The Journal of Open Source Software

excitingtools: An exciting Workflow Tool

Alexander Buccheri [1,2], **Fabian Peschel** [1], **Benedikt Maurer** [1], **Mara Voiculescu** [1], **Daniel T. Speckhard** [1], **Hannah Kleine** [1], **Elisa Stephan** [1], **Martin Kuban** [1], **and Claudia Draxl** [1]

1 Humboldt-Universität zu Berlin, Berlin, Germany **2** Max Planck Institute for the Structure and Dynamics of Matter, Hamburg, Germany

DOI: 10.21105/joss.05148

Software

- Review
- Repository
- Archive

Editor: Richard Gowers

Reviewers:

- @fjmartinmartinez
- @hmacdope

Submitted: 05 January 2023
Published: 03 May 2023

Introduction

exciting (Gulans et al., 2014) is a full-potential, all-electron, density-functional theory (DFT) code, which uses linearized augmented plane waves plus local orbitals (LAPWs and LOs, respectively) for valence and semi-core electrons, and explicitly treats core electrons via the Dirac equation. This basis set provides a systematic path for reaching the complete-basis-set limit, relying only on well-controlled numerical approximations. The high precision of the LAPW basis set makes **exciting** well-suited for the generation of benchmark-quality data, and to serve as a reference for other DFT implementations, especially those relying on pseudopotentials.

It has been shown that **exciting** is able to achieve microhartree precision for total energies in DFT calculations (Gulans et al., 2018). Also in G_0W_0 calculations, the complete-basis-set limit was attained (Nabok et al., 2016). Its high precision has also been unequivocally demonstrated in the so-called Δ test (Lejaeghere et al., 2016), which compared the relative precision of several DFT codes for a benchmark set of 71 elemental crystals.

Statement of Need

Materials databases such as NOMAD (Draxl & Scheffler, 2019), AFLOW (Curtarolo et al., 2012), and the Materials Project (Jain et al., 2013) host millions of DFT results. The majority of data have been computed with pseudopotential-based DFT codes, and is thus lacking validation with more precise methods. There is a strong need to provide databases with benchmark-quality results, which serve to give an indication of the precision one can achieve in a given material property, with a specific method and settings. For scientists and engineers who wish to compute specific properties to some required precision, having an indication of the optimal settings and suitable DFT approximations is extremely valuable. Beyond the ground state, materials databases for excited state calculations are, in general, strongly lacking. The generation of large amounts of excited state data will require both reliable ground state calculations as inputs and analogous benchmark-quality calculations. Moreover, machine learning models that predict material properties would greatly benefit from the availability of higher-fidelity data sets for a range of systems (Dong et al., 2022; Draxl & Scheffler, 2020; Toniato et al., 2021).

With demand for more calculations of higher precision and increased complexity, comes the need for more complex workflows, handled in a systematic, automated manner. To illustrate this point in the context of **exciting**, the choice of the LAPW basis and systematic convergence of calculations, even at the ground state level of theory, is more involved than with plane wave or Gaussian type orbital (GTO) basis sets. For example, one is free to choose any non-overlapping radii for the muffin-tin spheres of each atomic species, and any number of LAPWs and LOs. And for each of these basis functions, one is also free to choose the matching

order of radial functions, orbital angular momenta, and trial energy parameters associated with them (Andersen, 1975; Blaha et al., 2020; Gulans et al., 2014). This is before performing the conventional convergence tests, as done in all DFT calculations.

In order to perform systematically-converged, reproducible, benchmark-quality calculations for ground and excited state phenomena (the latter of which will typically include one or more ground state calculations), a framework is required to assist in the selection of calculation parameters, the simplification of input file generation, and the post-processing of results. Furthermore, this framework needs to be interoperable with the ecosystem of existing workflow tools. These challenges are met by *excitingtools*.

Summary

excitingtools is a Python package that provides a high-level frontend for the **exciting** all-electron DFT package, integrating all aspects of performing a calculation into a single program. *excitingtools* has been developed with interoperability in mind, and supports the use of Atomic Simulation Environment (ASE) (Larsen et al., 2017). Its serialized input classes and output parsers allow it to be used with higher-level workflow managers such as Jobflow (Ganose et al., 2022), Atomate (Mathew et al., 2017), and Atomic Simulation Recipes (ASR) (Gjerding et al., 2021).

excitingtools is an essential utility for simplifying the use of **exciting**, enabling greater user control over calculations. Whilst ASE consists of parsers and calculators, its API is largely restricted to ground state energies and forces. *excitingtools* exposes more functionality allowing users, for instance, to analyze their results of different SCF cycles in a calculation, or perform and parse excited state (e.g. GW) calculations. *excitingtools* enables automation of complex convergence calculations, facilitates high-throughput studies, and forms the building blocks of higher-level workflow managers, all of which are prerequisites for moving DFT codes towards exascale calculations (Gavini et al., 2022). Furthermore, *excitingtools* is under active development and follows a continuous integration/deployment model, such that new features and updates are delivered several times a year.

Features

- *excitingtools* allows the user to quickly create a class object with given key-value pairs in Python, to create input files for **exciting** in an automated manner.
  - This avoids the need for users to manually configure inputs, which is error-prone, and alleviates the need to frequently write single-purpose scripts.
  - ASE's `Atoms` class is accepted as a structure input.
- *excitingtools* provides parsers for fifty **exciting** output file formats.
  - Parsing **exciting** previously required downloading the NOMAD parsers which return custom objects, containing copious metadata. This is unnecessary for **exciting** users, and prevents straightforward numerical comparison of parsed results.
- API interoperability and serializable data structures allow easy integration with workflow managers.
  - These features allow the user to create simulation input files, run simulations and analyze data with Python, paving the way to high-throughput calculations with **exciting**.

Acknowledgements

The authors would like to thank Ask Hjorth Larsen and David Waroquiers for fruitful discussions. This work received funding from the European Union's Horizon 2020 research and innovation program under the grant agreement Nº 951786 (NOMAD CoE) and the German Research

Foundation (DFG) through the CRC 1404 (FONDA), project 414984028. Daniel Speckhard acknowledges support by the IMPRS for Elementary Processes in Physical Chemistry.

References

Andersen, O. K. (1975). Linear methods in band theory. *Physical Review B*, *12*(8), 3060. https://doi.org/10.1103/physrevb.12.3060

Blaha, P., Schwarz, K., Tran, F., Laskowski, R., Madsen, G. K., & Marks, L. D. (2020). WIEN2k: An APW+ lo program for calculating the properties of solids. *The Journal of Chemical Physics*, *152*(7), 074101. https://doi.org/10.1063/1.5143061

Curtarolo, S., Setyawan, W., Hart, G. L., Jahnatek, M., Chepulskii, R. V., Taylor, R. H., Wang, S., Xue, J., Yang, K., Levy, O., & others. (2012). AFLOW: An automatic framework for high-throughput materials discovery. *Computational Materials Science*, *58*, 218–226. https://doi.org/10.1016/j.commatsci.2012.02.005

Dong, J., Zhao, M., Liu, Y., Su, Y., & Zeng, X. (2022). Deep learning in retrosynthesis planning: Datasets, models and tools. *Briefings in Bioinformatics*, *23*(1), bbab391. https://doi.org/10.1093/bib/bbab391

Draxl, C., & Scheffler, M. (2019). The NOMAD laboratory: From data sharing to artificial intelligence. *Journal of Physics: Materials*, *2*(3), 036001. https://doi.org/10.1088/2515-7639/ab13bb

Draxl, C., & Scheffler, M. (2020). Big data-driven materials science and its FAIR data infrastructure. *Handbook of Materials Modeling: Methods: Theory and Modeling*, 49–73. https://doi.org/10.1007/978-3-319-44677-6_104

Ganose, A., Jain, A., Rignanese, G.-M., Waroquiers, D., Petretto, G., & others. (2022). Jobflow. In *GitHub repository*. GitHub. https://github.com/materialsproject/jobflow

Gavini, V., Baroni, S., Blum, V., Bowler, D. R., Buccheri, A., Chelikowsky, J. R., Das, S., Dawson, W., Delugas, P., Dogan, M., & others. (2022). Roadmap on electronic structure codes in the exascale era. *arXiv Preprint arXiv:2209.12747*. https://doi.org/10.48550/arXiv.2209.12747

Gjerding, M., Skovhus, T., Rasmussen, A., Bertoldo, F., Larsen, A. H., Mortensen, J. J., & Thygesen, K. S. (2021). Atomic simulation recipes: A python framework and library for automated workflows. *Computational Materials Science*, *199*, 110731. https://doi.org/10.1016/j.commatsci.2021.110731

Gulans, A., Kontur, S., Meisenbichler, C., Nabok, D., Pavone, P., Rigamonti, S., Sagmeister, S., Werner, U., & Draxl, C. (2014). Exciting: A full-potential all-electron package implementing density-functional theory and many-body perturbation theory. *Journal of Physics: Condensed Matter*, *26*(36), 363202. https://doi.org/10.1088/0953-8984/26/36/363202

Gulans, A., Kozhevnikov, A., & Draxl, C. (2018). Microhartree precision in density functional theory calculations. *Physical Review B*, *97*(16), 161105. https://doi.org/10.1103/physrevb.97.161105

Jain, A., Ong, S. P., Hautier, G., Chen, W., Richards, W. D., Dacek, S., Cholia, S., Gunter, D., Skinner, D., Ceder, G., & others. (2013). Commentary: The materials project: A materials genome approach to accelerating materials innovation. *APL Materials*, *1*(1), 011002. https://doi.org/10.1063/1.4812323

Larsen, A. H., Mortensen, J. J., Blomqvist, J., Castelli, I. E., Christensen, R., Dułak, M., Friis, J., Groves, M. N., Hammer, B., Hargus, C., & others. (2017). The atomic simulation

environment—a python library for working with atoms. *Journal of Physics: Condensed Matter*, *29*(27), 273002. https://doi.org/10.1088/1361-648X/aa680e

Lejaeghere, K., Bihlmayer, G., Björkman, T., Blaha, P., Blügel, S., Blum, V., Caliste, D., Castelli, I. E., Clark, S. J., Dal Corso, A., & others. (2016). Reproducibility in density functional theory calculations of solids. *Science*, *351*(6280), aad3000. https://doi.org/10.1126/science.aad3000

Mathew, K., Montoya, J. H., Faghaninia, A., Dwarakanath, S., Aykol, M., Tang, H., Chu, I., Smidt, T., Bocklund, B., Horton, M., & others. (2017). Atomate: A high-level interface to generate, execute, and analyze computational materials science workflows. *Computational Materials Science*, *139*, 140–152. https://doi.org/10.1016/j.commatsci.2017.07.030

Nabok, D., Gulans, A., & Draxl, C. (2016). Accurate all-electron G0W0 quasiparticle energies employing the full-potential augmented plane-wave method. *Physical Review B*, *94*(3), 035118. https://doi.org/10.1103/PhysRevB.94.035118

Toniato, A., Schwaller, P., Cardinale, A., Geluykens, J., & Laino, T. (2021). Unassisted noise reduction of chemical reaction datasets. *Nature Machine Intelligence*, *3*(6), 485–494. https://doi.org/10.1038/s42256-021-00319-w