DFT METHODS

Reproducibility in density functional theory calculations of solids


The widespread popularity of density functional theory has given rise to an extensive range of dedicated codes for predicting molecular and crystalline properties. However, each code implements the formalism in a different way, raising questions about the reproducibility of such predictions. We report the results of a community-wide effort that compared 15 solid-state codes, using 40 different potentials or basis set types, to assess the quality of the Perdew-Burke-Ernzerhof equations of state for 71 elemental crystals. We conclude that predictions from recent codes and pseudopotentials agree very well, with pairwise comparisons of the Perdew-Burke-Ernzerhof equations for 71 elemental crystals. We conclude that predictions from recent codes and pseudopotentials agree very well, with pairwise differences that are comparable to those between different high-precision experiments. Older methods, however, have less precise agreement. Our benchmark provides a framework for users and developers to document the precision of new applications and methodological improvements.

Scientific results are expected to be reproducible. When the same study is repeated independently, it should reach the same conclusions. Nevertheless, some recent articles have shown that reproducibility is not self-evident. A widely resounding Science article (1), for example, demonstrated a lack of reproducibility among published psychology experiments. Although the hard sciences are believed to perform better in this respect, concerns about reproducibility have emerged in these fields as well (2–4). The issue is of particular interest when computer programs are involved. Undocumented approximations or undetected bugs can lead to wrong conclusions (5). In areas where academic codes compete with commercial software, the unavailability of source code can hinder assessment of the relevance of conclusions (6, 7).

Density functional theory (DFT) calculations (8, 9) are a prominent example of an area that depends on the development and appropriate use of complex software. With rigorous foundations in the quantum theory of matter, DFT describes the structure and properties of molecules and solids at the atomic scale. Over the years, many academic groups have developed implementations of DFT in computer 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 ranging from metalurgy to drug design. Moreover, DFT calculations are used nowadays to build large databases (11, 12) and in multiscale calculations in which they serve as one part of the tool chain (13, 14). The precision of DFT codes thus underlies the scientific credibility and reproducibility of a substantial fraction of current work in the natural and engineering sciences, and therefore it has implications that reach far beyond the traditional electronic-structure research community.

The main idea of DFT is to solve the intratable many-particle Schrödinger equation by replacing the complete electron wave function with the much simpler ground-state electron density as the fundamental variable. Although this formulation 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).

Once a particular exchange-correlation functional has been chosen, the mathematical problem is completely specified as a set of Kohn-Sham equations, whose solution yields orbitals and energies from which the total electronic energy can be evaluated. A variety of such numerical solution schemes have been implemented in different computer codes. Comparisons of their performance are much less frequent or extensive than those of exchange-correlation functionals, however (21, 24–29). One might reasonably expect that because they solve the same equations, they all produce similar answers for a given crystal structure, but a glance at the literature shows that this assumption is by no means always true. Figure 1 demonstrates that even for a well-studied material such as silicon, deviations between predictions from different codes (the “precision”) are of the same order of magnitude as the deviation from the 0 K experimental value (the “accuracy”) (26, 30). Because all of the codes shown 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 because of approximations that are unrelated to the exchange-correlation functional. These approximations decrease the computational load but limit the precision.

What level of precision can we now achieve? Discussion of precision-related issues is uncommon in reports of solid-state DFT studies. The reproducibility of predictions is sometimes checked by cross-validation with other codes (21, 24–28), but we are not aware of any systematic assessments of precision (also called “verification”), even though such studies would reinforce confidence in practical DFT calculations.

As a group of 69 code developers and expert users, we determined the error bar associated with energy-versus-volume (E(V)) predictions of elemental solids by running the same benchmark protocol with various DFT codes. Parameters of these equations of state (EOS), such as the lattice parameter or the bulk modulus, are commonly used for accuracy assessments (15–19). By considering elemental solids, we have established a broad and comprehensive test for precision. Elemental solids have a wide range of chemical environments and constitute a reasonable first approximation to sampling the broad compositional space of multicomponent systems. 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 approaches 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 solutions should yield identical results, irrespective of the form of the basis provided that the basis set is complete. However, achieving technical convergence of the complete Kohn-Sham problem is not feasible in practice. Consider silicon, whose electronic structure is schematically illustrated in Fig. 2. The Aufbau principle requires first populating the lowest energy level, which 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 that is better suited to describe localized atomic-like states. Core orbitals may be either computed in an isolated atom environment, with their effect on valence transferred unaltered to the crystal, or relaxed self-consistently in the full crystal field. They can moreover be treated using a relativistic Hamiltonian, which is 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 that are based on plane-wave expansions or uniform real-space grids, the oscillatory behavior near the nucleus cannot be accurately represented because of 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 (gray zones in Fig. 2) and may conserve the core-region charge [norm-conserving pseudopotential (NCP)] or it may break the core-orbital localization [linear muffin-tin orbital (LMTO) (40) methods] or treat core and valence states on equal footing (e.g., by using numerical atomic-like orbitals) (41, 42). Avoiding pseudization 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 leads to a variety of methods as well, which depend on the choice of numerical methods, differ in many other details. Because each all-electron or pseudization method has its own fundamental advantages, it is highly desirable to achieve high precision for all of them.

The $\Delta$ matrix

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

$$
\Delta_{i}(a,b) = \sqrt{\frac{1}{N_{0}} \sum_{V_{0}} \left[ E_{a}(V) - E_{b}(V) \right]^{2} dV}
$$

A comparison of $\Delta_{i}$ values allows the expression of EOS differences as a single number, and a small $\Delta_{i}$ automatically implies small deviations between equilibrium volumes, bulk moduli, or any other EOS-derived observables. The overall difference $\Delta$ between methods $a$ and $b$ is...
obtained by averaging $\Delta$ over all 71 crystals in the benchmark set. Alternative definitions of $\Delta$ essentially render the same information (27, 28).

In this work, we applied the original $\Delta$ protocol to 40 DFT implementations of the Perdew-Burke-Ernzerhof (PBE) functional (43). Appropriate numerical settings were determined separately for each method, ensuring converged results. In all calculations, valence and semi-core electrons were treated on a scalar-relativistic level, because not all codes support spin-orbit coupling. This is not a limitation, because the aim is to compare codes with each other rather than to experiment. We do not elaborate here on speed and memory requirements, for which we refer to the documentation of the respective codes.

Figure 4 presents an overview of the most important $\Delta$ values, categorized by method: all-electron, PAW, ultrasoft pseudopotentials, and norm-conserving pseudopotentials. Approaches with a similar intrinsic precision are clustered together in this way. Both the full results and the most important numerical settings are included in tables S3 to 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 endeavors in full-output databases (44, 45) or workflow scripting (46, 47), but this capacity is not yet available for several of the codes used in this study. We have, however, tried to provide generation scripts for as many methods as possible (48), and we emphasize the need for such tools as an important future direction.

Comparing all-electron methods

Although the definition of $\Delta$ does not favor a particular reference, it is instructive to first examine the $\Delta$ values with respect to all-electron methods (Fig. 4). They generally come at a computationally higher cost, but all-electron approaches are often considered to be a standard for DFT calculations, because they implement the potential without pseudization. By comparing pseudopotential or PAW methods with all-electron codes, we can therefore get an idea of the error bar associated with each pseudization scheme. The $\Delta$ values between different all-electron methods reflect the remaining discrepancies, such as a different treatment of the scalar-relativistic terms or small differences in numerical methods.

To gain some insight into typical values of $\Delta$, we should first establish which values for $\Delta$ can be qualified as "small," so that we know which results can be considered equivalent. A first indication comes from converting differences between high-precision measurements of EOS parameters into a $\Delta$ format. Comparing the high-quality experimental data of Holzapfel et al. for Cu, Ag, and Au (49) with those of Kittel (50) and Knittle (51), for example, shows a small difference $\Delta_{\text{exp}}$ of 1.0 meV per atom. Because the average all-electron $\Delta$ for these materials is only 0.8 meV per atom, this implies that the precision of many DFT codes outperforms experimental precision.

Secondly, we also considered the differences between codes in terms of commonly reported EOS parameters. The 1.0 meV-per-atom maximum $\Delta$ among all-electron codes (Fig. 4, top) corresponds to an average volume deviation of 0.14 Å³ per atom (0.38%) or a median deviation of 0.05 Å³ per 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 with the scatter on experimental values, which can amount to up to 35% for the bulk moduli of the rare earth metals (for instance, see (52)), these values are very small. The difference between EOS obtained by independent all-electron codes is hence smaller than the spread between independent experimental EOS. We conclude that, unless some elements deviate substantially from the overall trend, codes with a
The black curve depicts the quadratic energy difference between two EOS \((E_1 - E_2)^2\), where the subscripts correspond to the two codes shown, and \(\Delta\) corresponds to the root-mean-square average. This is demonstrated by the shaded area, which is equally large above and below the \(\Delta^2\) line.

**Table 1. Agreement between osmium crystal predictions at nearly identical settings.** The top group includes \(\Delta\) values for the osmium crystal (in millielectron volts per atom) produced by four APW+lo calculations that 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. The bottom group includes the corresponding equilibrium volumes \(V_0\), bulk moduli \(B_0\), and bulk modulus derivatives \(B_1\).

<table>
<thead>
<tr>
<th></th>
<th>Elk</th>
<th>FLEUR</th>
<th>WIEN2K</th>
<th>exciting</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\Delta)(Elk)</td>
<td>–</td>
<td>0.03</td>
<td>0.02</td>
<td>0.20</td>
</tr>
<tr>
<td>(\Delta)(FLEUR)</td>
<td>0.03</td>
<td>–</td>
<td>0.04</td>
<td>0.22</td>
</tr>
<tr>
<td>(\Delta)(WIEN2K)</td>
<td>0.02</td>
<td>0.04</td>
<td>–</td>
<td>0.18</td>
</tr>
<tr>
<td>(\Delta)(exciting)</td>
<td>0.20</td>
<td>0.22</td>
<td>0.18</td>
<td>–</td>
</tr>
<tr>
<td>(V_0) (Å(^3) per atom)</td>
<td>14.276</td>
<td>14.276</td>
<td>14.276</td>
<td>14.274</td>
</tr>
<tr>
<td>(B_0) (GPa)</td>
<td>397.5</td>
<td>397.9</td>
<td>397.6</td>
<td>397.4</td>
</tr>
<tr>
<td>(B_1) (unitless)</td>
<td>4.86</td>
<td>4.89</td>
<td>4.83</td>
<td>4.82</td>
</tr>
</tbody>
</table>

**Table 2. Precision evolution of PAW and pseudopotential sets over time.** The \(\Delta\) values are expressed as an average over the all-electron methods (in millielectron volts per atom) and are listed chronologically per code. The corresponding code settings and the DFT-predicted EOS parameters are listed in tables S17, S19 to S26, S30, S31, and S33. The most recent potentials are the ones used to generate the data shown in Fig. 4.

<table>
<thead>
<tr>
<th></th>
<th>(\Delta) versus AE</th>
</tr>
</thead>
<tbody>
<tr>
<td>JTH01/ABINIT</td>
<td>2013</td>
</tr>
<tr>
<td>JTH02/ABINIT</td>
<td>2014</td>
</tr>
<tr>
<td>Vdb/CASTEP</td>
<td>1998</td>
</tr>
<tr>
<td>OTFG7/CASTEP</td>
<td>2013</td>
</tr>
<tr>
<td>OTFG9/CASTEP</td>
<td>2015</td>
</tr>
<tr>
<td>GPAW06/GPAW</td>
<td>2010</td>
</tr>
<tr>
<td>GPAW09/GPAW</td>
<td>2012</td>
</tr>
<tr>
<td>PSlib031/GE</td>
<td>2013</td>
</tr>
<tr>
<td>PSlib100/GE</td>
<td>2013</td>
</tr>
<tr>
<td>VASP2007/VASP</td>
<td>2007</td>
</tr>
<tr>
<td>VASP2012/VASP</td>
<td>2012</td>
</tr>
<tr>
<td>VASP2007/VASP</td>
<td>2015</td>
</tr>
</tbody>
</table>

The above-mentioned differences correspond to the best attainable precision for each all-electron code, using highly converged or “ultimate” computational settings. However, particular choices for these settings may still slightly change the \(\Delta\) values. It is not always necessary to set such stringent requirements, because efficient codes are able to perform well with less-than-perfect settings. Nevertheless, the difference between default- and ultimate-precision EOS may sometimes reach a few millielectron volts per atom (table S2). To eliminate the effect of numerical convergence altogether, we used the osmium crystal to test whether it is 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 in this case was to choose input settings as consistently as possible (using the same basis functions, grids, and other parameters). Comparing four APW+lo (augmented plane waves plus local orbitals) calculations in this way yielded the results in Table 1. Whereas numerical noise in various subroutines gives rise to fluctuations of only 0.02 to 0.04 meV per atom, the larger deviation of ~0.2 meV per atom in comparisons involving the code known as “exciting” can partly be attributed to a different scalar-relativistic treatment of the valence electrons in this code. There is no single universal method to account for the relativistic change of the electron mass in the kinetic energy. The “exciting” code uses the infinite-order regular approximation (53), whereas the other three APW+lo codes use the Koelling-Harmon scheme (54). A third possibility is to use the atomic zero-order regular approximation, as was done in the FHI-aims code package (tables S5 to S7) (42, 55).

**Comparing (pseudo)potential libraries**

In comparison with all-electron codes, pseudization approaches are generally faster, because 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 sometimes may not be available because of the increased complexity of the implementation. However, pseudization approaches all perform very well in terms of precision when compared with all-electron results (Fig. 4). For EOS, the precision of current potentials is able to compete with that of all-electron methods, yielding \(\Delta\) values of about 1 meV per atom, with a low approaching 0.3 meV per 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 \(\Delta\) value is smaller for newer potential sets. Moreover, older potentials such as the Troullier-Martins FHI98pp norm-conserving set in ABINIT or the Vanderbilt-type ultrasoft sets in Dacapo and CIPRSTEM all have a substantially larger \(\Delta\) (Fig. 4). This evolution is an evidence of internal quality-control mechanisms used by developers of potentials in the past, as well as of additional, more recent efforts based on the \(\Delta\) gauge (e.g., the Jollet-Torrent-Holzwarth (JTH) and Standard Solid-State Pseudopotentials at DOE).
Conclusions and outlook

Solid-state DFT codes have evolved considerably. The change from small and personalized codes to widespread general-purpose packages has pushed developers to aim for the best possible precision. Whereas past DFT-PBE literature on the lattice parameter of silicon indicated a spread of 0.05 Å, the most recent versions of the implementations discussed here agree on this value within 0.01 Å (Fig. 1 and tables S3 to S42). By comparing codes on a more detailed level using the $\Delta$ gauge, we have found the most recent methods to yield nearly indistinguishable EOS, with the associated error bar comparable to that between different high-precision experiments. This underpins the validity of recent DFT EOS results and confirms that correctly converged calculations yield reliable predictions. The implications are moreover relevant throughout the multidisciplinary set of fields that build upon DFT results, ranging from the physical to the 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 pairwise comparison of a wide range of codes and methods. It is now possible to verify whether any newly developed methodology can reach the same precision described here, and new DFT applications can be shown to use a methodology or/and potentials that were screened in this work.

Future work can examine the reproducibility of different codes even further. Such work might involve larger benchmark sets (describing different atomic environments per element), other functionals, or for the development of new computational approaches. This work might therefore substantially accelerate methodological advances in solid-state DFT.

Future work can examine the reproducibility of different codes even further. Such work might involve larger benchmark sets (describing different atomic environments per element), other functionals, an exhaustive comparison of different relativistic treatments, and/or 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 EOS parameters provides the most important pass-fail test of the quality of different implementations of Kohn-Sham theory. A method that is not able to reach

(SSSP libraries). The considerable difference in the older potentials, even for the predefined structures in this relatively simple test set, provides a compelling argument to use only the most recent potential files of a given code.

In addition to the comparison with all-electron codes, it is also interesting to assess how the same PAW or pseudopotential recipes are implemented in different ways. When both the GPAW and ABINIT codes use the GPAW 0.9 PAW set, for example, they agree to within a $\Delta$ of 0.6 meV per atom. A similar correspondence is found for the Schlipf-Gygi 2015-01-24 optimized norm-conserving Vanderbilt pseudopotentials (ONCVSP) (0.3 meV per atom between Quantum ESPRESSO and CASTEP), the Garrity-Bennett-Rabe-Vanderbilt (GBRV) 1.4 ultrasoft pseudopotentials (0.3 meV per atom between Quantum ESPRESSO and CASTEP) and the GBRV 1.2 set (0.7 meV per atom between PAW potentials in ABINIT and ultrasoft potentials in Quantum ESPRESSO). In this case, too, the small $\Delta$ 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, most of which are expected to be of the same order of magnitude or smaller than the differences among all-electron codes (1 meV per atom at most).

**Fig. 4.** $\Delta$ values for comparisons between the most important DFT methods considered (in millielectron volts per atom). Shown are comparisons of all-electron (AE), PAW, ultrasoft (USPP), and norm-conserving pseudopotential (NCPP) results with all-electron results (methods are listed in alphabetical order in each category). The labels for each method stand for code, code/specification (AE), or potential set/code (PAW, USPP, and NCPP) and are explained in full in tables S3 to S42. The color coding illustrates the range from small (green) to large (red) $\Delta$ 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 EOS parameters behind these numbers are included in tables S3 to S42, and fig. S1 provides a full $\Delta$ matrix for all methods mentioned in this article.
This study relied on the collective efforts of a large group of developers and expert users to make pairwise comparisons of widely used DFT codes. We compared 40 DFT methods in terms of Δ, which expresses the root-mean-square difference between the EOS of two codes, averaged over a benchmark set of 71 elemental crystals (Eq. 1). Our approach together with the details about the codes used, is described further in the supplementary materials. The reported settings yield highly converged results but may not be necessary for typical DFT applications. In particular, the use of sometimes very small electronic smearing widths requires much higher numbers of k-points than routine DFT calculations warrant.

REFERENCES AND NOTES

24. K. Rayanpura, A. Elmamouni, J. Peveling, J. Neugebauer, Supplementary materials. The reported settings are described further in the supplementary materials. The reported settings yield highly converged results but may not be necessary for typical DFT applications. In particular, the use of sometimes very small electronic smearing widths requires much higher numbers of k-points than routine DFT calculations warrant.
14/19-057); the U.S. NSF (grant DMR-14-08838); the Swedish Research Council; the Knut and Alice Wallenberg Foundation (grants 2013.0020 and 2012.0031); the Fund for Scientific Research–Flanders (FWO) (project no. G0E0116N); and the U.S. Department of Energy (grant DE-SC0008938). N.A.W.H. was supported by U.S. NSF grant DMR-1105485. J.A.F.-L. acknowledges financial support from the European Union’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 (grant 265468) and the COMP Centre of Excellence. C.D., A.G., and S.L. acknowledge support from the Deutsche Forschungsgemeinschaft (DFG) and the Einstein Foundation, Berlin. M.Sche. and C.D. received funding from the European Union’s Horizon 2020 research and innovation program under grant agreement no. 676580 with The Novel Materials Discovery (NOMAD) Laboratory, a European Center of Excellence. A.D.C., S.d.G., and E.K. acknowledge support from the Italian Ministry of Education, Universities, and Research (MIUR) through PRIN (Projects of National Interest) 2010–2011 (registration no. 20105ZZTSE_000). P.J.H., D.B.J., and M.I.J.P. are grateful for financial support by EPSRC under UK Car-Parrinello (UKCP) grant EP/KO19564/1. J.R.Y. acknowledge support from the Collaborative Computational Project for NMR Crystallography under EPSRC grant EP/I010501/1. W.P. acknowledges funding by FWO. D.J. is grateful for financial support by EPSRC under grant EP/J017639/1. S.Sa. acknowledges 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 SNSF’s National Centre of Competence in Research MARVEL. G.K. and P.B. acknowledge support by the Austrian Science Fund, project SFB-F41 (ViCoM). S.C. acknowledges financial support from OCAS NV by an OCAS-endowed chair at Ghent University. Computational resources were as follows: The Ghent University contributed used the Stevin Supercomputer Infrastructure at Ghent University, which is funded by Ghent University, FWO, and the Flemish Government (Economy, Science, and Innovation Department). The Université Catholique de Louvain contributed used the Tier-1 supercomputer of the Fédération Wallonie-Bruxelles (funded by Walloon Region under grant agreement no. 1117545), the Centre de Calcul Intensif et de Stockage de Masse–Université Catholique de Louvain supercomputing facilities, and the Consortium des Équipements de Calcul Interuniversitaires Fédération Wallonie-Bruxelles (CÉCI) (funded by the FRS-FNRS under convention 2.5020.11). The Science and Technology Facilities Council, Scientific Computing Department’s SCARF (Scientific Computing Application Resource for Facilities) cluster was used for the CASTEP calculations. The Basel University and École Polytechnique Fédérale de Lausanne contributed used the Swiss National Supercomputing Center in Lugano, Finland’s IT Centre for Science was used for the RSPt calculations. K.L. and F.T. thank C. Becker for instructive discussions on the comparison of atomic-scale simulations. K.L. and S.C. thank W. Dewitte for drafting the summary figure. S.J.C., P.J.H., C.J.P., M.I.J.P., K.B., and J.R.Y. declare the receipt of income from commercial sales of CASTEP by Biovia. N.M. and M.Sche. are members of the Board of Trustees of the Psi-k Electronic Structure Network. P.G. is director of the Quantum ESPRESSO Foundation, and N.M. is a representative member. X.G., D.R.H., M.T., D.C., F.J., and G.M.R. are members of the Advisory Board of ABINIT, an organization that develops and publishes open-source software related to this article. Commercial software is identified to specify procedures. Such identification does not imply recommendation by the National Institute of Standards and Technology. Atomic Simulation Environment scripts (46) for several of the codes are available online (48). All data are listed in tables S3 to S42.

SUPPLEMENTARY MATERIALS
www.sciencemag.org/content/351/6280/aad3000/suppl/DC1
Materials and Methods
Fig. S1
Tables S1 to S42
References (66–115)
27 August 2015; accepted 19 February 2016
10.1126/science.aad3000