MM 31: Data-driven Materials Science: Big Data and Worksflows

Materials Properties and more

Time: Thursday 15:00–18:00 Location: H10

MM 31.1 Thu 15:00 H10

Thermodynamic stability of the materials in the Materials Cloud three-dimensional crystals database (MC3D) — •Timo Reents[1,2], Marnik Bercx[1], and Giovanni Pizzi[1,2] — [1]Laboratory for Materials Simulations (LMS) and National Centre for Computational Design and Discovery of Novel Materials (MARVEL), Paul Scherrer Institut (PSI), CH-5232 Villigen PSI, Switzerland — [2]École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

High-throughput studies based on ab initio methods such as Density Functional Theory (DFT) enable the analysis of physical properties across a broad chemical space. Here, we present the Materials Cloud three-dimensional crystals database (MC3D), a DFT optimized and curated structural database of experimentally known inorganic crystals. All calculations are managed and driven by the AiiDA [1, 2] workflow engine, allowing to browse the full provenance graph and to share the results in the Materials Cloud [3]. We introduce the protocols behind MC3D, the new frontend, and we then focus on the thermodynamic stability. To improve the agreement between the theoretical and experimental thermodynamic stability, we apply empirical [4] and machine-learning [5] based corrections, and improve upon them, discussing the agreement with experimental data on stability.

[1] Huber, S.P. et al., Sci Data, 2020, 7, 300.
[2] Uhrin, M. et al., Comp. Mat. Sci., 2021, 187, 110086.
[3] Talirz, L. et al., Sci Data 7, 299 (2020).
[4] Stevanović, V. et al., Phys. Rev. B, 2012, 85, 115104.
[5] Gong, S. et al., JACS Au, 2022, 2, 1964-1977.

MM 31.2 Thu 15:15 H10

high-throughput computation and machine learning modeling of magnetic moments and Mössbauer spectroscopy for Fe-based intermetallics — •Bo Zhao, Xiankang Tang, and Hongbin Zhang — Institute of Materials Science, Technische Universität Darmstadt, Otto-Berndt-Str. 3, 64287 Darmstadt, Germany

Understanding the relationship between the local crystalline environment and magnetic properties is a fundamental challenge in condensed matter physics and materials science. This study explores this relationship in Fe-based intermetallic compounds, focusing on the magnetic moments and Mössbauer parameters of iron atoms, including the isomer shift, electric field gradient, and magnetic hyperfine field. High-throughput calculations and machine learning techniques are employed to predict magnetic properties based on local atomic structures, using smooth overlap of atomic positions (SOAP) as local descriptors. The results first reveal the sparsity of relevant materials in the Materials Project database. Leveraging high-throughput, system-specific data, the study demonstrates strong correlations between local atomic environments and magnetic properties, achieved through machine learning models. Furthermore, the limitations of symmetry-invariant descriptors in predicting tensor-like properties, such as the electric field gradient, are highlighted. By incorporating a graph-based equivariant autoencoder, the model achieves improved predictions by effectively capturing the symmetry of local environments.

MM 31.3 Thu 15:30 H10

Advanced Machine Learning of ^{17}O NMR in Non-Magnetic Oxides: High-Throughput Calculation, Prototype Compound Analysis, and Transfer Learning — •Zhiyuan Li, Bo Zhao, Hongbin Zhang, and Yixuan Zhang — Institute of Materials Science, TU Darmstadt, 64287 Darmstadt Germany

The study of ^{17}O NMR spectroscopy is crucial for understanding the local structure of oxides, where the naturally occurring NMR-active oxygen isotope, ^{17}O, provides unique insights into local environments due to its large chemical shift range and quadrupolar nature. In this work, we present a high-throughput workflow integrating AiiDa and CASTEP to calculate the NMR parameters of over 7100 compounds from the Materials Pro ject database, followed by utilizing machine learning models to predict ^{17}O NMR parameters. Furthermore, taking $BaTiO_3$ as an example, we identify prototypical ABO_3 crystal structures, construct $BaTiO_3$ analogs via substitution, perform ab initio molecular dynamics simulations to generate 3000 perturbated structures, and evaluate the NMR parameters. The results of our machine learning modeling with such additional dataset reveal that incorporating perturbated structures enhances the accuracy of the machine learning model. Moreover, by leveraging transfer learning, using previously trained model from our high-throughput dataset, the predictivity for the newly generated $BaTiO_3$ analogs can be further improved.

MM 31.4 Thu 15:45 H10

Advancing chemical shielding predictions in organic solids — •Matthias Kellner and Michele Ceriotti — École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland

In this presentation, we showcase our recent advancements in machine learning for predicting chemical shieldings in organic solids. Leveraging symmetry-adapted machine learning models, our updated infrastructure facilitates the accurate prediction of chemical shielding anisotropy and enables structure optimization driven by chemical shielding gradients. We will highlight how integrating machine learning potentials with property prediction models provides unique insights into atomistic processes, offering a powerful framework for exploring the complex behavior of organic materials.

MM 31.5 Thu 16:00 H10

Active learning workflow for mixed-halide perovskite stability and electronic band-structure — •Tim Bechtel[1,2], Santiago Rigamonti[1], and Claudia Draxl[1,2] — [1]Humboldt-Universität zu Berlin, Germany — [2]Max Planck Institute for Solid-State Research, Stuttgart, Germany

Mixed-halide perovskites are promising materials for stable and efficient light harvesting and emission applications, and their composition can be tailored to match relevant regions of the light spectrum [1]. Theoretical predictions from first-principles calculations can provide insight into stability, ground-state properties, and electronic structure [2,3]. Comparison with experimental results for "real" materials is, however, challenging. For example, the consideration of chemical (dis)order requires huge supercells, which is computationally out of reach with state-of-the-art methodology. For the family of $CsPb(Cl_xBr_yI_{1-x-y})_3$ compounds, we bridge this gap with an active learning workflow. It is based on a fine-tuned machine learning interatomic potential [4] that interpolates between already seen compositions, and actively explores new composition ranges. This approach allows for data-efficient predictions of stability through finite-temperature phase diagrams and optical properties for a wide range of compositions.

[1] H. Näsström, PhD Thesis, https://doi.org/10.18452/24939
[2] F. Pan, *et al.*; https://doi.org/10.1021/acs.chemmater.4c00571
[3] J. Laakso, *et al.*; https://doi.org/10.1103/PhysRevMaterials.6.113801
[4] I. Batatia, *et al.*; https://doi.org/10.48550/arXiv.2401.00096

15 min. break

MM 31.6 Thu 16:30 H10

Towards Multi-Fidelity Machine Learning Using Robust Density Functional Tight Binding Models — •Mengnan Cui[1,2], Karsten Reuter[1], and Johannes T. Margraf[2] — [1]Fritz-Haber-Institut der MPG, Berlin, Germany — [2]University of Bayreuth, Physical Chemistry V: Theory and Machine Learning

Machine learning has revolutionized the atomistic simulation of molecules and materials, offering unparalleled computational speed with high accuracy. However, its performance depends heavily on the quality and quantity of training data, presenting challenges due to the scarcity of high-fidelity datasets (beyond semilocal DFT). This study investigates transfer learning (TL) across multiple fidelities for molecules and solids, examining the role of fidelity levels and configuration/chemical space overlap in pre-training and fine-tuning. This reveals negative transfer driven by noise from low-fidelity methods like DFTB, which can significantly impact fine-tuned models. Despite this, multi-fidelity approaches consistently outperform single-fidelity learning and, in some cases, even surpass TL based on foundation models by leveraging an optimal overlap of pre-training and fine-tuning chemical spaces.

MM 31.7 Thu 16:45 H10

Enhancing FAIR Data Management with Automated Visualization of Calculations — •N. Daelman[1], E. Boydas[1], B. Mohr[1], J.M. Pizarro[1], T. Bereau[2], C. Draxl[1], L.M. Ghiringhelli[3], M. Girard[4], D. Usvyat[5], R. Valenti[6], S. Botti[7], and J.F. Rudzinski[1] — [1]CSMB, HU Berlin — [2]ITP, Heidelberg Uni. — [3]Dept. of Mater. Sci. and Eng., FAU Erlangen — [4]Max Planck Inst. for Poly. Res., Mainz — [5]Inst. für Chem., HU Berlin — [6]Inst. für Theor. Phys., GU Frankfurt/M — [7]RC-FEMS, Ruhr Uni. Bochum

In contrast to data science packages, first-degree data post-processing tends to lock people into silos built around a particular simulation software. NOMAD [nomad-lab.eu][1] is an open-source and community-driven data infrastructure that breaks open these silos by extracting scientific data from over 60 code packages into a code-agnostic schema within a research data management (RDM) ecosystem [2]. This talk showcases NOMAD*s new visualization features at various levels of RDM. At the level of individual calculations, NOMAD provides now more detailed electronic structure visualizations and fast, dynamic rendering of heavy files. Automated visualization does not imply, however, a lack of customizability. NOMAD provides support for tailored figures and larger-scale specialization via an accessible plugin-based system. At the level of research projects, NOMAD allows for quick monitoring of the data coverage via a fully customizable dashboard.

[1] Scheidgen, M. *et al.*, JOSS **8**, 5388 (2023).
[2] Scheffler, M. *et al.*, Nature **604**, 635-642 (2022).

MM 31.8 Thu 17:00 H10

NOMAD CAMELS: An Open-Source Solution for Creating FAIR Data from Experiments — •Alexander Fuchs[1,2], Johannes Lehmeyer[1,2], Michael Krieger[1,2], and Heiko Weber[1,2] — [1]Lehrstuhl für Angewandte Physik, Friedrich-Alexander-Universität Erlangen-Nürnberg — [2]FAIRmat Consortium

NOMAD CAMELS is a configurable open-source measurement software. It is suited to control experiments and records fully self-describing experimental data. It has its origins in the field of experimental physics where a wide variety of measurement instruments are used in frequently changing experimental setups and measurement protocols. CAMELS provides a graphical user interface (GUI) which allows the user to configure experiments without the need of programming skills or deep understanding of instrument communication. CAMELS translates user-defined measurement protocols into stand-alone executable Python code for full transparency of the actual measurement sequences. Metadata inflow from Electronic Lab Notebooks (ELNs) and data output into such is well supported for a seamless workflow. CAMELS is designed with a focus on full recording of data and metadata aligned with the NeXus ontology. When shared with others, data produced with CAMELS allow full understanding of the measurement and the resulting data in accordance with the FAIR principles.

MM 31.9 Thu 17:15 H10

Databases of Fermi surfaces and de Haas-van Alphen oscillation frequencies from first principles simulations — •Nataliya Paulish[1], Junfeng Qiao[2], and Giovanni Pizzi[1] — [1]PSI Center for Scientific Computing, Theory and Data, 5232 Villigen PSI, Switzerland — [2]Theory and Simulation of Materials (THEOS), and National Centre for Computational Design and Discovery of Novel Materials (MARVEL), École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland

The Fermi surface (FS) of a metal separates occupied from unoccupied electronic states. Knowing its shape is crucial to understanding the electronic properties of the material. Accurate simulation of the FS requires a very dense sampling of the Brillouin zone, and thus direct density functional theory (DFT) calculations are limited by their computational cost. To overcome this difficulty, we use interpolation from a basis of spatially localized projectability disentangled Wannier functions (PDWFs) - a recently developed algorithm for automated Wannierization [1]. Using this algorithm, FSs were generated for over 7'000 inorganic metals. We also computed de Haas-van Alphen frequencies associated with each FS, enabling direct comparison of our simulations with experiments. The procedure is fully automated using the AiiDA workflow engine [2]. Our database will be published openly online and browsable on the Materials Cloud MC3D section (https://mc3d.materialscloud.org).

[1] J. Qiao, G. Pizzi, N. Marzari, npj Comput Mater 9, 208 (2023)
[2] S. P. Huber et al., Scientific data 7, 1 (2020)

MM 31.10 Thu 17:30 H10

A systematic benchmark of G_0W_0 calculations — •Marc Thieme, Max Grossmann, Malte Grunert, and Erich Runge — Technische Universität Ilmenau, Ilmenau, Germany

Accurate and efficient ab initio electronic structure calculations of semiconductors and insulators are a prerequisite for building large, high-quality databases for machine learning (ML). However, the "optimal" choice (speed vs. accuracy) of the approximations used, i.e. the exchange-correlation (XC) functional for density functional theory (DFT) calculations or a particular many-body perturbation theory, remains unclear. A systematic benchmark of band gaps of solids using several different DFT XC functionals by Borlido et al [1,2] showed that hybrid functionals perform exceptionally well and seem to be the functionals of choice. The present study addresses the question of whether G_0W_0 calculations provide a sufficient increase in accuracy to justify their increased computational cost compared to simpler DFT calculations with hybrid functionals. We calculate the band gaps for about 300 materials using the G_0W_0 method starting from LDA/PBE DFT calculations. The deviations between G_0W_0 and experimental band gaps are systematically compared with those of the best hybrid functionals

[1] Borlido et al., J. Chem. Theory Comput. 15, 9 (2019)
[2] Borlido et al., npj Comput. Mater. 6, 96 (2020)

MM 31.11 Thu 17:45 H10

Machine Learning-Assisted Design of Magnetic Materials: Predicting Properties for not purely ternary $Nd_2Fe_{14}B$ — •Manuel Enns, Daniel Urban, Wolfgang Körner, and Christian Elsässer — Fraunhofer IWM, Wöhlerstraße 11, 79108 Freiburg, Germany

$Nd_2Fe_{14}B$-based hard-magnetic materials are widely used for strong permanent magnets. Their re-use and recycling after the end of the magnet's life cycle opens the question of the degradation of the magnetic properties due to the incorporation of unintentional impurity elements originating from the recycling procedures. In this talk, we present a data-mining and machine-learning (ML) approach using kernel-based learning methods to predict the influence of impurity atoms in $Nd_2Fe_{14}B$-based materials. The magnetic-property data used for training and testing the ML model were obtained by a combinatorial high-throughput screening (HTS) using density-functional theory calculations. We demonstrate that our ML approach can accurately predict the saturation magnetization, the uniaxial anisotropy constant, and the formation energy for $Nd_2Fe_{14}B$ with impurities added by recycling.