Extrapolation to the complete basis-set limit in density-functional theory using statistical learning

Daniel Speckhard[1,3], Christian Carbogno[2], Luca Ghiringhelli[4], Claudia Draxl[1,3] , Matthias Scheffler[2]
[1]Humboldt University of Berlin, Zum Grossen Windkanal 2, Berlin, Germany
[2]Fritz-Haber Institute of the Max Planck Society, Faradayweg 6-8, Berlin, Germany
[3]Max Planck Institute for Solid State Research, Heisenbergstrasse 1, Stuttgart, Germany
[4]Friedrich-Alexander Universität Erlangen-Nürnberg, Dr.-Mack-Str. 77, 90762 Fürth, Germany

dts@physik.hu-berlin.de

The numerical precision of density-functional-theory (DFT) calculations depends on a variety of computational parameters, one of the most critical being the basis-set size [1, 2, 3]. The ultimate precision is reached with an infinitely large basis set, i.e., in the complete basis-set (CBS) limit. Our aim is to find a machine-learning model that extrapolates finite basis-size calculations to the CBS limit. Quantile random forests (QRF) and symbolic regression, applying the SISSO approach, are used to estimate total energies, lattice parameters, and band gaps as a function of the basis-set size. The QRF model outperforms previous approaches in the literature for both codes, while SISSO outperforms the random-forest model for the **exciting** code [4]. Our approach also provides prediction intervals, which quantify the models' uncertainty. Uncertainty estimates for the total energies would provide users useful information about the precision of these calculations and how the data can be reused/re-purposed [5]. We evaluate our work on datasets consisting of 63 binary solids and 3000 binary semiconductors, respectively.

References

[1] Carbogno et al., npj Computational Materials 8, 69, (2022)
[2] Kraus, P., J. Chem. Theory Comput., 16, 5712–5722 (2020)
[3] Lejaeghere et al., Science 351, aad3000 (2016)
[4] Speckhard et al., Physical Review Materials, Accepted (2024)
[5] Draxl, C. & Scheffler, M., J. Physics: Mater. 2, 036001273

Figures

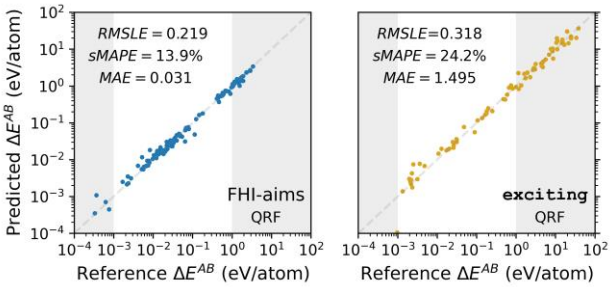

Figure 1. Predictions of total-energy differences to the CBS of the QRF model on FHI-aims (left) and **exciting** (right) data (63 binaries of different basis-set size precision), plotted against the respective DFT results. Relevant error metrics are added to help interpret the quality of the fit. Note the logarithmic axes. The region of DFT calculated ΔE^{AB} values between 1 meV/atom and 1 eV/atom is plotted with a lighter shade since these data are of particular interest to DFT practitioners.