MM 18: SYMD contributed

Time: Wednesday 10:15–13:15 Location: H23

MM 18.1 Wed 10:15 H23

Relation Between Element Specific Chemistry and Basis Set Size of Machine Learned Interatomic Potentials — •Haitham Gaafer, Jan Janssen, and Joerg Neugebauer — Max Planck Institute for Sustainable Materials, Düsseldorf, Germany

Machine learned interatomic potentials (MLIP) have gained popularity in materials science for their scalability and accuracy on par with the Density Functional Theory (DFT) training data. Based on the linear scaling with the number of neighbors, the primary focus in the recent years was increasing the flexibility of MLIPs to further improve their accuracy, as demonstrated by approaches such as Neural Network potentials and the Atomic Cluster Expansion (ACE). The Bessel functions and Chebyshev polynomials gained popularity as basis sets to represent the atomic bonds and orbitals. Nonetheless, the connection between an MLIP's basis set and its capability to represent the chemical complexity of various elements is not yet well understood.

In this study, we use ACE, as implemented in the Pacemaker software package, to investigate three non-magnetic transition metals (i.e., Al, Au, and Cu). For each element, we parameterize computationally efficient ACE potentials based on a minimal basis set to achieve a given root-mean-square error (RMSE) across both training and testing datasets. We find that the complexity of the MLIP primarily depends on the scaling of the per-atom energy distribution rather than the chemical complexity of the elements. Consequently, it is primarily a numerical effect rather than a chemical effect.

MM 18.2 Wed 10:30 H23

Physics-Based Generative Models: Enhanced Structure-Property Sampling in Inverse Materials Design — •Patricia König[1], Nicolas Bergmann[1], Piero Coronica[2], Chiara Panosetti[1], Hanna Türk[1], Karsten Reuter[1], and Christoph Scheurer[1] — [1]Fritz-Haber-Institut der MPG, Berlin — [2]Max Planck Computational and Data Facility, Munich

Data-driven approaches for the inverse design of novel materials with desired properties have become a key tool in materials discovery. Here, we introduce a framework using physics-based Generative Adversarial Networks for enhanced structure-property sampling via latent space design.

We are interested in sampling structures of two chemical systems associated with different relevant physical quantities, like the work function in the electrochemical adsorption of iodide and hydroxide on copper surfaces, and the oxygen chemical potential in the CO to CO_2 conversion over an amorphous RuO_2 catalyst. As part of our framework, we track and evaluate the structural diversity and convergence of our generator with machine-learning interatomic potentials and quantitative metrics. This enables a high throughput and cost-effective evaluation of structural guesses and their related properties to leverage the full potential of generative models. Concluding, we are showing on two model systems how to explore a vast chemical space of datasets with sparse areas, particularly structures with high free energies in transition states and diverse amorphous surface structures, thereby advancing the understanding and design of novel materials.

MM 18.3 Wed 10:45 H23

Active learning-based automated construction of Hamiltonian for structural phase transitions: a case study on $BaTiO_3$ — •Mian Dai[1], Yixuan Zhang[1], Nuno Fortunato[1], Peng Chen[2], and Hongbin Zhang[1] — [1]Institute of Materials Science, Technical University of Darmstadt, Darmstadt 64287, Germany — [2]Physics Department and Institute for Nanoscience and Engineering, University of Arkansas, Fayetteville, Arkansas 72701, USA

The effective Hamiltonians have been widely applied to simulate the phase transitions in polarizable materials, with coefficients obtained by fitting to accurate first-principles calculations. However, it is tedious to generate distorted structures with symmetry constraints, in particular when high-ordered terms are considered. In this work, we implement and apply a Bayesian optimization-based approach to sample the potential energy surface, automating the Hamiltonian construction by selecting distorted structures via active learning. Taking $BaTiO_3$ (BTO) as an example, we demonstrate that the Hamiltonian can be obtained using fewer than 30 distorted structures. Follow-up Monte Carlo simulations can reproduce the structural phase transition temperatures of BTO, comparable to experimental values with an error $<$ 10%. Our approach can be straightforwardly applied on other polarizable materials and paves the way for quantitative atomistic modelling of diffusionless phase transitions.

MM 18.4 Wed 11:00 H23

Comparing linear and deep learning surrogate models of materials electronic structure — •Valdas Vitartas, Chen Qian, James Kermode, and Reinhard Maurer — University of Warwick, Coventry, UK

The self-consistent electronic Hamiltonian matrix from Density Functional Theory (DFT) gives access to the electronic band structure and the density of states of a material, albeit at a large computational cost. Over recent years, several surrogate models based on linear parametrization and deep learning have been proposed to efficiently learn the electronic Hamiltonian as a function of the configuration and composition of materials. In this work, we compare two such models, the ACEhamiltonians [npj Comput. Mater. 8, 158] and MACE-H. Both provide a representation of the Hamiltonian in atomic orbital basis in terms of an equivariant many-body expansion of local atomic environments. In the case of ACEhamiltonians, the model parametrization is linear; for MACE-H, the representation serves as input to a message-passing neural network. The models are trained on reduced, valence-only Hamiltonian matrices for bulk gold and silicon generated from all-electron DFT via an approximately eigenspectrum-conserving transformation. We discuss the inherent strengths and weaknesses of the models by illustrating their accuracy, performance, data efficiency, and their ability to predict electronic quantities of interest for out-of-distribution configurations.

MM 18.5 Wed 11:15 H23

MACE-H: Equivariant Hamiltonian prediction with many-body expansion message passing — •Chen Qian, Valdas Vitartas, James Kermode, and Reinhard J. Maurer — University of Warwick, UK

The machine learning prediction of Kohn-Sham Density Functional Theory (DFT) Hamiltonians has the potential to accelerate the prediction of electronic properties, such as electronic band structures and electron-phonon coupling, while avoiding computationally expensive self-consistent field iterations. We introduce the MACE-H graph neural network, which combines the MACE body-order expansion message passing scheme with node-degree expansion blocks to efficiently generate messages that incorporate all relevant SO(3) irreducible representations. This model achieves high accuracy and high computational efficiency in capturing the local chemical environment. We demonstrate the model performance using several open materials benchmark datasets for 2D materials, achieving sub-meV prediction errors on matrix elements. Moreover, we discuss how the many-body expansion achieves higher data efficiency and examine its effect on out-of-distribution prediction for nanostructures featuring long-range interactions. To assess prediction outputs, we analyze the correlation between errors and hermiticity. The high computational efficiency and accuracy make the model a good candidate for electronic structure prediction in large-scale systems and high-throughput material screening.

15 min. break

MM 18.6 Wed 11:45 H23

Inferring Structure-Property Relationships with Artificial Intelligence: A Lignin Case Study — •Matthias Stosiek and Patrick Rinke — Technical University Munich, Munich, Germany

The potential of lignin as an abundant, underutilized biopolymer is increasingly being realized. A key challenge for the targeted production of lignins remains the poorly understood relation between lignin properties and its complex structure. Artificial intelligence (AI) methods could reveal such structure-function relationships but remain elusive in biomaterials research.

Structurally diverse lignins are extracted from birch wood combining the Aqua Solv Omni (AqSO) biorefinery process and AI-guided data acquisition[1]. Each lignin sample is characterized with 2D nuclear magnetic resonance (NMR) spectroscopy. A total of 95 collected NMR spectra are complemented with measurements of key lignin properties such as the antioxidant activity.

To establish structure-function relationships, we first correlate regions of the NMR spectra with the corresponding property measurements. Subsequently, we use RFR feature importance analysis to identify structural features that correlate with each property and provide a chemical interpretation of our findings. For instance, we find that a higher number of β-O-4 bonds leads to a lower surface tension in water indicating a more linear lignin structure. Our structure-inference approach is designed to be general and applicable to a wide range of materials and characterization data.

[1] D. Diment et al., ChemSusChem 2024, e202401711.

MM 18.7 Wed 12:00 H23

Machine learning electrostatics: Open challenges from batteries to proteins — •Max Veit — Technische Universität Darmstadt, DE

Long-range interactions such as electrostatics have long been a concern in developing accurate, efficient machine learning potential energy surfaces (ML-PES). Such potentials have now become established as a powerful technique allowing simulations of complex structures and processes with unprecedented realism and accuracy. However, the most widespread and successful methods to date do not incorporate any interactions beyond a fixed cutoff range, typically a few coordination shells. First, we need to ask the question of how methods with such an obvious limitation can be so successful, even applied to systems where long-range electrostatic interactions are known to be relevant. Second, we need to ask what approach, among the many proposed over the last decade or so, is the most appropriate if we want to incorporate long-range interactions in an accurate, efficient, and physically appropriate way. We investigate these questions in the context of a technologically relevant, experimentally accessible test system: lithium-intercalated graphite or nearly-graphitic nanoporous carbon. We first discuss the characteristics of this system that make it uniquely suited to machine learning simulation, then turn to the difficulties involved in defining what exactly makes a “good” electrostatic model, or long-range model in general, in the context of machine learning potentials, and finally discuss the implications for other systems – such as complex biomolecules – just out of the current reach of ML-PES simulations.

MM 18.8 Wed 12:15 H23

Data driven prediction of relative stability of binary and ternary TCP phases. — •Mariano Forti, Ralf Drautz, and Thomas Hammerschmidt — Interdisciplinary centre for advanced materials simulation, Ruhr-University Bochum

The study of precipitation of topological close packed(TCP) phases is of primary importance for the performance of superalloys. However, the structural complexity of these intermetallic compounds and the chemical complexity of the superalloys with typically up to ten elements hampers the exhaustive sampling of chemical space by density-functional theory (DFT) calculations. We overcome the related computational limitations by combining machine learning (ML) techniques with descriptors of the local atomic environment of the TCP phases and the use of interatomic potentials to predict phase properties with high precision. We illustrate our methodology studying the relative stability of the complex phases R, P, M and δ in binary and ternary systems produced from the main components in Co, Ni and Fe based superalloys.

MM 18.9 Wed 12:30 H23

Inverse Materials Design with Large Language Models — •Jan Janssen and Joerg Neugebauer — MPI for Sustainable Materials, Düsseldorf, Germany

Large language models (LLM) are trained on a vast amount of scientific literature to learn the included semantic, conceptional, and statistical relationships. The LLM applies these relationships to generate responses in natural language based on the context of the conversation. This raises the question: Can a LLM replace a scientist? Or how does the thought process of a scientist differ from the statistical approach of the LLM? Can the LLM make us better scientists?

We benchmark the capabilities of current LLMs to design new materials using atomistic simulations. While the required Python programming is challenging for the LLM and suffers from hallucination, this can be addressed with an agent-based approach by providing the LLM with a series of simulation workflows for the pyiron workflow framework. With these simulation workflows the LLM is not only capable to calculate material properties but can also invert the process and leverage statistical models to identify alloying compositions which match a pre-defined materials property, enabling inverse materials design.

Our benchmarks highlight the importance of developing scientific workflows. The more a workflow reduces the technical and scientific complexity of studying a given materials property the easier it is to use for LLMs and scientists alike. In this way LLMs also help us as scientists to validate and improve our scientific workflows. https://github.com/jan-janssen/LangSim

MM 18.10 Wed 12:45 H23

Workflow Utilities within the NOMAD Infrastructure: Lowering the Barrier to FAIR Data Management for Computational Materials Science — •J.F. Rudzinski[1], E. Boydas[1], N. Daelman[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], and S. Botti[7] — [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

NOMAD [nomad-lab.eu] [1] is an open-source, community-driven data infrastructure, focusing on materials science data. The NOMAD software can automatically extract data from the output of over 60 simulation codes, has been extensively expanded to support advanced many-body calculations and classical molecular dynamics simulations, and allows straightforward specialization via a rapidly developing plugin-based ecosystem. Both standardized and custom complex simulation workflows not only streamline data provenance and analysis but also facilitate the curation of AI-ready datasets. This contribution will focus on recently developed workflow functionalities and utilities within the NOMAD infrastructure. These advances enable highthroughput interfacing with the NOMAD repository, opening improved discovery pipelines by leveraging the benefits of NOMAD’s comprehensive and FAIR-compliant data management system [2].

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

MM 18.11 Wed 13:00 H23

Freedom of design: towards in silico design of molecules with desired quantum-mechanical properties — •Leonardo Medrano Sandonas[1], Alessio Fallani[2], Julian Cremer[3], Alexandre Tkatchenko[2], and Gianaurelio Cuniberti[1] — [1]TUD Dresden University of Technology, Germany. — [2]University of Luxembourg, Luxembourg. — [3]Pfizer Worldwide R&D, Germany.

The rational in silico design of chemical compounds requires a deep understanding of both the structure-property and property-property relationships that exist across chemical compound space (CCS), as well as efficient methodologies for defining an inverse property-to-structure mapping. In this presentation, we will discuss these relationships in the CCS sector spanned by small [Sci. Data 8, 43 (2021)] and large [Sci. Data 11, 742 (2024)] drug-like molecules, highlighting the existence of the ”Freedom of design” principle [Chem. Sci. 14, 10702 (2023)]. The insights gained are subsequently leveraged to design molecules with desired properties. To this end, we first developed a variational autoencoder (VAE) approach and demonstrated that CCS can be parameterized using a finite set of quantum-mechanical (QM) properties [Nat. Commun. 15, 6061 (2024)]. We showcased the capabilities of this method by conditionally generating de novo molecular structures, interpolating transition paths for chemical reactions, and providing insightful insights into property-structure relationships. We expect our work will contribute to the development of advanced generative frameworks that enhance the in silico design and identification of molecules for specific chemical processes.