Error Estimates for Solid-State Density-Functional Theory Predictions: An Overview by Means of the Ground-State Elemental Crystals

K. Lejaeghere , V. Van Speybroeck , G. Van Oost & S. Cottenier

Taylor & Francis
Taylor & Francis Group

Error Estimates for Solid-State Density-Functional Theory Predictions: An Overview by Means of the Ground-State Elemental Crystals

K. Lejaeghere,[1] V. Van Speybroeck,[1] G. Van Oost,[2] and S. Cottenier[1,3,*]

[1]Center for Molecular Modeling, Ghent University, Technologiepark 903, BE-9052 Zwijnaarde, Belgium
[2]Department of Applied Physics, Ghent University, Sint-Pietersnieuwstraat 41, BE-9000 Ghent, Belgium
[3]Department of Materials Science and Engineering, Ghent University, Technologiepark 903, BE-9052 Zwijnaarde, Belgium

Predictions of observable properties by density-functional theory calculations (DFT) are used increasingly often by experimental condensed-matter physicists and materials engineers as data. These predictions are used to analyze recent measurements, or to plan future experiments in a rational way. Increasingly more experimental scientists in these fields therefore face the natural question: what is the expected error for such a first-principles prediction? Information and experience about this question is implicitly available in the computational community, scattered over two decades of literature. The present review aims to summarize and quantify this implicit knowledge. This eventually leads to a practical protocol that allows any scientist—experimental or theoretical—to determine justifiable error estimates for many basic property predictions, without having to perform additional DFT calculations.

A central role is played by a large and diverse test set of crystalline solids, containing all ground-state elemental crystals (except most lanthanides). For several properties of each crystal, the difference between DFT results and experimental values is assessed. We discuss trends in these deviations and review explanations suggested in the literature.

A prerequisite for such an error analysis is that different implementations of the same first-principles formalism provide the same predictions. Therefore, the reproducibility of predictions across several mainstream methods and codes is discussed too. A quality factor Δ expresses the spread in predictions from two distinct DFT implementations by a single number. To compare the PAW method to the highly accurate APW+lo approach, a code assessment of VASP and GPAW (PAW) with respect to WIEN2k (APW+lo) yields Δ-values of 1.9 and 3.3 meV/atom, respectively. In both cases the PAW potentials recommended by the respective codes have been used. These differences are an order of magnitude smaller than the typical difference with experiment, and therefore predictions by APW+lo and PAW are for practical purposes identical.

Keywords error estimate, density-functional theory, benchmarking, code comparison

Table of Contents

1. INTRODUCTION

Density-functional theory (DFT)[1,2] remains one of the most popular methods to treat both model systems and realistic materials in a quantum mechanical way.[3–8] In condensed-matter physics, DFT is not only used to understand the observed behavior of solids, but increasingly more to predict characteristics of compounds that have not yet been determined experimentally (for example in Ma et al.,[9] Hautier et al.,[10] and Jain et al.[11]). In both cases, the first-principles results provide a point of reference, either to analyze data from measurements or to plan future experiments. It is therefore essential to have a quantitative idea of the expected deviation between a DFT prediction of a certain property and the corresponding experimental value. Error estimates are routinely provided in experimental physics, but in DFT applications this is a much less common practice. When confronted with a disagreement between theory and experiment, one usually resorts to higher-order levels of theory instead.[12–14]

As such, DFT is an exact reformulation of quantum physics, and does not involve any approximation. From a purely theoretical point of view, it should lead to exact predictions, with no need for an error estimate. In practice, however, one requires an educated guess for an essential ingredient of DFT: the exchange-correlation functional (hereafter referred to as "functional"). Apart from this main approximation, there are some other features that go beyond DFT in the way it is usually applied, such as the failure of the Born-Oppenheimer approximation[15,16] or high-Z radiative and other corrections from quantum electrodynamics.[17–19] They affect the results to some extent as well, but they are generally much less important than the

particular choice of the exchange-correlation functional. For any of these choices, DFT predictions will not agree perfectly with experimental observations. Deviations of this kind will be referred to as the *intrinsic error* for a particular functional.

Regardless of the difference with experiment, however, all predictions should be independent of how the DFT (Kohn-Sham) equations are solved numerically: each of the many available DFT implementations should give identical results for the same functional. In reality, there will be some scatter in the predictions of different codes,[20,21] as each of them introduces a distinct amount of numerical noise. This second source of fluctuations leads to a *numerical error*. It is therefore also important to assess to what extent predicted properties vary among DFT approaches.

The goal of this article is to quantify the knowledge about these two kinds of DFT errors—intrinsic and numerical—and, where relevant, to review the physics behind them.

A way to obtain insight into computational errors is by means of benchmark studies, examining the performance of different implementations and functionals for a large set of materials and properties. For atomic and molecular benchmark sets the intrinsic errors have already been assessed in great detail, often with the aim of selecting the best functional for a particular property.[20,22–26] Similar studies exist for solid-state DFT as well,[27–42] but they are mostly limited to a small number of properties and/or compounds. In addition, their focus is often on understanding the differences between functionals, so they do not lead to quantitative and universally applicable error estimates with respect to experiment. A benchmark that is really comprehensive

should meet two criteria: the number of elements that is included in the test set should be sufficiently large and the crystal structures in the set should be sufficiently diverse. This guarantees the transferability of the benchmark conclusions with respect to both the intrinsic and numerical errors. A natural choice to construct such an extensive test set emerges from the periodic table of elements. By taking the ground-state crystal structures of all elements,[43] the two criteria for a comprehensive solid-state benchmark set are simultaneously fulfilled. All elements are included, thus trivially fulfilling the first requirement. In addition, the corresponding crystal structures range from simple hexagonal and cubic configurations to low-symmetry geometries, like orthorhombic and monoclinic cells. Of course, extrapolating the obtained insights to more complex materials, such as multicomponent compounds, requires some care. It is not impossible, however, as will be shown, for example, in Section 3.2.

An additional advantage of using all elemental crystals follows from the periodic table's inherent ability to display trends and correlations. The systematic behavior of observable quantities along periods or groups is well known, but the deviations between DFT predictions and experimental values also appear to follow such trends (Section 3). In particular, the largest errors are restricted to distinct regions. So apart from providing a complete test set, the classification of elements in the periodic table allows for an easy visualization and interpretation of the data. Furthermore, the elemental materials are among the best known and most studied materials on Earth. Experimental data collections are, hence, rather easy to find, and one can assume with sufficient confidence that the reported data are accurate.

By means of the ground-state elemental crystals, the present review offers an overview of the power and limitations of solid-state DFT calculations. Although this knowledge is often implicitly available in the computational community, most of it is scattered over two decades of literature. The current work therefore summarizes and quantifies this implicit knowledge into a practical protocol, that will allow any scientist—experimental or theoretical—to provide justifiable error estimates for many basic property predictions. Intrinsic errors follow from a statistical analysis of the deviations between DFT predictions and experimental values for a given functional. Subsets of materials for which the deviations are particularly large are identified, and the reasons for this behavior are discussed. Numerical errors, on the other hand, express to what extent two independent DFT approaches produce identical predictions. We will focus on the correspondence between two particular methods, APW+lo and PAW, by means of representative mainstream codes. For the PAW method, the use of different atomic potentials can have large effects, but by considering only the sets of potentials recommended by each code, we hope to establish a general idea of the PAW error.

Instead of performing an extensive level-of-theory study for various functionals, we will focus on one typical example within the generalized gradient approximation of DFT (GGA). For this, the PBE functional is chosen, because it is known to yield good results for solids of a wide range of elements and properties.[39] Moreover, its popularity[44] guarantees the comparibility and applicability of our results. Other GGA functionals are expected to display approximately the same behavior, except maybe for very specific material classes. Kurth et al.,[27] for instance, showed how four GGA functionals provide similar trends for both the equilibrium volume and the bulk modulus. Of course, the determination of the error estimates is not limited to GGAs. The presented methodology is also applicable to other functionals (such as LDA or hybrid functionals) or first-principles approaches (such as Hartree-Fock, GW or RPA[45]).

This article is organized as follows. Section 2 describes the computational procedure for all properties under consideration and discusses the prerequisites for a sound comparison between theory and experiment. Within Section 3 the differences between DFT-GGA predictions and experimental values are assessed (intrinsic errors), whereas Section 4 focuses on the method and code dependence of the theoretically determined properties (numerical errors).

2. PREDICTING EXPERIMENTAL PROPERTIES BY MEANS OF DFT

2.1 Computational Recipes

DFT computations for five distinct sets of materials properties will be discussed. They can be divided into energetic (ΔE_{coh}) and elastic quantities (V_0, B_0, B_1, C_{ij}).[46] Of course many more properties may be determined by means of DFT, but the quantities introduced here are directly available from straightforward total energy calculations.

The **cohesive energy** or atomization energy ΔE_{coh} is a popular benchmark quantity.[20,22,27–29,32,33,39] Expressed as an energy difference per atom, it is defined as

$$\Delta E_{coh} = -(E_0 - E_{at}). \qquad [1]$$

Here, E_0 represents the energy per atom of the compound under investigation in its ground state, i.e., at 0 K and without external stress. One can determine it by means of a standard pressure optimization procedure, or by fitting a few $E(V)$ data points to an empirical equation of state (EOS) and extracting the equilibrium energy analytically. In this article, the latter option was chosen, using a common third-order Birch-Murnaghan relation:[47]

$$E(V) = E_0 + \frac{9V_0 B_0}{16} \left\{ \left[\left(\frac{V_0}{V} \right)^{2/3} - 1 \right]^3 B_1 \right. \\ \left. + \left[\left(\frac{V_0}{V} \right)^{2/3} - 1 \right]^2 \left[6 - 4 \left(\frac{V_0}{V} \right)^{2/3} \right] \right\}. \quad [2]$$

V_0 represents the equilibrium volume, B_0 the bulk modulus, and B_1 its pressure derivative. Other equations of state exist as well, but no significant difference with respect to ground-state properties is to be expected.

E_{at}, on the other hand, is the energy of one isolated atom in its electronic ground state. Since many solid-state DFT codes only allow for the use of periodic boundary conditions, the isolated atom needs to be calculated in a periodic unit cell as well. In the present computations, a free atom is placed in a big orthorhombic unit cell, such that every atom is surrounded by at least 15 Å of vacuum. In this way, one can sufficiently suppress spurious interactions between periodic images (<1 meV). The orthorhombic symmetry is chosen over, e.g., a simpler cubic one, to avoid physically incorrect spherical states. After all, the use of a unit cell forces the electron density to assume the same symmetry as the lattice.[48] This is, in most cases, only possible by means of partial occupation of the different electron orbitals, which is not physical. Lowering the crystal symmetry counteracts this phenomenon and should lead to strictly integer occupation numbers. However, some atoms end up with partially filled states, even when this approach is applied. In such cases, the occupation numbers have to be fixed manually before looking for the usual, self-consistent solution. In this article, only experimental ground-state electron configurations are used: even when DFT predicts a different configuration to be more stable[49] (e.g., for W), the experimental occupation numbers are taken in order to guarantee a meaningful comparison to measurements. Only for spin-orbit coupled calculations for the Pb atom is it not possible to impose the experimental electronic state. The PBE ground state $^1S_0$ is therefore used, and not the experimental $^3P_0$ state.

The negative sign in Equation (1) causes positive cohesive energies to correspond to stable phases (with respect to atomic decohesion). The other sign convention, however, is commonly used as well.

One of the key physical properties of a given compound is its **volume**. In first-principles calculations the equilibrium volume per atom V_0 can be obtained easily. One either employs an optimization routine or fits some $E(V)$ points to an empirical equation of state. This is similar to the procedure used to determine E_0, and again the latter option is chosen in this work.

The **bulk modulus** is closely related to the $E(V)$ behavior as well. It is proportional to the curvature of the equation of state at the equilibrium volume:

$$B_0 = -V \frac{\partial P}{\partial V}\bigg|_{V=V_0} = V \frac{\partial^2 E}{\partial V^2}\bigg|_{V=V_0}. \qquad [3]$$

It represents the resistance of the unloaded material to volume change, and hence to uniform pressure. Because it is linked to the curvature of the $E(V)$ relation, B_0 is a numerically sensitive quantity. A small deviation at a few data points is already able to change its value noticeably, especially when the bulk modulus is small (shallow EOS). This is increasingly so when only a narrow volume range is inspected.

B_1 stands for the **derivative of the bulk modulus** with respect to pressure, evaluated at the equilibrium volume:

$$B_1 = \frac{\partial B}{\partial P}\bigg|_{V=V_0} = \frac{\partial}{\partial P}\left(V\frac{\partial^2 E}{\partial V^2}\right)\bigg|_{V=V_0}. \qquad [4]$$

It is a third-order derivative of the energy and hence describes effects that are one order higher even than the bulk modulus. It is related to the volume-dependence of the $E(V)$ curvature. B_1 is therefore the most sensitive elastic quantity discussed in this study. Again, both the bulk modulus and its pressure derivative are obtained from fitting an EOS to calculated $E(V)$ data points.

The mechanical behavior of a crystal cannot be described solely by means of the bulk modulus. When anisotropic deformations are applied, other **elastic constants** come into play as well. The full set of these constants makes up the stiffness matrix $\boldsymbol{C}$. It represents a tensor of rank 2 and relates (small) cell strains to the corresponding stresses via Hooke's law $\boldsymbol{\sigma} = \boldsymbol{C} \cdot \boldsymbol{\epsilon}$. C is a symmetric 6×6 matrix, containing 21 independent constants at the most. In the case of hexagonal crystals, five distinct values remain (C_{11}, C_{12}, C_{33}, C_{13}, and C_{44}), while for cubic compounds there are only three (C_{11}, C_{12}, and C_{44}). The C_{ij} parameters can also be translated into more general elastic moduli, such as Young's modulus E, the shear modulus G, and Poisson's ratio ν. Even the bulk modulus can be obtained from a simple combination of the C_{ij}. In addition, the elastic constants are known to relate to structural stability and various other important physical properties.[50, 51]

Several methods are available to obtain the elastic constants from first principles, either by relating energy and strain[52] or stress and strain.[53, 54] In most cases, a stress-based procedure is preferred, because it is inherently faster. However, it requires an *ab initio* code that can determine the stress tensor. In a first step the cell pressure components are then extracted for a minimal set of deformed geometries. Together with the corresponding strains, this results in a system of linear equations. Solving that system yields the required elastic constants. When it is important to obtain an accurate value of C_{ij}, one should construct an overdetermined system, by applying the same strain sets at different magnitudes. The elastic constants can then be retrieved by using a least-squares method.

2.2. Comparing Theory and Experiment

When a DFT prediction is compared to a number from experiment, the corresponding ambient conditions should be as identical as possible. This means that, in the first place, the experimental result should refer to 0 K. Moreover, the measurement should be corrected for zero-point vibrational effects, which are not present in standard DFT calculations. The following paragraphs discuss how to extrapolate the experimental values to absolute zero and correct them for zero-point vibrations.

For the **cohesive energy** it takes little effort to match up theory and experiment consistently. Experimental data at low temperatures are, in most cases, available. Only the zero-point

energy ζ hinders a direct comparison between 0 K and experiment. From quantum mechanics this quantity is known to be $\frac{3}{2}\hbar\langle\omega\rangle$, with $\langle\omega\rangle$ the average phonon frequency. The latter can be estimated from Debye theory, where it is proportional to the maximum vibrational frequency, and hence to the Debye temperature Θ_D. The zero-point energy correction becomes[55]

$$\zeta = \frac{9}{8}k_B\Theta_D. \qquad [5]$$

Theoretical cohesive energies can only be compared to experiment if this contribution is added to the experimental values (added, due to the chosen sign convention in Equation (1)).

When no experimental value is available, Θ_D can be estimated. Here, the Debye-Grüneisen approximation[56]

$$\Theta_D = 0.617\frac{\hbar}{k_B}(6\pi^2)^{1/3}V_0^{1/6}\left(\frac{B_0}{M}\right)^{1/2} \qquad [6]$$

will be used. Both V_0 and the mass M are expressed per particle, corresponding to a single atom for most materials. For dimeric crystals, however, the diatomic molecule is chosen as a unit of repetition. The regular, room temperature experimental values for B_0 and V_0 are filled in, except when the difference with low temperature results (see further) is significant. This is the case for H, Cl, Br, and I.

Thermal **volume** corrections consist of two parts. Assuming to have a room temperature measurement at one's disposal, the first step consists in accounting for thermal expansion from absolute zero to ambient temperature:

$$\frac{\Delta V^{(1)}}{V} = \int_0^{T_{rt}} \alpha_V(T)\mathrm{d}T; \qquad [7]$$

$\alpha_V(T)$ represents the temperature-dependent volume expansion coefficient. It is zero at 0 K and $\alpha_{V,rt}$ at room temperature (T_{rt}). Since $\Delta V^{(1)}$ constitutes only a small correction with respect to the total volume V, Equation (7) will be approximated here as

$$\frac{\Delta V^{(1)}}{V} \approx \int_0^{T_{rt}} \alpha_{V,rt}\frac{T}{T_{rt}}\mathrm{d}T = \frac{\alpha_{V,rt}T_{rt}}{2}. \qquad [8]$$

In a limited number of cases the experimental expansion coefficient is not known. It can then be estimated from an empirical correlation to the "moleculization" energy:[57]

$$\alpha_{V,rt} = 3 \times \frac{48.14 \cdot 10^{-6}\mathrm{eV/K/atom}}{\Delta E_{mol}}. \qquad [9]$$

ΔE_{mol} is defined as the energy difference per atom between the crystalline material and its gas-like molecules. For elements with an atomic gas phase, it reduces to the atomization energy (cohesive energy). In the absence of experimental data on both $\alpha_{V,rt}$ and ΔE_{mol}, Equation (9) is completed with DFT values.

A second modification is again due to zero-point effects. Because of the volume-dependence of the zero-point energy ζ, the equilibrium volume is shifted slightly. According to Alchagirov et al.,[37,55] this small difference per atom amounts to

$$\Delta V^{(2)} = \frac{(B_1 - 1)\zeta}{2B_0} = \frac{9}{16}(B_1 - 1)\frac{k_B\Theta_D}{B_0}. \qquad [10]$$

Dacorogna and Cohen[58] proposed an alternative definition of the zero-point volume shift. They obtain a similar formula, but with B_1 instead of B_1-1 in Equation (10). However, the mathematical expression is preceded by some significant simplifications. When calculating zero-point effects it is therefore advisable to use Equation (10) instead, especially when B_1 is small.

For the **bulk modulus** thermal effects should be taken into account as well. A first contribution originates in the thermal expansion of the material. Similar to $\Delta V^{(1)}$, a correction $\Delta B^{(1)}$ can be determined too. Roughly approximating the relevant behavior, one can write:[59]

$$\Delta B^{(1)} = B_1 \cdot P(\Delta V^{(1)}) = -B_0B_1\frac{\Delta V^{(1)}}{V_0}. \qquad [11]$$

On the other hand, the effect of zero-point vibrations on the bulk modulus boils down to[55]

$$\frac{\Delta B^{(2)}}{B_0} = -\frac{\Delta V^{(2)}}{V_0}\left[\frac{1}{2}(B_1 - 1)\right.$$
$$\left. + \frac{2}{B_1 - 1}\left(\frac{2}{9} - \frac{1}{3}B_1 - \frac{1}{2}B_0B_2\right)\right], \qquad [12]$$

where B_2 stands for the second-order derivative of the bulk modulus with respect to pressure. It is a highly sensitive parameter and very difficult to extract from a few $E(V)$ data points. In addition, B_2 is not included in Equation (2) and a higher-order Birch-Murnaghan fit should be applied. Instead, the present work will use the intrinsic Birch-Murnaghan value:

$$(B_0B_2)^{BM} = B_0\frac{\partial^2}{\partial P^2}\left(V\frac{\partial^2 E^{BM}}{\partial V^2}\right)\Bigg|_{V=V_0}$$
$$= -\frac{143}{9} + 7B_1 - B_1^2. \qquad [13]$$

There are other possibilities as well,[55] ranging from a different equation of state to an accurate numerical determination of B_2. In order to establish the small correction $\Delta B^{(2)}$, however, this more consistent approach suffices.

Since it is already hard to accurately measure a high-order parameter like B_1 or to determine it from first principles, zero temperature modifications will often be negligible compared to experimental or computational errors. B_1 is therefore not adjusted to incorporate thermal expansion or zero-point effects.

No thermal corrections are applied to the elastic constants C_{ij} either. One can, however, imagine a modification similar to that of Dacorogna and Cohen[58] for the bulk modulus:

$$\Delta C_{ij}^{(m)} = \frac{\partial C_{ij}}{\partial P} \cdot P(\Delta V^{(m)})$$
$$= -B_0\frac{\partial C_{ij}}{\partial P}\frac{\Delta V^{(m)}}{V_0} \qquad [14]$$

with $m = 1$ to account for thermal expansion and $m = 2$ for zero-point effects. Unfortunately, experimental data about the pressure derivative of the elastic constants are scarce.

3. INTRINSIC ERRORS

3.1. Test Set Preparation

In order to establish statistically justified intrinsic error estimates, the ground-state elemental crystals at 0 K will be used as a benchmark set. Pettifor[60] lists these crystal structures, based on an overview by Villars and Daams.[43] However, in some cases literature suggests another phase to be even more stable at low temperatures. In order to ensure the use of 0 K cell geometries as much as possible, such an alternate structure is taken for boron,[61] nitrogen,[62] oxygen,[63] and sulfur.[64] Table 1 presents an overview of all structures used in the current test set.

Using a 0 K benchmark set entails two distinct advantages. On the one hand, some elements only crystallize just above absolute zero. Collecting both 300 K and 0 K compounds in one set might then seem a bit inconsistent. On the other hand, this approach facilitates the extrapolation from the experimental temperature to 0 K, as there are no phase transformations along the way.

All structures are considered in their stress-free ground state. This means that, when the space group allows some freedom in the internal positions, an optimization with respect to the total energy is necessary. This optimization procedure calls for a fast and well-accepted DFT algorithm. The projector augmented-wave method[65,66] (PAW) as implemented in VASP[8,67] (version 5.2.2) fulfills both criteria.

The elemental crystal structures have therefore been relaxed by means of this code, using the recommended PAW atomic potentials listed in the manual.[68] A force convergence criterion of 0.01 eV/Å was set. All calculations were performed using the tetrahedron method with Blöchl corrections,[69] while the reciprocal space was sampled by means of a Monkhorst-Pack grid.[70] Further computational details for the calculations are given in the online Supplementary Material, posted on the publisher's website.

The equilibrium structure has been obtained in two stages. For the determination of the equilibrium volume a uniformly spaced 13-point EOS (up to $V_0 \pm 6\%$) was calculated and fitted to a least-squares third-order Birch-Murnaghan relation (see Section 2.1). Only for a number of shallow $E(V)$ curves—in particular for H, N, S, the halogens, and the noble gases—an increased volume range turned out to be necessary. For each of the 13 crystal volumes, the atomic positions and the cell shape have been individually optimized. In a second step, the crystal has been reinitialized at the fitted V_0 and has then been optimized again.

These optimized crystal structures form the definitive test set (submitted to the COD[71] and ICSD[72] crystallographic databases). For each of them, most of the properties discussed in Section 2.1 have been determined in order to quantify the difference between PBE and experimental values (Section 3.2.3). The DFT part of the comparison has been performed by means of

TABLE 1
Ground-state crystal structures for all elements up to radon. Both the space group number and the Pearson notation are given (with hR*x* standing for *x* atoms in the *hexagonal* setting of the rhombohedral unit cell)

H																	He
194																	194
hP4																	hP2
Li	Be											B	C	N	O	F	Ne
166	194											166	194	205	12	15	225
hR9	hP2											hR36	hP4	cP8	mS4	mS8	cF4
Na	Mg											Al	Si	P	S	Cl	Ar
166	194											225	227	64	70	64	225
hR9	hP2											cF4	cF8	oS8	oF128	oS8	cF4
K	Ca	Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn	Ga	Ge	As	Se	Br	Kr
229	225	194	194	229	229	217	229	194	225	225	194	64	227	166	152	64	225
cI2	cF4	hP2	hP2	cI2	cI2	cI58	cI2	hP2	cF4	cF4	hP2	oS8	cF8	hR6	hP3	oS8	cF4
Rb	Sr	Y	Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag	Cd	In	Sn	Sb	Te	I	Xe
229	225	194	194	229	229	194	194	225	225	225	194	139	227	166	152	64	225
cI2	cF4	hP2	hP2	cI2	cI2	hP2	hP2	cF4	cF4	cF4	hP2	tI2	cF8	hR6	hP3	oS8	cF4
Cs	Ba	Lu	Hf	Ta	W	Re	Os	Ir	Pt	Au	Hg	Tl	Pb	Bi	Po	At	Rn
229	229	194	194	229	229	194	194	225	225	225	139	194	225	166	221		225
cI2	cI2	hP2	hP2	cI2	cI2	hP2	hP2	cF4	cF4	cF4	tI2	hP2	cF4	hR6	cP1		cF4

VASP, using the settings mentioned earlier. They allow for the convergence of all energy differences up to a few meV per atom at the most. For O and Cr (antiferromagnetic), Mn (ferrimagnetic), Fe, Co, and Ni (ferromagnetic), spin polarization has been taken into account, while for the heaviest elements (as from Lu) spin-orbit contributions have been incorporated. At that point relativistic effects beyond the scalar-relativistic approach become important, as will be shown later (Table 7).

The analysis in Section 3.2.3 will not show the raw calculated data, but will rather elaborate on the deviation between theory and experiment. The first-principles results and the thermally corrected experimental numbers[40,43,46,62,73–105] have been included in the Supplementary Material. A tabulation of calculated and experimental values for the elastic constants C_{ij} was published before by Shang et al.[40] for most of the present benchmark set. In Section 3.2.3 their data are used. Only for the experimental numbers of Ba did we find more realistic results elsewhere.[98,99] Since the authors only considered bcc, fcc, and hcp structures, this implies that for Li and Na a different geometry was applied than in the rest of this work (bcc instead of hR9). Moreover their results are based on a PW91 functional,[106,107] rather than the PBE approximation employed in the rest of this work. Although these GGA approaches yield different results in a few situations,[108] they are in most cases very similar and for the elastic constants no significant deviation is expected.

3.2. Statistical Analysis

3.2.1. Linear Regression

Benchmark studies usually analyze the difference between DFT and experiment statistically. The most common characteristics investigated are the mean error (signed) and the mean absolute error (unsigned). However, this approach implicitly assumes that the offset between DFT predictions and experimental results is the same for large and small values. For strictly positive quantities a relative shift seems more reasonable. The present analysis, therefore, explicitly treats relative deviations, in addition to the remaining scatter on that trend. This is done by means of a linear regression between DFT data and experimental results.

The linear regression is performed by means of a least-squares fit, from which we obtain the slope as well as the scatter with respect to the regression line.[109] The model hence presumes that a perfect correlation between experiment X and theory T exists, distorted by a random error ϵ centered around a zero mean: $X = \beta T + \epsilon$. If the exact exchange-correlation functional were known, it would approximately lead to $\epsilon = 0$ and $\beta = 1$. In practice, comparing the least-squares estimate of β to 1 offers a good measure of any systematic deviations, while the standard deviation of ϵ (also denoted as standard error of the regression or SER) expresses the residual error bar (see Figure 1).

Although the experimental community commonly employs the nomenclature "systematic errors" and "nonsystematic errors," we will avoid using the second term. Nonsystematic errors
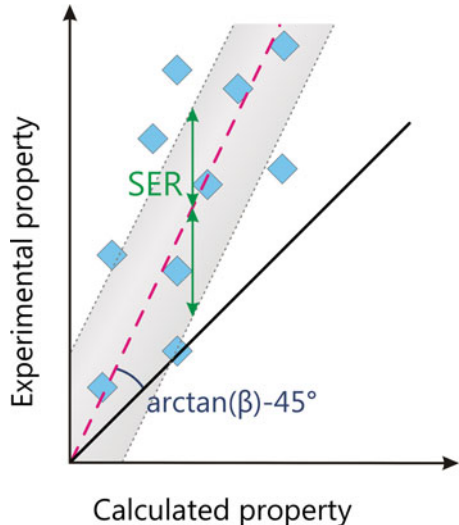

FIG. 1. The intrinsic error between experiment and DFT can be decomposed in a systematic deviation and a residual error bar. Systematic differences are quantified by comparing the linear regression line $X = \beta T$ (red, dashed) to the bisector $X = T$ (black, full). The residual error bar (green arrows) is defined as the standard deviation of the regression errors (SER). (Color figure available online.)

suggest a certain degree of randomness, which is not present in DFT. When one performs the same experiment several times, the results are spread around a mean value. In DFT, repeating the same calculation always yields identical results. In contrast to experimental error analysis, the spread of deviations in DFT results becomes apparent only when predictions for many different compounds are compared to experiment, i.e., when a benchmark set is used. The DFT scatter is then caused by some (subsets of) crystals being described better by the functional than others. Intrinsic error bars are hence fundamentally different from experimental error bars as well. The systematicness of DFT also appears when studying results within a certain family of materials: similar systems behave almost identically, proving the DFT spread not to behave randomly at all. An excellent example is found in literature, where the correspondence with experiment displays much less scatter when chemically similar compounds are involved.[42]

As an additional side note, one should be aware that experimental error bars have not been taken into account in any way. In fact, the statistical model should be $X + \eta = \beta T + \epsilon$, where η represents an additional (but uncorrelated) zero-mean perturbation. This does not only affect the comparison between individual DFT and experimental results, but also influences the values of the intrinsic systematic deviations and residual error bars that are presented in this work. By considering a test set that is sufficiently large, such as in the present study, one may hope that these effects level out. In addition, the elemental crystals belong to the most intensively studied materials, such that the

experimental errors are much smaller than for regular compounds. Without full knowledge of the experimental errors, however, we can only state that the SER provides an upper limit for the real PBE spread σ_ϵ. A possible solution consists in comparing DFT values to results from highly accurate many-body techniques instead. Such high-precision data are not available for many of the materials considered here, however, and to calculate them ourselves exceeds the scope of the current article.

3.2.2. Eliminating Outliers

A full statistical analysis of all elemental data cannot be performed straightforwardly. Some subsets of elements strongly distort the agreement between DFT and experiment. More meaningful error estimates are obtained when the most striking outliers are removed from the data set. Since the deviating behavior is often caused by a bad description of some underlying physical mechanism, most of them will be grouped in subsets of similar compounds. Instead of removing one outlier at a time from the data set, we choose only to remove entire structure types at once. This way, individual materials that behave well for the wrong reasons, are excluded as well, avoiding bias towards smaller errors.

A decomposition of the test set into eight subsets is proposed, based on some common physical properties of the corresponding elemental crystals (Figure 2). They are: (1) alkali and alkaline earth metals; (2) nonmagnetic transition metals; (3) magnetic materials; (4) correlation-dominated materials; (5) high-coordination p block compounds; (6) low-coordination
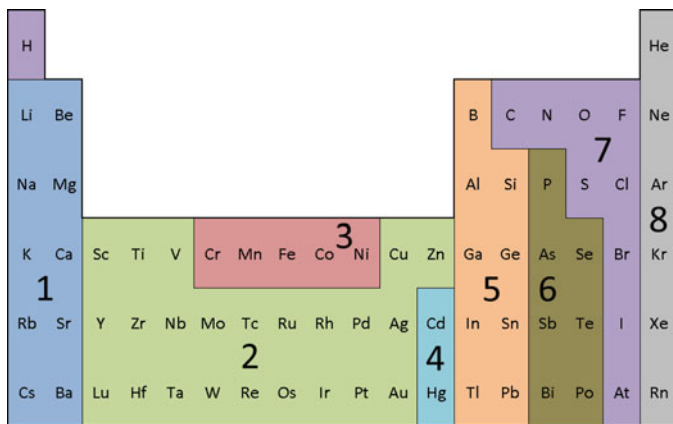

FIG. 2. Decomposition of the periodic table into smaller subsets of elements, based on common physical properties of the corresponding ground-state crystals: (1) alkali and alkaline earth metals; (2) nonmagnetic transition metals; (3) magnetic materials; (4) correlation-dominated materials; (5) high-coordination p block compounds; (6) low-coordination p block compounds; (7) molecular crystals; and (8) noble gases. Subsets 7 and 8 correspond to materials where dispersion interactions are essential. (Color figure available online.)

p block compounds; (7) molecular crystals; and (8) noble gases. Obviously, for some boundary elements, the most appropriate subset can be matter for discussion, but the classification in Figure 2 explains most trends for the intrinsic errors in a satisfactory manner (see further).

These eight subsets of elemental materials are representative for more complex (multicomponent) crystals as well. They provide prototype systems for particular bond types and physical phenomena, such as London dispersion (subsets 7 and 8), magnetism (subset 3), and electronic correlation (subset 4). Observations of DFT performance for these eight subsets will therefore carry over to multicomponent compounds.

In order to eliminate deviating subsets in an objective way, the following procedure has been used. All subsets from Fig. 2 that have half or more of their elements differing significantly from the dominating trend, have been excluded. A two-sided p-value of 10% is maintained. In other words, a data point is considered to deviate substantially when the (signed) relative residual error (expressing the difference between the DFT regression value and the experimental one) belongs to the outer 10% of a normal distribution. This approach was repeated in an iterative way: after the elimination of each subset the significance criterion has been reestablished, until no deviating subsets remained. For solids belonging to an excluded category, PBE is not expected to provide reliable property predictions.

This selection criterion has been visualized in Tables 2 and 3. For each elemental crystal the relative residual error $|x_{exp} - x_{reg}|/x_{exp}$ is shown. Large numbers suggest a significant deviation from the regression line and hence allow for the recognition of outliers. Because these differences are displayed in the shape of the periodic table, they allow for easy identification of deviating subsets[*]. A color code has been added to improve intuition, with the darkest shades corresponding to the largest deviations. The deviations with respect to the elastic constants represent the mean absolute errors over C_{11}, C_{12}, C_{33}, C_{13}, and C_{44}.

Another way to visualize the assessment procedure is presented in Figure 3. In order to get a better view on systematic deviations, both the linear regression line (dashed red) and the first quadrant bisector (full black, representing a perfect match between theory and experiment) have been added for all accepted elements (see Table 4). As an additional quality indicator, the Pearson product-moment correlation coefficient r has been included for each property,[113] with $r = 1$ indicating a perfect positive correlation. Data points corresponding to omitted

---

[*]The graphical representations in this work employ the conventional, medium-form periodic table. Hydrogen is kept in group IA (contrary to the vivid discussion in, e.g., Scerri[110]) and lutetium in group IIIB.[111] This increases the intuitive character of the results. Insights should be conveyed at a glance and many researchers are most familiar with the standard format of the periodic table. For the same reason, the two-dimensional representation is preferred over some 1D alternatives.[112]

TABLE 2

Relative residual errors (of the VASP-PBE regression results with respect to the thermally corrected experimental values) for ΔE_{coh} [46,73,74] (green), V_0 [43,75–82] (red), and B_0 [46,62,83–90] (blue) of the elemental crystals. The darkest shades correspond to the largest errors (color figure available online)

ΔE_{coh}

H																	He
Li 4%	Be 8%											B 6%	C 5%	N 6%	O 17%	F 40%	Ne 31%
Na 4%	Mg 2%											Al 0%	Si 3%	P 1%	S 6%	Cl 3%	Ar 7%
K 8%	Ca 3%	Sc 5%	Ti 12%	V 11%	Cr 3%	Mn 30%	Fe 12%	Co 16%	Ni 8%	Cu 1%	Zn 20%	Ga 7%	Ge 4%	As 1%	Se 6%	Br 10%	Kr 80%
Rb 10%	Sr 7%	Y 5%	Zr 0%	Nb 9%	Mo 9%	Tc 1%	Ru 2%	Rh 2%	Pd 6%	Ag 16%	Cd 37%	In 9%	Sn 0%	Sb 2%	Te 10%	I 15%	Xe 84%
Cs 11%	Ba 1%	Lu 8%	Hf 3%	Ta 2%	W 1%	Re 2%	Os 3%	Ir 0%	Pt 9%	Au 19%	Hg 69%	Tl 20%	Pb 4%	Bi 5%	Po 8%	At	Rn 81%

V_0

H 155%																	He 82%
Li 4%	Be 2%											B 2%	C 39%	N 36%	O 16%	F 30%	Ne 33%
Na 4%	Mg 3%											Al 2%	Si 1%	P 11%	S 51%	Cl 37%	Ar 43%
K 1%	Ca 5%	Sc 4%	Ti 4%	V 6%	Cr 4%	Mn 11%	Fe 6%	Co 4%	Ni 2%	Cu 0%	Zn 0%	Ga 2%	Ge 3%	As 3%	Se 10%	Br 20%	Kr 45%
Rb 1%	Sr 5%	Y 4%	Zr 2%	Nb 2%	Mo 1%	Tc 1%	Ru 1%	Rh 1%	Pd 2%	Ag 3%	Cd 6%	In 3%	Sn 5%	Sb 3%	Te 1%	I 30%	Xe 49%
Cs 2%	Ba 2%	Lu 3%	Hf 3%	Ta 2%	W 1%	Re 1%	Os 1%	Ir 0%	Pt 1%	Au 3%	Hg 22%	Tl 7%	Pb 4%	Bi 5%	Po 3%	At	Rn

B_0

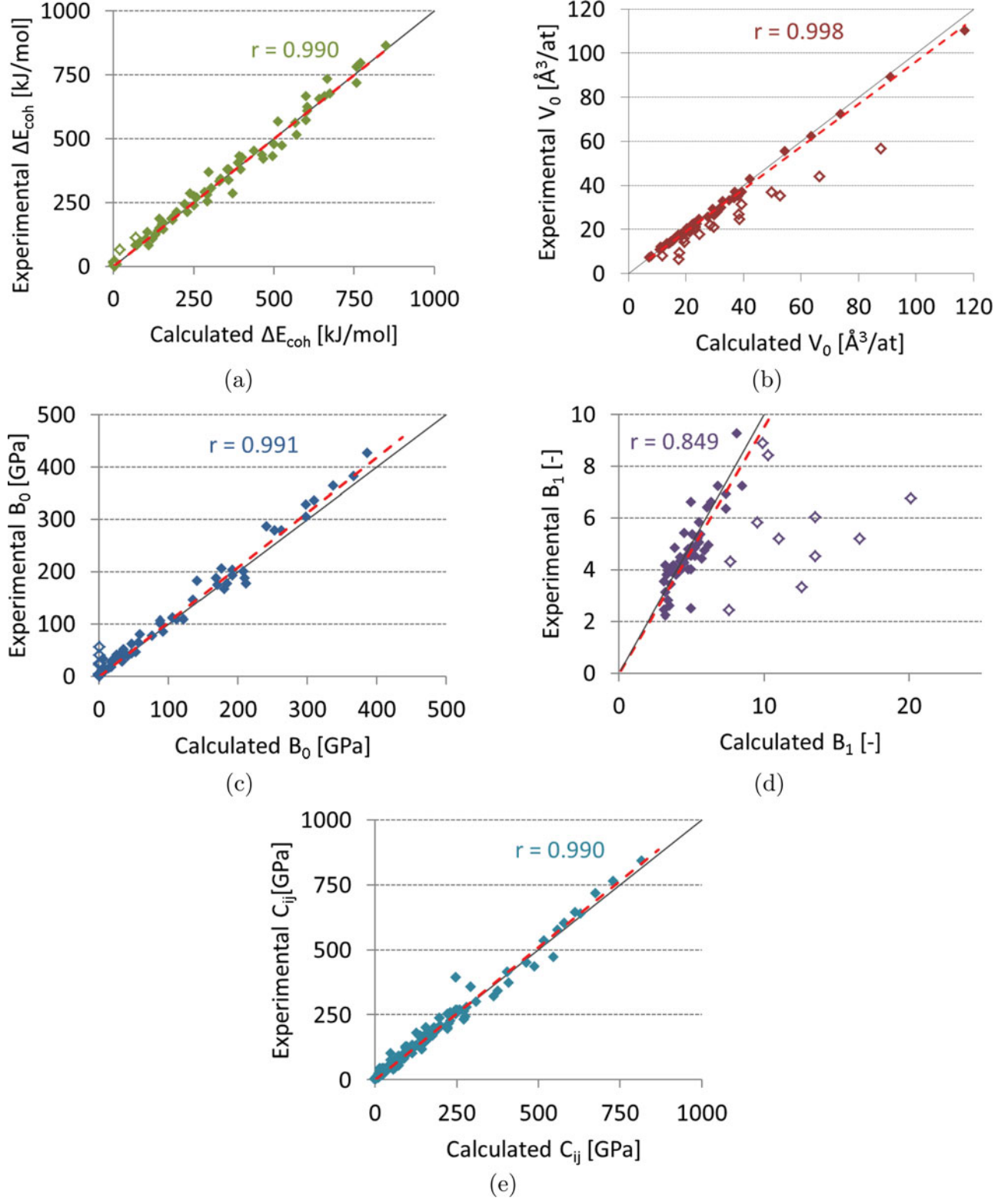| H 8%																	He 125%
Li 11%	Be 19%											B 19%	C 97%	N 70%	O	F	Ne 60%
Na 2%	Mg 1%											Al 5%	Si 8%	P 77%	S 98%	Cl 97%	Ar 75%
K 1%	Ca 16%	Sc 28%	Ti 10%	V 15%	Cr 9%	Mn 27%	Fe 11%	Co 10%	Ni 6%	Cu 1%	Zn 6%	Ga 16%	Ge 22%	As 35%	Se 57%	Br 94%	Kr 83%
Rb 18%	Sr 3%	Y 16%	Zr 16%	Nb 4%	Mo 0%	Tc 3%	Ru 3%	Rh 4%	Pd 6%	Ag 11%	Cd 24%	In 16%	Sn 12%	Sb 18%	Te 29%	I 92%	Xe 85%
Cs 11%	Ba 12%	Lu 18%	Hf 2%	Ta 0%	W 5%	Re 1%	Os 4%	Ir 2%	Pt 11%	Au 19%	Hg	Tl 25%	Pb 18%	Bi 26%	Po 30%	At	Rn

FIG. 3. Linear regression (dashed red) between the (thermally corrected) experimental and theoretical results (VASP-PBE / -PW91, see text) for the cohesive energy,[46,74] equilibrium volume,[43,75–82] bulk modulus,[46,62,83–90] pressure derivative of the bulk modulus,[91–97] and elastic constants.[40,98,99] The full, black line stands for $x_{exp} = x_{th}$. r represents the Pearson product-moment correlation coefficient for all elements included in the regression (filled symbols). The criterion for excluding certain elements from the fit (open symbols) is discussed in the text. (Color figure available online.)

TABLE 3

Relative residual errors (of the VASP-PBE/-PW91 regression results with respect to the thermally corrected experimental values) for B_1[91–97] (purple) and C_{ij}[40,98,99] (PW91, cyan) of the elemental crystals. The darkest shades correspond to the largest errors (color figure available online)

B_1

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
H																	He
Li 14%	Be											B	C 6%	N	O	F	Ne 16%
Na 25%	Mg 22%											Al 1%	Si 8%	P 187%	S 184%	Cl 102%	Ar 10%
K 14%	Ca 1%	Sc 16%	Ti 1%	V 12%	Cr 2%	Mn 8%	Fe 2%	Co 2%	Ni 18%	Cu 2%	Zn 25%	Ga	Ge 3%	As 263%	Se 57%	Br 204%	Kr 12%
Rb 9%	Sr 92%	Y 39%	Zr 31%	Nb 5%	Mo 4%	Tc	Ru 29%	Rh 11%	Pd 5%	Ag 19%	Cd 21%	In 1%	Sn 15%	Sb 70%	Te 17%	I 115%	Xe 12%
Cs 16%	Ba 22%	Lu	Hf 17%	Ta 1%	W 6%	Re 20%	Os 6%	Ir 3%	Pt 0%	Au 10%	Hg	Tl 9%	Pb 9%	Bi 203%	Po	At	Rn

C_{ij}

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
H																	He
Li 7%	Be 6%											B	C	N	O	F	Ne
Na 5%	Mg 8%											Al 8%	Si	P	S	Cl	Ar
K 7%	Ca 2%	Sc 6%	Ti 9%	V 29%	Cr 31%	Mn	Fe 15%	Co 12%	Ni 7%	Cu 2%	Zn 26%	Ga	Ge	As	Se	Br	Kr
Rb 6%	Sr 10%	Y 9%	Zr 10%	Nb 14%	Mo 7%	Tc 16%	Ru 3%	Rh 1%	Pd 9%	Ag 12%	Cd 23%	In	Sn	Sb	Te	I	Xe
Cs 11%	Ba 15%	Lu	Hf 6%	Ta 6%	W 3%	Re 8%	Os 4%	Ir 4%	Pt 17%	Au 23%	Hg	Tl 24%	Pb 36%	Bi	Po	At	Rn

subsets, on the other hand, have been represented by an open symbol.

3.2.3. Predicting Experiment

Using all crystals that survived this selection procedure (filled symbols in Figure 3), a least-squares linear regression can now be computed for all properties from Section 2.1. Table 4 summarizes the resulting intrinsic errors in terms of relative systematic deviations $(1 - \beta)$ and residual error bars (SER). Systematic errors are mentioned for PBE with respect to experiment, so a positive number implies PBE to overestimate that property. Each percentage is expressed relative to the PBE result, allowing for a straightforward calculation of the regression value $(x_{reg} = x_{th} - (1 - \beta)x_{th})$. Between brackets the significance level of $\beta = 1$ is mentioned. It represents the two-sided p-value when a null hypothesis of $\beta = 1$ is assumed: if there

really were no deviation at all, a small p-value would indicate that finding an even more extreme result would be highly unlikely. For the residual error bars a 95% confidence interval is given. The last column lists the subsets of elements that were excluded from the regression analysis by the selection procedure described before. For this the naming convention from Figure 2 has been used.

This statistical treatment makes several implicit assumptions. One of the most important premises is the use of a relative error over a constant offset. After all, for strictly positive quantities, such as V_0, an invariable shift seems counterintuitive. The impact of such an error is much larger when the investigated property is small. In addition, when using relative systematic deviations, the difference from $\beta = 1$ indeed matters. For the equilibrium volume ($p = 6 \cdot 10^{-11}$) and the bulk modulus ($p = 5 \cdot 10^{-4}$) the deviation from the bisector $x_{exp} = x_{th}$ is clearly significant. For other properties this is not always that obvious, but due to physical connections with V_0 and B_0, it is relevant to consider systematic deviations there as well.

A second remark concerns the nature of the intrinsic residual error bars. The numbers in Table 4 were computed assuming a normal distribution for the random error ϵ. This results in an absolute residual error bar. Indeed, a Pearson's χ^2-test does not contradict the applicability of a normal distribution to the intrinsic random errors. A null hypothesis assuming a Gaussian distribution around zero yields a 1-sided p-value of 0.21 for the volume. Within the DFT community, however, relative error bars are often implicity assumed. As a consequence, small volumes are expected to be predicted much more accurately, for example. According to our analysis, the matter appears to be not that simple, as is shown in Figure 4. In Figure 4a, the relative residual errors are plotted against the volume obtained by the least-squares fit. The overall decreasing trend suggests that the
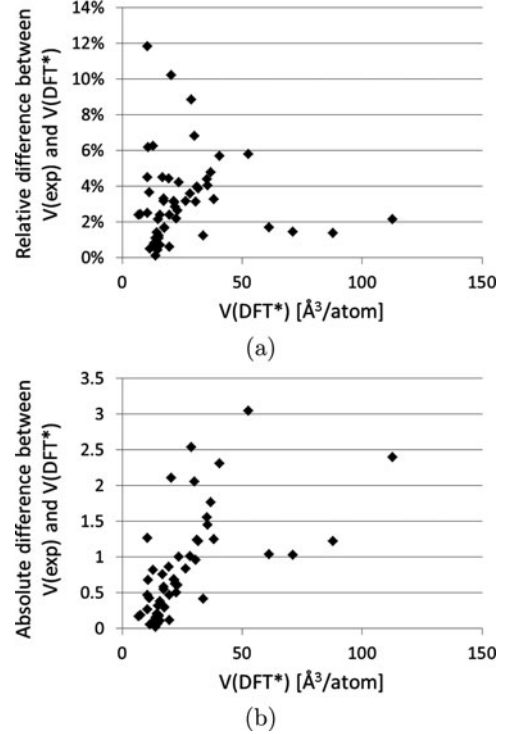

FIG. 4. Relative (a) and absolute (b) differences of the DFT regression data (DFT*) with respect to thermally corrected experimental values[43,75–82] for the equilibrium volume.

DFT errors are best described in terms of absolute error bars. Figure 4b displays the absolute residual errors instead. A rough linear correlation emerges, that implies that on the contrary a relative error bar is more appropriate. Both conclusions are compatible, by assuming a relative residual error bar of about 3% for small to median volumes (<50 Å^3/atom) and an absolute residual error bar of approximately 1.5 Å^3/atom for larger volumes (with the exception of a few outliers[†]). The number of data points is not sufficiently large for this finding to be really convincing, however. We therefore prefer to adopt the overall absolute error bar of 1.1 Å^3/atom (Table 4) as being valid for all volumes. Admittedly, for most small volumes the intrinsic residual error bars are then overestimated.

For any given material, the information in Table 4 can now be used to determine a meaningful estimate of a certain property. As an example, we consider the bulk modulus of diamond. This material is not included in the test set (the ground-state crystal

TABLE 4

Systematic deviations $1-\beta$ and intrinsic error bars (SER) for the VASP-PBE/-PW91 (see text) properties presented in Tables 2 and 3, compared with experiment. The significance of the systematic deviation from $x_{exp} = x_{th}$ is indicated between brackets, by means of the two-sided p-value (low p-values for high significance). For the standard error of the regression a 95% confidence interval is given in terms of the upper and lower limit (superscript and subscript). Subsets containing a lot of outliers have been excluded from the data set by means of the procedure mentioned in the text (notation from Figure 2)

	$1-\beta$	SER	Excluded
ΔE_{coh} [kJ/mol]	-0.0% (0.99)	30^{+7}_{-4}	4, 8
V_0 [Å^3/atom]	$+3.6\%$ (10^{-10})	$1.1^{+0.2}_{-0.2}$	4, 7, 8
B_0[GPa]	-4.9% (10^{-3})	15^{+4}_{-2}	7, 8
B_1[–]	$+4.8\%$ (0.03)	$0.7^{+0.2}_{-0.1}$	6, 7
C_{ij} [GPa][40]	-2.0% (0.01)	23^{+3}_{-2}	—

[†]The fitting error is primarily correlated to the bond type. Although badly performing subsets of materials have already been excluded from the test set, some crystal types are described significantly better by DFT than others, especially those with strong bonds (covalent p-bonds, half-filled d-shell elements). Their strong bonds also lead to more compact structures, which explains the smaller errors for smaller volumes (Figure 4b). Other subsets of crystals with similar volumes perform worse, however, and give rise to the larger relative errors in Figure 4a.

structure of carbon is graphite), but it can be assigned to subset 5, the high-coordination *p*-block compounds (similar to Si, Ge, and Sn). Subset 5 does not belong to one of the excluded subsets for B_0, and the intrinsic error for the bulk modulus prediction of diamond should therefore be representative. A bare VASP-PBE computation yields 434.8 GPa. When taking into account that PBE bulk moduli are systematically too small by 4.9% (Table 4), this value increases to 456.0 GPa. By means of Equations (12), (10), and (6), zero-point corrections (-17.6 GPa) are added back in, yielding 438.4 GPa. Using the appropriate intrinsic residual error bar of 15 GPa (Table 4), the final result becomes 438 ± 15 GPa. This is the most accurate DFT-PBE prediction of the experimental bulk modulus at 0 K, including an error bar on the computed value. In comparison, the experimental value extrapolated to absolute zero (Equations (11) and (8)) amounts to 443 GPa.[46] This number remains neatly within the error bar and is indeed closer to the regression-corrected bulk modulus than to the bare DFT value. A similar procedure can be used for all properties in Table 4. Table 5 offers a few more examples.

Table 4 is based on elemental solids only. One has to verify that the results from this statistical analysis are transferable to multicomponent materials. A good test case in that respect is the collection of thirty-one binary compounds for which Haas et al.[37,38] calculated lattice parameters by means of PBE. When we take their DFT results, the experimental volume falls outside the confidence interval for seven crystals. In all of these cases, the PBE volume is too large. By taking into account the systematic overestimation of 3.6%, however, only in two cases the experimental value exceeds the predicted range. Both of these compounds are ionically bound (NaF and NaCl), a bond type that has not been considered in any of the proposed subsets (Figure 2). Although the multicomponent test set may therefore not contain all materials types and although it only relates to the atomic volumes, these results already strongly support the transferability of our error estimates.

For one example from Haas et al., GaAs, the intrinsic error contributions are listed in Table 5. It illustrates that the systematic deviation really matters if it is the goal to get as close as possible to the experimental value. This also appears from the mean absolute difference between experiment and the 31 theoretical predictions by Haas et al. When one does not ap-

ply the relative deviation from Table 4, this number amounts to $0.72\,\sigma_V$ (0.78 Å³/atom), while for the regression values it is only $0.37\,\sigma_V$ (0.40 Å³/atom).

3.3. Agreement with Experiment

3.3.1. Errors per Materials Type

Because Tables 2 and 3 are shaped like the periodic table, the color code immediately allows singling out the areas where PBE breaks down. It leads to a number of subsets (Figure 2) which can be eliminated from the test set. They are listed in Table 4. For these elements, PBE performs significantly worse than usual, mostly because some key physical phenomenon is not described (well) by the functional. In this subsection, these localized error zones are discussed in more detail, as well as the mechanisms on which they are based.

A first, well-known example of the failure of PBE is the class of **dispersion-governed compounds**. Although some more advanced DFT approaches address this issue specifically,[114,115] regular GGA functionals do not describe London forces. This translates into a decreased cohesion, and hence an inflated volume and underestimated bulk modulus (see also Section 3.3.2).

Although London forces have been demonstrated to play a role in other structures as well,[116–118] the most important crystals that suffer from this shortcoming, belong to the nonmetals (subsets 7 and 8). They include the noble gases, the dimeric crystals, graphite, and sulfur. In these materials the London dispersion interaction governs the bonding between atoms, diatomic molecules, graphene sheets, and 8-membered rings, respectively. Nevertheless, it is essential to realize that both the element type and the crystal structure contribute to the importance of dispersion. It is perfectly plausible that a certain element behaves badly in structure A, while there are no problems when it assumes structure B. This can be illustrated nicely by means of carbon. The dispersion forces between the graphene layers in graphite give rise to a large discrepancy between DFT and experiment. Diamond on the other hand follows the same behavior as neighboring (semi)metallic elements or even outperforms them (Table 6).

For the molecular crystals (subset 7) the PBE cohesive energy is larger than the experimental value, contrary to the expectation. This is due to the overestimation of the intramolecular

TABLE 5

Demonstration of the use of intrinsic errors (Table 4) to improve DFT-PBE predictions and assess their reliability

	V_0(W) [Å³/atom]	B_0(diamond) [GPa]	V_0(GaAs) [Å³/atom]
PBE (bare)	16.28	434.8	23.73[37]
systematic deviation	15.69 (-3.6%)	456.0 ($+4.9\%$)	22.87 (-3.6%)
zero-point correction	15.71 ($+0.02$)	438.4 (-17.6)	23.00 ($+0.13$)[37]
residual error bar	**15.7 $\pm$ 1.1**	**438 $\pm$ 15**	**23.0 $\pm$ 1.1**
experiment (0 K)	**15.8**[37]	**443**[46]	**22.5**[37]

TABLE 6

Residual errors (of the VASP-PBE regression results with respect to the zero-kelvin extrapolated experimental values[43,46,73,74,83]) for two allotropes of carbon

	ΔE_{coh} [kJ/mol]	V_0 [Å³/at]	B_0[GPa]
graphite	39 (5%)	3.2 (39%)	54 (97%)
diamond	13 (2%)	0.1 (2%)	5 (1%)

bond strength (see, e.g., Lany[36] and Table II of Paier et al.[20]), which covers up any influence of the lack of dispersion. Elastic properties, on the other hand, are in most cases not affected by intramolecular effects and show a similar behavior as for the remaining nonmetals.

The **magnetic materials** (subset 3) stand out as well, predominantly with respect to ΔE_{coh}. Although the use of the generalized gradient approximation and a correct atomic reference state have already reduced the gap between theory and experiment substantially,[48] the remaining difference cannot be neglected. Current GGA funtionals are not able to describe magnetic compounds very well. Manganese illustrates this nicely. Its intricate magnetic state[119] has been approximated by assuming only collinear magnetism, but this does not explain the observed differences. The cohesive energy, for example, would be higher in its correct ground state, leading to an even more pronounced deviation from experiment. An explanation is found with Singh and Ashkenazi,[120] who noticed that GGAs overestimate the magnetic energy. This is caused by the increased number of degrees of freedom in spin-polarized systems (two spins), while the number of physical relations the GGA must fulfill stays the same.

The discrepancies between theory and experiment are not caused by the DFT functional alone, however. For some magnetic elements the applied thermal extrapolations are no longer valid, because of phase transition effects in the vicinity of the Curie or Néel temperature. Experimental chromium is a good example, displaying large magnetic distortions of the thermal expansion coefficient near 311 K.[121] A relation as simple as that of Equation (8) cannot capture these complex underlying phenomena.

The transition metals with (nearly) full d shells sometimes deviate from experiment as well (subset 4). The effects are smaller than for the previous two classes of materials, but they are unmistakably present, especially in terms of the cohesive energy and the elastic constants. One can attribute this phenomenon to **electronic correlation**. For Zn, Cd, and Hg a full-fledged many-body treatment has indeed convincingly shown the influence of d electron correlation on ΔE_{coh} and the potential energy landscape.[13,122,123] Data in Tables 2 and 3 even imply that similar (but smaller) effects show up in other elements at the end of the d block, such as in Pd, Ag, Pt, and Au. In noble metals, dispersion phenomena play an important role too,[117] however, and it is not immediately clear how much of the remaining discrepancy

can be attributed to electronic correlation. Since the influence of correlation in these elements appears limited, only Cd an Hg have been assigned to subset 4.

It seems that at the end of the d block, the high number of localized d electron pairs in combination with a small interatomic distance and a close-packed environment enhances correlation effects. Moreover, Philipsen and Baerends[48] suggested that at the very beginning and the very end of the $3d$ transition metals the GGA exchange energy drops, causing the electron correlation to gain importance. Any fortuitous cancellation of errors between exchange and correlation has therefore disappeared for the transition metals at the border of the d block.

These correlation effects appear to be of a mainly anisotropic nature, since all elastic constants C_{ij} are affected, but the deviations of V_0 and B_0 are less pronounced. This is also suggested by Wedig et al.[123] for Zn and Cd, where a different interlayer and intralayer behavior is observed.

Relativistic effects are expected to strongly influence heavy elements. VASP therefore makes use of the scalar-relativistic Kohn-Sham equations by default.[124] The major remaining contribution is due to spin-orbit coupling. However, it is shown by Philipsen and Baerends[125] that this does not change physical properties substantially. Only for gold and bismuth a distinct change is reported, but without closing the gap between theory and experiment entirely. The remaining difference for Au is primarily due to correlation and dispersion effects, as was already suggested above. For the $6p$ elements, on the other hand, spin-orbit coupling really plays an important role. Some key properties for the $5d$ and $6p$ compounds have been calculated, both with and without spin-orbit coupling (Table 7). It is immediately clear that, starting from the end of the $5d$ block, a spin-orbit treatment becomes indispensable. Hence, for all $5d$ and $6p$ elements this contribution has been included, except for C_{ij}.[40]

3.3.2. Errors per Property

The previous section shows that PBE is not "complete": some features just cannot be described by a simple GGA functional. One can exclude the affected materials (outliers) beforehand, however, and limit the analysis to those cases where PBE should perform well. The intrinsic errors from Table 4 are applicable to these crystals. Table 4 then shows that the PBE error estimates largely depend on what property is considered. The behavior of the residual error bar and the systematic deviation from experiment can be traced back to both the functional and the numerical determination of that particular property. Nevertheless, it is important to note that, although the overall error estimate can be linked to theoretical aspects, the correspondence to experiment for a single compound depends on the experimental accuracy as well. This is especially true for higher-order properties (such as the elastic constants and their derivatives), which are generally measured at a lower precision than those from (quasi)direct measurements (the lattice constants or

	ΔE_{coh}		V_0		B_0	
	n-SO	SO	n-SO	SO	n-SO	SO
Ag	16%	16%	3%	3%	11%	11%
Lu	7%	8%	3%	3%	18%	18%
Hf	1%	3%	3%	3%	2%	2%
Ta	1%	2%	2%	2%	1%	0%
W	0%	1%	1%	1%	3%	5%
Re	3%	2%	2%	1%	3%	1%
Os	1%	3%	0%	1%	1%	4%
Ir	4%	0%	0%	0%	0%	2%
Pt	7%	9%	1%	1%	9%	11%
Au	22%	19%	4%	3%	21%	19%
Hg	75%	69%	29%	22%		
Tl	6%	20%	9%	7%	27%	25%
Pb	44%	4%	4%	4%	10%	18%
Bi	15%	5%	2%	5%	17%	26%
Po	59%	8%	2%	3%	73%	30%
Rn	83%	81%				

the cohesive energy, for example). This is illustrated by the sometimes large spread on the data in Knittle's overview of B_1 values.[91]

From a computational viewpoint, however, the equilibrium **volumes** offer the best results among all considered quantities. After eliminating the outliers (listed in Table 4 and represented in Figure 3b by open symbols), an almost perfect correlation is obtained. Even so, the regression line does not coincide with the first quadrant bisector. Table 4 shows that the cell volumes are consistently too large by approximately 4%. This deviation is a well-known property of any GGA,[126] including PBE. It originates in a systematic underestimation of the bond strength (underbinding), resulting in slightly larger volumes. More particularly, GGAs favor inhomogeneous systems, with large (reduced) density gradients. Small unit cells, which have a more evenly distributed electron density, are therefore energetically less preferable. This phenomenon especially affects open structures, where the high-gradient tails of the valence electron orbitals become nonnegligible.[127]

Immediately linked to this observation is the underestimation of the **bulk modulus**. More weakly bound structures will be more easily compressible, leading to smaller B_0 values. Just like the too large predicted volumes, it is common behavior for GGA functionals.[126] On the other hand, PBE bulk moduli are predicted with a larger uncertainty than the volumes. The intrinsic residual error remains in most cases below 10–15%,

however (Table 2). The magnitude of this difference is mostly due to the sensitivity of the $E(V)$ curvature.

Since B_0 and the other **elastic constants** are closely related, the intrinsic errors with respect to the C_{ij} parameters are of a comparable scale. The bulk moduli are larger on average, which leads to slightly smaller relative errors (Tables 2 and 3). However, a good correlation is found in both cases, with a similar value of r for the elastic constants (Figure 3e) and the bulk moduli (Figure 3c).

For the **cohesive energy** PBE yields very good results as well. The intrinsic error bar of 30 kJ/mol is of the same order of magnitude as the rms error found by Lany[36] for PBE heats of formation of semiconductors and insulators (0.24 eV/atom). It is therefore representative for PBE energy differences between chemically different compounds. As mentioned before, for similar systems the intrinsic residual error bar is much smaller.[42] This also explains the success of evolutionary algorithms. They are based on energy differences of the order of a few meV per atom,[128–131] but some results have already been confirmed experimentally nevertheless.[9]

Contrary to V_0 and B_0, there is now no systematic under- or overestimation compared to experiment. The typical underbinding of GGA does not show as conclusively in ΔE_{coh}. This is due to the magnetic materials and the molecular crystals. As mentioned before, GGA functionals bias solutions towards magnetism for the former and overestimate the intramolecular contribution in the latter. In both cases this causes the cohesive energy to oppose the dominating trend. Without the crystals from subsets 3 and 7 in the test set, the cohesive energies would have been underestimated by 2% instead (p-value of 0.008). This behavior is in accordance with the expected underbinding of GGA.

Since the **bulk modulus derivative** B_1 is a higher-order parameter than B_0, the errors are expected to be one order worse as well. Although this is certainly the case, eliminating the outliers substantially improves the results (Figure 3d). However, even when they are removed, the resulting correlation coefficient ($r = 0.849$) remains significantly lower than for any other property already discussed.

B_1 appears to be overestimated with respect to experiment. This systematic deviation is significant, although the p-value may not show it conclusively (Table 4). It is again caused by GGA underbinding. As mentioned before, large volumes are favored due to their substantial density gradients. GGA hence lowers the energies of bigger cells most and straightens out the equation of state. This causes the $E(V)$ line to alter its decreasing curvature even more rapidly, increasing the rate of change of the bulk modulus with pressure (and volume), B_1. It also explains the deviating behavior of crystals with a low coordination, such as the molecular crystals. In these compounds the tails of the electron wave functions dominate the interstitial space, leading to considerable density gradients. The increase of the sensitive parameter B_1 is then enhanced even further.

4. NUMERICAL ERRORS

The previous section describes the intrinsic PBE errors for five different properties, based on a statistical treatment. They are used in a protocol which allows experimentalists and theoreticians to correct the bare DFT-PBE values for the observed systematic deviation from experiment and which quantifies the uncertainty on the obtained predictions (Table 5). A prerequisite for such a protocol is that different DFT implementations provide the same predictions: using different algorithms to solve the same (Kohn-Sham) equations should ideally lead to identical solutions. In practice, different amounts of noise are inevitably introduced in the predictions, even when numerical convergence has been achieved for each individual code. This scatter is due to several aspects of the solution algorithm. It can be due to its nature (e.g., the kind of basis set or the frozen-core approximation), its specific ingredients (e.g., the chosen pseudopotential) or its use of particular routines for standard tasks (e.g., Fourier transform routines). In order to guarantee the reproducibility of the intrinsic errors in Table 4, one needs to examine to what extent these issues affect the DFT computations. Once again, a reliable benchmark can be established using the ground-state elemental crystals.

The current section describes a procedure to express the difference between predictions from independent solid-state DFT approaches in a quantitative way, yielding a numerical error estimate. It will be used to examine differences between the PAW and APW+lo method, representing both methods by suitable mainstream codes (see further). However, the difference between codes can be attributed to other aspects as well, as was mentioned earlier. The influence of standard task routines is most likely small, but we will also use two PAW codes with different PAW atomic potentials, which can have quite drastic effects on the DFT results. All codes are therefore considered here with their recommended potentials. These can be thought of as representative for the quality of the investigated code. Although it is not our primary goal here, the procedure that will be described can also be used to select better performing PAW atomic potentials. The same holds for pseudopotentials in the case of plane-wave codes.

4.1. Test Set Preparation

The present implementation assessment starts from one reference code, the all-electron program WIEN2k[132] (version 11.1). It uses the APW+lo basis set,[133, 134] which is considered to be a standard for the numerical accuracy of solid-state DFT. WIEN2k predictions can therefore be considered to yield the exact results for a given functional, as long as numerical accuracy is achieved[37] (large basis set and dense k-mesh, see Supplementary Material for more details). Two codes are compared to this reference code: VASP[8,67] (version 5.2.2) and GPAW[135–137] (version 0.8.0), both using the PAW method.[65] GPAW calculates all wave functions, densities, and potentials as grid-based quantities, while VASP uses a plane-wave basis set. All calculations

with these programs are performed by means of the potentials recommended by the respective developers: the 2010 recommended PAW potentials for VASP and the 0.6 atomic set-ups for GPAW. Detailed computational parameters are summarized in the Supplementary Material.

For reasons of uniformity and comparability, the same PBE functional has been selected for all three codes. It is used in a protocol that seeks to evaluate a particular DFT approach in an easily reproducible manner. The VASP-optimized ground-state crystal (Section 3.1) serves as a starting point for each computation and from a 7-point equation of state (0.94–1.06 V_0) the properties of interest (E_0, V_0, B_0, and B_1) are extracted. All geometries are kept frozen (the cell shape and relative atomic positions are kept fixed at their initial values), instead of allowing for relaxation changes. This not only lowers the computational load, it also restricts the code evaluation to the implementation of DFT-PBE itself. Indeed, the task of optimizing the cell shape or internal positions belongs to another computational layer, on top of the task of solving the DFT equations for a given rigid geometry. This section aims to examine how different implementations compare with respect to the DFT-PBE procedure only. It does not intend to study how close every individual approach comes to experiment.

In the same spirit, some other modifications of the hitherto employed test set have been made. All calculations have been limited to the scalar-relativistic part (using the Koelling-Harmon approach[124]). By neglecting the spin-orbit contribution, an additional secondary algorithm implementation is avoided. The computational procedure also becomes more uniform this way, since all elements are now treated on equal terms. Because no spin-orbit coupling is added to the system's Hamiltonian, it suffices to use non spin-orbit geometries as a starting point for the seven-point equation of state.

A simplified unit cell has been selected for Mn and S as an additional means of lowering the computational effort. Manganese is treated in an antiferromagnetic fcc phase (space group 225, cF4), while for sulfur the β Po phase is imposed (space group 166 or hR3). These geometries are physically relevant, as they can be found in the Mn and S phase diagram, respectively.[138,139]

All other elements have been kept at the structure previously optimized by VASP (Section 3.1), in order to conserve the large diversity of the input set. The CIF files for all crystals in this code benchmark set are available in the Supplementary Material.

4.2. Agreement between Implementations

The procedure mentioned above results in a large collection of numbers for each code (71 elements × 4 properties). It is not convenient to compare them directly, however. Because of the different units involved, a coherent approach would require the use of relative deviations. Tables 2 and 3, on the other hand, show that each property corresponds to a different magnitude of relative error. This scale is mainly determined by the computational procedure and therefore does not alter substantially
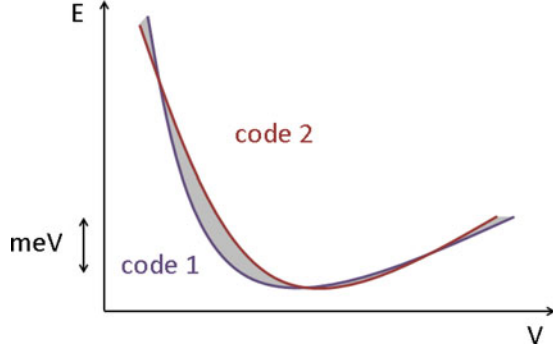
FIG. 5. The EOS parameters can differ significantly, while the $E(V)$ curves themselves are very similar. In that case the area between the two functions is a better indicator of the overall deviation. (Color figure available online.)

when shifting from a code-experiment comparison to an inter-code assessment. A single numerical error value, expressing the difference between two particular DFT methods by means of one number, can be obtained by applying a weighted average. As all properties of interest depend on the equation of state, it is most straightforward to compare the $E(V)$ curves produced by different approaches directly. The dispersion-governed compounds illustrate this strategy well. Since their $E(V)$ curves are very shallow, small deviations in the bulk modulus will inflate the relative error considerably. However, the equations of state as such can be very similar, the two curves at no point differing by more than a few meV per atom (Figure 5). For that reason a numerical error estimate Δ is defined as follows:

$$\Delta = \left\langle \sqrt{\frac{\int \Delta E^2(V) \mathrm{d}V}{\Delta V}} \right\rangle. \quad [15]$$

In other words, the rms energy difference between the $E(V)$ curves of these particular programs is averaged over all elemental crystals. Δ hence provides an intuitive measure of the energy distance between equations of state.

Because different codes sometimes employ different reference energies E_0, depending on the concept, all equations of state are set to zero at their equilibrium volume. An alternative solution would entail the calculation of cohesive energies, in order to provide a common reference for the equilibrium energy. However, not all programs allow for an easy manipulation of the electronic configuration of atoms. Moreover, the computational load would increase considerably.

The computation of Δ can be automated quite easily. The fitted Birch-Murnaghan equation allows Equation (15) to be written in an analytical form. Only V_0, B_0, and B_1 are then needed for both codes under investigation. The resulting expressions have been added in the Appendix for convenience. The WIEN2k data necessary for a code comparison have been provided in the Supplementary Material.

The interval of integration is linked to the reference data. In view of how the $E(V)$ parameters are determined, the intercode difference is to be integrated between $V_{0,\text{WIEN 2k}} \pm$

6%. ΔV hence corresponds to $0.12\,V_{0,\text{WIEN 2k}}$. By definition, $\Delta(\text{APW+lo})_{(\text{WIEN2k})}$ becomes zero.

The rms energy differences between the equations of state predicted by APW+lo$_{(\text{WIEN2k})}$ and PAW$_{(\text{VASP})}$, or APW+lo$_{(\text{WIEN2k})}$ and PAW$_{(\text{GPAW})}$, are represented in Table 9. They show that most critical elements are characterized by approximately half-filled d levels. Such numerical errors can amount to up to 8.3 meV/atom for PAW$_{(\text{VASP})}$ (Tc) and 20.9 meV/atom for PAW$_{(\text{GPAW})}$ (Ru). This agrees with physical intuition, because these crystals are among the least compressible. Their equations of state are very steep, and relatively small modifications of the parameters can strongly change the energy. The least sensitive elements are for the same reason located near the alkali metals and the noble gases (0–0.7 meV/atom numerical error) (see Table 8). Only in comparison to experiment the latter group of materials stands out, but this is because PBE grossly overestimates the rare-gas volumes.

When averaging the numbers in Table 9 over all elements, the numerical error of each DFT approach can be determined for the given set of recommended PAW potentials. Δ is 1.9 meV/atom for PAW$_{(\text{VASP})}$, while for PAW$_{(\text{GPAW})}$ it is 3.3 meV/atom. This agreement between implementations is an order of magnitude better than the difference with experimental results. To show this, a similar energy difference between DFT-PBE and experiment is computed. It uses experimental values as the reference situation, while the method under test is the full-fledged version of PAW$_{(\text{VASP})}$. This means that the $E(V)$ parameters have been taken from Tables 2 and 3. The deviations per element are presented in Tab. 9, leading up to a Δ-factor of 23.5 meV/atom. This difference in magnitude can also be observed with the $E(V)$ characteristics themselves. Figure 6 shows the distribution of volume errors between two codes and with respect to experiment. Again, the spread is much larger in the latter case.

$\Delta(\text{PAW})_{(\text{VASP})}$ does not change noticeably when the number of elements is reduced to that of PAW$_{(\text{GPAW})}$. This shows that GPAW and VASP, while both using the same PAW method, do not produce entirely identical results. This variation most likely originates in the different quality of the atomic potentials and the different type of basis functions used. However, in comparison to

TABLE 8
Comparison between codes for two extreme situations: large (Tc an Ru) and small (Ar) numerical errors Δ_i. V_0 is given in Å^3/atom, B_0 in GPa, and Δ_i in meV/atom. B_1 is dimensionless

		V_0	B_0	B_1	Δ_i
Tc	APW+lo$_{(\text{WIEN2k})}$	14.47	301.4	4.56	0
	PAW$_{(\text{VASP})}$	14.60	298.5	4.55	8.3
Ru	APW+lo$_{(\text{WIEN2k})}$	13.81	315.4	4.96	0
	PAW$_{(\text{GPAW})}$	14.09	310.9	4.87	20.9
Ar	APW+lo$_{(\text{WIEN2k})}$	52.21	0.7	7.84	0
	PAW$_{(\text{VASP})}$	52.65	0.8	7.35	0.1
	PAW$_{(\text{GPAW})}$	52.66	0.8	3.27	0.1

TABLE 9

Rms energy differences Δ_i between the equations of state predicted by APW+lo$_{\text{(WIEN2k)}}$ and PAW$_{\text{(VASP)}}$ (green), APW+lo$_{\text{(WIEN2k)}}$ and PAW$_{\text{(GPAW)}}$ (red), and experiment and PAW$_{\text{(VASP)}}$ (blue) for the ground-state elemental crystals. All values are expressed in meV per atom. The darkest shades correspond to the largest errors. The average numerical error Δ is shown for each code at the header of the table (color figure available online)

$\Delta(\text{PAW})_{\text{(VASP)}} = 1.9$ meV/atom

H																	He
0.1																	0.0
Li 0.2	Be 0.1											B 0.3	C 0.3	N 10.6	O 8.3	F 1.5	Ne 0.1
Na 0.0	Mg 0.7											Al 0.3	Si 2.0	P 3.8	S 3.3	Cl 4.0	Ar 0.1
K 0.1	Ca 0.2	Sc 0.4	Ti 0.9	V 1.3	Cr 3.1	Mn 1.4	Fe 3.4	Co 3.4	Ni 2.0	Cu 0.4	Zn 0.3	Ga 0.2	Ge 2.4	As 1.7	Se 1.5	Br 1.5	Kr 0.1
Rb 0.1	Sr 0.1	Y 0.5	Zr 2.7	Nb 7.3	Mo 5.5	Tc 8.3	Ru 2.3	Rh 5.4	Pd 4.4	Ag 4.1	Cd 1.4	In 0.4	Sn 0.2	Sb 0.1	Te 0.5	I 0.9	Xe 0.1
Cs 0.3	Ba 0.7	Lu 4.3	Hf 1.2	Ta 1.0	W 3.5	Re 4.3	Os 3.8	Ir 1.9	Pt 2.5	Au 5.9	Hg 0.5	Tl 0.4	Pb 0.6	Bi 0.4	Po 0.4	At	Rn 0.0

$\Delta(\text{PAW})_{\text{(GPAW)}} = 3.3$ meV/atom

H																	He
0.2																	0.0
Li 0.2	Be 2.6											B 0.2	C 0.3	N 0.2	O 10.4	F 1.0	Ne 0.2
Na 0.1	Mg 0.3											Al 0.2	Si 0.6	P 1.2	S 2.2	Cl 0.9	Ar 0.1
K 0.0	Ca 1.0	Sc 0.4	Ti 3.7	V 9.3	Cr 2.3	Mn 3.7	Fe 1.0	Co 1.3	Ni 3.1	Cu 2.5	Zn 0.4	Ga 1.7	Ge 1.5	As 1.0	Se 2.3	Br 2.8	Kr 0.0
Rb 0.1	Sr 1.5	Y	Zr 1.2	Nb 3.2	Mo 3.8	Tc	Ru 20.9	Rh 14.5	Pd 3.3	Ag 4.4	Cd 0.1	In 0.3	Sn 0.1	Sb	Te 8.4	I 3.3	Xe
Cs 0.3	Ba 0.6	Lu	Hf	Ta 10.2	W 19.3	Re	Os 18.5	Ir 13.2	Pt 6.3	Au 6.8	Hg	Tl	Pb 0.8	Bi 0.4	Po	At	Rn

$\Delta(\text{exp}) = 23.5$ meV/atom

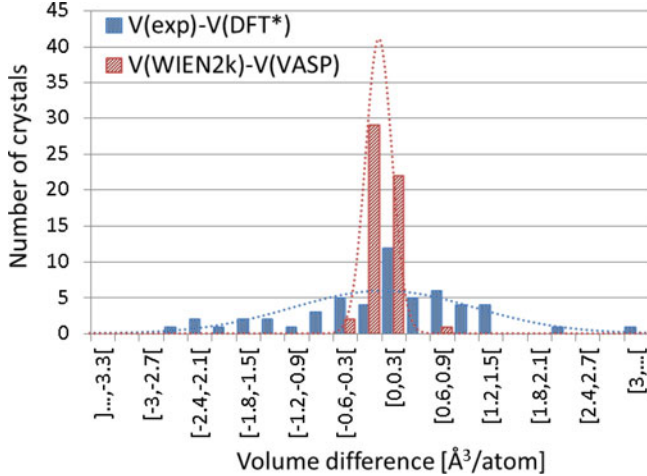| H																	He
Li 0.4	Be											B	C 17.6	N	O	F	Ne 17.1
Na 0.2	Mg 1.5											Al 3.9	Si 13.7	P 19.0	S 44.0	Cl 46.1	Ar 38.6
K 1.4	Ca 3.0	Sc 1.7	Ti 3.0	V 13.1	Cr 1.6	Mn 47.7	Fe 10.6	Co 3.3	Ni 6.2	Cu 13.5	Zn 7.6	Ga	Ge 28.2	As 17.1	Se 11.1	Br 22.3	Kr 54.6
Rb 1.3	Sr 2.4	Y 0.7	Zr 8.2	Nb 14.8	Mo 26.5	Tc	Ru 30.4	Rh 44.6	Pd 49.1	Ag 39.3	Cd 34.0	In 23.6	Sn 40.7	Sb 25.1	Te 9.9	I 68.1	Xe 59.8
Cs 4.0	Ba 2.5	Lu	Hf 2.8	Ta 16.8	W 35.6	Re 29.9	Os 65.2	Ir 50.7	Pt 54.9	Au 64.0	Hg	Tl 42.4	Pb 32.9	Bi 32.0	Po	At	Rn

FIG. 6. Intrinsic (PBE regression vs. experiment) and numerical errors (PAW$_{(VASP)}$ vs. APW+lo$_{(WIEN2k)}$) for the equilibrium volume of the ground-state elemental crystals, using the subsets of elements that have been shown to perform well for PBE (see Table 4). A normal distribution has been fitted to both data sets (dotted line). (Color figure available online.)

experiment, the differences are negligible. The intrinsic residual error bars and regression slopes provided in Table 4 can therefore be applied to DFT-PBE results, irrespective of which approach was used to calculate them.

This comparison of three DFT implementations can easily be extended. Ideally every solid-state DFT approach should be tested in the same way, and have its Δ-value computed. As such tests are preferably performed by specialists in the individual codes, all input CIF files have been made available in the Supplementary Material and at arXiv,[140] as well as some ready-made post-processing scripts. In addition, the ASE developers have implemented a framework for performing the necessary calculations.[135,141] On the CMM website[142] an updated overview will be maintained of all Δ-factors reported to us. Such information not only provides insight into the reproducibility of the intrinsic errors of Table 4, but can also guide users to select a method for a specific task, at least as far as accuracy of energy vs. volume relations is concerned.

5. CONCLUSIONS

Using the ground-state elemental crystals as a test set, DFT-PBE computational errors have been reviewed. Errors intrinsic to the functional were quantified for five materials properties, describing energetic (ΔE_{coh}) and elastic (V_0, B_0, B_1, C_{ij}) quantities. They explain the deviation of DFT predictions from experiment. Numerical errors, due to the implementation of the DFT scheme into a computer code, were studied for the PAW method (VASP and GPAW), and were expressed with respect to the reference APW+lo method (WIEN2k). Both types of errors have been discussed for PBE, one of the most widely applied functionals in solid-state DFT. The results are expected to be representative of GGA in general.

Each of the five properties has been assessed with respect to the ground-state elemental crystals. The correspondence to experiment has been analyzed statistically, leading to a decomposition of the intrinsic error into systematic deviations and residual error bars. These intrinsic errors have been shown to agree with some generally known GGA traits. The typical underbinding of GGA has been reproduced and quantified, for example. Table 10 presents a summary of our results, as well as a similar analysis for other properties that are available from data sets in the literature. Contrary to Table 4, however, Table 10 presents systematic deviations in terms of the experimental value: $(x_{th} - x_{exp})/x_{exp} = 1/\beta - 1$. This expresses more intuitively to what extent DFT varies from experiment: a $1/\beta - 1$ of $+1\%$ for example means that PBE overestimates the experimental result by 1%.

Based on the quantification of intrinsic errors, a computational recipe has been presented which allows to correct bare DFT-PBE results for the systematic deviation from experiment, and which attaches meaningful error estimates to the obtained predictions (Table 5). An examination of 31 binary compounds not included in the benchmark set[37,38] indicates that our analysis carries over to multicomponent crystals. Errors can hence be estimated straightforwardly for PBE predictions already available from literature.

The overall agreement between VASP-PBE and experiment is quite good, but some subsets of elements perform better than others. DFT predictions for magnetic materials and correlation-dominated compounds deviate significantly from experimental values, for example, especially with respect to the cohesive energy. Long-range interaction is another issue. Although some solutions exist to incorporate London dispersion into DFT, such as the DFT-D[114] or vdW-DF2 method,[115] regular GGAs do not describe dispersion-governed crystal types well. Bulk moduli are found to be consistently underestimated, while predictions for both the volume and the pressure derivative of the bulk modulus are systematically too large. Results for heavy elements are acceptable as long as spin-orbit coupling is added, starting from the 6p block. Based on these observations, some general guidelines have been summarized in Table 10 as to what categories of materials will not be described well. Some classes were not or only marginally represented in the elemental benchmark set, such as ionic or strongly correlated compounds. For these, a similar study using an extended benchmark set, by including some binary ionic compounds and transition metal oxides, would be useful.

All conclusions with respect to the intrinsic PBE errors can only be universally applicable when it does not matter how the DFT formalism is implemented. Such numerical errors should be much smaller than intrinsic ones. By means of a quality factor Δ, which conveys exactly this information, APW+lo$_{(WIEN2k)}$ has been compared to PAW$_{(VASP)}$ ($\Delta = 1.9$ meV/atom) and PAW$_{(GPAW)}$ ($\Delta = 3.3$ meV/atom), both for their recommended sets of atomic potentials. The rms energy distance between equations of state from different methods indeed appears to be an order of magnitude smaller than the gap between theory and experiment (see also Table 10). The intrinsic systematic

TABLE 10
Systematic deviations $1/\beta - 1$ and residual error bars for DFT-GGA predictions compared with experiment (intrinsic errors) and between codes (numerical errors). Δ represents the average rms energy difference between the equations of states of two codes. All data were (re)analyzed in the present study, except for the error bar for ΔE_{evol}, which is based on a proof of principle (see Section 3.3.2). Subsets of materials to which the error estimates do not apply are mentioned explicitly

	Intrinsic errors (this work, VASP-PBE)		
	systematic deviation	residual error bar	not applicable to:
ΔE_{coh} [kJ/mol]	-0.0%	30	dominant correlation; pure dispersion
V_0 [Å^3/atom]	$+3.8\%$	1.1	dominant correlation; dispersion
B_0 [GPa]	-4.7%	15	dispersion
B_1 [$-$]	$+5.0\%$	0.7	low coordination number
	Intrinsic errors (other works, VASP-PW91/-PBE/-PBE+U)		
C_{ij} [GPa][40]	-2.0%	23	
ΔE_{form} [kJ/mol][36]	-13.1%	15	
ΔE_{react} [kJ/mol][42]	-4.8%	3.2	chemically dissimilar
ΔE_{evol} [kJ/mol][9,128–131]		$\lesssim 1$	chemically or structurally dissimilar
	Numerical errors (this work, PBE)		
Δ(PAW)$_{(VASP)}$ [kJ/mol]		0.19	—
Δ(PAW)$_{(GPAW)}$ [kJ/mol]		0.32	—
⋮		⋮	

deviations and residual error bars presented in Table 10 can hence be applied to PBE predictions regardless of the computational approach. This is useful when discussing the implications of DFT results in an experimental context. This accuracy review is to be considered as a starting point only.

The presented statistical procedure is applicable to other functionals or methods as well. It would be useful to determine the intrinsic systematic deviations and residual error bars for, e.g., LDA or hybrid functionals. A comparison to results from high-level many-body techniques would even allow for the elimination of the influence of experimental errors. Another extension would be to take into account experimental error bars in the statistical analysis. For the relevant properties of the elemental materials, this requires an extensive literature search for the most accurately known values and their error bars. These error bars are not commonly available in tabulations in the literature and are therefore beyond the scope of this work. With respect to the assessment of numerical errors, we invite both code developers and users to determine the quality factor Δ for their code as well. It not only guarantees the transferability of the intrinsic errors to all codes, but also provides a criterion to evaluate the accuracy of a particular DFT approach. A comprehensive list of Δ-factors[142] can then serve as a guideline through the maze of available DFT methods.

SUPPLEMENTARY MATERIAL

See Supplementary Material posted on the publisher's website, for computational details, the full calculated and experimental data and the files necessary for a code comparison.

REFERENCES

1. P. Hohenberg and W. Kohn, Inhomogeneous electron gas, *Phys. Rev.*, **136**, B864–B871 (1964).

2. W. Kohn and L. J. Sham, Self-consistent equations including exchange and correlation effects, *Phys. Rev.,* **140**, A1133–A1138 (1965).

3. M. D. Segall, P. J. D. Lindan, M. J. Probert, C. J. Pickard, P. J. Hasnip, S. J. Clark, and M. C. Payne, First-principles simulation: ideas, illustrations and the CASTEP code, *J. Phys. Condens. Matt.*, **14**, 2717–2744 (2002).

4. W. K. Liu, E. G. Karpov, S. Zhang, and H. S. Park, An introduction to computational nanomechanics and materials, *Comput. Meth. Appl. Mech. Eng.*, **193**, 1529–1578 (2004).

5. R. M. Martin, *Electronic Structure: Basic Theory and Practical Methods*, Cambridge University Press, Cambridge, 2004.

6. K. Capelle, A bird's-eye view of density-functional theory, *Braz. J. Phys.*, **36**, 1318–1343 (2006).

7. J. Hafner, C. Wolverton, and G. Ceder (Guest Editors), Topical issue of the MRS Bulletin: 'Density Functional Theory in Materials Research', *MRS Bull.*, **31** (2006).

8. J. Hafner, *Ab-Initio* Simulations of Materials Using VASP: Density-Functional Theory and Beyond, *J. Comput. Chem.*, **29**, 2044–2078 (2008).

9. Y. Ma, M. Eremets, A. R. Oganov, Y. Xie, I. Trojan, S. Medvedev, A. O. Lyakhov, M. Valle, and V. Prakapenka, Transparent dense sodium, *Nature*, **458**, 182–185 (2009).

10. G. Hautier, A. Jain, S. P. Ong, B. Kang, C. Moore, R. Doe, and G. Ceder, Phosphates as lithium-ion battery cathodes: An evaluation based on high-throughput *ab initio* calculations, *Chem. Mater.*, **23**, 3495–3508 (2011).

11. A. Jain, G. Hautier, C. J. Moore, S. P. Ong, C. C. Fischer, T. Mueller, K. A. Persson, and G. Ceder, A high-throughput infrastructure for density functional theory calculations, *Comp. Mater. Sci.*, **50**, 2295–2310 (2011).

12. L. Schimka, J. Harl, A. Stroppa, A. Grüneis, M. Marsman, F. Mittendorfer, and G. Kresse, Accurate surface and adsorption energies from many-body perturbation theory, *Nature Mater.*, **9**, 741–744 (2010).

13. N. Gaston, D. Andrae, B. Paulus, U. Wedig, and M. Jansen, Understanding the hcp anisotropy in Cd and Zn: the role of electron correlation in determining the potential energy surface, *Phys. Chem. Chem. Phys.*, **12**, 681–687 (2010).

14. A. Bil, B. Kolb, R. Atkinson, D. G. Pettifor, T. Thonhauser, and A. N. Kolmogorov, van der Waals interactions in the ground state of $Mg(BH_4)_2$ from density functional theory, *Phys. Rev. B*, **83**, 224103 (2011).

15. D. N. Beratan and J. J. Hopfield, Failure of the Born-Oppenheimer and Franck-Condon approximations for long distance electron transfer rate calculations, *J. Chem. Phys.*, **81**, 5753–5759 (1984).

16. I. B. Bersuker, Limitations of density functional theory in application to degenerate states, *J. Comput. Chem.*, **18**, 260–267 (1997).

17. P. Pyykkö, The physics behind chemistry and the periodic table, *Chem. Rev.*, **112**, 371–384 (2012).

18. E. Engel, A. Facco Bonetti, S. Keller, I. Andrejkovics, and R. M. Dreizler, Relativistic optimized-potential method: Exact transverse exchange and Møller-Plesset-based correlation potential, *Phys. Rev. A*, **58**, 964–992 (1998).

19. E. Engel and U. Lechner, Exchange-correlation and QED effects from a density functional based level shift approach, *Chem. Phys.*, **311**, 209–226 (2005).

20. J. Paier, R. Hirschl, M. Marsman, and G. Kresse, The Perdew-Burke-Ernzerhof exchange-correlation functional applied to the G2-1 test set using a plane-wave basis set, *J. Chem. Phys.*, **122**, 234102 (2005).

21. A. Kiejna, G. Kresse, J. Rogal, A. De Sarkar, K. Reuter, and M. Scheffler, Comparison of the full-potential and frozen-core approximation approaches to density-functional calculations of surfaces, *Phys. Rev. B*, **73**, 035404 (2006).

22. J. P. Perdew, K. Burke, and M. Ernzerhof, Generalized gradient approximation made simple, *Phys. Rev. Lett.*, **77**, 3865–3868 (1996).

23. S. Kotochigova, Z. H. Levine, E. L. Shirley, M. D. Stiles, and C. W. Clark, Local-density-functional calculations of the energy of atoms, *Phys. Rev. A*, **55**, 191–199 (1997).

24. S. Kotochigova, Z. H. Levine, E. L. Shirley, M. D. Stiles, and C. W. Clark, Erratum: Local-density-functional calculations of the energy of atoms, *Phys. Rev. A*, **56**, 5191–5192 (1997).

25. Y. Zhao and D. G. Truhlar, Density functionals with broad applicability in chemistry, *Acc. Chem. Res.*, **41**, 157–167 (2008).

26. L. Goerigk and S. Grimme, A thorough benchmark of density functional methods for general main group thermochemistry, kinetics, and noncovalent interactions, *Phys. Chem. Chem. Phys.*, **13**, 6670–6688 (2011).

27. S. Kurth, J. P. Perdew, and P. Blaha, Molecular and solid-state tests of density functional approximations: LSD, GGAs and meta-GGAs, *Int. J. Quantum Chem.*, **75**, 889–909 (1999).

28. L. Vitos, B. Johansson, J. Kollár, and H. L. Skriver, Exchange energy in the local Airy gas approximation, *Phys. Rev. B*, **62**, 10046–10050 (2000).

29. V. N. Staroverov, G. E. Scuseria, J. Tao, and J. P. Perdew, Tests of a ladder of density functionals for bulk solids and surfaces, *Phys. Rev. B*, **69**, 075102 (2004).

30. V. N. Staroverov, G. E. Scuseria, J. Tao, and J. P. Perdew, Erratum: tests of a ladder of density functionals for bulk solids and surfaces, *Phys. Rev. B*, **78**, 239907(E) (2008).

31. J. Heyd, J. E. Peralta, G. E. Scuseria, and R. L. Martin, Energy band gaps and lattice parameters evaluated with the Heyd-Scuseria-Ernzerhof screened hybrid functional, *J. Chem. Phys.*, **123**, 174101 (2005).

32. J. Paier, M. Marsman, K. Hummer, G. Kresse, I. C. Gerber, and J. G. Ángyán, Screened hybrid density functionals applied to solids, *J. Chem. Phys.*, **124**, 154709 (2006).

33. F. Tran, R. Laskowski, P. Blaha, and K. Schwarz, Performance on molecules, surfaces, and solids of the Wu-Cohen GGA exchange-correlation energy functional, *Phys. Rev. B*, **75**, 115131 (2007).

34. B. Grabowski, T. Hickel, and J. Neugebauer, *Ab initio* study of the thermodynamic properties of nonmagnetic elementary fcc metals: exchange-correlation-related error bars and chemical trends, *Phys. Rev. B*, **76**, 024309 (2007).

35. M. Ropo, K. Kokko, and L. Vitos, Assessing the Perdew-Burke-Ernzerhof exchange-correlation density functional revised for metallic bulk and surface systems, *Phys. Rev. B*, **77**, 195445 (2008).

36. S. Lany, Semiconductor thermochemistry in density functional calculations, *Phys. Rev. B*, **78**, 245207 (2008).

37. P. Haas, F. Tran, and P. Blaha, Calculation of the lattice constant of solids with semilocal functionals, *Phys. Rev. B*, **79**, 085104 (2009).

38. P. Haas, F. Tran, and P. Blaha, Erratum: Calculation of the lattice constant of solids with semilocal functionals, *Phys. Rev. B*, **79**, 209902(E) (2009).

39. G. I. Csonka, J. P. Perdew, A. Ruzsinszky, P. H.T. Philipsen, S. Lebègue, J. Paier, O.A. Vydrov, and J. G. Ángyán, Assessing the performance of recent density functionals for bulk solids, *Phys. Rev. B*, **79**, 155107 (2009).

40. S. L. Shang, A. Saengdeejing, Z. G. Mei, D. E. Kim, H. Zhang, S. Ganeshan, Y. Wang, and Z.K. Liu, First-principles calculations of pure elements: Equations of state and elastic stiffness constants, *Comput. Mater. Sci.*, **48**, 813–826 (2010).

41. R. Nazarov, T. Hickel, and J. Neugebauer, Vacancy formation energies in fcc metals: Influence of exchange-correlation functionals and correction schemes, *Phys. Rev. B*, **85**, 144118 (2012).

42. G. Hautier, S. P. Ong, A. Jain, C. J. Moore, and G. Ceder, Accuracy of density functional theory in predicting formation energies of ternary oxides from binary oxides and its implication on phase stability, *Phys. Rev. B*, **85**, 155208 (2012).

43. P. Villars and J. L.C. Daams, Atomic-environment classification of the chemical elements, *J. Alloys Comp.*, **197**, 177–196 (1993).

44. http://www.marcelswart.eu/dft-poll.

45. W. H. Dickhoff and D. Van Neck, *Many-Body Theory Exposed!*, 2nd ed., World Scientific, Singapore (2008).

46. C. Kittel, *Introduction to Solid State Physics*, 8th ed., John Wiley & Sons, Inc., New York (2005).

47. F. Birch, Finite elastic strain of cubic crystals, *Phys. Rev.*, **71**, 809–824 (1947).

48. P. H. T. Philipsen and E. J. Baerends, Cohesive energy of 3*d* transition metals: Density functional theory atomic and bulk calculations, *Phys. Rev. B*, **54**, 5326–5333 (1996).

49. E. J. Baerends, V. Branchadell, and M. Sodupe, Atomic reference energies for density functional calculations, *Chem. Phys. Lett.*, **265**, 481–489 (1997).

50. D. E. Kim, S. L. Shang, and Z. K. Liu, Effects of alloying elements on elastic properties of Ni by first-principles calculations, *Comput. Mater. Sci.*, **47**, 254–260 (2009).

51. N. Miao, B. Sa, J. Zhou, and Z. Sun, Theoretical investigation on the transition-metal borides with Ta_3B_4-type structure: A class of hard and refractory materials, *Comput. Mater. Sci.*, **50**, 1559–1566 (2011).

52. Y. Le Page and P. Saxe, Symmetry-general least-squares extraction of elastic coefficients from *ab initio* total energy calculations, *Phys. Rev. B*, **63**, 174103 (2001).

53. Y. Le Page and P. Saxe, Symmetry-general least-squares extraction of elastic data for strained materials from *ab initio* calculations of stress, *Phys. Rev. B*, **65**, 104104 (2002).

54. S. L. Shang, Y. Wang, and Z. K. Liu, First-principles elastic constants of α- and θ-Al_2O_3, *Appl. Phys. Lett.*, **90**, 101909 (2007).

55. A. B. Alchagirov, J. P. Perdew, J. C. Boettger, R. C. Albers, and C. Fiolhais, Energy and pressure versus volume: Equations of state motivated by the stabilized jellium model, *Phys. Rev. B*, **63**, 224115 (2001).

56. V. L. Moruzzi, J. F. Janak, and K. Schwarz, Calculated thermal properties of metals, *Phys. Rev. B*, **37**, 790–799 (1988).

57. Y. Tsuru, Y. Shinzato, Y. Saito, M. Shimazu, M. Shiono, and M. Morinaga, Estimation of linear thermal expansion coefficient from cohesive energy obtained by ab-initio calculation of metals and ceramics, *J. Ceram. Soc. Jpn.*, **118**, 241–245 (2010).

58. M. M. Dacorogna and M. L. Cohen, First-principles study of the structural properties of alkali metals, *Phys. Rev. B*, **34**, 4996–5002 (1986).

59. R. Gaudoin and W. M. C. Foulkes, *Ab initio* calculations of bulk moduli and comparison with experiment, *Phys. Rev. B*, **66**, 052104 (2002).

60. D. G. Pettifor, *Bonding and Structure of Molecules and Solids*, Oxford University Press, Oxford (1995).

61. A. Masago, K. Shirai, and H. Katayama-Yoshida, Crystal stability of α- and β- boron, *Phys. Rev. B*, **73**, 104102 (2006).

62. T. A. Scott, Solid and liquid nitrogen, *Phys. Rep.*, **27C**, 89–157 (1976).

63. Y. A. Freiman and H. J. Jodl, Solid oxygen, *Phys. Rep.*, **401**, 1–228 (2004).

64. W. I. F. David, R. M. Ibberson, S. F. J. Cox, and P. T. Wood, Order-disorder transition in monoclinic sulfur: a precise structural study by high-resolution neutron powder diffraction, *Acta Crystallogr. Sect. B*, **62**, 953–959 (2006).

65. P. E. Blöchl, Projector augmented-wave method, *Phys. Rev. B*, **50**, 17953–17979 (1994).

66. G. Kresse and D. Joubert, From ultrasoft pseudopotentials to the projector augmented-wave method, *Phys. Rev. B*, **59**, 1758–1775 (1999).

67. G. Kresse and J. Furthmüller, Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set, *Comput. Mater. Sci.*, **6**, 15–50 (1996).

68. http://cms.mpi.univie.ac.at/vasp/vasp/vasp.html (April 23, 2009 version).

69. P. E. Blöchl, O. Jepsen, and O.K. Andersen, Improved tetrahedron method for Brillouin-zone integrations, *Phys. Rev. B*, **49**, 16223–16233 (1994).

70. H. J. Monkhorst and J. D. Pack, Special points for Brillouin-zone integrations, *Phys. Rev. B*, **13**, 5188–5192 (1976).

71. S. Gražulis, D. Chateigner, R. T. Downs, A. F. T. Yokochi, M. Quirós, L. Lutterotti, E. Manakova, J. Butkus, P. Moeck, and A. Le Bail, Crystallography Open Database – an open-access collection of crystal structures, *J. Appl. Crystallogr.*, **42**, 726–729 (2009).

72. A. Belsky, M. Hellenbrandt, V. L. Karen, and P. Luksch, New developments in the Inorganic Crystal Structure Database (ICSD): accessibility in support of materials research and design, *Acta Crystallogr. Sect. B*, **58**, 364–369 (2002).

73. N. N. Greenwood and A. Earnshaw, *Chemistry of the Elements*, 2nd ed., Butterworth-Heinemann, Oxford (1997).

74. T. Tohei, A. Kuwabara, F. Oba, and I. Tanaka, Debye temperature and stiffness of carbon and boron nitride polymorphs from first principles calculations, *Phys. Rev. B*, **73**, 064304 (2006).

75. L. H. Bolz, M. E. Boyd, F. A. Mauer, and H. S. Peiser, A re-examination of the crystal structures of α and β nitrogen, *Acta Crystallogr.*, **12**, 247–248 (1959).

76. L. Meyer, C. S. Barrett, and S. C. Greer, Crystal structure of α-fluorine, *J. Chem. Phys.*, **49**, 1902–1907 (1968).

77. D. N. Batchelder, D. L. Losee, and R. O. Simmons, Measurements of lattice constant, thermal expansion, and isothermal compressibility of neon single crystals, *Phys. Rev.*, **162**, 767–775 (1967).

78. B. M. Powell, K. M. Heal, and B. H. Torrie, The temperature dependence of the crystal structures of the solid halogens, bromine and chlorine, *Mol. Phys.*, **53**, 929–939 (1984).

79. O. G. Peterson, D. N. Batchelder, and R. O. Simmons, Measurements of X-ray lattice constant, thermal expansivity, and isothermal compressibility of argon crystals, *Phys. Rev.*, **150**, 703–711 (1966).

80. D. L. Losee and R. O. Simmons, Thermal-expansion measurements and thermodynamics of solid krypton, *Phys. Rev.*, **172**, 944–957 (1968).

81. D. R. Sears and H. P. Klug, Density and expansivity of solid xenon, *J. Chem. Phys.*, **37**, 3002–3006 (1962).

82. C. A. Swenson, Phase transition in solid mercury, *Phys. Rev.*, **111**, 82–91 (1958).

83. E. Kim and C. Chen, Calculation of bulk modulus for highly anisotropic materials, *Phys. Lett. A*, **326**, 442–448 (2004).

84. Y. Endoh, G. Shirane, and J. Skalyo, Jr., Lattice dynamics of solid neon at 6.5 and 23.7 K, *Phys. Rev. B*, **11**, 1681–1688 (1975).

85. E.-F. Düsing, W. A. Grosshans, and W. B. Holzapfel, Equation of state of solid chlorine and bromine, *J. Phys. Colloques (Paris)*, **45**, 203–206 (1984).

86. G. J. Keeler and D. N. Batchelder, Measurement of the elastic constants of argon from 3 to 77°K, *J. Phys. C*, **3**, 510–522 (1970).

87. H. Fujihisa and K. Takemura, Stability and the equation of state of α-manganese under ultrahigh pressure, *Phys. Rev. B*, **52**, 13257–13260 (1995).

88. J. Skalyo, Jr., Y. Endoh, and G. Shirane, Inelastic neutron scattering from solid krypton at 10°K, *Phys. Rev. B*, **9**, 1797–1803 (1974).

89. D. L. Price, J. M. Rowe, and R. M. Nicklow, Lattice dynamics of grey tin and indium antimonide, *Phys. Rev. B*, **3**, 1268–1279 (1971).

90. N. A. Lurie, G. Shirane, and J. Skalyo, Jr., Temperature dependence of the zero-sound elastic constants of crystalline xenon, *Phys. Rev. B*, **9**, 2661–2669 (1974).

91. E. Knittle, Static compression measurements of equations of state, in *Mineral Physics and Crystallography: A Handbook of Physical Constants*, Ahrens, T. J., Ed., American Geophysical Union Press, 98–142 (1995).

92. S. S. Batsanov, *Effects of Explosions on Materials: Modification and Synthesis under High Pressure Shock Compression*, Springer, New York (1994).

93. M. W. Guinan and D. J. Steinberg, Pressure and temperature derivatives of the isotropic polycrystalline shear modulus for 65 elements, *J. Phys. Chem. Solids*, **35**, 1501–1512 (1974).

94. C. S. Menoni, J. Z. Hu, and I. L. Spain, Germanium at high pressures, *Phys. Rev. B*, **34**, 362–368 (1986).

95. H. J. Beister, K. Strössner, and K. Syassen, Rhombohedral to simple-cubic phase transition in arsenic under pressure, *Phys. Rev. B*, **41**, 5535–5543 (1990).

96. T. Kenichi, Bulk modulus of osmium: High-pressure powder x-ray diffraction experiments under quasihydrostatic conditions, *Phys. Rev. B*, **70**, 012101 (2004).

97. J. H. Li, S. H. Liang, H. B. Guo, and B. X. Liu, Four-parameter equation of state and determination of the thermal and mechanical properties of metals, *J. Alloys Compd.*, **431**, 23–31 (2007).

98. L. G. Hector, Jr., J. F. Herbst, W. Wolf, P. Saxe, and G. Kresse, *Ab initio* thermodynamic and elastic properties of alkaline-earth metals and their hydrides, *Phys. Rev. B*, **76**, 014121 (2007).

99. U. Buchenau, M. Heiroth, H. R. Schober, J. Evers, and G. Oehlinger, Lattice dynamics of strontium and barium, *Phys. Rev. B*, **30**, 3502–3505 (1984).

100. D. R. Lide (ed.) *CRC Handbook of Chemistry and Physics,* 89th ed., CRC Press, Boca Raton, Florida (2008–2009).

101. J. B. Nelson and D. P. Riley, The thermal expansion of graphite from 15°C. to 800°C.: Part I. Experimental, *Proc. Phys. Soc. London*, **57**, 477–486 (1945).

102. R. W. Keyes, The electrical properties of black phosphorus, *Phys. Rev.*, **92**, 580–584 (1953).

103. J. Wallis, I. Sigalas, and S. Hart, Determination of the thermal expansion of orthorhombic sulfur, *J. Appl. Crystallogr.*, **19**, 273–274 (1986).

104. R. R. Pawar and V. T. Deshpande, Anisotropic thermal expansion of arsenic, *J. Mater. Sci.*, **5**, 1061–1062 (1970).

105. M. S. Anderson and C. A. Swenson, Experimental compressions for sodium, potassium, and rubidium metals to 20 kbar from 4.2 to 300 K, *Phys. Rev. B*, **28**, 5395–5418 (1983).

106. J. P. Perdew, J. A. Chevary, S. H. Vosko, K. A. Jackson, M. R. Pederson, D. J. Singh, and C. Fiolhais, Atoms, molecules, solids, and surfaces: Applications of the generalized gradient approximation for exchange and correlation, *Phys. Rev. B*, **46**, 6671–6687 (1992).

107. J. P. Perdew, J. A. Chevary, S. H. Vosko, K. A. Jackson, M. R. Pederson, D. J. Singh, and C. Fiolhais, Erratum: Atoms, molecules, solids, and surfaces: Applications of the generalized gradient approximation for exchange and correlation, *Phys. Rev. B*, **48**, 4978 (1993).

108. A. E. Mattsson, R. Armiento, P. A. Schultz, and T. R. Mattsson, Nonequivalence of the generalized gradient approximations PBE and PW91, *Phys. Rev. B*, **73**, 195123 (2006).

109. M. H. DeGroot and M. J. Schervish, *Probability and Statistics,* 3rd ed., Addison-Wesley, Boston (2002).

110. E. Scerri, Explaining the periodic table, and the role of chemical triads, *Found. Chem.*, **12**, 69–83 (2010).

111. E. Scerri, Which elements belong in group 3? *J. Chem. Educ.*, **86**, 1188 (2009).

112. P. Villars, K. Brandenburg, M. Berndt, S. LeClair, A. Jackson, Y.-H. Pao, B. Igelnik, M. Oxley, B. Bakshi, P. Chen, and S. Iwata, Binary, ternary and quaternary compound former/nonformer prediction via Mendeleev number, *J. Alloys Compd.*, **317-318**, 26–38 (2001).

113. J. L. Rodgers and W. A. Nicewander, Thirteen ways to look at the correlation coefficient, *Amer. Statist.*, **42**, 59–66 (1988).

114. S. Grimme, Semiempirical GGA-type density functional constructed with a long-range dispersion correction, *J. Comput. Chem.*, **27**, 1787–1799 (2006).

115. K. Lee, E. D. Murray, L. Kong, B. I. Lundqvist, and D. C. Langreth, Higher-accuracy van der Waals density functional, *Phys. Rev. B*, **82**, 081101(R) (2010).

116. N. G. Pace, G. A. Saunders, and Z. Sümengen, The elastic constants and interatomic binding forces in arsenic, *J. Phys. Chem. Solids*, **31**, 1467–1476 (1970).

117. D. D. Richardson and J. Mahanty, Van der Waals contribution to the binding energy of noble metals, *J. Phys. C Solid State Phys.* **10**, 3971–3976 (1977).

118. S. Appalakondaiah, G. Vaitheeswaran, S. Lebègue, N. E. Christensen, and A. Svane, Effect of van der Waals interactions on the structural and elastic properties of black phosphorus, *Phys. Rev. B*, **86**, 035105 (2012).

119. D. Hobbs, J. Hafner, and D. Spišák, Understanding the complex metallic element Mn. I. Crystalline and noncollinear magnetic structure of α-Mn, *Phys. Rev. B*, **68**, 014407 (2003).

120. D. J. Singh and J. Ashkenazi, Magnetism with generalized-gradient-approximation density functionals, *Phys. Rev. B*, **46**, 11570–11577 (1992).

121. R. B. Roberts, G. K. White, and E. Fawcett, Thermal expansion of Cr and Cr-V alloys, *Physica B (Utrecht)*, **119**, 63–67 (1983).

122. N. Gaston, B. Paulus, K. Rosciszewski, P. Schwerdtfeger, and H. Stoll, Lattice structure of mercury: Influence of electronic correlation, *Phys. Rev. B*, **74**, 094102 (2006).

123. U. Wedig, M. Jansen, B. Paulus, K. Rosciszewski, and P. Sony, Structural and electronic properties of Mg, Zn, and Cd from Hartree-Fock and density functional calculations including hybrid functionals, *Phys. Rev. B*, **75**, 205123 (2007).

124. D. D. Koelling and B. N. Harmon, A technique for relativistic spin-polarized calculations, *J. Phys. C Solid State*, **10**, 3107–3114 (1977).

125. P. H.T. Philipsen and E. J. Baerends, Relativistic calculations to assess the ability of the generalized gradient approximation to reproduce trends in cohesive properties of solids, *Phys. Rev. B*, **61**, 1773–1778 (2000).

126. V. Ozoliņš and M. Körling, Full-potential calculations using the generalized gradient approximation: Structural properties of transition metals, *Phys. Rev. B*, **48**, 18304–18307 (1993).

127. P. Haas, F. Tran, P. Blaha, K. Schwarz, and R. Laskowski, Insight into the performance of GGA functionals for solid-state calculations, *Phys. Rev. B*, **80**, 195109 (2009).

128. A. R. Oganov and C. W. Glass, Crystal structure prediction using *ab initio* evolutionary techniques: Principles and applications, *J. Chem. Phys.*, **124**, 244704 (2006).

129. N. L. Abraham and M. I. J. Probert, A periodic genetic algorithm with real-space representation for crystal structure and polymorph prediction, *Phys. Rev. B*, **73**, 224104 (2006).

130. N. L. Abraham and M. I. J. Probert, Improved real-space genetic algorithm for crystal structure and polymorph prediction, *Phys. Rev. B*, **77**, 134117 (2008).

131. A. N. Kolmogorov, S. Shah, E. R. Margine, A. F. Bialon, T. Hammerschmidt, and R. Drautz, New superconducting and semiconducting Fe-B compounds predicted with an *ab initio* evolutionary search, *Phys. Rev. Lett.*, **105**, 217003 (2010).

132. P. Blaha, K. Schwarz, G. K. H. Madsen, D. Kvasnicka, and J. Luitz, WIEN2k, An augmented plane wave + local orbitals program for calculating crystal properties, Karlheinz Schwarz, Techn. Universität Wien, Austria (1999).

133. E. Sjöstedt, L. Nordström, and D. J. Singh, An alternative way of linearizing the augmented plane-wave method, *Solid State Commun.*, **114**, 15–20 (2000).

134. G. K. H. Madsen, P. Blaha, K. Schwarz, E. Sjöstedt, and L. Nordström, Efficient linearization of the augmented plane-wave method, *Phys. Rev. B*, **64**, 195134 (2001).

135. S. R. Bahn and K. W. Jacobsen, An object-oriented scripting interface to a legacy electronic structure code, *Comput. Sci. Eng.*, **4**, 56–66 (2002).

136. J. J. Mortensen, L. B. Hansen, and K. W. Jacobsen, Real-space grid implementation of the projector augmented wave method, *Phys. Rev. B*, **71**, 035109 (2005).

137. J. Enkovaara, C. Rostgaard, J. J. Mortensen, J. Chen, M. Dułak, L. Ferrighi, J. Gavnholt, C. Glinsvad, V. Haikola, H. A. Hansen, H. H. Kristoffersen, M. Kuisma, A. H. Larsen, L. Lehtovaara, M. Ljungberg, O. Lopez-Acevedo, P. G. Moses, J. Ojanen, T. Olsen, V. Petzold, N. A. Romero, J. Stausholm-Møller, M. Strange, G. A. Tritsaris, M. Vanin, M. Walter, B. Hammer, H. Häkkinen, G. K.H. Madsen, R. M. Nieminen, J. K. Nørskov, M. Puska, T. T. Rantala, J. Schiøtz, K. S. Thygesen, and K. W. Jacobsen, Electronic structure calculations with GPAW: a real-space implementation of the projector augmented-wave method, *J. Phys. Condens. Matt.*, **22**, 253202 (2010).

138. E. Rapoport and G. C. Kennedy, Phase diagram of manganese to 40 kbars, *J. Phys. Chem. Solids*, **27**, 93–98 (1966).

139. H. Luo, R. G. Greene, and A. L. Ruoff, β-Po phase of sulfur at 162 GPa: X-ray diffraction study to 212 GPa, *Phys. Rev. Lett.*, **71**, 2943–2946 (1993).

140. K. Lejaeghere, V. Van Speybroeck, G. Van Oost, and S. Cottenier, The ground state elemental crystals as a benchmark set for solid state DFT: intrinsic accuracy and code comparison, (2012), arXiv:1204.2733 [cond-mat.mtrl-sci].

141. M. Dułak, Private Communication.

142. http://molmod.ugent.be/DeltaCodesDFT.

APPENDIX: CALCULATING THE Δ-FACTOR

For a particular compound, the energy difference between the equations of state of WIEN2k (w) and a code under investigation (c) can be evaluated analytically, using the Birch-Murnaghan relation of Equation (2). Some mathematical manipulations yield

$$\int_{V_i}^{V_f} (E^c(V) - E^w(V))^2 \mathrm{d}V = F(V_f) - F(V_i), \qquad [A.16]$$

where the primitive function $F(V)$ can be written as a power series in $V^{-1/3}$:

$$F(V) = \sum_{n=-2}^{4} x_n V^{-(2n+1)/3}. \qquad [A.17]$$

The coefficients x_n are given by

$$x_n = -\frac{3}{2n+1} \sum_{i+j=n+2} \left(a_i^c - a_i^w\right)\left(a_j^c - a_j^w\right) \qquad [A.18]$$

with $i, j \in \{0, 1, 2, 3\}$. Then x_{-1}, for example, becomes

$$x_{-1} = 6\left(a_1^c - a_1^w\right)\left(a_0^c - a_0^w\right). \qquad [A.19]$$

The constants a_i are the coefficients of the Birch-Murnaghan equation in its polynomial form:

$$a_3 = \frac{9V_0^3 B_0}{16}(B_1 - 4) \qquad [A.20]$$

$$a_2 = \frac{9V_0^{7/3} B_0}{16}(14 - 3B_1) \qquad [A.21]$$

$$a_0 = \frac{9V_0^{5/3} B_0}{16}(3B_1 - 16) \qquad [A.22]$$

$$a_0 = \frac{9V_0 B_0}{16}(6 - B_1). \qquad [A.23]$$

When evaluating Equation (A.18), a_i^c stands for the coefficient of the code under test, while a_i^w means it corresponds to the WIEN2k equation of state.