MM 4: Data Driven Material Science: Big Data and Workflows I

Time: Monday 10:15–13:00 Location: C 243

MM 4.1 Mon 10:15 C 243

Investigating Structural Descriptors for High-Dimensional Neural Network Potentials — •Moritz R. Schäfer[1,2], Moritz Gubler[3], Stefan Goedecker[3], and Jörg Behler[1,2] — [1]Theoretische Chemie II, Ruhr-Universität Bochum, Germany — [2]Research Center Chemical Sciences and Sustainability, Research Alliance Ruhr, Germany — [3]Department of Physics, University of Basel, Switzerland

High-dimensional neural network potentials (HDNNPs) are a well established technique to efficiently compute energies and forces akin to ab initio standards for conducting extensive molecular dynamics simulations of intricate systems in high dimensions. This method expresses the total energy from environment-specific atomic energy contributions, with the option to incorporate electrostatic interactions utilizing flexible atomic charges. The reliability of both components significantly depends on the accuracy of the structural descriptors used to define the atomic environments. Here, we combine atom-centered symmetry functions with the newly introduced overlap matrix descriptor. Furthermore, we analyze the strengths and weaknesses of each descriptor, providing insights through demonstrations on benchmark systems.

MM 4.2 Mon 10:30 C 243

Universally Accurate or Specifically Inadequate? Stress-Testing General Purpose Machine Learning Interatomic Potentials — •Konstantin Jakob[1], Karsten Reuter[1], and Johannes T. Margraf[1,2] — [1]Fritz-Haber-Insttitut der MPG, Berlin — [2]Universität Bayreuth

Machine learning interatomic potentials (MLIPs) have revolutionized the field of atomistic materials simulation, both due to their remarkable accuracy - when trained adequately - and their computational efficiency compared to established *ab initio* methods. Very recently, several general purpose MLIPs have been reported, which are broadly applicable across the periodic table. These represent a fascinating opportunity for materials discovery, provided that they are robust and transferable. In this context, metastability and polymorphism pose significant challenges, as the underlying training data sets cannot cover the full space of such structures and compositions. In order to stress test current general purpose MLIPs, we evaluate models based on the M3GNet and MACE architectures on a unique set of inorganic, crystalline materials generated by atom substitutions. Validating these two models, we shine light on both successes and drawbacks of using general purpose MLIPs and evaluate the opportunities that further research can hold.

MM 4.3 Mon 10:45 C 243

Pressure-transferable neural network models for density-functional theory — •Timothy Callow[1], Lenz Fiedler[1], Normand Modine[2], and Attila Cangi[1] — [1]Center for Advanced Systems Understanding, Helmholtz-Zentrum Dresden-Rossendorf, Untermarkt 20, Görlitz, 02826, Saxony, Germany — [2]Computational Materials and Data Science, Sandia National Laboratories, 1515 Eubank Blvd, Albuquerque, 87123, NM, USA

Density functional theory (DFT) is well-known as the workhorse of electronic structure calculations in materials science and quantum chemistry. However, its applications stretch beyond these traditionally-studied fields, such as to the warm-dense matter (WDM) regime. Under WDM conditions, there are different challenges to consider (compared to ambient conditions) when using DFT. Namely, the electronic structure problem must be solved (i) for large particle numbers, (ii) for a range of temperatures, and (iii) for a range of pressures. Promising solutions were demonstrated for problems (i) and (ii) [1,2] using a recently-developed workflow to machine-learn the local density of states (LDOS) [3]. In this talk, we discuss our progress in developing a solution for problem (iii). This problem presents additional challenges because the LDOS varies quite significantly with changes in the pressure, making it a difficult problem for neural network models.

[1] L Fiedler et al., npj Comput Mater 9, 115 (2023) [2] L Fiedler et al., Phys. Rev. B 108, 125146 (2023) [3] J. A. Ellis et al., Phys. Rev. B 104, 035120 (2021)

MM 4.4 Mon 11:00 C 243

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

The Density Functional Tight Binding (DFTB) approach allows electronic structure based simulations at length and time scales far beyond what is possible with first-principles methods. Unfortunately, the sparse availability of DFTB parameters across the periodic table is a significant barrier to the use in many cases. To this end, we introduce a systematic workflow for the robust parameterization of DFTB across the periodic table. Specifically, the parameters for most elements are obtained via Bayesian Optimization on a consistent set of real and artificial elemental solids, spanning a wide range of coordination environments. Elements which do not form stable elemental solids are treated separately, using representative binary compounds as reference. In seeking to bridge the gap towards higher-level quantum mechanical theories, the use of the DFTB parameterization in multi-fidelity machine learning is explored.

MM 4.5 Mon 11:15 C 243

Application of Question Answering method to extract information from materials science literature — •Matilda Sipilä[1], Farrokh Mehryary[2], Emil Nuutinen[2], Sampo Pyysalo[2], Filip Ginter[2], and Milica Todorović[1] — [1]Department of Mechanical and Materials Engineering, University of Turku, Turku, Finland — [2]Department of Computing, University of Turku, Turku, Finland

Scientific text is a promising source of data in materials science, and there is ongoing research how to utilise textual data in materials discovery. In addition to the more established approaches like named entity recognition or dictionary based methods, new machine learning tools such as question answering (QA) are becoming available. The advantages of this method are that it is easy to scale and that it does not require manual text labeling or annotating work, but there may be some loss in precision compared to other methods.

We tested the performance of the QA method on the well-known task of information extraction. We extracted bandgap values of halide perovskite materials from scientific literature. Large language models were tuned towards a specific QA task and then used to select the correct answer for the question about materials properties. In comparison to more established methods, the QA method performed well and we were able to extract correct information from text. This information can be used to map the space of materials properties and find promising new materials solutions. The potential in QA method lies in versatility, accessibility and scalability, since it is easy to use even for researchers with no previous knowledge of language technology.

15 min. break

MM 4.6 Mon 11:45 C 243

Transferable interatomic potential of water with the atomic cluster expansion — •Eslam Ibrahim, Yury Lysogorskiy, and Ralf Drautz — ICAMS, Ruhr Universität Bochum, 44780 Bochum, Germany

We present a transferable parameterization of water using the Atomic Cluster Expansion (ACE). Our approach efficiently samples liquid water by employing static calculations of various ice phases. The active learning feature of ACE-based D-optimality algorithm is utilized to select relevant water configurations, circumventing computational challenges associated with ab-initio molecular dynamics (AIMD) simulations. Our results demonstrate that ACE descriptors enable a potential fitted solely on ice structures to provide a very good description of liquid water. The developed potential shows remarkable agreement with first-principles references, accurately capturing structural and dynamic properties of liquid water. This includes pair correlation functions, covalent bonding profiles, hydrogen bonding profiles, diffusion coefficient, and thermodynamic properties like the melting point of water. This work introduces an efficient sampling technique for machine learning potentials in water simulations, along with a transferable interatomic potential that rivals the accuracy of ab-initio references. This advancement enhances our understanding of water's behavior at the atomic level and opens new avenues for studying complex aqueous systems.

MM 4.7 Mon 12:00 C 243

Automatic extraction and analysis of dislocations in atom probe tomography data using skeletonization — •Alaukik Saxena, Baptiste Gault, and Christoph Freysoldt — Max-Planck-Institut für Eisenforschung GmbH, Düsseldorf 40237, Germany

Atom probe tomography (APT) is a unique technique that provides 3D elemental distribution with a near-atomic resolution for a given material. Extracting and analyzing microstructural features in 3D APT data is challenging and time-intensive, given their complex morphology. Here, we introduce a workflow to systematically extract linear microstructural features, particularly dislocations, from the APT data. The workflow extracts isosurface meshes from APT data and, as a preprocessing step, filters them using principal component analysis (PCA) to find geometrically anisotropic microstructural features. Further, a topology analysis concept called skeletonisation is applied to extract the linear graphs or skeletons of each mesh. Since the skeleton encapsulates the underlying geometry of a mesh, it is used to identify and segment linear features or dislocation segments even in very complex microstructures containing, for example, dislocation networks. This enables a robust composition and geometric analysis of dislocations in APT data. Additionally, the workflow integrates crystallographic data from APT to determine dislocation orientation in the crystal coordinate system. Overall, this advanced workflow significantly reduces manual effort and opens new possibilities for high-throughput studies in material science.

MM 4.8 Mon 12:15 C 243

Stable diffusion based microstructure reconstruction and generation — •Yixuan Zhang[1], Teng Long[2], and Hongbin Zhang[1] — [1]Institute of Materials Science, Technical University of Darmstadt, 64287, Darmstadt, Germany — [2]School of Materials Science and Engineering, Shandong University, 250061, Jinan, China

In recent years, the reconstruction and generation of microstructures have become pivotal in understanding and predicting the mechanical and functional properties of materials. This study introduces a novel approach to microstructure reconstruction based on stable diffusion models. Our implementation employs a stable diffusion model to capture the intricate patterns and features inherent in microstructures, which can be adapted to further refine reconstructed the phase and grain orientation of microstructures, ensuring their statistical and morphological fidelity to the original samples. The model is trained using a comprehensive dataset of 500,000 synthetic micrographs, ensuring the model's robustness and versatility across various material classes. Our results demonstrate that our approach outperforms conventional methods in terms of accuracy, speed, and adaptability. The reconstructed microstructures exhibit remarkable similarity to their counterparts, both qualitatively and quantitatively. Furthermore, the generative capabilities of our model pave the way for optimizing novel microstructures, aiding in the design of materials with desired properties.

MM 4.9 Mon 12:30 C 243

Automatic Generation of Atomic Structure Datasets for Machine Learning Potentials: Alloys and Applicatoin to Mg/Al/Ca — •Marvin Poul[1], Liam Huber[2], and Joerg Neugebauer[1] — [1]Max-Planck- Institut für Eisenforschung, Düsseldorf, Germany — [2]Grey Haven Solutions, Victoria, Canada

We extend a recently proposed strategy for automatically generating training data for machine learning interatomic potentials (MLIP) to alloys.[1]

It is based on small periodic structures (around ten atoms) of various concentrations that are sampled from all crystallographic space groups. These structures are minimized and then again randomly perturbed in positions and cell shape around the obtained local minima. This procedure akin to ab initio random structure search yields samples around the relevant parts of phase space without prior knowledge automatically. Only the cell stoichiometry and the magnitude of the random perturbations remain hyperparameters in this approach.

We explore the natural question of how well potentials can extrapolate in the combinatorically large concentration space and test that they accurately describe structures near the convex hull as well as larger super cells of random alloys. Finally we verify the potentials on binary phase diagrams (and defect phase diagrams) in the ternary Mg/Al/Ca system.

This opens the way for automatic parametrization of MLIPs, promising to bring ab initio accuracy to a large number of systems at scale.

[1]: https://doi.org/10.1103/PhysRevB.107.104103

MM 4.10 Mon 12:45 C 243

Physics-informed neural network for predicting the Gibbs free energy — •Clement Paulson[1], Amin Sakic[2], Vedant Dave[3], Elmar Rueckert[3], Ronald Schnitzer[1], and David Holec[1] — [1]CDL KnowDAS, Department of Materials Science, Montanuniversität Leoben, Austria — [2]Department of Materials Science, Montanuniversität Leoben, Austria — [3]CPS Lab, Montanuniversit Leoben, Austria

We employ a physics-informed neural network approach in conjunction with the CALPHAD formalism to determine the Gibbs free energy of alloys. The Gibbs free energy, essential for CALPHAD simulations, is determined by predicting the Redlich-Kister parameter using a composite neural network utilizing novel descriptors derived from the atomic, composition-based, and thermodynamic properties of elements. The composite neural network comprises a low-fidelity network trained on CALPHAD-generated mixing enthalpies and a high-fidelity network trained on experimental mixing enthalpies. These two models are further connected to a physics-informed neural network, which determines the Redlich-Kister parameters. The predicted Redlich-Kister parameters can then be directly implemented into a thermodynamic database file for immediate use with existing CALPHAD software. This approach holds promise for expediting materials development and phase stability determination. Comparative experimental results highlight the accuracy and potential of this deep learning-based method, offering a novel path for forecasting the Gibbs free energy in multicomponent systems and accelerating the development of databases.