Ab initio based thermal property predictions at a low cost: An error analysis

Kurt Lejaeghere,[1] Jan Jaeken,[1] Veronique Van Speybroeck,[1] and Stefaan Cottenier[1,2]
[1]*Center for Molecular Modeling, Ghent University, Technologiepark 903, BE-9052 Zwijnaarde, Belgium*
[2]*Department of Materials Science and Engineering, Ghent University, Technologiepark 903, BE-9052 Zwijnaarde, Belgium*
(Received 8 November 2013; published 28 January 2014)

Ab initio calculations often do not straightforwardly yield the thermal properties of a material yet. It requires considerable computational efforts, for example, to predict the volumetric thermal expansion coefficient α_V or the melting temperature T_m from first principles. An alternative is to use semiempirical approaches. They relate the experimental values to first-principles predictors via fits or approximative models. Before applying such methods, however, it is of paramount importance to be aware of the expected errors. We therefore quantify these errors at the density-functional theory level using the Perdew-Burke-Ernzerhof functional for several semiempirical approximations of α_V and T_m, and compare them to the errors from fully *ab initio* methods, which are computationally more intensive. We base our conclusions on a benchmark set of 71 ground-state elemental crystals. For the thermal expansion coefficient, it appears that simple quasiharmonic theory, in combination with different approximations to the Grüneisen parameter, provides a similar overall accuracy as exhaustive first-principles phonon calculations. For the melting temperature, expensive *ab initio* molecular-dynamics simulations still outperform semiempirical methods.

DOI: 10.1103/PhysRevB.89.014304 PACS number(s): 65.40.De, 64.70.dj, 06.20.Dk, 71.15.Mb

I. INTRODUCTION

Density-functional theory (DFT) [1–3] is used increasingly often to support experimental research. It moreover allows predicting a compound's characteristics without raw materials or expensive apparatus. Even conditions inaccessible to experiment, such as those in the Earth's core, are not particularly more troublesome than ambient pressure and temperature [4]. Unfortunately, not all properties directly follow from simple DFT simulations. Static DFT calculations are practically limited to 0 K, complicating the determination of thermal properties.

Two such properties are the volumetric thermal expansion coefficient α_V and the melting temperature T_m. They can be predicted from an *ab initio* approach, but only by going beyond single-point DFT computations, using either high-temperature molecular dynamics [5,6] or phonon-based calculations [7]. These quickly become computationally intensive. Alternatively, semiempirical approximations provide the same information at a much lower cost [8–12]. They significantly reduce the calculation time and effort by relating α_V and T_m to much simpler DFT predictors, such as the cohesive energy or the pressure derivative of the bulk modulus. These methods are based on approximative models or fits to experiment.

A semiempirical approach is often the only pragmatic way to ensure a fast interaction between experimental and computational research. This is especially true when the design of new materials is concerned. Only when the resulting accuracy does not suffice or when the physical mechanisms behind a particular compound are unclear, may more expensive calculations be justified. Semiempirical approximations do lead to additional errors, however. These errors originate from the less-than-perfect correlation between the investigated property and its predictor. In particular, one simple expression often does not fully represent the physics behind a material's behavior. Describing it with a predictor hence oversimplifies reality, causing semiempirical predictions to deviate from the experimental values.

The semiempirical relations for α_V and T_m should therefore be treated with caution. Although they expedite calculations substantially, this benefit must be weighed against the introduction of additional (non-DFT) errors. To decide how useful a particular semiempirical approach is, its errors need to be quantified. For that reason, we performed an error analysis similar to that in Ref. [13], which was dedicated to properties directly from DFT, such as crystal volume or cohesive energy. The procedure investigates the agreement between experimental values and (semi)theoretical predictions for 71 elemental crystals and characterizes the remaining discrepancies in terms of systematic deviations and residual error bars. Applied to α_V and T_m, such a procedure allows comparing the different semiempirical approximations and evaluating the difference with strictly first-principles methods.

The remainder of this paper is therefore structured as follows. Section II outlines which purely theoretical approaches and semiempirical relations can be employed to predict α_V and T_m, while Sec. III elaborates on the computational aspects of our predictor calculations and error analyses. These methods are then applied in Sec. IV to provide reliable error estimates for the semiempirical approaches. The results are also compared to more advanced predictions. Section V summarizes the most important conclusions.

II. HIGH-LEVEL THEORY AND SEMIEMPIRICAL ALTERNATIVES

A. Thermal expansion coefficient

The thermal expansion coefficient nicely illustrates that a purely theoretical framework leads to computationally expensive simulations. High-precision results are obtained in the quasiharmonic approximation, by calculating the volume-dependent phonon spectrum $\omega_i(V)$ as a function of wave vector and mode [7]. Using thermodynamical formulas, this allows establishing the temperature-dependent free energy, which reaches a minimum at its equilibrium volume V_0. The shift of

V_0 as a function of T yields the volumetric thermal expansion coefficient $\alpha_V(T)$. However, the phonon spectrum consists of the eigenvalues of the dynamical matrix, and this matrix includes the derivatives of the forces with respect to all atomic displacements. Unit cells with M inequivalent atoms therefore require the assessment of three independent displacements for each of these atoms. For small unit cells, periodic boundary conditions moreover cause the displaced atoms to interact with their periodic images. To prevent such undesired interactions, the unit cell needs to be expanded to several times its original size. Finally, only stringent computational settings yield numerically converged forces. The overall result is a set of (at least) $3M$ supercell calculations, each of which is computationally intensive due to its considerable cell size and required accuracy.

The method proposed by Tsuru *et al.* [8] offers a much simpler alternative. Inspired by the link between lattice expansion and crystalline cohesion, the authors fitted inverse relations between the experimental linear thermal expansion coefficient $\alpha_l(T)$ and the DFT-PW91 [14,15] cohesive energy ΔE_{coh}. At room temperature (T_{rt}), they obtained

$$\alpha_V^{Ts} = 3\,\alpha_l^{Ts} = 3\,\frac{48.14\times10^{-6}\ \text{eV/K/atom}}{\Delta E_{\text{coh}}} \tag{1}$$

by fitting to a number of pure metals, as well as binary oxides, nitrides, borides, and carbides.

Other semiempirical approaches are based on thermodynamical considerations. In the quasiharmonic approximation, it can be shown that [7]

$$\alpha_V = \frac{\gamma C_V}{B_0 V_0} \tag{2}$$

with

$$\gamma = \frac{\sum \gamma_i C_{V,i}}{C_V} \tag{3}$$

and

$$\gamma_i = -\frac{\partial \ln \omega_i}{\partial \ln V}. \tag{4}$$

Here, B_0 stands for the equilibrium bulk modulus, C_V for the isochoric heat capacity, and γ for the overall Grüneisen parameter. The index i again labels the different phonon modes, such that within the 1D Einstein model:

$$C_{V,i} = k_B\left(\frac{\hbar\omega_i}{k_B T}\right)^2 \frac{\exp\left(\frac{\hbar\omega_i}{k_B T}\right)}{\left[\exp\left(\frac{\hbar\omega_i}{k_B T}\right)-1\right]^2} \tag{5}$$

(with k_B the Boltmann constant). The key quantities in Eq. (2), however, are the Grüneisen parameters γ_i, which express the volume dependence of the phonon frequencies and can hence be considered as a measure of the anharmonicity of the crystal. Indeed, purely harmonic crystals only contain volume-independent phonons, so thermal expansion does not occur there [$\alpha_V(T) = 0, \forall\, T$].

Equation (2) can be directly completed with information from phonon calculations, using Eq. (4). More approximatively, however, it is also possible to determine γ from the pressure derivative of the bulk modulus B_1. Such a relation effectively replaces thermal effects by an overall (pressurelike) bulk effect and therefore does not require computing phonons. Several formulations are available for $\gamma(B_1)$, the most popular ones being those of Slater [9], Dugdale and MacDonald [10], and Vashchenko and Zubarev [11]:

$$\gamma^S = -\tfrac{1}{6} + \tfrac{1}{2}B_1, \tag{6}$$

$$\gamma^{\text{DM}} = -\tfrac{1}{2} + \tfrac{1}{2}B_1, \tag{7}$$

$$\gamma^{\text{VZ}} = -\tfrac{5}{6} + \tfrac{1}{2}B_1. \tag{8}$$

They yield

$$\alpha_V^S = \left(-\frac{1}{6} + \frac{1}{2}B_1\right)\frac{C_V}{B_0 V_0}, \tag{9}$$

$$\alpha_V^{\text{DM}} = \left(-\frac{1}{2} + \frac{1}{2}B_1\right)\frac{C_V}{B_0 V_0}, \tag{10}$$

$$\alpha_V^{\text{VZ}} = \left(-\frac{5}{6} + \frac{1}{2}B_1\right)\frac{C_V}{B_0 V_0}. \tag{11}$$

γ can also be derived from the relation between a material's binding energy and atomic separation. By employing a universal binding-energy relation for metals [16], Guinea and co-workers [12] found an expression that does not explicitly depend on B_1:

$$\gamma^G = 0.38\,\frac{r_{\text{WS},0}}{l_{\text{TF}}} \tag{12}$$

with $r_{\text{WS},0}$ the equilibrium Wigner-Seitz radius and l_{TF} the Thomas-Fermi screening length:

$$l_{\text{TF}} = \sqrt{\frac{\Delta E_{\text{coh}}}{12\pi r_{\text{WS},0} B_0}}. \tag{13}$$

Completing Eq. (2) with Eq. (12) then yields a relation similar to that of Tsuru *et al.* [Eq. (1)]:

$$\alpha_V^G = 10.35\left(\frac{l_{\text{TF}}}{r_{\text{WS},0}}\right)\left(\frac{k_B}{\Delta E_{\text{coh}}}\right). \tag{14}$$

Contrary to Eq. (1), however, the prefactor is now compound-dependent.

B. Melting temperature

Just like α_V, the melting temperature T_m cannot be obtained from a simple DFT calculation either. Moreover, the quasiharmonic approximation fails at temperatures that high, so even most phonon-based results become inadequate. *Ab initio* melting temperatures therefore require molecular-dynamics calculations of the solid and/or liquid sample near the melting point. Several versions of this strategy are available. It is possible to investigate the evolution of a one- or two-phase sample into a solid or liquid monophase [5], for example, and map the phase diagram by changing the ambient conditions. Alternatively, the coexistence between both phases can be studied as well [6], maintaining a biphase sample at all times. In addition to these two techniques, other approaches allow establishing the melting temperature too [17], but they have

until now not been possible without using fitted potentials, and can therefore not be designated as purely "first principles" yet.

In the monophase methods, where only one phase remains at the end of the simulation, T_m follows from tuning the temperature and pressure, and monitoring the resulting state (solid or liquid). This can be done using NVE, NVT, or NPT ensembles, where either the particle number N, the volume V, the energy E, the temperature T or the pressure P are held constant. The melting temperature is then found as the transition point between the two states of matter. It has been shown that the most accurate monophase results are obtained when the initial configuration contains a solid-liquid interface [18], the so-called two-phase approach [19–21]. When this is not the case, nucleation of the competing phase is hindered, leading to superheated or undercooled states and hence to substantial errors on the predicted melting temperature. One-phase calculations do offer the smallest computational effort, however, and two strategies exist to remedy the superheating issue in solid-to-liquid simulations. The Z method was proposed by Belonoshko and co-workers [22] and retrieves the *ab initio* melting temperature by letting the system relax from its superheated melting point to the NVE ensemble equilibrium [23]. The method is named after the characteristic shape of the corresponding isochores. The void method, on the other hand, introduces voids into the solid sample [24], facilitating the melting process in a similar way as the solid-liquid interface does in the two-phase approach. However, as far as we know, void-induced melting calculations have not yet been attempted at the DFT level.

Contrary to the monophase methods, the necessary melting information can also be extracted from the first-principles solid-liquid equilibrium. This is the so-called coexistence method [25]. Here, an NVE [6] or NPH [26] ensemble is simulated for a biphase medium (containing both the solid and the liquid). With the NVE ensemble, each starting energy characterizes another point (P,T) of the phase equilibrium. This allows establishing the melting curve point by point. The NPH ensemble, on the other hand, allows to straightforwardly select a predefined pressure, for which the corresponding melting temperature is obtained.

In stark contrast to these rather intricate procedures, the melting temperature is known to correlate well with the cohesive energy [27]. Both properties are closely linked to the material's bond strength, and ΔE_{coh} readily follows from routine DFT calculations. A theoretical proportionality has been proposed by Guinea and co-workers [12], based on their universal binding-energy relation for metals [16]. They assumed the rms displacement of the atoms at T_m to equal the Thomas-Fermi screening length and obtained

$$T_m = 0.032\frac{\Delta E_{\text{coh}}}{k_B}. \tag{15}$$

In addition to this theoretical equation, more empirical forms exist as well. Li *et al*., for example, found a proportionality similar to Eq. (15) not only to be valid for pure metals, but even to approximate the melting behavior of intermetallic structures such as CsCl-type compounds [28] and Laves phases [29]. Tateno mentions a proportionality between the cohesive energy and the melting temperature too, but he includes additional materials properties in the prefactor [30]. In this study, we restricted the ΔE_{coh} proportionality to a relation similar to Eq. (15), i.e., a single linear function of the cohesive energy with a constant prefactor.

The melting temperature also correlates well with elastic moduli. Indeed, just like the cohesive energy, they express the resistance of the material against deformation, which is needed to initiate melting. Fine *et al*., for example, used this connection to predict $\langle C_{ii}\rangle$ from T_m [31]. Kittel moreover observes that "melting temperatures and bulk moduli vary roughly as the cohesive energies" [27], which implies that bulk moduli and melting temperatures are related. In this work, we therefore examined B_0 as a predictor for T_m as well [see further, Eq. (17)].

III. METHODOLOGY

We evaluated the performance of all semiempirical relations for α_V and T_m (Sec. II) using the error analysis method by Lejaeghere *et al*. [13]. This procedure is based on a large test set, containing 71 ground-state elemental crystals up to radon (not including the lanthanides), for which the correspondence with experiment is checked. It results in quantitative estimates for the systematic and residual errors.

A. Computational details

For the elemental test set, all calculations were performed using the projector augmented-wave method (PAW) [32,33] in the Vienna *ab initio* simulation package (VASP) [34,35] (version 5.2.2). Spin polarization was taken into account for Fe, Co, and Ni (ferromagnetic), O and Cr (antiferromagnetic), and Mn (ferrimagnetic), while for the heaviest elements (as from the 5*d* transition metals), spin-orbit coupling was included. We used the atomic potentials recommended by the VASP manual [36], which we also listed in Ref. [37]. For the description of exchange and correlation, the functional by Perdew, Burke, and Ernzerhof (PBE) [38] was selected. As it is one of the most popular functionals to date [39], an error assessment for PBE is of more use to the community than for most other functionals. Admittedly, Eq. (1) was originally fitted to PW91 data, but results from this functional are very similar to those from PBE [38], so we directly applied Eq. (1) to PBE as well.

To ensure that all error estimates are independent of the code settings, it is essential to numerically converge each DFT-based property. The cutoff energy was therefore set at 400 eV, except for He, B, C, N, O, F, and Ne, where it was 600 eV. We moreover took $6750/N$ k points in the first Brillouin zone for each N-atom cell, using a Monkhorst-Pack grid [40]. The shape and atomic positions of each unit cell were relaxed until all forces were smaller than 10^{-2} eV/Å, and the self-consistent electronic cycles were converged up to 10^{-4} eV. Combined, such settings led to a numerical convergence of 1 meV/atom in the total energy. These results were largely independent of the used code as well, since different implementations seldom change the equations of state more than a few meV/atom [13,41].

Using these computational parameters, α_V and T_m were determined semiempirically for each of the test set crystals. This required only a limited number of DFT predictors (Sec. II): the cohesive energy ΔE_{coh}, the bulk modulus B_0, the equilibrium volume per atom V_0, and the pressure derivative of the bulk modulus B_1. When necessary, the

isochoric heat capacity C_V was set to its high-temperature limit $3k_B$. For some elements the cohesive energy was slightly modified. In particular, all semiempirical relations use ΔE_{coh} as a predictor for the crystalline cohesion [Eqs. (1), (14), and (15)], but for dimeric crystals, it is mainly the cohesion between diatomic molecules that matters. These entities are maintained throughout expansion, until even after melting. A good predictor for α_V or T_m should therefore incorporate that information. To emphasize the difference from ΔE_{coh}, we named this quantity the "moleculization" energy ΔE_{mol}:

$$\Delta E_{\text{mol}} = -(E_{\text{solid}} - E_{\text{gaslike}}). \tag{16}$$

For most compounds, it does correspond to the atomization (or cohesive) energy, but for dimeric crystals, it yields a more suitable predictor.

The moleculization energy was determined for all elemental compounds by subtracting the energy of isolated atoms or molecules from that of the equilibrium crystal structure. Since VASP employs periodic boundary conditions, that isolated particle can only be simulated in a periodic environment as well. All atoms and molecules were therefore calculated in an orthorhombic cell of approximately $15 \times 15 \times 15$ Å^3. These dimensions allowed to sufficiently suppress the unphysical interaction between periodic images (<1 meV/atom). In most cases this yielded reliable single-particle energies, but when fractional energy-level occupancies were found or when the predicted ground-state electron configuration differed from experiment, the self-consistent cycle was manually constrained. Only for the spin-orbit-coupled calculation of the Pb atom, this was not possible, so there the PBE ground state $^1S_0$ was used instead of the experimental $^3P_0$ state.

The mechanical parameters V_0, B_0, and B_1 were computed from a 13-point equation of state, homogeneously spaced between $V = 0.94V_0$ and $V = 1.06V_0$. Each of these crystals was optimized with respect to the cell shape and ionic positions, followed by a single-point energy calculation. The data points were then fitted to a four-parameter Birch-Murnaghan equation [42], from which the relevant properties were extracted. Although some semiempirical equations were meant to be completed with room-temperature values, such as Eq. (2), we assumed thermal corrections on the input data to play only a minor role.

B. Error analysis

For both properties, α_V and T_m, reliable error estimates were determined by applying a least-squares linear regression $Y = \beta_1 X + \beta_0 + \epsilon$ to the experimental values Y as a function of the (DFT-based) semiempirical predictions X [43] (see Fig. 1). The regression slope β_1 and the intercept β_0 express the systematic deviation between experiment and theory, while the standard error of the regression (SER), equal to the standard deviation of the zero-centered, normally distributed error ϵ, provides an error bar for the fit.

If numerical convergence is achieved in all DFT calculations, then β_1, β_0, and the SER are intrinsic to the used model, with an ideal model yielding $\beta_1 = 1$, $\beta_0 = 0$, and SER = 0. In practice, however, there is both a systematic deviation and a residual scatter, which can be attributed to the semiempirical relation and to the DFT predictions themselves. Some data

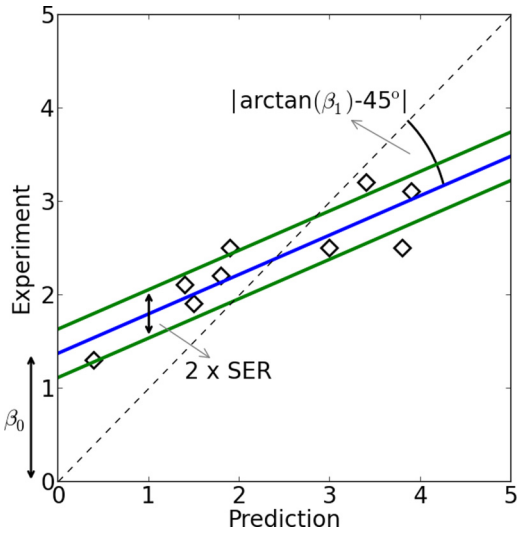


FIG. 1. (Color online) The error on semiempirical predictions can be quantified by the systematic deviation (β_0 and β_1) and the residual error bar (SER), which follow from a least-squares linear regression between the experimental and theoretical values. For this hypothetical data set, the middle full line (in blue) represents the regression curve, and the dashed line represents the first-quadrant bisector.

points even deviate so strongly that they become outliers with respect to the general trend. The resulting distortion of the linear regression substantially complicates the search for a realistic fit. To deal with such deviating data points, an exclusion criterion was developed to identify materials which the semiempirical predictions fail to appropriately describe. A flowchart of the method is presented in Fig. 2, consisting of two parts: as it is hard to automatically identify outliers when they distort the regression curve, a RANSAC procedure [44] was first performed to minimize the influence of outliers on the fit (upper part of Fig. 2); RANSACs are hard to combine with statistical considerations, however, so the resulting regression line was then used as a starting point for a more refined removal of outliers (bottom row of Fig. 2).

RANSAC algorithms identify the most probable regression curve by comparing many inlier-outlier combinations. To avoid bias, that process is randomized. For our 50- to 70-point data sets, we randomly took 100 two-point samples (Fig. 2: Nos. 1 to 100, first row) and collected inliers (filled symbols) in an iterative fashion (reversed arrows). This was done by enforcing a threshold of seven times the SER (using the overall SER in the first step) and adding all points that lay closer to the fit (second row). Then the regression line and the SER were re-established (third row), and the procedure was repeated until no more points could be added. Using such a flexible threshold allows detecting the most important outliers, while losing none of the inliers. Afterwards, the quality of each of the 100 fits was defined as the median error of the inliers, and only the fit with the smallest median of squared residuals was withheld.

The resulting set of preliminary inliers was used to remove additional outliers in a statistically rigorous way, again working in an iterative and elementwise fashion (bottom row of Fig. 2): we excluded the compound with the most extreme externally studentized residual and redid the fit, for as long as that residual belonged to the outer 0.1% of a

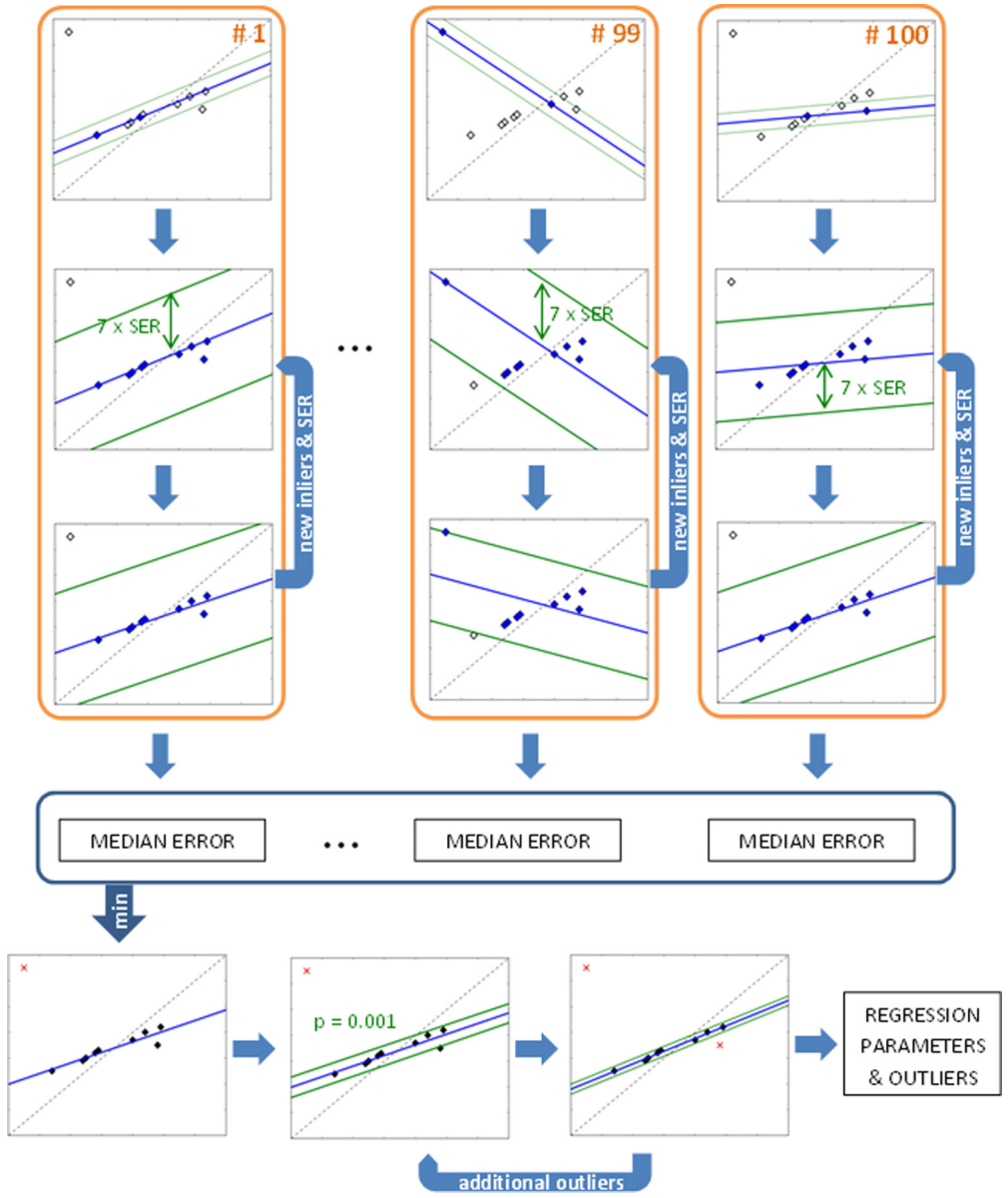


FIG. 2. (Color online) Flowchart of the applied error analysis procedure for a hypothetical data set. First a RANSAC approach [44] is used (top four rows), comparing 100 randomly initiated regression lines (Nos. 1 to 100) and retaining the one with the smallest median error. Afterwards, that resulting fit is employed for a statistically more rigourous outlier exclusion (bottom row). A more elaborate explanation is provided in the text.

Student's t distribution (two-sided p value) [43]. Although such a stringent significance threshold might not get rid of all outliers, it prevents inliers from being excluded and the remaining data set already yields reliable estimates for the regression parameters.

IV. RESULTS AND DISCUSSION

A. Thermal expansion coefficient

By fitting experimental results [45–51] to predicted ones, we estimated the errors introduced by the different semiempirical α_V approximations. This yielded a measure of the systematic and residual errors (Sec. III B). The results are summarized in Table I and are essentially independent of the used code (see Ref. [37] for a comparison of error estimates for the Slater approximation using VASP and GPAW [52–54]). Individual residual errors are shown in Tables II and III and Fig. 3, also indicating the outliers, which were excluded from the linear regression (hatched elements or open symbols). A complete overview of all theoretical and experimental numbers is presented in Ref. [37].

1. Elementwise deviations

Table I and Fig. 3 confirm that the correspondence between the different semiempirical predictions and experiment is not perfect. First, there is some systematic deviation, as most regression slopes and intercepts deviate from the ideal values ($\beta_1 = 1, \beta_0 = 0$). The two-sided p values with respect to a null hypothesis of $\beta_1 = 1$, for example, are smaller than 0.1% for

TABLE I. Error estimates for the semiempirically predicted volumetric thermal expansion coefficients α_V (in 10^{-5} K^{-1}) and melting temperatures T_m (in degrees Kelvin), based on a least-squares regression $Y = \beta_1 X + \beta_0$ between experimental data Y and semiempirical data X for the ground-state elemental crystals. 95% confidence intervals have been provided for the standard error of the regression (SER) and the regression parameters β_1 and β_0. The columns headed $P(\beta_1 = 1)$ and $P(\beta_0 = 0)$ represent the two-sided probabilities that, if in fact $\beta_1 = 1$ or $\beta_0 = 0$, respectively, the current regression parameters are obtained. The final column displays the number of outliers (see Sec. III B) with respect to the total number of test set crystals.

Approximation		SER	β_1	β_0	$P(\beta_1 = 1)$	$P(\beta_0 = 0)$	Number of outliers
α_V (Slater)	[Eq. (9)]	$1.7^{+0.4}_{-0.3}$	0.852 ± 0.041	-0.6 ± 0.8	4×10^{-9}	0.12	10/56
α_V (Dugdale-MacDonald)	[Eq. (10)]	$1.8^{+0.5}_{-0.3}$	1.056 ± 0.051	-0.7 ± 0.8	0.03	0.07	10/56
α_V (Vashchenko-Zubarev)	[Eq. (11)]	$2.1^{+0.5}_{-0.4}$	1.354 ± 0.072	-0.8 ± 1.0	6×10^{-13}	0.10	10/56
α_V (Guinea *et al.*)	[Eq. (14)]	$1.9^{+0.5}_{-0.3}$	0.894 ± 0.037	-0.8 ± 0.8	4×10^{-7}	0.05	2/56
α_V (Tsuru *et al.*)	[Eq. (1)]	$2.4^{+0.6}_{-0.4}$	1.324 ± 0.061	-1.4 ± 1.0	1×10^{-14}	0.01	3/56
T_m (via B_0)	[Eq. (17)]	466^{+96}_{-68}	1.000 ± 0.003	0 ± 198	1.00	1.00	0/68
T_m (Guinea *et al.*)	[Eq. (15)]	283^{+58}_{-41}	1.042 ± 0.003	8 ± 113	0.00	0.89	0/68

all but one of the α_V approximations: if the correct regression slope were 1 nevertheless, then the probability of obtaining test set data as extreme as ours would be less than 0.1%. Second, the remaining scatter cannot be neglected either, as appears from the residual errors and outliers (Tables II and III). Indeed, all SERs lie between 1.7 and 2.5×10^{-5} K^{-1}, which is in some cases of the same order of magnitude as the expansion coefficients themselves (see Fig. 3). To explain the largest deviations, we look at a few elements in more detail.

Several approximations of the thermal expansion coefficient rely on accurate DFT values for the volume and the bulk modulus. However, for some subsets of elements, PBE is not able to predict those properties correctly [13]. This is particularly the case for the correlation-dominated materials, the molecular crystals and the noble gases. Indeed, C, S, and Cd were indicated as outliers or have large residual errors in all α_V approximations based on Eq. (2). For the Slater, Dugdale-MacDonald, and Vashchenko-Zubarev approaches, reliable B_1 values are needed as well. Given PBE's inaccurate B_1 predictions for the low-coordination p block compounds and Sr, these α_V methods do not adequately describe P, As, Se, Sb, Te, Bi, and Sr either. Finally, the approximations of Guinea *et al.* and Tsuru *et al.* depend on the cohesive energy. For ΔE_{coh}, the correlation-dominated compounds and the noble gases fall short. It explains why the α_V predictions for Cd, and to a lesser extent Zn, deviate considerably.

However, not all differences between the theoretical and the experimental results are due to the accuracy with which the DFT-based predictors are determined. The semiempirical approximations themselves can also introduce errors, as they do for Cr, for example. We cannot expect any of the investigated approximations to yield meaningful results for this material, as it displays a magnetic phase transformation around room temperature [55]. Such phenomena are not included in any of the equations, which purely relate thermal expansion to vibrational or bonding types of interactions. Strikingly, the diamond-type structures (Si, Ge and Sn) yield unreliable results as well, while from a DFT point of view, the necessary predictors are determined up to a high degree of accuracy. Here too, the experimental expansion behavior contains some strange features, with negative α_V values over a considerable temperature range, but explicit phonon calculations are able to reproduce it (see later, Table V). This indicates that the semiempirical approaches are to blame. They fail to describe the negative Grüneisen parameters of the low-frequency phonon modes, which give rise to the initial low-temperature compression [56]. While the experimental results for γ incorporate these negative contributions, the semiempirical approximations do not. An estimate for γ that is based on bulk properties, only represents the behavior of average phonons, rather than the average phonon behavior: at low temperatures, the low-energy phonons are occupied the most, and their behavior can deviate significantly from the behavior of an average phonon. Bulk γ values hence fail when the low-frequency phonon modes differ too much from the rest. Negative Grüneisen parameters are not only found with the diamond-type compounds, but occur whenever bonds are strengthened during expansion. This happens in certain bending modes [56,57], leading to phonon frequencies that increase with volume. However, overall bond strengthening only occurs when the vibrational mode does not hinder neighboring atoms. The diamond structure, with its tetragonal environment, is ideally suited for such a behavior, but so are compounds with a low coordination number in general. The bad thermal expansion predictions of Se, Sb, and such are therefore as much due to the semiempirical relations themselves as to the unreliable PBE values for B_1. Using experimental values for B_1 does indeed not improve the quality of the predictions (see Tables 2.5 and 2.6 in Ref. [37]).

2. *Comparison between methods*

Despite similar shortcomings in each of the α_V approximations, there is a clear difference in overall quality. Systematic deviations can be corrected for—this is equivalent to proposing a new semiempirical equation—but the remaining scatter is intrinsic to the particular set of predictors and the form of the semiempirical function. The SERs (Table I) suggest that the relations by Slater, Dugdale and MacDonald, and Guinea *et al.* yield the most accurate results. These SERs are not always based on the same number of data points, however. In particular, less outliers are excluded from the regression when the methods of Guinea *et al.* and Tsuru *et al.* are evaluated, even though they have some classic outliers only slightly below the exclusion threshold (see Sec. III B). Indeed, Table III

TABLE II. (Color online) (Absolute) elementwise deviations between the experimental [45–51] room-temperature volumetric thermal expansion coefficients $\alpha_V(T_{rt})$ and least-squares linear regression results based on semiempirical predictions in the Slater [Eq. (9)], Dugdale-MacDonald [Eq. (10)], and Vashchenko-Zubarev [Eq. (11)] approximations (in 10^{-5} K^{-1}). The darkest shades correspond to the largest errors (common color code for Tables II and III) and hatched areas indicate outliers.

$\alpha_{V,rt}$(Slater)

H																	He
													B	C	N	O	F	Ne
Li	Be												1.1	928				
3.4	1.3												Al	Si	P	S	Cl	Ar
Na	Mg												1.6	2.5	140	2530		
3.7	0.4												Ga	Ge	As	Se	Br	Kr
K	Ca	Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn		1.8	3.3	32	106		
3.2	0.4	0.4	0.2	0.6	3.9	2.6	0.3	1.2	0.8	0.5	1.1		In	Sn	Sb	Te	I	Xe
Rb	Sr	Y	Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag	Cd		0.6	3.8	9	23		
0.6	6.3	0.1	0.3	0.7	0.3	1.0	0.6	0.7	0.6	0.2	2.4		Tl	Pb	Bi	Po	At	Rn
Cs	Ba	Lu	Hf	Ta	W	Re	Os	Ir	Pt	Au	Hg		1.5	2.0	10	1.1		
6.9	2.0	0.5	0.1	0.7	0.5	1.1	0.6	0.8	0.9	0.9								

$\alpha_{V,rt}$(DM)

H																	He
													B	C	N	O	F	Ne
Li	Be												1.0	1070				
2.3	0.9												Al	Si	P	S	Cl	Ar
Na	Mg												1.5	2.4	166	3034		
4.6	0.5												Ga	Ge	As	Se	Br	Kr
K	Ca	Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn		2.0	3.5	38	125		
3.5	0.9	0.1	0.4	0.8	4.5	2.2	0.2	1.2	0.7	0.3	1.9		In	Sn	Sb	Te	I	Xe
Rb	Sr	Y	Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag	Cd		0.1	4.0	11	27		
0.8	7.0	0.4	0.1	0.9	0.4	1.1	0.7	0.6	0.5	0.2	3.5		Tl	Pb	Bi	Po	At	Rn
Cs	Ba	Lu	Hf	Ta	W	Re	Os	Ir	Pt	Au	Hg		2.2	1.7	12	0.7		
7.9	1.3	0.3	0.3	0.8	0.6	1.2	0.7	0.8	0.8	0.7								

$\alpha_{V,rt}$(VZ)

H																	He
													B	C	N	O	F	Ne
Li	Be												0.9	1270				
0.3	0.3												Al	Si	P	S	Cl	Ar
Na	Mg												1.3	2.4	204	3760		
6.3	0.7												Ga	Ge	As	Se	Br	Kr
K	Ca	Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn		2.3	3.8	47	150		
4.3	1.6	0.2	0.6	0.9	5.4	1.7	0.1	1.1	0.5	0.0	3.0		In	Sn	Sb	Te	I	Xe
Rb	Sr	Y	Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag	Cd		1.0	4.3	13	33		
1.6	7.9	0.8	0.1	1.0	0.5	1.1	0.6	0.5	0.2	0.8	5.0		Tl	Pb	Bi	Po	At	Rn
Cs	Ba	Lu	Hf	Ta	W	Re	Os	Ir	Pt	Au	Hg		3.3	1.2	15	0.2		
10	0.2	0.0	0.5	0.9	0.7	1.2	0.7	0.7	0.7	0.3								

shows that the low-coordination compounds and molecular crystals also perform badly for these methods. To compare all approximations on an equal footing, we looked at the SERs of the common inliers [58] (Table IV, set II). These reduced to 1.7, 1.7, 1.9, 1.6, and 1.9×10^{-5} K^{-1} for the Slater, Dugdale-MacDonald, Vashchenko-Zubarev, Guinea and Tsuru relations respectively. Among the different approximations to α_V, the one by Guinea *et al.* therefore performs best. It incorporates

TABLE III. (Color online) (Absolute) elementwise deviations between the experimental [45–51] room-temperature volumetric thermal expansion coefficients $\alpha_V(T_{rt})$ and least-squares linear regression results based on semiempirical predictions in the approximations of Guinea *et al*. [Eq. (14)] and Tsuru *et al*. [Eq. (1)] (in 10^{-5} K^{-1}). The darkest shades correspond to the largest errors (common color code for Tables II and III) and hatched areas indicate outliers.

$\alpha_{V,rt}$(Guinea)

H																	He
Li	Be											B	C	N	O	F	Ne
1.2	1.4											0.1	25				
Na	Mg											Al	Si	P	S	Cl	Ar
2.9	1.2											2.7	2.1	3.7	26		
K	Ca	Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn	Ga	Ge	As	Se	Br	Kr
3.7	1.4	0.6	0.1	0.5	1.4	3.8	1.0	1.6	1.5	1.3	0.9	0.4	2.1	2.3	0.7		
Rb	Sr	Y	Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag	Cd	In	Sn	Sb	Te	I	Xe
0.4	3.2	0.5	0.3	0.7	0.2	1.0	0.7	1.0	0.9	1.2	3.4	3.4	2.8	2.4	2.8		
Cs	Ba	Lu	Hf	Ta	W	Re	Os	Ir	Pt	Au	Hg	Tl	Pb	Bi	Po	At	Rn
3.8	3.3	0.7	0.2	0.7	0.5	1.1	0.8	0.9	1.1	1.3		0.3	2.4	3.1	0.7		

$\alpha_{V,rt}$(Tsuru)

H																	He
Li	Be											B	C	N	O	F	Ne
3.3	0.4											0.8	1.3				
Na	Mg											Al	Si	P	S	Cl	Ar
5.0	3.9											2.7	2.1	5.8	20		
K	Ca	Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn	Ga	Ge	As	Se	Br	Kr
4.4	1.9	0.2	0.4	0.6	1.9	2.9	1.0	1.5	1.4	0.8	6.9	0.5	2.0	0.0	11		
Rb	Sr	Y	Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag	Cd	In	Sn	Sb	Te	I	Xe
0.2	3.7	0.0	0.0	0.8	0.3	0.7	0.4	0.6	0.3	0.7	15	2.7	3.1	2.4	1.5		
Cs	Ba	Lu	Hf	Ta	W	Re	Os	Ir	Pt	Au	Hg	Tl	Pb	Bi	Po	At	Rn
3.7	2.6	0.3	0.1	0.8	0.5	0.8	0.5	0.5	0.4	0.5		2.3	0.2	3.8	3.4		

the largest number of bulk properties and is not based on B_1, which can not always be reliably obtained from PBE. The other methods do rely on B_1 or contain too few input parameters to fully describe thermal expansion.

3. *Source of the errors*

As mentioned earlier, the SER reflects the influence of two distinct effects. On the one hand, the use of DFT-PBE values for the predictors introduces some errors, since the agreement with experiment is not perfect. On the other hand, no semiempirical method can capture all aspects of reality perfectly. Both issues give rise to an appreciable deviation from the regression line. Their respective contributions to the total SER can be investigated by replacing the PBE predictors by experimental numbers [e.g., by completing Eq. (1) with experimental cohesive energies]. The resulting error is then entirely due to the (semi)empirical relation. By comparing to the original SERs, the influence of using DFT values instead of experimental ones can be evaluated. Such an analysis has been performed in Table IV. For a fair comparison between the PBE- and experiment-based predictions, equal test sets were considered by only assessing elements that are inliers for all of the methods (Table IV, set III). These common inliers are listed in Ref. [37].

The resulting SERs show that taking the predictors from PBE or from experiment does not change the overall scatter on the regression line much. In all cases, the values are very similar, which suggests that the main contribution to the SER is due to the semiempirical relation. The table also shows that the specific error value strongly depends on the size of the used test set. This is because ideal inliers or outliers do not exist, and smaller test sets will inevitably decrease the corresponding error. For the smallest considered set (set III), all semiempirical relations perform equally well. The method of Guinea *et al*. still has the widest applicability, however, since its SERs for sets II and III lie very closely together.

4. *Comparison to high-level theory*

Although some semiempirical approximations are better than others, fully first-principles methods are still expected to yield the most accurate results. The applicability of semiempirical methods hence largely depends on the difference with

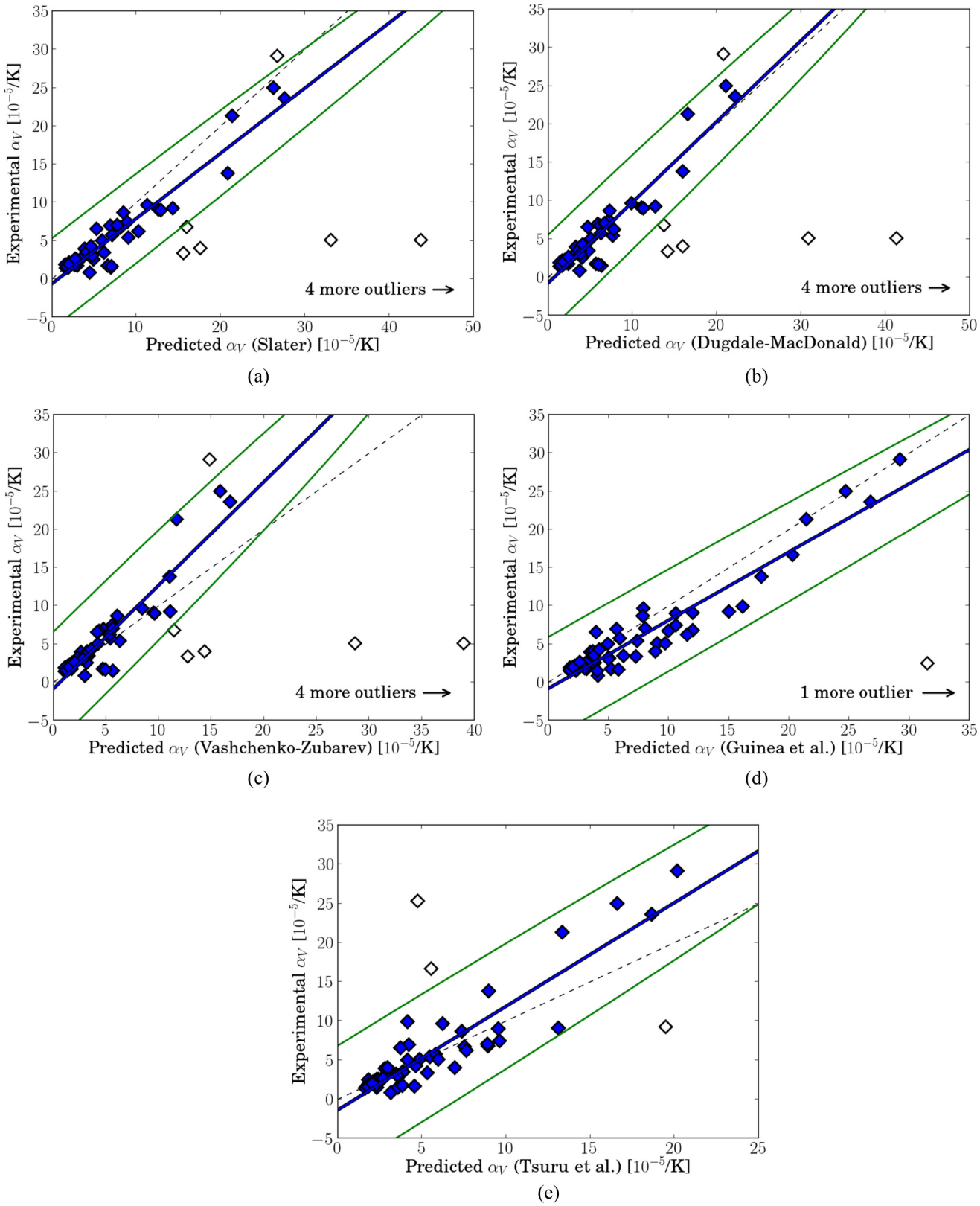


FIG. 3. (Color online) Least-squares linear regression curves (middle full line, in blue) between experimental [45–51] and semiempirically predicted values for the room-temperature volumetric thermal expansion coefficients of the ground-state elemental crystals. Outliers are depicted as open symbols and correspond with two-sided p values smaller than 0.1% for the externally studentized residuals (upper and lower full lines, in green). The dashed line represents the first-quadrant bisector, with which the regression line would coincide if all predictions were perfect.

TABLE IV. Standard errors of the regression (SER) for the semiempirically predicted volumetric thermal expansion coefficients α_V (in 10^{-5} K^{-1}) and melting temperatures T_m (in degrees Kelvin), based on a least-squares regression $Y = \beta_1 X + \beta_0$ between experimental data Y and semiempirical data X for the ground-state elemental crystals. Set I represents the full data set for each of the PBE-based methods (excluding their respective outliers) and is identical to Table I. Sets II and III, on the other hand, only contain the common inliers between all PBE-based predictions (II) or between all PBE- and experiment-based predictions (III) of either α_V or T_m (these sets are listed in Ref. [37]).

Approximation		PBE - set I	PBE - set II	PBE - set III	Exp - set III
α_V (Slater)	[Eq. (9)]	1.7	1.7	1.4	1.3
α_V (Dugdale-MacDonald)	[Eq. (10)]	1.8	1.7	1.4	1.3
α_V (Vashchenko-Zubarev)	[Eq. (11)]	2.1	1.9	1.5	1.5
α_V (Guinea *et al.*)	[Eq. (14)]	1.9	1.6	1.5	1.6
α_V (Tsuru *et al.*)	[Eq. (1)]	2.4	1.9	1.4	1.4
T_m (via B_0)	[Eq. (17)]	466	466	473	496
T_m (Guinea *et al.*)	[Eq. (15)]	283	283	295	282

those high-accuracy predictions. As mentioned in Sec. II A, the thermal expansion coefficient can be determined more rigorously from quasiharmonic phonon theory. Grabowski *et al.*, for example, applied it to the nonmagnetic fcc elemental metals using the PBE functional [59]. They still noticed considerable errors, however, especially for silver and gold. These relatively large errors for the phonon approach were also observed by Souvatzis and Eriksson [60], using PW91 calculations for the 4*d* transition metals. In both studies, the errors on the phonon predictions were of the same magnitude or even larger than the errors on most semiempirical predictions discussed here. Souvatzis and Eriksson suggested the LDA functional to outperform GGAs for transition metals, but only by fortuitous cancellation of errors. In fact, only more advanced functionals substantially improve phonon-based results for the transition metals [61]. PBE predictions are more reliable for non-transition-metal compounds, as is shown by Al and Pb in the study of Grabowski *et al.*, and by our own results for Si and K (Table V, using phonopy [62]; see Ref. [37] for further computational details).

Table V compares some PBE-based phonon results to semiempirical predictions. Systematic deviations were determined and corrected for by establishing a linear regression between these 12 experimental and theoretical data points. Hence, instead of the phonon-based or semiempirical values X themselves, the regression-corrected numbers $\beta_1 X + \beta_0$ are listed. Even for the phonon approach, a small systematic deviation was found, since GGA functionals are known to underbind crystals [63], yielding too large volumes and expansion coefficients.

At first sight, the SER for quasiharmonic phonon theory seems to be best, but this impression is entirely due to the results for silicon. Indeed, as discussed before, all semiempirical approaches fail to reliably represent its thermal expansion. When excluding the influence of Si on the SER, the semiempirical methods offer a similar (or even better) accuracy as the PBE-based phonon predictions, even for nontransition metals (Table V). It is therefore important to realize that explicit phonon calculations only have added value when the semiempirical relations break down and when the PBE functional itself is still valid (e.g., when no dispersion interactions are present). In all other cases, the thermal expansion coefficient can safely be calculated from a semiempirical approach, as this often provides the same degree of accuracy, but at a much smaller computational cost. When that accuracy is insufficient, phonon calculations may prove more suitable, but only by exploring better functionals [61]. The difference in computational effort will increase even further, however.

B. Melting temperature

The melting temperature can be estimated from both the cohesive energy [Eq. (15)] [12] and the bulk modulus [27]. To our knowledge, no explicit relation has been published for the

TABLE V. Volumetric thermal expansion coefficient (in 10^{-5} K^{-1}) according to experiment [45], the (regression-corrected) quasiharmonic phonon approach and the (regression-corrected) approximations by Slater [Eq. (9)], Dugdale-MacDonald [Eq. (10)], Vashchenko-Zubarev [Eq. (11)], Guinea *et al.* [Eq. (14)], and Tsuru *et al.* [Eq. (1)]. The applied linear regressions $Y = \beta_1 X + \beta_0$ are based on the experimental numbers Y and the raw theoretical values X for these 12 compounds, and correct for systematical deviations. Transition metals are distinguished from other element types.

Compound	Exp [45]	QHA	Slater	DM	VZ	Guinea	Tsuru
Ni	4.02	3.23 [64]	3.79	3.74	3.67	3.52	2.70
Cu	4.95	4.76 [59]	5.17	5.24	5.37	4.81	4.55
Rh	2.46	2.44 [59]	2.12	1.97	1.73	2.35	1.85
Pd	3.54	3.60 [59]	3.41	3.41	3.43	3.63	4.14
Ag	5.67	6.57 [59]	6.34	6.70	7.30	5.81	7.17
Ir	1.92	1.90 [59]	1.38	1.16	0.81	1.77	1.22
Pt	2.64	2.94 [59]	2.08	1.95	1.73	2.40	2.26
Au	4.26	5.82 [59]	3.91	4.03	4.24	4.04	5.32
Al	6.93	6.10 [59]	6.18	6.21	6.27	5.55	4.63
Pb	8.67	7.84 [59]	7.65	7.94	8.43	7.80	9.69
Si	0.78	0.67	3.78	3.59	3.28	3.95	2.99
K	24.99	24.95	25.03	24.88	24.55	25.20	24.31
Applied β_1		0.869	0.978	1.230	1.652	1.031	1.588
Applied β_0		−0.1	−0.6	−1.0	−1.7	−0.3	−2.0
SER with Si		0.73	1.10	1.07	1.12	1.15	1.35
SER[a] without Si		0.77	0.58	0.64	0.83	0.59	1.22

[a]Strictly speaking, the systematic deviation should be redetermined when the SER without Si is required. However, this one crystal does not influence the regression line much, and we are only interested in a measure of the effect of Si on the SER.

TABLE VI. (Color online) (Absolute) elementwise deviations between the experimental [27,65,66] melting temperatures T_m and least-squares linear regression results based on semiempirical predictions in the approximations of Eq. (17) and of Guinea *et al*. [Eq. (15)] (in degrees Kelvin). The darkest shades correspond to the largest errors (common color code).

$T_m(B_0)$

H																	He
472																	
Li	Be											B	C	N	O	F	Ne
142	76											162		426	440	442	467
Na	Mg											Al	Si	P	S	Cl	Ar
175	141											181	479	341	97	322	405
K	Ca	Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn	Ga	Ge	As	Se	Br	Kr
176	487	887	544	237	204	698	191	416	338	242	263	574	248	398	25	230	372
Rb	Sr	Y	Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag	Cd	In	Sn	Sb	Te	I	Xe
193	468	980	881	864	264	444	497	325	32	23	189	345	270	170	95	112	326
Cs	Ba	Lu	Hf	Ta	W	Re	Os	Ir	Pt	Au	Hg	Tl	Pb	Bi	Po	At	Rn
197	447	1068	1145	1233	777	18	337	516	413	300	354	124	178	122	233		

$T_m(\Delta E_{coh})$

H																	He
3																	
Li	Be											B	C	N	O	F	Ne
176	112											49		50	38	37	9
Na	Mg											Al	Si	P	S	Cl	Ar
56	332											402	83	474	798	146	39
K	Ca	Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn	Ga	Ge	As	Se	Br	Kr
8	362	203	178	94	560	22	74	228	149	5	258	724	243	68	522	209	98
Rb	Sr	Y	Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag	Cd	In	Sn	Sb	Te	I	Xe
5	409	179	302	59	471	171	62	47	386	264	299	473	731	151	218	308	142
Cs	Ba	Lu	Hf	Ta	W	Re	Os	Ir	Pt	Au	Hg	Tl	Pb	Bi	Po	At	Rn
17	263	344	65	214	270	401	210	1	23	127	144	17	163	267	109		

latter, however. Because the data suggest a linear correlation (Fig. 2.3 in Ref. [37]), we fitted a linear regression between the experimental values of T_m [27,65,66] and the PBE values of B_0, yielding

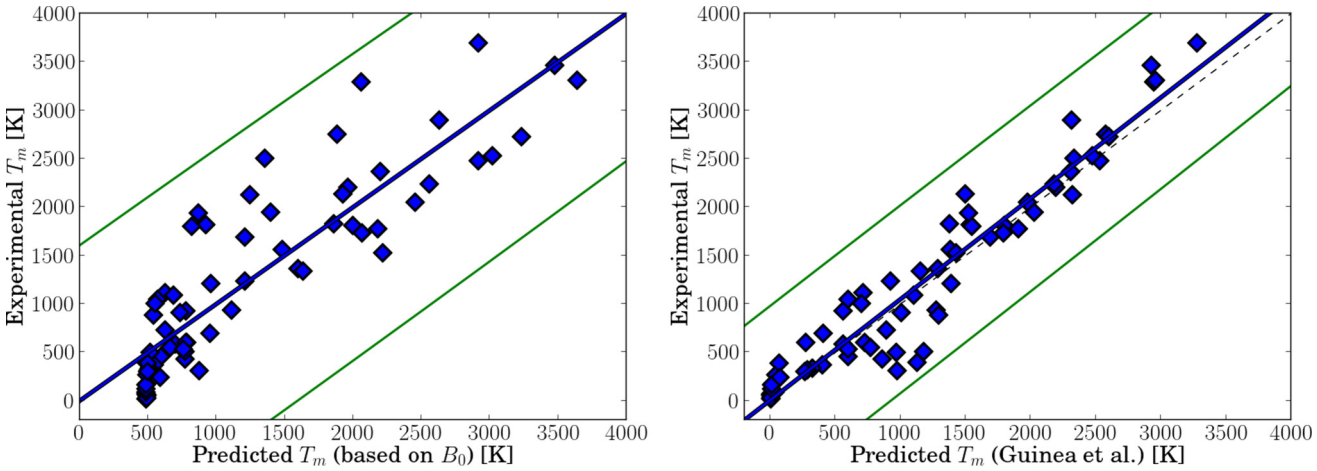
$$T_m = 482.8 + 8.172\, B_0. \tag{17}$$

Of course, there will be no systematic deviation for this equation by definition, but the effects of using PBE are expected to be similar as with the bulk modulus [13]. All other error characteristics remain useful, and are listed in Table I, together with the errors for $T_m(\Delta E_{\text{coh}})$. These values again do not depend much on the used code [37]. The individual residual errors are shown in Table VI and are visualized in Fig. 4.

1. *Comparison between methods*

Predicting a material's melting temperature from B_0 clearly yields less accurate results than using the cohesive energy. The SER is almost double in the first case (Table I). One possible reason is that Eq. (17) employs the relation between T_m and B_0 to capture the link between elasticity and melting behavior, while this link could also be expressed more generally. To check this, we fitted the melting temperature to a linear combination of the elastic constants C_{ij}, using PBE data for the cubic and hexagonal elemental crystals (see Table 2.8 in Ref. [37]). The resulting SER was indeed smaller than for a fit to B_0 (489 K instead of 512 K), but only slightly. We can therefore assume the relation between the melting temperature and the bulk modulus to be sufficiently representative of the physical connection between melting and elasticity.

The different quality of the two semiempirical approaches to T_m must hence be explained differently. Since the effect of using DFT-based predictors is again small (Table IV), that difference is mainly due to the semiempirical approaches themselves. In that respect, it is important to realize that melting goes beyond a purely elastic behavior. When a compound starts melting, the supplied heat needs to break all bonds. This is a process which is well described by the cohesive energy, since it relates to the initial and final stages of the bond breaking. Elastic properties only describe the initial stage of bond stretching. To account for non-negligible deformations, anharmonic effects need to be included [12,67], and a function of B_0 does not contain that information.



(a) (b)

FIG. 4. (Color online) Least-squares linear regression curves (middle full line, in blue) between experimental [27,65,66] and semiempirically predicted values for the melting temperatures of the ground-state elemental crystals. Outliers are depicted as open symbols and correspond with two-sided p values smaller than 0.1% for the externally studentized residuals (upper and lower full lines, in green). The dashed line represents the first-quadrant bisector, with which the regression line would coincide if all predictions were perfect.

2. *Comparison to high-level theory*

Not only the comparison between approximations is of interest, but so is their overall quality. This again requires looking at fully first-principles methods. Unfortunately, the number of high-quality results for the melting temperature is quite small due to the huge computational effort associated with *ab initio* molecular dynamics (AIMD). Table VII lists relevant data for five elemental compounds (i.e., based on a PBE or PW91 functional and at ambient pressure). We only provided values from the coexistence method (see Sec. II B), as these are the most frequently encountered AIMD results, and because it might be imprudent to treat all high-level approaches on the same footing. To correct for systematic deviations, we fitted separate regression lines to these five materials. Even for AIMD results, some deviation was found (and corrected for), since PBE underbinds crystals and therefore affects the simulated melting behavior. Note that the AIMD value for Li is based on the bcc structure (which is the structure at melting), while the semiempirical predictions relate to an hR9 crystal. The cohesive energy and the bulk modulus are very similar in both structures, however (a 0.1 kJ/mol and a 0.2 GPa difference, respectively), leading to almost identical melting temperature predictions.

Table VII shows that most T_m calculations are quite accurate, especially those from molecular dynamics or the relation by Guinea *et al*. Indeed, not only do the (regression-corrected) AIMD results correspond almost perfectly to the experimental values, the quality of Eq. (15) is also extremely good, particularly when the simplicity of the underlying computation is considered. Contrary to the predictions of α_V, however, the more fundamental method (in this case AIMD) remains the most accurate for T_m.

TABLE VII. Melting temperature (in degrees Kelvin) from experiment, (regression-corrected) fully *ab initio* molecular dynamics calculations (AIMD), and (regression-corrected) semiempirical relations in terms of ΔE_{coh} [Guinea *et al*., Eq. (15)] and B_0 [Eq. (17)]. The applied linear regressions $Y = \beta_1 X + \beta_0$ are based on the experimental numbers Y and the raw theoretical values X for these five compounds, and correct for systematical deviations.

Compound	Exp [27]	AIMD	Guinea	via B_0
Li	454	522 [68]	300	535
Al	934	870[a] [6]	1153	1314
Si	1687	1664 [69]	1676	1455
Ta	3293	3316 [70]	3257	2735
W	3695	3690 [71]	3676	4023
Applied β_1		1.052	1.258	1.502
Applied β_0		43	−450	−360
SER		57	156	456

[a]PW91 instead of PBE functional.

V. CONCLUSIONS

Using a test set with all ground-state elemental crystals (up to Rn, not including the lanthanides), we quantified the accuracy of several semiempirical approaches to the thermal expansion coefficient α_V and the melting temperature T_m. The discrepancy between theoretical predictions and experimental values was expressed in terms of a systematic and a residual error. The systematic deviation was determined from the slope of a linear regression curve between experiment and theory, while the remaining error bar was defined as the standard error of the regression. Outliers were flagged based on their externally studentized residuals, allowing us to identify those compounds for which the α_V or T_m predictions were certainly not reliable anymore. This unreliability could be due to the failure of the semiempirical approach itself or to the limited accuracy of the underlying DFT predictor (with the largest contribution from the former effect, see Table IV). The results from the error analyses are presented in Table I and are summarized below.

For the thermal expansion coefficient, we assessed multiple predictions based on the Grüneisen parameter γ, as well as an inverse proportionality to the cohesive energy. They all yielded very good results, particularly the former approximations, with the best behavior when γ was derived from a universal binding-energy relation for metals [12]. There were a few exceptions, however, as some elemental crystals did not perform well for any of the semiempirical approaches. Several deviations from experiment were caused by the failure of the used PBE functional (when dispersion interactions were present, for example, such as for graphite). On the other hand, it was also shown that neither an expression in γ nor in ΔE_{coh} could reproduce effects which were caused by anomalous low-frequency phonon modes, as most notably happens in Si. In such materials, more advanced calculations are necessary, e.g., using quasiharmonic phonon theory. Nevertheless, except for these few special cases, the accuracy of most semiempirical approximations is as good or even better than for explicit phonon calculations (Table V).

The melting temperature was approximated by both a relation to the cohesive energy and to the bulk modulus, with the latter method yielding the largest residual errors. B_0 only relates to elastic phenomena, while melting involves strongly anharmonic effects. By contrast, the quality of the correlation to ΔE_{coh} was quite good. Requiring only a limited computational effort, its results lay about 300 K from the experimental values. *Ab initio* molecular dynamics does offer even better results (Table VII), but at a much higher cost.

Fully first-principles calculations are often regarded as the only means to obtain high-accuracy predictions for materials properties. Nevertheless, some semiempirical approximations offer a cheap alternative with the same quality of results (such as for α_V) or with only marginally larger errors (such as for T_m). These methods are particularly interesting for materials design, where time can be an issue.

ACKNOWLEDGMENTS

This work is supported by the Fund for Scientific Research–Flanders (FWO) by means of a fellowship and a project (G.0402.11N) and by the Research Board of Ghent University. Stefaan Cottenier acknowledges financial support from OCAS NV by an OCAS-endowed chair at Ghent University. Calculations were carried out using the Stevin Supercomputer Infrastructure at Ghent University, funded by Ghent University, the Hercules Foundation, and the Flemish Government (EWI Department).

[1] P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964).
[2] W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).
[3] R. Martin, *Electronic Structure: Basic Theory and Practical Methods* (Cambridge University Press, Cambridge, 2004).
[4] D. Alfè, M. J. Gillan, and G. D. Price, Nature (London) **401**, 462 (1999).
[5] T. Ogitsu, E. Schwegler, F. Gygi, and G. Galli, Phys. Rev. Lett. **91**, 175502 (2003).
[6] D. Alfè, Phys. Rev. B **68**, 064423 (2003).
[7] M. T. Dove, *Introduction to lattice dynamics* (Cambridge University Press, Cambridge, 1993).
[8] Y. Tsuru, Y. Shinzato, Y. Saito, M. Shimazu, M. Shiono, and M. Morinaga, J. Ceram. Soc. Jpn. **118**, 241 (2010).
[9] J. C. Slater, Phys. Rev. **57**, 744 (1940).
[10] J. S. Dugdale and D. K. C. MacDonald, Phys. Rev. **89**, 832 (1953).
[11] V. Ya. Vashchenko and V. N. Zubarev, Sov. Phys. Solid State **5**, 653 (1963).
[12] F. Guinea, J. H. Rose, J. R. Smith, and J. Ferrante, Appl. Phys. Lett. **44**, 53 (1984).
[13] K. Lejaeghere, V. Van Speybroeck, G. Van Oost, and S. Cottenier, Crit. Rev. Solid State **39**, 1 (2014).
[14] J. P. Perdew, J. A. Chevary, S. H. Vosko, K. A. Jackson, M. R. Pederson, D. J. Singh, and C. Fiolhais, Phys. Rev. B **46**, 6671 (1992).
[15] J. P. Perdew, J. A. Chevary, S. H. Vosko, K. A. Jackson, M. R. Pederson, D. J. Singh, and C. Fiolhais, Phys. Rev. B **48**, 4978 (1993).
[16] J. H. Rose, J. Ferrante, and J. R. Smith, Phys. Rev. Lett. **47**, 675 (1981).
[17] D. Alfè, M. J. Gillan, and G. D. Price, J. Chem. Phys. **116**, 6170 (2002).
[18] J. Bouchet, F. Bottin, G. Jomard, and G. Zérah, Phys. Rev. B **80**, 094102 (2009).
[19] J. D. Kubicki and A. C. Lasaga, Am. J. Sci. **292**, 153 (1992).
[20] J. Mei and J. W. Davenport, Phys. Rev. B **46**, 21 (1992).
[21] A. B. Belonoshko, Geochim. Cosmochim. Acta **58**, 4039 (1994).
[22] A. B. Belonoshko, N. V. Skorodumova, A. Rosengren, and B. Johansson, Phys. Rev. B **73**, 012201 (2006).
[23] A. B. Belonoshko, L. Burakovsky, S. P. Chen, B. Johansson, A. S. Mikhaylushkin, D. L. Preston, S. I. Simak, and D. C. Swift, Phys. Rev. Lett. **100**, 135701 (2008).
[24] J. F. Lutsko, D. Wolf, S. R. Phillpot, and S. Yip, Phys. Rev. B **40**, 2841 (1989).
[25] A. J. C. Ladd and L. V. Woodcock, Mol. Phys. **36**, 611 (1978).
[26] S. Yoo, X. C. Zeng, and S. S. Xantheas, J. Chem. Phys. **130**, 221102 (2009).
[27] C. Kittel, *Introduction to Solid State Physics*, 8th ed. (Wiley, Hoboken, 2005).
[28] C. Li and P. Wu, Chem. Mater. **14**, 4833 (2002).
[29] C. Li, J. L. Hoe, and P. Wu, J. Phys. Chem. Solids **64**, 201 (2003).
[30] J. Tateno, Solid State Commun. **10**, 61 (1972).
[31] M. E. Fine, L. D. Brown, and H. L. Marcus, Scr. Metall. **18**, 951 (1984).
[32] P. E. Blöchl, Phys. Rev. B **50**, 17953 (1994).
[33] G. Kresse and D. Joubert, Phys. Rev. B **59**, 1758 (1999).
[34] G. Kresse and J. Furthmüller, Comput. Mater. Sci. **6**, 15 (1996).
[35] J. Hafner, J. Comput. Chem. **29**, 2044 (2008).
[36] VASP *manual*, http://www.vasp.at (April 23, 2009 version).
[37] See Supplementary Material at http://link.aps.org/supplemental/10.1103/PhysRevB.89.014304 for more computational details, as well as the full calculated and experimental data.
[38] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).

[39] DFT poll, http://www.marcelswart.eu/dft-poll
[40] H. J. Monkhorst and J. D. Pack, Phys. Rev. B **13**, 5188 (1976).
[41] Comparing Solid State DFT Codes, Basis Sets and Potentials, http://molmod.ugent.be/DeltaCodesDFT
[42] F. Birch, Phys. Rev. **71**, 809 (1947).
[43] N. R. Draper and H. Smith, *Applied regression analysis*, 3rd ed., Wiley Series in Probability and Statistics (Wiley-Interscience, New York, 1998).
[44] M. A. Fischler and R. C. Bolles, Comm. ACM **24**, 381 (1981).
[45] *CRC Handbook of Chemistry and Physics*, edited by D. Lide, 89th ed. (CRC Press, Boca Raton, Florida, 2008-2009).
[46] J. B. Nelson and D. P. Riley, Proc. Phys. Soc. London **57**, 477 (1945).
[47] R. W. Keyes, Phys. Rev. **92**, 580 (1953).
[48] J. Wallis, I. Sigalas, and S. Hart, J. Appl. Crystallogr. **19**, 273 (1986).
[49] R. R. Pawar and V. T. Deshpande, J. Mater. Sci. **5**, 1061 (1970).
[50] M. S. Anderson and C. A. Swenson, Phys. Rev. B **28**, 5395 (1983).
[51] J. A. C. Marples and C. C. Koch, Phys. Lett. A **41**, 307 (1972).
[52] S. R. Bahn and K.W. Jacobsen, Comput. Sci. Eng. **4**, 56 (2002).
[53] J. J. Mortensen, L. B. Hansen, and K. W. Jacobsen, Phys. Rev. B **71**, 035109 (2005).
[54] 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, J. Phys.: Condens. Matter **22**, 253202 (2010).
[55] G. K. White, Aust. J. Phys. **14**, 359 (1961).
[56] C. H. Xu, C. Z. Wang, C. T. Chan, and K. M. Ho, Phys. Rev. B **43**, 5024 (1991).
[57] C. Lind, Materials **5**, 1125 (2012).
[58] In total, 12 possible outliers were excluded from all data sets, 11 of which are indicated in Tables II and III. The last one is Zn, which only appears after the first 11 outliers are removed from the data set of α_V(Tsuru). It is therefore omitted from the other data sets as well.
[59] B. Grabowski, T. Hickel, and J. Neugebauer, Phys. Rev. B **76**, 024309 (2007).
[60] P. Souvatzis and O. Eriksson, Phys. Rev. B **77**, 024110 (2008).
[61] A. Dal Corso, J. Phys.: Condens. Matter **25**, 145401 (2013).
[62] A. Togo, F. Oba, and I. Tanaka, Phys. Rev. B **78**, 134106 (2008).
[63] V. Ozoliņš and M. Körling, Phys. Rev. B **48**, 18304 (1993).
[64] M. Pozzo and D. Alfè, Phys. Rev. B **88**, 024111 (2013).
[65] N. N. Greenwood and A. Earnshaw, *Chemistry of the Elements*, 2nd ed. (Butterworth-Heinemann, Oxford, 1997).
[66] R. L. Mills and E. R. Grilly, Phys. Rev. **101**, 1246 (1956).
[67] J. H. Rose, J. R. Smith, F. Guinea, and J. Ferrante, Phys. Rev. B **29**, 2963 (1984).
[68] E. R. Hernández, A. Rodriguez-Prieto, A. Bergara, and D. Alfè, Phys. Rev. Lett. **104**, 185701 (2010).
[69] S. Yoo, S. S. Xantheas, and X. C. Zeng, Chem. Phys. Lett. **481**, 88 (2009).
[70] L. G. Wang and A. van de Walle, Phys. Chem. Chem. Phys. **14**, 1529 (2012).
[71] L. G. Wang, A. van de Walle, and D. Alfè, Phys. Rev. B **84**, 092102 (2011).