MM 43: Data Driven Material Science: Big Data and Workflows V

Time: Wednesday 15:45–18:00 Location: C 243

MM 43.1 Wed 15:45 C 243

The MALA Package - Transferable and Scalable Electronic Structure Simulations Powered by Machine Learning — •Lenz Fiedler and Attila Cangi — Helmholtz-Zentrum Dresden-Rossendorf / CASUS

Interactions between electrons and nuclei determine all materials properties, and modeling these interactions is of paramount importance to pressing scientific questions. However, even with the most advanced electronic structure tools, such as density functional theory (DFT), electronic structure simulations are usually restricted to a few thousand atoms.

Machine-learning DFT (ML-DFT) tackles this challenge by providing rapid access to observables of interest. Most current ML-DFT methodologies focus on the mapping between ionic configurations and scalar observables, rather than a full prediction of electronic structure. The Materials Learning Algorithms (MALA) python package addresses this gap, by providing a user-friendly framework to construct ML-DFT models that give access to a range of electronic structure observables, such as the electronic density, DOS and total free energy. The usefulness of these models across phase boundaries [1], multiple length scales [2] and temperature ranges [3] has been amply demonstrated. In this talk, a general introduction to the framework along with its computational properties and capabilities is given.

[1] J. A. Ellis *et. al*, Phys. Rev. B, 2021, **104**, 035120 ** [2] L. Fiedler *et. al*, npj Comput. Mater., 2022, **9**, 115

[3] L. Fiedler *et. al*, Phys. Rev. B, 2023, **108**, 125146

MM 43.2 Wed 16:00 C 243

A robust, simple and efficient algorithm to converge GW calculations — •Max Grossmann, Malte Grunert, and Erich Runge — Theoretische Physik I, TU Ilmenau, Germany

In the era of high-throughput computation for materials science and machine learning, the demand for heuristics and efficient convergence schemes in ab initio calculations has become essential to speed up computational workflows. A prime example of computationally demanding ab initio material-science calculations is the GW method, which is crucial for accurate band gap predictions and the subsequent computation of dielectric functions using TDDFT/BSE. The main hurdle within GW calculations lies in the convergence of two interdependent parameters in the dynamic screening W which must be optimized simultaneously. To overcome this obstacle we introduce a straightforward, "cheap first, expensive later" coordinate-search algorithm, which is tested against a commonly used state-of-the-art method on a diverse dataset of 50 semiconducting and insulating solids. Additionally, the practical independence between the **k**-point grid and parameters in W is checked using both algorithms on five different **k**-meshes of increasing density for all investigated materials, starting at **Γ**-only calculations. The results show how to converge GW calculation in a robust, simple and efficient way.

MM 43.3 Wed 16:15 C 243

FAIR Data Management for Computational Materials Science using NOMAD — •Luca M. Ghiringhelli[1], Joseph F. Rudzinski[2], José M. Pizarro[2], Nathan Daelman[2], and Silvana Botti[3] — [1]Department of Materials Science and Engineering, Friedrich-Alexander-Universität, Erlangen-Nürnberg — [2]Institut für Physik und IRIS-Adlershof, Humboldt-Universität zu Berlin, Berlin — [3]RC-FEMS and Faculty of Physics, Ruhr University Bochum, Bochum

NOMAD [nomad-lab.eu][1, 2] is an open-source data infrastructure for materials science data. Originally built as a repository for DFT calculations, NOMAD has been extensively developed over the past 2 years to support a wide range of both computational and experimental data. Additionally, NOMAD now includes a general workflow support that not only streamlines data provenance and analysis but also facilitates the curation of AI-ready datasets. This talk will demonstrate how these features, along with NOMAD's adherence to the FAIR principles (Findability, Accessibility, Interoperability, Reusability), provide a powerful framework for enhancing data utility and discovery. I will highlight how this FAIR-compliant perspective distinguishes NOMAD from other Big-Data infrastructures, e.g., allowing users to specify their data quality needs. Finally, I will provide an outlook of NOMAD's potential for creating a cohesive, interconnected, and economical scientific data landscape.

[1] Scheidgen, M. *et al.*, JOSS **8**, 5388 (2023).

[2] Scheffler, M. *et al.*, Nature **604**, 635-642 (2022).

MM 43.4 Wed 16:30 C 243

A many-body framework for long-range interactions in atomistic machine learning — •Kevin Kazuki Huguenin-Dumittan, Philip Robin Loche, and Michele Ceriotti — EPFL, Lausanne, Switzerland

Many properties of matter, from chemical bonds to the band structure in solids, arise from the quantum nature of electrons. Accurate atomistic modeling of such systems thus requires the use of quantum approaches, which can computationally become prohibitively expensive compared to their classical counterparts. Machine learning based approaches have seen a surge in interest over the past decade to bridge the gap between the two worlds by providing near-quantum accuracy at a cost scaling similarly to classical methods. One systematic source of error in most such models is the assumption of locality, which neglects important long-range interactions ranging from electrostatics to more complex effective interactions between macromolecules. In this talk, we present a long-range framework that is a significant improvement over the previously introduced LODE descriptors, which (1) scales effectively linearly with system size, (2) captures a broad family of long-range interactions beyond the Coulomb potential, (3) can systematically generate many-body features beyond just pair potentials and (4) smoothly and fully integrates with preexisting machine learning frameworks with little extra effort including the computation of gradients or equivariant models for tensorial properties.

MM 43.5 Wed 16:45 C 243

Automated prediction of Fermi surfaces from first principles — •Nataliya Paulish[1], Junfeng Qiao[2], and Giovanni Pizzi[1,2] — [1]Paul Scherrer Institut (PSI), Villigen, Switzerland — [2]École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

Knowing the shape of the Fermi surface (FS) and the energy dispersion in its vicinity is crucial to understand the electronic properties of materials and identify materials relevant for applications. Experimental methods to measure FSs are very expensive and time-consuming, and accurate theoretical predictions would help to get deeper insights from the experimental data. Direct first-principles calculations of the FS, requiring very dense sampling in reciprocal space, are thus limited by the computational cost. To accelerate the calculations, we use interpolation with Maximally Localized Wannier Functions (MLWFs), powered by our new method that allows fully automated calculation of MLWFs [1]. We first validate the numerical approach by comparing our simulation results with literature data for de Haas-van Alphen (dHvA) oscillation frequencies and investigate the main sources of numerical errors. We then use our high-throughput setup, with our code implemented as an AiiDA [2] workflow, to create a large database of Fermi surfaces and dHvA oscillation frequencies of 3D inorganic metals, starting from high-symmetry systems.

[1] J. Qiao *et al.*, npj Comput. Mater. 9, 208 (2023)

[2] S. P. Huber *et al.*, Scientific Data, 7, 300 (2020)

15 min. break

MM 43.6 Wed 17:15 C 243

Accurate and Efficient Protocols for High-Throughput Computational Materials Science — •Gabriel M. Nascimento[1], Flaviano José dos Santos[2], Marnik Bercx[2], Davide Grassano[1], Giovanni Pizzi[2], and Gabriel de Miranda Nascimento[1,2] — [1]École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland — [2]Laboratory for Materials Simulations (LMS), Paul Scherrer Insititute, Villigen, Switzerland

Advances in simulation methods and workflow engines, and an ever-increasing computational power have enabled the paradigm change in computational materials science of high-throughput (HT) discovery. A significant challenge in these efforts is to automate the choice of parameters in simulation codes, where a consistent balance of accuracy and computational cost is required. We propose a rigorous methodology to assess the quality and sensitivity of self-consistent density-functional theory (DFT) calculations with respect to smearing and Brillouin zone

sampling of 2D and 3D crystalline compounds. Our results enable the construction of unified protocols for the choice of parameters that guarantee k-point sampling convergence and control of entropic systematic errors for calculations within most classes of materials, while minimizing the required overall wall time. These protocols are integrated into automated workflows for DFT calculations using Quantum ESPRESSO, accessible through the open-source AiiDA framework. This work streamlines access to state-of-the-art computational frameworks, fostering HT research and supporting the development of computational materials databases.

MM 43.7 Wed 17:30 C 243

Defect Phase Diagrams for Grain Boundaries in Mg: Automized workflows for chemical trends — •Prince Mathews[1], Rebecca Janisch[2], Jörg Neugebauer[1], and Tilmann Hickel[1,3] — [1]Max-Planck-Institut für Eisenforschung, Düsseldorf, Germany — [2]ICAMS, Ruhr Universität Bochum, Germany — [3]Federal Institute for Materials Research and Testing (BAM), Berlin, Germany

The design of lattice defects is a crucial component for processing tailored materials as they control the mechanical properties and corrosion behavior. In order to achieve this, the framework of defect phase diagrams is a powerful and knowledge-based design strategy, which can be used to manipulate defect phases using defect-property relationships. Defect phase diagrams use the chemical potential as the key variable as it has the same value at the defect as well as in the adjacent bulk. With pyiron, an integrated development environment for computational materials science, an automatic workflow has been developed for calculating defect phase diagrams for binary systems using ab-initio simulations. The workflow is used for the example of Ga and Ca segregation to $\Sigma 7$ [0001] 21.78° (sym. plane 12-30) Mg grain boundaries. Chemical trends for the relative stability of different atomic configurations at the grain boundaries are derived. Connecting the workflow with experimental datasets, a good agreement has been confirmed. The advantages of automized workflows for the defect phase diagrams is then demonstrated for the extension of the considerations to finite temperatures and free energies calculated within the quasi-harmonic approximation.

MM 43.8 Wed 17:45 C 243

High-level hybrid workflow-based approach for characterizing magnetocaloric materials: concurrent computational and experimental validation — •Simon Bekemeier[1], Alisa Chirkova[1], and Christian Schröder[1,2] — [1]Hochschule Bielefeld - University of Applied Sciences and Arts — [2]Faculty of Physics, Bielefeld University

Processing of expanding experimental and computational datasets often requires multiple tools applied in a sequence before the main result is retrieved. Data treatment, its transfer between different tools and sharing are key points in nowadays materials development and digitalization processes. A study of magnetocaloric materials produces a moderate size dataset, from which certain values need to be calculated (ΔS, magnetic entropy change). Obtaining it from first principles requires several computational steps, with the output of the previous step transferred to the next one. In this work, a workflow for the ΔS-calculation from both experimental and theoretical data is developed. The computational tools are accessed through the pyiron©IDE that includes atomistic simulations using density functional theory as well as newly implemented CINOLA spin-dynamics and magnetic transition evaluation from thermodynamics. As a proof of concept, we show results for Ni and Gd exhibiting a second-order magnetic phase transition and for representative model materials with a first-order phase transition. Our approach has an advantage of the acceleration and automatization of the data transformation and treatment within a single study; furthermore, it offers experimentalists a convenient option to perform powerful simulations on their own.