

How to verify the precision of density-functional-theory implementations via reproducible and universal workflows

Emanuele Bosoni¹, Louis Beal², Marnik Bercx³, Peter Blaha⁴, Stefan Blügel⁵, Jens Bröder^{5,6}, Martin Callsen^{7,8,9}, Stefaan Cottenier^{7,8}, Augustin Degomme², Vladimir Dikan¹, Kristjan Eimre³, Espen Flage-Larsen^{10,11}, Marco Fornari¹², Alberto Garcia¹, Luigi Genovese², Matteo Giantomassi¹³, Sebastiaan P. Huber^{3,14}, Henning Janssen⁵, Georg Kastlunger¹⁵, Matthias Krack¹⁶, Georg Kresse^{17,18}, Thomas D. Kühne^{19,20}, Kurt Lejaeghere^{8,21}, Georg K. H. Madsen⁴, Martijn Marsman^{17,18}, Nicola Marzari^{3,16}, Gregor Michalicek⁵, Hossein Mirhosseini²², Tiziano M. A. Müller²³, Guido Petretto¹³, Chris J. Pickard^{24,25}, Samuel Poncé¹³, Gian-Marco Rignanese¹³, Oleg Rubel²⁶, Thomas Ruh^{4,8}, Michael Sluydts^{7,8,27}, Danny E.P. Vanpoucke^{7,28}, Sudarshan Vijay¹⁵, Michael Wolloch^{17,18}, Daniel Wortmann⁵, Aliaksandr V. Yakutovich²⁹, Jusong Yu^{3,16}, Austin Zadoks³, Bonan Zhu^{30,31}, and Giovanni Pizzi^{3,16,*}

¹Institut de Ciència de Materials de Barcelona, ICMAB-CSIC, Campus UAB, 08193 Bellaterra, Spain

²Univ. Grenoble-Alpes, CEA, IRIG-MEM-L_Sim, 38000 Grenoble, France

³Theory and Simulation of Materials (THEOS) and National Centre for Computational Design and Discovery of Novel Materials (MARVEL), École Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland

⁴Institute for Materials Chemistry, Technical University of Vienna, Getreidemarkt 9/165-TC, A-1060 Vienna, Austria

⁵Peter Grünberg Institut and Institute for Advanced Simulation, Forschungszentrum Jülich and JARA, D-52425 Jülich, Germany

⁶Institute for Advanced Simulation, Materials Data Science and Informatics (IAS-9), Forschungszentrum Jülich, D-52425 Jülich, Germany

⁷Department of Electromechanical, Systems and Metal Engineering, Ghent University, Belgium

⁸Center for Molecular Modeling (CMM), Ghent University, Belgium

⁹Institute of Atomic and Molecular Sciences, Academia Sinica, Taipei 10617, Taiwan

¹⁰Norwegian EuroHPC Competence Center, Sigma2 AS, Norway

¹¹SINTEF Industry, Materials Physics, Oslo, Norway

¹²Department of Physics and Science of Advanced Materials Program, Central Michigan University, Mount Pleasant, Michigan 48859, USA

¹³Institut de la Matière Condensée et des Nanosciences (IMCN), Université catholique de Louvain, Chemin des Étoiles 8, Louvain-la-Neuve 1348, Belgium

¹⁴National Centre of Competence in Research (NCCR) Catalysis, École Polytechnique Fédérale de Lausanne (EPFL), CH-1015, Lausanne, Switzerland

¹⁵Center for Catalysis Theory (Cattheory), Department of Physics, Technical University of Denmark (DTU), 2800 Kongens Lyngby, Denmark

¹⁶Laboratory for Materials Simulations (LMS), Paul Scherrer Institut (PSI), CH-5232 Villigen PSI, Switzerland

¹⁷University of Vienna, Faculty of Physics and Center for Computational Materials Science, Kolingasse 14-16, A-1090 Vienna, Austria

¹⁸VASP Software GmbH, Sensengasse 8, A-1090 Vienna, Austria

¹⁹Center for Advanced Systems Understanding (CASUS) and Helmholtz-Zentrum Dresden-Rossendorf, D-02826 Görlitz, Germany

²⁰Paderborn Center for Parallel Computing (PC2) and Center for Sustainable Systems Design, University of Paderborn, D-33098 Paderborn, Germany

²¹OCAS NV/ArcelorMittal Global R&D Gent, Pres. J. F. Kennedylaan 3, Zelzate B-9060, Belgium

²²Dynamics of Condensed Matter, Chair of Theoretical Chemistry, University of Paderborn, D-33098 Paderborn, Germany

²³HPE HPC EMEA Research Lab, CH-4051 Basel, Switzerland

²⁴Department of Materials Science & Metallurgy, University of Cambridge, 27 Charles Babbage Road, Cambridge

CB3 0FS, United Kingdom

²⁵Advanced Institute for Materials Research, Tohoku University 2-1-1 Katahira, Aoba, Sendai, 980-8577, Japan

²⁶Department of Materials Science and Engineering, McMaster University, 1280 Main Street West, Hamilton, Ontario L8S 4L8, Canada

²⁷ePotentia, Frans van Dijckstraat 59, 2100 Deurne Antwerpen, Belgium

²⁸Institute for Materials Research (IMO-IMOMEC), UHasselt - Hasselt University, Belgium

²⁹Swiss Federal Laboratories for Materials Science and Technology (Empa), nanotech@surfaces laboratory, CH-8600 Dübendorf, Switzerland

³⁰Department of Chemistry, University College London, 20 Gordon St, Bloomsbury, London WC1H 0AJ, United Kingdom

³¹The Faraday Institution, Didcot OX11 0RA, United Kingdom

*e-mail: giovanni.pizzi@psi.ch

ABSTRACT

In the past decades many density-functional theory methods and codes adopting periodic boundary conditions have been developed and are now extensively used in condensed matter physics and materials science research. Only in 2016, however, their precision (i.e., to which extent properties computed with different codes agree among each other) was systematically assessed on elemental crystals: a first crucial step to evaluate the reliability of such computations. We discuss here general recommendations for verification studies aiming at further testing precision and transferability of density-functional-theory computational approaches and codes. We illustrate such recommendations using a greatly expanded protocol covering the whole periodic table from Z=1 to 96 and characterizing 10 prototypical cubic compounds for each element: 4 unaries and 6 oxides, spanning a wide range of coordination numbers and oxidation states. The primary outcome is a reference dataset of 960 equations of state cross-checked between two all-electron codes, then used to verify and improve nine pseudopotential-based approaches. Such effort is facilitated by deploying AiiDA common workflows that perform automatic input parameter selection, provide identical input/output interfaces across codes, and ensure full reproducibility. Finally, we discuss the extent to which the current results for total energies can be reused for different goals (e.g., obtaining formation energies).

Key points:

- Verification efforts are critical to assess the reliability of density-functional theory (DFT) simulations and provide results with properly quantified uncertainties.
- Developing standard computation protocols to perform verification studies and publishing curated and FAIR reference datasets can significantly facilitate their use to improve codes and computational approaches.
- The use of fully automated workflows with common interfaces between codes can guarantee uniformity, transferability, and reproducibility of results.
- A careful description of the numerical and methodological details needed to compare with the reference datasets is essential; we discuss and illustrate this point with a dataset of 960 all-electron equations of state.
- Reference datasets should always include an explanation of the target property for which they were generated, and a discussion of their limits of applicability.
- Further extensions of DFT verification efforts are needed to cover more functionals, more computational approaches, and the treatment of magnetic and relativistic (spin-orbit) effects. They should also aim at concurrently delivering optimized protocols that, not only target ultimate precision, but also optimize the computational cost for a target accuracy.

Website summary: Verification efforts of DFT calculations are of crucial importance to evaluate the reliability of simulation results. We discuss general recommendations for performing such studies and illustrate them with an all-electron reference dataset of 960 equations of state covering the whole periodic table (hydrogen to curium). The importance of verification for the improvement of pseudopotential codes is also demonstrated.

The fast improvement of hardware, methods, and tools for **density-functional theory (DFT)** calculations in periodic boundary conditions has greatly advanced the field of condensed matter physics and computational materials science, paving the way for an effective use of the “materials design process” that accelerates the discovery, development and deployment of new materials thanks to the aid of simulations^{1,2}. Efficient software infrastructures^{3–13} facilitate, nowadays, large high-throughput calculations of a panoply of material properties which are often made available to the public in large repositories^{14–21}. Most datasets aspire to be findable, accessible, interoperable, and reusable (FAIR)²² in order to accelerate materials discovery, possibly with the aid of machine learning. They are queryable with *ad hoc* application programming interfaces (APIs) or, for many of them, via a single common API thanks to the recent efforts of the OPTIMADE²³ consortium. However, full integration of different data is often limited by considerations related to uncertainty quantification^{1,24–27}. In this work, we discuss recommendations on how to quantify to which extent properties (total energies and derived quantities) obtained by different DFT codes agree among each other.

In principle, DFT applies the fundamental laws of quantum physics to predict properties of a material, with no other inputs than the chemical composition and the crystal structure. In reality, the electronic-structure calculations involve a variety of choices to solve the equations prescribed by DFT and introduce several levels of approximation. Those choices, reflected in the resulting data, range from the specific flavor of DFT (e.g., the approach used for the exchange-correlation functional) to the discretization assumptions (e.g., the basis set), to the specific computational parameters needed by the codes. Some approaches are more reliable, and therefore often slower, while others make more substantial approximations in order to gain computational speed and enable the study of systems with more atoms. Furthermore, even when formally the same choices have been made in different codes, these may provide slightly different results due to the details of their implementations. The importance of verifying the precision of codes has been long recognized²⁸. Despite this, when considering DFT codes adopting periodic boundary conditions, a first systematic assessment of their precision was performed in 2016, where the consistency of 40 computational approaches was assessed by calculating the **equation of state (EOS)** (i.e., the energy-versus-volume curve) for a test set of 71 elemental crystals^{29,30}. This so-called “ Δ -project” led to the conclusion that the mainstream codes were in very good agreement with each other, which was not the case a decade before. Despite being already a large project by itself, the “ Δ -project” was only the first step towards a careful verification of DFT calculations, which requires a much larger diversity of structural and chemical variables, as also discussed in the outlook of Ref. 29.

In this Expert Recommendation, we list a set of guiding principles to perform new verification studies of DFT calculations (see Box 1), as well as a recommendation (see Box 2) for users of DFT codes, encouraging them to refer to quantitative sources on the reliability and precision of the codes and computational approaches used in their publications. In order to illustrate these recommendations, we create a curated reference set of highly converged results for the EOS of 960 crystals, using two independent state-of-the-art **all-electron (AE) DFT** codes (FLEUR^{31,32} and WIEN2k^{33,34}). These 960 crystals cover all elements and a wide variety of structural and chemical environments in the form of four unary compounds and six oxides. The resulting data are shared on the Materials Cloud¹⁹ according to the FAIR²² principles. A key feature of our work is that the thousands of computations performed are implemented within a reproducible and automatic infrastructure. Specifically, the launching and management of all the DFT calculations is carried out using AiiDA^{3,4,35}. The choice of code-specific inputs and numerical parameters (called “protocols” in the following) are implemented in the publicly available `aidda-common-workflows` (ACWF) package^{36,37} together with a number of error handlers to recover automatically from typical failure modes of each code. This setup enables to easily generate new datasets and to extend the current work for the verification of other computational approaches (see also Box 4).

As we discuss later, some choices of numerical parameters (such as the smearing type and size, or the k-point integration mesh) must be performed consistently in order to make correct use of the dataset. The suggestions regarding how to use our reference dataset are summarized in Box 3. One of our recommendations for verification efforts is to develop metrics to quantify discrepancies between codes that depend on physically measurable quantities. We implement this recommendation by defining two new metrics (in addition to the Δ metric introduced in Ref. 38) to facilitate quantitative comparison of EOS results for pairs of codes or computational approaches, and we discuss their benefits.

Using these metrics, we then compare the EOS results of our reference dataset to the results obtained by a number of pseudopotential codes. The latter are designed to enhance computational efficiency by considering explicitly only “valence” electrons, which contribute to bonding^{39–41}. The codes considered here are: ABINIT^{13,42,43}, BigDFT⁴⁴, CASTEP⁴⁵, CP2K^{46,47}, GPAW^{48,49}, QUANTUM ESPRESSO^{50,51}, SIESTA^{52,53}, the SIRIUS⁵⁴ library (via its CP2K interface) and VASP^{55,56}. The numerical basis sets implemented in these codes include plane waves, Gaussians combined with plane waves, Daubechies wavelets, and atomic orbitals. For this reason, we do not label our results simply with the code name, but with a short string also indicating a few additional relevant parameters to better specify the details of the computational approach. We stress that the aim of this study is not to provide a ranking or to evaluate the quality of different codes, but to illustrate with a few examples the value of curated datasets generated following our recommendations. In particular, we illustrate its use to improve existing pseudopotentials and to assess the consistency of results of several computational approaches to compute the EOS within DFT.

Expert Recommendation Box 1, Summary of recommendations to perform verification studies of DFT calculations

- Quantitatively estimate the precision of DFT computational approaches and implementations with respect to exact numerical results. Provide adequate details of the verification protocols to ensure reproducibility of the results and a correct reuse in data-driven research, e.g., clarifying their range of applicability and specifying which parameters need to be fixed — independent of the approach — to ensure comparable results.
- Develop fully automated workflows to guarantee uniformity and transferability of parameters between computational approaches. This includes the definition and use of “standard protocols”, i.e., automated selection of numerical parameters — often specific to each computational approach — that can ensure numerically precise results.
- Publish curated reference datasets from systematic verification studies. Facilitate their use to improve other codes by making the datasets FAIR: findable and accessible on open repositories, interoperable by using standard formats and clear annotations, and reusable by specifying all parameters needed to reproduce the results. See Box 3 for an example.
- Organize the reference data in appropriate subsets by recognizing the diversity of focus and the non-uniform capabilities of available computational approaches (e.g., if some systems require additional effort to be supported by all codes).

Expert Recommendation Box 2, Summary of recommendations for users of DFT codes

- When publishing research that makes use of DFT codes, refer as much as possible to quantitative sources that document the precision of the numerical implementation (all-electron vs. pseudopotential, basis-set type and size, . . .).
- Equally important is a validation statement that refers to the accuracy of the chosen exchange–correlation functional to correctly and accurately address the physics at hand. Note that, however, this is beyond the scope of the current Expert Recommendation focusing on the precision of numerical implementations.
- Always cite the exact pseudopotentials that are used in published simulations, including the exchange–correlation functional, the library from which they were obtained and the exact library version, together with all the essential numerical parameters of the calculations (e.g., k-point integration mesh and smearing, basis set type and size or plane-wave cutoffs, . . .). Lack of this information results in essentially non-reproducible simulations.

Finally, in our Outlook, we discuss a set of recommendations (summarized in Box 4) on future extensions of verification efforts. On the one hand, we suggest to cover more exchange–correlation functionals, computational approaches, and treatment of magnetic and relativistic (spin-orbit) effects. On the other hand, we highlight how future studies should not only target ultimate precision, but also aim at delivering protocols that optimize the computational cost for a target accuracy. We stress that, in this Expert Recommendation, we limit all discussions to verification efforts: i.e., investigating code precision, that is, how codes reproduce the ideal theoretical results given by DFT (e.g., with a given choice of exchange–correlation functional). We do not discuss validation, i.e., accuracy with respect to the experimental results. While this is also a highly relevant topic (and we briefly mention it in Box 2), it is beyond the scope of this Expert Recommendation.

AE reference dataset for EOS parameters

In this section we discuss our reference dataset of EOS calculations, that we use to illustrate, with a practical example, how to implement the recommendations of Box 1. The results are obtained with the AE codes FLEUR and WIEN2k, using the PBE⁵⁷ exchange–correlation functional. The two codes use the linearized augmented plane waves plus local orbitals method, but differ in details of the basis set and some computational setup parameters.

Crystal-structures dataset

We compute the EOS on a dataset of 960 cubic crystal structures. In order to provide a chemically comprehensive dataset, we consider all elements in the periodic table from $Z = 1$ (hydrogen) to $Z = 96$ (curium). Furthermore, we systematically scan structural diversity and investigate the transferability to more complex chemical environments by examining, for each element, 4 mono-elemental cubic crystals (“unaries dataset”) and six cubic oxides (“oxides dataset”).

Specifically, the unaries dataset considers all elements in the **face-centered cubic (FCC)**, **body-centered cubic (BCC)**, **simple cubic (SC)** and diamond crystal structure, thus covering a wide range of coordination numbers (12, 8, 6, and 4, respectively); a total of 384 systems. More information about the crystal structures are presented in the Supplementary Information (SI) Table S1.1. The oxides dataset is composed of six cubic oxides for each of the 96 elements X, with chemical formula X_2O , XO , X_2O_3 , XO_2 , X_2O_5 and XO_3 , thus totaling 576 additional structures, whose crystal structures are detailed in SI Table S1.2. The oxide stoichiometries are chosen such that the formal oxidation state of the element considered varies from +1 to +6. We note that the actual oxidation state is typically different (see the discussion with a Hirshfeld-I^{58–60} analysis in SI Sec. S2), but shows a good correlation with calculated charges being on average about half the formal oxidation state. The X-O distance varies rather systematically over these 6 oxides (see SI Fig. S1.3), typically with the smallest distance for XO_3 and the largest for X_2O . This indicates that XO_3 could be a proxy for systems with very short bond lengths, such as in high-pressure studies. The two datasets of unaries and oxides (jointly called “full dataset”; a total of 960 systems) complement each other in covering chemical and structural variety for each element.

In addition to the criteria above, all structures have been chosen to be cubic and such that forces on all atoms are zero by symmetry. Therefore, the only free parameter is the unit cell volume V or, equivalently, the lattice parameter. As a consequence, the **EOS** results can be compared with any code able to compute total energies, with no requirement on the capability of computing forces and stresses.

It is important to note that most structures are not stable in nature (in particular under our constraint of cubic spacegroup symmetry). Still, they can be used to assess that all codes reproduce the same **DFT** result, with the advantage of providing a consistent set across the whole periodic table.

Computation of EOS parameters and comparison between AE codes

The **EOS** has been traditionally used to determine the computational parameters and study convergence of **DFT** calculations. By fitting the **DFT** energy vs. cell volume to an **EOS**, it is possible to extract the theoretical predictions of the equilibrium volume V_0 , the bulk modulus B_0 , and its derivative with respect to the pressure, B_1 . The Birch–Murnaghan **EOS**⁶¹

$$E(V) = E_0 + \frac{9V_0B_0}{16} \left\{ \left[\left(\frac{V_0}{V} \right)^{\frac{2}{3}} - 1 \right]^3 B_1 + \left[\left(\frac{V_0}{V} \right)^{\frac{2}{3}} - 1 \right]^2 \left[6 - 4 \left(\frac{V_0}{V} \right)^{\frac{2}{3}} \right] \right\} \quad (1)$$

was used in the Δ -project^{29,30} and we follow the same approach by performing a fit of $E(V)$ of Eq. (1) using calculations of the total energy corresponding to 7 equidistant constant volumes between 94% and 106% of a reference central volume \tilde{V}_0 (for each Structure). We emphasize that the results are quite sensitive to the precise choice of volume range, reference central volume, and even of fitting algorithm, as we discuss in SI Sec. S3. In this work, the reference central volumes \tilde{V}_0 for each of the 960 crystals have been chosen after an iterative process of performing more and more accurate simulations with the two **AE** codes considered here, until the difference between the reference central volume and the equilibrium volume of the **EOS** fit was smaller than the 2% volume spacing between total-energy calculations. These central reference volumes are tabulated in SI Sec. S1 and the corresponding crystal-structure files are available in the data entry associated to this Recommendation⁶². These volumes have no physical significance, but for precise comparison between computational approaches, each of them should use the same reference volumes.

The results obtained with the **AE** codes FLEUR and WIEN2k constitute our reference data. Figure 1 shows the distributions of the percentage difference between FLEUR and WIEN2k for V_0 , B_0 and B_1 with respect to their average, for instance the V_0 difference (in %) is given by:

$$100 \cdot \frac{V_0^{\text{WIEN2k}} - V_0^{\text{FLEUR}}}{(V_0^{\text{WIEN2k}} + V_0^{\text{FLEUR}})/2}. \quad (2)$$

Although the histograms do not carry material-specific information, they clearly highlight the agreement between the two **AE** codes. The relative difference on the equilibrium volume is below 0.3% for all the materials except for 5 oxides (see SI Sec. S4 and the raw data in Ref. 62 for the full dataset). The discrepancies for B_0 and B_1 are larger; this is not surprising, because they originate from higher derivatives of the **EOS** curves (see also discussion in SI Sec. S3). We emphasize that these values, obtained after careful convergence of all numerical parameters related to the basis-set choices in the two codes, are of extremely high precision, with a spread that can even be an order of magnitude smaller than the typical discrepancies that we observe between pseudopotential codes (see discussion later).

The complete list of numerical parameters used for the **AE** calculations is presented in SI Sec. S5. We highlight here that the exact choice of the electronic-state smearing and of the k-point integration mesh, as well as the specific quantity considered as the energy $E(V)$ (internal energy, or free energy including the entropic smearing contribution as we do here), are of crucial importance for a reliable comparison among codes and must be fully consistent; therefore, we discuss those in detail in section “Using the All-Electron Reference Dataset” and in Box 3.

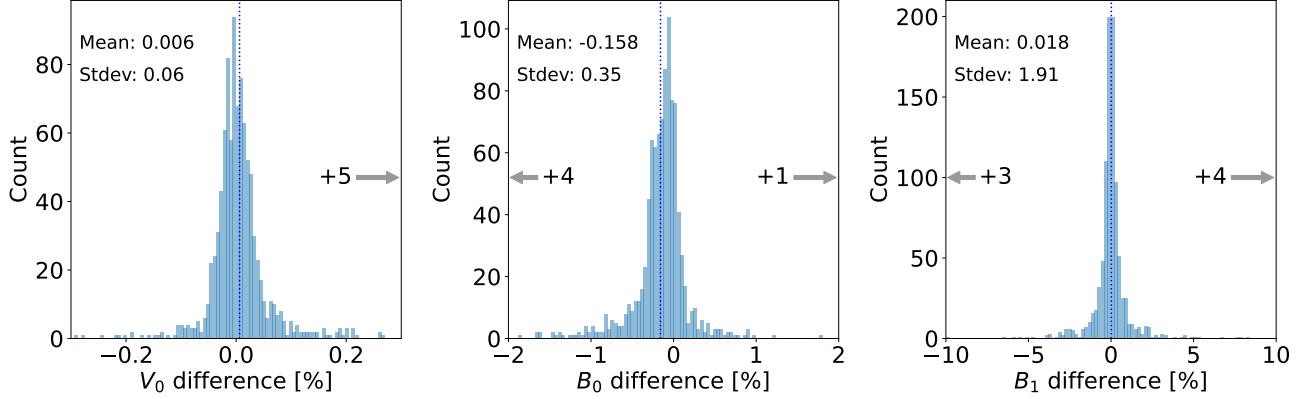


Figure 1. Histograms of the percentage difference between the results of the two all-electron codes (FLEUR and WIEN2k) with respect to their average for the three parameters of the EOS: V_0 , B_0 and B_1 , for the full dataset of unaries and oxides. Positive values indicate larger values for WIEN2k with respect to FLEUR. Mean and standard deviation (stdev) of the distributions are reported on the top left of each panel. The number near the arrows indicate the number of outliers outside of the x -axis range. The relative difference on V_0 is below 0.1% for 93% of all structures in our dataset; the relative difference on B_0 is below 1% for 97% of the structures; and the relative difference on B_1 is below 2% for 92% of the structures. The 5 outliers for V_0 are NeO_3 (0.302%), RbO_3 (0.343%), Cs_2O_5 (0.323%), Fr_2O_5 (0.645%) and Ra_2O_5 (0.333%), corresponding to lattice parameters for FLEUR/WIEN2k of $4.320/4.324\text{\AA}$, $4.783/4.789\text{\AA}$, $6.247/6.254\text{\AA}$, $6.120/6.133\text{\AA}$ and $6.238/6.244\text{\AA}$, respectively.

Average AE dataset: the reference for further studies

In addition to the data for each of the two codes and in order to provide a single comparison reference, we also provide a “reference average all-electron dataset” obtained by averaging the values of V_0 , B_0 , and B_1 for each of the 960 systems in the full dataset. The corresponding values are in SI Sec. S4 and published according to the FAIR principles in Ref. 62. Considering the very good agreement between the two codes, this average dataset constitutes an excellent reference, and we use this average to compare with the pseudopotential codes in Section “Comparison with Pseudopotential-based Computational Approaches”. In addition, if error bars are desired, the spread between the results of the two AE codes can be used as an estimate of our dataset precision.

Metrics for EOS comparison

In Refs. 29, 30, the “ Δ ” metric was used to compare the EOS computed with two different DFT computational approaches a and b . There, $\Delta = \Delta(a, b)$ was defined as:

$$\Delta(a, b) = \sqrt{\frac{1}{V_M - V_m} \int_{V_m}^{V_M} [E_a(V) - E_b(V)]^2 dV}, \quad (3)$$

where $E_a(V)$ and $E_b(V)$ are the Birch–Murnaghan fits of the data points obtained from approaches a and b respectively, the two EOS have been lined up with respect to their minimum energy, and as discussed earlier the integral spans a $\pm 6\%$ volume range centered at a central volume \tilde{V}_0 (with \tilde{V}_0 values tabulated in SI Sec. S1), i.e., $V_m = 0.94\tilde{V}_0$ and $V_M = 1.06\tilde{V}_0$.

The use of a single metric to compare two EOS curves simplifies the data analysis, since it can be used instead of the difference of the Birch–Murnaghan parameters V_0 , B_0 , and B_1 , as we did in Fig. 1. However the value of $\Delta(a, b)$, that has the units of energy, has the shortcoming of being too sensitive to the value of the bulk modulus of the material: visually similar discrepancies between two curves result in larger Δ values for materials with larger B_0 . This was already recognized in Ref. 63, where a modified metric Δ_1 was suggested, renormalized to a reference value of V_0 and B_0 . In addition, the $E_a(V)$ and $E_b(V)$ quantities in Eq. (3) are typically renormalized by the number of atoms in the unit cell, to provide a “ Δ/atom ” metric, independent of the choice of the simulation cell size. Since we expand our analysis to two-component oxides, generalizations might be required (e.g., by normalizing instead per formula unit).

We propose and recommend here two new metrics that we label ε and ν , and we discuss their pros and cons. We first define

the following shorthand notation for the integral average of a quantity $f(V)$ over the volume range $[V_m, V_M]$:

$$\langle f \rangle = \frac{1}{V_M - V_m} \int_{V_m}^{V_M} f(V) dV. \quad (4)$$

Using this notation, we can simply write $\Delta(a, b) = \sqrt{\langle [E_a(V) - E_b(V)]^2 \rangle}$. The first metric $\varepsilon(a, b)$ that we define is a renormalized dimensionless version of Δ :

$$\varepsilon(a, b) = \sqrt{\frac{\langle [E_a(V) - E_b(V)]^2 \rangle}{\sqrt{\langle [E_a(V) - \langle E_a \rangle]^2 \rangle} \sqrt{\langle [E_b(V) - \langle E_b \rangle]^2 \rangle}}}. \quad (5)$$

This metric, similarly to the Δ_1 of Ref. 63 or the subsequently defined Δ_{rel} available in the DeltaCodesDFT package³⁰, is insensitive to the magnitude of the bulk modulus (see SI Sec. S6). In addition, it is independent of the use of a “per-formula-unit” or “per-atom” definition of the EOS (see SI Sec. S6). Therefore, $\varepsilon(a, b)$ provides a uniform metric across the variety of structural and chemical environments under investigation, given the requirement that it must be calculated with the same relative volume range for every material. As the list of central reference volumes has been fixed (see SI Sec. S1), and as we use the same $\pm 6\%$ volume range as in Ref. 29, 30, the 960 intervals $[V_m, V_M]$ are unambiguously defined. We highlight, in passing, that the discrete form of Eq. (5), i.e.:

$$\varepsilon(a, b) = \sqrt{\frac{\sum_i [E_a(V_i) - E_b(V_i)]^2}{\sqrt{\sum_i [E_a(V_i) - \langle E_a \rangle]^2} \sqrt{\sum_i [E_b(V_i) - \langle E_b \rangle]^2}}} \quad (6)$$

where the index i runs over the explicit calculations of $E(V)$ from DFT, provides a reasonably good approximation to the value of Eq. (5) as long as the minima of the $E_a(V_i)$ and $E_b(V_i)$ data points are aligned on the energy scale, with the advantage that it can be used to directly compare raw DFT total energy data without requiring an EOS fitting. Nevertheless, we stress that in the rest of this work we use the expression of Eq. (5) and not its discrete version of Eq. (6). Eq. (6) is grounded in the definition of the coefficient of determination (or R^2) in statistics⁶⁴ as a fraction of variance unexplained. We can interpret the value of $1 - \varepsilon^2$ as the coefficient of determination (i.e. $1 - \varepsilon^2 \approx R^2$) in a situation when one EOS $E_a(V)$ is treated as a fit for the other EOS $E_b(V)$. More precisely, since we want to define a symmetric metric $\varepsilon(a, b) = \varepsilon(b, a)$, our ε^2 is the $1 - R^2$ value that one would obtain treating $E_a(V)$ as a fit for $E_b(V)$ and vice versa with the geometric mean of both data variances. We note that the interpretation $R^2 \approx 1 - \varepsilon^2$ holds in very good approximation when the value of ε is much smaller than one (i.e., for very similar $E(V)$ curves).

We discuss the sensitivity of ε to perturbations of the Birch–Murnaghan parameters in SI Sec. S7. The main outcome is that ε is mostly sensitive to the variations of V_0 , and much less of B_0 and B_1 . For some applications, however, some of the parameters are more relevant than others (e.g., if one is mostly interested in accurate bulk moduli). For these cases, we recommend (see also Box 4) to define metrics of discrepancy that depend directly on physically measurable quantities. Since an EOS is very well described by the three parameters V_0 , B_0 and B_1 , we suggest a second metric v that directly captures the relative difference of these three parameters between two computational approaches a and b , using appropriate weights w_{V_0} , w_{B_0} , w_{B_1} :

$$v_{w_{V_0}, w_{B_0}, w_{B_1}}(a, b) = 100 \sqrt{\sum_{Y=V_0, B_0, B_1} \left[w_Y \cdot \frac{Y_a - Y_b}{(Y_a + Y_b)/2} \right]^2}, \quad (7)$$

where, for instance, $(V_0)_a$ indicates the value of V_0 obtained by fitting the data of approach a , and so on. The (arbitrary) prefactor 100 is chosen to obtain values with similar order of magnitude to those of ε . Furthermore, it also helps in interpreting the value of v as an estimate of percentage errors (rather than relative errors) on the fit parameters. We highlight that v depends on the weights (w_Y), that in turn could be chosen to satisfy particular applications. Here, since we aim to be application-agnostic, we choose weights based only on the sensitivity of our fitting procedure to random numerical noise applied to the energy values of the EOS data points. The detailed procedure to determine the weights is described in SI Sec. S3; we just report here the final choice $w_{V_0} = 1$, $w_{B_0} = 1/20$ and $w_{B_1} = 1/400$. Intuitively, the reduced weights are consistent with the expected increase of numerical uncertainty propagated in the fit when estimating higher-order derivatives of the EOS. Similarly to ε , also these weights depend on the volume range of the datapoints used for the EOS fit, as well as the specific choice of the fitting algorithm (see details in SI Sec. S3). In the rest of the manuscript, we refer to v assuming this specific choice of weights, i.e., $v \equiv v_{1, 1/20, 1/400}$. In SI Sec. S7 we also discuss an intuitive interpretation of the v metric: it is the percentage error on the equilibrium volume between the two approaches a and b , when B_0 and B_1 are the same in the two approaches; otherwise, when

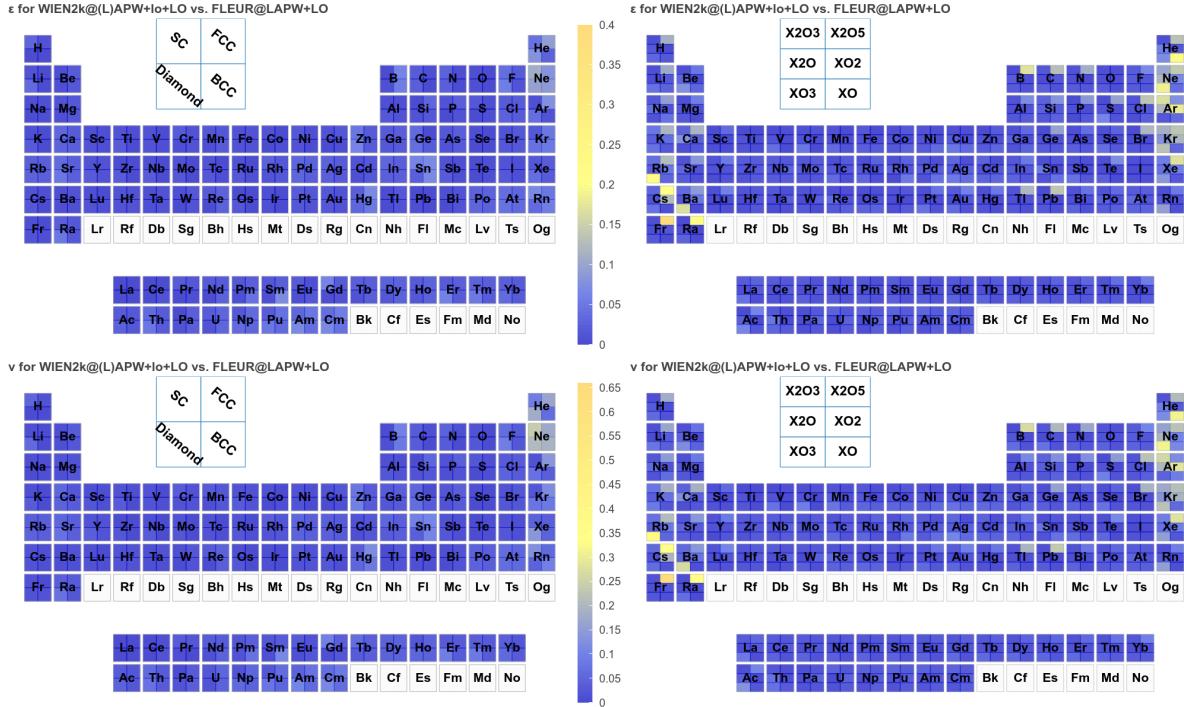


Figure 2. Discrepancy between the AE results (obtained with WIEN2k and FLEUR) in our reference dataset, measured either with the ε metric (top panels) or the ν metric (bottom panels), for all 96 elements considered. Each square for a given element is subdivided in 4 (6) in the left (right) panel, each referring to one of the unary (oxide) structures, as indicated in the legend presented in each panel. The color scale is the same for each pair (unaries and oxides) of periodic tables for the same metric (ε or ν). All structures are within our threshold for good agreement except Cs_2O_5 ($\varepsilon = 0.20$, $\nu = 0.33$), Fr_2O_5 ($\varepsilon = 0.40$, $\nu = 0.66$), Ra_2O_5 ($\varepsilon = 0.21$, $\nu = 0.33$) and RbO_3 ($\varepsilon = 0.21$, $\nu = 0.37$).

B_0 and B_1 differ, it corresponds to an equivalent percentage error on V_0 that would result in quantitatively similar changes to the EOS curve, within the $\pm 6\%$ volume range considered here.

The metrics ε and ν allow to compare a pair of codes for each material in the dataset. Fig. 2 reports the results for the pair (FLEUR, WIEN2k) across the entire set of structures under investigation in the form of periodic tables, enabling a quick identification of the most problematic elements in each set. For instance, as one might expect the agreement is generally worse for noble gases: weakly bonded systems with a very small bulk modulus and thus more susceptible to numerical errors due to the choice of the basis set and of other computational parameters.

We emphasize that with our choice of the volume range and weights for ν , the two metrics provide very consistent information, highlighting the importance of properly defining metrics based on physically measurable quantities, and on careful analysis of the error propagation of the fitting procedure, as we recommend in Box 4 (see also SI Sec. S7, where we discuss quantitatively the effect of perturbations on the EOS parameters to the values of ε , ν and Δ). Indeed, although ε and ν are constructed according to quite different principles, they turn out to contain nearly identical information (in SI Sec. S8 we show that they are to a good extent linearly correlated for $\nu \lesssim 1$, with $\nu \approx 1.65\varepsilon$). This has the consequence that periodic tables for ε or ν will be almost identical if the range of the color scale is taken according to this linear correlation (as it is the case, e.g., in Fig. 2 and is discussed in more detail in SI Sec. S9).

Finally, we identify (and report in Box 3, see also discussion in SI Sec. S7) indicative thresholds on ε and ν to represent an excellent agreement between two EOS curves if $\varepsilon \lesssim 0.06$ or $\nu \lesssim 0.1$, or a good agreement (noticeable, but still relatively small) if $\varepsilon \lesssim 0.2$ or $\nu \lesssim 0.33$. As discussed earlier, we can interpret the two thresholds $\varepsilon = 0.06$ ($\varepsilon = 0.2$) approximately as a determination coefficient $R^2 \approx 1 - 0.06^2 = 0.9964$ ($R^2 \approx 1 - 0.2^2 = 0.96$) if one EOS is treated as a fit for the other. The data from the two AE codes shows an overall excellent agreement: only four systems out of 960 have one or both metrics outside of the “good-agreement” range (Cs_2O_5 , Fr_2O_5 , Ra_2O_5 and RbO_3) when comparing the two AE codes of our reference dataset, and 883 out of 960 systems have an excellent agreement for both ε and ν according to the thresholds discussed above.

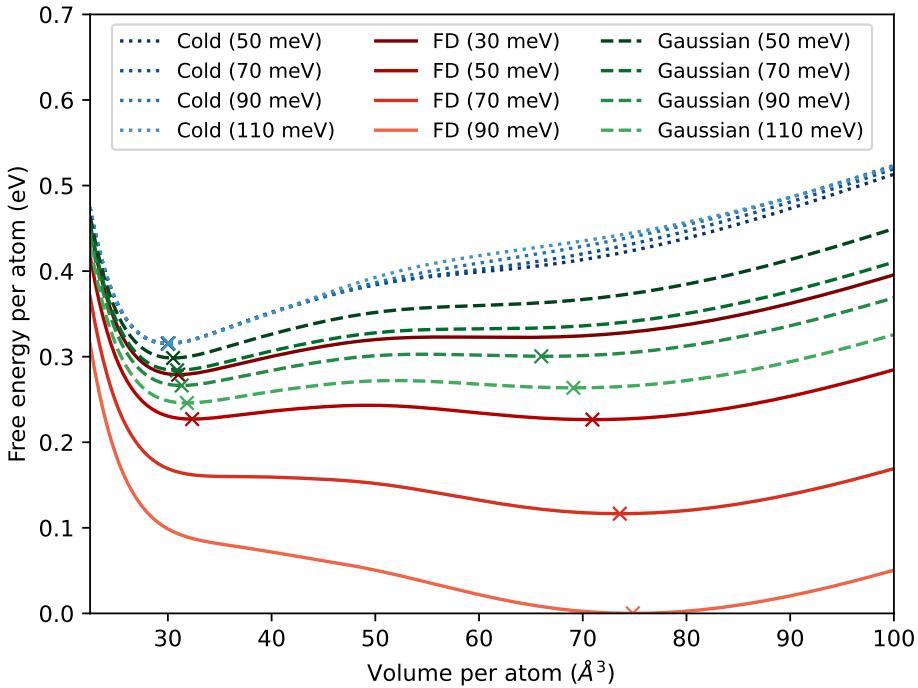


Figure 3. Effect of different choices of smearing on the EOS of erbium in the diamond structure computed with the QUANTUM ESPRESSO code. Er in the diamond structure is one of the systems in which the effect of smearing is most pronounced. In the legend, “Cold”, “FD” and “Gaussian” indicate, respectively, cold smearing⁶⁵, Fermi–Dirac smearing and Gaussian smearing. Two alternative minimum-energy solutions can exist for the FD and Gaussian smearings at very different volumes (as indicated by the crosses, when local minima exist in the curves); which one is selected depends on the choice of smearing and broadening. Note also the much reduced sensitivity of the cold smearing to the broadening temperature, and how FD and Gaussian smearings are essentially equivalent after a renormalisation of the FD broadening by a factor ≈ 2.565 , as discussed in Ref. 66.

Using the AE reference dataset

In this section we detail several aspects that must be carefully considered when using the reference dataset presented in the previous section. We then show how the dataset has been used to evaluate the precision of several computational approaches based on pseudopotentials and to improve a number of pseudopotential libraries.

Recommendations on how to use the dataset

When comparing our reference dataset with results from other codes, either for verification purposes or as a reference to improve basis sets and pseudopotentials, it is essential to use the same approximations (such as the exchange–correlation functional or the treatment of spin) and numerical choices (smearing and k-point integration mesh), as these parameters significantly affect the EOS results. Therefore, we discuss here (and summarize in Box 3) specific recommendations on which parameters should not be changed when generating new data to compare with.

All calculations are performed in periodic boundary conditions using the PBE⁵⁷ exchange–correlation functional, without including spin-polarization effects (non-magnetic calculations) and within a scalar-relativistic approximation (no spin-orbit coupling) for the orbitals treated as valence states. The reciprocal-space integration is performed with a Monkhorst–Pack uniform k-point grid including the Γ point, chosen as the smallest integration mesh guaranteeing a linear spacing of at most 0.06 Å^{-1} in each of the three reciprocal-space directions for the smallest volume, and the same set of k-points (in scaled units) for all other volumes. A Fermi–Dirac smearing of electronic states with a broadening of 0.0045 Ry ($\approx 61.2 \text{ meV}$) is used in all cases, requiring the high-density k-points sampling mentioned above. In addition, the quantity $E(V)$ that is fitted with the Birch–Murnaghan expression of Eq. (1) is not the internal energy, but the free energy that includes the entropic contribution $-TS$ introduced by the smearing (where T is the effective temperature given by the smearing broadening).

We stress that, in general, two codes using a different smearing distribution are expected to return comparable results only in the limit of an infinite number of k-points and an infinitesimal smearing. However, for the purpose of verification, we do not need to reach this computationally expensive limit, provided that the same parameters among codes are chosen. As a consequence, our results should not be considered a prediction of the zero-temperature (i.e., no smearing) limit. We still highlight, however, that our choice of the k-point density results in a very dense and almost converged integration mesh (for fixed broadening): e.g., all values of V_0 computed with WIEN2k change by less than 0.07% when comparing with a denser k-point integration mesh with linear spacing of 0.045 \AA^{-1} , except in two cases (RbO₃: 3.7% change, and HeO, 0.16% change). More details are reported in SI Sec. S10.

To emphasize the sensitivity of the EOS to the choice of smearing, we show in Fig. 3 one of the most pathological cases of our dataset, erbium in the diamond crystal structure. In this case, the EOS does not have a simple shape but displays instead, for the case of Fermi–Dirac and Gaussian smearing, two minima at very different volumes. Which one is favored in energy depends on the type of smearing and the value of the broadening. This behavior can be explained by the presence of a set of narrow f bands close to the Fermi level, shown in SI Sec. S11, whose filling strongly depends on the smearing. If we are after an improved erbium pseudopotential, trying to optimize it with a different smearing (and thus possibly for a different minimum) will result in an incorrect pseudopotential. A similar reasoning holds for the choice of using the free energy instead of the internal energy for the EOS (see SI Sec. S12). We also highlight that we adopted a scalar-relativistic treatment of valence electrons for our dataset. In most pseudopotential codes, this can be applied only for the valence electrons that are treated explicitly, while the treatment for the core electrons is implicitly included in the pseudopotential used. Even for AE codes, electrons are typically partitioned into a core (treated fully relativistically) and a valence set (treated scalar relativistically). We highlight that the two AE codes used in this work do not adopt the same core/valence assignment for all crystals (see SI Sec. S5), yet they agree very well, illustrating that the core/valence assignment might not lead to ambiguities in the calculated results, provided all other numerical parameters are chosen consistently.

Finally, we note that many additional code-specific parameters exist, such as the type and size of the basis set or the pseudopotential family. These choices are implemented in our automated common workflows³⁶ and can be selected using a new protocol defined for this work and named verification-PBE-v1. Specific details for each code are reported in SI Sec. S5 for the AE codes, and in SI Sec. S13 for all the other codes.

Before showing an example of the comparison of our reference AE dataset with nine computational approaches based on pseudopotentials, we discuss an additional recommendation. Our dataset was generated to provide reference EOS for each of the 960 structures. One might be tempted to reuse our dataset for different purposes. For instance, since the values of the minimum energy of the EOS curves are also available from the fits, one could imagine using them to compare total energies of various oxides of the same element X, estimating their relative stability and the corresponding formation energies. However, while this approach often results in sensible values, some notable cases lead to significantly off results, even by 1 eV/atom (see results in SI Sec. S18). The reason is that we designed our workflows and protocols for the EOS, in order to guarantee that simulation parameters are chosen consistently for all volumes of a given material, but this is not necessarily true among different materials. As an example, since oxides of the same element might have very different interatomic spacings, the choices of atomic radii (and the corresponding core/valence separation) for the AE codes might be different in different systems, which precludes direct comparison between total energies. From a more general point of view, one needs to be aware of the context in which data was produced, and consider implications and limitations when using them for different applications, as we discuss in the Outlook section.

Comparison with pseudopotential-based computational approaches

Using the recommendations of the previous section and of Box 3, we now compare our reference dataset with the results obtained with nine computational approaches based on pseudopotentials. As discussed earlier, each approach is not only defined by the choice of the code, but also by the pseudopotentials used (and, where applicable, by the type of basis set). Therefore, we summarize here briefly the meaning of the labels used for every computational approach. The two AE codes, FLEUR and WIEN2k, are labeled with their code name, followed by an indication of the basis set they use: FLEUR@LAPW+LO and WIEN2k@ (L) APW+lo+LO, respectively (see SI Sec. S5 for more details). All other labels also include, at the end, the name of the pseudopotential library that was used. In particular: ABINIT@PW | PseudoDojo-v0.5 indicates the ABINIT code, adopting a plane-wave (PW) basis set, using norm-conserving pseudopotentials from the PseudoDojo standard library version 0.5^{68,69}; BigDFT@DW | HGH-K (Valence) indicates a (partial) set of structures with valence-only Hartwigsen–Goedecker–Hutter pseudopotentials⁷⁰ calculated with the BigDFT code, adopting a basis set of Daubechies wavelets (DW), CASTEP@PW | C19MK2 indicates the CASTEP code using on-the-fly generated core-corrected ultrasoft pseudopotentials from the C19 library with updated settings for the f block elements, CP2K/Quickstep@TZV2P | GTH indicates the CP2K Quickstep code using Goedecker–Teter–Hutter pseudopotentials^{71,72} and a molecularly optimized TZV2P-type basis set⁷³, GPAW@PW | PAW-v0.9.20000 indicates the GPAW^{48,49} code used in its plane-wave mode using GPAW’s PAW

Expert Recommendation Box 3, Summary of details to properly compare with the reference dataset presented in this work

- Use the PBE exchange–correlation functional, do not include spin-polarization effects, and consider a scalar-relativistic treatment of (valence) electrons.
- Use Fermi–Dirac smearing with a value of 0.0045 Ry. While this choice does not lead to zero-smearing results (which would require extrapolation and extremely dense k-point integration meshes), using the same values ensures that results are comparable. In extreme cases, using a different smearing may affect significantly the equilibrium volume and the overall shape of the EOS.
- Compute the equations of state (EOS) using as the proper variational functional the free energy $E - TS$, where E is the internal energy and $-TS$ is the smearing-energy entropic contribution. Other choices, such as the internal energy E , or the extrapolated energy for zero smearing (e.g., the expression $E - TS/2$, valid for Fermi–Dirac or Gaussian smearings^{66,67}) can result in significant changes of the EOS, including large variations of the minimum-energy volume.
- Use the same protocol to fit the EOS curves: 7 equally spaced points in a volume range of $\pm 6\%$ around the specified central volume. With these choices, values $\varepsilon \lesssim 0.06$ or $v \lesssim 0.1$ can be considered to indicate an excellent agreement, and $\varepsilon \lesssim 0.2$ or $v \lesssim 0.33$ a good agreement (with a noticeable, but still relatively small discrepancy between them). A different volume range will affect these thresholds and require a different choice of weights for v to capture differences that are not purely statistical in nature. In addition, a different volume range will affect the k-point integration mesh (see next point).
- Use the exact same choice of the k-point integration mesh: regular grid including the Γ point and the smallest mesh guaranteeing a spacing between points of at most 0.06 \AA^{-1} along each of the three reciprocal-space directions for the smallest volume, and the same set of k-points (in scaled units) for all other volumes. This is typically converged for most systems and ensures that results can be compared even in the rare case of an unconverged k-point integration mesh.
- Do not transfer the choices performed for this reference dataset to a different context, since it might lead to incorrect conclusions. For instance, extracting formation energies from our reference dataset can provide inaccurate results, since the parameters used in our simulations are guaranteed to be consistent only for different volumes of the same material, but not necessarily among different materials.

pseudopotentials included in the setup release 0.9.20000⁷⁴, Quantum ESPRESSO@PW|SSSP-prec-v1.3 indicates the QUANTUM ESPRESSO code using the Standard Solid-State Pseudopotentials (SSSP) library (PBE precision version 1.3)^{75,76}, SIESTA@AtOrOptDiamond|PseudoDojo-v0.4 indicates the SIESTA code using norm-conserving pseudopotentials from the PseudoDojo standard library version 0.4 in psml format^{68,69,77} and localized basis sets in which the orbitals for each element are taken from a partial optimization, considering just the unary Diamond structure for that element (therefore no optimization for the chemical environment of each material has been performed), SIRIUS/CP2K@PW|SSSP-prec-v1.2 indicates the SIRIUS library code (run via its interface to CP2K) using the SSSP pseudopotential library (PBE precision version 1.2)⁷⁵ and VASP@PW|GW-PAW54* indicates the VASP code (v6.3) using the PAW GW PBE pseudopotentials released in the dataset potpaw_PBE.54, except for the lanthanides (see Section “Pseudopotentials Improvement”). The exact versions of the codes and libraries, together with the other code-specific choices implemented in the verification-PBE-v1 protocol, are detailed in SI Sec. S13. The choices of computational approaches (for each code) listed above have been selected by the workflow developers of each code, trying to identify converged parameters and limiting to choices commonly available to users (or in some cases improving upon them, as we discuss later in the section “Pseudopotentials improvement”).

The results are presented in Fig. 4 in the form of box-and-whisker plots for the percentage error of V_0 , B_0 and B_1 , with respect to the reference average AE dataset. Applying one of our general recommendations of Box 1, we partition our results in three groups (considering separately rare earths and/or heavy elements), in order to highlight the non-uniform capabilities of the various computational approaches. Indeed, the narrow bands originating from the localized f electrons are very challenging to be described accurately with plain DFT⁷⁸. Therefore, often pseudopotentials for these elements are not available and thus several approaches cannot produce data for rare earths. Even when available, those pseudopotentials might be less tested and thus deliver a lower precision. By separating the results, we also enable a fairer comparison of approaches for the common set of elements (from H to Bi excluding the lanthanide elements from La to Lu).

Our results show that different numerical approaches have different precision; in general, the spread of the parameters of

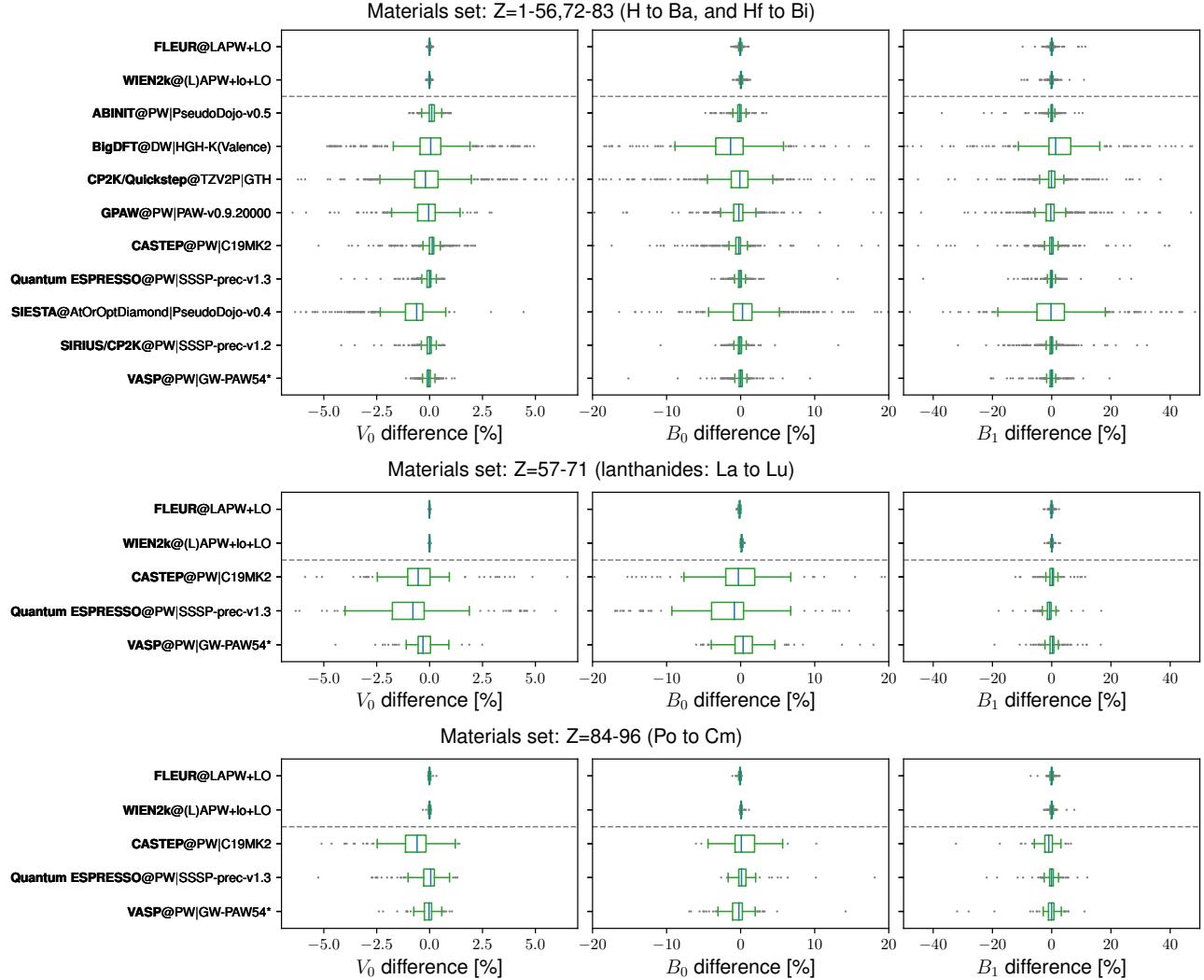


Figure 4. Box-and-whisker plots comparing the V_0 , B_0 and B_1 discrepancy of each computational approach involved in this work with respect to the average all-electron reference dataset. The two AE codes are also reported at the top of the plot above the dashed line, for comparison. In the box-and-whisker plots, the central blue line represents the median and the box extends between the first quartile Q_1 and the third quartile Q_3 . The “whiskers” extend between $Q_1 - 1.5 \cdot \text{IQR}$ and $Q_3 + 1.5 \cdot \text{IQR}$ (where IQR is the inter-quartile range $Q_3 - Q_1$). Outliers are represented as grey points. Note that some of the outliers are outside of the visible axis range in order to facilitate comparison between codes on the same axis range. Each row corresponds to a different subset of materials, where only the computational approaches that could compute those materials are included (since not all approaches include pseudopotentials for rare-earths). Specifically, the top row includes all materials from H to Bi excluding the lanthanide elements from La to Lu (68 elements in total). For this set, 295 crystals are missing for BigDFT@DW | HGH-K (Valence), Na (FCC) is missing for CP2K/Quickstep@TZV2P | GTH, Hg (FCC) and RbO₃ are missing for SIESTA@AtOrOptDiamond | PseudoDojo-v0.4, and all 10 crystals containing Tc are missing for GPAW@PW | PAW-v0.9.20000. The central row reports the results for lanthanides only (from La to Lu), and the bottom row for all materials from Po to Cm (i.e., heavy elements, including actinides up to Cm).

pseudopotential approaches are significantly larger than those between our two **AE** codes. In addition, the results indicate that additional work is required to obtain a high precision with approaches employing localized basis sets (that, on the other hand, are typically faster and scale better with system size) with respect to those using a plane-wave basis set (in the present case, the approaches listed above using the ABINIT, CASTEP, GPAW, QUANTUM ESPRESSO, VASP and SIRIUS/CP2K codes). Indeed, while a plane-wave basis set can be tuned with a single numerical parameter (the energy cutoff), systematically improving localized basis sets (and the associated pseudopotentials) requires dedicated efforts, that we recommend in Box 4. Verification projects such as this will facilitate these efforts by providing appropriate benchmarks. Another example of verification is presented in SI Sec. S14, where we discuss the agreement of different codes adopting the same computational approach, in particular with the same plane-wave basis set and the same pseudopotential library. In this case, the results show an agreement that is similar in precision to the one between the two **AE** codes.

Periodic tables (similar to Fig. 2) for each code are provided in SI Sec. S9, allowing for a closer inspection of the results resolved per chemical element and crystal-structure type. These tables also show that using a larger crystal-structure set (960 systems here) with respect to the set of 71 of Ref. 29 helps in highlighting possible shortcomings of pseudopotentials, as we discuss in more detail in SI Sec. S15. The results of each code are also available in Ref. 62, and can be visually displayed and compared directly online on the Materials Cloud¹⁹ at <https://acwf-verification.materialscloud.org>.

Pseudopotentials improvement

Curated datasets such as the one presented in this Recommendation can drive efforts to improve pseudopotentials, ultimately delivering more precise computational approaches. To illustrate this, we briefly summarize examples of pseudopotential enhancements that we performed to improve the comparison with our **AE** results (more technical details are discussed in SI Sec. S16) and used in the generation of the data of Fig. 4.

The results for ABINIT@PW | PseudoDojo-v0.5 for elements around the 4f block (from Te to Ba, and from Tl to Rn) were not giving ideal agreement using available pseudopotentials from PseudoDojo (version 0.4). In almost all cases, we found that the accuracy of the pseudopotentials is significantly improved by including a projector for the unbound *f* state, at the expense of an increase of the computational cost when applying the non-local part of the Hamiltonian V_{nl} (this can, however, be mitigated by the use of Legendre polynomials). Without this projector, the local part of the pseudopotential cannot reproduce the all-electron scattering properties of the *f* angular momentum (see SI Sec. S16.1). This led to the creation of a new PseudoDojo table (version 0.5), used here.

For CASTEP@PW | C19MK2, starting from the on-the-fly pseudopotential generation settings for the built-in C19 library, pseudopotentials for the lanthanide and actinide elements were improved by systematically changing the core radii, adding additional projectors, and adding fractional occupations of states that are empty in the reference atomic configurations. While making these changes did result in improvements, we note that no iterative optimization has been carried out to fit to the **AE** results.

For Quantum ESPRESSO@PW | SSSP-prec-v1.3, the pseudopotentials of SSSP PBE Precision version 1.1.2⁷⁹ have been updated for elements Na, Cu, Cs, Cd, Ba, As, Te, I, Hg, Ne, Ar, Kr, Xe, Rn; these new pseudopotentials have been released in the new SSSP PBE Precision 1.2⁸⁰. The new pseudopotentials have been selected by re-verifying the precision of pseudopotentials from various external libraries against the **AE** reference dataset discussed in this Recommendation, and replacing those displaying significant discrepancies with pseudopotentials from other libraries that displayed a better agreement (lower ϵ and v). Moreover, in SSSP PBE Precision version 1.3⁷⁶ (used here) new pseudopotentials have been included for actinides (Th-Lr) from Ref. 81, as well as for Ac, At, Ra, and Fr from PSlibrary⁸².

For VASP@PW | GW-PAW54*, the latest available PAW potential set (version 5.4) was improved by reducing by about 20% the core radii for lanthanides (other than La, Ce, and Lu). Furthermore, placing two electrons in the 6s shell, half an electron in the 5d shell and the rest in the *f* shell led to the most balanced description. For Tm, Er, and Yb, three *f* projectors were required to accurately describe the *f* scattering properties. The optimization was continued until very accurate scattering properties were obtained and agreement with very small core potentials was excellent, in turn resulting in a significant improvement of the agreement with the **AE** reference dataset.

Outlook

This work constitutes a next step in a grand scheme of actions aiming at controlling the numerical aspects of electronic structure calculations, where the diversity of computational approaches and codes provides an opportunity for pairwise verification. Compared to earlier work^{29,30}, we define here more discriminative metrics (crystals where two approaches would agree according to Δ might agree less according to ϵ or v , see SI Fig. S8.1a,b) and consider many more crystals, leading to more stringent testing. While major conclusions based on previous work remain valid (see SI Sec. S15 and SI Sec. S17), the dataset presented here—together with the clear set of recommendations on how to reuse the data—provides a more refined and valuable reference for verification, uncertainty quantification and pseudopotential optimization.

Expert Recommendation Box 4, Summary of recommendations to extend the verification effort presented in this work

- Extend the current study to more computational approaches (codes, basis or pseudopotential sets, etc.), adopting the same reference crystal-structure set presented here.
- Extend the current study to more properties (forces, phonons, Kohn–Sham band structures, formation energies, . . .). Choose properties (and materials) that maximize the number of codes that can compute them (e.g., here for the EOS, only single-point DFT simulations are required; forces and stresses were not used).
- Investigate the generality of optimal protocols and develop new ones for each property being computed, generalizing how to select a consistent set of parameters for multiple runs. For instance, for the EOS it is important to use the same k-point integration mesh at all volumes, but for a formation energy one wants a mesh that has reached a threshold accuracy for each component. In addition, use the same core/valence assignment, core radii and any other approach-specific precaution needed to compare total energies of different crystals.
- Create additional curated sets needed to generate improved pseudopotentials. E.g., extend to fully relativistic simulations and consider other exchange–correlation functionals in addition to PBE, such as the local-density approximation (LDA) and PBEsol, but also a selection of hybrids and meta-generalized-gradient approximations (meta-GGAs) for instance. Provide also curated sets needed to generate core-hole pseudopotentials for the simulation of core-level spectroscopies.
- Beside targeting improved pseudopotentials, develop dedicated efforts to optimize localized basis sets when these cannot be systematically improved by just tuning one or a few numerical parameters.
- Develop new protocols aiming at “good-enough data”: i.e., not only targeting ultimate numerical precision (needed for verification), but also optimizing the computational cost for a target accuracy.
- Disseminate these protocols to the broad simulation community to optimize energy and CPU time and to expand the computational feasibility of DFT computational approaches in high-throughput studies or for expensive post-DFT methods (e.g., many-body perturbation theory).
- For new verification protocols, define metrics (such as ε and v discussed here for the EOS) that depend on physically measurable quantities. Using such metrics, that condense in a single quantity the precision of computational approaches on a property of interest, one can easily define precision thresholds, compare approaches quantitatively, and evaluate the uniformity of results in a dataset. If fitting procedures are needed, assess the robustness of the chosen algorithms and estimate the uncertainty on the fitted parameters, using the results to define error bars.

Additionally, by formulating recommendations on how to perform further validation studies, and by providing and sharing universal common workflow interfaces (based on the AiiDA workflow infrastructure) to reproduce our calculations and perform new ones, we facilitate the community in taking new steps towards a better control of the uncertainty quantification in electronic structure calculations. There are several directions in which those steps could be taken. First, we recommend the creation of similar datasets for other commonly used exchange–correlation functionals (such as LDA and PBEsol, but possibly also a selection of hybrid and meta-GGA functionals), as well as for fully relativistic simulations. For these studies, we recommend to use the same initial set of crystal structures discussed here, possibly only adapting the central point of the volume interval $[V_m, V_M]$ if the equilibrium volume V_0 for the functional does not lie anymore roughly in the middle of the interval. Indeed, the set is fairly complete and systematic, and using the same structures facilitates the comparison between different computational approaches and approximations. In addition, we recommend to test and verify codes also for magnetic materials.

Once such datasets are available, efforts to further improve pseudopotentials (and basis sets) should be initiated or continued, with the aim of making the results easily available to the broad simulation community. One useful outcome could be, for instance, the generation of new reliable fully relativistic pseudopotential datasets for rare-earths (especially of the norm-conserving type, often required by many codes computing advanced materials properties). Another relevant example, involving also the generation of additional bespoke AE reference datasets, is the generation of pseudopotentials with a hole in the core, needed to predict the outcome of X-ray photoelectron spectroscopy (XPS) or X-ray absorption spectroscopy (XAS) experiments.

As a note, we highlight here that some of our structures are unrealistic. When generating a new pseudopotential for a given chemical element, one might want to accept a compromise and not reproduce precisely the EOS of all 10 unaries and oxides, in order to obtain a computationally cheaper pseudopotential (e.g., with less projectors, more electrons in the core, or requiring a

smaller energy cutoff), as long as the results are precise enough for the intended applications.

Other properties beyond the EOS are relevant to characterize materials and might benefit from tailored verification efforts; these include, e.g., formation energies, electronic band structures and phonon frequencies. As we already highlighted, the simulation protocols might be significantly different for each property. We therefore recommend that these protocols are well designed, documented and discussed, together with their limit of applicability. In particular, especially if limiting to a scalar-relativistic approach as we did here, we recommend to further investigate the relevance of the choice of which electrons are included in the core or in the valence, as this can be of higher relevance than for the EOS (e.g., for formation energies, see also SI Sec. S18). Moreover, new metrics should be designed to quantitatively compare results, ideally directly dependent on physically measurable quantities. Error propagation through any fitting procedure or data analysis should be carefully assessed, as we did in SI Sec. S3, to be able to define appropriate error bars.

Finally, we emphasize that while the goal here was ultimate precision in order to provide a reference dataset and obtain the best agreement possible between computational approaches, in real simulations one needs to optimize also the computational cost for a target accuracy, to obtain “good-enough data” for their scientific purpose. This is especially true for high-throughput runs or when the DFT simulations are the first step of more expensive post-DFT methods. We thus encourage to develop protocols to automatically define or select optimally converged parameters that at the same time minimize energy and CPU time, and then disseminate these to the whole community, so that they become easily accessible and usable by a broad range of users.

References

1. Alberi, K. *et al.* The 2019 materials by design roadmap. *J. Phys. D: Appl. Phys.* **52**, 013001, DOI: [10.1088/1361-6463/ad926](https://doi.org/10.1088/1361-6463/ad926) (2018).
2. Marzari, N., Ferretti, A. & Wolverton, C. Electronic-structure methods for materials design. *Nat. Mater.* **20**, 736–749, DOI: [10.1038/s41563-021-01013-3](https://doi.org/10.1038/s41563-021-01013-3) (2021).
3. Pizzi, G., Cepellotti, A., Sabatini, R., Marzari, N. & Kozinsky, B. AiiDA: automated interactive infrastructure and database for computational science. *Comput. Mater. Sci.* **111**, 218–230, DOI: [10.1016/j.commatsci.2015.09.013](https://doi.org/10.1016/j.commatsci.2015.09.013) (2016).
4. Huber, S. P. *et al.* AiiDA 1.0, a scalable computational infrastructure for automated reproducible workflows and data provenance. *Sci. Data* **7**, DOI: [10.1038/s41597-020-00638-4](https://doi.org/10.1038/s41597-020-00638-4) (2020).
5. Ong, S. P. *et al.* Python materials genomics (pymatgen): A robust, open-source python library for materials analysis. *Comput. Mater. Sci.* **68**, 314–319, DOI: <https://doi.org/10.1016/j.commatsci.2012.10.028> (2013).
6. Jain, A. *et al.* Fireworks: a dynamic workflow system designed for high-throughput applications. *Concurr. Comput. Pract. Exp.* **27**, 5037–5059, DOI: [10.1002/cpe.3505](https://doi.org/10.1002/cpe.3505) (2015).
7. Mathew, K. *et al.* Atomate: A high-level interface to generate, execute, and analyze computational materials science workflows. *Comput. Mater. Sci.* **139**, 140–152, DOI: [10.1016/j.commatsci.2017.07.030](https://doi.org/10.1016/j.commatsci.2017.07.030) (2017).
8. Bahn, S. R. & Jacobsen, K. W. An object-oriented scripting interface to a legacy electronic structure code. *Comput. Sci. Eng.* **4**, 56–66, DOI: [10.1109/5992.998641](https://doi.org/10.1109/5992.998641) (2002).
9. Hjorth Larsen, A. *et al.* The atomic simulation environment—a Python library for working with atoms. *J. Physics: Condens. Matter* **29**, 273002, DOI: [10.1088/1361-648X/AA680E](https://doi.org/10.1088/1361-648X/AA680E) (2017).
10. Curtarolo, S. *et al.* Aflow: An automatic framework for high-throughput materials discovery. *Comput. Mater. Sci.* **58**, 218–226, DOI: <https://doi.org/10.1016/j.commatsci.2012.02.005> (2012).
11. Janssen, J. *et al.* pyiron: An integrated development environment for computational materials science. *Comput. Mater. Sci.* **163**, 24 – 36, DOI: <https://doi.org/10.1016/j.commatsci.2018.07.043> (2019).
12. Armiento, R. *Database-Driven High-Throughput Calculations and Machine Learning Models for Materials Design*, 377–395 (Springer International Publishing, Cham, 2020).
13. Gonze, X. *et al.* The Abinit project: Impact, environment and recent developments. *Comput. Phys. Commun.* **248**, 107042, DOI: [10.1016/j.cpc.2019.107042](https://doi.org/10.1016/j.cpc.2019.107042) (2020).
14. Jain, A. *et al.* Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL Mater.* **1**, 011002, DOI: [10.1063/1.4812323](https://doi.org/10.1063/1.4812323) (2013).
15. Landis, D. D. *et al.* The Computational Materials Repository. *Comput. Sci. & Eng.* **14**, 51–57, DOI: [10.1109/MCSE.2012.16](https://doi.org/10.1109/MCSE.2012.16) (2012).
16. Kirklin, S. *et al.* The open quantum materials database (oqmd): assessing the accuracy of dft formation energies. *npj Comput. Mater.* **1**, 15010, DOI: [10.1038/npjcompumats.2015.10](https://doi.org/10.1038/npjcompumats.2015.10) (2015).
17. Merkys, A. *et al.* A posteriori metadata from automated provenance tracking: Integration of AiiDA and TCOD. *J. Cheminformatics* **9**, 56, DOI: [10.1186/s13321-017-0242-y](https://doi.org/10.1186/s13321-017-0242-y) (2017). [1706.08704v3](https://doi.org/10.1186/s13321-017-0242-y).
18. Stevanović, V., Lany, S., Zhang, X. & Zunger, A. Correcting density functional theory for accurate predictions of compound enthalpies of formation: Fitted elemental-phase reference energies. *Phys. Rev. B* **85**, 115104, DOI: [10.1103/PhysRevB.85.115104](https://doi.org/10.1103/PhysRevB.85.115104) (2012).
19. Talirz, L. *et al.* Materials cloud, a platform for open computational science. *Sci. Data* **7**, DOI: [10.1038/s41597-020-00637-5](https://doi.org/10.1038/s41597-020-00637-5) (2020).
20. Draxl, C. & Scheffler, M. Nomad: The fair concept for big data-driven materials science. *MRS Bull.* **43**, 676–682, DOI: [10.1557/mrs.2018.208](https://doi.org/10.1557/mrs.2018.208) (2018).
21. Curtarolo, S. *et al.* Aflowlib.org: A distributed materials properties repository from high-throughput ab initio calculations. *Comput. Mater. Sci.* **58**, 227–235, DOI: <https://doi.org/10.1016/j.commatsci.2012.02.002> (2012).
22. Wilkinson, M. D. *et al.* The FAIR guiding principles for scientific data management and stewardship. *Sci. Data* **3**, DOI: [10.1038/sdata.2016.18](https://doi.org/10.1038/sdata.2016.18) (2016).
23. Andersen, C. W. *et al.* OPTIMADE, an API for exchanging materials data. *Sci. Data* **8**, DOI: [10.1038/s41597-021-00974-z](https://doi.org/10.1038/s41597-021-00974-z) (2021).

24. IEEE Standard for System, Software, and Hardware Verification and Validation. *IEEE Std 1012-2016 (Revision IEEE Std 1012-2012/ Incorporates IEEE Std 1012-2016/Cor1-2017)* 1–260, DOI: [10.1109/IEEESTD.2017.8055462](https://doi.org/10.1109/IEEESTD.2017.8055462) (2017).
25. Wang, A. *et al.* A framework for quantifying uncertainty in dft energy corrections. *Sci. Reports* **11**, 15496, DOI: [10.1038/s41598-021-94550-5](https://doi.org/10.1038/s41598-021-94550-5) (2021).
26. Carbogno, C. *et al.* Numerical quality control for dft-based materials databases. *npj Comput. Mater.* **8**, 69, DOI: [10.1038/s41524-022-00744-4](https://doi.org/10.1038/s41524-022-00744-4) (2022).
27. Poncé, S. *et al.* Verification of first-principles codes: Comparison of total energies, phonon frequencies, electron–phonon coupling and zero-point motion correction to the gap between abinit and qe/yambo. *Comput. Mater. Sci.* **83**, 341–348, DOI: <https://doi.org/10.1016/j.commatsci.2013.11.031> (2014).
28. Pople, J. Nobel Lecture. NobelPrize.org, <https://www.nobelprize.org/prizes/chemistry/1998/pople/lecture/>.
29. Lejaeghere, K. *et al.* Reproducibility in density functional theory calculations of solids. *Science* **351**, aad3000, DOI: [10.1126/science.aad3000](https://doi.org/10.1126/science.aad3000) (2016).
30. 71-crystal benchmark results from ref. 29 plus data collected later on, <https://molmod.ugent.be/deltacodesdft>.
31. The FLEUR project. <https://www.flapw.de/>.
32. Wortmann, D. *et al.* Fleur. Zenodo, DOI: [10.5281/zenodo.7576163](https://doi.org/10.5281/zenodo.7576163) (2023).
33. Blaha, P. *et al.* Wien2k: An augmented plane wave plus local orbitals program for calculating crystal properties. <http://www.wien2k.at/>.
34. Blaha, P. *et al.* Wien2k: An apw+lo program for calculating the properties of solids. *J. Chem. Phys.* **152**, 074101, DOI: [10.1063/1.5143061](https://doi.org/10.1063/1.5143061) (2020).
35. Uhrin, M., Huber, S. P., Yu, J., Marzari, N. & Pizzi, G. Workflows in AiiDA: Engineering a high-throughput, event-based engine for robust and modular computational workflows. *Comput. Mater. Sci.* **187**, 110086, DOI: [10.1016/j.commatsci.2020.110086](https://doi.org/10.1016/j.commatsci.2020.110086) (2021).
36. Huber, S. P. *et al.* Common workflows for computing material properties using different quantum engines. *npj Comput. Mater.* **7**, 136, DOI: [10.1038/s41524-021-00594-6](https://doi.org/10.1038/s41524-021-00594-6) (2021).
37. Huber, S. P. *et al.* AiiDA common workflows (ACWF) package (2021). <https://github.com/aiidateam/aiida-common-workflows>.
38. Lejaeghere, K., Speybroeck, V. V., Oost, G. V. & Cottenier, S. Error Estimates for Solid-State Density-Functional Theory Predictions: An Overview by Means of the Ground-State Elemental Crystals. *Critical Rev. Solid State Mater. Sci.* **39**, 1–24, DOI: [10.1080/10408436.2013.772503](https://doi.org/10.1080/10408436.2013.772503) (2013).
39. Gross, G. & Pastori Parravicini, G. *Solid State Physics* (Academic Press, Amsterdam, 2013), second edn.
40. Martin, R. M. *Electronic Structure: Basic Theory and Practical Methods* (Cambridge University Press, Cambridge, 2013), second edn.
41. Cohen, M. L. & Louie, S. G. *Fundamentals of Condensed Matter Physics* (Cambridge University Press, Cambridge, 2016).
42. Gonze, X. *et al.* Recent developments in the ABINIT software package. *Comput. Phys. Commun.* **205**, 106–131, DOI: [10.1016/j.cpc.2016.04.003](https://doi.org/10.1016/j.cpc.2016.04.003) (2016).
43. Romero, A. H. *et al.* ABINIT: Overview and focus on selected capabilities. *The J. Chem. Phys.* **152**, 124102, DOI: [10.1063/1.5144261](https://doi.org/10.1063/1.5144261) (2020).
44. Ratcliff, L. E. *et al.* Flexibilities of wavelets as a computational basis set for large-scale electronic structure calculations. *The J. Chem. Phys.* **152**, 194110, DOI: [10.1063/5.0004792](https://doi.org/10.1063/5.0004792) (2020).
45. Clark, S. J. *et al.* First principles methods using CASTEP. *Zeitschrift für Kristallographie - Cryst. Mater.* **220**, DOI: [10.1524/zkri.220.5.567.65075](https://doi.org/10.1524/zkri.220.5.567.65075) (2005).
46. The CP2K simulation package. <https://www.cp2k.org>.
47. Kühne, T. D. *et al.* CP2k: An electronic structure and molecular dynamics software package - quickstep: Efficient and accurate electronic structure calculations. *The J. Chem. Phys.* **152**, 194103, DOI: [10.1063/5.0007045](https://doi.org/10.1063/5.0007045) (2020).
48. Mortensen, J. J., Hansen, L. B. & Jacobsen, K. W. Real-space grid implementation of the projector augmented wave method. *Phys. Rev. B* **71**, 035109, DOI: [10.1103/PhysRevB.71.035109](https://doi.org/10.1103/PhysRevB.71.035109) (2005).

49. Enkovaara, J. *et al.* Electronic structure calculations with GPAW: a real-space implementation of the projector augmented-wave method. *J. Physics: Condens. Matter* **22**, 253202 (2010).
50. Giannozzi, P. *et al.* QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials. *J. Physics: Condens. Matter* **21**, 395502, DOI: [10.1088/0953-8984/21/39/395502](https://doi.org/10.1088/0953-8984/21/39/395502) (2009).
51. Giannozzi, P. *et al.* Advanced capabilities for materials modelling with quantum ESPRESSO. *J. Physics: Condens. Matter* **29**, 465901, DOI: [10.1088/1361-648x/aa8f79](https://doi.org/10.1088/1361-648x/aa8f79) (2017).
52. Soler, J. M. *et al.* The SIESTA method for ab initio order-N materials simulation. *J. Physics: Condens. Matter* **14**, 2745–2779, DOI: [10.1088/0953-8984/14/11/302](https://doi.org/10.1088/0953-8984/14/11/302) (2002).
53. García, A. *et al.* Siesta: Recent developments and applications. *The J. Chem. Phys.* **152**, 204108, DOI: [10.1063/5.0005077](https://doi.org/10.1063/5.0005077) (2020).
54. The SIRIUS domain-specific library for electronic-structure calculations. <https://github.com/electronic-structure/SIRIUS>.
55. Kresse, G. & Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **54**, 11169–11186, DOI: [10.1103/physrevb.54.11169](https://doi.org/10.1103/physrevb.54.11169) (1996).
56. Kresse, G. & Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B* **59**, 1758–1775, DOI: [10.1103/physrevb.59.1758](https://doi.org/10.1103/physrevb.59.1758) (1999).
57. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865–3868, DOI: [10.1103/physrevlett.77.3865](https://doi.org/10.1103/physrevlett.77.3865) (1996).
58. Vanpoucke, D. E. P., Bultinck, P. & Van Driessche, I. Extending Hirshfeld-I to bulk and periodic materials. *J. Comput. Chem.* **34**, 405–417, DOI: [10.1002/jcc.23088](https://doi.org/10.1002/jcc.23088) (2013).
59. Vanpoucke, D. E. P., Van Driessche, I. & Bultinck, P. Reply to ‘Comment on “Extending Hirshfeld-I to bulk and periodic materials”’. *J. Comput. Chem.* **34**, 422–427, DOI: [10.1002/jcc.23193](https://doi.org/10.1002/jcc.23193) (2013).
60. Bultinck, P., Van Alsenoy, C., Ayers, P. W. & Carbó-Dorca, R. Critical analysis and extension of the hirshfeld atoms in molecules. *J. Chem. Phys.* **126**, 144111, DOI: [10.1063/1.2715563](https://doi.org/10.1063/1.2715563) (2007).
61. Birch, F. Finite elastic strain of cubic crystals. *Phys. Rev.* **71**, 809–824, DOI: [10.1103/PhysRev.71.809](https://doi.org/10.1103/PhysRev.71.809) (1947).
62. Bosoni, E. *et al.* How to verify the precision of density-functional-theory implementations via reproducible and universal workflows. *Mater. Cloud Arch.* DOI: <https://doi.org/10.24435/materialscloud:s4-3h> (2023). <https://doi.org/10.24435/materialscloud:s4-3h>.
63. Jollet, F., Torrent, M. & Holzwarth, N. Generation of projector augmented-wave atomic data: A 71 element validated table in the XML format. *Comput. Phys. Commun.* **185**, 1246–1254, DOI: [10.1016/j.cpc.2013.12.023](https://doi.org/10.1016/j.cpc.2013.12.023) (2014).
64. Glantz, S. A., Slinger, B. K. & Neilands, T. B. *Selecting the “Best” Regression Model*, in *Primer of Applied Regression and Analysis of Variance, third edition* (McGraw-Hill Education, New York, 2017).
65. Marzari, N., Vanderbilt, D., De Vita, A. & Payne, M. C. Thermal Contraction and Disordering of the Al(110) Surface. *Phys. Rev. Lett.* **82**, 3296–3299, DOI: [10.1103/PhysRevLett.82.3296](https://doi.org/10.1103/PhysRevLett.82.3296) (1999).
66. dos Santos, F. J. & Marzari, N. Fermi energy determination for advanced smearing techniques. *arXiv* **2212.07988**, DOI: [10.48550/ARXIV.2212.07988](https://doi.org/10.48550/ARXIV.2212.07988) (2022).
67. Gillan, M. J. Calculation of the vacancy formation energy in aluminium. *J. Phys.. Cond. Matter* **1**, 689, DOI: [10.1088/0953-8984/1/4/005](https://doi.org/10.1088/0953-8984/1/4/005) (1989).
68. van Setten, M. *et al.* The PseudoDojo: Training and grading a 85 element optimized norm-conserving pseudopotential table. *Comput. Phys. Commun.* **226**, 39–54, DOI: [10.1016/j.cpc.2018.01.012](https://doi.org/10.1016/j.cpc.2018.01.012) (2018).
69. The PseudoDojo website. <http://www.pseudo-dojo.org/>.
70. Hartwigsen, C., Goedecker, S. & Hutter, J. Relativistic separable dual-space Gaussian pseudopotentials from H to Rn. *Phys. Rev. B* **58**, 3641–3662, DOI: [10.1103/PhysRevB.58.3641](https://doi.org/10.1103/PhysRevB.58.3641) (1998).
71. Goedecker, S., Teter, M. & Hutter, J. Separable dual-space Gaussian pseudopotentials. *Phys. Rev. B* **54**, 1703 (1996).
72. Krack, M. Pseudopotentials for H to Kr optimized for gradient-corrected exchange-correlation functionals. *Theor. Chem. Accounts* **114**, 145–152 (2005).
73. VandeVondele, J. & Hutter, J. Gaussian basis sets for accurate calculations on molecular systems in gas and condensed phases. *The J. chemical physics* **127**, 114105 (2007).

74. GPAW atomic PAW setups. <https://wiki.fysik.dtu.dk/gpaw/setupssetups.html#atomic-paw-setups>.
75. Prandini, G., Marrazzo, A., Castelli, I. E., Mounet, N. & Marzari, N. Precision and efficiency in solid-state pseudopotential calculations. *npj Comput. Mater.* **4**, DOI: [10.1038/s41524-018-0127-2](https://doi.org/10.1038/s41524-018-0127-2) (2018).
76. Prandini, G. *et al.* A standard solid state pseudopotentials (sssp) library optimized for precision and efficiency. *Mater. Cloud Arch.* **2023.65** (2022). Version v11, <https://doi.org/10.24435/materialscloud:f3-ym>.
77. García, A., Verstraete, M. J., Pouillon, Y. n. & Junquera, J. The psml format and library for norm-conserving pseudopotential data curation and interoperability. *Comput. Phys. Commun.* **227**, 51 – 71, DOI: <https://doi.org/10.1016/j.cpc.2018.02.011> (2018).
78. Topsakal, M. & Wentzcovitch, R. Accurate projected augmented wave (paw) datasets for rare-earth elements (re=la–lu). *Comput. Mater. Sci.* **95**, 263–270, DOI: <https://doi.org/10.1016/j.commatsci.2014.07.030> (2014).
79. Prandini, G. *et al.* A standard solid state pseudopotentials (sssp) library optimized for precision and efficiency. *Mater. Cloud Arch.* **2021.76** (2021). Version v7, <https://doi.org/10.24435/materialscloud:rz-77>.
80. Prandini, G. *et al.* A standard solid state pseudopotentials (sssp) library optimized for precision and efficiency. *Mater. Cloud Arch.* **2022.159** (2022). Version v8, <https://doi.org/10.24435/materialscloud:3v-xt>.
81. Sachs, M. *et al.* Dft-guided crystal structure redetermination and lattice dynamics of the intermetallic actinoid compound uir. *Inorg. Chem.* **60**, 16686–16699 (2021).
82. Dal Corso, A. Pseudopotentials periodic table: From H to Pu. *Comput. Mater. Sci.* **95**, 337–350 (2014).
83. Thörnig, P. JURECA: Data Centric and Booster Modules implementing the Modular Supercomputing Architecture at Jülich Supercomputing Centre. *J. large-scale research facilities JLSRF* **7**, 182, DOI: [10.17815/jlsrf-7-182](https://doi.org/10.17815/jlsrf-7-182) (2021).
84. Igor Levin. NIST Inorganic Crystal Structure Database (ICSD), National Institute of Standards and Technology (2018).
85. Kokalj, A. Xcrysden—a new program for displaying crystalline structures and electron densities. *J. Mol. Graph. Model.* **17**, 176–179, DOI: [https://doi.org/10.1016/S1093-3263\(99\)00028-5](https://doi.org/10.1016/S1093-3263(99)00028-5) (1999).
86. Mulliken, R. S. Electronic structures of molecules xi. electroaffinity, molecular orbitals and dipole moments. *J. Chem. Phys.* **3**, 573–585, DOI: [10.1063/1.1749731](https://doi.org/10.1063/1.1749731) (1935).
87. Mulliken, R. S. Electronic population analysis on lcao–mo molecular wave functions. i. *J. Chem. Phys.* **23**, 1833–1840, DOI: [10.1063/1.1740588](https://doi.org/10.1063/1.1740588) (1955).
88. Hirshfeld, F. L. Bonded-atom fragments for describing molecular charge densities. *Theor. Chim. Acta* **44**, 129–138, DOI: [10.1007/BF00549096](https://doi.org/10.1007/BF00549096) (1977).
89. Bader, R. F. W. A quantum theory of molecular structure and its applications. *Chem. Rev.* **91**, 893–928, DOI: [10.1021/cr00005a013](https://doi.org/10.1021/cr00005a013) (1991).
90. Becke, A. D. A multicenter numerical integration scheme for polyatomic molecules. *J. Chem. Phys.* **88**, 2547–2553, DOI: [10.1063/1.454033](https://doi.org/10.1063/1.454033) (1988).
91. Vanpoucke, D.E.P. HIVE-tools. <https://github.com/DannyVanpoucke/HIVE4-tools>.
92. Lebedev, V. I. & Laikov, D. A quadrature formula for the sphere of the 131st algebraic order of accuracy. *Doklady Math.* **59**, 477–481 (1988).
93. Bultinck, P., Ayers, P. W., Fias, S., Tiels, K. & Van Alsenoy, C. Uniqueness and basis set dependence of iterative hirshfeld charges. *Chem. Phys. Lett.* **444**, 205–208, DOI: <https://doi.org/10.1016/j.cplett.2007.07.014> (2007).
94. Wolffis, J. J., Vanpoucke, D. E. P., Sharma, A., Lawler, K. V. & Forster, P. M. Predicting partial atomic charges in siliceous zeolites. *Microporous Mesoporous Mater.* **277**, 184–196, DOI: [10.1016/j.micromeso.2018.10.028](https://doi.org/10.1016/j.micromeso.2018.10.028) (2019).
95. Bueken, B. *et al.* A Flexible Photoactive Titanium Metal–Organic Framework Based on a $[\text{Ti}_3^{IV}(\mu_3\text{-O})(\text{O})_2(\text{COO})_6]$ Cluster. *Angew. Chem. Int. Ed.* **54**, 13912–13917, DOI: [10.1002/anie.201505512](https://doi.org/10.1002/anie.201505512) (2015).
96. Verstraelen, T., Sukhomlinov, S. V., Van Speybroeck, V., Waroquier, M. & Smirnov, K. S. Computation of charge distribution and electrostatic potential in silicates with the use of chemical potential equalization models. *J. Phys. Chem. C* **116**, 490–504, DOI: [10.1021/jp210129r](https://doi.org/10.1021/jp210129r) (2012).
97. 2 - Chemical Periodicity and the Periodic Table. In Greenwood, N. & Earnshaw, A. (eds.) *Chemistry of the Elements*, 27–28, DOI: <https://doi.org/10.1016/B978-0-7506-3365-9.50008-0> (Butterworth-Heinemann, Oxford, 1997), second edn.

98. Lejaeghere, K., Vanduyfhuys, L., Verstraelen, T., Van Speybroeck, V. & Cottenier, S. Is the error on first-principles volume predictions absolute or relative? *Comput. Mater. Sci.* **117**, 390–396, DOI: <https://doi.org/10.1016/j.commatsci.2016.01.039> (2016).
99. Andersen, O. K. Linear methods in band theory. *Phys. Rev. B* **12**, 3060–3083, DOI: [10.1103/PhysRevB.12.3060](https://doi.org/10.1103/PhysRevB.12.3060) (1975).
100. Wimmer, E., Krakauer, H., Weinert, M. & Freeman, A. J. Full-potential self-consistent linearized-augmented-plane-wave method for calculating the electronic structure of molecules and surfaces: O₂ molecule. *Phys. Rev. B* **24**, 864–875, DOI: [10.1103/PhysRevB.24.864](https://doi.org/10.1103/PhysRevB.24.864) (1981).
101. Singh, D. Ground-state properties of lanthanum: Treatment of extended-core states. *Phys. Rev. B* **43**, 6388–6392, DOI: [10.1103/PhysRevB.43.6388](https://doi.org/10.1103/PhysRevB.43.6388) (1991).
102. Michalicek, G., Betzinger, M., Friedrich, C. & Blügel, S. Elimination of the linearization error and improved basis-set convergence within the FLAPW method. *Comput. Phys. Commun.* **184**, 2670 – 2679, DOI: <http://dx.doi.org/10.1016/j.cpc.2013.07.002> (2013).
103. Friedrich, C., Schindlmayr, A., Blügel, S. & Kotani, T. Elimination of the linearization error in GW calculations based on the linearized augmented-plane-wave method. *Phys. Rev. B* **74**, 045104, DOI: [10.1103/PhysRevB.74.045104](https://doi.org/10.1103/PhysRevB.74.045104) (2006).
104. Betzinger, M., Friedrich, C., Blügel, S. & Görling, A. Local exact exchange potentials within the all-electron FLAPW method and a comparison with pseudopotential results. *Phys. Rev. B* **83**, 045105, DOI: [10.1103/PhysRevB.83.045105](https://doi.org/10.1103/PhysRevB.83.045105) (2011).
105. Bröder, J., Wortmann, D. & Blügel, S. Using the AiiDA-FLEUR package for all-electron ab initio electronic structure data generation and processing in materials science. In *In Extreme Data Workshop 2018 Proceedings*, vol. 40 of *IAS Series*, p 43–48 (Forschungszentrum Jülich, Jülich, 2019).
106. Bröder, J. *et al.* JuDFTteam/aiida-fleur: AiiDA-FLEUR. Zenodo, DOI: [10.5281/zenodo.6420726](https://doi.org/10.5281/zenodo.6420726) (2022).
107. Madsen, G. K. H., Blaha, P., Schwarz, K., Sjöstedt, E. & Nordström, L. Efficient linearization of the augmented plane-wave method. *Phys. Rev. B* **64**, 195134, DOI: [10.1103/PhysRevB.64.195134](https://doi.org/10.1103/PhysRevB.64.195134) (2001).
108. Karsai, F., Tran, F. & Blaha, P. On the importance of local orbitals using second energy derivatives for *d* and *f* electrons. *Comput. Phys. Commun.* **220**, 230, DOI: [10.1016/j.cpc.2017.07.008](https://doi.org/10.1016/j.cpc.2017.07.008) (2017).
109. Lippert, G., Parrinello, M. & Hutter, J. A hybrid Gaussian and plane wave density functional scheme. *Mol. Phys.* **92**, 477–488 (1997).
110. Larsen, A. H., Vanin, M., Mortensen, J. J., Thygesen, K. S. & Jacobsen, K. W. Localized atomic basis set in the projector augmented wave method. *Phys. Rev. B* **80**, 195112, DOI: [10.1103/PhysRevB.80.195112](https://doi.org/10.1103/PhysRevB.80.195112) (2009).
111. Setup generation — GPAW (2022). https://wiki.fysik.dtu.dk/gpaw/setups/generation_of_setups.html.
112. Kresse, G. & Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B* **59**, 1758–1775, DOI: [10.1103/PhysRevB.59.1758](https://doi.org/10.1103/PhysRevB.59.1758) (1999).
113. Hamann, D. R. Optimized norm-conserving vanderbilt pseudopotentials. *Phys. Rev. B* **88**, 085117, DOI: [10.1103/physrevb.88.085117](https://doi.org/10.1103/physrevb.88.085117) (2013).
114. Schlipf, M. & Gygi, F. Optimization algorithm for the generation of oncv pseudopotentials. *Comput. Phys. Commun.* **196**, 36–44 (2015).
115. Garrity, K. F., Bennett, J. W., Rabe, K. M. & Vanderbilt, D. Pseudopotentials for high-throughput dft calculations. *Comput. Mater. Sci.* **81**, 446–452 (2014).
116. Vanderbilt, D. Soft self-consistent pseudopotentials in a generalized eigenvalue formalism. *Phys. Rev. B* **41**, 7892 (1990).

Acknowledgments

This work is supported in part by the European Union’s Horizon 2020 research and innovation programme under grant agreement No. 824143 (European MaX Centre of Excellence “Materials design at the Exascale”) and by NCCR MARVEL, a National Centre of Competence in Research, funded by the Swiss National Science Foundation (grant number 205602). For the purpose of Open Access, a CC BY public copyright licence is applied to any Author Accepted Manuscript (AAM) version arising from this submission.

We acknowledge Flaviano José dos Santos for useful discussions on the analysis of the smearing types and k-point convergence, and Xavier Gonze, Marc Torrent, and François Jollet for useful discussions on PAW pseudopotentials.

M.F. and N.M. acknowledge the contribution of Sadasivan Shankar in early discussions about the use of 6 prototype oxides as general platform to explore the transferability of pseudopotentials.

Work at ICMAB (E.B., A.G., V.D.) is supported by the Severo Ochoa Centers of Excellence Program (MCIN CEX2019-000917-S), by grant PGC2018-096955-B-C44 of MCIN/AEI/10.13039/501100011033, “ERDF A way of making Europe”, and by GenCat 2017SGR1506. We also thank the Barcelona Supercomputer Center for computational resources. V.D. acknowledges support from DOC-FAM, European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 754397.

O.R. acknowledges travel support from WIEN2k (Technical University of Vienna).

The Jülich team (S.B., J.B., H.J., G.M., D.W) acknowledges support by the Joint Lab Virtual Materials Design (JL-VMD) of the Forschungszentrum Jülich, the Helmholtz Platform for Research Software Engineering - Preparatory Study (HIRSE_PS), and we gratefully acknowledge the computing time granted through JARA on the supercomputers JURECA⁸³ at Forschungszentrum Jülich and CLAIX at RWTH Aachen University.

H.M and T.D.K (University of Paderborn) gratefully acknowledge the Gauss Centre for Supercomputing e.V. (www.gauss-centre.eu) for funding this project by providing computing time on the GCS Supercomputer JUWELS at Jülich Supercomputing Centre (JSC).

S.P. and G.-M.R. (Université catholique de Louvain) acknowledge support from the F.R.S.-FNRS. Computational resources have been provided by the PRACE-21 resources MareNostrum at BSC-CNS and by the Consortium des Équipements de Calcul Intensif (CÉCI), funded by the Fonds de la Recherche Scientifique de Belgique (F.R.S.-FNRS) under Grant No. 2.5020.11 and by the Walloon Region as well as computational resources awarded on the Belgian share of the EuroHPC LUMI supercomputer.

G.Ka. and S.V. received funding from the VILLUM Centre for the Science of Sustainable Fuels and Chemicals (9455) from VILLUM FONDEN. Computational resources were provided by the Niflheim supercomputing cluster at the Technical University of Denmark (DTU). They also thank Jens Jørgen Mortensen and Ask H. Larsen for the valuable discussions on optimizing the workflow for the GPAW code.

S.C. acknowledges financial support from OCAS NV by an OCAS-endowed chair at Ghent University. The computational resources and services used at Ghent University were provided the VSC (Flemish Supercomputer Center), funded by the Research Foundation Flanders (FWO) and the Flemish Government – department EWI.

M.W. gratefully acknowledges computational resources provided by the Vienna Scientific Cluster (VSC). This research was funded in part by the Austrian Science Fund (FWF) [P 32711].

E.F.L. would like to acknowledge resources provided by Sigma2 - the National Infrastructure for High Performance Computing and Data Storage in Norway and support from the Norwegian Research Infrastructure Services (NRIS).

B.Z. is grateful to the UK Materials and Molecular Modelling Hub for computational resources, which is partially funded by EPSRC (EP/P020194/1 and EP/T022213/1) and acknowledge the use of the UCL Myriad and Kathleen High Performance Computing Facility (Myriad@UCL, Kathleen@UCL), and associated support services, in the completion of this work.

N.M., G.Pi., and A.G. acknowledge support from the European Union’s Horizon 2020 research and innovation programme under grant agreement No. 957189 (BIG-MAP), also part of the BATTERY 2030+ initiative under grant agreement No. 957213.

G.P., J.Y. and G.-M.R. acknowledge support by the Swiss National Science Foundation (SNSF) and by the Fonds de la Recherche Scientifique de Belgique (F.R.S.-FNRS) through the “FISH4DIET” Project (SNSF grant 200021E_206190 and F.R.S.-FNRS grant T.0179.22).

G.P. acknowledges support by the Open Research Data Program of the ETH Board, under the Establish project “PREMISE”.

J.Y. acknowledges support from the European Union’s Horizon 2020 research and innovation programme under grant agreement No. 760173 (MARKETPLACE).

Author contributions

S.C. and K.L. developed an initial early design of this research topic, and analyzed together with M.C. the first exploratory datasets. M.S. was responsible for job, queue and data management in the first exploratory phase. E.B. and G.Pi. contributed the idea to use the AiiDA and `aiida-common-workflow` infrastructure to carry on the thousands of DFT simulations required by the project. E.B. and G.Pi. coordinated the whole project.

M.F. and N.M. contributed the idea of using six oxides to test pseudopotentials across different coordinations and chemistries. P.B., G.M., and O.R. performed the iterative refinement of the input parameters for the AE calculations, that ultimately resulted in the generation of the central volumes of our dataset. O.R. proposed the ϵ metric. N.M. proposed the v metric. S.P. raised the issue of the smearing selection, that ultimately led to the decision of a fixed k-point integration mesh and smearing broadening. K.E. contributed to the data analysis and the conversion of the data into a dynamic website. D.E.P.V. contributed the Hirshfeld-I calculated charges and their analysis. E.B., M.W. and G.Pi. performed the analysis of the error propagation in the fit and the estimation of the parameters of the v metric. E.B., S.C., O.R., and G.Pi. analyzed in detail the dependency and sensitivity of the metrics Δ , ϵ and v .

A.Z. and S.P. developed the ABINIT implementation of the common workflow which relies on the `aiida-abinit` plugin developed and maintained by A.Z., G.Pe. and S.P. M.G. created new pseudopotentials used for ABINIT and improved the parameter profile. A.Z. and S.P. performed the ABINIT calculations, including verification tests. The work on ABINIT was supervised by G.M.R. and S.P.

L.B., A.D. and L.G. contributed to the BigDFT-related parts of the work. A.D. developed the BigDFT implementation of the common workflows. L.B. and A.D. generated the BigDFT results under the supervision of L.G.

B.Z. developed the CASTEP implementation of the common workflow which relies on the `aiida-castep` plugin which is also maintained by B.Z., and performed all CASTEP simulations. C.J.P. created new on-the-fly generated pseudopotentials using the verification tests performed by B.Z.

M.K., T.D.K., H.M., T.M.A.M. and A.V.Y. contributed to all CP2K-related parts of this work. A.V.Y. implemented the workflows and performed preliminary calculations. The workflows rely on the `aiida-cp2k` plugin developed by A.V.Y., T.M.A.M., and others. M.K. performed preliminary calculations, created new pseudopotentials and contributed to the design of the protocol, T.D.K. contributed to the CP2K setup, discussion of the results and supervised the calculations, H.M. conducted all AiiDA calculations and analyzed the results, T.M.A.M. provided implementations of CP2K input and output parsers.

S.B., J.B., H.J., G.M. and D.W. contributed the FLEUR-related parts of this work¹. G.M. hereby developed the parameter profile and performed the calculations. J.B. and H.J. adapted and extended the AiiDA-FLEUR plugin and the related parts of the AiiDA common-workflows package. S.B. and D.W. contributed to the analysis and discussion of the FLEUR results.

G.Ka. and S.V. contributed to the GPAW-related parts of the work. S.V. developed the GPAW implementation of the common workflows, which relies on the `aiida-ase` plugin and ran the calculations. G.Ka. and S.V. analyzed the GPAW calculations.

M.B., S.P.H., N.M., J.Y. and G.Pi. contributed to all QUANTUM ESPRESSO-related parts of this work. M.B. and S.P.H. developed the QUANTUM ESPRESSO implementation of the common workflow which relies on the `aiida-quantumespresso` plugin developed and maintained by M.B., S.P.H., G.Pi. and others. J.Y. generated and tested new pseudopotentials used for QUANTUM ESPRESSO. The work on QUANTUM ESPRESSO was supervised by G.Pi. and N.M.

H.M. and T.D.K. contributed to all SIRIUS/CP2K-related parts of this work. The workflows rely on the `aiida-cp2k` plugin developed by A.V.Y., T.M.A.M., and others. H.M. conducted all AiiDA calculations and analyzed the results. T.D.K. contributed to the SIRIUS/CP2K setup, discussion of the results and supervised the calculations.

E.B., V.D. and A.G. contributed to the SIESTA-related parts of the work. E.B. developed the SIESTA implementation of the common workflows, that relies on the `aiida-siesta` plugin developed by E.B., A.G., V.D. and others. E.B. generated the SIESTA results in collaboration with A.G.

M.W., M.M. and E.F.L. performed the execution and analysis of the VASP-related workflows used to generate the data for this work. G.Kr. generated updated potentials for the lanthanides. E.F.L. maintains the VASP implementation of the common workflows project and the `aiida-vasp` plugin (developed by a community of contributors, see full contributor list in the plugin documentation) which is used to execute the VASP calculations.

P.B., G.K.H.M., O.R. and T.R. contributed the WIEN2k-related parts of this work. T.R. performed preliminary calculations, P.B. created the setup of the WIEN2k calculations and supervised and analyzed the results, G.K.H.M. contributed the conversion of AiiDA structures to a WIEN2k struct file, and O.R. developed the `aiida-wien2k` plugin and performed all AiiDA-WIEN2k calculations.

E.B., M.F. and G.Pi. wrote the first version of the manuscript, and all authors contributed to the editing and revision of the manuscript.

Competing interests

G.Pe. and G.-M.R. are shareholders and Directors of Matgenix SRL. G.Kr. is shareholder of the VASP Software GmbH, and M.W. and M.M. are part-time employees of the VASP Software GmbH. C.J.P. is an author of the CASTEP code and receives income from its commercial sales.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Code availability

The source code of the common workflows is released under the MIT open-source license and is made available on GitHub (<https://github.com/aiidateam/aiida-common-workflows>). It is also distributed as an installable package through the Python

¹Corresponding contact person for the FLEUR contributions: Gregor Michalicek, g.michalicek@fz-juelich.de

Package Index (<https://pypi.org/project/aiida-common-workflows>). The source code of the scripts to generate the plots is released under the MIT open-source license and is made available on GitHub (<https://github.com/aiidateam/acwf-verification-scripts>). All codes to generate the figures of this paper are available in the data entry of Ref. 62.

Data availability

The data and the scripts used to create all the images in this work are available on the Materials Cloud Archive⁶². Note that the data includes the entire AiiDA provenance graph of each workflow execution presented in the main text (including therefore all input files and output files of all simulations, as well as their logical relationship, in AiiDA format), as well as the curated data that is extracted from that database in order to produce the images.

Supplementary information for “How to verify the precision of density-functional-theory implementations via reproducible and universal workflows”

Emanuele Bosoni, Louis Beal, Marnik Bercx, Peter Blaha, Stefan Blügel, Jens Bröder, Martin Callsen, Stefaan Cottenier, Augustin Degomme, Vladimir Dikan, Kristjan Eimre, Espen Flage-Larsen, Marco Fornari, Alberto Garcia, Luigi Genovese, Matteo Giantomassi, Sebastiaan P. Huber, Henning Janssen, Georg Kastlunger, Matthias Krack, Georg Kresse, Thomas D. Kühne, Kurt Lejaeghere, Georg K. H. Madsen, Martijn Marsman, Nicola Marzari, Gregor Michalicek, Hossein Mirhosseini, Tiziano M. A. Müller, Guido Petretto, Chris J. Pickard, Samuel Poncé, Gian-Marco Rignanese, Oleg Rubel, Thomas Ruh, Michael Sluydts, Danny E. P. Vanpoucke, Sudarshan Vijay, Michael Wolloch, Daniel Wortmann, Aliaksandr V. Yakutovich, Jusong Yu, Austin Zadoks, Bonan Zhu, Giovanni Pizzi

S1 Structures under investigation

This section reports some details on the crystal structures used in the verification study. As already explained in the main text, we consider two subsets: the “unaries dataset” and the “oxides dataset”.

The “unaries dataset” consists of 4 monoelemental cubic crystals for every element from $Z=1$ (hydrogen) to $Z=96$ (curium), in the well-known structures face-centered cubic, body-centered cubic, simple cubic and in the diamond structure. The details of each of the 4 monoelemental cubic crystals are described in Table S1.1, together with the indication of a prototype belonging to each category. A visualization of the crystal structures is reported in SI Fig. S1.1.

Table S1.1. Description of the four unary crystals under investigation, with a prototypical crystalline example and the corresponding ID from the ICSD database⁸⁴. The quantity l is the length of the primitive-cell lattice vectors, a the cubic conventional-cell side, and d_{nn} is the nearest-neighbor distance.

	Prototype (ICSD number)	Space group	Wyckoff site (site symmetry)	Coordination number	l	d_{nn}
FCC	Al (43423)	Fm $\bar{3}$ m (225)	4a (m $\bar{3}$ m)	12	$a/\sqrt{2}$	$a/\sqrt{2}$
BCC	V (43420)	Im $\bar{3}$ m (229)	2a (m $\bar{3}$ m)	8	$\sqrt{3}a/2$	$\sqrt{3}a/2$
SC	α -Po (43211)	Pm $\bar{3}$ m (221)	1a (m $\bar{3}$ m)	6	a	a
Diamond	C (diamond) (28857)	Fd $\bar{3}$ m (227)	8a ($\bar{4}$ 3m)	4	$a/\sqrt{2}$	$\sqrt{3}a/4$

The “oxides dataset” is composed by six cubic oxides with chemical formula X_2O , XO , X_2O_3 , XO_2 , X_2O_5 and XO_3 , where X goes also in this case from hydrogen to curium. The details of each of these structures are reported in Table S1.2, that also includes the formal oxidation number that is expected for X in the structure (we stress that the actual oxidation state is different from the formal charge, see SI Sec. S2). A visualization of the crystal structures is reported in SI Fig. S1.2.

For every material, the primitive cell is provided as input of the verification study (except when X is oxygen; in this case, the same cell is used as for all other oxides where X is different from oxygen, even if in the case of oxygen some of these cells might not be the smallest primitive cell, due to the increased symmetry). However, some codes might prefer to perform the actual calculation of the EOS on the cubic conventional cell, or use the actual primitive cell in the case of $X =$ oxygen. The actual number of atoms in the simulation is reported inside the JSON files with the results of this verification study, available in Ref. 62.

In the following, we will often refer to quantities (energy, volume, etc.) per formula unit. To avoid ambiguity, we explicitly list here the number of atoms in the formula unit for each of our 10 prototypes: FCC (1), BCC(1), SC(1), Diamond (2), X_2O (3), XO (2), X_2O_3 (5), XO_2 (3), X_2O_5 (7), XO_3 (4). We highlight that these numbers also correspond to the number of atoms in the primitive cell, except for X_2O_3 and X_2O_5 that have 10 and 14 atoms in the primitive cell, respectively.

Finally, we report the central volumes for all 960 structures used for the calculation of the EOS data in SI Table S1.3. In order to compare results, the same central volumes (and the same volume range of $\pm 6\%$, with 7 points) should be used when generating additional datasets. To visualize the data, we report in SI Fig. S1.3 the distance of the X atom from its closest (oxygen) neighbor, across the whole periodic table and for the 10 prototypes.

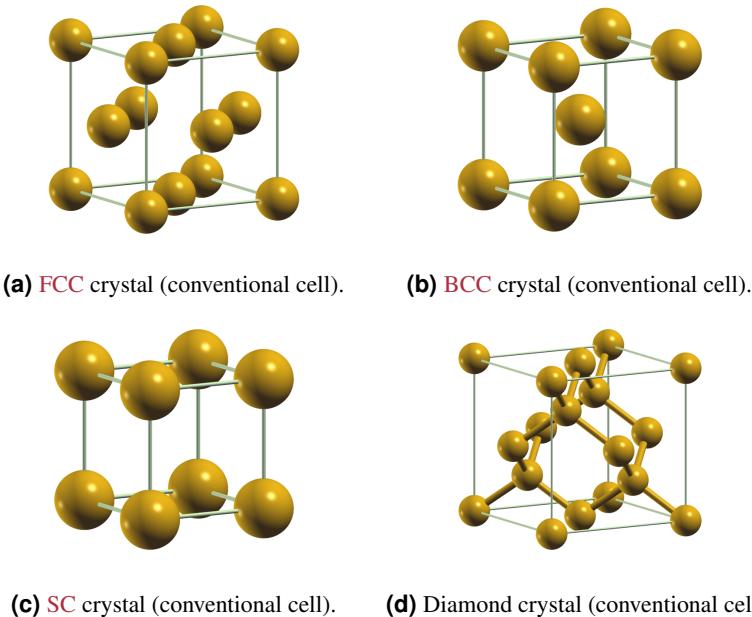


Figure S1.1. Conventional cells of the 4 unary prototypes used in this work. Images generated using XCrysDen⁸⁵.

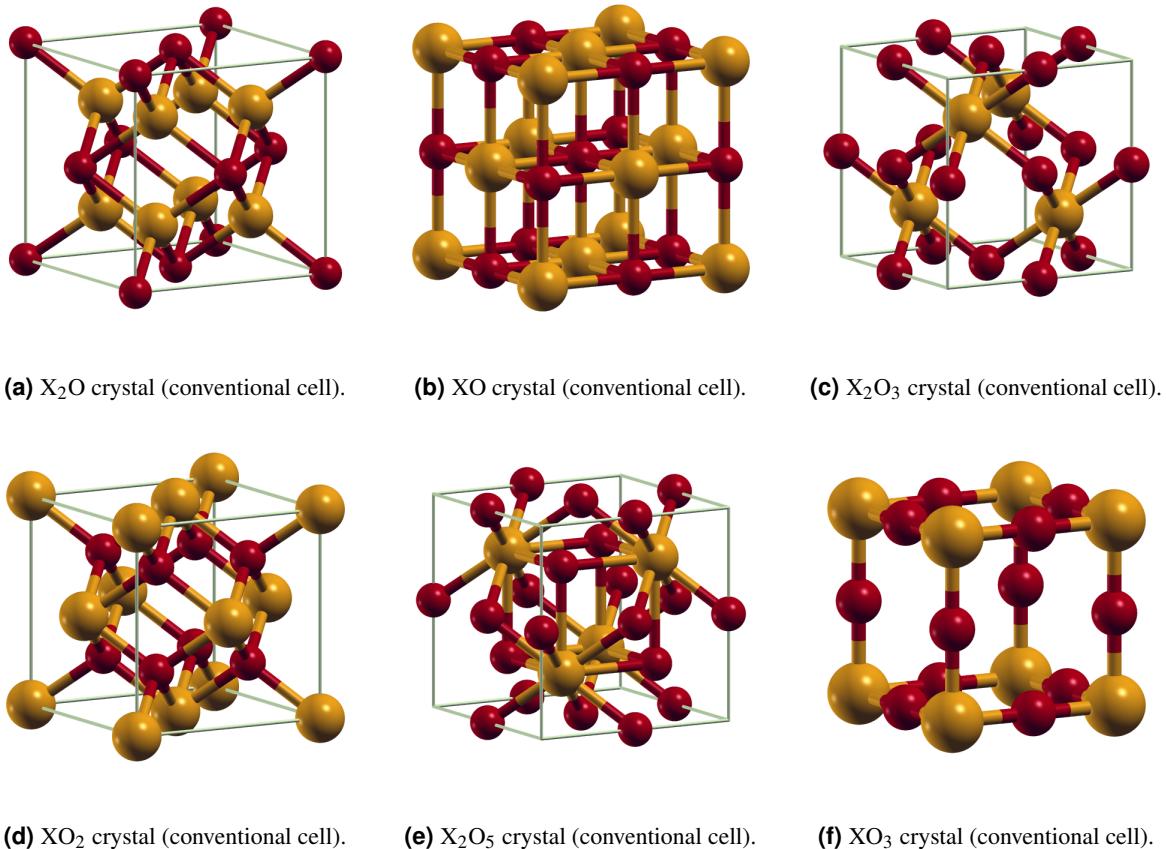


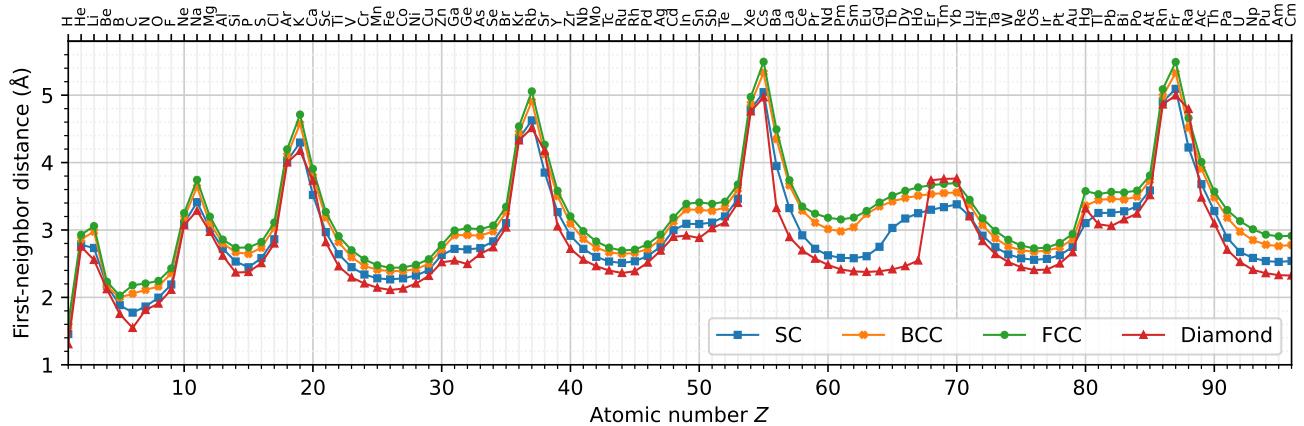
Figure S1.2. Conventional cells of the 6 oxide prototypes used in this work. Oxygen atoms are represented as red atoms, while X atoms as orange atoms. Images generated using XCrysDen⁸⁵.

Table S1.2. Description of the crystal structure of the six cubic oxides, with a prototypical crystalline example and the corresponding ID from the ICSD database⁸⁴. The quantity l is the length of the primitive-cell lattice vectors, a the cubic conventional-cell side, and d_{nn} is the distance of the X atom to its nearest-neighbor (oxygen) atom.

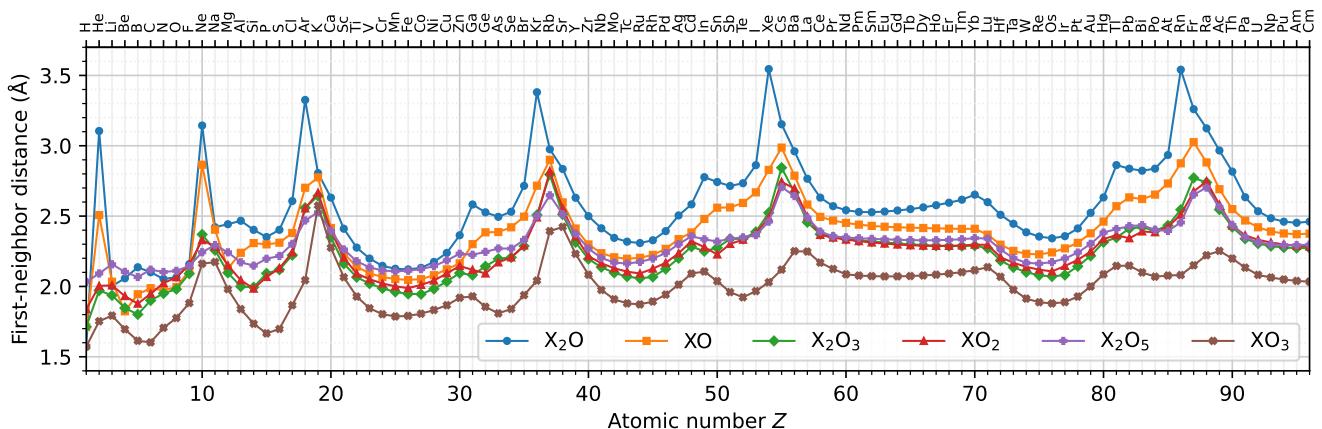
	Formal oxidation of X	Prototype (ICSD number)	Space group	Wyckoff site site symmetry of X	Coordination of X	l	d_{nn}
X ₂ O	+1	Na ₂ O (60435)	Fm $\bar{3}m$ (225)	X=8c, O=4a -43m	4	$a/\sqrt{2}$	$a\sqrt{3}/4$
XO	+2	NaCl (18189)	Fm $\bar{3}m$ (225)	X=4a, O=4b m-3m	6	$a/\sqrt{2}$	$a/2$
X ₂ O ₃	+3	Ag ₂ O ₃ (15999)	Pn $\bar{3}m$ (224)	X=4b, O=6d -3m	6	a	$a\sqrt{3}/4$
XO ₂	+4	ZrO ₂ (105553)	Fm $\bar{3}m$ (225)	X=4a, O=8c m-3m	8	$a/\sqrt{2}$	$a\sqrt{3}/4$
X ₂ O ₅	+5	—	Pn $\bar{3}m$ (224)	X=4b, O=4c+6d -3m	6	a	$a\sqrt{3}/4$
XO ₃	+6	ReO ₃ (647352)	Pm $\bar{3}m$ (221)	X=1a, O=3d m-3m	6	a	$a/2$

Table S1.3. Table with the central volumes used for the calculation of the EOS datapoints. Volumes are expressed in Å³ per formula unit (see definition of the formula unit in SI Sec. S1). The reference structures having these central volumes are available in Ref. 62.

	FCC	BCC	SC	Diamond	X ₂ O	X ₂ O ₅	XO ₂	X ₂ O ₃	XO	XO ₃
H	2.96383	2.96392	3.08364	6.84867	11.96463	51.67201	19.10477	30.91989	10.02535	31.07589
He	17.83621	18.12465	21.38414	64.32268	92.12727	56.39662	24.77499	47.12856	31.51877	43.04663
Li	20.21287	20.26593	20.40472	51.36159	24.72203	61.84374	24.92385	44.77362	16.82806	46.06591
Be	7.87403	7.81517	10.26455	29.37901	26.83653	57.32171	22.18348	38.79261	12.12488	39.00478
B	5.89415	6.14152	6.6991	16.62498	30.02954	54.37317	20.37072	35.98325	14.71073	33.62805
C	7.31505	6.69648	5.60717	11.39533	28.55083	58.60995	22.75464	42.26604	15.70062	32.91818
N	7.60577	7.23569	6.48497	18.35414	26.67062	57.22806	25.50994	45.6704	15.3358	39.68942
O	8.00192	7.7972	7.94802	21.36412	27.11225	57.93205	27.11223	47.87245	15.90618	44.79083
F	10.14406	10.08192	10.52097	29.004	30.96209	60.78463	30.83431	56.31851	18.85462	53.30451
Ne	24.26591	24.70382	29.68755	89.0995	95.7889	69.51806	38.90827	81.94501	47.11781	80.77935
Na	37.10691	36.99607	39.7588	108.85725	43.71875	74.35023	36.67184	70.84884	27.7619	82.17703
Mg	23.11539	22.93116	27.59134	80.79618	44.83726	69.448	30.59043	56.77165	19.2497	62.13962
Al	16.48998	16.92508	20.17082	55.21453	46.20985	63.07075	26.30127	49.23828	22.4582	49.74088
Si	14.4803	14.66715	16.23082	40.92143	42.69026	61.00716	24.05335	48.82219	24.59467	41.777
P	14.58744	14.29218	14.60552	41.27791	40.01409	65.27141	27.20059	56.43141	24.30168	37.00066
S	15.88301	15.73711	17.19755	48.58823	42.6917	67.07762	29.73936	58.62747	24.68481	39.1327
Cl	21.29569	21.46346	23.45819	67.51736	54.55214	75.16606	34.69697	67.62641	26.95723	52.00364
Ar	52.33201	53.50562	65.51497	198.12738	113.32949	92.69732	51.32183	103.20479	39.41786	68.32473
K	73.99534	73.80511	79.47128	224.24589	68.00111	99.39122	58.36793	113.79592	42.74207	136.60539
Ca	42.20189	42.15587	43.70466	160.093	56.09468	84.42406	42.31153	79.93599	28.18987	94.27512
Sc	24.6858	24.88426	26.11875	68.84236	43.1186	71.37218	32.98765	62.21668	22.3317	70.50445
Ti	17.39633	17.26807	18.40137	45.88764	36.3163	63.72954	28.1791	54.26031	19.61645	57.35797
V	13.9076	13.46008	14.68791	37.2808	32.71378	59.96306	26.57435	50.7807	18.31774	50.25823
Cr	11.89373	11.55544	12.80026	33.09544	30.44097	58.12757	25.42189	48.27373	17.6635	46.89194
Mn	10.75345	10.78666	11.9053	30.35903	29.56957	57.65839	24.64946	46.45968	17.2926	45.62581
Fe	10.26671	10.50643	11.65681	28.93599	29.39392	58.00272	24.1901	45.43045	17.13572	45.9389
Co	10.31329	10.54766	11.90016	29.77725	29.83543	58.94975	25.03188	45.30703	17.27042	47.13241
Ni	10.83846	10.90046	12.56525	33.01991	31.63922	60.71118	26.20001	47.99648	17.99342	49.20696
Cu	11.96066	12.00521	13.9456	38.3573	34.5396	64.24158	28.23319	51.90164	19.10594	52.00277
Zn	15.15266	15.35236	18.21695	49.37022	40.67075	68.51634	30.3891	56.63194	20.27908	56.55698
Ga	18.89945	19.1961	20.12767	50.86204	53.04849	67.83879	29.26061	55.49587	24.35231	57.48398
Ge	19.61105	19.26408	19.92661	47.84474	49.67078	69.75991	28.19282	60.60137	27.12181	51.09293
As	19.25156	19.06952	20.35448	57.08944	47.81491	72.02994	31.50589	65.30072	27.1941	47.3416
Se	20.38999	20.33796	22.67444	63.49704	49.91744	71.9333	33.07453	66.14337	28.35245	49.87193
Br	26.41028	26.78091	29.81278	86.15579	61.62718	77.98233	37.01169	73.66672	31.10869	58.2754
Kr	66.18624	67.66229	82.81744	250.49559	118.94061	96.19024	47.53319	97.16926	40.08181	67.9248
Rb	91.38789	91.27765	99.14298	283.10731	81.07467	114.34806	69.0703	134.5233	48.78267	109.64128



(a) First-neighbor distance for the unaries dataset.



(b) First-neighbor distance of the X atom for the oxides dataset (the first neighbor is, in all cases considered here, an oxygen atom).

Figure S1.3. First-neighbor distance of the X atom to its closest neighbor for all 960 systems in our dataset.

S2 Hirshfeld-I charges

Six different oxide crystals are imposed, in order to force the element X into 6 different formal oxidation states. The hope is that this will bring each element into 6 chemically sufficiently different environments. In this section, we analyze whether this expectation has been realized. This is done by monitoring the Hirshfeld-I charges throughout this oxide set, as a proxy for the chemical environment.

S2.1 Hirshfeld-I atoms-in-molecules methodology

Atomic charges have been calculated within the context of an atoms-in-molecules (AIM) approach. The basic goal of these approaches is to divide the electrons, or more specifically the electron density, of a multi-atom system into subunits associated with chemical atoms. One can either start from the calculated wavefunctions (e.g., Mulliken charges^{86,87}) or from the electron density distribution (EDD) (e.g., Hirshfeld⁸⁸ or Bader⁸⁹ charges). In this work, the iterative Hirshfeld approach (HI), which is an improvement of the Hirshfeld approach, is used. This approach alleviates the dependence on the chosen initial references of the original Hirshfeld method^{58–60}. In Hirshfeld (and other stockholder) methods, the EDD in each point in space is divided over all the nearby AIM, in contrast to, for example, the Bader method⁸⁹, which assigns the entire electron density of a given point in space to a single AIM. This gives rise to smooth AIM which overlap in real-space. The Hirshfeld weights for an atom A are defined as:

$$w_A^H(r) = \frac{\rho_A^{AIM}(r)}{\rho_{mol}(r)}, \quad (S1)$$

where $\rho_A^{AIM}(r)$ and $\rho_{mol}(r)$ are the EDDs for the AIM and the molecule respectively. This however creates a circular reference, as the AIM EDD is calculated using the Hirshfeld weights. As a solution, Hirshfeld suggested the use of a reference EDD defined as the spherical average of the EDD of the free atom in a chosen reference state⁸⁸. To make sure weights at every point in space remain normalized to unity, the molecular density is replaced by the sum of the atomic reference EDD, giving rise to a so-called *promolecular* EDD. Within this setup, one has to chose suitable atomic reference states, and it was found that for different atomic reference states, different atomic charges were obtained. This issue was resolved by Bultinck *et al.*⁶⁰, who proposed an extension of the scheme by iterative modification of the reference state. Starting with, for example, neutral reference EDDs, $\rho_A^0(r)$, the $w_A^H(r)$ are calculated. From these the AIM EDDs are calculated as:

$$\rho_A^{AIM}(r) = \frac{\rho_A^0(r)}{\rho_{promol}^0(r)} \rho_{mol}(r). \quad (S2)$$

With these AIM EDDs the atomic charge, x , is calculated through integration over the entire system. In the following step, the reference EDD is constructed as the linear interpolation of EDDs with atomic charge $I < x < I + 1$, with I the integer value of the ionic charge lower than x . Using these reference EDDs, $\rho_A^x(r)$, new weights, AIM EDDs, and atomic charges are calculated. This scheme is then iterated to convergence of the atomic charges.

For periodic systems, the problem of the infinite size of the system is resolved by only considering the atoms of the unit cell for calculation of the charges and periodic copies which are “nearby”⁵⁸. The Hirshfeld weights are calculated for all grid points (of an atom centered Becke grid⁹⁰) associated with the atoms of the unit cell, and all other grid points which are located in the same spatial region. Atoms contributing to the weights are thus the unit cell atoms, as well as periodic copies within a limited range^{58,59}. Furthermore, it was found that the EDDs of the valence electrons (i.e., all electrons not included in the frozen core) can be used without loss of quality compared to all-electron EDDs, while using a much coarser grid.

S2.2 Computational settings

In this work, HI-charges are calculated using the previous implementation for periodic systems,^{58,59} as found in the HIVE package⁹¹. The calculations for generating the EDDs are performed using the VASP package. Reference atomic densities are calculated using a small unit cell of $20 \times 20 \times 20 \text{ \AA}^3$ for the cations, while a large $40 \times 40 \times 40 \text{ \AA}^3$ cell is used for the tail correction of the anions. The plane wave kinetic energy cut off is set to 1000 eV. The EDDs of the oxide and unary systems are obtained from static calculations using a $33 \times 33 \times 33$ Γ -centered k-point integration mesh and a kinetic energy cut off of 1000 eV, using the PBE functional as defined in S16.5. The atomic charges of the systems are calculated using the HI partitioning scheme with a charge convergence criterion of 1.0×10^{-4} electron. Charges are integrated on a logarithmic radial grid with atom-centered spherical shells of 1202 Lebedev–Laikov grid points^{90,92}.

S2.3 Discussion of Hirshfeld-I atomic charges and their relation to formal oxidation states

By imposing the topology of the oxides, every element X should exist in a predetermined formal oxidation state covering all integer values from +1 to +6. This formal oxidation state, however does not correspond one-to-one to the local configuration

of the charge density around element X that describes how element X binds to the surrounding oxygens. This is particularly relevant for those oxides that are ‘*exotic*’ (e.g., hydrogen in HO_3 has a formal oxidation state of +6, whereas even in a complete ionic picture hydrogen can only donate a single electron). In order to survey the actual chemical environment of every element X, we therefore calculated the Hirshfeld-I charges for X in all oxides and unaries (not shown), near the equilibrium volume^{58–60}. As any other AIM scheme to define charges, Hirshfeld-I charges have their limitations. A common limitation all AIM schemes have to deal with is the fact that atomic charge is not a quantum mechanical observable. As such, there exists no absolute true value to find. The choice of the specific AIM scheme to calculate charges is therefore guided by the wish to satisfy other requirements. Attractive features of the Hirshfeld-I charges in this context are: They are (1) basis-set independent⁹³, (2) very robust, meaning charges are not structure dependent if the chemical environment remains the same, while very sensitive to changes in the chemical environment or oxidation state^{94–96}, (3) and large, though always smaller than the formal charge.

The results in SI Fig. S2.1 show the variation of the oxygen and metal charges over the periodic table as function of the formal oxidation state. These pictures reflect several trends that are intuitively expected: for X_2O (formal oxidation state +1), the elements with a HI charge closest to 1 are the alkali elements, while for XO (formal oxidation state +2) the elements with a HI charge closest to 2 are the earth-alkaline elements. On the other hand, the HI charges are clearly limited and are often about one half of the formal oxidation state. For instance, for XO_3 (formal oxidation state +6), the HI charges are often in the range 2.5–3.5.

Looking at the entire distribution of the atomic charges over the entire oxide dataset, (see SI Fig. S2.2) shows that the average metal charge gradually shifts to higher charges with increasing formal charge. In case of the oxygen charges, note that the oxygen charge decreases in size with increasing formal charge of the metal, which is a consequence of the fact that the formal charge is never actually fully transferred. Visualizing the results per formal oxidation state of the oxide system (see SI Fig. S2.3) shows an increasing trend, as expected. More interestingly, if the materials are split in two subsets – those with a ‘reasonable’ formal oxidation state for X and those with an ‘exotic’ formal oxidation stated for X (see caption of SI Fig. S2.3) – it becomes clear that for the ‘reasonable’ subset the calculated Hirshfeld-I charge is on average about half of the formal charge. In the case of the other subset, a lower value is found. Taken together, the different features of SI Fig. S2.3 consistently express that even though the nominal formal charges are not obtained, the six different crystal structures for the oxides give rise to systematically different chemical environments. This was exactly the purpose of imposing these six different oxide crystal structures.

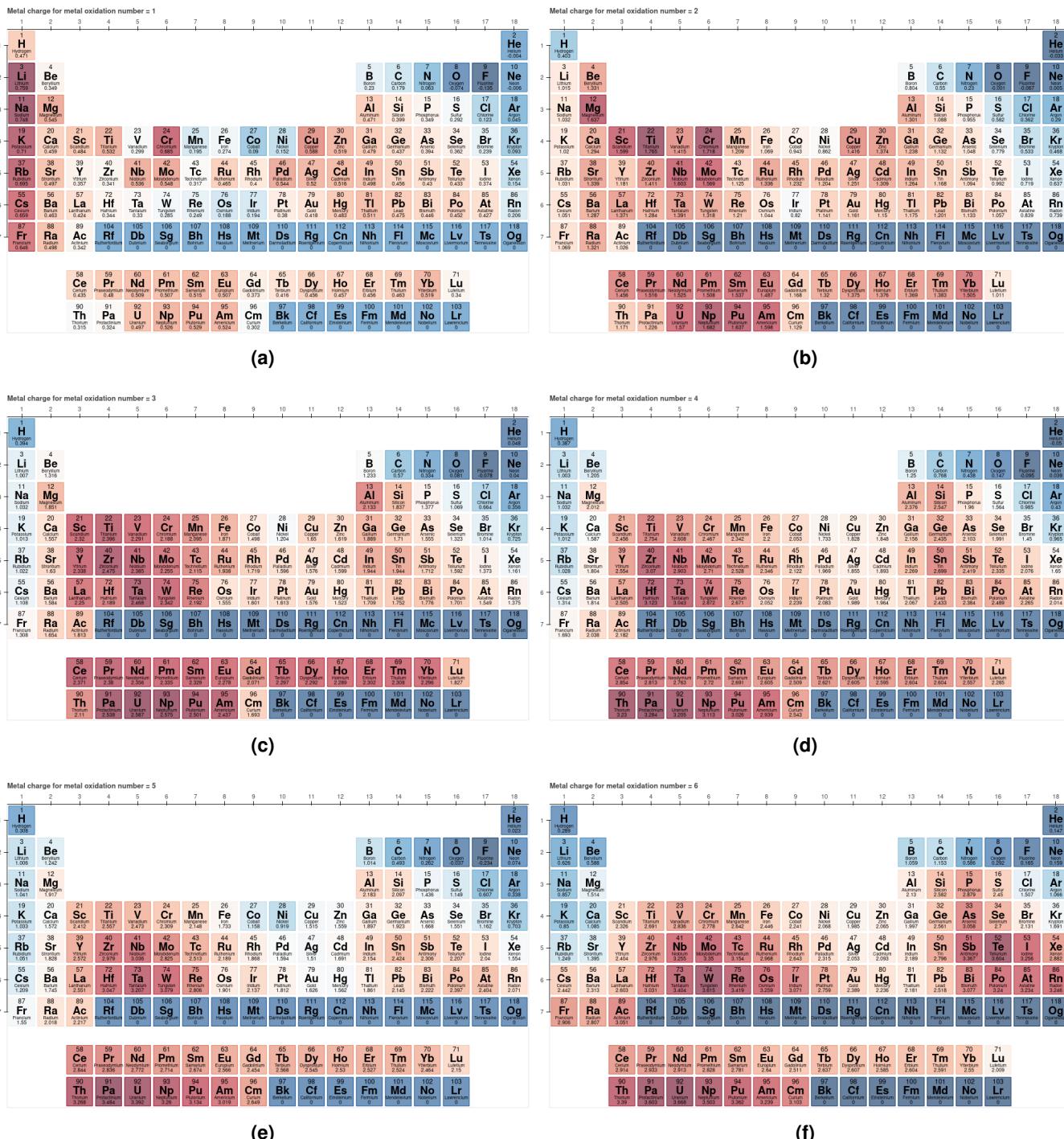


Figure S2.1. The average Hirshfeld-I charge on the metal atom for the oxides dataset according to the formal charge. Formal charge +1 (a), +2 (b), +3 (c), +4 (d), +5 (e), and +6 (f). Brownish colors indicate elements that have a Hirshfeld-I charge that is in line with their formal charge.

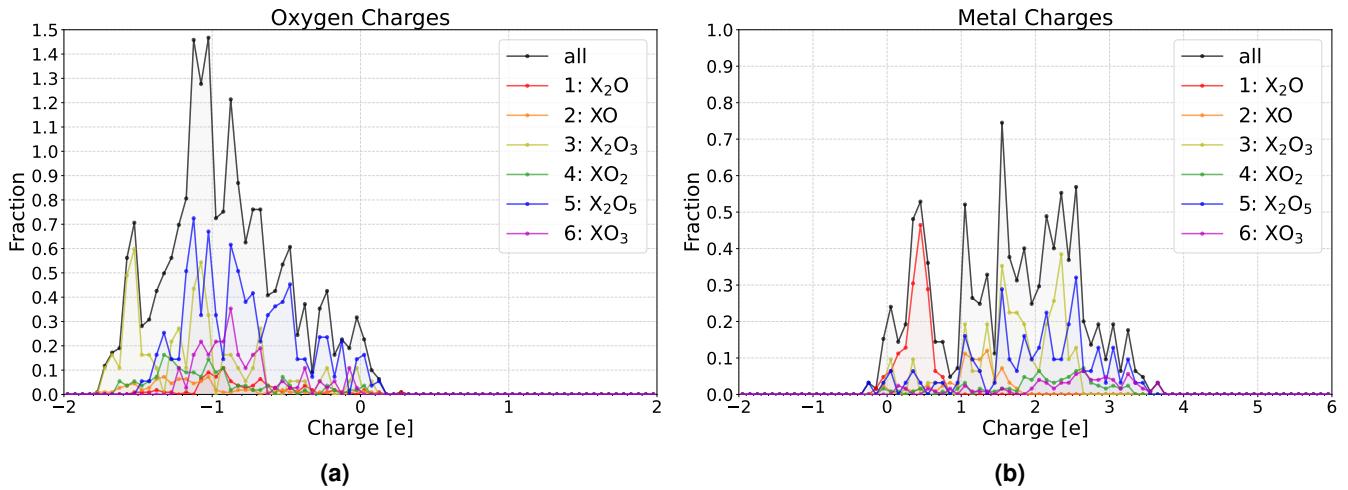


Figure S2.2. (a) Distribution of oxygen charges according to the Hirshfeld-I method for the oxides dataset. (b) Distribution of metal charges according to the Hirshfeld-I method for the oxides dataset.

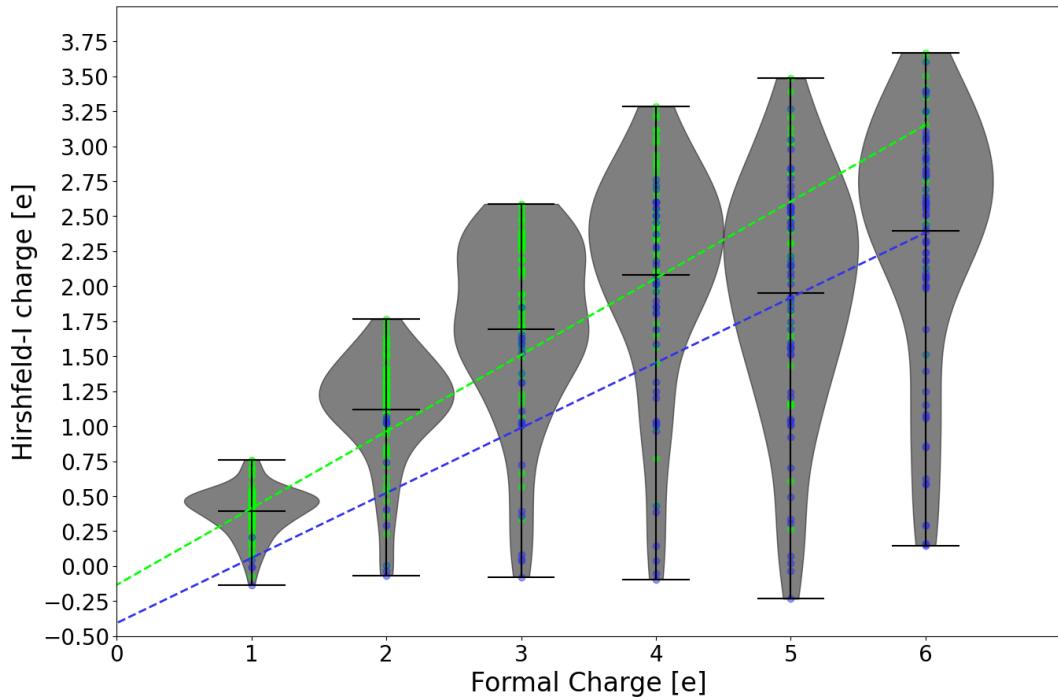


Figure S2.3. Violin plot of the HI charges of X in the oxides, as function of the formal oxidation state. Green dots and green linear fit: all oxides for which the formal oxidation state of X in this oxide is less than or equal to the maximal common formal oxidation state of X (as listed in Ref. 97). Blue dots and blue linear fit: all oxides for which the formal oxidation state of X in this oxide is larger than the maximal common formal oxidation state of X (as listed in Ref. 97). The green subset has therefore all oxides for which the formal oxidation state of X is ‘reasonable’, the blue subset represents oxides for which the formal oxidation state of X is ‘exotic’.

S3 Determination of the weights of the metric v based on the error propagation on the Birch–Murnaghan fit parameters

In this Section, we motivate the choice of weights $w_{V_0} = 1$, $w_{B_0} = 1/20$ and $w_{B_1} = 1/400$ discussed in the text for the v metric.

When computing the [EOS](#) curves $E(V)$, the results from any simulation are affected by numerical noise, originating from many different sources (finiteness of the k-point integration mesh, basis set discretization, thresholds to stop the self-consistent convergence cycle, …). When these points are fitted to a Birch–Murnaghan equation of state, the error propagates to the resulting fit parameters. This has been investigated in detail in Ref. 98, and several of the observations mentioned underneath are in line with the conclusions reached there. Intuitively, one can already expect that the numerical error will be larger for those parameters that are associated with higher-order derivatives. For instance, V_0 is the minimum of the [EOS](#) curve (i.e., the zero of the first derivative) is expected to be affected by a smaller error with respect to B_0 that is related to the curvature of the [EOS](#) curve close to the minimum (thus, to its second-order derivative). A first observation is that the error on all parameters will increase for increasing input noise on the energy datapoints. However, our goal in this section is not to quantify the error on each of these properties independently, but rather to understand if the error on pairs of fit parameters is related. In particular we will show that errors on B_0 (B_1) are typically 20 (400) times larger than those on V_0 ; by arbitrarily setting $w_{V_0} = 1$ (a change to this would result only in a global multiplicative factor), this will justify our final choice of weights.

We extract these relative weights using the following approach. We start from our reference [AE](#) dataset and consider, for each of the 960 materials, the fitted parameters V_0^{ref} , B_0^{ref} and B_1^{ref} . Rather than using the datapoints from the [AE](#) simulations, however, we generate a new “perfect” dataset (i.e., not affected by any numerical noise) by creating, for every curve, 7 fictitious points lying exactly on the Birch–Murnaghan curve, with the same volume spacing as discussed in the main text (spacing of 2% in volume between 94% and 106% of the tabulated central volume). This removes from our analysis any existing numerical noise of the [AE](#) simulations that is due to the numerical approximations in the two specific codes, rather than originating from the fitting procedure. We then select a reference average numerical error n_σ for the energy value of each point, and randomly displace each energy by a random value following a normal distribution with zero mean and standard deviation n_σ . We fit these noisy datapoints with the Birch–Murnaghan curve, thus obtaining fitted values of V_0 , B_0 and B_1 , that will be different from the initial reference ones V_0^{ref} , B_0^{ref} and B_1^{ref} . We therefore compute the relative errors on each of them with respect to their average (similar to what is done for ϵ and v): e.g., for the minimum volume $\eta_{V_0} = (V_0 - V_0^{\text{ref}})/((V_0 + V_0^{\text{ref}})/2)$, for the bulk modulus $\eta_{B_0} = (B_0 - B_0^{\text{ref}})/((B_0 + B_0^{\text{ref}})/2)$ and for the derivative of the bulk modulus $\eta_{B_1} = (B_1 - B_1^{\text{ref}})/((B_1 + B_1^{\text{ref}})/2)$. We repeat the procedure for N_s random samples, and finally compute the average of the absolute value of the three relative errors η_{V_0} , η_{B_0} and η_{B_1} on the N_s samples (for each material):

$$\bar{\eta}_{V_0} = \sum_{i=1}^{N_s} \frac{|\eta_{V_0}(i)|}{N_s} \quad (\text{S3})$$

and similarly for $\bar{\eta}_{B_0}$ and $\bar{\eta}_{B_1}$, where i denotes each of the individual independent random noise samples. The values $\bar{\eta}_{V_0}$, $\bar{\eta}_{B_0}$ and $\bar{\eta}_{B_1}$ quantify the typical average errors on the three fit parameters for a numerical noise of magnitude n_s . By producing histograms of the three quantities over the whole dataset of 960 structures, we obtain a peaked distribution that represents the range of values typical of our materials dataset ($V_0 \approx 3 - 400 \text{ \AA}^3$, $B_0 \approx 0.001 - 2.7 \text{ eV/\AA}^3$ and $B_1 \approx 0.4 - 12$). The position of the peak, as we expected, depends on the noise magnitude n_σ . As we discussed, however, we do not consider the histograms of these three quantities but we produce, instead, histograms for the two “relative” quantities

$$\frac{\bar{\eta}_{B_0}}{\bar{\eta}_{V_0}} \quad \text{and} \quad \frac{\bar{\eta}_{B_1}}{\bar{\eta}_{V_0}}. \quad (\text{S4})$$

The positions of the peaks of these histograms will represent the quantities we wish to determine: the typical ratio of numerical error on pairs of fit parameters. The results of our simulations can be summarized as follows:

- $N_s = 100$ samples are already enough to converge the statistics and the histograms for our goal of identifying the peaks of the histograms;
- the positions of the peaks of $\bar{\eta}_{V_0}$, $\bar{\eta}_{B_0}$ and $\bar{\eta}_{B_1}$ are roughly proportional to the input noise n_σ ; however, the position of the peaks of $\frac{\bar{\eta}_{B_0}}{\bar{\eta}_{V_0}}$ and $\frac{\bar{\eta}_{B_1}}{\bar{\eta}_{V_0}}$ are, to a good approximation, independent of n_σ for the noises that we considered (in the range $10^{-4} - 10^{-6} \text{ eV}$) (see also Ref. 98 for similar conclusions on the Δ metric);
- for our choice of volume range (94%–106%), the two histograms (see SI Fig. S3.1) display clear peaks at positions that can be rounded to 20 and 400, respectively. The peak positions are consistent when considering independently unaries and oxides (even if the spread of the peaks is different in the two cases). Hence, we choose the weights as $w_{V_0} = 1$, $w_{B_0} = 1/20$ and $w_{B_1} = 1/400$.

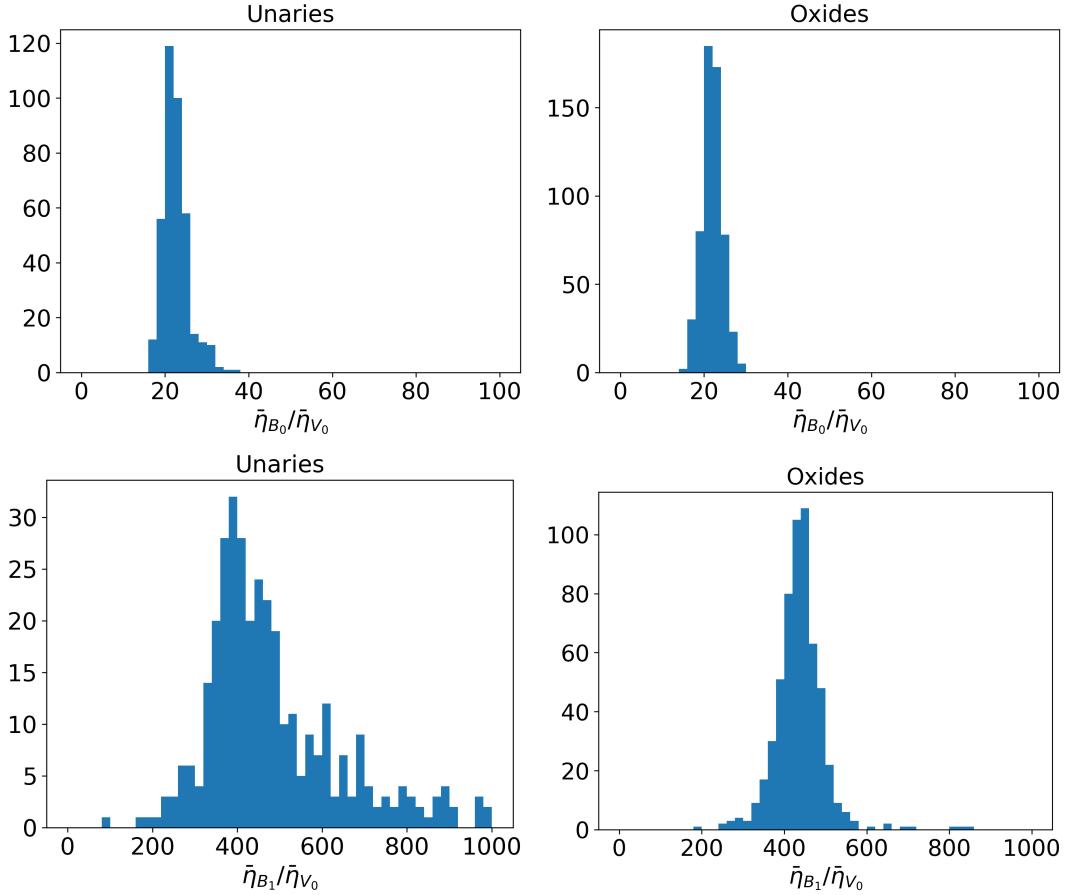


Figure S3.1. Histograms of the typical error ratios of B_0 vs. V_0 ($\bar{\eta}_{B_0}/\bar{\eta}_{V_0}$, top row) and of B_1 vs. V_0 ($\bar{\eta}_{B_1}/\bar{\eta}_{V_0}$, bottom row) for the unaries set (left column) and the oxides set (right column). The simulations were run for $N_s = 100$ random samples and a standard deviation on the error of the energy on the datapoints of $n_\sigma = 10^{-5}$ eV. The histograms indicate that the error of B_0 (B_1) is approximately 20 (400) times larger than the error on V_0 , justifying our choice of weights for the metric v .

- The peak positions are insensitive to the number of datapoints, as long as the total volume range is not modified. Instead, they change significantly if the volume range is changed. For instance, using a volume range of 90%–110% would result in values closer to 15 and 200 for the two ratios, respectively. This can also be intuitively explained: B_1 , for instance, is related to the non-parabolicity of the Birch–Murnaghan curve away from its minimum. If we consider a very small volume range, the curve will be very close to parabolic, and we therefore expect a large error on B_1 since the fit has very little information on the non-parabolic behavior. For larger volume ranges, the curve starts to deviate significantly from a parabola, thus providing more information to the fitting algorithm on the actual value of B_1 , in turn resulting into a smaller relative error on B_1 vs. B_0 or V_0 .
- The stability of the fit, especially on B_1 , is significantly affected by the choice of fitting algorithm. For instance, we realized that if one uses the `optimize.curve_fit` subroutine of SciPy (<https://www.scipy.org>), which is *not* the algorithm used in this work, the choice of the fitting starting point is very important, and we also observe that iterating the procedure a few times (using the results of the previous step as starting points for the next fit) improves the stability. Instead, the function used in this work (that is the same also used in Ref. 29, 30) is a non-iterative fitting algorithm that proves to be much more robust.

S4 Reference all-electron results for V_0 , B_0 and B_1

This section reports the complete reference dataset of the EOS parameters obtained with the two all-electron codes FLEUR and WIEN2k, and the absolute value of their percentage difference (that we indicate with η). Moreover, it reports the averaged parameters among them, that constitutes our reference average dataset presented in this manuscript. Data is divided in 10 tables, one for each crystal structure (4 unaries and 6 oxides). The agreement for V_0 is within 0.3% for all materials except Cs₂O₅ (0.323%), Fr₂O₅ (0.645%), Ra₂O₅ (0.333%), NeO₃ (0.302%) and RbO₃ (0.343%). Not surprisingly, these are 5 crystals with very small bulk moduli B_0 : it has been shown in Ref. 98 that the error in the volume scales inversely with the value of the bulk modulus.

Parameters are expressed per formula unit (see also SI Sec. S1). Note that, for X₂O₃ and for X₂O₅, the primitive cell has twice the number of atoms (10 and 14, respectively) than the number of atoms in the formula unit (5 and 7, respectively). This is reflected in a factor of 0.5 in the volumes reported in this table with respect to the volume of the unit cells in the input files available in Ref. 62.

Table S4.1. Table with all calculated EOS parameters for the FCC structures obtained with FLEUR and WIEN2k.

	FLEUR			WIEN2k			Abs. percentage difference [%]			Average set		
	V_0 [Å ³]	B_0 [eV/Å ³]	B_1	V_0 [Å ³]	B_0 [eV/Å ³]	B_1	$ \eta(V_0) $	$ \eta(B_0) $	$ \eta(B_1) $	V_0 [Å ³]	B_0 [eV/Å ³]	B_1
H	2.9651	0.6768	3.2596	2.9646	0.6764	3.2598	0.017	0.048	0.007	2.9648	0.6766	3.2597
He	17.7867	0.0053	6.208	17.7585	0.0054	6.6423	0.159	0.404	6.759	17.7726	0.0053	6.4251
Li	20.2246	0.0862	3.3304	20.2243	0.0862	3.3298	0.001	0.001	0.019	20.2245	0.0862	3.3301
Be	7.8705	0.7437	3.4867	7.8728	0.7437	3.471	0.029	0.006	0.449	7.8716	0.7437	3.4789
B	5.8886	1.6978	3.7481	5.8945	1.6993	3.7554	0.099	0.088	0.196	5.8916	1.6985	3.7517
C	7.3221	0.9364	3.5885	7.3212	0.9358	3.5906	0.012	0.07	0.057	7.3216	0.9361	3.5896
N	7.6017	1.1174	4.1748	7.6009	1.1161	4.1633	0.011	0.119	0.276	7.6013	1.1167	4.169
O	7.9988	0.8755	4.7639	7.9987	0.8745	4.7427	0.002	0.115	0.446	7.9988	0.875	4.7533
F	10.1454	0.3137	5.6468	10.1486	0.3134	5.6241	0.031	0.111	0.404	10.147	0.3135	5.6355
Ne	24.2855	0.0079	7.5394	24.3201	0.0077	7.2489	0.142	1.852	3.929	24.3028	0.0078	7.3941
Na	37.1012	0.0482	3.6866	37.0968	0.0481	3.6838	0.012	0.025	0.076	37.099	0.0482	3.6852
Mg	23.1275	0.2192	3.9959	23.123	0.2193	3.9966	0.02	0.065	0.019	23.1252	0.2193	3.9963
Al	16.4943	0.4838	4.6231	16.4964	0.4838	4.6233	0.012	0.01	0.004	16.4954	0.4838	4.6232
Si	14.4856	0.5171	4.3331	14.4788	0.5174	4.3282	0.048	0.049	0.113	14.4822	0.5173	4.3307
P	14.5664	0.5752	3.9579	14.5607	0.5759	4.0276	0.039	0.112	1.744	14.5636	0.5756	3.9927
S	15.88	0.4921	4.308	15.8814	0.4917	4.3089	0.009	0.089	0.023	15.8807	0.4919	4.3085
Cl	21.2849	0.1893	5.442	21.2915	0.189	5.439	0.031	0.169	0.055	21.2882	0.1891	5.4405
Ar	52.3124	0.0047	6.8685	52.2404	0.0047	7.4623	0.138	0.284	8.287	52.2764	0.0047	7.1654
K	74.0038	0.0221	3.7672	74.0051	0.0221	3.7758	0.002	0.093	0.229	74.0044	0.0221	3.7715
Ca	42.1888	0.1084	3.2754	42.1998	0.1084	3.2907	0.026	0.002	0.465	42.1943	0.1084	3.283
Sc	24.6877	0.3188	3.1588	24.686	0.3187	3.1581	0.007	0.029	0.023	24.6869	0.3187	3.1584
Ti	17.3952	0.6709	3.4556	17.3939	0.6707	3.4529	0.008	0.021	0.079	17.3946	0.6708	3.4543
V	13.9046	1.0993	3.8722	13.9053	1.0994	3.8706	0.005	0.01	0.041	13.905	1.0993	3.8714
Cr	11.8849	1.4816	4.1479	11.887	1.4825	4.1167	0.018	0.059	0.756	11.8859	1.4821	4.1323
Mn	10.7458	1.7527	4.4245	10.7485	1.7506	4.4042	0.026	0.119	0.461	10.7471	1.7516	4.4144
Fe	10.2594	1.7835	4.6058	10.261	1.7813	4.6159	0.015	0.126	0.218	10.2602	1.7824	4.6109
Co	10.3081	1.5919	4.7528	10.3087	1.5909	4.7557	0.006	0.065	0.061	10.3084	1.5914	4.7542
Ni	10.8354	1.2626	4.8979	10.8344	1.2609	4.8916	0.009	0.134	0.129	10.8349	1.2617	4.8947
Cu	11.9534	0.8805	5.052	11.951	0.8795	5.0594	0.02	0.109	0.145	11.9522	0.88	5.0557
Zn	15.1641	0.4322	5.2768	15.1599	0.4331	5.2674	0.027	0.212	0.177	15.162	0.4326	5.2721
Ga	18.9499	0.3032	4.8071	18.9431	0.3038	4.8034	0.035	0.165	0.078	18.9465	0.3035	4.8053
Ge	19.5823	0.3927	4.3284	19.5826	0.3929	4.3261	0.002	0.053	0.054	19.5825	0.3928	4.3273
As	19.3209	0.493	4.3003	19.3143	0.4928	4.3653	0.034	0.038	1.5	19.3176	0.4929	4.3328
Se	20.3778	0.4333	4.6567	20.3779	0.4333	4.6647	0.0	0.017	0.172	20.3779	0.4333	4.6607
Br	26.4146	0.1723	5.5508	26.4204	0.172	5.5512	0.022	0.137	0.006	26.4175	0.1722	5.551
Kr	66.0139	0.004	7.5013	66.0696	0.004	7.2722	0.084	1.635	3.101	66.0418	0.004	7.3867
Rb	91.4121	0.0172	3.7661	91.4428	0.0172	3.7729	0.034	0.114	0.18	91.4275	0.0172	3.7695
Sr	54.882	0.0728	3.2918	54.9025	0.0727	3.296	0.037	0.07	0.129	54.8923	0.0728	3.2939
Y	32.4691	0.243	3.2686	32.474	0.2428	3.2699	0.015	0.062	0.04	32.4715	0.2429	3.2692
Zr	23.2137	0.5626	3.2904	23.2131	0.5624	3.288	0.002	0.037	0.072	23.2134	0.5625	3.2892
Nb	18.7683	1.0213	3.656	18.7673	1.021	3.6527	0.005	0.027	0.09	18.7678	1.0211	3.6544
Mo	16.0351	1.4875	4.0608	16.0351	1.4874	4.0598	0.0	0.013	0.024	16.0351	1.4875	4.0603
Tc	14.5123	1.8563	4.4992	14.513	1.8564	4.4999	0.005	0.006	0.015	14.5127	1.8564	4.4996
Ru	13.8362	1.9096	4.8614	13.837	1.9097	4.8617	0.006	0.001	0.006	13.8366	1.9096	4.8616
Rh	14.0519	1.6017	5.1987	14.049	1.6037	5.1824	0.021	0.122	0.313	14.0505	1.6027	5.1906
Pd	15.3284	1.0486	5.5494	15.3224	1.0488	5.5242	0.039	0.019	0.454	15.3254	1.0487	5.5368
Ag	17.8411	0.5681	5.8386	17.836	0.5676	5.8264	0.029	0.101	0.209	17.8386	0.5678	5.8325
Cd	22.8435	0.261	5.9936	22.839	0.2608	5.9998	0.02	0.088	0.103	22.8413	0.2609	5.9967

Table S4.10. (continued) Table with all calculated EOS parameters for the XO_3 structures obtained with FLEUR and WIEN2k.

	FLEUR			WIEN2k			Abs. percentage difference [%]			Average set		
	V_0 [\AA^3]	B_0 [eV/ \AA^3]	B_1	V_0 [\AA^3]	B_0 [eV/ \AA^3]	B_1	$ \eta(V_0) $	$ \eta(B_0) $	$ \eta(B_1) $	V_0 [\AA^3]	B_0 [eV/ \AA^3]	B_1
Ra	87.7112	0.2318	8.0555	87.7466	0.2294	8.1243	0.04	1.042	0.85	87.7289	0.2306	8.0899
Ac	91.3919	0.2075	5.8257	91.492	0.2046	5.6485	0.11	1.398	3.09	91.442	0.206	5.7371
Th	84.9418	0.3566	5.4148	84.985	0.353	5.3017	0.051	1.021	2.11	84.9634	0.3548	5.3583
Pa	77.7028	0.5786	5.6105	77.7052	0.577	5.6618	0.003	0.275	0.91	77.704	0.5778	5.6362
U	72.2283	0.929	4.2343	72.2313	0.9266	4.2328	0.004	0.259	0.034	72.2298	0.9278	4.2335
Np	70.2882	0.8987	4.1862	70.2877	0.8955	4.2317	0.001	0.362	1.08	70.288	0.8971	4.209
Pu	68.8403	0.8615	4.2029	68.8377	0.8588	4.2618	0.004	0.318	1.391	68.839	0.8602	4.2323
Am	67.8603	0.8185	4.1875	67.8468	0.8165	4.3118	0.02	0.245	2.926	67.8536	0.8175	4.2497

S5 Simulation parameters for FLEUR and WIEN2k

The AE codes used for producing the reference results in this work implement the (linearized) augmented-plane-wave + local orbitals ((L)APW+LO) method. This method is named after a family of different, but related basis sets, each having different characteristics and demands on cutoff parameters. It is based on a partitioning of the unit cell into muffin-tin (MT) spheres at the atom positions and an interstitial region in between the spheres. The MT sphere radii hereby have to be adapted to each structure to avoid overlapping spheres, but are typically chosen to be nearly touching, because large spheres reduce demands on the LAPW basis set cutoff parameter. The exact choice of the radii typically differs between (L)APW+LO codes.

The choice of the core/valence electron separation is related to the MT sphere radii. To avoid instabilities in the calculations and to obtain precise results, the extent of the core-electron states beyond the MT sphere boundary has to be limited. The AE codes employed here share the same core/valence separation for many structures, but differ in this choice for a considerable amount of other structures. Taking into account the overall excellent agreement between the results from the two codes it can be deduced that there is no significant dependence of the results on the differing description of physics for core and valence electrons, as long as the core/valence separation is within a reasonable range.

The valence electrons are represented by the (L)APW+LO basis chosen by the respective code. The different possible choices of basis sets nevertheless share a common set of parameters with which they are specified. These are cutoff parameters for the basis set size and the angular momentum expansion in the MT spheres, as well as energy parameters defining the linearization centers for each atom. Additionally, the local orbital setup has to be defined for each MT sphere. The two AE codes employed in this work make use of different kinds of (L)APW+LO basis sets and their choices for the setup aspects discussed here differ strongly. A sketch on the parameter setup recipes is provided in the following two subsections. A detailed list of the setup parameter choices for each structure is available in the supplementary data, as referenced in the respective sections below.

S5.1 FLEUR

The FLEUR code^{31,32} is an open-source implementation of the all-electron full-potential linearized augmented-plane-wave (FLAPW) method^{99,100}. The calculations with this code make use of a conventional LAPW basis in combination with local orbitals (LOs) of different types to describe semicore states¹⁰¹ and to eliminate the linearization error for the valence states^{102–104}. The employed setup profile defines global parameters, identical for all calculations and element-specific parameters. Further parameters are automatically adapted to each investigated structure.

In the context of this profile, global parameters are the reciprocal LAPW basis set cutoff parameter $K_{\max} = 5.0 \text{ } a_0^{-1}$ and the plane-wave cutoff parameters $G_{\max} = G_{\max, \text{XC}} = 25.0 \text{ } a_0^{-1}$, where a_0 is the Bohr radius. The latter cutoff parameter covers the expansions of the plane-wave part of the density and the potential, the interstitial-region indicator function, and the exchange-correlation contribution to the potential. The Fermi–Dirac smearing and the k-point density are also fixed to the common choices of this work. Element-specific setup aspects include the core-valence separation of the electron states and the LO setup. The radii of the MT spheres centered on atom α , R_{MT}^{α} , are adapted to the smallest unit cell within an equation-of-states (EOS) workflow. For this, element-specific initial MT sphere radii are expanded to cover up to about 92% of the distance between the atoms, with a limit of a maximal MT radius of $2.66 \text{ } a_0$. The procedure also implies an adaption of the angular momentum cutoffs in the spheres to $l_{\max}^{\alpha} \approx K_{\max} R_{\text{MT}}^{\alpha}$. A detailed description of the different parameters is available in the FLEUR user guide^{31,32}.

The parameter profile is the result of an iterative refinement process aiming at precision and stability on the basis of the structures investigated in this work, i.e., structures with a wide range of neighboring atom distances. However, it is not designed to provide an absolute convergence of the total energy, which may still be affected by structure-adapted numerical parameters like the MT radii. The determination of these parameters for each structure on the basis of the smallest unit cell solves this issue for the EOS workflow. Comparing the total-energy results from the EOS workflows for different structures may also yield reasonable numbers, but these do not reflect the precision capabilities of the code when used in an adequate way to perform such a comparison, see also discussion in SI Sec. S18.

With the exception of the k-point integration-mesh generation, the used parameter profile is implemented in the openly available releases of FLEUR starting with the MaX-R6.1 release. The results presented here have been obtained using the development version as of 2022/03/31. The profile is employed by invoking the FLEUR input generator with the command line option “-profile oxides_validation”. Beyond using this profile, the AiiDA common workflows package³⁶ protocol “verification-pbe-v1” in combination with AiiDA-FLEUR^{105,106} also sets the k-point integration mesh.

The resulting parametrization for each structure is discussed in the all-electron-setups folder of the supplementary data available in Ref. 62, in particular in the files

- setup-oxides-verification-PBE-v1-fleur.json
- setup-unaries-verification-PBE-v1-fleur.json

that are documented in the file `all-electron-data.md`.

In an ongoing effort the data from this work is also used to define further FLEUR parameter profiles for different precision levels and computational effort. This is done by reducing in a controlled way cutoff parameters and changing other aspects of the setup and relating the corresponding results to those presented here.

S5.2 WIEN2k

The WIEN2k calculations^{33,34} employ the (linearized) augmented plane wave + local orbitals ((L)APW+lo) method¹⁰⁷ with additional local orbitals¹⁰¹ (LOs) for states with an energy 1 Ry below the Fermi energy (semicore states) and high-derivative LOs (HDLOs)¹⁰⁸ for all “chemical” l values (except when there is already a semicore LO). The APW+lo basis set is used for all “chemical” l values (s, p, d or f - depending on the atom), while LAPW is used for higher angular momentum up to $l_{\max} = 10$.

The WIEN2k calculations have been initialized (see the WIEN2k users guide³³) for the smallest volume of each case using:

```
init_lapw -b -prec 3 -nokshift -fermits 0.0045 -red 3 -numk -1 0.0317506
```

For subsequent volumes we use:

```
init_lapw -b -prec 3 -nokshift -fermits 0.0045 -red Element:RMT -numk 0 kx ky kz -fft ix iy iz
```

where `Element:RMT`, `kx ky kz`, and `ix iy iz` are inserted from the output of the first volume to ensure identical parameters.

This high-precision setup limits the maximal atomic sphere radius R_{MT}^{α} to $2.35 a_0$, but otherwise sets the sphere sizes automatically depending on nearest neighbor distances and type of atom (largest for f elements, intermediate for d elements and smallest for sp elements). Also the plane wave cutoff $R_{\text{MT}, \min}^{\alpha} K_{\max}$ is set automatically depending on the type of atom and the smallest atomic sphere radius $R_{\text{MT}, \min}^{\alpha}$ and varies from 7.08 (H_2O_5) to 11. All states with an energy above -6 Ry or with a charge density of more than $0.01 e^-$ outside the atomic sphere are considered as valence states and treated scalar-relativistically, while lower energy states are considered as core and solved numerically with a radial symmetric Dirac equation. Note that with this choice the definition of core states for an element may change depending on its R_{MT}^{α} and this makes the calculation of formation energies in certain cases unrealistic as discussed also in SI Sec. S18. An SCF cycle was considered converged when both the change in total energy was less than 10^{-6} Ry and the change in the electron charge density within R_{MT}^{α} was less than $10^{-6} e$.

The charge density and potential inside spheres is expanded into lattice harmonics up to $L_{\max} = 6$ and for the non-spherical Hamiltonian matrix elements the angular momentum of the wave functions is restricted to $l_{\max}^{hs} = 8$. In the interstitial region the density/potential is expanded into a Fourier series with cutoff parameter $G_{\max} = 25 a_0^{-1}$ (except for alkali metals, noble gases, and Hg, where $G_{\max} = 40 a_0^{-1}$).

The resulting parametrization for each structure is discussed in the `all-electron-setups` folder of the supplementary data available in Ref. 62, in particular in the files

- `setup-oxides-verification-PBE-v1-wien2k.json`
- `setup-unaries-verification-PBE-v1-wien2k.json`

that are documented in the file `all-electron-data.md`.

S6 Dependence of the metrics on the size of the simulation cell and on bond stiffness

The equation of state is often expressed in terms of the absolute energy E and volume V of the simulation cell, but other times quantities "per-formula-unit" or "per-atom" are considered. We want here to demonstrate that the new metrics ε and v , introduced here, are intrinsic quantities, i.e., they are independent of the simulation cell size (while the original Δ metric is extensive).

Dependence on the number of atoms in the simulation cell: We first show that the original metric Δ is an extensive quantity that depends linearly on the number of atoms (or, equivalently, on the volume) of the simulation cell.

Let us consider a supercell where the number of atoms is increased by a factor of C . The volume and the total energy (assuming $E_0 = 0$) will scale accordingly as

$$E' = CE, \quad V' = CV. \quad (\text{S5})$$

The integrand in Eq. (3) for the Δ metric will scale as

$$[E'_a(V') - E'_b(V')]^2 dV' = C^3 [E_a(V) - E_b(V)]^2 dV \quad (\text{S6})$$

and the denominator as

$$V'_M - V'_m = C(V_M - V_m). \quad (\text{S7})$$

The effect of increasing the number of atoms on the Δ metric is

$$\begin{aligned} \Delta'(a, b) &= \sqrt{\frac{1}{V'_M - V'_m} \int_{V'_m}^{V'_M} [E'_a(V') - E'_b(V')]^2 dV'} = \sqrt{\frac{C^3}{C(V_M - V_m)} \int_{V_m}^{V_M} [E_a(V) - E_b(V)]^2 dV} \\ &= C \sqrt{\frac{1}{V_M - V_m} \int_{V_m}^{V_M} [E_a(V) - E_b(V)]^2 dV} = C\Delta(a, b). \end{aligned} \quad (\text{S8})$$

Thus, Δ scales *linearly* with the number of atoms in the simulation cell. This shortcoming is typically addressed by computing the Δ metric renormalized per atom.

For the ε metric we proceed in the same manner to demonstrate, instead, its independence from the simulation cell size.

$$\begin{aligned} \varepsilon'(a, b) &= \frac{\int_{V'_m}^{V'_M} [E'_a(V') - E'_b(V')]^2 dV'}{\sqrt{\left(\int_{V'_m}^{V'_M} [E'_a(V') - \langle E'_a \rangle]^2 dV'\right) \left(\int_{V'_m}^{V'_M} [E'_b(V') - \langle E'_b \rangle]^2 dV'\right)}} \\ &= \frac{C^3 \int_{V_m}^{V_M} [E_a(V) - E_b(V)]^2 dV}{\sqrt{\left(C^3 \int_{V_m}^{V_M} [E_a(V) - \langle E_a \rangle]^2 dV\right) \left(C^3 \int_{V_m}^{V_M} [E_b(V) - \langle E_b \rangle]^2 dV\right)}} \\ &= \frac{\int_{V_m}^{V_M} [E_a(V) - E_b(V)]^2 dV}{\sqrt{\left(\int_{V_m}^{V_M} [E_a(V) - \langle E_a \rangle]^2 dV\right) \left(\int_{V_m}^{V_M} [E_b(V) - \langle E_b \rangle]^2 dV\right)}} = \varepsilon(a, b). \end{aligned} \quad (\text{S9})$$

Since C cancels out, we have proven that ε is independent of the number of atoms in a simulation cell considered.

Analogously, it is also easy to see that v is also an intrinsic quantity, independent of the number of atoms in the simulation cell: indeed, v is defined as a function of the relative errors of the parameters V_0 , B_0 and B_1 , that are all intrinsic quantities.

We stress that this means that, while for Δ is recommended to report it normalized (e.g., per atom, or per formula unit), ε and v should *not* be normalized.

Sensitivity to the value of the bulk modulus: Let us compare the metrics obtained comparing results for two different materials. We assume that the first material has bulk modulus B_0 and the second is identical except for its bulk modulus, that is scaled by factor of C , i.e., $B'_0 = CB_0$. We assume that there is no other difference between the two materials (same V_0 and B_1) and we are considering the same simulation volume (or number of atoms in the simulation cell). The total energy of the second material will then scale as

$$E' = CE \quad (\text{S10})$$

according to Eq. (1) and assuming $E_0 = 0$ (minimum energy of both materials have been shifted to zero). The integrand in Eq. (3) for the Δ metric scales as

$$[E'_a(V) - E'_b(V)]^2 dV = C^2 [E_a(V) - E_b(V)]^2 dV. \quad (\text{S11})$$

Following similar steps as for the number of atoms above, we arrive to the conclusion that Δ scales also *linearly* with the bulk modulus, while the ε metric is insensitive to the it. Similarly, also v is insensitive to the reference value of the bulk modulus of the material, since it only depends on the relative change of the bulk modulus B_0 between the two systems, that does not depend on the factor C .

We stress that this fact does not mean that ε or v will not capture a difference between bulk moduli when comparing two computational approaches a and b . It means, instead, that two datasets with a similar discrepancy in the bulk moduli (say 2%) will result in the same ε or v irrespective of overall stiffness of their chemical bonds (i.e., their bulk modulus). We highlight that this shortcoming of the Δ metric was already recognized in the literature and addressed by defining a modified metric Δ_1 in Ref. 63.

S7 Sensitivity of Δ , ε and v to perturbations of the EOS parameters and choice of thresholds for excellent and good agreement

Any of the three metrics Δ , ε or v expresses the difference between two EOS curves by a single number. It is not *a priori* obvious, however, when those numbers can be considered small or large, and which features of the EOS have the largest impact on the value. In this section, we address these points.

In SI Fig. S7.1 we compare the EOS of a hypothetical material with the EOS obtained after four different perturbations of the material (see caption for details). The values for Δ , ε and v that express the difference between the original and perturbed EOS are listed for each case. Analyzing these results allows to associate typical orders of magnitude to each of the metrics, and quantify their variation with respect to changes in V_0 , B_0 , and B_1 .

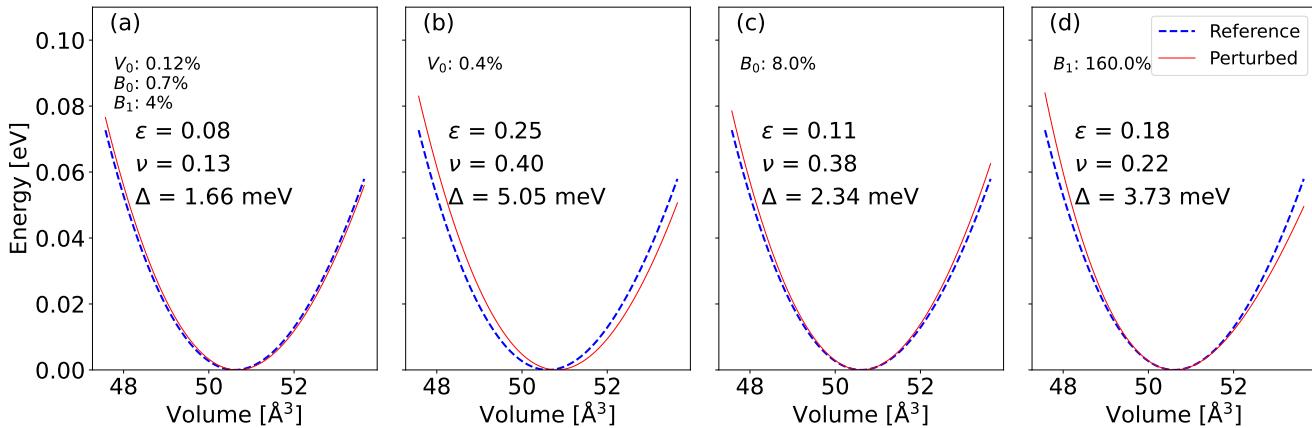


Figure S7.1. EOS for a hypothetical material with $V_0 = 50.61 \text{ \AA}^3$ per formula unit, $B_0 = 0.71 \text{ eV}/\text{\AA}^3$, and $B_1 = 4.67$ (“reference” curve, dashed blue), compared with the EOS where a perturbation has been applied to some of the EOS-defining parameters (“perturbed” curve, solid red). Each panel reports the resulting value for the three metrics ε , v and Δ obtained comparing the two curves. The perturbed parameters and the magnitude of the perturbation are also indicated in each panel. In particular, in panel (a) a perturbation of 0.12% is applied to V_0 , of 0.7% to B_0 and of 4% to B_1 . Panels (b), (c) and (d) present the cases when a perturbation is applied independently to V_0 (0.4%), B_0 (8%) or B_1 (160%), respectively, while the other parameters are kept unchanged.

The hypothetical reference material that is used in SI Fig. S7.1 has $V_0 = 50.61 \text{ \AA}^3$ per formula unit, $B_0 = 0.71 \text{ eV}/\text{\AA}^3$, and $B_1 = 4.67$. These values are obtained as averages over the entire crystals set and thus represent a hypothetical “average” EOS.

In panel (a), a perturbation is applied to all three parameters, namely 0.12% to V_0 , of 0.7% to B_0 and of 4% to B_1 . These values are twice the standard deviations of the discrepancies between the two AE codes in our reference dataset, see Fig. 1 in the main text. Since the two EOS curves are almost undistinguishable, this result highlights the high level of agreement between our two AE codes. Based on this observation, we define a qualitative range of $\varepsilon \lesssim 0.06$ or $v \lesssim 0.1$ for which we consider two codes display excellent agreement. The threshold of $\varepsilon = 0.06$ (approximately) corresponds to a determination coefficient $R^2 \approx 1 - \varepsilon^2 = 0.9964$ when one EOS curve is considered as a fit to the other.

In panels (b), (c) and (d), instead, a larger perturbation is applied to only one of the three parameters V_0 , B_0 or B_1 , respectively. The perturbation to V_0 is chosen as 0.4%, and the magnitude of the perturbations for B_0 and B_1 is scaled following the inverse ratios 1/20/400 of the weights for the v metric (see SI Sec. S3). This results in visually similar discrepancies between the two curves. This is an expected result, and is another way to interpret the results discussed in SI Sec. S3 for the weights of v . Indeed, those weights were obtained by inferring the error propagated on the fitted parameters from a given amount of random noise on the datasets; the inverse weights can be conversely interpreted, intuitively, as the relative magnitude of the perturbation to each of the parameters required to induce similar changes to the EOS curve. These panels help us make a number of observations:

- All metrics (Δ and ε intrinsically, and v by explicit definition of the weights) give a much stronger importance to changes of the equilibrium volume V_0 than to changes of the other parameters. This is a positive feature of the metrics, as the EOS shape is mostly sensitive to V_0 as well. The metric v has the additional advantage, as already discussed, that weights can be tuned to give more importance to other parameters, if an application requires it.
- With our definitions of ε and v , the two metric often return similar values for a given pair of EOS curves, with ε typically

slightly smaller (in the SI Sec. S8, we actually identify an approximate proportionality ratio between the two, valid for small values of the metrics).

- The values of the metrics on these last three panels allow us to define threshold values for noticeable, but still relatively small, changes between two EOS curves: $\varepsilon \lesssim 0.2$ or $v \lesssim 0.33$. Therefore, we take the ranges of $0.06 < \varepsilon \lesssim 0.20$ or $0.10 < v \lesssim 0.33$ as the signature for good (but not excellent) agreement between two codes. The upper end of this range $\varepsilon = 0.20$ (approximately) corresponds to a determination coefficient $R^2 \approx 1 - \varepsilon^2 = 0.96$ when one EOS curve is considered as a fit to the other.
- We can assign an intuitive meaning to the metric v . If two EOS only differ in the equilibrium volume, its numerical value corresponds to the percentage error on the equilibrium volume between the two curves. If also B_0 and B_1 change, then v will also take into account the discrepancies on these two parameters, rescaled so that similar contributions to v result to similar quantitative changes to the EOS curve (in the volume range of interest, $\pm 6\%$ in this work).
- We note that in panels (b), (c) and (d) the value of v is not exactly 0.4 as one might naively expect, because the perturbation that we apply refers to the reference curve, but the v metric is defined in a symmetric way, with percentage differences with respect to the average of the two curves.
- SI Fig. S7.2, finally, illustrates the clear disagreement between EOS curves when $\varepsilon \geq 1.0$ or $v \geq 1.65$ (these values are used as upper limit for the colorbars of the figures in SI Sec. S9). As a note, $\varepsilon = 1$ is an estimator for the situation where the coefficient of determination $R^2 \approx 1 - \varepsilon^2$ starts to be negative (even if the approximation $R^2 \approx 1 - \varepsilon^2$ does not hold exactly anymore for such large values of ε). We highlight that a negative R^2 value indicates that a horizontal line at the average value of the data provides a better fit than the actual fit function. It is clear from SI Fig. S7.2 that in such cases there is no agreement at all between the results of two codes. Therefore, when $\varepsilon > 1.0$ or $v > 1.65$, two codes are said to be clearly different in SI Sec. S9.

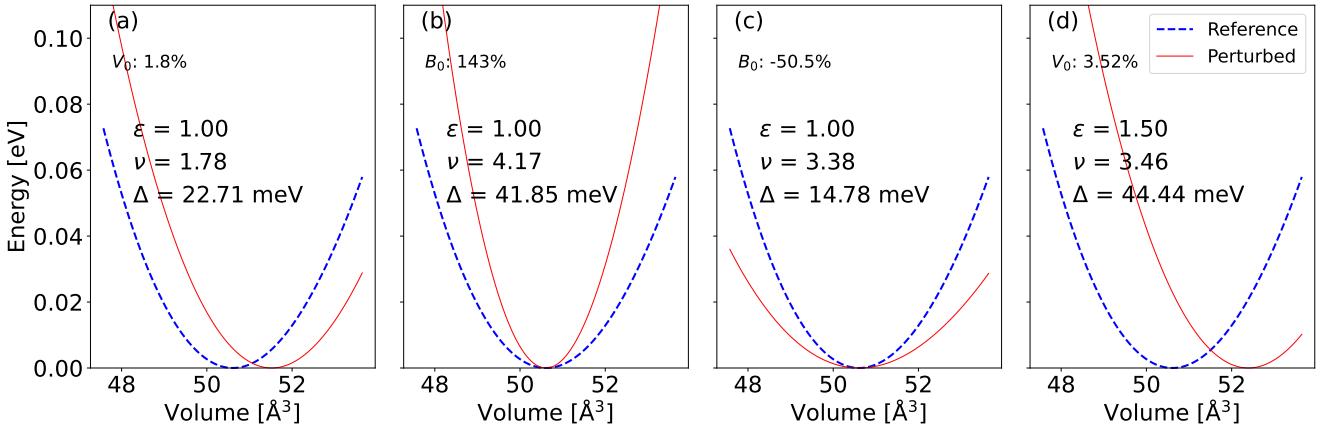


Figure S7.2. EOS for a hypothetical material with $V_0 = 50.61 \text{ \AA}^3$ per formula unit, $B_0 = 0.71 \text{ eV}/\text{\AA}^3$, and $B_1 = 4.67$ (“reference” curve, dashed blue), compared with the EOS where a perturbation has been applied to some of the EOS-defining parameters (“perturbed” curve, solid red). With respect to SI Fig. S7.1, we highlight here typical values for the metrics for large changes of the parameters that make the EOS curves clearly different. Each panel reports the resulting value for the three metrics ε , v and Δ obtained comparing the two curves. The perturbed parameters and the magnitude of the perturbation are also indicated in each panel. In particular, in panel (a) a perturbation is applied to V_0 to obtain a value of $\varepsilon = 1$; panels (b) and (c) apply a (positive and negative, respectively) perturbation to B_0 resulting in $\varepsilon = 1$, and finally in panel (d) an even larger perturbation is applied to V_0 to obtain a value of $\varepsilon = 1.5$.

S8 Mutual correlation between the metrics

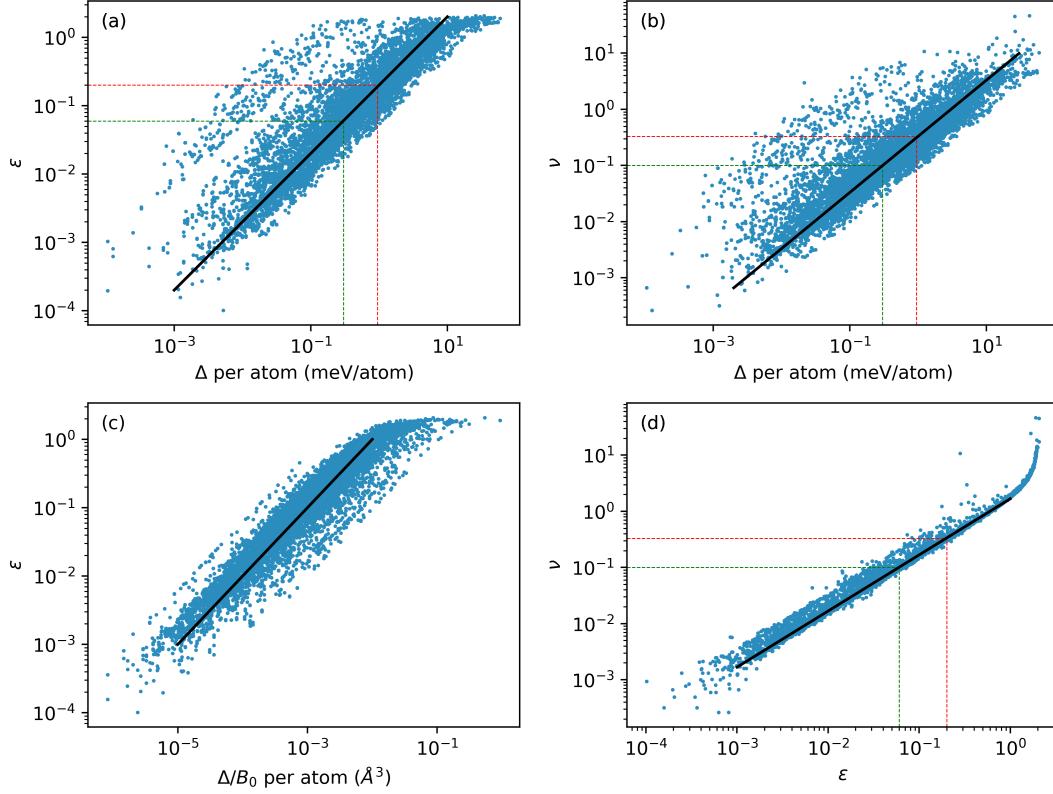


Figure S8.1. Cross-correlation between the Δ , ε , and v metrics for the entire data (unaries and oxides) presented in the main text. Black lines are helpers to indicate a linear relation between the quantities (slope of 1 in a log-log plot). Dashed lines on ε and v axes show the “excellent” (green) and “good” (red) agreement thresholds recommended in the main text. From the plots, these correspond to values of Δ of approximately 0.3 and 0.95 meV/atom, respectively.

In order to assess the correlation between the various metrics, we present in SI Fig. S8.1 for all codes mentioned in the main text their Δ , ε or v with respect to the reference EOS, for each of the unaries and oxides. These Δ , ε or v are plotted against one of the other metrics. The results support the presence of an approximately linear correlation between the three metrics (Δ , ε , and v). However, there is more scattering in the correlation between ε and Δ (and, similarly, between v and Δ), while ε and v agree more consistently on a global scale.

We first note that the scattering between ε and Δ can be reduced by normalizing Δ by the bulk modulus, similar to the Δ_1 metric introduced in Ref. 63, as shown in SI Fig. S8.1(c). We then observe that the new metrics ε and v are instead almost linearly correlated when their values are $\lesssim 1$; for larger discrepancies, the values of v tend to grow faster than the values of ε , i.e., ε becomes relatively less sensitive to further small changes to the EOS curves if they are already significantly different.

This almost linear correlation can be justified with some approximations. Let us consider the simple case of two parabolic EOS curves with the same B_0 , differing only in the equilibrium volume V_0 . This is a valid approximation, since we discussed above that both metrics are mostly sensitive to changes of V_0 rather than B_0 or B_1 . If we call $2V_R = V_M - V_m$ the volume range for the integration in ε (see also Eq. (4) in the main text for the definition of V_m and V_M) and \tilde{V} the average volume of the two curves (with actual minima for $V_0 = \tilde{V} \pm \Delta V$), we obtain $v = 200 \frac{\Delta V}{\tilde{V}}$. We now consider the limit in which $\Delta V \ll V_R$ (i.e., of a small discrepancy of the two curves in the volume range of interest, corresponding to the regime of small ε and v in which our data show an almost linear relation between the two metrics).

It is then straightforward to show that $\varepsilon \approx 2\sqrt{15} \frac{\Delta V}{V_R}$. Indeed, writing the two curves as $E_{1,2}(V) = A(V - \tilde{V} \pm \Delta V)^2$, with A an appropriate coefficient (the same for both curves with our assumptions of same bulk modulus), we get that the integral in the

numerator of ε , $\langle [E_1(V) - E_2(V)]^2 \rangle$, is given by:

$$\frac{1}{2V_R} \int_{\tilde{V}-V_R}^{\tilde{V}+V_R} [A(V - \tilde{V} - \Delta V)^2 - A(V - \tilde{V} + \Delta V)^2]^2 dV = \frac{1}{V_R} \int_{\tilde{V}}^{\tilde{V}+V_R} [A(V - \tilde{V} - \Delta V)^2 - A(V - \tilde{V} + \Delta V)^2]^2 dV \quad (\text{S12})$$

$$= \frac{1}{V_R} \int_{\tilde{V}}^{\tilde{V}+V_R} [4A(V - \tilde{V})\Delta V]^2 dV = \frac{16A^2\Delta V^2 V_R^2}{3}. \quad (\text{S13})$$

Similarly, we can obtain (using our assumption $\Delta V \ll V_R$) that

$$\langle E_1(V) \rangle = \frac{1}{2V_R} \int_{\tilde{V}-V_R}^{\tilde{V}+V_R} A(V - \tilde{V} - \Delta V)^2 dV \approx \frac{1}{V_R} \int_{\tilde{V}}^{\tilde{V}+V_R} A(V - \tilde{V})^2 dV = \frac{AV_R^2}{3}, \quad (\text{S14})$$

and $\langle E_2(V) \rangle = \langle E_1(V) \rangle$. Using similar steps, one can also obtain $\langle [E_1(V) - \langle E_1 \rangle]^2 \rangle = \langle [E_2(V) - \langle E_2 \rangle]^2 \rangle \approx \frac{4}{45}A^2V_R^4$. Putting all results together, we obtain the final result $\varepsilon \approx 2\sqrt{15}\frac{\Delta V}{V_R}$.

Finally, considering our choice of a $\pm 6\%$ volume range ($V_R \approx 0.06\tilde{V}$) gives $\varepsilon \approx \frac{100}{3}\sqrt{15}\frac{\Delta V}{\tilde{V}}$. Therefore, the two metrics are linearly dependent with the ratio of

$$\left. \frac{v}{\varepsilon} \right|_{\Delta V} = \frac{6}{\sqrt{15}} \approx 1.55, \quad (\text{S15})$$

which also matches well with the ratio of v and ε calculated for a specific value of V_0 perturbation shown in SI Fig. S7.1(b).

A similar analysis can be performed for two parabolic **EOS** curves with the same V_0 , but differing in B_0 . In this case, $v = 10 \frac{\Delta B_0}{B_0}$, where \tilde{B}_0 is the average B_0 of the two curves and $\pm \Delta B_0$ is the difference from the average value for the two curves. It can be shown, similarly to the case of differing V_0 , that in this case $\varepsilon \approx 3 \frac{\Delta B_0}{B_0}$. The resulting ratio of the two metrics now becomes

$$\left. \frac{v}{\varepsilon} \right|_{\Delta B_0} \approx \frac{10}{3} \approx 3.33, \quad (\text{S16})$$

which also matches well with the ratio of v and ε in SI Fig. S7.1(c). The difference by a factor of 2 between (S15) and (S16) shows that, compared to ε , v is 2 times more sensitive to B_0 variation relative to V_0 variation.

To assess the typical ratio of v and ε in our dataset, in SI Fig. S8.2 we show a cross-correlation plot between v and ε for the entire dataset of calculated crystals. One can see that for smaller values of the metrics ($v < 1$), the relation is approximately linear, and a numerical fit gives the slope of $\frac{v}{\varepsilon} \approx 1.65$, which is close to 1.55 found in Eq. (S15). This is expected, as both metrics give larger weight to V_0 errors compared to B_0 or B_1 errors, and in our dataset errors on V_0 (once rescaled with these weights) dominate over B_0 and B_1 errors. This is also visible, for instance, from the histograms of Fig. 1 in the main text, where the ratios of the standard deviation of the histograms on V_0 , B_0 and B_1 do not follow the $1, \frac{1}{20}, \frac{1}{400}$ ratio of v .

The cross correlation plots allow to establish a data-driven relation between different metrics. For instance, the two boundaries $\varepsilon = 0.06$ and 0.2 selected in this project as “excellent” and “good” agreement between two **EOS**s, and the corresponding thresholds for v (0.10 and 0.33) have been chosen according to these cross correlations. In addition, we can see from SI Fig. S8.1a that these thresholds translate, for Δ , approximately $\Delta \approx 0.3$ and 0.95 meV/atom. This result is comparable with the average Δ across **AE** codes $\langle \Delta \rangle = 0.5 - 0.9$ meV/atom obtained in the earlier benchmark^{29,30} for monoelemental solids, and is consistent with the conclusion obtained there that $\Delta = 1$ meV/atom is a threshold under which one can speak about good agreement (for materials with bulk moduli that are not particularly small). This consistency between former and present benchmarks is about the metric; we refer to SI Sec. S15 for a demonstration of consistency regarding the crystal set, and an illustration of the added value of the present benchmark study.

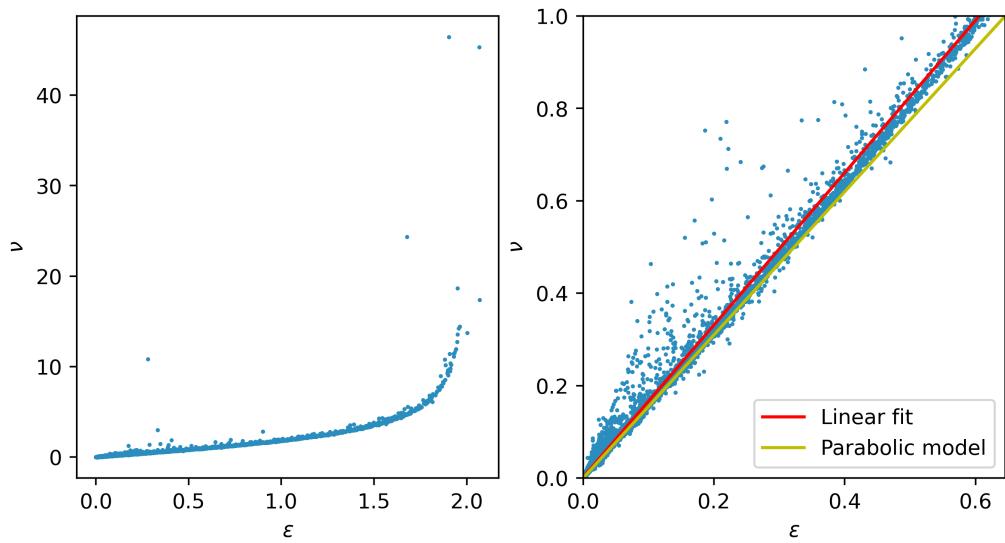


Figure S8.2. Cross-correlation between the ε and ν metrics on a linear-axes plot. Left panel: all data; right panel: zoom on the data with $\nu < 1$ where the relation is approximately linear, together with a linear fit of the data (red) resulting in a slope of ≈ 1.65 , and a line with a slope ≈ 1.55 (yellow curve) taken from the parabolic model.

S9 Detailed results for all computational approaches

In this section, we report the comparison of each of the computational approaches considered in the main text with the average all-electron reference dataset, using both metrics ε and v . For each metric, the same colorbar is used for all approaches, based upon the ranges of agreement identified in SI Sec. S7 (in addition, the ratio of the threshold values for ε and v is in agreement with their approximate linear relationship, see SI Sec. S8):

- “excellent agreement” ($\varepsilon \leq 0.06, v \leq 0.10$): a very dark shade of blue (not evolving very much over this narrow interval);
- “good agreement” ($0.06 < \varepsilon \leq 0.20, 0.10 < v \leq 0.33$): color evolving from a dark shade of blue to yellow as the values of ε or v increase;
- threshold for good agreement ($\varepsilon = 0.20, v = 0.33$): yellow;
- “noticeably different” ($0.20 < \varepsilon \leq 1.0, 0.33 < v \leq 1.65$): color evolving from yellow to red as the values of ε or v increase;
- “clearly different” ($\varepsilon > 1.0, v > 1.65$): one uniform darker shade of red, regardless of the value.

Crystals that were not computed are left in white. The caption of every plot mentions the number of crystals belonging to each of these categories. The results for all the codes are shown in SI Figs. S9.1 to S9.11.

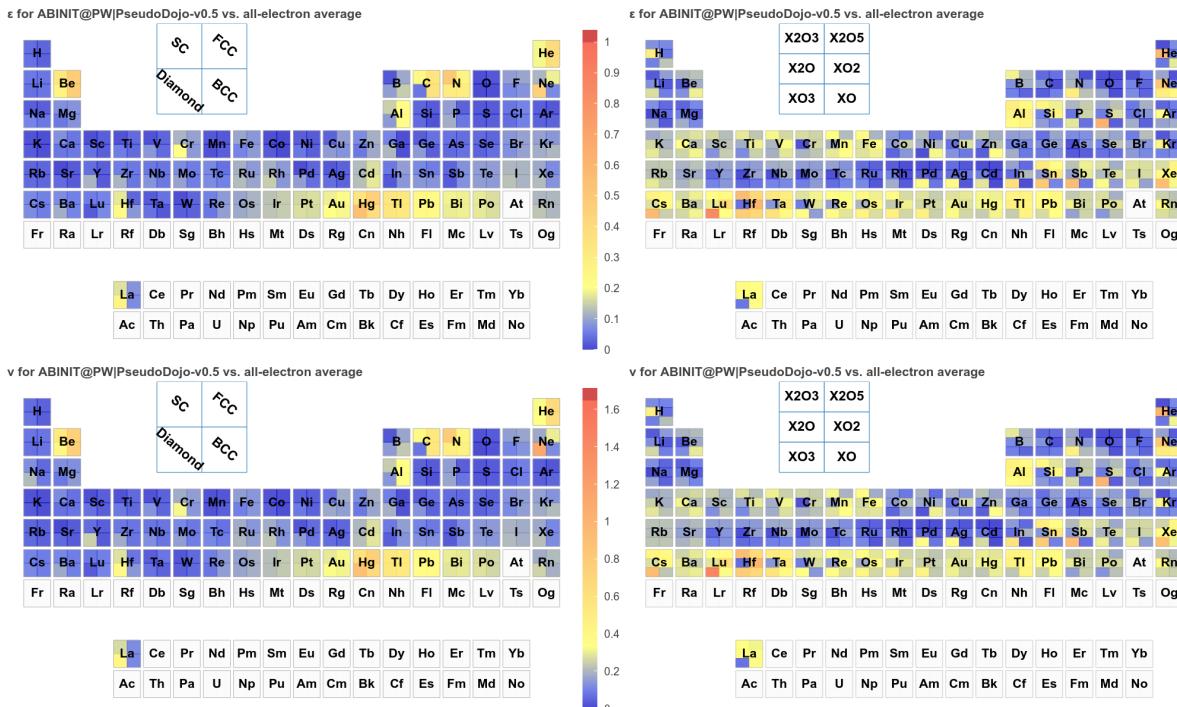


Figure S9.1. Value of the comparison metrics ε (top) and v (bottom) for ABINIT@PW|PseudoDojo-v0.5 with respect to the average all-electron reference dataset. Left panels: unaries; right panels: oxides. 720 out of 960 crystals were calculated. The number of crystals that land in the excellent, good, noticeably different, and clearly different agreement ranges for the ε metric are 232, 377, 111, 0, respectively. For the v metric, they are 244, 378, 98, 0, respectively.

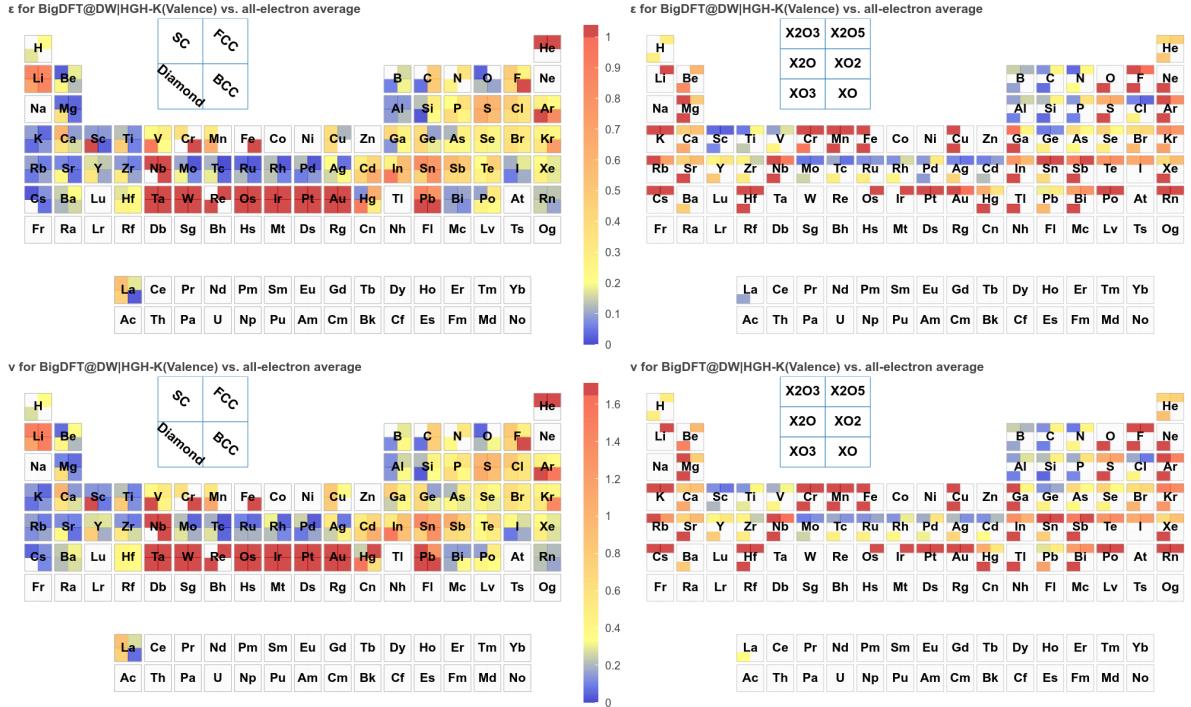


Figure S9.2. Value of the comparison metrics ϵ (top) and ν (bottom) for BigDFT@DW| HGH-K (Valence) with respect to the average all-electron reference dataset. Left panels: unaries; right panels: oxides. 402 out of 960 crystals were calculated. The number of crystals that land in the excellent, good, noticeably different, and clearly different agreement ranges for the ϵ metric are 45, 97, 173, 87, respectively. For the ν metric, they are 29, 106, 173, 94, respectively.

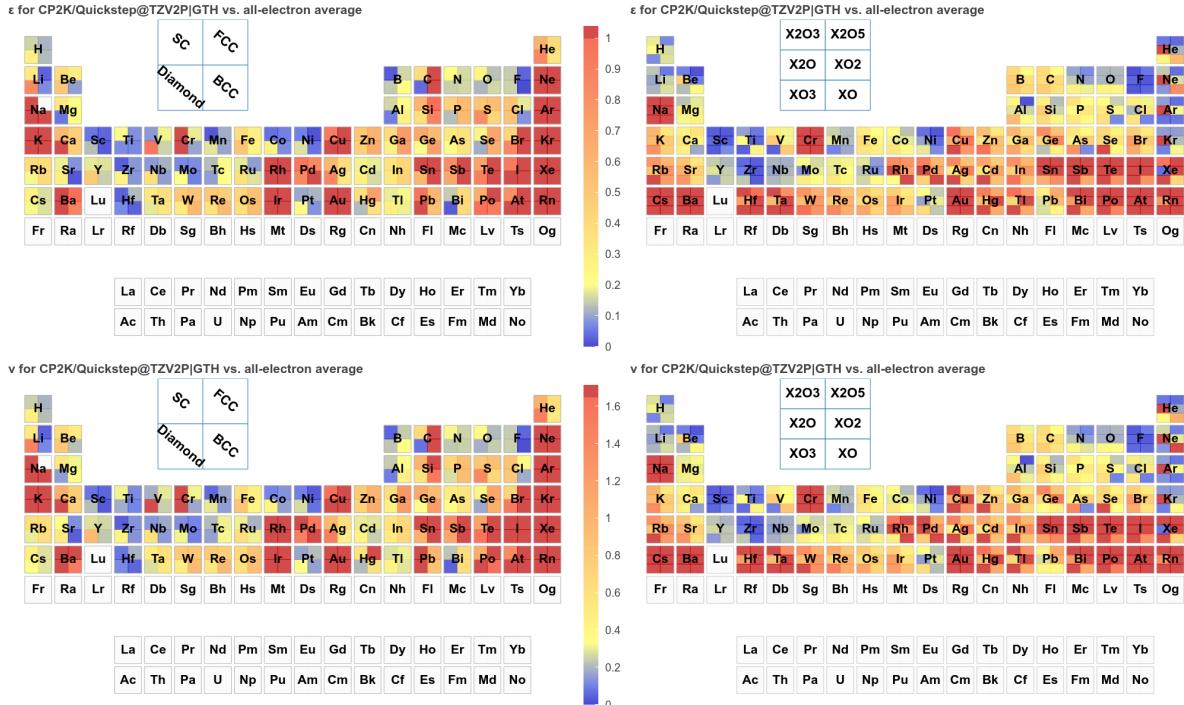


Figure S9.3. Value of the comparison metrics ε (top) and v (bottom) for CP2K/Quickstep@TZV2P | GTH with respect to the average all-electron reference dataset. Left panels: unaries; right panels: oxides. 709 out of 960 crystals were calculated. The number of crystals that land in the excellent, good, noticeably different, and clearly different agreement ranges for the ε metric are 57, 171, 317, 164, respectively. For the v metric, they are 55, 169, 302, 183, respectively.

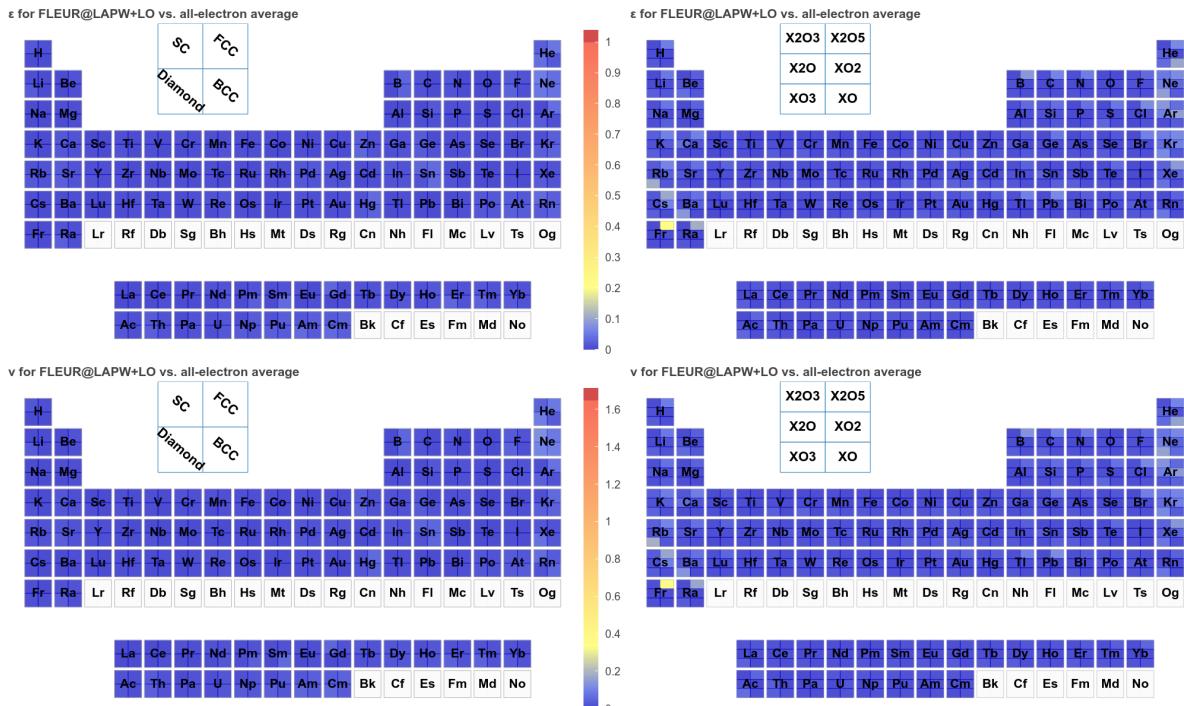


Figure S9.4. Value of the comparison metrics ε (top) and v (bottom) for FLEUR@LAPW+LO with respect to the average all-electron reference dataset. Left panels: unaries; right panels: oxides. 960 out of 960 crystals were calculated. The number of crystals that land in the excellent, good, noticeably different, and clearly different agreement ranges for the ε metric are 936, 23, 1, 0, respectively. For the v metric, they are 938, 22, 0, 0, respectively.

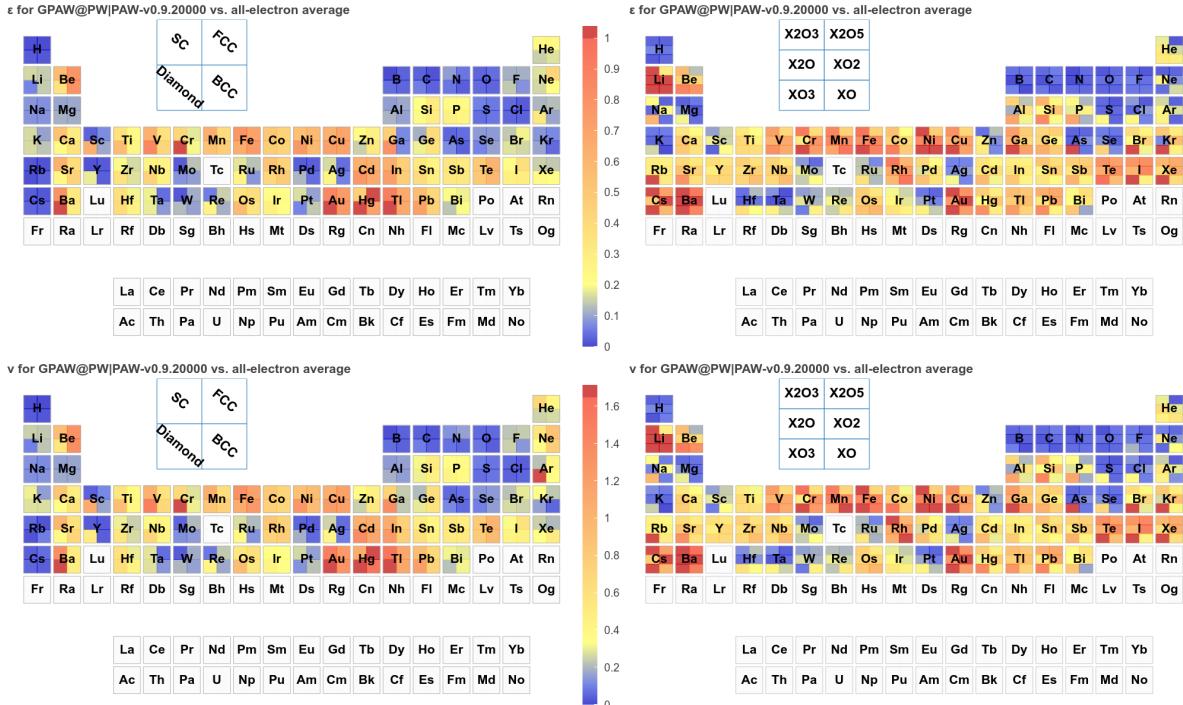


Figure S9.5. Value of the comparison metrics ε (top) and v (bottom) for GPAW@PW | PAW-v0.9.20000 with respect to the average all-electron reference dataset. Left panels: unaries; right panels: oxides. 670 out of 960 crystals were calculated. The number of crystals that land in the excellent, good, noticeably different, and clearly different agreement ranges for the ε metric are 130, 156, 350, 34, respectively. For the v metric, they are 128, 155, 347, 40, respectively.

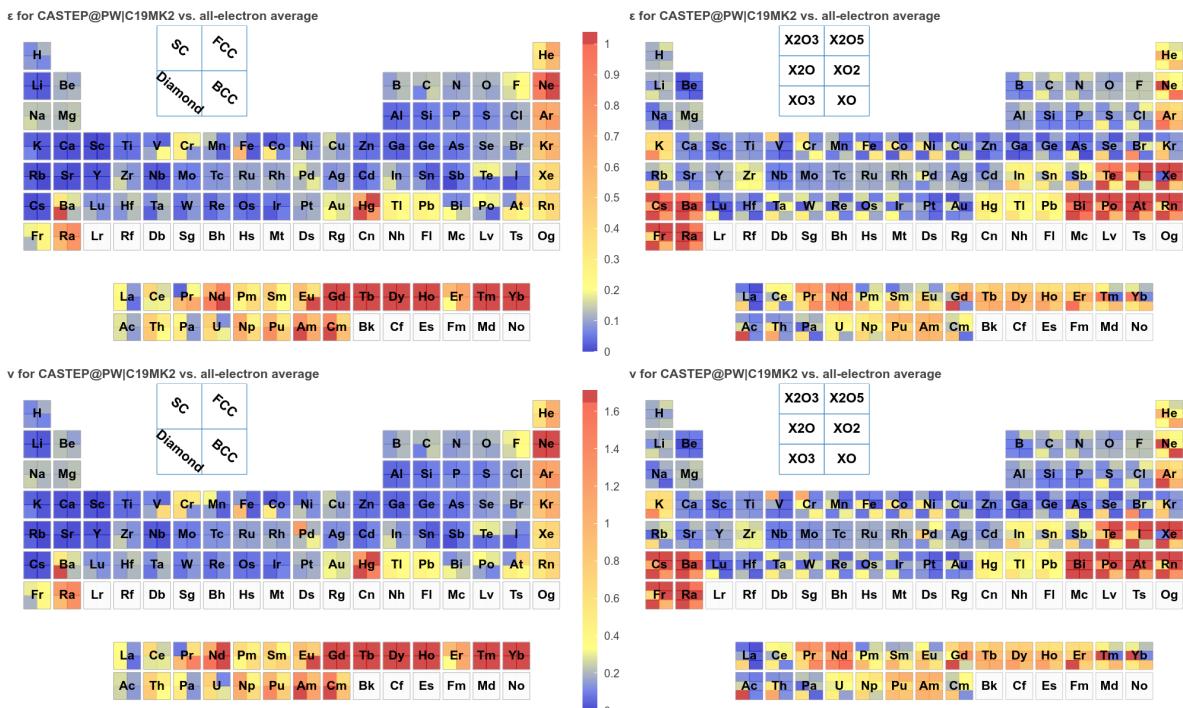


Figure S9.6. Value of the comparison metrics ε (top) and v (bottom) for CASTEP@PW | C19MK2 with respect to the average all-electron reference dataset. Left panels: unaries; right panels: oxides. 960 out of 960 crystals were calculated. The number of crystals that land in the excellent, good, noticeably different, and clearly different agreement ranges for the ε metric are 197, 410, 277, 76, respectively. For the v metric, they are 206, 399, 267, 88, respectively.

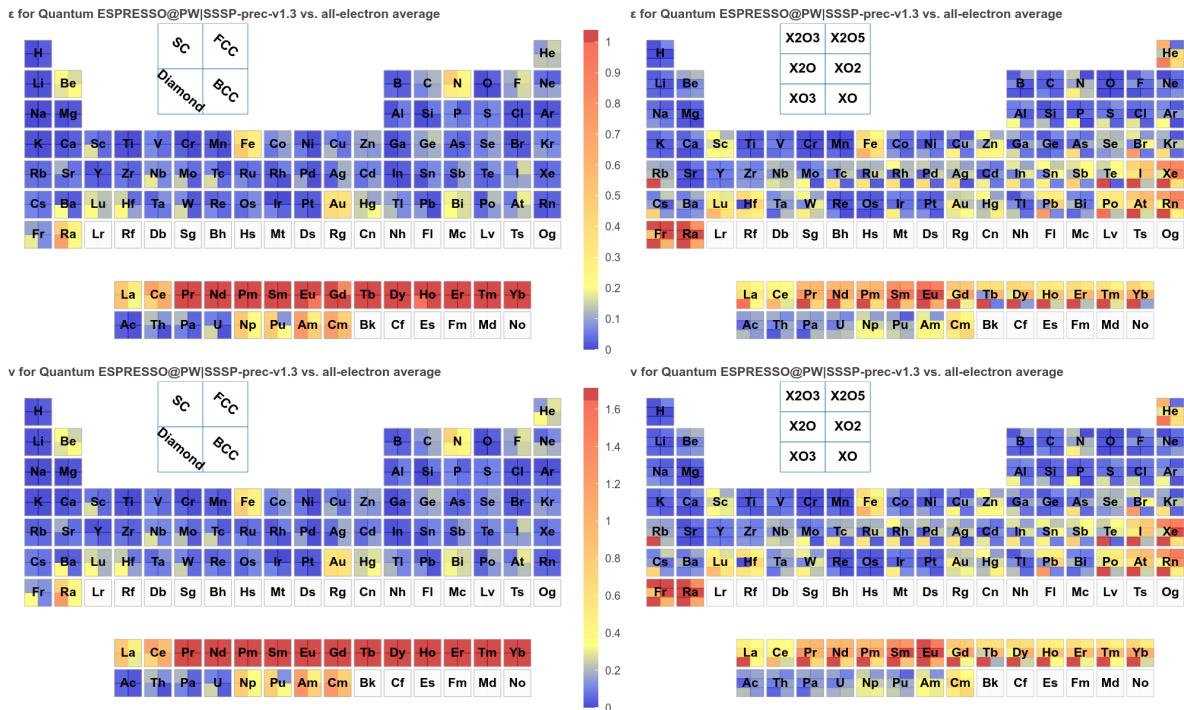


Figure S9.7. Value of the comparison metrics ε (top) and v (bottom) for Quantum ESPRESSO@PW/SSSP-prec-v1.3 with respect to the average all-electron reference dataset. Left panels: unaries; right panels: oxides. 960 out of 960 crystals were calculated. The number of crystals that land in the excellent, good, noticeably different, and clearly different agreement ranges for the ε metric are 388, 300, 199, 73, respectively. For the v metric, they are 395, 300, 184, 81, respectively.

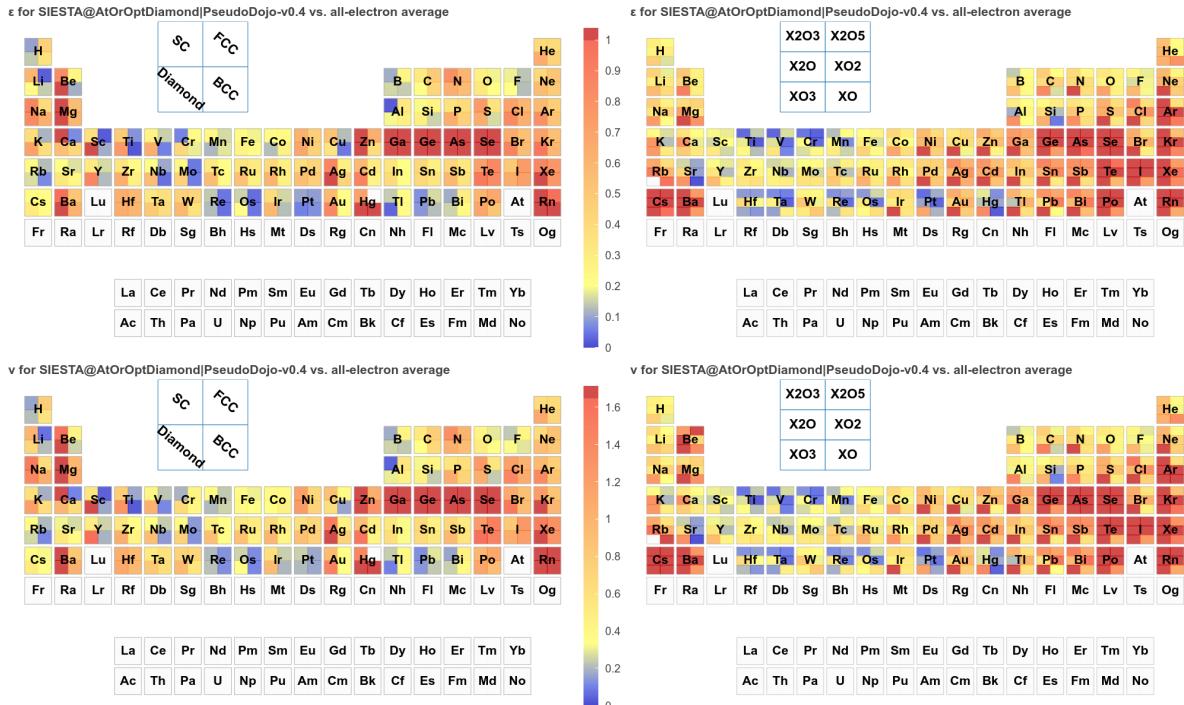


Figure S9.8. Value of the comparison metrics ϵ (top) and ν (bottom) for

Siesta@AtOrOptDiamond|PseudoDojo-v0.4 with respect to the average all-electron reference dataset. Left panels: unaries; right panels: oxides. 698 out of 960 crystals were calculated. The number of crystals that land in the excellent, good, noticeably different, and clearly different agreement ranges for the ϵ metric are 30, 117, 444, 107, respectively. For the ν metric, they are 18, 137, 424, 119, respectively.

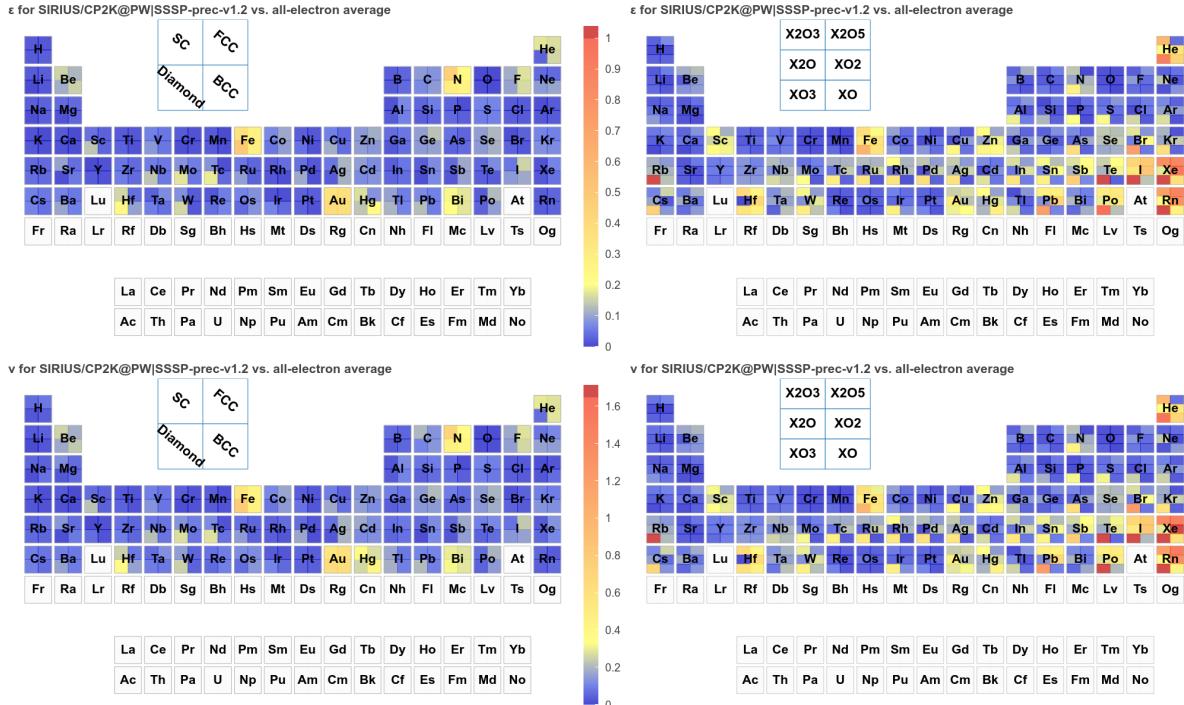


Figure S9.9. Value of the comparison metrics ε (top) and ν (bottom) for SIRIUS/CP2K@PW|SSSP-prec-v1.2 with respect to the average all-electron reference dataset. Left panels: unaries; right panels: oxides. 700 out of 960 crystals were calculated. The number of crystals that land in the excellent, good, noticeably different, and clearly different agreement ranges for the ε metric are 363, 251, 81, 5, respectively. For the ν metric, they are 374, 247, 72, 7, respectively.

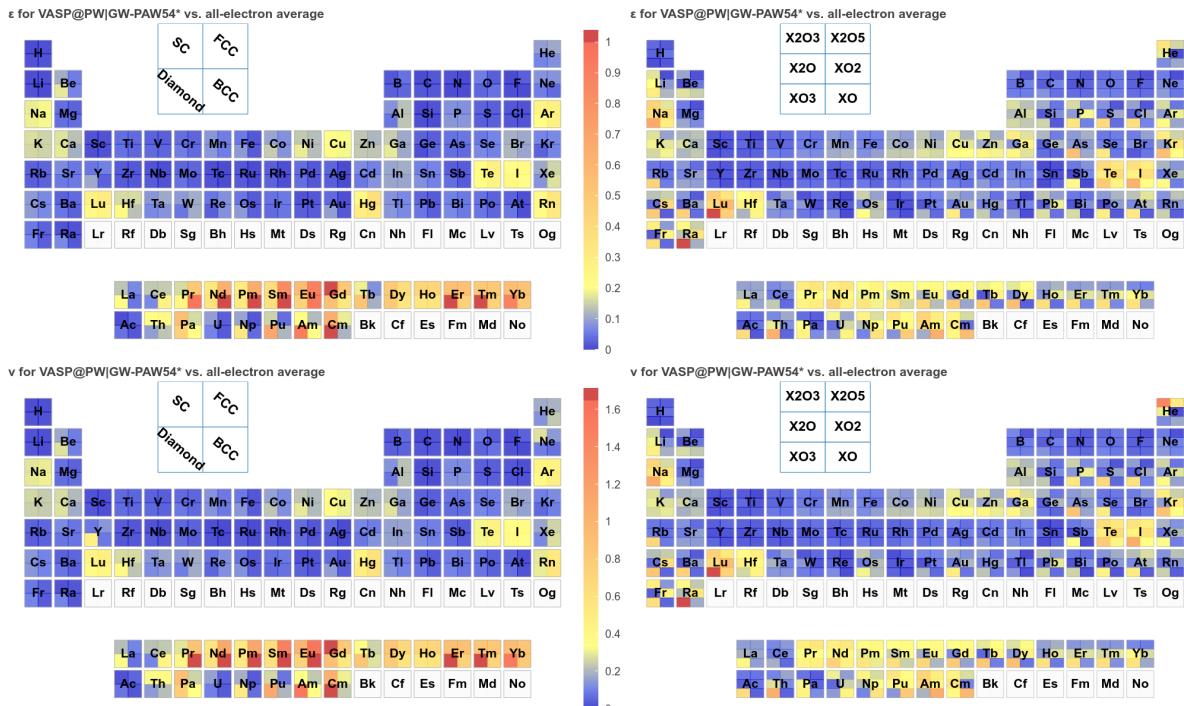


Figure S9.10. Value of the comparison metrics ε (top) and ν (bottom) for VASP@PW|GW-PAW54* with respect to the average all-electron reference dataset. Left panels: unaries; right panels: oxides. 960 out of 960 crystals were calculated. The number of crystals that land in the excellent, good, noticeably different, and clearly different agreement ranges for the ε metric are 403, 348, 200, 9, respectively. For the ν metric, they are 419, 341, 189, 11, respectively.

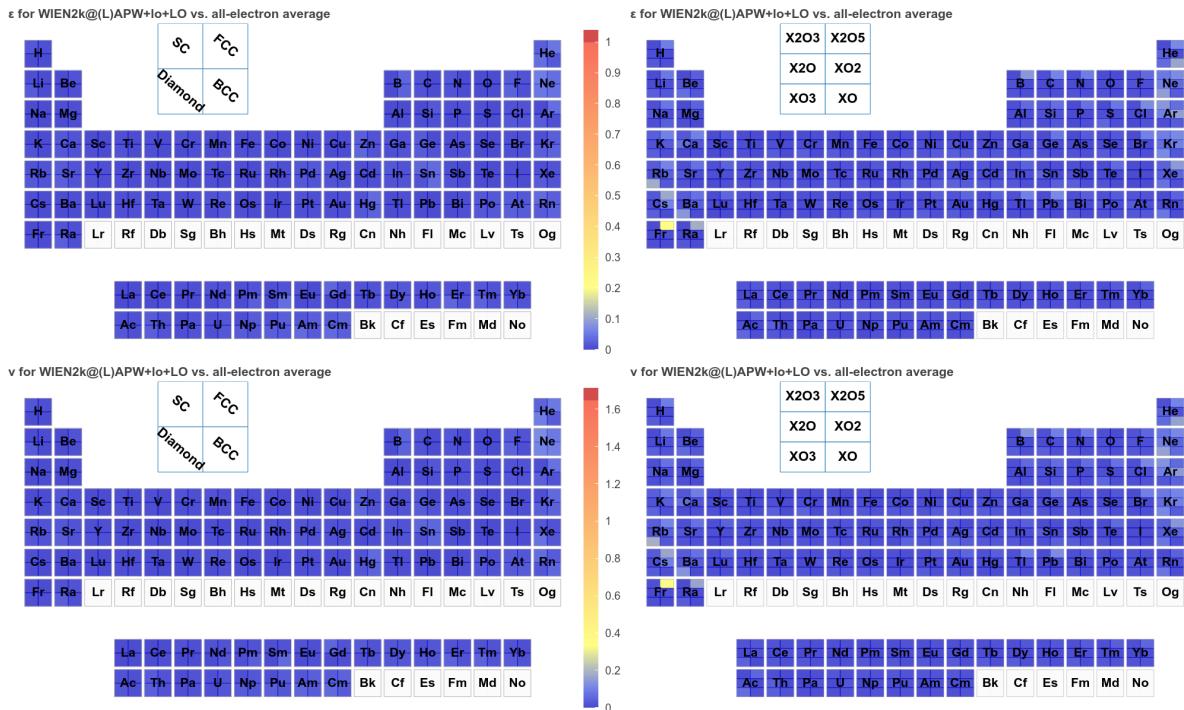


Figure S9.11. Value of the comparison metrics ε (top) and ν (bottom) for WIEN2k@L APW+lo+LO with respect to the average all-electron reference dataset. Left panels: unaries; right panels: oxides. 960 out of 960 crystals were calculated. The number of crystals that land in the excellent, good, noticeably different, and clearly different agreement ranges for the ε metric are 936, 23, 1, 0, respectively. For the ν metric, they are 938, 22, 0, 0, respectively.

S10 Smearing and k-point convergence

For the study presented in this paper, a fixed choice of k-point integration mesh and smearing has been implemented. In particular, the k-point mesh is a uniform regular grid including the Γ point, that guarantees a linear spacing of 0.06 \AA^{-1} in each of the three reciprocal-space directions, and the smearing is a Fermi–Dirac type with broadening of 0.0045 Ry . This choice of parameters is essential in order to compare with the reference dataset presented in this manuscript, as explained in Box 3 of our recommendations. In this section, we present the reasoning for our choice of parameters.

The Fermi–Dirac smearing is widely used in the community and it is implemented in all the codes that participate in the study. The choice of the broadening has been made according to the recommendation of Ref. 66: the smearing parameter should not be too small to avoid sampling errors, nor too large to prevent systematic deviations due to the dependency of the total energy on the smearing broadening. The latter problem is explained in details in Ref. 66, that shows the quadratic dependence of the free energy with respect to the Fermi–Dirac smearing temperature. The problem of sampling error is instead demonstrated in SI Fig. S10.1, where we analyze a FCC aluminum crystal (conventional cell with 4 atoms) with an atom displaced by 0.1 \AA with respect to its equilibrium position. The figure reports the magnitude of the force on the displaced atom as a function of the k-point integration mesh and smearing broadening. The instability of the force for very small broadening is clearly visible. Approaching the zero limit of the smearing, it becomes more and more necessary to have a dense k-point integration mesh in order to maintain 0.001 eV/\AA convergence on the forces. The sampling error is not a peculiar feature of the Fermi–Dirac smearing; any other smearing type suffers from this drawback, as demonstrated in Ref. 66. Our choice of smearing broadening ($0.0045 \text{ Ry} \approx 61 \text{ meV}$) lays on the extreme right of SI Fig. S10.1 and in this region a k-point integration mesh of $36 \times 36 \times 36$ is sufficient to obtain a converged value of the force within 0.001 eV/\AA . At the same time, 0.0045 Ry is a small enough value to reduce to a minimum issues due to the dependency of the total energy on the smearing broadening. In any case, for the goal of verification, the exact value is not so important, as long as all codes perform the very same choice of smearing and k-point integration mesh.

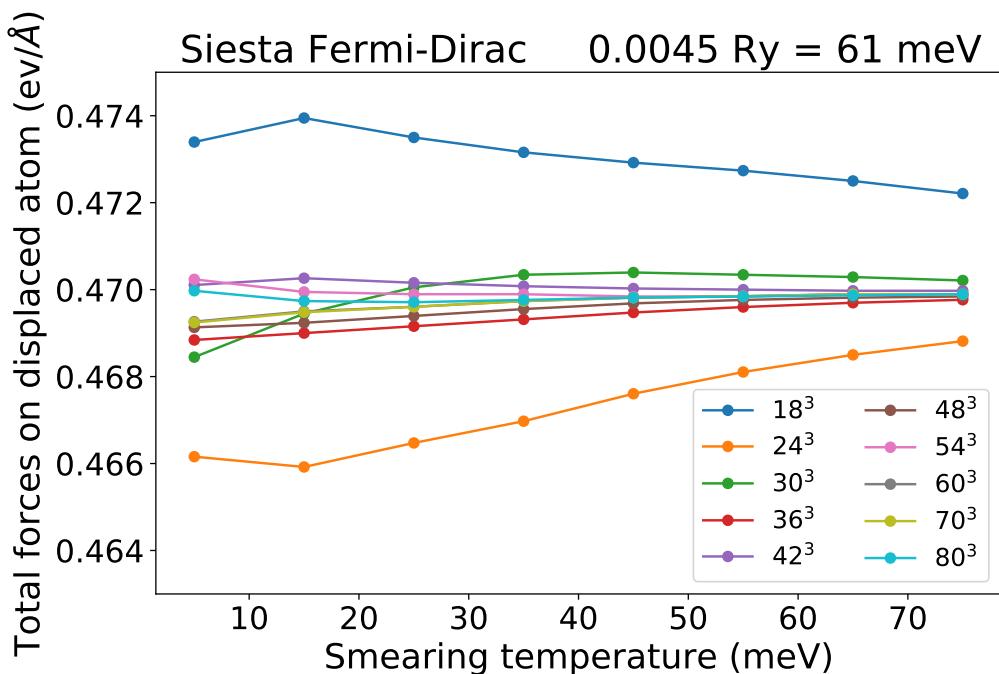


Figure S10.1. Total force on a displaced atom as a function of the smearing temperature and the k-point integration mesh. The system under investigation is an FCC Al with an atom displaced by 0.1 \AA in the x direction. The volume per atom is 16.47 \AA^3 , for which a distance of 0.06 \AA^{-1} between k-points correspond to a $26 \times 26 \times 26$ mesh. The smearing is a Fermi–Dirac smearing. Calculations of the forces are made with SIESTA and DZP (double-zeta polarized) basis set.

With our choice of the smearing broadening, we expect a rapid convergence of the electronic-structure properties with respect to the the k-point density. We test this assumption looking at the effect of the k-point integration mesh on the estimation of the EOS parameters. Using WIEN2k results, we compare calculations with k-point distance of 0.06 \AA^{-1} and 0.045 \AA^{-1} for all materials in the study. The comparison is reported in SI Fig. S10.2. This figure shows an overall discrepancy that is

significantly smaller with respect to the FLEUR-WIEN2k comparison presented in Fig. 1 in the main text (note that the x-axis range is half of the one in Fig. 1). Looking at the histograms, we can estimate that the overall agreement is at least a factor of 2 better for B_1 with respect to the FLEUR-WIEN2k comparison, more than a factor of 4 better for B_0 and even one order of magnitude better for V_0 . The same conclusion cannot be drawn looking at the standard deviations reported in the figure, due to the presence of two important outliers: RbO₃ (3.7% difference in V_0 , -142.29% difference in B_0 and 149.35% difference in B_1) and HeO (0.16% difference in V_0 , 3.33% difference in B_0 and -7.90% difference in B_1). RbO₃ and HeO are the only two materials that are not converged with a k-point distance of 0.06 Å⁻¹. All other materials are converged within 0.07% of V_0 . It is interesting to notice that, even though RbO₃ and HeO are not converged, their discrepancy is not so dramatic when comparing FLEUR and WIEN2k (Fig. 1 in the main text). This, once more, justifies our recommendation of adopting the same k-point integration mesh for all computational approaches.

For completeness, we mention that the k-point mesh comparison of SI Fig. S10.2 has been performed on the crystal-structure set described in the main text for the unaries. For the oxides set, instead, we present results of a previous iteration of the volume refinement. Therefore the structures used for the oxides calculation in SI Fig. S10.2 have central volumes that slightly differs with respect to the ones used in the main text.

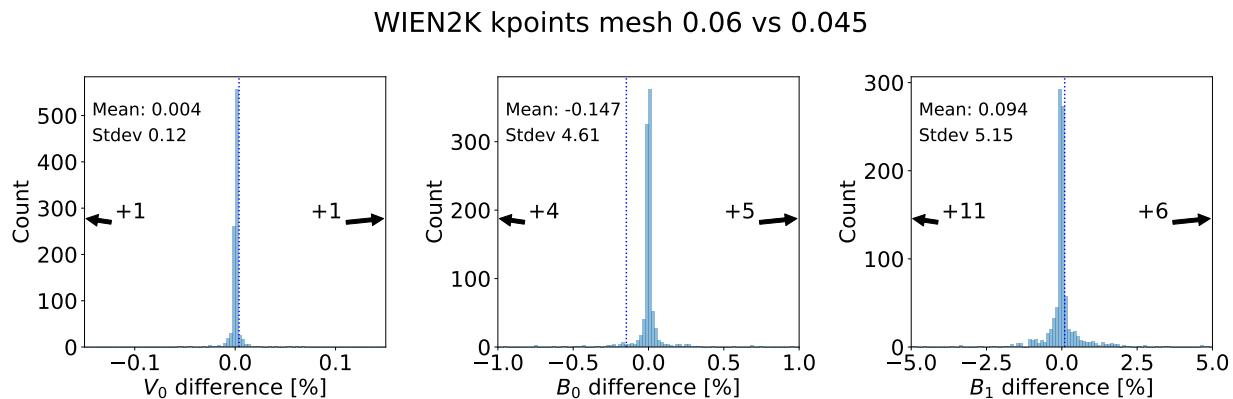


Figure S10.2. Histograms of the percentage difference between the results obtained with a k-point integration mesh with linear density 0.06 Å⁻¹ and 0.045 Å⁻¹ for the three EOS parameters V_0 , B_0 , and B_1 . Results are obtained with WIEN2k code. Numbers close to the arrows indicate outliers beyond the x-axis range.

S11 Band structure of erbium in the diamond crystal structure

In SI Fig. S11.1 we show the band structure of erbium in the diamond crystal structure, obtained using the initial crystal structure (see SI Table S1.3) with conventional cubic lattice parameter of 8.6296 Å. The simulation has been run with the QUANTUM ESPRESSO code (version 7.0) using the PBE pseudopotential from the SSSP PBE Precision 1.2 library.⁷⁸, and the recommended cutoffs of 40 Ry and 320 Ry for the wavefunctions and charge density, respectively. A $9 \times 9 \times 9$ k-point integration mesh is chosen (note: this is less dense than the mesh recommended in the main text, but sufficient to demonstrate qualitatively the key features of the band structure, that is the goal of this section), and a Fermi–Dirac smearing with 0.0045 Ry of broadening. Input and output files are available in Ref. 62.

The band structure clearly displays a set of “almost flat” bands (with a dispersion of ~ 0.1 eV) very close to the Fermi energy in the range between -0.2 eV and 0 eV, originating from the f states of erbium. For this specific calculation and volume, these f bands are just below the Fermi level, but their position can shift with volume and cross the Fermi level, significantly affecting the nature of the occupied states in the material. This explains the unconventional shape of the EOS displayed in the main text in Fig. 3.

An interesting note is that the location of the f bands, determining the lowest-energy minimum of the EOS, also depends on the value Z , i.e., on the column of the periodic table. For our choice of parameters (k-point integration mesh and smearing) erbium is at the boundary between elements favoring the minimum at lower volume (for smaller Z) and elements favoring the minimum at higher volume (for higher Z). This is clearly visible in SI Fig. S1.3a as a jump of the first-neighbor distance for the diamond structures for elements before and after erbium.

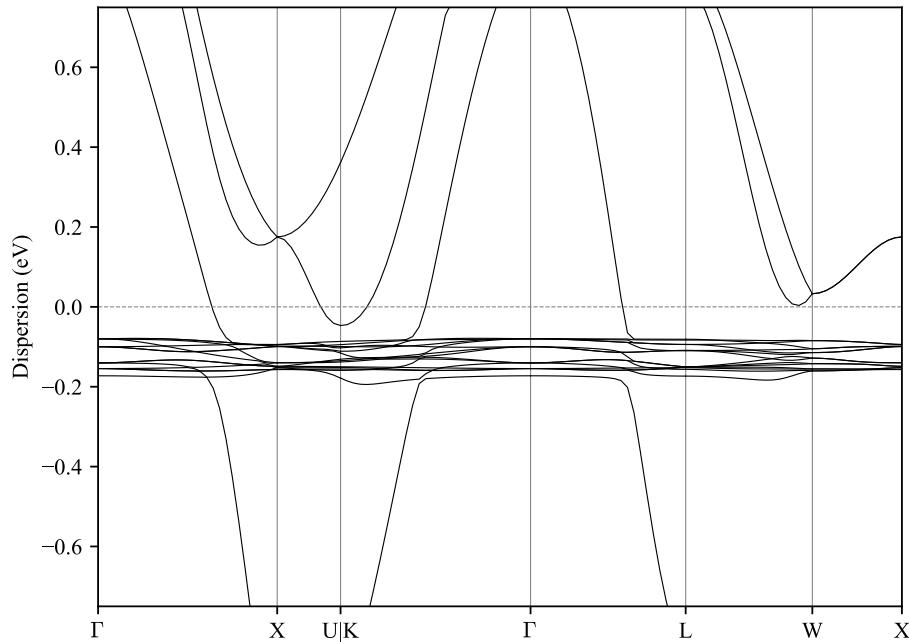


Figure S11.1. Band structure of erbium in the diamond crystal structure. The zero of energy is set at the Fermi level.

S12 Total energy versus free energy

Here we want to show a couple of examples where the exact choice of the energy to be computed in the **EOS**, namely the internal energy E , the free energy $E - TS$ (where $-TS$ is the smearing contribution) or the approximation $E - TS/2$, can significantly affect the **EOS** curves.

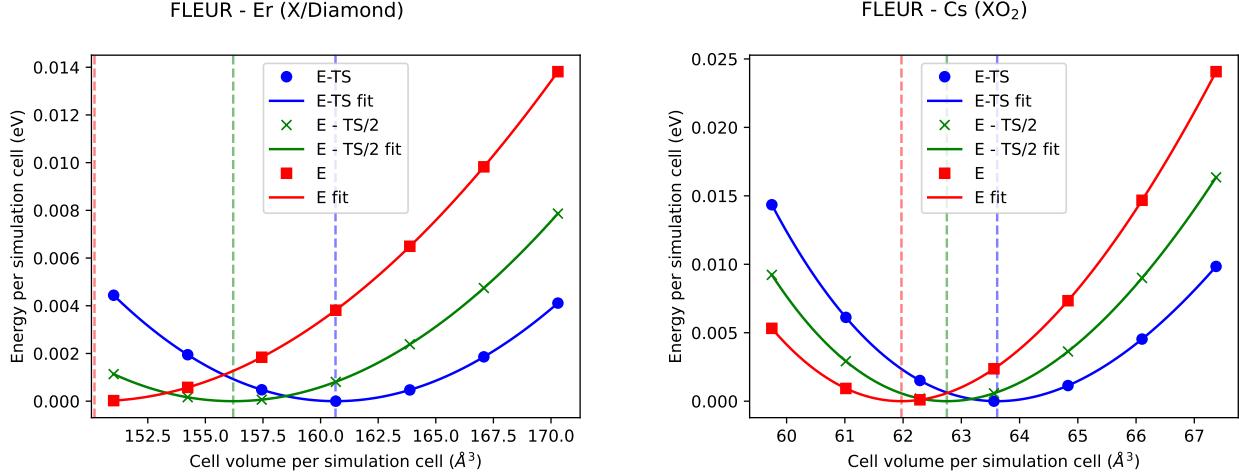


Figure S12.1. Comparison of **EOS** curves for two systems computed using the FLEUR code, using different quantities on the energy axis. Left panel: Er in the diamond structure; right panel: CsO_2 .

While in many cases the choice of one of these three quantities does not have any effect on the curves (e.g., in the case of large-gap insulators), in SI Fig. S12.1 we show two **EOS** curves that we selected as they show a significant deviation (even larger deviations exist for other systems in our dataset). The first case is erbium in the diamond structure. We already discussed its band structure in SI Sec. S11: we expect that there are significant flat f bands (thus with high density of states) crossing the Fermi level as a function of volume; therefore, the $-TS$ contribution will also depend on volume, since it originates from the contribution of bands within a small energy range (comparable to the chosen smearing broadening) from the Fermi level.

One does not need, however, to consider such pathological cases: even other systems might show important discrepancies between the various curves if the density of states at the Fermi level changes significantly as a function of volume. This is illustrated for instance by CsO_2 (right panel in SI Fig. S12.1), whose band structure for two different volumes is reported in SI Fig. S12.2. Here, we see that bands with saddle points or almost-flat bands can be found within $\pm\sigma$ from the Fermi level (with σ being the value of smearing broadening recommended here). These bands shift with volume and therefore their contribution to $-TS$ will be a function of volume.

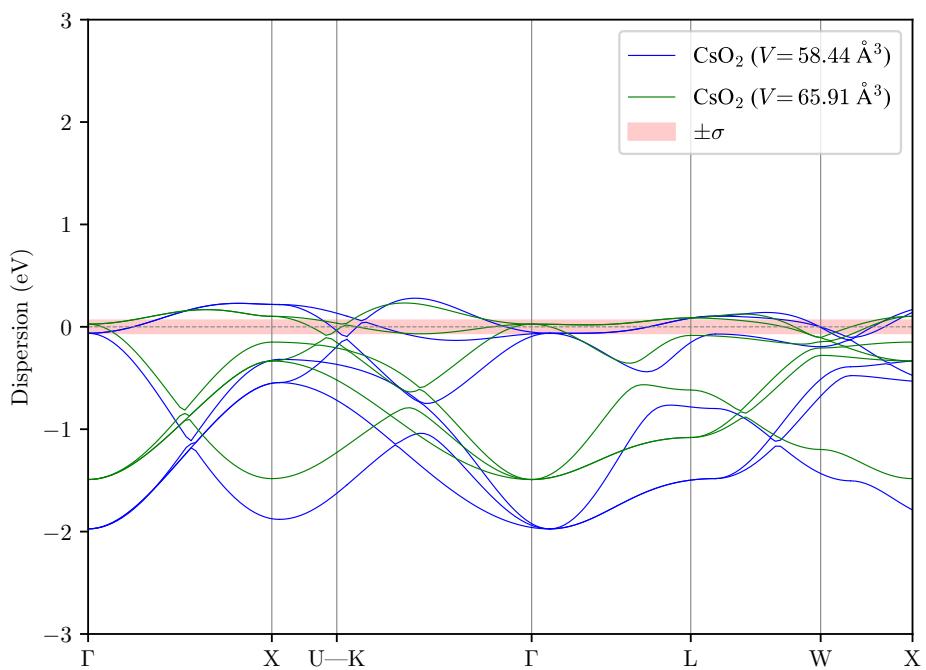


Figure S12.2. Band structure of CsO_2 for two different volumes computed with the QUANTUM ESPRESSO code. The shaded red area indicates an energy range of $\pm 0.0045 \text{ Ry}$, that is the smearing value recommended in this work.

S13 Code specific parameters for the pseudopotential codes

In this section, we discuss the technical choices adopted for the verification work and implemented, for each code, in the verification-PBE-v1 protocol. We remind the reader that a number of parameters (such as the smearing type and broadening or the k-point integration mesh) have been fixed for all codes (see also Box 3 in the main text). We also remind that the parameters for the two all-electron codes are discussed instead in SI Sec. S5.

S13.1 ABINIT@PW|PseudoDojo-v0.5

The ABINIT calculations were performed with the version 9.6.2 of the code^{13,42,43} and v0.2a2 of the aiida-abinit plugin. All calculations were run with `tolvrs = 1e-10`, Fermi–Dirac smearing (`occopt = 3`) of 0.0045 Rydberg (`tsmear = 0.00225`), a reasonable number of empty bands (`fband = 2`), and a minimal k-spacing of 0.06 Å⁻¹. Norm-conserving pseudopotentials from the PseudoDojo-v0.5 scalar-relativistic PBE standard library have been employed; the “high” stringency recommended energy cutoffs were used. The RMM-DIIS diagonalization algorithm⁵⁵ (`rmm_diis = 1`) was used for calculations with norm-conserving scalar-relativistic PBE pseudopotentials because of its improved computational efficiency over the default CG method. Note that `rmm_diis = 1` means that the first four SCF iterations are performed with the CG method in order to obtain reasonably good trial states before changing to the RMM-DIIS method. RMM-DIIS is more efficient although less stable than the CG algorithm as there is no explicit orthogonalization while optimizing the trial states. This led to approximately a 0.7% error rate (108 out of 15460 total from all calculations performed in the process of this study) in the calculations, however in all cases, running the failed calculations with CG resulted in successful convergence.

S13.2 BigDFT@DW|HGH-K (Valence)

The version of the BigDFT code employed for these calculation is the 1.9.2. For the great majority of the structures presented here, the pseudopotentials employed in the calculations are norm-conserving Hartwigsen–Goedecker–Hutter⁷⁰ of the Krack family⁷² (HGH-K). For this verification campaign, rather than choosing the most precise pseudopotentials for a given element, we employed the default pseudopotentials with the least possible number of valence electrons. We have therefore used a set which provides an overestimation of the precision error of this pseudopotential family. For comparison we have also included some of the semicore pseudopotentials, see SI Sec. S16.2. The BigDFT code formalism employs Daubechies wavelets basis sets to express the Kohn–Sham (KS) orbitals, which enable to reach precise converged results for a given pseudopotential with moderate effort with respect to the number of degrees of freedom employed. The wavelets grid spacing was set to a value of 0.3 a_0 , with all the high-resolution degrees of freedom activated, and the k-point integration mesh correspond to a equivalent length of 94 a_0 . Density mixing scheme was employed for electronic convergence, reached for a threshold voxel accuracy of 10⁻¹² atomic units. Also, 120 empty Kohn–Sham states were included for each k-point. Symmetry operations were also included to limit the calculations to the irreducible k-points. At the time of developing the workflow, the BigDFT code was migrating its user interface into the PyBigDFT python module, which provides a user interface to the underlying executable. Therefore, to avoid issues in the API modifications, we fixed the PyBigDFT version to a beta release in the aiida-bigdft plugin, which activated limited features with respect to the stable version available nowadays. In particular, not all the structures which were defined with a non-orthogonal unit-cell were transformed in an orthogonal supercell, required by the code. This compatibility problem resulted in less structures treated by this approach. It is planned to release a stable version of the plugin compatible with the AiiDA 2.x API.

S13.3 CASTEP@PW|C19MK2

CASTEP is a plane-wave pseudopotential code⁴⁵, the 20.1.1 version is used in this work. Calculation parameters closely follow the “precision” setting in the initial common workflow implementation³⁶. The cut off energy is fixed at 800 eV for all calculations since energy comparison is needed between different chemical systems. The reciprocal space sampling is done through Γ centered Monkhorst–Pack grids with a fixed spacing of 0.06 Å⁻¹ (i.e., 0.00954929 2πÅ⁻¹), in line with other codes. On-the-fly generated (OTFG) core-corrected ultrasoft pseudopotentials from the library C19 is used for the study except for the f-block elements (see section S16.3 for more details). The C19 library is aimed for general use with a balance between precision and speed, and it has been the default potential library since CASTEP version 19.1.1. The modified pseudopotential generation strings for f-block elements are tabulated in Table S16.1. The energy convergence threshold for electronic minimization is set to 1×10^{-8} eV per atom.

S13.4 CP2K/Quickstep@TZV2P|GTH

The DFT module QUICKSTEP of the open-source simulation package CP2K is an implementation of the Gaussian and plane wave (GPW) and the all-electron Gaussian augmented plane wave approaches^{46,47}. Therein, the Kohn–Sham orbitals are represented by contracted Gaussian basis functions, whereas the electronic charge density is expanded in plane waves¹⁰⁹. For

the former, an accurate molecularly optimized triple- ζ basis set with two additional sets of polarization functions (TZV2P-MOLOPT) is employed⁷³, whereas for the latter a density cutoff of 2400 Ry is utilized, which differs from a conventional plane wave cutoff by a factor of four. Due to its GPW method, however, CP2K/QUICKSTEP is rather insensitive with respect to high density cutoffs. Furthermore, four multi-grids are used to ensure an efficient mapping of product Gaussians onto the real-space integration grids, so that wide and smooth Gaussian functions are mapped onto a coarser grid than narrow and sharp Gaussians. To control which product Gaussians are mapped onto which level of the multi-grid, a relative cutoff of 80 Ry is applied that defines the plane wave cutoff of a reference grid covered by a Gaussian with unit standard deviation. Separable and norm-conserving Goedecker–Teter–Hutter-type pseudopotentials including scalar relativistic effects are used to describe the interactions between the valence electrons and the ionic cores^{71,72}.

S13.5 GPAW@PW|PAW-v0.9.20000

GPAW^{48,49} is an open-source DFT code developed at the Technical University of Denmark (DTU) and other universities and computer centers, originally created for combining a homogeneous grid basis set with the projector augmented wave (PAW) method. Today, the code also provides a linear combination of atomic orbitals (LCAO) basis¹¹⁰ and a plane wave mode. The latter has been applied in this study with a plane-wave cutoff of 800 eV for all calculations, with the exception of systems containing noble gases, where a cutoff of 1200 eV combined with a tighter energy and density convergence was applied. As with other codes, Fermi–Dirac smearing of 0.06122 eV was used. Besides the mentioned parameters, default values (as per GPAW v. 21) were used for all other keywords needed to perform the calculation in order to ensure that the results reflect the most representative user experience. The applied PAW potentials, included in GPAW’s PAW potential suite, were specifically created for the PBE exchange correlation functional, by applying GPAW’s setup creator¹¹¹. In particular, we use the pseudopotentials included in the setup release 0.9.20000 available at <https://wiki.fysik.dtu.dk/gpaw/setups/setups.html#atomic-paw-setups>. GPAW is tightly linked to the atomic simulation environment (ASE)^{8,9}, which handles the user interface and is developed independently.

S13.6 Quantum ESPRESSO@PW|SSSP-prec-v1.3

All calculations have been run using version 7.0 of the Quantum ESPRESSO code and version 3.5.1 of the AiiDA Quantum ESPRESSO plugin. For the results presented in the main text, all pseudopotentials were selected from the SSSP PBE Precision 1.3⁷⁶ library, with plane-wave cutoffs corresponding to the largest recommended value from the elements in each structure. In accordance with the verification-PBE-v1 protocol, the Brillouin zone sampling was performed using Γ -centered meshes with a spacing of 0.06 Å and a Fermi–Dirac smearing of 0.0045 Ry. All other inputs parameters were set via the precision protocol as described in the SI of Ref. 37, most importantly the energy convergence threshold was set to a very strict $0.1 \cdot 10^{-9}$ Ry per atom. For the comparison with ABINIT and CASTEP discussed in S14, the pseudopotentials were selected from the PseudoDojo SR PBE standard set, version 0.4, in UPF format, after a small modification of the .upf files as described in S14. Plane-wave cutoffs were obtained from the "high" stringency hints provided by the PseudoDojo table, all other computational parameters were unchanged.

S13.7 SIESTA@AtOrOptDiamond|PseudoDojo-v0.4

The calculations presented in this work have been carried out with Siesta version Max-1.2.0 (<https://gitlab.com/siesta-project/siesta/-/tags/MaX-1.2.0>) powered by the aiida-siesta plugin version 1.2.0. Pseudopotentials from the PseudoDojo FR standard set, version 0.4, in PSML format, have been employed. The real-space cutoff for the representation of charge densities and potentials is fixed at 900 Ry. The recommended course of action regarding basis sets in Siesta is to perform an optimization considering the key features of the chemical environment of each system. In this project, we have not carried out the optimization for all 960 systems. Instead, we have attempted a partial, per element, optimization, considering only the unary diamond crystals at their central volume. The orbitals thus optimized for each element are then reused for all the other unary and oxide structures involving that element. The optimization starts from a TZDP basis with the addition of an extra f orbital. For alkali metals and alkaline earth metals the addition of a d orbital shell is also necessary. The optimization is performed with the Nelder–Mead algorithm (multidimensional optimization without derivatives), having as variables the first-zeta radius of each orbital and the split norm parameter that controls the ratio between the first and subsequent zetas. We foresee using the information garnered in this verification study to develop further heuristics and guide the development of fully automatic methods to generate basis sets taking into account appropriate chemical environment descriptors.

S13.8 SIRIUS/CP2K@PW|SSSP-prec-v1.2

SIRIUS is a domain-specific library, which implements pseudopotential plane wave and full potential linearized augmented plane wave methods and is designed for GPU acceleration⁵⁴. As such it brings additional functionalities to CP2K such as collinear and non-collinear magnetic systems with or without spin-orbit coupling. It is written in C++14 with the MPI,

OpenMP, and CUDA/ROCm programming models. As shown previously, SIRIUS/CP2K allows for energy conserving *ab-initio* molecular dynamics simulations with a constant shift in the order of μHa compared to QUANTUM ESPRESSO reference calculations⁴⁷. All SIRIUS/CP2K simulations were performed using the pseudopotentials of SSSP PBE Precision 1.2⁸⁰ together with a plane-wave cutoff of 55 a.u.⁻¹ for the density and potential, as well as 10 a.u.⁻¹ for $|\mathbf{k} + \mathbf{G}|$, respectively.

S13.9 VASP@PW | GW-PAW54*

All VASP results in this work have been obtained with VASP version 6.3.0 and AiiDA-VASP 2.1.0, using preferentially the GW VASP PBE potential set version 54.

The input parameters explicitly set for this work were (defaults in brackets): PREC = Accurate(Normal), EDIFF = 1E – 7(1E – 4), ALGO = Normal(Normal), NELM = 300(60), and LMAXMIX = 6(2). The last setting ensures that all electronic states up to quantum numbers $l = 6$ are included in the density mixer, and it is necessary to change this in order to converge some *d*-electron systems and most lanthanides. The plane wave cutoff was fixed at 1000 eV (~ 73.50 Ry) for all calculations. The PAW method was used¹¹².

We used the recommended GW potential sets, whenever possible, with exception of oxygen, where we chose the hard O_h_GW potential instead of the recommended O_GW since it improved results for the oxides dataset. This potential is required if short bonds to oxygen atoms are encountered. The recommended plane wave cutoff energy for this potential is 765 eV, however, for high precision studies we recommend to increase the cutoff in VASP by 30% yielding the employed 1000 eV. For the elements where no GW potential is supplied, we use the recommended standard PBE PAW potential. (The recommendations were taken from the [VASP-wiki](#) on April 22nd, 2022).

S14 Precision of plane-wave codes when using the same pseudopotential library

In order to assess how much codes implementing the same computational approaches agree among each other, we compare the ε metric among a subset of the plane-wave codes considered in this work, when using the same pseudopotential library. We consider two different cases. In the first case, we compare three different codes (ABINIT, QUANTUM ESPRESSO, and CASTEP) using the PseudoDojo v0.4 pseudopotential library. In the second case, we compare QUANTUM ESPRESSO and SIRIUS/CP2K using the SSSP PBE precision v1.2 pseudopotential library.

For the first case (PseudoDojo v0.4 library), an initial comparison highlighted some discrepancies among the codes for Cu, Zn, and Ne. After investigation, we found that these discrepancies stem from differences in the numerical treatment of the form factors, the local part of the pseudopotential, the model core charge density or the beta projectors with spherical Bessel functions (usage of spline or not, different integration methods, etc.).

While we have not precisely identified the main root of the effect, which would require additional investigation, it is clear that the treatment of the long range part of the pseudopotential is rather sensitive to the implementation details. For this reason, we manually truncated the radial mesh reported in the pseudopotential files (in UPF format, normally truncated at 10 Bohr) after 6 Bohr for all elements, as it is done in the psp8 format used by ABINIT, in order to have the same radial mesh in all codes. We verified that, except for the three cases mentioned above, this truncation had no visible effects on the results when compared to all-electron results. In the figures below, the tables for which the radial mesh has been truncated will be referred to as “trim” as a suffix to the approach label reported in the periodic-table titles.

Using this simple truncation to avoid numerical instabilities, we show in SI Figs. S14.1, S14.2 and S14.3 the six possible pairwise comparisons between the three codes. Remarkably, except for two noble gases, the agreement is excellent (we have used the same color scale for ε as in the main manuscript). This agreement is on average even better than the agreement between AE codes of this work (see Fig. 1 of the main text). We highlight however that this is expected, since the plane-wave codes use the same basis set and, in this SI Section, an almost identical set of numerical parameters, while the two AE codes differ in the details of the basis set and numerical parameters.

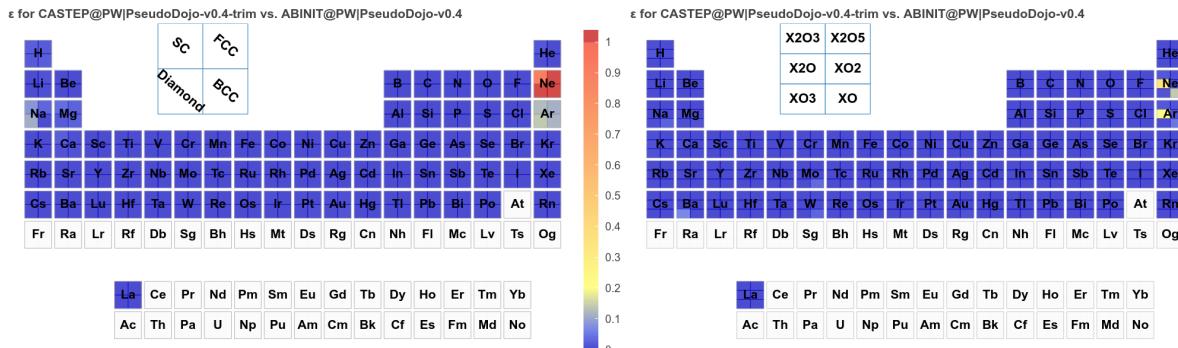


Figure S14.1. Value of the comparison metric ε between the CASTEP code (with truncated radial mesh) and the ABINIT code for the unaries and oxides set using the same PseudoDojo v0.4 library.

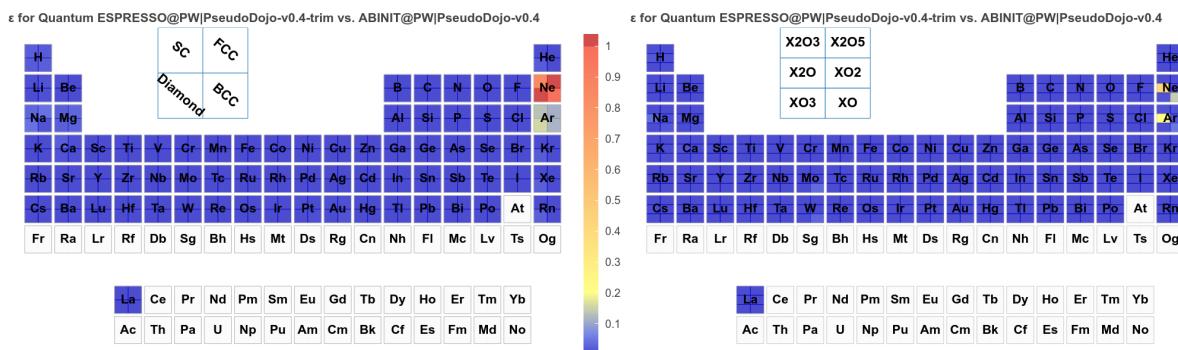


Figure S14.2. Value of the comparison metric ε between the QUANTUM ESPRESSO code (with truncated radial mesh) and the ABINIT code for the unaries and oxides set using the same PseudoDojo v0.4 library.

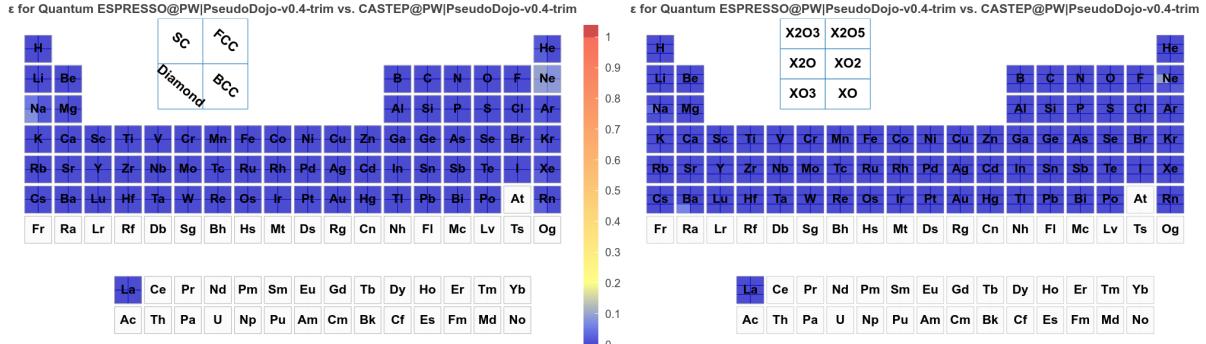


Figure S14.3. Value of the comparison metric ϵ between the QUANTUM ESPRESSO and CASTEP codes (both with truncated radial mesh) for the unaries and oxides set using the same PseudoDojo v0.4 library.

In the second case, we compared the SSSP PBE precision v1.2 pseudopotential library for QUANTUM ESPRESSO and SIRIUS/CP2K. The comparison of the ϵ metric is shown in SI Fig. S14.4. The match is very good for all systems except for Ba-diamond and RbO₃. We note here that in the main text, we used the SSSP PBE precision v1.3 for QUANTUM ESPRESSO, but we stress that it is equivalent to v1.2 for the subset of chemical elements considered in this section (the difference being that v1.3 also includes actinides).

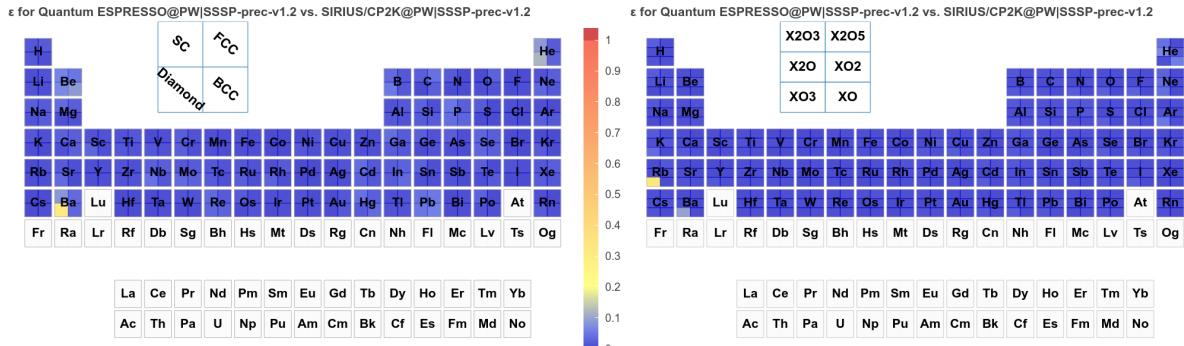


Figure S14.4. Value of the comparison metric ϵ between the QUANTUM ESPRESSO and SIRIUS/CP2K software for the unaries and oxides set using the same SSSP PBE precision v1.2 library.

The overall excellent agreement displayed in this SI Section demonstrates that different codes implementing the same computational approach (including, in addition to the basis set, also the pseudopotential library and other computational parameters) can reproduce the same results. In addition, our investigation helps identifying the remaining numerical aspects that might produce different outcomes and that merit further investigation, such as the truncation of the radial mesh discussed here.

S15 Transferability of conclusions from the previous smaller crystal set to the current 960 crystal set

In this section we want to show that even if two codes provide nearly identical results for the 71 crystal set of Refs. 29, 30, they will not necessarily provide identical results for the larger 960 crystal set of the present paper, demonstrating the value of the larger crystal-structure set.

To perform a quantitative comparison using the same computational approach, we use only data obtained as part of the present study. That means that we do not use the 71 crystal set (that has several crystals that are not part of the present set), but rather the set of 29 crystals common to both sets, as listed in SI Sec. S17, as a proxy for the 71 crystal set. SI Fig. S15.1 shows on the horizontal axis the ϵ averaged over these 29 crystals for any pair of approaches or codes used in this work, and on the vertical axis the ϵ for the same pairs but now averaged over all crystals (up to 960 crystals). The overall correlation in SI Fig. S15.1 shows that the 29 crystal set does indeed capture part of the information. However, as expected, the data points fall above the $y = x$ line, demonstrating that the large set probes relevant behavior that is not probed in the small sets previously used. In particular, there are code pairs (see shaded area in SI Fig. S15.1) that have a mutual $\epsilon \leq 0.2$ for the 29 unaries (good agreement), yet have a larger ϵ for the large set (and similarly for $\epsilon \leq 0.06$, that signals excellent agreement). SI Fig. S15.2 shows two examples when this happens: panels (a) and (c) demonstrate that while values for ϵ are small on the 29 crystals of the small set (boxed cases in SI Fig. S15.2(a,c)), other crystals outside this set may contribute to a larger average ϵ . This is typically the case for lanthanides, not included in the earlier 71-crystal benchmark (that stopped at Rn). Moreover, some elements lead to a low ϵ for the unaries, but to a larger ϵ for the oxides (e.g., Cs-Ba-Fr-Ra and Te-I-Xe-Bi-Po-At-Rn for CASTEP).

The overall conclusion of this analysis is that the 71-crystal set of unaries – or the 29-crystal set as its proxy – gives a fair first assessment of the comparative behavior of two DFT codes, while the complete 960-crystal set of unaries and oxides provides a more detailed comparison, both because more elements and more structures per element are included. Conclusions that were based on the benchmarks of Refs. 29, 30 will therefore still hold, yet can be refined. For testing newly developed pseudopotential libraries, a stepwise approach can be implemented: a first quick test on one unary per element will reveal the largest deviations; once these are fixed, the whole dataset of unaries and oxides can be used to hunt for smaller deviations.

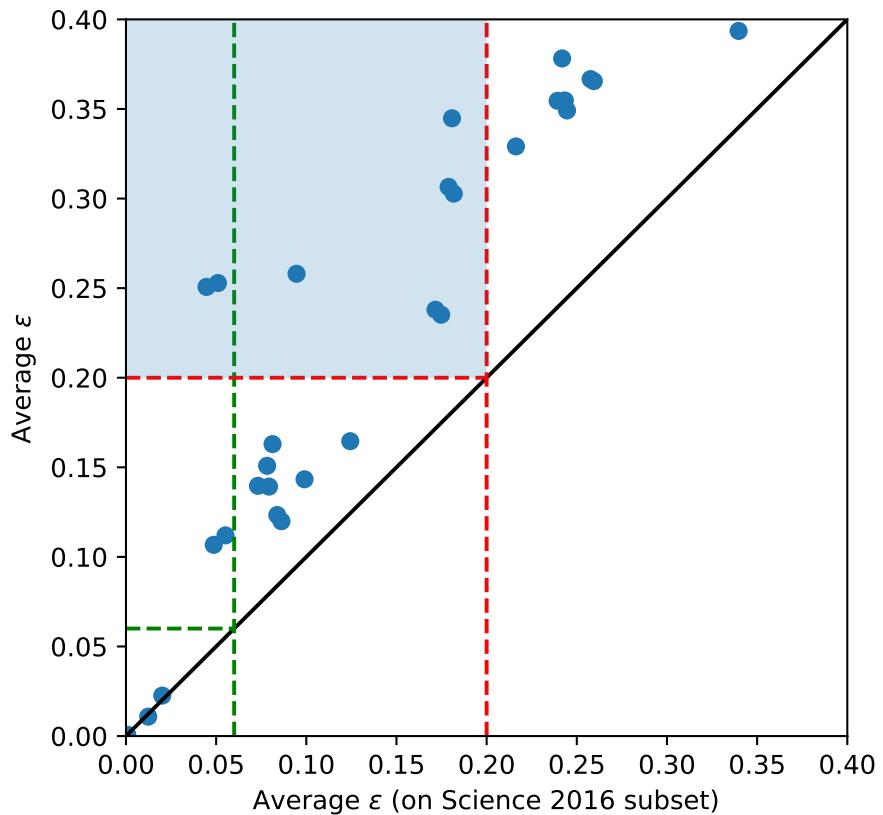


Figure S15.1. Vertical axis: ε averaged over all crystals of the unary and oxide set computed in this work, for all pairs of computational approaches considered in the main text. Horizontal axis: ε averaged over the 29 crystals from these 960 that appear as well in the 71 crystal set from Refs. 29, 30. The horizontal and vertical lines indicate the threshold values for excellent (green) and good (red) agreement, as discussed in the main text. The region above the horizontal red line and to the left of the vertical red line, highlighted by the shaded blue area, indicates pairs that are in good agreement according to the small crystal-structure set, but less so when considering the full set.

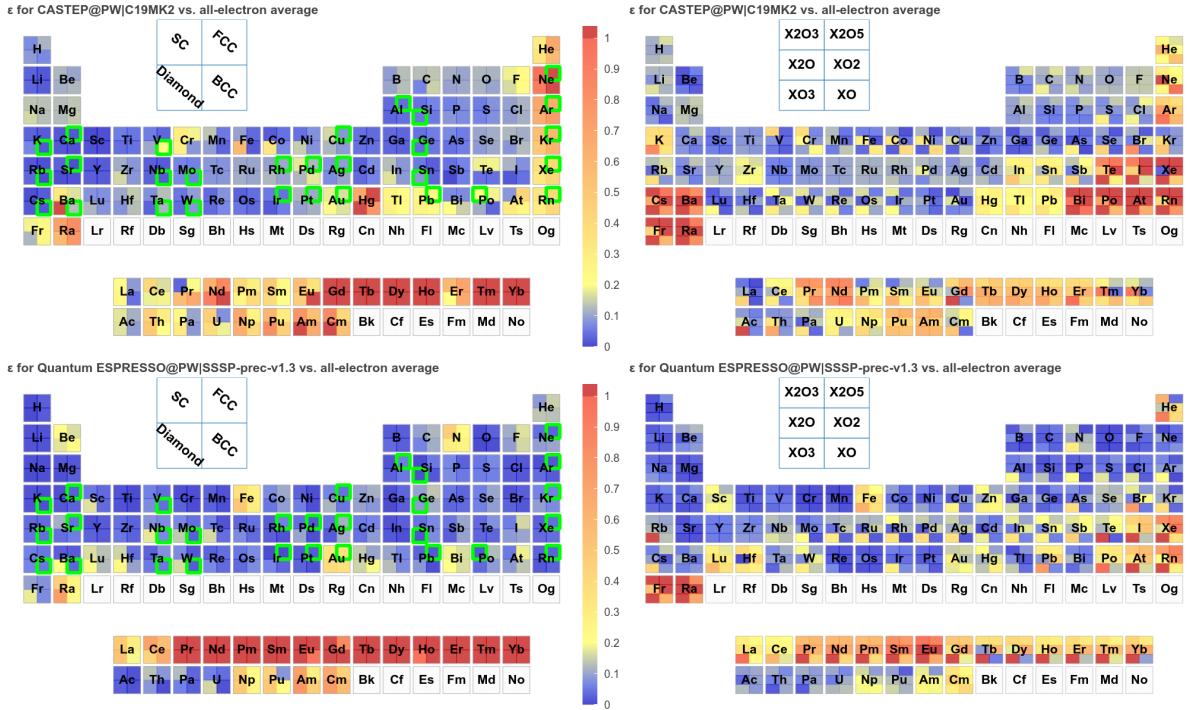


Figure S15.2. Same periodic tables for the comparison metric ε as in SI S9 for codes CASTEP and QUANTUM ESPRESSO, where the 29 crystal structures that appear also in the 71-crystal set of Ref. 29 are highlighted. While the 29 structures display values with excellent agreement, other structures result in a less good agreement. These two examples illustrate some of the cases in SI Fig. S15.1 for which the agreement is very good on the small crystal set, yet less good on the large crystal set.

S16 Additional pseudopotential datasets

This section discusses more in detail some additional datasets obtained using the same codes of the main text, but different basis sets or pseudopotential families, as well as the comparison with earlier versions of pseudopotentials, before those that were optimized here.

S16.1 ABINIT

During the process of computing the oxide verification equations of state using the initial standard norm-conserving scalar-relativistic PBE PseudoDojo (version 0.4)^{68,113}, we observed that the results for around 11 of the pseudopotentials were not in as good agreement with the all-electron equations of state as other elements from the same pseudopotential family. This led to an investigation into possible improvements to the pseudopotentials for Ba, Bi, I, Pb, Po, Rb, Rn, S, Te, Tl, and Xe. With the exception of S, we found that the accuracy of the pseudopotentials is significantly improved by including a projector for the *f* channel. In the original version, indeed, the local part of the pseudopotential was not able to reproduce the all-electron scattering properties of the *f* channel in the empty region.

A particularly severe case is the one of Ba which is shown in SI Fig. S16.1(a) where the pseudized *f* channel (black dashed line) does not reproduce well the all-electron reference. As a consequence, the EOS for the stable BCC phase of Ba is slightly off (SI Fig. S16.1(c)) but the one for BaO₃ is completely wrong, as shown in SI Fig. S16.1(d). This issue is fixed by the additional projector as shown in SI Fig. S16.1(b), which then gives excellent agreement with the all-electron EOS.

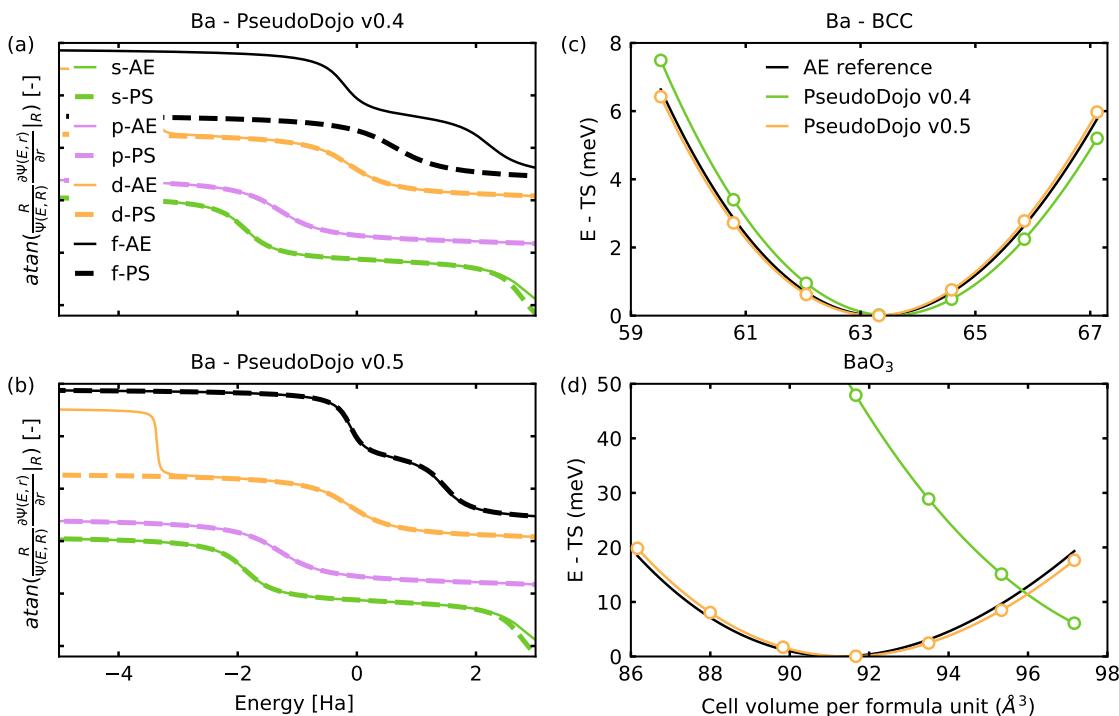


Figure S16.1. Comparison of the scattering properties of the all-electron (AE) atom and the pseudized (PS) Hamiltonian for two different Ba pseudopotentials generated without (with) an *f* projector. The subfigures (a) and (b) show the arctangent of the *l*-dependent logarithmic derivative computed for some *R* greater than the pseudization radius, where ψ is the solution of the non-local radial equation regular at the origin (a) without and (b) with *f* projector. Since the transferability of a pseudopotential is directly related to the capability of reproducing the AE logarithmic derivative over a wide range of energies, the version with *f* projector is expected to provide more accurate results and we provide two equation of state example in (c) for BCC barium and in (d) for BaO₃ where the pseudopotential with the additional projector is more accurate than the one without. This is true for all 10 cases tested in this study.

It should be noted that the inclusion of the *f* projector increases the computational cost associated to the application of the non-local part of the Hamiltonian V_{nl} . This is especially true if the computation of $V_{nl}|\psi\rangle$ is obtained by projecting the wavefunction over spherical Harmonics Y_{lm} . We stress, however, that in ABINIT the projection is implemented by expressing the sum over *m* in terms of Legendre polynomials (useylm = 0 input variable, default option when norm-conserving pseudopotentials are employed). In this case, the computational cost of including an additional projector for *l* is not so high

because, contrary to the case when spherical harmonics are used, the number of floating point operations required to apply a projector does not scale with the total number of magnetic quantum numbers $2l + 1$. This study led to the creation and adoption of 11 new pseudopotentials for Ba, Bi, I, Pb, Po, Rb, Rn, S, Te, Tl, and Xe in a new PseudoDojo (version 0.5). Because these pseudopotentials have not been subject to a convergence study with respect to all-electron results, 20 Hartree were added to the “high” stringency recommended cutoffs from ONCVPSP for safety.

In addition to norm-conserving pseudopotentials, we also investigated the JTH PAW PBE v1.1⁶³ table which has improved versions of the pseudopotentials for H, Li, Si, Cu, Zn, Ga, Cd, Sb, Lu, Os, Ir and Bi with respect to v1.0 used in the previous AiiDA common workflows study. Through testing, it was noted that the recommended kinetic energy cutoff values for these PAW pseudopotentials were not sufficient for the desired level of agreement with all-electron codes, so twice the “high” recommended cutoffs were used. The conjugate-gradient (CG) diagonalization algorithm was used for all calculations with PAW pseudopotentials. Note that there do exist pseudopotentials for lanthanide- and actinide-series elements in this PAW table but they have not been verified against all-electron reference results. Because of this and because the PAW potentials are included primarily as a point of comparison against the norm-conserving PseudoDojo family (which does not provide lanthanide nor actinide elements), these elements are not included in the reported results and analysis.

We therefore compare in SI Fig. S16.2 the performance of the three PBE pseudopotential libraries tested here using ABINIT: the norm-conserving PseudoDojo (version 0.4), the norm-conserving PseudoDojo (version 0.5) and the PAW JTH v1.1 tables. We conclude that PseudoDojo (version 0.5) is the most precise one and is therefore used in the main manuscript.

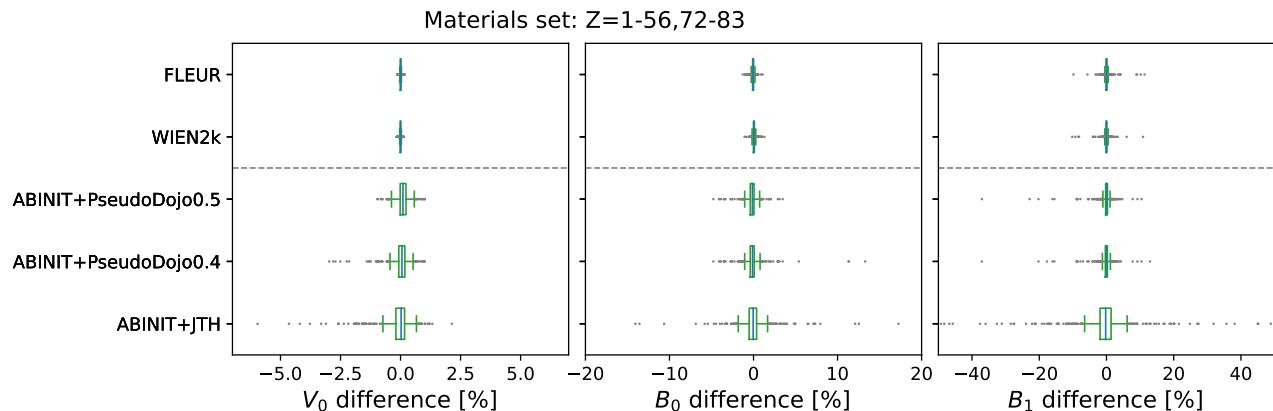


Figure S16.2. Comparison of the three tested PBE pseudopotential tables using the ABINIT software.

In summary, unary and oxide verification results were calculated for ABINIT with three sets of pseudopotentials: JTH PAW PBE v1.1, PseudoDojo norm-conserving standard scalar-relativistic PBE v0.4, and a new PseudoDojo norm-conserving standard scalar-relativistic PBE v0.5 based on v0.4 with improved potentials for the 10 elements listed above. The latter set is reported in the main text of the manuscript.

S16.2 BigDFT

Data production using BigDFT showed clear outliers in comparison to reference codes (FLEUR/WIEN2k) using HGH-K Valence only pseudopotentials. We therefore recalculated the EOS using semicore pseudopotentials for all the crystals for which this type of pseudopotential was available and that showed a difference in volume $\geq 0.2 \text{ \AA}^3$ with respect to the reference AE data. The new calculations resulted, in most cases, in a significant shift towards a closer agreement with the reference AE, as shown in SI Fig. S16.3.

S16.3 CASTEP

A unique feature of CASTEP is that the pseudopotentials are typically generated on-the-fly during the calculations, although file-based potentials are still supported. Each on-the-fly generated (OTFG) potential is defined using a compact configuration string. An OTFG library is a collection of predefined configurations. Many such libraries are built into the CASTEP executable itself, with different focuses. In this study, the pseudopotentials used are provided by the C19 library, except for the lanthanide and actinide elements. For these elements, a new set of pseudopotentials are generated to improve agreement with the all-electron reference data, which had not been available during the development of the original C19 library.

The configuration strings for these elements are tabulated in Table S16.1. Each field is separated by “|”. The first field is the local angular momentum channel. The second is the core radius in atomic units and the next three are the recommended cut off energies (Ha) corresponding to the *Coarse*, *Medium*, and *Fine* settings for CASTEP, which do not affect a calculation if a cut

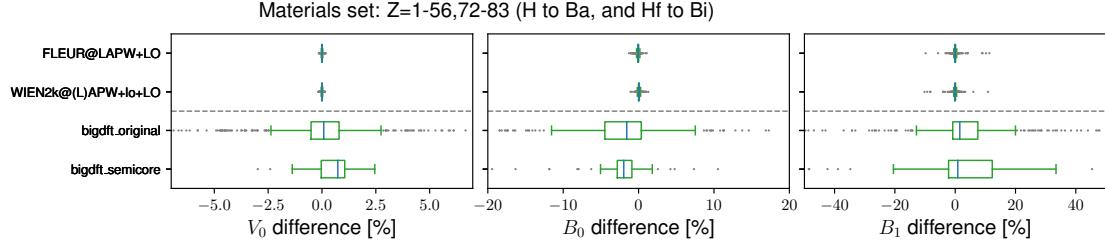


Figure S16.3. Comparative plot displaying the improvement obtained by using HGH-K (Semicore) pseudopotentials over Valence only pseudopotentials.

off energy is specified explicitly (as in this study). This is followed by the orbitals to be pseudized, which are separated by ":". Each orbital is defined by a *nl* number and may have suffixes to indicate what kind of projectors should be used. For example, "60U" indicates that a single ultrasoft projector should be used for the 6s channel (the default is two ultrasoft projectors). The "NN" suffix indicates that two norm-conserving projectors should be included. The "U2U2" suffix indicates two ultrasoft projectors each with a core radius of 2.0. The "L" suffix pins the local channel, and the "P" suffix indicates that the pseudized channel is not represented by an explicit projector or a local channel. The parameter "qc" inside the brackets controls the smoothness of the potential. Occupations of atomic states can be further modified by settings inside a curly bracket.

Element	Old settings	New settings
La	2 2.3 5 6 7 50U:60:51:52 (qc=4.5)	2 2.3 5 6 7 50U:60:51:52:43{4f0.1} (qc=4.5)
Ce	2 2.1 9 10 11 50U:60:51:43:52L (qc=6)	2 2.2 8 9 10 50U:60:51:52:43{5d0.1} (qc=4.5)
Pr	2 2.1 9 10 11 50U:60:51:43 (qc=6)	2 2.1 10 12 13 50U:60:51:52:43{5d0.1} (qc=5)
Nd	2 2.1 9 10 11 50U:60:51:43 (qc=6)	2 2.1 10 12 13 50U:60:51:52:43{5d0.1} (qc=5)
Pm	2 2.1 10 12 13 50U:60:51:43 (qc=6)	2 2.1 8 9 11 50U:60:51:52:43{5d0.1,4f4} (qc=5.5)
Sm	2 2.1 10 12 13 50U:60:51:43 (qc=6)	2 2.1 9 10 12 50U:60:51:52:43{5d0.1,4f5} (qc=5.5)
Eu	2 2.1 10 12 13 50U:60:51:43 (qc=6)	2 2.1 9 10 12 50U:60:51:52:43{5d0.1,4f6} (qc=5.5)
Gd	2 2.1 10 12 13 50U:60:51:52L:43 (qc=6)	3 2.1 9 10 12 50U:60:51:52:43 (qc=5.5)
Tb	2 2.1 10 12 13 50U:60:51:43 (qc=6)	2 2.2 12 13 15 50U:60:51:52:43{5d0.1} (qc=5)
Dy	2 1.9 12 14 16 50U:60:51:43 (qc=6.5)	2 2.0 12 13 15 50U:60:51:52:43{5d0.1} (qc=6.5)
Ho	2 1.9 12 14 16 50U:60:51:43 (qc=6.5)	2 2.0 12 13 15 50U:60:51:52:43{5d0.1} (qc=6.5)
Er	2 2.1 10 12 13 50U:60:51:43{6s0.5} (qc=6)	2 2.1 10 12 13 50U:60:51:52:43{6s0.1,5d0.1} (qc=6)
Tm	2 2.1 10 12 13 50U:60:51:43{4f12} (qc=6)	2 2.1 10 12 13 50U:60:51:52:43{5d0.1,4f12} (qc=6)
Yb	2 2.1 10 12 13 50U:60:51:43{4f13} (qc=6)	2 2.1 10 12 13 50U:60:51:52:43{5d0.1,4f13} (qc=6)
Ac	2 2.5 5 6 7 60U:70NN:61:62L	2 2.4 7 7 9 60U:70U2U2:61:62:53{6d0.1,5f0.1} (qc=5)
Th	2 2.5 7 7 9 60U:70NN:61:62	2 2.2 7 7 9 60U:70U2U2:61:62:53{5f0.1} (qc=5)
Pa	2 2.1 9 10 11 60U:70:61:53:62P (qc=6)	2 2.2 8 9 10 60U:70U2U2:61:62:53 (qc=5)
U	2 2.1 10 12 13 60U:70:61:53:62P (qc=6)	2 2.2 8 9 10 60U:70U2U2:61:62:53 (qc=5)
Np	2 2.1 10 12 13 60U:70:61:53:62P (qc=6)	2 2.2 9 10 12 60U:70U2U2:61:62:53 (qc=5)
Pu	2 2.1 10 12 13 60U:70:61:53:62P{6d1,7s1} (qc=6)	2 2.2 9 10 12 60U:70U2U2:61:62:53{6d0.1} (qc=5.5)
Am	2 2.1 10 12 13 60U:70:61:53 (qc=6)	2 2.2 9 10 12 60U:70U2U2:61:62:53{6d0.1} (qc=5.5)
Cm	2 2.1 10 12 13 60U:70:61:53:62L (qc=6)	2 2.2 9 10 12 60U:70U2U2:61:62:53 (qc=5.5)

Table S16.1. Configuration strings of the on-the-fly pseudopotential generation before and after the update for the lanthanide and actinide elements involved in this study.

Examples EOS curves for La and Ce are displayed in SI Fig. S16.4, showing that the updated pseudopotentials agree better with the all electron data. The improvements may vary for other elements with *f* electrons. One should note that a better fit to the all electron results does not necessarily mean smaller errors compared to the experimental results. This is due to the inherent self-interaction errors in semi-local DFT that are present in the description of *f* electrons.

For some elements, the improvements are achieved through the inclusion of additional orbital states and l channels/projectors such as La, where previously the 4*f* channel was neglected. In other cases, the reference atomic calculations include partially occupied atomic states that are otherwise empty in the original configurations. For example {5d0.1} adds 0.1 electrons to the 5*d* channel. In some cases, the updated potentials contain increased core radii and are made softer via a decreased "qc" value as far as possible. These modifications are applied consistently for elements that neighbors to each other in the periodic table. We also want to emphasize that the settings used in this study are one-shot updates based on the C19 library, rather than the outcomes of iterative optimizations. With the help of the automated test framework and publicly available all-electron reference data, it should be easy to adjust the strings and test them rigorously for further improvements, as required.

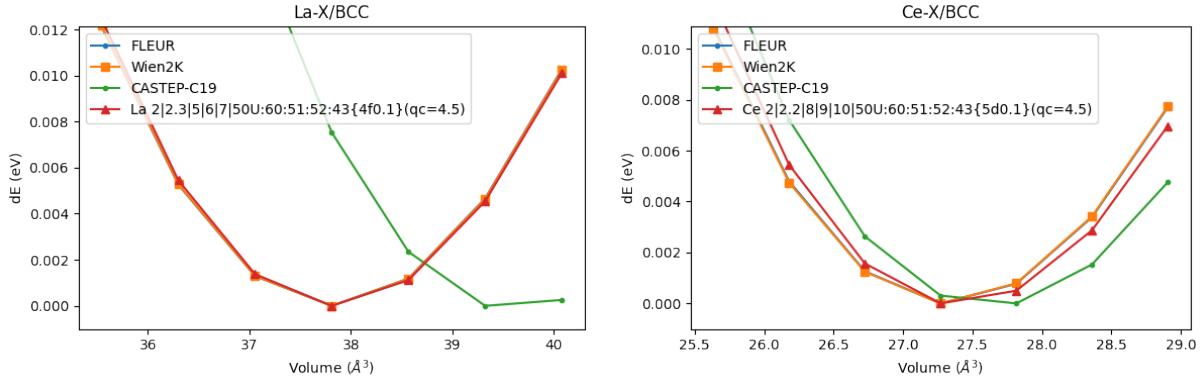


Figure S16.4. Comparing EOS curves of La and Ce in the BCC configuration. The updated pseudopotentials agree better with the all-electron data compared with those from the C19.

S16.4 Quantum ESPRESSO

SSSP⁷⁵ is a library of pseudopotentials that undergoes rigorous verification and ranking procedures. It contains two distinct families of pseudopotentials. The first family (“efficiency”) is composed of relatively soft pseudopotentials that are still sufficiently precise for use in high-throughput calculations. The second family (“precision”) contains pseudopotentials that are extremely precise with respect to all-electron references, even if more computationally expensive. In our QUANTUM ESPRESSO calculations, aiming at high precision, we have therefore used the SSSP “precision” library. Version 1.1.2 of SSSP, available before this work, was verified only on the unary configurations presented in Ref. 29, 30. With the additional results of this work, we have identified that certain pseudopotentials were not the best selection, and that more precise pseudopotentials are available in libraries that were not included in the previous generation of the SSSP library.

Element	SSSP precision v1.1.2	SSSP precision v1.3
Te	6 US GBRV v1 uspp	6 US PSL v1.0.0-low ld1
Na	9 NC PD v4-std oncvpsp3	9 PAW PSL v1.0.0-low ld1
Cu	19 NC PD v4-std oncvpsp3	11 PAW PSL v1.0.0-low ld1
Cs	9 US GBRV v1 uspp	9 NC PD v4-str oncvpsp3
Cd	12 US PSL v0.3.1 ld1	20 PAW PSL v1.0.0-high ld1
Ba	10 PAW PSL v1.0.0-high ld1	10 NC PD v5 oncvpsp4
As	5 US PSL v0.2 ld1	15 NC PD v4-std oncvpsp3
I	17 PAW PSL v1.0.0-high ld1	17 NC PD v4-std oncvpsp3
Hg	20 NC SG15 v0 oncvpsp3	12 US GBRV v1 uspp
Ne	8 NC SG15 v0 oncvpsp3	8 PAW PSL v1.0.0-high ld1
Ar	8 NC SG15 v0 oncvpsp3	8 PAW PSL v1.0.0-high ld1
Kr	8 NC SG15 v0 oncvpsp3	18 PAW PSL v1.0.0-high ld1
Xe	18 NC SG15 v0 oncvpsp3	18 PAW PSL v1.0.0-high ld1
Rn	18 NC SG15 v0 oncvpsp3	18 PAW PSL v1.0.0-high ld1
Ir	15 NC GBRV v1 uspp	31 US PSL v1.0.0-high ld1

Table S16.2. List of pseudopotentials modified between SSSP v1.1.2 and SSSP v1.3. Each element in the table is composed of 5 strings separated by a | symbol, respectively: the number of electrons in the valence (Z), the type of pseudopotential (NC: norm-conserving, PAW: projector-augmented wave, US: ultrasoft), the source library (SG15: from Ref. 114, PSL: PSlibrary⁸², GBRV: from Ref. 115, PD: PseudoDojo⁶⁸), an internal version number that identifies the pseudopotential inside the given source library and the code used to generate them (oncvpsp3, oncvpsp4: version 3 and version 4 of the ONCVPSP code¹¹³, ld1: the ld1.x code of Quantum ESPRESSO⁵¹, uspp: the UltraSoft PseudoPotential (USPP) generation code¹¹⁶).

The pseudopotentials that were updated are summarized in Table S16.2. In particular, the pseudopotentials of the noble gases (Ne, Ar, Kr, Xe, Rn) were previously obtained using the SG15¹¹⁴ library; in SSSP v1.3, we have replaced them with “PAW-high” pseudopotentials from the PSlibrary⁸² based on our more recent verification results as they provide significantly better agreement with the all-electron reference EOS curves. An example of the improvement in the case of NeO_2 is shown in SI Fig. S16.5.

The pseudopotentials of I, Hg were obtained from SG15¹¹⁴ as well. In SSSP v1.3, the pseudopotential of I is replaced with

the one from PseudoDojo library and the pseudopotential of Hg is replaced with the one from GBRV¹¹⁵ library. It is not only because these two pseudopotentials from SG15 library are less precise, but also because these two pseudopotentials lead to the electronic step convergence issue during the calculation. Using new pseudopotentials makes all equation of state calculation finished without issues.

The SSSP v1.1.2 libraries for Ir have been updated with the latest pseudopotential from Pslibrary US v1.0.0 in the “high” family⁸². It should be noted that the original Ir pseudopotential from the GBRV library contains a ghost state at 10eV.

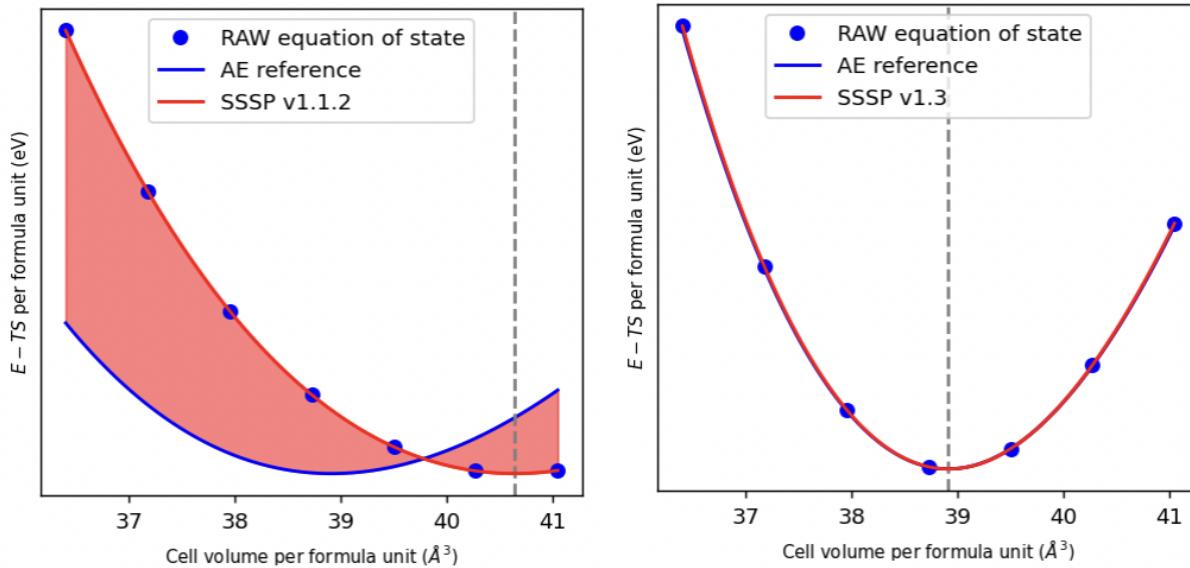


Figure S16.5. Left: EOS of NeO_2 using the Ne pseudopotential from SSSP v1.1.2. Right: EOS of NeO_2 using the Ne pseudopotential from SSSP v1.3.

For Ba, in SSSP v1.3 we select the new pseudopotential from the PseudoDojo v0.5 library⁶⁹ (generated in the context of this work, see SI Sec. S16.1) that includes an f projector.

For Te, Na, and Cu we consider in SSSP v1.3 the PAW or ultrasoft pseudopotentials from the Pslibrary⁸² in the “low” family. These pseudopotentials have fewer semicore states and larger cut-off radii, making them possibly less accurate, but optimized for lower kinetic energy cut-offs. Tests on ten configurations, including oxides, revealed that some pseudopotentials from the “low” family are actually even more accurate than those from the “high” family (the latter have more semicore states and smaller cut-off radii). We note that Pslibrary suggests to use “high” pseudopotentials only for special applications, while “low” ones can yield sufficient precision for regular calculations. As an example of the results with the new pseudos, we show the results for TeO_2 for the two versions of SSSP in SI Fig. S16.6.

Starting from v1.3, the SSSP library also includes pseudopotentials for actinides (Th-Lr) developed in Ref. 81 as well as the pseudopotentials for Ac, At, Ra, and Fr from the “high” family of Pslibrary⁸².

Finally, in the case of Cs, As, while the test of the SSSP v1.1.2 pseudopotentials on unaries was resulting in good-quality EOS curves, we obtained significant disagreements for oxides. We therefore replaced the corresponding pseudopotentials with others that provided more precise agreements. We illustrate the improvements in the case of Cs in SI Fig. S16.7.

S16.5 VASP

The lanthanides Pr, Nd, Pm, Sm, Eu, Gd, Tb, Dy, Ho, Er, Tm, and Yb were updated during the preparation of the data, since the initial ones showed quite significant deviations from the all-electron results. For elements containing $4f$ electrons, VASP recommends to use potentials with the f electrons placed in the frozen core, to avoid the well known self-interaction errors resulting from DFT. VASP therefore provides well-tested potentials with frozen f electrons for the lanthanides, with a valency of 2 or 3 (for Er, Eu, and Yb, both valences are available as separate potentials). Since this study has settled on treating f electrons explicitly as valence, we used the potentials that place all f electrons in the valence. Using semi-local functionals, these potentials lead to significant over-binding and too small unit cells (compared to experiment as well as compared to PAW potentials that place the f electrons in the core). However, the aim of this work is code comparison primarily. The lanthanide potentials used in this work have been generated by G. Kresse. They are using a much smaller core radius of 2.2 a.u. than the previous versions. Additionally, the reference electronic configuration was altered, by placing 0.5 electrons (instead of 1

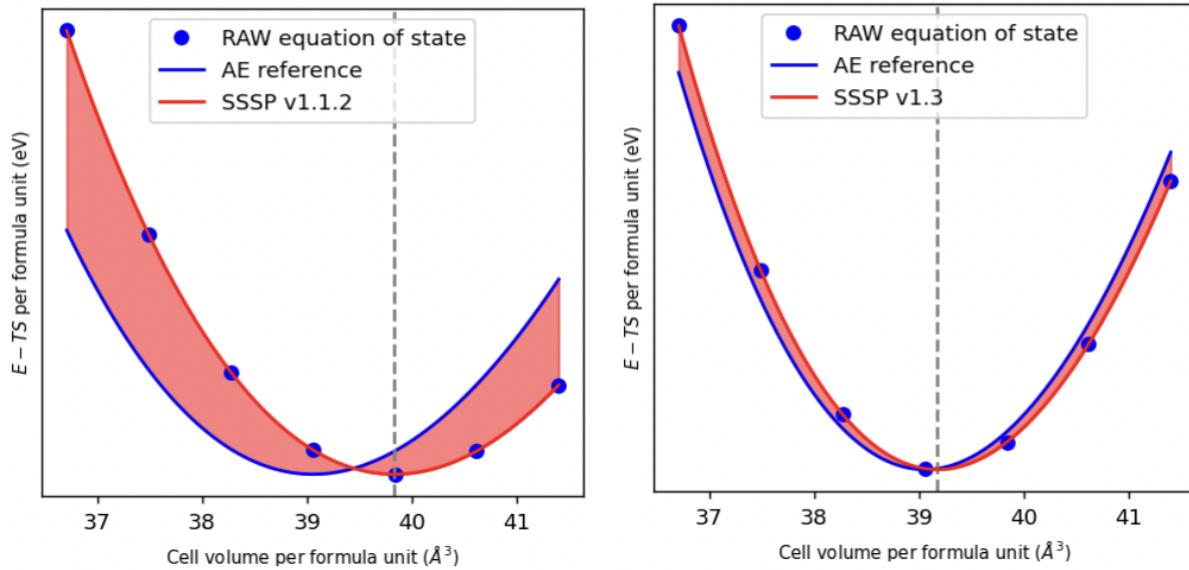


Figure S16.6. Left: EOS of TeO_2 using the Te pseudopotential from SSSP v1.1.2. Right: EOS of TeO_2 using the Te pseudopotential from SSSP v1.3.

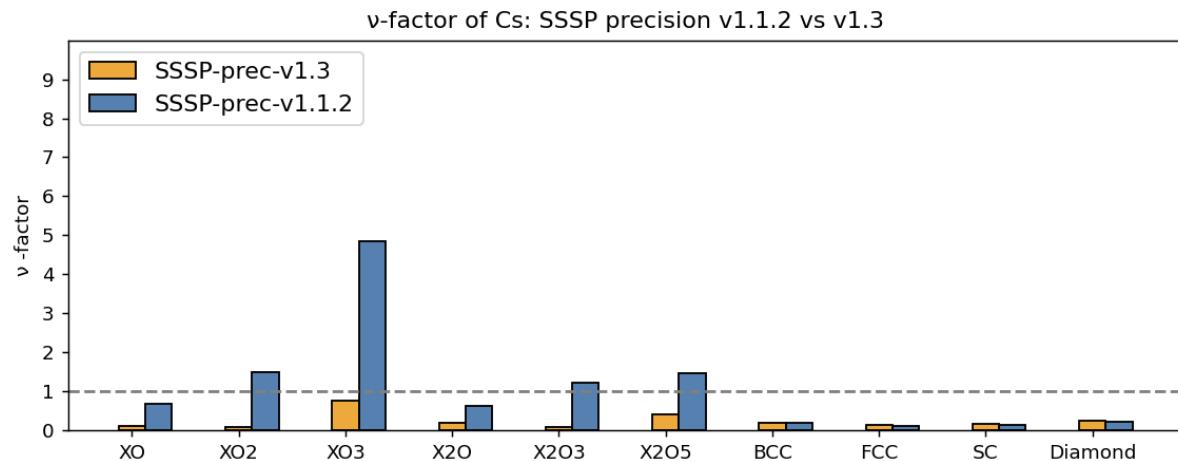


Figure S16.7. Comparison (via the v metric) between QUANTUM ESPRESSO results and the all-electron reference from WIEN2k for the different crystal structures of Cs, revealing that while the old pseudopotential was already quite precise for unaries, only the new pseudopotential in SSSP v1.3 (from the PseudoDojo library) provides high precision results for oxides.

electron) from the $4f$ shell into the $5d$ shell. Generally, two projectors were used for the f shells. Due to contracting f shells towards the right of the series (Tm, Er, and Yb) a third projector was found to be necessary to obtain an accurate description of the f scattering properties. The new lanthanide potentials will be released on the VASP portal in the 6.4 PBE PAW potential set with an $_h$ suffix. For future reference the specific potential mapping that was used is RECOMMENDED_ACWF_LANTH, where LANTH indicates that the new lanthanide potentials have been used.

For the unaries we have compared the elevated settings optimized for this study with the default settings of VASP (keeping only $\text{LMAXMIX} = 6$) and the chosen recommended GW potential set with the recommended PBE potential set. The resulting ϵ metrics are plotted in SI Fig. S16.8.

In the first panel of SI Fig. S16.8 we used the recommended PAW potentials for PBE (as given on the VASP website, for the lanthanides the new PAW potentials were used as detailed above), with the default cutoffs (dependent on the system) and precision setting ($\text{PREC} = \text{Normal}$). In panel (b) we used the same potentials, but $\text{PREC} = \text{Accurate}$ and a plane wave cutoff of 800 eV alongside the other settings detailed in section S13.9, while panel (c) corresponds to the final results presented in the

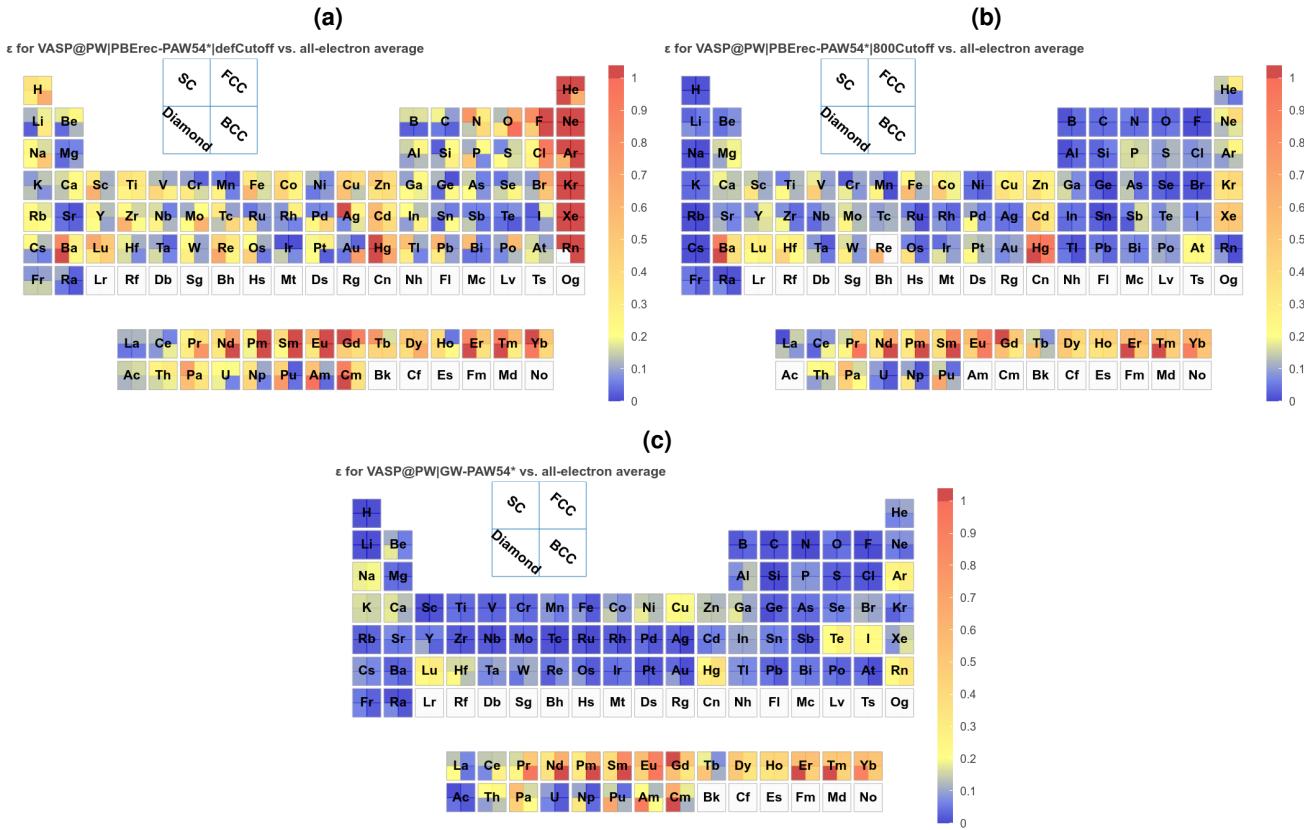


Figure S16.8. Value of the comparison metric ϵ for the unaries using three different settings for VASP with the averaged AE results as reference. The datasets in (a) and (b) use the recommended PBE potentials. Computational parameters are mostly left at default values in (a), while (b) increases the energy cutoff to 800 eV and adopts the other parameters described in section S13.9. In (c) we show the final dataset with recommended GW potentials, if available, and 1000 eV energy cutoff.

paper using the final settings (1000 eV) and the recommended GW potentials.

Clearly, changing the default parameters substantially improves the agreement with the AE reference calculations. The noble gases in particular are dramatically improved by using the more accurate settings. But many other weakly bonded structures are improved as well. If then also the GW potentials are used, some elements do improve again by quite a lot (Ba, Hg, Xe, Cd, . . .), while others improve only slightly (e.g., P, S, As, . . .) and some get slightly worse (e.g., Na, K, Rn, . . .). For the oxides, the improvement is significant for materials with short oxygen bonds, when using the O_h_GW potential instead of the standard PBE one (not shown).

S17 Consistency check with previous benchmarks for the all-electron data

All calculations in this work have been run independently and from scratch by expert users of the respective codes, following a strict protocol. As the all-electron calculations WIEN2k and FLEUR serve as references here, it is useful to compare them to results obtained for the same crystals in an earlier benchmark^{29,30}. Table S17.1 shows this comparison for the parameters of the Birch-Murnaghan EOS: both for FLEUR and WIEN2k, there are only small relative differences between the parameters obtained in Refs. 29,30 and the ones obtained in the present work. The only crystal with significant relative deviations is Ne (FCC), which is not surprising given its very shallow equation of state (small value of the bulk modulus, which is known⁹⁸ to lead to a large uncertainty in the volume). This demonstrates good agreement between the previous and current data, obtained from independent calculations.

A different view on the same data is represented in SI Fig. S17.1. It shows on the horizontal axis the difference between the FLEUR and WIEN2k results for these 29 crystals, expressed by the Δ metric, as obtained in Refs. 29,30. On the vertical axis, the Δ metric for the same crystal and the same two codes is shown, now using the data obtained in the present work. Although the range on the horizontal axis does not extend much beyond the threshold of good agreement of $\Delta = 1$ meV/at that was used in Refs. 29,30, the range on the vertical axis is significantly smaller.

The conclusion of this analysis is that although the agreement between the all-electron codes FLEUR and WIEN2k was already very good in Refs. 29,30 ($\Delta \leq 1$ meV/at), and although the relative differences in the observable properties that can be derived from the EOS are very small (Tab. S17.1), the agreement between the two all-electron codes in the present work is definitely even better than it was in previous works (SI Fig. S17.1).

Table S17.1. Table comparing the V_0 , B_0 and B_1 parameters for the subset of 29 structures of those suggested in Ref. 29,30 that are also present in our set of unaries, namely those that have a cubic FCC, BCC, SC or Diamond structure and that were treated without spin polarization. In particular, Fe(BCC), Mn(FCC), Ni(FCC) and Cr(BCC) have not been included because they are treated including spin polarization in Ref. 29,30; all other structures not shown here were not considered in a cubic structure in Ref. 29,30. The table reports the absolute percentage error on each parameter with respect to the data for the same code and the same crystal reported in Ref. 29,30.

Element	Structure	FLEUR			WIEN2k		
		V_0 error [%]	B_0 error [%]	B_1 error [%]	V_0 error [%]	B_0 error [%]	B_1 error [%]
Ag	FCC	0.01	1.59	1.98	0.06	0.86	6.97
Al	FCC	0.01	1.22	5.69	0.10	0.72	1.15
Ar	FCC	0.55	5.61	29.04	0.28	1.28	2.71
Au	FCC	0.20	1.62	10.40	0.02	0.27	2.95
Ba	BCC	0.16	2.18	9.86	0.26	0.05	30.67
Ca	FCC	0.11	0.24	5.91	0.00	1.50	0.59
Cs	BCC	0.19	0.39	4.82	0.20	1.45	38.97
Cu	FCC	0.16	0.06	0.71	0.05	0.10	4.14
Ge	Diamond	0.09	0.34	2.68	0.02	0.31	2.60
Ir	FCC	0.00	0.82	0.07	0.03	0.12	1.38
K	BCC	0.15	0.47	0.47	0.14	0.14	21.68
Kr	FCC	0.37	4.43	38.54	0.62	5.21	35.59
Mo	BCC	0.09	0.03	5.32	0.04	0.12	2.87
Nb	BCC	0.12	0.59	12.32	0.03	0.97	3.71
Ne	FCC	2.69	0.41	41.55	0.29	13.74	99.20
Pb	FCC	0.15	0.45	1.86	0.09	0.16	4.52
Pd	FCC	0.03	0.77	3.34	0.08	0.35	0.65
Po	SC	0.03	0.60	9.91	0.04	0.13	0.39
Pt	FCC	0.19	1.19	6.52	0.08	0.52	0.02
Rb	BCC	0.07	0.86	0.83	0.39	0.55	53.53
Rh	FCC	0.07	0.63	0.91	0.07	0.34	2.65
Rn	FCC	0.04	7.79	14.45	0.48	4.23	34.33
Si	Diamond	0.04	0.03	0.68	0.03	0.02	0.07
Sn	Diamond	0.01	0.47	3.56	0.06	0.71	4.94
Sr	FCC	0.71	3.27	37.98	0.68	3.43	5.89
Ta	BCC	0.02	1.62	8.91	0.03	1.11	1.17
V	BCC	0.07	1.20	1.87	0.07	0.16	2.31
W	BCC	0.04	0.72	6.96	0.03	0.04	2.63
Xe	FCC	0.11	4.06	6.18	0.36	2.00	12.28

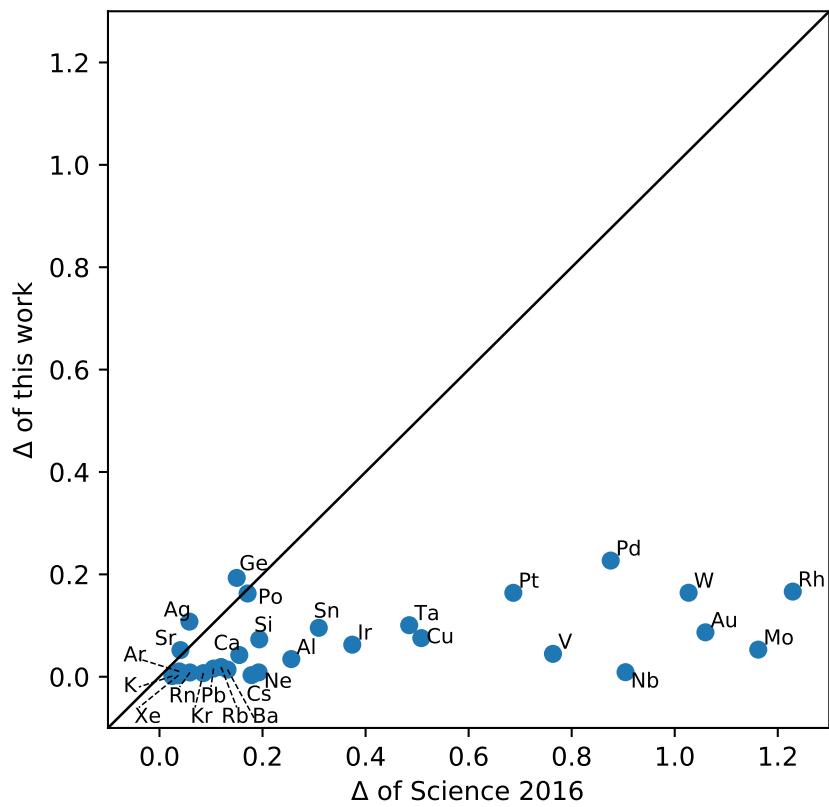


Figure S17.1. Correlation of the Δ metric on the 29 crystals listed in Tab. S17.1. x axis: Δ metric for these crystals between FLEUR and WIEN2k, using the results from Refs. 29, 30. y axis: same metric Δ for the same crystals and the same two codes, but using data from the present work. The black solid line indicates $y = x$. We note that $\Delta \leq 1$ meV/atom was considered in Ref. 29 to indicate a good agreement.

S18 Discrepancies of formation energies computed from the current dataset

In SI Fig. S18.1 we report a histogram of the difference of the formation energy obtained using the data for the two all-electron codes FLEUR and WIEN2k, computed from the minimum-energy value of the EOS data curves. The histogram is obtained considering the formation energy of all X_nO_m oxides, using the lowest-energy unary of element X and of oxygen as the two endpoints (in the case of oxygen, the lowest-energy non-magnetic unary in our dataset is the simple cubic structure). The majority of the datapoints are in the visible x axis range, i.e., with an (absolute) discrepancy smaller than 50 meV/atom. Nevertheless, several outliers are present: 52 out of the 576 materials considered have a discrepancy larger than 50 meV/atom. The most outstanding outliers are AtO_3 , PoO_3 , BiO_3 , Am_2O , Pu_2O , AmO and PuO . They have a discrepancy larger than 1 eV/atom. For compounds containing Am and Pu, the discrepancy is also due by the fact that the unary identified as having lowest energy is different between the two codes.

As we discuss in the main text, the reason for this discrepancy is that our workflows have been designed to guarantee consistent simulation parameters among calculations for a given material at different volumes. However, when considering different structures, we did not enforce any consistency between simulation parameters, e.g., the choices of atomic radii for the AE codes might be different in different systems. Especially, changes to the core/valence separation from structure to structure may lead to larger discrepancies in this comparison, because the differing relativistic descriptions of core and valence electrons lead to different energy contributions. Therefore, we recommend not to use our dataset to generate plots like the one of SI Fig. S18.1 or, more generally, to avoid performing data analysis that considers energy differences between different structures. In this case, instead, one should design new appropriate workflows that can ensure the consistency of simulation parameters among all relevant calculations.

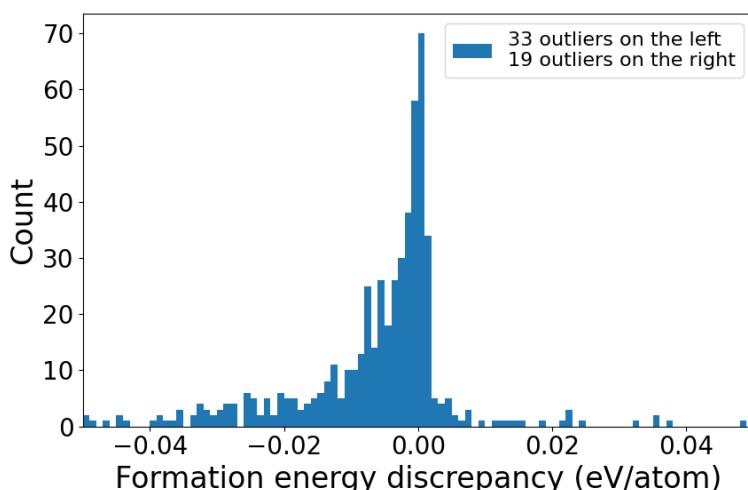


Figure S18.1. Histogram of the discrepancy between the formation energy obtained from our reference dataset for the two codes FLEUR and WIEN2k. We note that no correction is applied to the formation energies (as it is typically done, for instance, for oxygen²⁵). However, since we are only considering differences in formation energies between two codes, these corrections cancel out so they do not need to be considered explicitly. The number of outliers outside of the visible x axis range is reported in the top right corner of the figure.