MM 3: Data-driven Materials Science: Big Data and Worksflows

Machine Learning, Potential Development

Time: Monday 10:15–13:00 Location: H10

MM 3.1 Mon 10:15 H10

Benchmarking DFT Functionals at Finite Temperature with ASSYST and MLIPs — •Marvin Poul and Jörg Neugebauer — Max-Planck-Institut für Nachhaltige Materialien

A key ingredient to the accuracy of Density Functional Theory (DFT) calculations is the chosen approximation to the exchange-correlation functional. Local Density Approximation (LDA) and Generalized Gradient Approximation (GGA) calculations often bracket experimental observations, but systematic exploration of the behavior of different density functionals is hindered by the high computational cost of DFT in realistic applications, especially concerning finite temperature properties. Using the ASSYST[1] method, we automatically generate unary, general purpose Atomic Cluster Expansion (ACE) Machine Learning Interatomic Potentials (MLIPs) for a range of metals using LDA, PBE and r2SCAN functionals. The key advantage of ASSYST lies in the small cells ($\leq$ 10 atoms per cell) that it generates as training data. This allows us to relabel the data using different functionals very efficiently. We then use these potentials to calculate melting curves, thermal expansion, and formation energies of various defects (grain boundaries, surfaces, point defects) to systematically assess strengths and weaknesses of the DFT functionals. In general, we find good agreement with corresponding DFT results, showing that ASSYST can reliably create transferable potentials for metals at DFT accuracy.

[1]: https://www.researchsquare.com/article/rs-4732459/v1

MM 3.2 Mon 10:30 H10

Assessing the role of physical constraints in machine learning potentials — •Marcel F. Langer, Sergey N. Pozdnyakov, Filippo Bigi, and Michele Ceriotti — Laboratory of Computational Science and Modeling (COSMO) and National Centre for Computational Design and Discovery of Novel Materials (MARVEL), Institute of Materials, École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland

Machine learning potentials, which approximate the potential energy surface of atomistic systems to enable larger and longer simulations than first-principles methods, have advanced rapidly in recent decades. Much of this development has been driven by the increasingly sophisticated treatment of physical symmetries, in particular invariances, in the underlying machine learning models. However, the rise of so-called unconstrained models, which replace exact invariance with learned approximations, has sparked debate over this approach. Some models even choose to directly predict forces, rendering the resulting force fields non-conservative. We investigate the effectiveness of such models and evaluate the impact of disregarding physical constraints for practical simulations. In particular, we study the effects of breaking rotational symmetry in a machine-learning potential for water [1] and discuss the potential consequences of direct force predictions.

[1]: M.F. Langer, S.N. Pozdnyakov, and M. Ceriotti, Mach. Learn.: Sci. Technol. 5 04LT01 (2024)

MM 3.3 Mon 10:45 H10

Fast and flexible range-separated models for atomistic machine learning — •Philip Loche, Marcel F. Langer, and Michele Ceriotti — Laboratory of Computational Science and Modeling (COSMO), Institute of Materials, École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland

Most machine learning (ML) interatomic potentials rely on a locality ansatz, decomposing energy into short-ranged, atom-centered contributions. This limits their ability to describe problems dominated by long-range physical effects, such as electrostatics. We present a framework integrating established algorithms for non-bonded interactions – including Ewald summation, PME, and P3M – into atomistic ML. Reference implementations are provided in PyTorch and JAX. Beyond Coulomb potentials, we introduce Exterior Potential Features for general long-range ML applications. Our modular libraries enable accurate physical force evaluations, seamless integration with local ML schemes via automatic differentiation, and flexible architectures for advanced models. We benchmark these tools for molecular dynamics, range-separated ML potentials, and long-range atomic descriptors.

MM 3.4 Mon 11:00 H10

Beyond Numerical Hessians: Applications for Higher Order Derivatives in Machine Learning Interatomic Potentials — •Nils Gönnheimer[1,2], Karsten Reuter[1], and Johannes T. Margraf[2] — [1]University of Bayreuth — [2]Fritz-Haber-Institut der MPG, Berlin

The development of machine learning interatomic potentials (MLIPs) has revolutionized computational chemistry by enhancing the accuracy of empirical force fields while retaining a large computational speed-up compared to first-principles calculations. Despite these advancements, calculating Hessian matrices remains challenging due to the lack of analytical second-order derivatives, necessitating the use of computationally expensive finite difference methods (which can lead to numerical instabilities because of rounding errors). Automatic differentiation (AD) offers a promising approach to reducing this computational effort and making the calculation of Hessian matrices more efficient and accurate. In this contribution, we discuss the implementation of AD Hessians in the equivariant MACE framework. This new methodology finds applications in screening the heat capacities of metal-organic frameworks (MOFs) and in the calculation of infrared (IR) spectra, which are an ubiquitous tool for molecular characterization.

MM 3.5 Mon 11:15 H10

Diversity-Driven Active Learning of Interatomic Potentials for Reaction Network Exploration — •Francesco Cannizzaro, King Chun Lai, Patricia Poths, Sebastian Matera, Vanessa J. Bukas, and Karsten Reuter — Fritz-Haber-Institut der MPG, Berlin

We present an automatic workflow for the simultaneous active learning of Machine-Learning Interatomic Potentials (MLIPs) and exploration of complex networks of activated events. This workflow consists of alternating periods of training and the generation of candidate structures for the enrichment of the training set using the recently developed Automatic Process Explorer (APE) [1]. This allows us to determine elementary processes and corresponding barriers without the need of human supervision. From the output of the APE explorations, we identify maximally diverse atomic structures utilizing the DECAF fuzzy classification algorithm [2] and add only these to the training set. We exemplify this strategy on carbon intercalation in Pd, using GAP and MACE as MLIP frameworks. We find that this diversity driven approach outperforms state-of-the-art training set designs based on molecular dynamics for finding activated events and corresponding barriers. Particularly, our workflow performs very well in reducing outliers, which is of utmost importance for activated event dynamics since this is often controlled by only a few barriers.

[1] Lai *et al.*, ChemRxiv, https://doi.org/10.26434/chemrxiv-2024-jbzr7 (2024).

[2] Lai *et al.*, J. Chem. Phys. **159**, 024129 (2023).

15 min. break

MM 3.6 Mon 11:45 H10

Accelerating Materials Exploration with Active Machine Learning: Integrating SISSO with FHI-aims — Yi Yao[1,2], Lucas Foppa[1], Akhil Sugathan Nair[1], Andrei Sobolev[1,2], •Konstantin Lion[1,2], Sebastian Kokott[1,2], and Matthias Scheffler[1] — [1]NOMAD Laboratory at the Fritz Haber Institute of the Max Planck Society, Berlin, Germany — [2]Molecular Simulations from First Principles e.V., Berlin, Germany

We present a user-friendly web application for active learning-based materials exploration with the goal of broadening the usability of AI tools. The platform integrates the SISSO (Sure Independence Screening and Sparsifying Operator) method [J. Chem. Phys. 159, 114110 (2023)] with FHI-aims software [Comp. Phys. Commun. 180, 2105 (2009)] to provide interpretable modeling and reliable property predictions. SISSO dynamically updates models during the exploration process, while FHI-aims ensures accurate all-electron density functional theory (DFT)-based calculations. The property prediction workflow is managed using the atomate2 library, providing many "standard" DFT workflows and efficient utilization of compute resources ranging from

local machines to cloud infrastructures. By leveraging SISSO-based uncertainty prediction, the application implements active learning to efficiently identify materials with desirable target properties. Two case studies, the exploration of the bulk modulus in perovskites and the prediction of stable oxides under harsh conditions, demonstrate the platform's ability to accelerate materials discovery.

MM 3.7 Mon 12:00 H10

Data-driven design of mechanically hard soft magnetic high-entropy alloys — •Mian Dai[1], Yixuan Zhang[1], Xiaoqing Li[2], Stephan Schönecker[2], Liuliu Han[3], Ruiwen Xie[1], Chen Shen[1], and Hongbin Zhang[1] — [1]Institute of Materials Science, Technical University of Darmstadt, Darmstadt, Germnay — [2]Department of Materials Science and Engineering, KTH - Royal Institute of Technology, Stockholm, Sweden — [3]Max Planck Institute for Sustainable Materials, Düsseldorf, Germany

The rational design of mechanically hard soft magnets, combining high hardness with magnetically soft properties, represents a critical frontier in materials science. Here, we introduce a comprehensive data-driven framework to navigate the vast compositional space of high-entropy alloys (HEAs) and identify candidates optimized for these dual functionalities. Utilizing a curated dataset of 1,842,628 density functional theory calculations, encompassing 45,886 quaternary and 414,771 quinary equimolar HEAs derived from 42 elements, we employ ensemble learning to synergistically integrate multiple predictive models. This methodology captures the relationships between composition, crystal structure, mechanical performance, and magnetic behavior, enabling the identification of alloys with a unique combination of high hardness and magnetic softness. Our framework establishes a robust pathway for the accelerated discovery of next-generation hard-soft magnetic materials, underscoring the transformative potential of data-driven strategies in materials design.

MM 3.8 Mon 12:15 H10

Autonomous optimization of coin-cell batteries and thin-film growth — •Edan Bainglass[1,6], Peter Kraus[2,5], Francisco Ramirez[3,6], Enea Svaluto-Ferro[2], Loris Ercole[3,6], Benjamin Kunz[2], Sebastiaan Huber[3,6], Nukorn Plainpan[2], Nikita Shepelin[1], Nicola Marzari[1,3,6], Corsin Battaglia[2,3,4], and Giovanni Pizzi[1,3,6] — [1]PSI, Villigen, Switzerland — [2]Empa, Dübendorf, Switzerland — [3]EPFL, Lausanne, Switzerland — [4]ETH Zurich, Zurich, Switzerland — [5]TUB, Berlin, Germany — [6]MARVEL, Switzerland

Advancements in materials science are increasingly driven by the integration of automation of both experiments and simulations, machine learning, and robust data management frameworks. In this talk, we discuss the integration of experimental systems with the AiiDA [1] workflow management system, both battery coin cell assembly and cycling [2], and for thin film growth by pulsed laser deposition (PLD). We discuss the ongoing integration of these platforms with the FINALES [3] fast intention-agnostic learning server towards fully autonomous optimization of battery end-of-life (EOL) performance. We also discuss preliminary results demonstrating the feasibility of autonomously optimizing the layer-by-layer thin-film growth with PLD. These case studies demonstrate the potential of automated workflows to accelerate the discovery and optimization of functional materials.

[1] S. P. Huber et al., Sci. data 7, 300 (2020)
[2] P. Kraus et al., J. Mat. Chem. A 12, 10773 (2024)
[3] M. Vogler et al. Adv. Ener. Mat. 2403263 (2024)

MM 3.9 Mon 12:30 H10

Learning Disorder in Generative Materials Discovery - Bridging Prediction and Experiment — •Konstantin Jakob[1], Aron Walsh[2], Karsten Reuter[1], and Johannes T. Margraf[1,3] — [1]Fritz-Haber-Institut der MPG, Berlin, Germany — [2]Imperial College London, London, UK — [3]University of Bayreuth, Bayreuth, Germany

In recent years, generative machine learning (ML) models have demonstrated tremendous potential for the design and discovery of new materials. This has led to extensive predictions of previously unknown, potentially stable inorganic materials. However, current models suffer from the fact that the underlying training data is purely based on density functional-calculations for small, ideal crystals. As a consequence, many of the supposedly new materials are in fact experimentally known as disordered crystals. In this work, we address this issue by performing a thorough analysis of crystal disorder in the experimental structures of the Inorganic Crystal Structure Database (ICSD). Based on this, we develop disorder classification models and representations that can predict the likelihood of disorder across chemical space. Eventually, these concepts will allow us to extend current generative models to realistic crystal systems and bridge the gap between prediction and experiment.

MM 3.10 Mon 12:45 H10

Materials-Discovery Workflows Guided by Symbolic Regression: Identifying Stable Oxides for Catalytic Applications — •Akhil S. Nair, Lucas Foppa, and Matthias Scheffler — The NOMAD Laboratory at the FHI of the Max Planck Society, Berlin, Germany

AI-driven workflows will accelerate materials discovery by efficiently guiding experiments or simulations towards materials with desired properties. However, probabilistic AI approaches commonly used in these workflows are limited by the relatively small size of high-quality datasets and they rely on typically unknown, low-dimensional representations. Herein, we discuss the recent advancements in applying symbolic regression based on the sure-independence screening and sparsifying operator (SISSO) approach within iterative frameworks for materials discovery. This involves an ensemble approach for the uncertainty quantification of SISSO models as well as the development of optimization strategies to efficiently explore promising regions of the materials space. These developments present an opportunity to integrate SISSO into sequential-learning workflows for materials discovery. Importantly, SISSO provides materials-property maps covering the entire materials space, further reducing the risk that the workflow misses promising materials that were overlooked in the initial dataset. We demonstrate the effectiveness of the SISSO-guided workflows by identifying stable oxides for catalytic applications.