

Numerical Quality Control for DFT-based Materials Databases

Christian Carbogno^{a*}, Kristian Sommer Thygesen^b, Björn Bieniek^a, Claudia Draxl^{c,a},
Luca Ghiringhelli^a, Andris Gulans^c, Oliver T. Hofmann^d, Karsten W. Jacobsen^b,
Sven Lubeck^c, Jens Jørgen Mortensen^b, Mikkel Strange^b, Elisabeth Wruss^d,
and Matthias Scheffler^a

- a. Fritz-Haber-Institut der Max-Planck-Gesellschaft,
Faradayweg 4–6, D-14195 Berlin, Germany
- b. Center for Atomic-scale Materials Design (CAMD), Department of Physics,
Technical University of Denmark. Fysikvej 1 2800 Kgs. Lyngby, Denmark
- c. Physics Department and IRIS Adlershof,
Humboldt-Universität zu Berlin, Zum Großen Windkanal 6, D-12489 Berlin, Germany
- d. Institute of Solid State Physics,
Graz University of Technology, NAWI Graz, Petergasse 16, 8010 Graz, Austria

* carbogno@fhi-berlin.mpg.de

Density-functional theory (DFT) has become an invaluable tool in materials science. Recently, the precision of different approaches has been scrutinized for the PBE functional using extremely accurate numerical settings [1]. However, little is yet known about code- and method-specific errors that arise under more commonly used numerical settings, e.g. in *high-throughput* calculations. This constitutes a severe issue, since it prevents repurposing DFT data created using different settings and/or codes, for instance the calculations stored in the NOMAD [2] or other repositories [3,4,5].

To overcome this, we study the convergence of different properties (geometries, total and relative energies) in four conceptually-different DFT codes (exciting, FHI-aims, GPAW, and VASP) for typical settings used in production calculations. Specifically, we discuss relative and absolute errors as a function of the numerical settings, e.g., basis sets and **k**-grids, for 71 elemental solids [1]. Using this data, we propose an analytical model that allows for error estimates for *any* compound, as we explicitly demonstrate for 73 binary and ternary solids. We show how the developed formalism can be incorporated into electronic structure theory databases so that data created using different settings and/or codes becomes quantitatively comparable [6], e.g., so to use it in machine-learning approaches. Eventually, we discuss the extensibility of our approach towards more complex materials properties.

- [1] K. Lejaeghere, *et al.*, *Science* **351**, aad3000 (2016).
- [2] L. M. Ghiringhelli, *et al.*, *NPJ Computational Materials* **3**, 46 (2017).
- [3] C. E. Calderon, *et al.*, *Comput. Mater. Sci.* **108**, 233 (2015).
- [4] A. Jain, *et al.*, *APL Materials* **1**, 011002 (2013).
- [5] J. E. Saal, *et al.*, *JOM* **65**, 1501 (2013).
- [6] <https://analytics-toolkit.nomad-coe.eu/>