = \tfrac{1}{2}(E_{\mathrm{slab}} - N \times E_{\mathrm{bulk}}) = \gamma A. \quad (6)$$

To differentiate the normalized surface energy from E_{surf}, it is usually indicated by γ (as was done in Sec. II A) and expressed in $\mathrm{J/m^2}$.

To correctly model a real surface, no interaction ought to be present between surfaces of neighboring slabs and the center of the slab should behave as bulk material. The first condition is fulfilled by making the vacuum region sufficiently large. The second condition is fulfilled when the work function and surface energy are converged with respect to slab thickness. Using Eq. (6) in this convergence test has a potential computational pitfall, however. Because the bulk and surface energies stem from different DFT calculations, possible numerical discrepancies, caused by different $\vec{k}$-point sampling for example, are enlarged by a factor N, the thickness of the slab [Eq. (6)]. A number of solutions have been proposed to avoid this divergence with respect to slab thickness. Boettger [42] proposed to derive the energy of a bulk layer by subtracting the energy of a slab with $N-1$ layers from that of a slab with N layers. Fiorentini and Methfessel [43] put forward the idea of performing a number of slab calculations and deriving the bulk energy from a linear fit to the energies. Another possible solution is to create a bulk unit cell which matches the orientation of the slab. This would allow for a perfect match in $\vec{k}$-point meshes.

By evaluating all of the above-described approaches to improve convergence of Eq. (6), it was concluded that none of them improved on simply performing a *very precise* bulk and slab calculation. Recently Singh-Miller *et al.* [14] came to the same conclusion when reviewing these different methods. The alternative methods are mainly useful when fine $\vec{k}$ meshing is unfeasible due to hardware limitations. When it is possible to apply highly converged sampling in reciprocal space, however, sticking to a simple bulk calculation is sufficient and avoids additional difficulties. It is, for example, not clear how thick the slab should be when using the Boettger method. Neither

is it easy to determine which thickness range is suitable for deriving the bulk energy from a linear fit, as per the method by Fiorentini and Methfessel.

A calculation of an atomic slab model is inherently limited to a surface of a specific orientation. How does one compare the corresponding anisotropic surface energy prediction by DFT with experimental data? Since Tyson and Miller presented their experimental data as valid for a surface of "general orientation," the ideal solution would be to calculate a number of different orientations sufficiently large to integrate over the two-dimensional orientation space. For materials with cubic symmetry, however, the {100}, {110}, and {111} surfaces actually suffice to perform the integration. By applying the terrace-ledges-kinks (TLK) description of a general surface [44], a ledge (or step) energy can be fitted to the transformation from one orientation to the other, and can subsequently be used to interpolate the surface energy to a general orientation. This approach produces good results for cubic materials because the above-mentioned simple surfaces already cover a total of 26 orientations. For materials with hexagonal, rhombohedral, or tetragonal symmetries a significantly larger number of DFT calculations is needed. Fitting the ledge energy is only valid for interpolating between two simulated orientations with a limited angular separation. Moreover, additional simulations involve more vicinal surfaces, requiring larger slab unit cells and thus a higher computational cost. For the sake of simplicity and consistency across different crystal structures, the basic arithmetic mean of three low-index surfaces was used (see further). Comparing this simple approach with the more accurate results from the TLK approach for the cubic materials, a decrease in surface energy of about -5.5% is observed. This shift is, however, almost completely material independent, ranging from -5.66% to -5.19%. For a comparison to experimental data, the near-constant deviation means the simple arithmetic mean is as valuable as the TLK approach (see Sec. II C).

Whereas the DFT surface energy is derived from the comparison of a bulk and a slab calculation [Eq. (6)], the work function can be directly obtained solely from the slab calculation. The work function is

$$\Phi = V_{\text{vac}} - E_{\text{F}}, \tag{7}$$

where E_{F} is the Fermi energy and V_{vac} corresponds to the local one-electron potential at a point in the vacuum where it becomes constant (i.e., where it is no longer affected by the presence of the surface). To obtain better convergence of the work function with respect to slab thickness, the method of macroscopic averages was proposed by Fall *et al.* [45]. In this method, the Fermi energy from a bulk calculation is used instead. However, for the simple nonpolar slabs in this work, their method does not result in a more reliable work function, provided the slab used for the calculation is sufficiently thick.

For many materials there is only a polycrystalline work function available, which corresponds to a surface containing different facets of varying crystallographic orientation. In a review of photoelectric experiments for obtaining work function data, Helander [46] concluded it is the facet with the lowest anisotropic work function which determines the measurement. This is confirmed by the review by Kawano, where there are both polycrystalline and anisotropic data available for 14 materials. The work functions for polycrystalline samples are on average only 90 meV higher than those of the surface orientation with the lowest work function.

The slab unit cells were constructed from optimized bulk geometries with the use of the ACONVASP software package and are provided in the Supplemental Material [47,48]. For the elemental materials with cubic crystal structures, the {100}, {110}, and {111} surfaces were taken into account. For the hcp and rhombohedral structures, the {0001}, $\{10\bar{1}0\}$, and $\{11\bar{2}0\}$ surfaces were simulated. The {100}, {110}, and {001} surfaces were selected for the tetragonal materials and the {100}, {010}, and {001} surfaces for orthorhombic Ga. The surface atoms were allowed to relax to acquire an optimized surface structure for all materials. The $(2 \times 1)\{111\}$ and the $(2 \times 1)\{100\}$ surface reconstructions for the diamond structures were also included. There are a number of possible reconstructions for the diamond structures, which were computationally evaluated by Stekolnikov *et al.* [49]. As the (2×1) reconstructions were always either the most stable or at most 6% higher in surface energy, these were the ones included in the present DFT calculations. For materials with more than one atom in the primitive unit cell, some surface orientations have more than one possible termination. In these cases, the energetically most favored termination—the one with the lowest surface energy—was applied to both ends of the slab. For Fe, Co, Ni, and Cr, magnetization of the surface was taken into account.

The Vienna *ab initio* simulation package (VASP) software [50–53] was used with VASP 5.2 recommended PAW potentials [54–56] for all LDA and PBE calculations. These potentials were recently shown to provide a similar precision as all-electron calculations [57]. Convergence tests were performed to determine the most appropriate meshing in reciprocal space, plane-wave cut-off energy, slab thickness, and number of surface layers allowed to relax. To efficiently determine the optimal settings, four electronically different materials were selected for each crystal structure (see Supplemental Material for convergence tests [48]). The most stringent settings required for this group of four materials were applied to all other materials with the same crystal structure. The numerical precision associated with these settings was 60 meV for work functions and 5% for surface energies in the worst cases, significantly lower than the experimental error bars discussed in Sec. II A. The plane-wave cut-off energy was set at 400 eV for all materials except for Li (800 eV) and C (600 eV). A convergence criterion of 10^{-8} eV was used for the electronic self-consistent cycle and of 10^{-6} eV for the structural optimization. The vacuum spacing between slabs was always chosen to be approximately 20 Å wide. The exchange and correlation contributions to the local one-electron potential were not taken into account to determine V_{vac}, as these converge to zero sufficiently far from the surface. During structure relaxations, Methfessel-Paxton smearing of first order [58] ($\sigma = 0.01$ eV) was used. During all other calculations, the tetrahedron method with Blöchl corrections was used [59] for all other calculations. All final slab geometries, settings, DFT results, and experimental data used for comparison are available in the Supplemental Material [48].

C. Statistical analysis

When presenting results of experimental measurements, it is common practice to provide confidence intervals. Guidelines for the determination of such intervals are summarized in the *Guide to the Expression of Uncertainty in Measurement* (GUM) [60]. Published computational results, on the other hand, are rarely accompanied by an error bar. Indeed, determining such an interval is not straightforward, as Civalleri *et al.* [61] illustrated by showing how the comparison of computational methods is affected by the chosen error measure. A formal framework proposed by Irikura *et al.* [62] aims to adhere to international standards as recommended by the GUM by removing the bias in the error. For a DFT-predicted materials property this entails selecting a representative test set and determining the average error. This predictable part of the DFT error can subsequently be distinguished from a material specific error and utilized to transform the calculated result to the expected true value, serving as an *a posteriori* calibration. This decomposition of the computational error has been applied to DFT results [19,20,63] and was recently reviewed by Pernot *et al.* [64]. Note that, whereas the present work discusses computational errors in terms of predictable and material-specific contributions, literature sometimes refers to systematic and random errors. Here the terms systematic and random are avoided, since all computational results are intrinsically deterministic.

To find the predictable error, the discrepancy between the DFT results and experimental data must be analyzed for trends. This boils down to a search for a general transfer function f which maps the computational results X_i onto the experimental values Y_i. The preferred option is a simple model for f with only a few parameters which are intuitively easy to understand. Because of the simple form of f and the complex nature of the errors, there will always remain some amount of scatter ϵ between the transformed DFT results $\hat{Y}_i = f(X_i)$ and the actual experimental results Y_i:

$$Y_i = f(X_i) + \epsilon_i = \hat{Y}_i + \epsilon_i. \tag{8}$$

The residual errors ϵ_i are the material-specific part of the error. They represent a true measure of the inadequacy of the method, as they cannot be removed by correcting for an overall bias. If ϵ_i were zero for all materials i, there would be a perfect mapping between DFT and experiment. This is why an average of three different anisotropic DFT-predicted surface energies is equally valuable as an isotropic surface energy derived by the more accurate TLK approach, as discussed in Sec. II B. The discrepancy between those two methods is entirely material independent and thus predictable. Consequently, the residual errors ϵ_i are the same for both methods.

Determining $\hat{Y} = f(X)$ is most conveniently achieved by proposing a function with p parameters $(\beta_0, \beta_1, \ldots, \beta_p) = \vec{\beta}$ and subsequently minimizing the sum of squares of the residual errors (SSR). For a zero-centered normal distribution of the residual errors ϵ_i, this yields the maximum likelihood estimate (MLE) for $\vec{\beta}$ [65],

$$\mathrm{SSR} = \sum_{i=1}^{n} (\hat{Y}_i - Y_i)^2 = \sum_{i=1}^{n} [f(X_i, \vec{\beta}) - Y_i]^2. \tag{9}$$

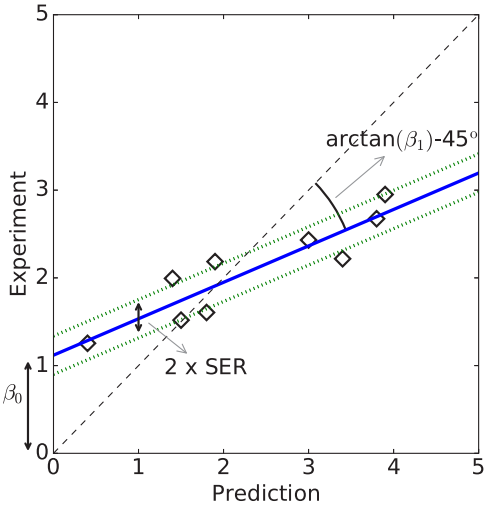

FIG. 3. An illustration of the meaning of the SER. It is a variable, residual error (green dotted line) on top of a predictable deviation (solid blue line).

The SSR is a metric for the total absolute deviation from the proposed model $f(X, \vec{\beta})$. Divided by the statistical degrees of freedom $n_{\mathrm{df}} = n - p$, it also forms an unbiased estimator for the variance σ_ϵ^2 on the ϵ_i. Its square root is referred to as the standard error on the regression (SER) [65] (Fig. 3):

$$\mathrm{SER} = \hat{\sigma_\epsilon} = \sqrt{\frac{\mathrm{SSR}}{n - p}}. \tag{10}$$

When there is an additional uncertainty on the experimental values Y_i, the residual error in Eq. (8) decomposes into $\epsilon_i = \epsilon_{i,\mathrm{DFT}} - \epsilon_{i,\mathrm{expt}}$. If all experimental values are equally precise, Eq. (9) and the fitting procedure can be used unaltered. If, on the other hand, $\epsilon_{i,\mathrm{expt}}$ differs for different X_i the MLE metric to obtain $\vec{\beta}$ slightly changes [65]:

$$\mathrm{SSR} = \sum_{i=1}^{n} \left(\frac{f(X_i, \vec{\beta}) - Y_i}{\sigma_{i,\mathrm{expt}}} \right)^2 = \sum_{i=1}^{n} w_i [f(X_i, \vec{\beta}) - Y_i]^2. \tag{11}$$

The added weights $w_i = 1/\sigma_{i,\mathrm{expt}}^2$ warp the SSR metric, emphasizing more precise experimental data. This in turn alters the interpretation of the SER. Multiplied by the local experimental deviation $\sigma_{i,\mathrm{expt}}$, it is now an estimate of the local residual error ϵ_i. Because the experimental imprecision is likely uncorrelated with shortcomings of the DFT model, one can write

$$\hat{\sigma}_{\epsilon,i}^2 = \sigma_{i,\mathrm{exp}}^2 \mathrm{SER}^2 = \sigma_{i,\mathrm{expt}}^2 + \hat{\sigma}_{i,\mathrm{DFT}}^2. \tag{12}$$

Note how Eq. (12) implies that the SER can never be expected to be smaller than one, as deviations caused by experiment are always present. The SER thus becomes a measure of the relative contribution of the DFT model to the total uncertainty. This confirms the importance of the discussion on experimental errors in Sec. II A. This is not only true for the quantities discussed in this work. Recently, Kirklin *et al.* [66] concluded that the mean absolute error on experimental formation energies was of comparable size

(0.082 eV/atom) to the average absolute difference between DFT and experiment (0.096 eV/atom).

By determining the parameters $\vec{\beta}$, a protocol can be created to transform any DFT-predicted value to the expected experimental one. The SER $\times\, \sigma_{i,\mathrm{expt}}$ obtained via $\vec{\beta}$ then serves as an error bar on this prediction. However, there are a number of important assumptions involved in the application of the least-squares method and the subsequent residual error analysis. First, the value of the derived function $f(X)$ depends on the soundness of the experimental data and the representativity of the test set for a general material. The meticulous determination of material-dependent experimental error bars (Sec. II A) and the large and varied test set should address those concerns. Second, all derived quantities in this section assume a zero-centered normal distribution for the residual errors $(\hat{Y}_i - Y_i)/\sigma_i$. By analyzing the residual errors, this assumption can be evaluated (Sec. III).

III. RESULTS

A. The character of the predictable error

In Sec. II C the transfer function f between theory and experiment, which contains the predictable error, was discussed for a general situation. To determine the most suitable form of f, Bayesian model selection (BMS) was used. The probabilities of different polynomial degrees were compared, by evaluating the model probability as described by Mana *et al.* [67], and are presented in Fig. 4. For the work function data, a linear form $f(X_i) = \hat{Y}_i = \beta_0 + \beta_1 X_i$ is the most likely model. For the surface energy data, on the other hand, adding a quadratic degree of freedom is an equally likely possibility. The possibility of a quadratic dependence is, however, chiefly caused by a different error behavior for some materials classes. The residual errors for the group 1 and group 2 materials are distinctly different from the other materials. Combined with their low surface energies (see Fig. 6), this causes the apparent quadratic dependence. When separate BMS analyses are performed, one for the 10 group 1 and group 2 materials and one for the 33 other materials, the linear form turns out to be the more probable model except for the PBE surface energies for the group 1 and group 2 materials, shown with the red columns in Fig. 4(d). In this case, the influence of the aberrant behavior of Be heavily influences the small group of only ten materials. Because the error on the DFT surface energy for a *general* material is not actually quadratic, but rather consists of two linear contributions, no quadratic term was added to f. This approach serves to make the statistical analysis on the current test transferable to a general material.

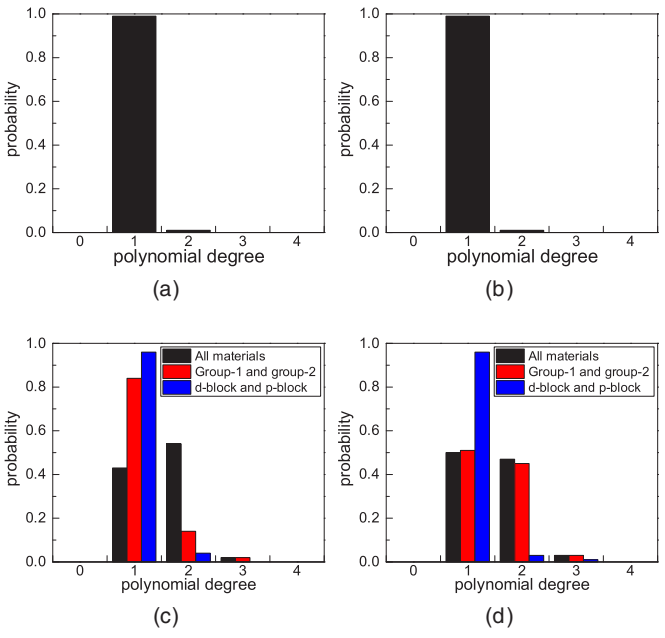

FIG. 4. Results of the Bayesian model selection (BMS) analysis. Left: LDA work functions (a) and surface energies (c). Right: PBE work functions (b) and surface energies (d). A linear fit is clearly the most likely model for both LDA and PBE work functions. For surface energies, the different error character of group 1 and group 2 materials as opposed to the *d* block and *p* block materials result in the quadratic model competing with the linear model.

B. Outlier analysis

In Sec. II C it was stressed that one of the inherent assumptions of the present statistical analysis is the zero-centered normality of the residual errors. Since all data sets contain fewer than 100 materials there are no meaningful quantitative tests available to check for normality. The most effective ways to check for non-normal behavior are then the simple histogram and the quantile-quantile plot (QQ plot). The latter is constructed with Blom's method for normal scores [68]. Both graphical tools also assist in spotting outliers in the data set. Excluding such outliers merely on the grounds of an aberrant residual error is unjustified, however. Indeed, as the residual error contains material-specific inaccuracies of the DFT result, outliers may provide information on the shortcomings of the theoretical prediction.

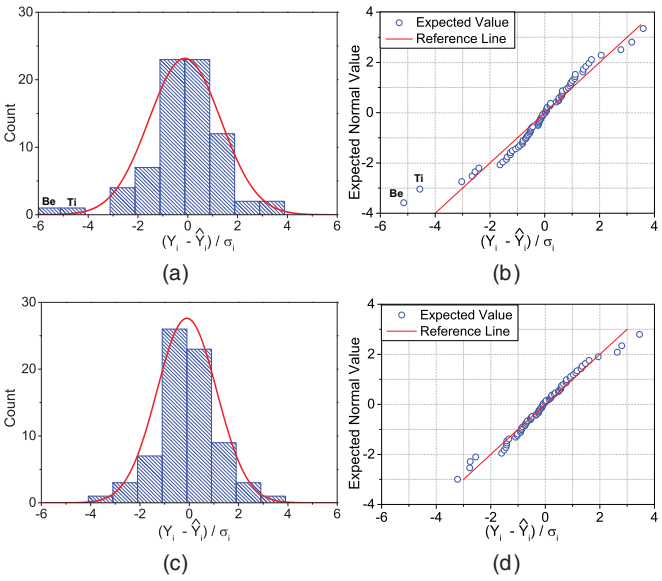

FIG. 5. Analysis of the residual errors for LDA work functions: histograms with fitted normal distributions before (a) and after (c) the removal of the outliers Be and Ti, accompanied by QQ plots before (b) and after (d) outlier removal.

In the histogram and QQ plot of the residuals [69] for the LDA work function (Fig. 5), two clear outliers can be observed: Be and Ti [see also Fig. 6(a)]. The same outliers can also be identified in the PBE work function data set [Fig. 6(b)]. The large residual error is most likely due to the surface crystallography of the samples used to measure the experimental work functions. Section II A stated that for a given material, the lowest anisotropic DFT work function corresponds to the polycrystalline result. This requires of course that the surface orientation with the lowest work function is present in the polycrystalline sample, a condition which seems unfulfilled in the case of Be and Ti. This is especially the case for Be, for which there are alternative work function data available [70,71] that vary significantly from the value of Michaelson. Both materials were therefore excluded from the analysis. Figure 5 demonstrates the importance of removing both outliers. The reduced data in Fig. 5(d) fit significantly better to a normal distribution than the full set in Fig. 5(b). For LDA and PBE the SER drops from 1.46 to 1.23 and 1.49 to 1.29, respectively. This reduction of the SER is not merely a consequence of outlier removal, it also indicates a much improved regression result for the remaining data. The DFT-predicted work functions of Zr and As also show quite a large discrepancy with respect to the experimental values (Fig. 6). These were materials for which the work function was presented as uncertain by Michaelson [40], therefore they carry no weight in the statistical analysis. As such, it is expected that the DFT prediction is more accurate than the uncertain experimental values for those materials.

Among the residual errors for the LDA surface energies, Cr and Ge are two clear outliers (see Fig. 6(c) and the Supplemental Material [48]). For both of these materials, the disagreement between theory and experiment is due to the breakdown of the approach of Tyson and Miller [36] to

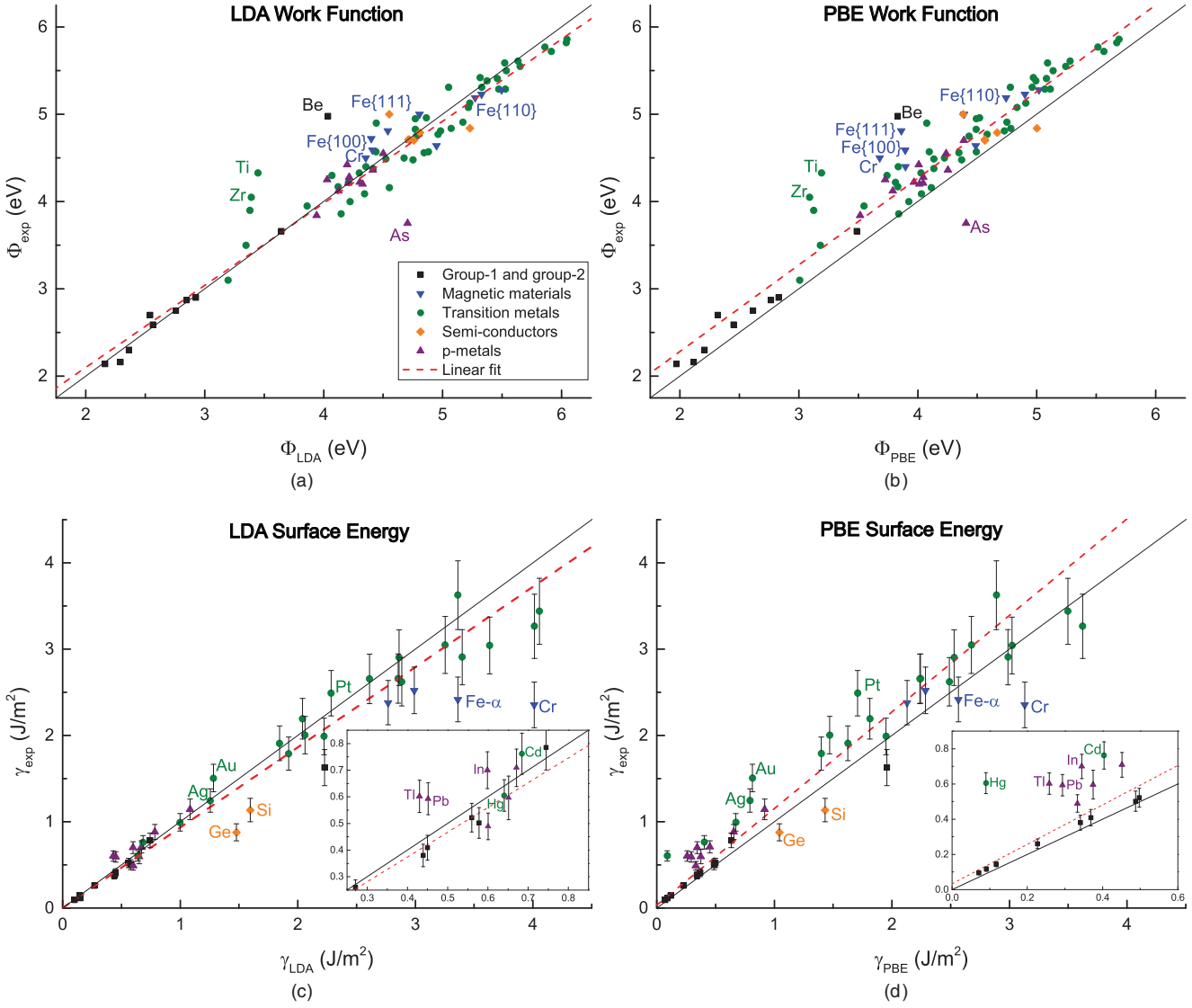

FIG. 6. Linear fits of experimental values to DFT predictions for all four evaluated data sets. Left: LDA work functions (a) and surface energies (c). Right: PBE work functions (b) and surface energies (d). For clarity, an inset is provided of the low surface energy range for both LDA (c) and PBE (d). The data points are color coded per material class [legend in (a)]. The linear fit is indicated by a red dashed line, with the first bisector in black, corresponding to the ideal theory, serving as reference. Materials of particular interest are indicated with a label.

obtain reliable experimental surface energies. Cr is the only material in the present test set which is antiferromagnetic [72]; no magnetic surface effects are included in the derivation by Tyson and Miller however. In the case of Ge, the significant surface reconstruction is probably the cause of the very large residual error. The approach by Tyson and Miller focuses specifically on metals, without taking surface reconstruction into account. For Si, of which the surface energy was also overestimated but not such a significant outlier, it is possible to consult additional experimental results, derived via other methods, to shed some light on this problem. Eaglesham *et al.* [73] derived surface energies for the low-index surface orientations of Si from the equilibrium shape of crystallites, combined with the cleavage energy for the Si{111} plane measured by Jaccodine [74]. With the latter value measured at $-196\,^{\circ}$C and the crystallite equilibrium shapes used to extrapolate the {111} surface energies to the other orientations, the values presented by Eaglesham *et al.* are good candidates for comparison with 0 K DFT values. These experimental values are in much better agreement with the DFT values than the semiempirical results obtained by Tyson and Miller [36]. The DFT surface energies are still larger than the experimental values, but only by about 3%–11% for PBE and 16%–23% for LDA. Moreover, the anisotropy found by Eaglesham *et al.* [73] in the crystallite shape, is reproduced by the DFT calculations. When alternative experimental data for Si{110} are also included into the comparison, it is found that no experimental method renders surface energies as low as the one presented by Tyson and Miller, although they range from 1.43 to 5.38 J/m^2 [75]. From this it can be concluded that the semiempirical derivation used by Tyson and Miller is unsuitable for obtaining the surface energy of the Si and Ge semiconductors. Excluding both Cr and Ge from the test set again ensures zero-centered normality of the residuals and improved regression results with lower SERs. To maintain consistency when comparing the LDA and PBE functionals, Cr and Ge are also removed from the PBE surface energy data set, although in this case they are less distinct outliers. This is not because these materials perform better with PBE than with LDA, but because PBE is less accurate for surface energies than LDA (see Sec. III C).

C. Regression analysis

Table I presents the error estimates resulting from the linear regressions for LDA and PBE work functions and surface energies. All four linear fits are shown in Fig. 6, with data points of particular interest accompanied by a label. The weighted residual errors for the work function and surface energies are indicated with a color code in Fig. 7, with darker shades highlighting larger residual errors. Some experimental work functions were deemed unreliable by Michaelson [40], so these carry zero weight in the regression analysis.

Both LDA and PBE produce very similar residual errors for work functions when compared to experiment. Regardless of the chosen functional, the SER is smaller than $\sqrt{2}$ [see Eq. (12)], which means that the DFT inaccuracy, after correcting for predictable errors, is expected to be smaller than the typical experimental uncertainty. These predictable errors differ somewhat between LDA and PBE, with a simpler

TABLE I. Error estimates from the linear regression between experimental and DFT values, for all four data sets. The coefficients β_0 and β_1 are accompanied by an expected error, obtained from the diagonal elements of the covariance matrix (see the Appendix A). A 95% confidence interval is provided for the SERs (see the Appendix A for derivation).

	β_0 (eV)	β_1	SER
Φ_{LDA}	0.22 ± 0.09	0.94 ± 0.02	$1.22^{1.47}_{1.05}$
Φ_{PBE}	0.30 ± 0.09	0.99 ± 0.02	$1.29^{1.55}_{1.11}$
	β_0 (J/m^2)	β_1	SER
γ_{LDA}	0.00 ± 0.01	0.93 ± 0.02	$1.36^{1.74}_{1.12}$
γ_{PBE}	0.03 ± 0.01	1.12 ± 0.02	$2.34^{2.98}_{1.92}$

character of the predictable error for PBE, as it solely consists of a 0.3 eV constant offset.

Although PBE is the higher level of theory, it does not appear to improve surface energies; LDA is clearly the more accurate functional in this respect, as evidenced by its significantly smaller SER. The predictable error derived from the regression shows an 11% underestimation of the surface energy by PBE ($\beta_1 = 1.12$), but an overestimation of 8% by the LDA functional ($\beta_1 = 0.93$). Such trends are to be expected: creating a surface can be crudely approximated by the breaking of atomic bonds [76] and it is a well-established observation in solid-state DFT that the LDA functional overestimates bond strength, whereas PBE underestimates it [77]. Consequently, the predictable errors on the DFT surface energy are transferable from predictable errors on the DFT cohesive energy [19], a bulk property. Such a connection is also present for the residual error. The Pearson correlation coefficient between the residual errors for the surface energy and the *relative* residual errors [78] for the equilibrium volume is -0.81. This suggests that the error on DFT surface energy predictions very often coincides with an error in predicting the correct equilibrium volume. Such a relation is not observed for work function errors.

The magnetic materials, Fe and especially Cr, are among the worst performing materials, their surface energies being severely overestimated by DFT. This ties in to the discussion of the validity of the approach by Tyson and Miller [36] when magnetic materials are concerned (Sec. III B). The surface energies for Fe, Co, and Ni are slightly improved by using PBE, but it is hard to discern whether this is due to improved DFT modeling or the manner in which the experimental data are derived. The surface energies of Ge and Si are also overestimated by both functionals, even though the DFT values presented in this work are somewhat lower than those presented by previous authors [49,79].

The SER for LDA surface energies is markedly lower than for PBE results, making the former stand out as the functional of choice for predictions of the surface energy. The worse general accuracy of PBE surface energies is mainly due to the poor description of materials in the right part of the d block and of some of the p metals. These materials are labeled in Figs. 6(c) and 6(d), and the trend can be clearly seen in Fig. 7. The loss of accuracy of PBE with respect to LDA for surface

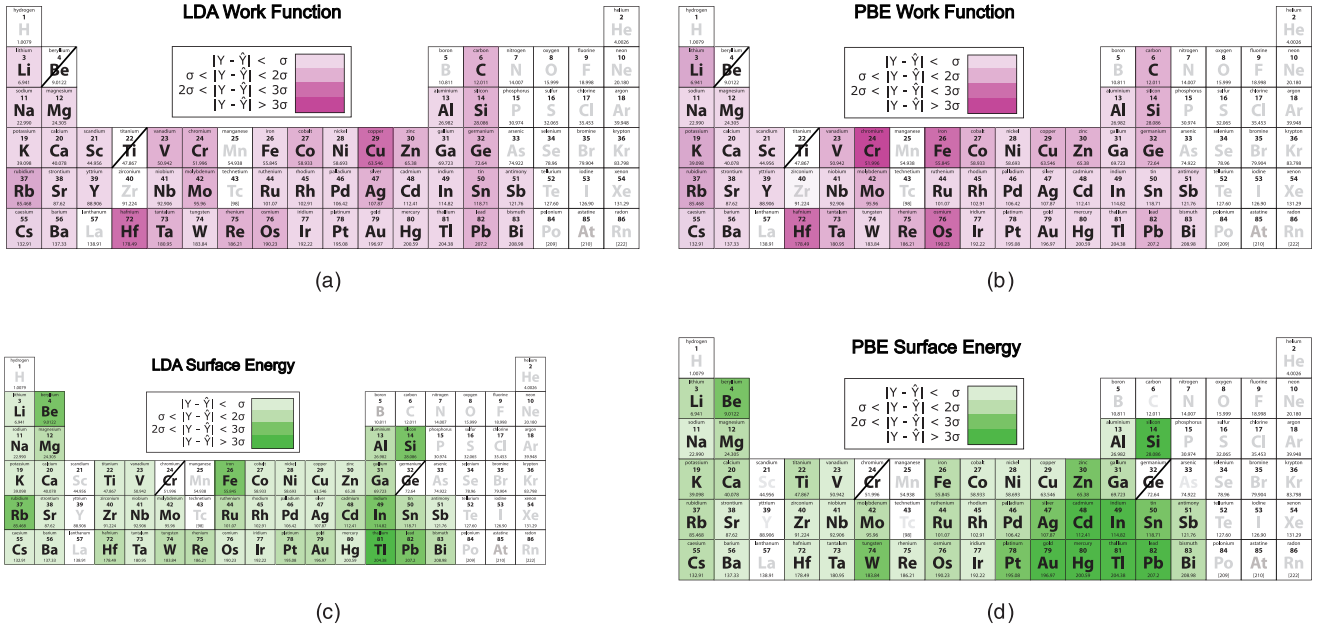

FIG. 7. Overview of the residual errors of the materials for which a comparison with experiment can be made. Top: LDA (a) and PBE (b) work functions. Bottom: LDA (c) and PBE (d) surface energies. A darker shade represents a larger absolute value of the weighted residual error ϵ_i. There are four shades in total, corresponding with the intervals 0–1, 1–2, 2–3, and 3–∞ (see legend). Be and Ti were removed from the work function data set, Cr and Ge were removed from the surface energy data set (see Sec. III B). Materials for which no comparison is made have a gray symbol. Mn, Be, Se, and Te were not calculated because of their crystallographic complexity. There are experimental work function data for Zr and As, but the accuracy of these values is unconfirmed [40].

energies has been discussed by both Kurth *et al.* [80] and Staroverov *et al.* [77], who evaluated the surface energy for a jellium surface. The separate contributions of the exchange and correlation energy to the surface energy were compared to highly accurate random-phase approximation (RPA) calculations that included a short-range GGA correction. They concluded that the reason why LDA provides more accurate surface energies than PBE is entirely fortuitous. Correlation contributions, such as dispersion, are inaccurate for both LDA and PBE, but the overestimation of the exchange contribution to the surface energy within LDA cancels that inaccuracy to a large degree. This explains why materials in the right part of the *d* block suffer from worse surface energy predictions within PBE, as the correlation energy offers a large contribution to the total energy for these materials. This is for example apparent from the extreme case of Hg [81]. In the case of the *p* metals, the overestimation of the exchange energy also corrects for the absence of dispersion contributions. A well-known example where this correction is relevant is the c/a of graphite, which is predicted very accurately by LDA but is overestimated by PBE [82,83]. When, as a test, all materials from group 12 onwards (those with a full *d* shell or a partially occupied *p* shell) and noble metals Pt, Ag, and Au are excluded from the test set, PBE and LDA perform equally well (Table II). For both functionals the SER is reduced to 0.98 and 1, respectively, suggesting their inaccuracy is very similar when the materials with a large correlation energy are excluded. Because these materials are outliers due to theoretical shortcomings in the DFT functionals, and not due to inaccuracies in the experimental result, they were not excluded from the test set.

Due to the deficient description of both exchange and correlation by semilocal functionals, an improved prediction of the surface energy can be achieved by describing both of them at a higher level. One of the possible approaches is quantum Monte Carlo (QMC), which was used by Alfè and Gillan [84] to calculate the surface energy of the MgO {100} surface. By comparing the QMC result to the DFT-LDA and DFT-PBE results, respectively, they concluded that LDA yields a better result. An alternative approach is to use exact exchange (EXX) combined with the many-electron perturbation theory extension known as the random phase approximation (RPA) to go beyond DFT [85,86]. Both Lazar and Otyepka [17] and Schimka *et al.* [87] recently obtained very accurate surface energies by using the EXX+RPA approach. By comparing their highly accurate calculated values with surface energies from cleavage experiments, they suggested the problem of unreliable surface energies as obtained by (semi)local DFT has been solved [77,80].

TABLE II. A comparison between the linear fits of LDA and PBE surface energies to experimental ones when Pt, Ag, Au, and all group 12 through group 15 materials are removed from the data set. Removing these materials with large correlation contributions results in LDA and PBE having a very similar accuracy.

	β_0 (J/m^2)	β_1	SER
γ_{LDA}	0.00 ± 0.01	0.90 ± 0.02	$1.00^{1.37}_{0.79}$
γ_{PBE}	0.02 ± 0.01	1.06 ± 0.03	$0.98^{1.35}_{0.77}$

However, EXX+RPA calculations are computationally very expensive. The current analysis highlights the problematic materials classes for which such an investment would be cost effective.

D. Validation of the model

In the previous section, predictable errors were derived that, when corrected for, enable a better agreement with experiment. These predictable errors remain to be validated. This means that the ansatz that the test set is representative for a general material should be evaluated. In a first step, this can be done internally (within the test set). By leaving one material out of the test set and repeating the linear regression between theory and experiment, the sensitivity of the regression to a single data point can be assessed. None of the four data sets contain a data point of which the removal significantly impacts the parameters $\vec{\beta}$. Not even the PBE surface energy of Hg, the removal of which results in a lowering of the SER to 2.00, alters the regression coefficients more than the expected deviation (Table I).

To externally validate the derived regression results, binary materials can be considered. As an example, DFT-predicted work functions of ZrC{100} [88] and Cu_3Pt{111} [89] are transformed using the regression results from Table I, provided with an error bar and compared to experimental results. The same comparison is made for the surface energy of MgO {100}, although the experimental data in this case originate from cleavage experiments [24,90,91]. The transformed DFT results and the corresponding experimental data are presented in Table III.

The error bars in the first and second column are calculated by propagating the expected error on β_0 and β_1 (which are small), combined with $\sigma_{i,\mathrm{DFT}}$, which is derived from the SER and Eq. (12). To apply Eq. (12), $\sigma_{i,\mathrm{expt}}$ must be known. This is of course not practically possible when DFT is used to *predict* a work function or surface energy. The standard experimental errors that were discussed in Sec. II A can be used for this purpose: $\sigma_{\mathrm{expt}} = 0.32$ eV for work functions and $\sigma_{\mathrm{expt}} = 0.11\gamma$ for surface energies, with γ the transformed DFT result.

For MgO and Cu_3Pt the transformed theoretical results are in excellent agreement with experiment, although the error bar for γ_{PBE} is rather large due to the larger SER. The transformed work function of Cu_3Pt is closer to the experimental value for both LDA and PBE than the bare calculated value. In the case of the surface energy of MgO, it is unclear whether the transformation is an improvement because of the large spread in experimental data. In contrast, the predicted work function of ZrC does not agree all that well with experiment. There is only one experimental value available for the {100} surface, but quite a few more are amassed for polycrystalline ZrC in the Kawano review [41]. They are spread out between 3.2 and 4.48 eV, which indicates that precise ZrC work functions are hard to obtain. In this light, as in the case of As and Zr, the theoretically predicted 4.32 eV might serve as a superior assessment of the ZrC{100} work function.

TABLE III. Comparison of the transformed DFT-predicted surface energy of MgO {100} and work functions of ZrC{100} and Cu_3Pt{111} surfaces to experimental data. Error bars are derived by propagating the error on the coefficients $\vec{\beta}$ and deriving $\sigma_{i,\mathrm{DFT}}$ from the SER. The direct results from DFT calculations are given in parentheses. The experimental error bars on the surface energies are those presented by the original authors. For the work functions, however, no such experimental error bar was given. The general experimental error of ±0.32 eV, derived from the data by Kawano [41], is given here.

	γ_{LDA} (J/m^2)	γ_{PBE} (J/m^2)	γ_{expt} (J/m^2)
MgO	1.16 ± 0.12 (1.25)	1.03 ± 0.24 (0.896)	1.15 ± 0.08 [90] 1.20 ± 0.12 [24] 1.33 ± 0.06 [91]
	Φ_{LDA} (eV)	Φ_{PBE} (eV)	Φ_{expt} (eV)
ZrC{100}	4.31 ± 0.25 (4.35)	4.32 ± 0.29 (4.06)	3.85 ± 0.32 [88]
Cu_3Pt{111}	5.18 ± 0.26 (4.93)	5.22 ± 0.29 (5.32)	5.40 ± 0.32 [89]

IV. CONCLUSION AND OUTLOOK

By comparing 43 experimental surface energies and 73 work functions to their DFT-predicted counterparts, the accuracy of the PBE and LDA functionals was quantified for surface properties. The deviation between the theoretical predictions and experiment was separated into a predictable and a residual contribution. The predictable error was determined with a weighted linear regression, fitting to a linear trend, and is valid for a general material. The residual error, on the other hand, stems from material-specific deviations from the general trend. It was quantified by the standard error on the regression (SER), which is derived from the residual errors for the entire test set. The SER is regarded as a good measure of theoretical inaccuracy and was used to assess the quality of DFT predictions with different functionals. The comparison of the DFT-predicted surface energies and work functions with experimental data is summarized in Fig. 6. The residual errors for both functionals are indicated with a color code in Fig. 7.

Both the LDA and PBE functional yield a SER smaller than $\sqrt{2}$ when the DFT-predicted work functions are compared to experiment. This means that, after correcting for predictable errors, both levels of theory are likely to yield work function results of comparable—even slightly better—quality than experimental results, as the error due to theoretical inaccuracy is expected to be smaller than the one resulting from experimental imprecision. The predictable part of the error has a very simple form for the PBE functional, as it only consists of a 30 meV constant offset. Both LDA and PBE, however, are excellent tools to predict the work function of a crystalline solid.

For surface energy predictions, there is a large discrepancy between the SER of the LDA and PBE functionals. This is mainly caused by poor performance of the PBE functional. Although it is a higher level of theory, it severely underestimates the surface energy of materials for which correlation contributions are important. In contrast, the LDA functional provides more accurate predictions for such materials. This agreement is, however, entirely fortuitous, as was discussed by Kurth *et al.* [80] and Staroverov *et al.* [77]. As such, LDA emerges as a more accurate method for calculating surface energies.

It would be very insightful to extend this research to other functionals, especially for surface energies. Other parametrizations of the GGA, such as the revised PBE (revPBE) [92] or PBEsol functional [93], are tuned for a better prediction of surface properties and could provide more accurate surface energies. More recent GGA functionals, such as the TM functional by Tao and Mo [94], might also provide improved surface energies. However, significant improvements, especially for correlation-governed materials, probably require a higher level of theory. The use of the random-phase approximation (RPA) could solve the severe underestimation of the surface energy for those materials, as was recently suggested in reports by Schimka *et al.* [87] and Lazar and Otyepka [17].

This research was limited to the VASP software package and the use of its recommended PAW potentials. A comparison with different simulation packages, as was performed by Lejaeghere *et al.* [19,57] for bulk properties, could uncover computational shortcomings not associated with the exchange-correlation functional. In addition, the inclusion of spin-orbit coupling, or even of fully relativistic effects, could provide valuable insights into DFT-related shortcomings for the heavier materials.

During this research, over 300 anisotropic work functions and surface energies were calculated. Most predictions, however, cannot be directly compared to experimental data. Moreover, for some materials the experimental values are scarce or unreliable, e.g., the work functions of As, Zr, and ZrC. In such cases, DFT results may provide a reliable estimate, after correcting for predictable errors. In addition, the development of more advanced experimental methods, especially for measuring surface energy, could provide a wealth of information to analyze the DFT shortcomings more precisely.

ACKNOWLEDGMENTS

This work benefited from financial support from the Research Board of Ghent University (BOF) and the Research Foundation Flanders (FWO) (projects No. G.0760.12 and No. G0E0116N). Sam De Waele and Stefaan Cottenier acknowledge financial support from OCAS NV by a funded Ph.D. position and by an OCAS-endowed chair at Ghent University, respectively. The computational resources and services used in this work were provided by the VSC (Flemish Supercomputer Center), funded by the Research Foundation Flanders (FWO) and the Flemish Government department EWI (Economy, Science and Innovation).

APPENDIX: WEIGHTED GENERAL LINEAR REGRESSION

This Appendix serves to provide all the statistical formulas used in this work. The estimators for the parameters $\vec{\beta}$ are found by minimizing Eq. (11). The result can be elegantly contracted in matrix form:

$$\vec{\beta} = (\mathbf{X}^{\mathrm{T}} \cdot \mathbf{W} \cdot \mathbf{X})^{-1} \cdot (\mathbf{X}^{\mathrm{T}} \cdot \mathbf{W} \cdot \vec{Y}).$$

For a fit to a polynomial of degree p, the (i,m) element of $\mathbf{X}$ is X_i^{m-1} (X_i being the ith individual data point) and $\mathbf{W}$ is a square diagonal matrix with dimensions $(p+1) \times (p+1)$ and $\mathbf{W}(i,i) = w_i$. $\vec{Y}$ is the column matrix with experimental values Y_i.

As $\hat{\vec{Y}} = \mathbf{X} \cdot \vec{\beta}$, the residuals $\vec{\epsilon} = \vec{Y} - \hat{\vec{Y}}$ can be expressed in terms of $\mathbf{X}$, $\mathbf{W}$, and $\vec{Y}$ as $\vec{\epsilon} = (1_n - \mathbf{P}) \cdot \vec{Y}$ with the hat matrix $\mathbf{P}$,

$$\mathbf{P} = \mathbf{X} \cdot (\mathbf{X}^{\mathrm{T}} \cdot \mathbf{W} \cdot \mathbf{X})^{-1} \cdot \mathbf{X}^{\mathrm{T}} \cdot \mathbf{W}.$$

From the residuals, the SSR can be obtained as $\mathrm{SSR} = \vec{\epsilon}^{\mathrm{T}} \vec{\epsilon}$.

The covariance matrix $\mathbf{V}$ from which the expected errors on the parameters can be derived for the general linear regression case is

$$\mathbf{V} = (\mathbf{X}^{\mathrm{T}} \cdot \mathbf{W} \cdot \mathbf{X})^{-1}.$$

A confidence interval for the SSE can be determined from SSE/σ^2, which is χ^2 distributed with $n-p$ degrees of freedom. From this, an interval for the SER can be derived.

Algebraically, a weighted linear regression is equivalent to a normal linear regression after the transformation

$$\mathbf{X}' = \mathbf{W}^{1/2} \cdot \mathbf{X},$$

$$\vec{Y}' = \mathbf{W}^{1/2} \cdot \vec{Y}.$$

This allows the use of the BMS algorithm as described by Mana *et al.* [67], where the conditional likelihood of a certain model $\mathbf{M}$, defined by the matrix $\mathbf{X}$, given the data $\vec{Y}$ can be calculated with

$$\mathrm{Prob}(\mathbf{M}|\vec{Y}) \propto \frac{\Gamma\left(\frac{n-l}{2}\right)\Gamma\left(\frac{l}{2}\right)}{|\hat{\vec{Y}}'|^l |\epsilon'|^{(n-l)}} - \frac{\Gamma\left(\frac{n-l}{2}\right)\Gamma\left(\frac{n}{2}\right) {}_2F_1\left(\frac{l}{2}, \frac{n-l}{2}; \frac{n+2-l}{2}; \frac{|\epsilon'|^2}{|\hat{\vec{Y}}'|^2}\right)}{|\hat{\vec{Y}}'|^n},$$

where Γ is the gamma function, ${}_2F_1$ is the regularized hypergeometric function, l is the number of parameters that are fitted in the regression ($l = p+1$), and $\hat{\vec{Y}}'$ and ϵ' are the predictions and the residuals resulting from the normal linear regression using $\mathbf{X}'$ and $\vec{Y}'$. If the model search space is restricted, the above equation can be used to assign a probability to each model.

[1] P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964).
[2] W. Kohn and L. Sham, Phys. Rev. **140**, A1133 (1965).
[3] N. Sukumar, G. Pilania, H. Zhu, and R. Ramprasad, *A Matter of Density: Exploring the Electron Density Concept in the Chemical, Biological, and Materials Sciences* (John Wiley and Sons, New York, 2012).
[4] J. Greeley, T. F. Jaramillo, J. Bonde, I. Chorkendorff, and J. K. Nørskov, Nat. Mater. **5**, 909 (2006).

[5] J. K. Nørskov, F. Abild-Pedersen, F. Studt, and T. Bligaard, Proc. Natl. Acad. Sci. **108**, 937 (2011).

[6] J. M. Kweun, C. Li, Y. Zheng, M. Cho, Y. Y. Kim, and K. Cho, Appl. Surf. Sci. **370**, 279 (2016).

[7] L. Vitos, A. V. Ruban, H. L. Skriver, and J. Kollár, Surf. Sci. **411**, 186 (1998).

[8] C. J. Fall, N. Binggeli, and A. Baldereschi, Phys. Rev. B **61**, 8489 (2000).

[9] I. G. Batyrev, J.-H. Cho, and L. Kleinman, Phys. Rev. B **63**, 172420 (2001).

[10] H. W. Hugosson, O. Eriksson, U. Jansson, A. V. Ruban, P. Souvatzis, and I. A. Abrikosov, Surf. Sci. **557**, 243 (2004).

[11] M. C. Holthausen, J. Comput. Chem. **26**, 1505 (2005).

[12] J. L. F. Da Silva, C. Stampfl, and M. Scheffler, Surf. Sci. **600**, 703 (2006).

[13] M. Ropo, K. Kokko, and L. Vitos, Phys. Rev. B **77**, 195445 (2008).

[14] N. E. Singh-Miller and N. Marzari, Phys. Rev. B **80**, 235407 (2009).

[15] L. Kornblum, P. Shekhter, Y. Slovatizky, Y. Amouyal, and M. Eizenberg, Phys. Rev. B **86**, 125305 (2012).

[16] J. Wang and S.-Q. Wang, Surf. Sci. **630**, 216 (2014).

[17] P. Lazar and M. Otyepka, Phys. Rev. B **91**, 115402 (2015).

[18] J. Depeng and W. Shaoqing, Acta Metall. Sin. **51**, 597 (2015).

[19] K. Lejaeghere, V. Van Speybroeck, G. Van Oost, and S. Cottenier, Crit. Rev. Solid State Mater. Sci. **39**, 1 (2013).

[20] K. Lejaeghere, J. Jaeken, V. Van Speybroeck, and S. Cottenier, Phys. Rev. B **89**, 014304 (2014).

[21] J. C. Heyraud and J. J. Métois, Surf. Sci. **177**, 213 (1986).

[22] C. Bombis, A. Emundts, M. Nowicki, and H. P. Bonzel, Surf. Sci. **511**, 83 (2002).

[23] H. P. Bonzel and M. Nowicki, Phys. Rev. B **70**, 245430 (2004).

[24] J. J. Gilman, J. Appl. Phys. **31**, 2208 (1960).

[25] C. T. Campbell, S. C. Parker, and D. E. Starr, Science **298**, 811 (2002).

[26] O. W. Richardson, *Thermionic Emission from Hot Bodies* (Watchmaker Publishing, Aberdeen, WA, 2003).

[27] Z. Zhang and J. T. Yates, Chem. Rev. **112**, 5520 (2012).

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

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

[30] T. H. Heumann and J. Johannisson, Acta Metall. **20**, 617 (1972).

[31] V. K. Kumikov, Mater. Sci. Eng. **60**, L23 (1983).

[32] C. Ozcan and N. Hasirci, J. Appl. Polym. Sci. **108**, 438 (2008).

[33] K. Kendall, N. McN.Alford, and J. D. Birchall, Nature (London) **325**, 794 (1987).

[34] K. Kendall, J. Phys. D: Appl. Phys. **23**, 1329 (1990).

[35] K. Kendall, Powder Technol. **66**, 101 (1991).

[36] W. R. Tyson and W. A. Miller, Surf. Sci. **62**, 267 (1977).

[37] W. A. Miller and G. A. Chadwick, Acta Metall. **15**, 607 (1967).

[38] W. Tyson, Canad. Metal. Quart. **14**, 307 (1975).

[39] This does not solely include atoms that are directly at the surface. Rather, it concerns all atoms sufficiently near to the surface that contribute to the surface entropy.

[40] H. B. Michaelson, J. Appl. Phys. **48**, 4729 (1977).

[41] H. Kawano, Prog. Surf. Sci. **83**, 1 (2008).

[42] J. C. Boettger, Phys. Rev. B **49**, 16798 (1994).

[43] V. Fiorentini and M. Methfessel, J. Phys.: Condens. Matter **8**, 6525 (1996).

[44] J. F. Nicholas, in *Structure*, edited by G. Chiarotti (Springer, Berlin, 1993), Vol. 24a, pp. 37–39.

[45] C. J. Fall, N. Binggeli, and A. Baldereschi, J. Phys.: Condens. Matter **11**, 2689 (1999).

[46] M. G. Helander, M. T. Greiner, Z. B. Wang, and Z. H. Lu, Appl. Surf. Sci. **256**, 2602 (2010).

[47] S. Curtarolo, W. Setyawan, G. L. W. Hart, M. Jahnatek, R. V. Chepulskii, R. H. Taylor, S. Wang, J. Xue, K. Yang, O. Levy, M. J. Mehl, H. T. Stokes, D. O. Demchenko, and D. Morgan, Comput. Mater. Sci. **58**, 218 (2012).

[48] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevB.94.235418 for crystallography information files (.cif's) of all the optimized surfaces and a document containing the experimental data, the DFT results with computational settings and the figures for the statistical analysis of the residuals.

[49] A. A. Stekolnikov, J. Furthmüller, and F. Bechstedt, Phys. Rev. B **65**, 115318 (2002).

[50] G. Kresse and J. Hafner, Phys. Rev. B **47**, 558 (1993).

[51] G. Kresse and J. Hafner, Phys. Rev. B **49**, 14251 (1994).

[52] G. Kresse and J. Furthmüller, Comput. Mater. Sci. **6**, 15 (1996).

[53] G. Kresse and J. Furthmüller, Phys. Rev. B **54**, 11169 (1996).

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

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

[56] G. Kresse, M. Marsman, and J. Furthmüller, VASP the GUIDE (2015).

[57] K. Lejaeghere, G. Bihlmayer, T. Björkman, P. Blaha, S. Blügel, V. Blum, D. Caliste, I. E. Castelli, S. J. Clark, A. D. Corso, S. d. Gironcoli, T. Deutsch, J. K. Dewhurst, I. D. Marco, C. Draxl, M. Dułak, O. Eriksson, J. A. Flores-Livas, K. F. Garrity, L. Genovese, P. Giannozzi, M. Giantomassi, S. Goedecker, X. Gonze, O. Grånäs, E. K. U. Gross, A. Gulans, F. Gygi, D. R. Hamann, P. J. Hasnip, N. A. W. Holzwarth, D. Iuşan, D. B. Jochym, F. Jollet, D. Jones, G. Kresse, K. Koepernik, E. Küçükbenli, Y. O. Kvashnin, I. L. M. Locht, S. Lubeck, M. Marsman, N. Marzari, U. Nitzsche, L. Nordström, T. Ozaki, L. Paulatto, C. J. Pickard, W. Poelmans, M. I. J. Probert, K. Refson, M. Richter, G.-M. Rignanese, S. Saha, M. Scheffler, M. Schlipf, K. Schwarz, S. Sharma, F. Tavazza, P. Thunström, A. Tkatchenko, M. Torrent, D. Vanderbilt, M. J. v. Setten, V. V. Speybroeck, J. M. Wills, J. R. Yates, G.-X. Zhang, and S. Cottenier, Science **351**, aad3000 (2016).

[58] M. Methfessel and A. T. Paxton, Phys. Rev. B **40**, 3616 (1989).

[59] P. E. Blöchl, O. Jepsen, and O. K. Andersen, Phys. Rev. B **49**, 16223 (1994).

[60] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, and OIML (2008).

[61] B. Kirchner, G. Maroulis, R. Paul, S. Roy, P. Sarkar, A. Savin, G. Wu, and M. Springborg, *Chemical Modelling: Applications and Theory* (Royal Society of Chemistry, London, 2012).

[62] K. K. Irikura, R. D. Johnson, III, and R. N. Kacker, Metrologia **41**, 369 (2004).

[63] G. Hautier, S. P. Ong, A. Jain, C. J. Moore, and G. Ceder, Phys. Rev. B **85**, 155208 (2012).

[64] P. Pernot, B. Civalleri, D. Presti, and A. Savin, J. Phys. Chem. A **119**, 5288 (2015).

[65] N. R. Draper and H. Smith, *Applied Regression Analysis*, 3rd ed., Wiley Series in Probability and Statistics (John Wiley and Sons, New York, 1998).

[66] S. Kirklin, J. E. Saal, B. Meredig, A. Thompson, J. W. Doak, M. Aykol, S. Rühl, and C. Wolverton, npj Comput. Mater. **1**, 15010 (2015).
[67] G. Mana, P. A. Giuliano Albo, and S. Lago, Measurement **55**, 564 (2014).
[68] D. G. Altman, *Practical Statistics for Medical Research*, 1st ed. (Chapman and Hall/CRC, FL, 1990).
[69] Studentized residuals are the preferred choice to check for outliers, whereas the normal residuals are most appropriate for checking for non-normal behavior. The studentized residuals for all the data evaluated in this work were always close to the normal residuals, yielding no different conclusions with respect to outliers.
[70] V. S. Fomenko and G. V. Samsonov, in *Handbook of Thermionic Properties*, edited by V. S. Fomenko and G. V. Samsonov (Springer, New York, 1974), pp. 5–68.
[71] A. K. Green and E. Bauer, Surf. Sci. **74**, 676 (1978).
[72] R. Soulairol, C.-C. Fu, and C. Barreteau, Phys. Rev. B **84**, 155402 (2011).
[73] D. J. Eaglesham, A. E. White, L. C. Feldman, N. Moriya, and D. C. Jacobson, Phys. Rev. Lett. **70**, 1643 (1993).
[74] R. J. Jaccodine, J. Electrochem. Soc. **110**, 524 (1963).
[75] A. Gleizer and D. Sherman, Int. J. Fract. **187**, 1 (2014).
[76] L. Z. Mezey and J. Giber, Jpn. J. Appl. Phys. **21**, 1569 (1982).
[77] V. N. Staroverov, G. E. Scuseria, J. Tao, and J. P. Perdew, Phys. Rev. B **69**, 075102 (2004).
[78] K. Lejaeghere, L. Vanduyfhuys, T. Verstraelen, V. Van Speybroeck, and S. Cottenier, Comput. Mater. Sci. **117**, 390 (2016).
[79] R. Pérez and P. Gumbsch, Phys. Rev. Lett. **84**, 5347 (2000).
[80] S. Kurth, J. P. Perdew, and P. Blaha, Int. J. Quantum Chem. **75**, 889 (1999).
[81] N. Gaston, B. Paulus, K. Rosciszewski, P. Schwerdtfeger, and H. Stoll, Phys. Rev. B **74**, 094102 (2006).
[82] N. Mounet and N. Marzari, Phys. Rev. B **71**, 205214 (2005).
[83] C. Lechner, B. Pannier, P. Baranek, N. C. Forero-Martinez, and H. Vach, J. Phys. Chem. C **120**, 5083 (2016).
[84] D. Alfè and M. J. Gillan, J. Phys.: Condens. Matter **18**, L435 (2006).
[85] H. Eshuis, J. E. Bates, and F. Furche, Theor. Chem. Acc. **131**, 1 (2012).
[86] X. Ren, P. Rinke, C. Joas, and M. Scheffler, J. Mater. Sci. **47**, 7447 (2012).
[87] L. Schimka, J. Harl, A. Stroppa, A. Grüneis, M. Marsman, F. Mittendorfer, and G. Kresse, Nat. Mater. **9**, 741 (2010).
[88] T. C. Tessner and P. R. Davis, J. Vac. Sci. Technol. A **11**, 1 (1993).
[89] U. Schneider, G. R. Castro, H. Busse, T. Janssens, J. Wesemann, and K. Wandelt, Surf. Sci. **269–270**, 316 (1992).
[90] A. R. C. Westwood and D. L. Goldheim, J. Appl. Phys. **34**, 3335 (1963).
[91] P. L. Gutshall and G. E. Gross, J. Appl. Phys. **36**, 2459 (1965).
[92] B. Hammer, L. B. Hansen, and J. K. Nørskov, Phys. Rev. B **59**, 7413 (1999).
[93] M. Fishman, H. L. Zhuang, K. Mathew, W. Dirschka, and R. G. Hennig, Phys. Rev. B **87**, 245402 (2013).
[94] J. Tao and Y. Mo, Phys. Rev. Lett. **117**, 073001 (2016).