Reproducibility in density-functional theory calculations of solids

Kurt Lejaeghere,[1*] Gustav Bihlmayer,[2] Torbjörn Björkman,[3,4] Peter Blaha,[5] Stefan Blügel,[2] Volker Blum,[6] Damien Caliste,[7,8] Ivano E. Castelli,[9] Stewart J. Clark,[10] Andrea Dal Corso,[11] Stefano de Gironcoli,[11] Thierry Deutsch,[7,8] John Kay Dewhurst,[12] Igor Di Marco,[13] Claudia Draxl,[14,15] Marcin Dułak,[16] Olle Eriksson,[13] José A. Flores-Livas,[12] Kevin F. Garrity,[17] Luigi Genovese,[7,8] Paolo Giannozzi,[18] Matteo Giantomassi,[19] Stefan Goedecker,[20] Xavier Gonze,[19] Oscar Grånäs,[13,21] E. K. U. Gross,[12] Andris Gulans,[14,15] François Gygi,[22] D. R. Hamann,[23,24] Phil J. Hasnip,[25] N. A. W. Holzwarth,[26] Diana Iușan,[13] Dominik B. Jochym,[27] François Jollet,[28] Daniel Jones,[29] Georg Kresse,[30] Klaus Koepernik,[31,32] Emine Küçükbenli,[11] Yaroslav O. Kvashnin,[13] Inka L. M. Locht,[13,33] Sven Lubeck,[14] Martijn Marsman,[30] Nicola Marzari,[9] Ulrike Nitzsche,[31] Lars Nordström,[13] Taisuke Ozaki,[34] Lorenzo Paulatto,[35] Chris J. Pickard,[36] Ward Poelmans,[1] Matt I. J. Probert,[25] Keith Refson,[37,38] Manuel Richter,[31,32] Gian-Marco Rignanese,[19] Santanu Saha,[20] Matthias Scheffler,[15,39] Martin Schlipf,[22] Karlheinz Schwarz,[5] Sangeeta Sharma,[12] Francesca Tavazza,[17] Patrik Thunström,[40] Alexandre Tkatchenko,[15] Marc Torrent,[28] David Vanderbilt,[23] Michiel J. van Setten,[19] Veronique Van Speybroeck,[1] John M. Wills,[41] Jonathan R. Yates,[29] Guo-Xu Zhang,[42] Stefaan Cottenier[1,43*]

[1]Center for Molecular Modeling, Ghent University, Technologiepark 903, BE-9052 Zwijnaarde, Belgium,

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

[3]Department of Physics, Åbo Akademi, FI-20500 Turku, Finland,

[4]COMP/Department of Applied Physics, Aalto University School of Science, P.O. Box 11100, FI-00076 Aalto, Finland,

[5]Institute of Materials Chemistry, Vienna University of Technology, Getreidemarkt 9/165-TC, A-1060 Vienna, Austria,
[6]Department of Mechanical Engineering and Materials Science, Duke University, Durham, NC 27708, USA,
[7]Université Grenoble Alpes, INAC-SP2M, L_Sim, F-38042 Grenoble, France,
[8]CEA, INAC-SP2M, L_Sim, F-38054 Grenoble, France,
[9]Theory and Simulation of Materials (THEOS), and National Centre for Computational Design and Discovery of Novel Materials (MARVEL), École Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland,
[10]Department of Physics, University of Durham, Durham DH1 3LE, United Kingdom,
[11]International School for Advanced Studies (SISSA) and DEMOCRITOS IOM-CNR Trieste, Via Bonomea 265, I-34136 Trieste, Italy,
[12]Max-Planck-Institut für Mikrostrukturphysik, Weinberg 2, D-06120 Halle, Germany,
[13]Department of Physics and Astronomy, Division of Materials Theory, Uppsala University, PO Box 516, SE-75120 Uppsala, Sweden,
[14]Institut für Physik and IRIS Adlershof, Humboldt-Universität zu Berlin, Zum Großen Windkanal 6, D-12489 Berlin, Germany,
[15]Fritz-Haber-Institut der Max-Planck-Gesellschaft, Faradayweg 4-6, D-14195 Berlin, Germany,
[16]Center for Atomic-scale Materials Design, Department of Physics, Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark,
[17]Material Measurement Laboratory, National Institute of Standards and Technology, 100 Bureau Drive, Stop 8553, Gaithersburg, MD 20899, USA,
[18]Department of Chemistry, Physics, and Environment, University of Udine, via delle Scienze 208, I-33100 Udine, Italy,
[19]Institute of Condensed Matter and Nanosciences - NAPS, Université catholique de Louvain, Chemin des étoiles 8, B-1348 Louvain-la-Neuve, Belgium,
[20]Institut für Physik, Universität Basel, Klingelbergstr. 82, CH-4056 Basel, Switzerland,
[21]School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138, USA,
[22]Department of Computer Science, University of California Davis, Davis, CA 95616, USA,
[23]Department of Physics and Astronomy, Rutgers University, Piscataway, NJ 08854-8019, USA,
[24]Mat-Sim Research LLC, P.O. Box 742, Murray Hill, NJ 07974, USA,
[25]Department of Physics, University of York, Heslington, York YO10 5DD, United Kingdom,
[26]Department of Physics, Wake Forest University, Winston-Salem, NC 27109, USA,
[27]Scientific Computing Department, Science and Technology Facilities Council, Rutherford Appleton Laboratory, Didcot OX11 0QX, United Kingdom,
[28]CEA, DAM, DIF, F-91297 Arpajon, France,
[29]Department of Materials, University of Oxford, 16 Parks Road, Oxford OX1 3PH, United Kingdom,

[30]University of Vienna, Faculty of Physics and Center for Computational Materials Science, Sensengasse 8/12, A-1090 Vienna, Austria,
[31]IFW Dresden e.V., P.O. Box 270 116, D-01171 Dresden, Germany,
[32]Dresden Center for Computational Materials Science (DCMS), TU Dresden, D-01069 Dresden, Germany,
[33]Institute for Molecules and Materials, Radboud University, Heyendaalseweg 135, 6525 AJ Nijmegen, the Netherlands,
[34]Institute for Solid State Physics, The University of Tokyo, Kashiwa 277-8581, Japan,
[35]Institut de Minéralogie, de Physique des Matériaux, et de Cosmochimie (IMPMC), Sorbonne Universités, UPMC University Paris 06, UMR CNRS 7590, Muséum National d'Histoire Naturelle, IRD UMR 206, 4 Place Jussieu, F-75005 Paris, France,
[36]Department of Materials Science & Metallurgy, University of Cambridge, 27 Charles Babbage Road, Cambridge CB3 0FS, United Kingdom,
[37]Department of Physics, Royal Holloway, University of London, Egham TW20 0EX, United Kingdom,
[38]ISIS Facility, Science and Technology Facilities Council, Rutherford Appleton Laboratory, Didcot OX11 0QX, United Kingdom,
[39]Department of Chemistry and Biochemistry and Materials Department, University of California, Santa Barbara, CA 93106-5050, USA,
[40]Institute for Solid State Physics, Vienna University of Technology, A-1040 Vienna, Austria,
[41]Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA,
[42]Institute of Theoretical and Simulational Chemistry, Academy of Fundamental and Interdisciplinary Sciences, Harbin Institute of Technology, Harbin 150080, People's Republic of China,
[43]Department of Materials Science and Engineering, Ghent University, Technologiepark 903, BE-9052 Zwijnaarde, Belgium

*To whom correspondence should be addressed;
E-mail: kurt.lejaeghere@ugent.be, stefaan.cottenier@ugent.be.

The widespread popularity of density-functional theory has given rise to a vast range of dedicated codes to predict molecular and crystalline properties. However, each code implements the formalism in a different way, raising questions on the reproducibility of such predictions. We report the results of a community-wide effort that compares 15 solid-state codes using 40 different potentials or basis set types assessing the quality of the Perdew-Burke-Ernzerhof equations of state for 71 elemental crystals. We conclude that pre-

dictions from recent codes and pseudopotentials agree very well, with pairwise differences comparable to those between different high-precision experiments. This does not hold for slightly older methods. The provided benchmark now presents a solid platform for users and developers to document precision in new applications and methodological improvements.

Scientific results are expected to be reproducible. When the same study is repeated independently, one needs to reach the same conclusions. Nevertheless, some recent articles have shown that reproducibility is not self-evident. A widely resounding Science article (*1*) has dramatically demonstrated the lack of reproducibility of psychology experiments. Although the more exact sciences are believed to perform better in this respect, concerns about reproducibility have emerged in those fields as well (*2–4*). The issue is of particular interest when computer programs are involved. Undocumented approximations or undetected bugs can lead to entirely wrong conclusions (*5*). In areas where academic codes compete with commercial software, company secrets sometimes hinder assessment of the relevance of the conclusions (*6, 7*).

Density-functional theory (DFT) calculations (*8, 9*) are a typical example of a field that entirely depends on the development and appropriate use of complex software. When implemented in a computer code, DFT provides a way to describe the behaviour of molecules and solids at the atomic scale. Over the years, many academic groups have developed their own DFT codes, and several of these have been adopted by large user communities. Commercial alternatives are entering this area as well. At present, more than 15.000 papers are published each year that make use of DFT codes (*10*), with applications varying from metallurgy to drug design. Moreover, DFT calculations are nowadays used to build large databases (*11, 12*), or in multi-scale calculations where DFT codes are one part of the tool chain (*13, 14*). It is no exaggeration to state that the precision of DFT codes underlies the scientific credibility and reproducibility of a substantial fraction of the current natural and engineering sciences, and thus reaches far beyond the traditional DFT community.

The main idea of density-functional theory is to solve the startlingly difficult many-particle Schrödinger equation by replacing the complete electron wave function by the much simpler ground-state electron density as fundamental variable. Although this reformulation is in principle exact, it is not fully known how the interaction between individual electrons should be transformed. As a result, the specific form of the unknown part of the interaction energy, the exchange-correlation functional, has been the focus of many investigations, leading to a plethora of available functionals in both solid-state physics (*15–19*) and quantum chemistry (*15, 20–23*).

When a particular exchange-correlation functional has been chosen, all that is left is to solve the so-called Kohn-Sham equations and evaluate the total energy. This too can be done in different ways, the performance and precision of which has been compared much less extensively (*21, 24–29*). Solving the Kohn-Sham equations might appear a purely numerical job, so one might assume all codes to produce similar answers to the same problem. A glance at the literature, however, shows that this assumption is by no means always true. Fig. 1 demonstrates that even for a well-studied material like silicon, predictions from different codes (the "precision") vary by the same order of magnitude as the deviation from the 0 K experimental value (*26*) (the "accuracy") (*30*). Because all codes depicted in Fig. 1 treat silicon at the same level of theory, using the same exchange-correlation functional, they yield the same accuracy by definition. However, the particular predictions vary from one code to another due to approximations unrelated to the exchange-correlation functional. These approximations decrease the computational load, but limit the precision.

What precision can we now achieve? Precision-related issues have scarcely been studied and the concept of error bars is all but non-existent in the solid-state DFT community. Although the reproducibility of predictions is sometimes checked by cross-validation with other codes (*21, 24–28*), we are not aware of systematic assessments of precision (also called "verification").

Such studies are nevertheless an important prerequisite to ensure the acceptance of practical DFT calculations. As a group of 69 code developers and expert users, we determine the error bar on energy-versus-volume ($E(V)$) predictions of elemental solids, running the same benchmark protocol with our different, preferred DFT codes. Several parameters of these equations of state (EOS), such as the lattice constant or the bulk modulus, are common targets in accuracy assessments (*15–19*). Moreover, elemental solids provide a broad and comprehensive test for precision. Unlike comparisons to experiment or other levels of theory, the approximations used differ mainly in a very local region around the nuclei. Errors on multicomponent materials may therefore initially be considered as the sum of the errors on the constituent elements; while not sampling every source of error present in a multicomponent system, it is likely that the elements capture much of it. Our effort has resulted in 18 602 DFT calculations, which we aimed to execute with a rigorously determined precision. This exercise might seem simple, but each code tackles the Kohn-Sham equations and subsequent energy evaluation in its own way, requiring different solutions to deal with difficulties in different parts of the computational procedure.

Kohn-Sham solution techniques

The Kohn-Sham equations describe a many-electron system in terms of a density built from single-particle wave functions. By expressing these wave functions as a linear combination of predefined basis functions, the Kohn-Sham equations reduce to a matrix equation, which can in principle be solved exactly. Their solution should yield identical results irrespective of the form of the basis functions, provided the basis set is complete. However, achieving technical convergence of the complete Kohn-Sham problem is in practice not feasible. For example, silicon has an electronic structure as schematically illustrated in Fig. 2: the Aufbau principle requires first populating the lowest energy level, which for silicon is the 1s band.

This is much lower in energy than the valence and conduction bands, and the localization of the orbitals close to the nuclei demands high spatial resolution. These core electrons do not contribute directly to chemical bonding, so they can be separated out and represented using a different basis, better suited to describe localized atomic-like states. Core orbitals may either be computed in an isolated atom environment and their effect on valence transferred unaltered to the crystal, or be relaxed self-consistently in the full crystal field. They can moreover be treated using a relativistic Hamiltonian, essential for core electrons in heavy atoms. Different relativistic schemes may lead to differences in the predicted $E(V)$ curves.

To stitch together a complete solution, the wave functions of the semi-core and valence electrons (2s 2p and 3s 3p, respectively, in the case of silicon) must be constructed to include the effect of orthogonality to the core electrons. This central problem can be solved in a number of different ways depending on the choice of numerical method. For methods based on plane-wave expansions or uniform real-space grids, the oscillatory behaviour cannot be accurately represented due to the limited spatial resolution. The need for unmanageably large basis sets can be mitigated by adding a carefully designed repulsive part to the Kohn-Sham potential, a so-called pseudopotential. This pseudopotential affects only a small region around the nuclei (grey zones in Fig. 2) and may conserve the core-region charge (norm-conserving pseudopotentials (*31, 32*)), giving rise to an analytically straightforward formalism, or break norm conservation by including a compensating augmentation charge (ultrasoft pseudopotentials (*33*)), allowing for smoother wavefunctions and hence smaller basis sets. Alternatively, the projector-augmented wave (PAW) approach defines an explicit transformation between the all-electron and pseudopotential wavefunctions using additional partial-wave basis functions (*34, 35*). This allows PAW codes to obtain good precision for small numbers of plane waves or large grid spacings, but choosing suitable partial-wave projectors is not trivial. We will refer to both pseudopotential and PAW methods as pseudization approaches. In contrast, all-electron methods explicitly construct basis functions that are restricted to a specific energy range ((L)APW (*36–39*), LMTO (*40*)), or treat core and valence states on equal footing, e.g., by using numerical atomic-like orbitals (*41, 42*). Dealing with the full potential enables better precision, but inevitably increases the computation time. In these codes, the ambiguity in solving the Kohn-Sham problem shifts from the choice of the pseudization scheme to the choice of the basis functions. This choice lead to a variety of methods as well, which, despite solving the same Kohn-Sham equations, differ in many other details. Because each of these methods has its own fundamental advantages, it is highly desirable to achieve high precision for all of them.

The Δ matrix

The case study for silicon (Fig. 1) demonstrates that different approaches to the potential or basis functions may lead to noticeably varying predictions, even for straightforward properties like the lattice constant. There is no absolute reference to compare these methods against, as each approach has its own intricacies and approximations. To determine whether the same results can be obtained irrespective of the code or (pseudo)potential, we instead present a large-scale, pair-wise code comparison using the Δ gauge. This criterion was formulated by Lejaeghere *et al.* (*26*) to quantify differences between DFT-predicted $E(V)$ profiles in an unequivocal way. They proposed a benchmark set of 71 elemental crystals and defined for every element i the quantity Δ_i as the root-mean-square difference between the equations of state of methods a and b over a $\pm 6\,\%$ interval around the equilibrium volume $V_{0,i}$. The calculated equations of state are lined up with respect to their minimum energy and compared in an interval symmetrical around the average equilibrium volume (see Fig. 3):

$$\Delta_i(a,b) = \sqrt{\frac{\int_{0.94V_{0,i}}^{1.06V_{0,i}} \left(E_{b,i}(V) - E_{a,i}(V)\right)^2 \,\mathrm{d}V}{0.12V_{0,i}}} \tag{1}$$

A comparison of Δ_i values allows the expression of EOS differences as a single number, and a small Δ_i automatically implies small deviations between equilibrium volumes, bulk moduli or any other EOS-derived observables as well. The overall difference Δ between methods a and b is obtained by averaging Δ_i over all 71 crystals in the benchmark set. Alternative definitions of Δ have recently appeared as well (*27, 28*), and essentially render the same information. In this work, we apply the original Δ protocol to 40 DFT implementations of the PBE functional (*43*). Appropriate numerical settings were determined for each method separately, ensuring converged results. In all calculations, valence and semicore electrons were treated on a scalar-relativistic level, as not all codes support spin-orbit coupling. This is not a limitation, as the aim is to compare codes to each other rather than to experiment. We do not elaborate much on speed and memory requirements, for which we refer to the documentation of the respective codes.

Fig. 4 presents an overview of the most important Δ values categorized into all-electron, PAW, ultrasoft pseudopotential and norm-conserving pseudopotential methods. Approaches with a similar intrinsic precision are in this way clustered together. Both the full results and the most important numerical settings have been included in Tables S3–S42. A complete specification would have to include code defaults and hard-coded values, so a reasonable compromise was chosen. A full specification could be realized by recent endeavours towards full-output databases (*44, 45*) or workflow scripting (*46, 47*), but this is not yet available for several of the codes treated here. We have however tried to provide generation scripts for as many methods as possible (*48*), and emphasize the need for such tools as an important future direction.

Comparing all-electron methods

Although the definition of Δ does not favour a particular reference, it is most instructive to first have a closer look at the Δ values with respect to all-electron methods (Fig. 4). They generally come at a computationally larger cost, but all-electron approaches to DFT are often considered to be a gold standard, as they implement the potential without pseudization. By comparing pseudopotential or PAW methods to all-electron codes, we therefore get an idea of the error bar on each pseudization scheme. The Δ values between different all-electron methods, on the other hand, reflect the remaining discrepancies, such as a different treatment of the scalar relativistic terms or small differences in numerical methods.

To gain some intuition into typical values of Δ, we should first establish which values for Δ can be qualified as 'small', leading to results that can be considered equivalent. A first indication comes from converting differences between high-precision measurements of equation-of-state parameters into a Δ format. Comparing the high-quality experimental data of Holzapfel *et al.* for Cu, Ag and Au (*49*) to those of Kittel (*50*) and Knittle (*51*), for example, marks a small difference Δ_{exp} of 1.0 meV/atom. Since the average all-electron Δ for these materials is only 0.8 meV/atom, this implies that the precision of many DFT codes outperforms experimental precision. Secondly, we also break down the differences between codes in terms of commonly reported equation-of-state parameters. The 1.0 meV/atom maximum Δ between all-electron codes (Fig. 4, top) corresponds to an average volume deviation of 0.14 Å^3/atom (0.38 %) or a median deviation of 0.05 Å^3/atom (0.24 %) over the entire 71-element test set. For the bulk modulus the average deviation is 1.6 GPa (4.0 %) and the median deviation 0.8 GPa (1.6 %). Compared to the scatter on experimental values, amounting to up to 35 % for the bulk moduli of the rare-earth metals, for instance (*52*), these values are very small. The difference between equations of state obtained by independent all-electron codes is hence smaller than the spread between independent experimental equations of state. We conclude that, unless some elements deviate significantly from the overall trend, codes with a mutual Δ of 1 or even 2 meV/atom can be deemed to yield indistinguishable equations of state for all practical purposes.

The above-mentioned differences correspond to the best attainable precision for each all-electron code, using highly converged or 'ultimate' computational settings. It is important to realize, however, that particular choices of these settings may still slightly change the Δ values. Conversely, it is not always necessary to set such stringent requirements, as efficient codes are able to perform well with less-than-perfect settings. Nevertheless, the difference between default- and 'ultimate'-precision equations of state may sometimes reach a few meV/atom (see Table S2). To eliminate the effect of numerical convergence altogether, we tested for the osmium crystal whether it was possible to obtain exactly the same result with different codes. Rather than aiming for the best representation of the ideal PBE results, as in the rest of this work, the goal was now to choose input settings as consistently as possible (using the same basis functions, grids and other parameters). Comparing four APW+lo calculations in this way yielded the results in Table 1. While numerical noise in various subroutines gives rise to fluctuations of only 0.02-0.04 meV/atom, the larger deviation of approximately 0.2 meV/atom in comparisons with respect to `exciting` can partly be attributed to a different scalar-relativistic treatment of the valence electrons in this code. Indeed, there is no single, universal method to account for the relativistic change of the electron mass in the kinetic energy. `exciting` uses the infinite-order regular approximation (IORA) (*53*), while the other three APW+lo codes use the Koelling-Harmon scheme (*54*). A third possibility is to use the zero-order regular approximation (ZORA), as was done in FHI-aims (*42, 55*) (see Tables S5–S7).

Comparing (pseudo)potential libraries

In comparison to all-electron codes, pseudization approaches are generally faster, as fewer states are considered and explicit construction and diagonalization of the Hamiltonian matrix is avoided. Among these, PAW and ultrasoft pseudopotentials require fewer basis functions than the norm-conserving variety, but advanced features such as linear response theory or hybrid functionals may sometimes not be available due to the increased complexity of the implementation. However, they all perform remarkably well in terms of precision when compared with all-electron results (see Fig. 4). For equations of state, the precision of current potentials is able to compete with that of all-electron methods, yielding Δ values of about 1 meV/atom, with a low approaching 0.3 meV/atom. This has not always been the case. As suggested by the example of silicon (Fig. 1), the available potentials have improved considerably over time. In Table 2 it can be seen that for several codes the Δ value is smaller for newer potential sets. Moreover, older potentials like the Troullier-Martins `FHI98pp/ABINIT` norm-conserving set, the `Vdb2/DACAPO` ultrasoft set and the `Vdb/CASTEP` ultrasoft set all have a substantially larger Δ (Fig. 4). This evolution is evidence of internal quality control mechanisms used by developers of potentials in the past, as well as more recently, of additional efforts based on the Δ gauge (e.g., the JTH and SSSP potential libraries). The striking difference with the older potentials, even for the predefined structures in this relatively simple test set, provides a compelling argument to only use the most recent potential files of a given code.

In addition to a comparison with all-electron codes, it is also interesting to assess how different codes implement the same PAW or pseudopotential recipes. When both GPAW and ABINIT use the GPAW 0.9 PAW set, for example, they agree to within a Δ of 0.6 meV/atom. A similar correspondence is found for the Schlipf-Gygi 2015-01-24 ONCVPSP norm-conserving pseudopotentials (0.3 meV/atom between QUANTUM ESPRESSO and CASTEP), the GBRV 1.4 ultrasoft pseudopotentials (0.3 meV/atom between QUANTUM ESPRESSO and CASTEP) and the GBRV 1.2 set (0.7 meV/atom between PAW potentials in ABINIT and ultrasoft potentials in QUANTUM ESPRESSO). Here too, the small Δ values indicate a good agreement between codes. This agreement moreover encompasses varying degrees of numerical convergence, differences in the numerical implementation of the particular potentials and computational differences beyond the pseudization scheme, where the latter are expected to be of the same order of magnitude or smaller than the differences between all-electron codes (1 meV/atom at most).

Conclusions and outlook

Solid-state DFT codes have evolved tremendously. The change from small and personal codes to widespread general-purpose packages has pushed developers to aim for the best possible precision. Whereas DFT-PBE literature on the lattice constant of silicon displayed a spread of 0.05 Å in the past, the most recent versions of the implementations discussed here agree on this value by 0.01 Å (see Fig. 1 and Tables S3–S42). By comparing codes on a more detailed level using the Δ gauge, we have indeed found most recent methods to yield nearly indistinguishable equations of state, with the remaining error bar comparable to that between different high-precision experiments. This underpins the validity of recent DFT EOS results and ensures that correctly converged calculations yield reliable predictions. The message moreover impacts all of the multi-disciplinary set of fields that build upon DFT results, ranging from physical to biological sciences.

In spite of the absence of one absolute reference code, we were able to improve and demonstrate the reproducibility of DFT results by means of a pair-wise comparison of a wide range of codes and methods. Any new methodology development can now verify whether it can reach the same precision, and new DFT applications can show to have used a method and/or potentials that were screened in that way. The data generated in the framework of this paper serve as a crucial enabler for such a reproducibility-driven paradigm shift, and future updates of available Δ values will be presented at http://molmod.ugent.be/deltacodesdft. The reproducibility of reported results also provides a sound basis for further improvement of the accuracy of DFT, i.e. the investigation of new DFT functionals, or for the development of new computational approaches. This work might therefore speed up methodological advances in solid-state DFT substantially.

There is scope for future work to check the reproducibility of different codes even further.

This might consider larger benchmark sets (describing different atomic environments per element), other functionals, an exhaustive comparison of different relativistic treatments, and a more detailed account of computational differences (using databases or scripts, for example). The precision of band gaps, magnetic anisotropies and other non-EOS properties would also be of interest. However, the current investigation of equation-of-state parameters provides the most important pass/fail test to the quality of different implementations of Kohn-Sham theory. A method that is not able to reach an acceptable precision with respect to the equations of state of the elemental crystals, will likely not fulfill even more stringent demands.

References and Notes

1. Open Science Collaboration, *Science* **349**, aac4716 (2015).

2. *Nature Mater.* **14**, 133 (2015).

3. W. F. van Gunsteren, *Angew. Chem. Int. Ed.* **52**, 118 (2013).

4. C. Drummond, *Proc. Evaluation Methods for Machine Learning Workshop at the 26th Int. Conf. Machine Learning* (Montreal, Canada, 2009).

5. G. Miller, *Science* **314**, 1856 (2006).

6. M. Vihinen, *Nature* **521**, 261 (2015).

7. D. C. Ince, L. Hatton, J. Graham-Cumming, *Nature* **482**, 485 (2012).

8. P. Hohenberg, W. Kohn, *Phys. Rev.* **136**, B864 (1964).

9. W. Kohn, L. J. Sham, *Phys. Rev.* **140**, A1133 (1965).

10. R. O. Jones, *Rev. Mod. Phys.* **87**, 897 (2015).

11. S. Curtarolo, *et al.*, *Comput. Mater. Sci.* **58**, 218 (2012).

12. A. Jain, *et al.*, *APL Mater.* **1**, 011002 (2013).

13. C.-C. Fu, J. Dalla Torre, F. Willaime, J.-L. Bocquet, A. Barbu, *Nature Mater.* **4**, 68 (2005).

14. M. Friák, *et al.*, *Eur. Phys. J. Plus* **126**, 101 (2011).

15. S. Kurth, J. P. Perdew, P. Blaha, *Int. J. Quantum Chem.* **75**, 889 (1999).

16. V. N. Staroverov, G. E. Scuseria, J. Tao, J. P. Perdew, *Phys. Rev. B* **69**, 075102 (2004).

17. P. Haas, F. Tran, P. Blaha, *Phys. Rev. B* **79**, 085104 (2009).

18. G. I. Csonka, *et al.*, *Phys. Rev. B* **79**, 155107 (2009).

19. P. Pernot, B. Civalleri, D. Presti, A. Savin, *J. Phys. Chem. A* **119**, 5288 (2015).

20. L. A. Curtiss, K. Raghavachari, P. C. Redfern, J. A. Pople, *J. Chem. Phys.* **106**, 1063 (1997).

21. J. Paier, R. Hirschl, M. Marsman, G. Kresse, *J. Chem. Phys.* **122**, 234102 (2005).

22. Y. Zhao, D. G. Truhlar, *Acc. Chem. Res.* **41**, 157 (2008).

23. L. Goerigk, S. Grimme, *Phys. Chem. Chem. Phys.* **13**, 6670 (2011).

24. A. Kiejna, *et al.*, *Phys. Rev. B* **73**, 035404 (2006).

25. B. Grabowski, T. Hickel, J. Neugebauer, *Phys. Rev. B* **76**, 024309 (2007).

26. K. Lejaeghere, V. Van Speybroeck, G. Van Oost, S. Cottenier, *Crit. Rev. Solid State* **39**, 1 (2014).

27. F. Jollet, M. Torrent, N. Holzwarth, *Comput. Phys. Commun.* **185**, 1246 (2014).

28. E. Küçükbenli, *et al.*, *arXiv:1404.3015 [cond-mat.mtrl-sci]* (2014).

29. S. Poncé, *et al.*, *Comput. Mater. Sci.* **83**, 341 (2014).

30. The studies of accuracy and precision for DFT calculations are often referred to as validation and verification (V&V), respectively (see http://esvv.cecam.org/).

31. D. R. Hamann, M. Schlüter, C. Chiang, *Phys. Rev. Lett.* **43**, 1494 (1979).

32. L. Kleinman, D. M. Bylander, *Phys. Rev. Lett.* **48**, 1425 (1982).

33. D. Vanderbilt, *Phys. Rev. B* **41**, 7892 (1990).

34. P. E. Blöchl, *Phys. Rev. B* **50**, 17953 (1994).

35. G. Kresse, D. Joubert, *Phys. Rev. B* **59**, 1758 (1999).

36. J. C. Slater, *Phys. Rev.* **51**, 846 (1937).

37. O. K. Andersen, *Phys. Rev. B* **12**, 3060 (1975).

38. E. Sjöstedt, L. Nordström, D. J. Singh, *Solid State Commun.* **114**, 15 (2000).

39. G. K. H. Madsen, P. Blaha, K. Schwarz, E. Sjöstedt, L. Nordström, *Phys. Rev. B* **64**, 195134 (2001).

40. J. M. Wills, *et al.*, *Full-Potential Electronic Structure Method. Energy and Force Calculations with Density Functional and Dynamical Mean Field Theory*, vol. 167 of *Springer Series in Solid-State Sciences* (Springer-Verlag, Berlin Heidelberg, 2010).

41. K. Koepernik, H. Eschrig, *Phys. Rev. B* **59**, 1743 (1999). http://www.fplo.de.

42. V. Blum, *et al.*, *Comput. Phys. Commun.* **180**, 2175 (2009).

43. J. Perdew, K. Burke, M. Ernzerhof, *Phys. Rev. Lett.* **77**, 3865 (1996).

44. G. Yuan, F. Gygi, *Comput. Sci. Disc.* **3**, 015004 (2010).

45. http://nomad-repository.eu.

46. S. R. Bahn, K. W. Jacobsen, *Comput. Sci. Eng.* **4**, 56 (2002). .

47. G. Pizzi, A. Cepellotti, R. Sabatini, N. Marzari, B. Kozinsky, *arXiv:1504.01163 [physics.comp-ph]* (2015). http://www.aiida.net/.

48. https://svn.fysik.dtu.dk/projects/gpaw/trunk/gpaw/test/big/dcdft/.

49. W. B. Holzapfel, M. Hartwig, W. Sievers, *J. Phys. Chem. Ref. Data* **30**, 515 (2001).

50. C. Kittel, *Introduction to Solid State Physics* (John Wiley & Sons, Inc, 2005), 8th edn.

51. E. Knittle, *Mineral Physics and Crystallography: A Handbook of Physical Constants*, T. Ahrens, ed. (American Geophysical Union Press, 1995), pp. 98–142.

52. W. A. Grosshans, W. B. Holzapfel, *Phys. Rev. B* **45**, 5171 (1992).

53. K. G. Dyall, E. van Lenthe, *J. Chem. Phys.* **111**, 1366 (1999).

54. D. D. Koelling, B. N. Harmon, *J. Phys. C: Solid State* **10**, 3107 (1977).

55. E. van Lenthe, E. Baerends, J. Snijders, *J. Chem. Phys.* **101**, 9783 (1994).

56. I.-H. Lee, R. M. Martin, *Phys. Rev. B* **56**, 7197 (1997).

57. T. Miyake, T. Ogitsu, S. Tsuneyuki, *Phys. Rev. Lett.* **81**, 1873 (1998).

58. W. Windl, M. M. Bunea, R. Stumpf, S. T. Dunham, M. P. Masquelier, *Phys. Rev. Lett.* **83**, 4345 (1999).

59. L. Vitos, B. Johansson, J. Kollár, H. Skriver, *Phys. Rev. B* **62**, 10046 (2000).

60. R. Miotto, G. P. Srivastava, A. C. Ferraz, *Phys. Rev. B* **63**, 125321 (2001).

61. L. E. Ramos, *et al.*, *Phys. Rev. B* **63**, 165210 (2001).

62. L. V. C. Assali, W. V. M. Machado, J. F. Justo, *Braz. J. Phys.* **34**, 602 (2004).

63. J. Heyd, J. E. Peralta, G. E. Scuseria, R. L. Martin, *J. Chem. Phys.* **123**, 174101 (2005).

64. J. Paier, *et al.*, *J. Chem. Phys.* **124**, 154709 (2006).

65. F. Tran, R. Laskowski, P. Blaha, K. Schwarz, *Phys. Rev. B* **75**, 115131 (2007).

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

67. F. Birch, *Phys. Rev.* **71**, 809 (1947).

68. http://elk.sourceforge.net/.

69. A. Gulans, *et al.*, *J. Phys.: Condens. Matter* **26**, 363202 (2014).

70. I. Y. Zhang, X. Ren, P. Rinke, V. Blum, M. Scheffler, *New J. Phys.* **15**, 123033 (2013).

71. http://www.flapw.de.

72. G. Lehmann, M. Taut, *Phys. Status Solidi B* **54**, 469 (1972).

73. S. Froyen, *Phys. Rev. B* **39**, 3168 (1989).

74. P. Blaha, K. Schwarz, G. K. H. Madsen, D. Kvasnicka, J. Luitz, WIEN2k, An Augmented Plane Wave + Local Orbitals Program for Calculating Crystal Properties (2001). (Karlheinz Schwarz, Techn. Universität Wien, Austria).

75. K. F. Garrity, J. W. Bennett, K. M. Rabe, D. Vanderbilt, *Comput. Mater. Sci.* **81**, 446 (2014). http://www.physics.rutgers.edu/gbrv/.

76. X. Gonze, *et al.*, *Comput. Mater. Sci.* **25**, 478 (2002).

77. X. Gonze, *et al.*, *Comput. Phys. Commun.* **180**, 2582 (2009).

78. M. Torrent, F. Jollet, F. Bottin, G. Zérah, X. Gonze, *Comput. Mater. Sci.* **42**, 337 (2008).

79. https://wiki.fysik.dtu.dk/gpaw/setups/setups.html.

80. J. J. Mortensen, L. B. Hansen, K. W. Jacobsen, *Phys. Rev. B* **71**, 035109 (2005).

81. J. Enkovaara, *et al.*, *J. Phys.: Condens. Matter* **22**, 253202 (2010).

82. http://pwpaw.wfu.edu/.

83. http://www.abinit.org/downloads/PAW2/.

84. http://www.qe-forge.org/gf/project/pslibrary/frs/.

85. A. Dal Corso, *Phys. Rev. B* **86**, 085135 (2012).

86. P. Giannozzi, *et al.*, *J. Phys.: Condens. Matter* **21**, 395502 (2009).

87. N. Marzari, D. Vanderbilt, A. De Vita, M. C. Payne, *Phys. Rev. Lett.* **82**, 3296 (1999).

88. A. Dal Corso, *Comput. Mater. Sci.* **95**, 337 (2014).

89. G. Kresse, J. Furthmüller, *Comput. Mater. Sci.* **6**, 15 (1996).

90. J. Hafner, *J. Comput. Chem.* **29**, 2044 (2008).

91. P. E. Blöchl, O. Jepsen, O. K. Andersen, *Phys. Rev. B* **49**, 16223 (1994).

92. S. J. Clark, *et al.*, *Z. Kristallogr.* **220**, 567 (2005).

93. http://materialscloud.org/sssp/.

94. D. R. Hamann, *Phys. Rev. B* **88**, 085117 (2013).

95. M. Schlipf, F. Gygi, *Comput. Phys. Commun.* **196**, 36 (2015).

96. https://wiki.fysik.dtu.dk/dacapo/.

97. https://wiki.fysik.dtu.dk/dacapo/Pseudopotential_Library/.

98. http://http://www.abinit.org/downloads/psp-links/psp-links/gga_fhi/.

99. M. Fuchs, M. Scheffler, *Comput. Phys. Commun.* **119**, 67 (1999).

100. http://cp2k.web.psi.ch/potentials/.

101. S. Goedecker, M. Teter, J. Hutter, *Phys. Rev. B* **54**, 1703 (1996).

102. C. Hartwigsen, S. Goedecker, J. Hutter, *Phys. Rev. B* **58**, 3641 (1998).

103. M. Krack, *Theor. Chem. Acc.* **114**, 145 (2005).

104. A. Willand, *et al.*, *J. Chem. Phys.* **138**, 104109 (2013).

105. http://bigdft.org/Wiki/index.php?title=New_Soft-Accurate_NLCC_pseudopotentials.

106. L. Genovese, *et al.*, *J. Chem. Phys.* **129**, 014109 (2008).

107. S. Mohr, *et al.*, *J. Chem. Phys.* **140**, 204110 (2014).

108. I. Morrison, D. M. Bylander, L. Kleinman, *Phys. Rev. B* **47**, 6728 (1993).

109. D. R. Hamann, *Phys. Rev. B* **40**, 2980 (1989).

110. T. Ozaki, *Phys. Rev. B* **67**, 155108 (2003).

111. T. Ozaki, H. Kino, *Phys. Rev. B* **69**, 195113 (2004).

112. T. Ozaki, H. Kino, *Phys. Rev. B* **72**, 045121 (2005).

113. http://www.openmx-square.org/.

114. http://www.mat-simresearch.com/.

115. http://www.pseudo-dojo.org/.

Acknowledgments

The present research benefited from financial support from the Research Board of Ghent University (BOF), the Fond de la Recherche Scientifique de Belgique (FRS-FNRS) through PDR Grants T.0238.13 - AIXPHO and T.1031.14 - HiT4FiT, the Communauté française de Belgique through the BATTAB project (ARC 14/19-057), NSF Grant DMR-14-08838, the Swedish Research Council (VR), the Knut and Allic Wallenberg (KAW) foundation, the Fund for Scientific Research–Flanders (FWO) (project number G0E0116N) and the US Department of Energy through grant DOE-BES DE-SC0008938..

The work by N.A.W.H. was supported by NSF grant no. DMR-1105485. J.A.F.-L. acknowledges financial support from the EU's 7th Framework Marie-Curie Scholarship Program within the "ExMaMa" Project (Project No. 329386). I.D.M., O.E., O.G., D.I., Y.O.K., I.L.M.L., and L.N. acknowledge support from eSSENCE. T.B. was supported by the Academy of Finland through Grant No. 263416 and the COMP Centre of Excellence. C.D., A.G., and S.L. acknowledge support from the DFG and the Einstein Foundation, Berlin. A.D.C., S.d.G. and E.K. acknowledge support from the Italian MIUR through PRIN 2010-11 (Registration number

20105ZZTSE_005). P.J.H., D.B.J. and M.I.J.P. are grateful for financial support by EPSRC under the UKCP Grant EP/K013564/1. C.J.P. and J.R.Y. acknowledge support from the Collaborative Computational Project for NMR Crystallography under EPSRC grant EP/J010510/1. W.P. acknowledges funding by FWO. D.J. is grateful for financial support by EPSRC under grant EP/J017639/1. S. Saha acknowledges the support from the Swiss National Science Foundation (SNF). G.-M.R. is thankful for personal financial support from FRS-FNRS. The work by I.E.C. and N.M. was supported by the NCCR MARVEL of the Swiss National Science Foundation. G.K. and P.B. acknowledge support by the Austrian Science Fund (FWF), Project SFB-F41 (ViCoM). S.C. acknowledges financial support from OCAS NV by an OCAS-endowed chair at Ghent University.

Computational resources were provided by: for the Ghent University contributors, the Stevin Supercomputer Infrastructure at Ghent University, funded by Ghent University, the Hercules Foundation, and the Flemish Government (EWI Department); for the Université catholique de Louvain contributors, the Tier-1 supercomputer of the Fédération Wallonie-Bruxelles, infrastructure funded by the Walloon Region under the grant agreement no. 1117545; the CISM/UCL supercomputing facilities and the Consortium des Équipements de Calcul Intensif en Fédération Wallonie-Bruxelles (CÉCI) funded by the FRS-FNRS under convention 2.5020.11; for the CASTEP calculations, the STFC Scientific Computing Department's SCARF cluster; for the Basel University and EPFL contributors, the Swiss National Supercomputing Center (CSCS) in Lugano; for the RSPt contributors, Finland's IT Centre for Science (CSC).

K.L. and F.T. would also like to thank Chandler Becker for instructive discussions on the comparison of atomic-scale simulations.

S.J.C., P.J.H., C.J.P., M.I.J.P., K.R. and J.R.Y. declare the receipt of royalties from commercial sales of CASTEP by Biovia.

Commercial software is identified to specify procedures. Such identification does not imply

recommendation by the National Institute of Standards and Technology.

Supplementary Materials

Materials and Methods

Fig. S1

Tables S1 to S42

References (*48, 66–115*)

Fig. 1. Historical evolution of the predicted equilibrium lattice parameter for silicon. All data points represent calculations within the DFT Perdew-Burke-Ernzerhof (PBE) framework. Values from literature (*15, 16, 18, 56–65*) are compared both to the predictions of different codes within this paper (data points from 2016 and inset; older methods or calculations with lower numerical settings have been depicted by empty symbols) and to the experimental value extrapolated to 0 K and corrected for zero-point effects (red line) (*26*). The concepts of precision and accuracy (see text) are illustrated graphically.

Fig. 2. Electronic states in solid silicon. The valence states are delocalized over the solid (green line), as the wave functions overlap from one atom to the next. The lowest-energy 1s state (red) is at an energy two orders of magnitude lower than the valence states, and is strongly localized near the nucleus with no overlap between the atoms. The grey regions around the atoms indicate approximately where the wavefunction, density and potential are smoothed in pseudized methods.

Fig. 3. Graphical representation of the Δ gauge. The black line depicts the quadratic energy difference between two equations of state, and Δ_i corresponds to the root-means-square average. This is demonstrated by the shaded area, which is equally large above and below the Δ_i^2 line.

Fig. 4. Δ-values between the most important DFT methods considered (in meV/atom). Comparison of all-electron (AE), PAW, ultrasoft (USPP) and norm-conserving pseudopotential codes (NCPP) to all-electron results, listed in alphabetical order per category. The tags stand for `code`, `code/specification` (AE) or `potential set/code` (PAW/USPP/NCPP), and are specified in full in Tables S3–S42. The colour code ranges from green over yellow to red (small to large Δ values). The mixed potential set SSSP was added to the ultrasoft category, in agreement with its prevalent potential type. Both the code settings and the DFT-predicted equation-of-state parameters behind these numbers have been included in Tables S3–S42, and

a full Δ matrix for all methods mentioned in this article is available in Fig. S1.

Table 1. Agreement between osmium crystal predictions at nearly identical settings. (Top) Δ_i values for the osmium crystal (in meV/atom) when four APW+lo calculations tried to mimic the same settings as well as possible. These settings are therefore different from the ones used for Fig. 4 and reported in Tables S3–S4, S8 and S15. (**Bottom**) The corresponding equilibrium volumes V_0, bulk moduli B_0 and bulk modulus derivatives B_1.

Table 2. Precision evolution of PAW and pseudopotential sets over time. The Δ-values are expressed as an average over the all-electron methods (in meV/atom) and are listed chronologically per code. Both the corresponding code settings and the DFT-predicted equation-of-state parameters have been listed in Tables S17, S19–S26, S30–S31 and S33. The most recent potentials are the ones used to generate the data shown in Fig. 4.