MM 53: Data Driven Material Science: Big Data and Workflows VI

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

MM 53.1 Thu 10:15 C 243

Scalable machine learning for predicting the electronic structure of matter — •Attila Cangi[1,2], Lenz Fiedler[1,2], Bartosz Brzoza[1,2], Karan Shah[1,2], Tim Callow[1,2], and Steve Schmerler[1] — [1]Helmholtz-Zentrum Dresden-Rossendorf, Dresden, Germany — [2]Center for Advanced Systems Understanding, Görlitz, Germany

I will present our recent progress in significantly speeding up density functional theory calculations with machine learning [1], for which we have developed the Materials Learning Algorithms framework [2]. Our findings illustrate significant improvements in calculation speed for metals at their melting point. Additionally, our use of automated machine learning has yielded significant reductions in computational resources required to identify optimal neural network architectures [3]. Most importantly, I will present our latest breakthrough, which enables fast neural-network driven electronic structure calculations for systems unattainable by conventional density functional theory calculations [4].

[1] L. Fiedler, K. Shah, M. Bussmann, A. Cangi, Phys. Rev. Materials, 6, 040301 (2022). [2] J. Ellis, L. Fiedler, G. Popoola, N. Modine, J. Stephens, A. Thompson, A. Cangi, S. Rajamanickam, Phys. Rev. B, 104, 035120 (2021). [3] L. Fiedler, N. Hoffmann, P. Mohammed, G. Popoola, T. Yovell, V. Oles, J. Austin Ellis, S. Rajamanickam, A. Cangi, Mach. Learn.: Sci. Technol., 3, 045008 (2022). [4] L. Fiedler, N. Modine, S. Schmerler, D. Vogel, G. Popoola, A. Thompson, S. Rajamanickam, A. Cangi, npj. Comput. Mater., 9, 115 (2023).

MM 53.2 Thu 10:30 C 243

SE(3)-Transformers for predicting the electronic structure of hydrogen molecules — •Bartosz Brzoza and Attila Cangi — CASUS/HZDR

In this work, we demonstrate the efficacy of a neural network model implemented as the Materials Learning Algorithms (MALA) package in predicting the electronic structure of a system of hydrogen molecules under various pressure and temperature conditions across the molecular liquid-solid phase boundary, demonstrating the potential of our methods for molecular systems. Additionally, we investigate the use of SE(3)-Transformer Graph Neural Networks to improve the generalizability and extrapolation capabilities of our models. Our results indicate that the MALA framework provides a powerful and efficient tool for accelerating Kohn-Sham density functional theory calculations in molecular systems. This work paves the way for future research in developing advanced machine-learning algorithms for accelerating electronic structure calculations both accurately and efficiently.

MM 53.3 Thu 10:45 C 243

Physics-Informed Machine Learning for Addressing Challenges in Static and Time-Dependent Density Functional Theory — •Karan Shah and Attila Cangi — Center for Advanced Systems Understanding, Helmholtz-Zentrum Dresden-Rossendorf, Görlitz, Germany

We explore the potential of Physics-Informed Machine Learning (ML) in addressing key computational tasks in both static and time-dependent Density Functional Theory (DFT/TDDFT). The talk will focus on two projects that employ advanced ML techniques, specifically Physics-Informed Neural Networks (PINNs) and Fourier Neural Operators (FNOs), to tackle these complex tasks.

In the first part, we examine the use of FNOs in addressing the density-to-potential inversion problem in static DFT. By employing these methods as alternatives to numerical inversion schemes, we demonstrate enhancements in predictive transferability and speed. We highlight the applications to exactly solvable systems, illustrating their potential as accurate and rapid data-driven surrogate models.

In the second part, we discuss the application of PINNs to accelerate TDDFT calculations. By incorporating the fundamental physical constraints of the TD Kohn-Sham equations directly into the learning process, we demonstrate the performance and generalisability of PINN solvers on the time evolution of model systems across varying system parameters, domains and energy states.

By integrating physics and ML, these projects shed light on promising new directions for addressing challenges in DFT and TDDFT.

MM 53.4 Thu 11:00 C 243

Datadriven thermodynamic modeling with CALPHAD — •Tobias Spitaler and Lorenz Romaner — Montanuniversität Leoben, Department Werkstoffwissenschaft, Leoben, Österreich

CALPHAD models and computational thermodynamics play an essential role in materials science and in the development of novel materials. In the CALPHAD method, thermodynamic quantities and phase diagrams are calculated from a parameterized model of the Gibbs free energy, which is stored in a thermodynamic database. The reliability of the calculated quantities relies on the correctness and quality of the thermodynamic database. With new computational tools and statistical methods, the database creation can be accelerated and uncertainty can be quantified, which is propagated from the input data to the quantities of interest.

We combine heterogeneous data from experiment and simulation in CALPHAD modeling and use statistical tools to propagate the uncertainty of the model parameters to quantities of interest. We demonstrate parameter optimization and uncertainty quantification in the phase diagram of selected systems (e.g. W-Ti, Fe-C). Uncertainty quantification of phase boundaries, invariant points and other quantities of interest are demonstrated. With the statistical methods regions with high uncertainty in the composition space can be identified and the potential experiments with the highest information proposed. With the data-driven and statistical approach to CALPHAD modeling, new thermodynamic databases can be obtained in a faster and more reproducible way.

MM 53.5 Thu 11:15 C 243

Complete Basis Set Limit Extrapolation in Density Functional Theory Calculations using Statistical Learning — •Daniel Speckhard[1,2], Claudia Draxl[1,2], and Matthias Scheffler[2,3] — [1]Humboldt-Universität zu Berlin, Physics Department and IRIS Adlershof, Berlin, Germany — [2]Max-Planck-Institut für Festkörperforschung, Stuttgart, Germany — [3]The NOMAD Laboratory at the FHI of the Max-Planck-Gesellschaft and IRIS-Adlershof of the Humboldt-Universität zu Berlin

The numerical precision of density-functional-theory (DFT) calculations depends on a variety of computational parameters, one of the most critical being the basis-set size. The ultimate precision is reached with an infinitely large basis set, i.e., in the complete basis-set (CBS) limit. Our aim is to find a machine-learning model that extrapolates finite basis-size calculations to the CBS limit. Quantile random forests and symbolic regression, applying the SISSO approach, are used to estimate total energies, lattice parameters, and band gaps as a function of the basis-set size. The random-forest model outperforms previous approaches in the literature for both codes, while SISSO outperforms the random-forest model for the `exciting` code. Our approach also provides prediction intervals, which quantify the models' uncertainty. We evaluate our work on datasets consisting of 63 binary solids and 4000 binary semiconductors, respectively.

[1] C. Carbogno *et al.*, *npj Comput. Mater.* **8**, *69 (2022).*

15 min. break

MM 53.6 Thu 11:45 C 243

High-resolution beyond the depth of field limit in 3D phase-contrast imaging using ptychographic multi-slice electron tomography — •Andrey Romanov[1], Min Gee Cho[2], Mary Cooper Scott[2,3], and Philipp Pelz[1] — [1]Institute of Micro- and Nanostructure Research (IMN) & Center for Nanoanalysis and Electron Microscopy (CENEM), Friedrich Alexander-Universität Erlangen-Nürnberg, IZNF, 91058 Erlangen, Germany — [2]University of California Berkeley, Berkeley, California 94720, United States — [3]Lawrence Berkeley National Laboratory, Berkeley, California 94720, United States

In electron microscopy, achieving atomic resolution in large volumes remains challenging, despite progress in optics and computational algorithms. Electron ptychography allows 3D information extraction from a single position- and momentum-resolved (4D-STEM) dataset, but axial resolution is limited to around 3nm. A novel approach involving tilt-series 4D-STEM data and subsequent ptychographic and tomographic reconstruction surpasses previous constraints. Here, we present the reconstruction of a phase-contrast volume three times

greater than the depth of field, enabling differentiation of individual atoms in all dimensions of a Co3O4 nanocube in an 18.2 nm side length volume. This advancement opens new possibilities for material exploration in atomic-level 3D imaging.

MM 53.7 Thu 12:00 C 243

Advancing In-Situ SEM Imaging: Integrating Deep Learning Super-Resolution for Accelerated Analysis — •Tom Reclik[1], Philipp Schumacher[1], Setareh Medghalchi[1], Maximilian Wollenweber[1], Ulrich Kerzel[2], and Sandra Korte-Kerzel[1] — [1]Institute for Physical Metallurgy and Materials Physics, RWTH Aachen University, Aachen, Germany — [2]Data Science and Artificial Intelligence in Materials and Geoscience, Faculty of Georesources and Materials Engineering, RWTH Aachen University, Aachen, Germany

Scanning Electron Microscopy (SEM) is pivotal in revealing intricate micro and nanoscale features across various research fields. However, obtaining large-area high-resolution SEM images presents challenges, including prolonged scanning durations and potential sample degradation due to extended electron beam exposure. This challenge is significantly amplified, when the time dimension in in-situ experiments is added. Here we present a new in-situ workflow, accelerating the imaging process step, by employing deep learning super-resolution algorithms coupled with the automated high resolution rescanning of points of interest. Our approach significantly speeds up the imaging process, thereby enabling the discoveries of new physics with a high statistical relevancy.

MM 53.8 Thu 12:15 C 243

Bayesian Optimization for High-Resolution Transmission Electron Microscopy — •Xiankang Tang, Yixuan Zhang, and Hongbin Zhang — Institute of Materials Science, TU Darmstadt, 64287 Darmstadt, Germany

High-resolution transmission electron microscopy (HRTEM) allows to study the atomic structure of solid materials with a resolution of sub-Angstrom. By matching experimental and simulated images, unknown experimental parameters and crystal structures can be determined. However, this process entails strong domain expertise and can be time consuming. In this work, we implement and apply a Bayesian optimization-based approach to improve the efficiency of the image matching process. After adopting properly defined loss functions capturing both the global and local image features, it is demonstrated that our approach can not only match the experimental and simulated images in terms of absolute image contrast, but also naturally identify the unknown parameters in the experiments. This method offers a new possibility for automated HRTEM image analysis.We acknowledge Dr. Lei Jin and Prof. Rafal Dunin-Borkowski from FZ Jülich for stimulating discussions.

MM 53.9 Thu 12:30 C 243

Pydidas: A new tool for automated X-ray diffraction data analysis — •Malte Storm, Anton Davydok, Peter Staron, and Christina Krywka — Helmholtz-Zentrum Hereon, Institute of Materials Physics, Max-Planck-Strasse 1, 21502 Geesthacht

The push towards faster experiments and generally the enhanced data rates at latest generation synchrotrons require new efforts on the analysis side to keep processing speed up with experimental speed. Especially fast *in situ* or *in operando* experiments generate huge data quantities and require fast feedback (in the form of data processing) during experiments to optimize experimental results of beamtimes.

Historically, data reduction and pre-processing of X-ray diffraction datasets has been very much burden to the user groups. To help existing users and attract new user groups who do not have the technical experience to perform the required processing steps, Hereon has developed the pydidas software [1] suite for X-ray diffraction data analysis.

Pydidas is explicitly designed to allow full data analysis in a single tool, to natively process modern data container types (HDF5) and to make efficient use of parallel computing. Data browsing and display, experiment calibration, workflow setup, processing and visualization are all available from within pydidas. Emphasis has been placed on an intuitive user interface and accessibility also for non-experts. Pydidas is open source software and publicly available.

[1] http://pydidas.hereon.de

MM 53.10 Thu 12:45 C 243

Deep learning-based feature detection on 2D X-ray scattering data for high throughput data analysis — •Alexander Hinderhofer, Vladimir Starostin, Constantin Voelter, Alexander Gerlach, and Frank Schreiber — Institut für Angewandte Physik, Universität Tübingen, Auf der Morgenstelle 10, 72076 Tübingen, Germany

In situ real-time grazing-incidence X-ray diffraction (GIXD) is a key technique for thin film structural characterization during sample preparation. In-situ GIXD can produce large amounts of data, on the scale of several thousand images per hour, frequently exceeding the capabilities of traditional data processing methods. We propose an automated pipeline for the analysis of GIXD images, based on the Faster Region-based Convolutional Network architecture for object detection, modified to conform to the specifics of the scattering data. The model exhibits high accuracy in detecting diffraction features on noisy patterns with various experimental artifacts. We demonstrate our method on real-time tracking of organic-inorganic perovskite structure crystallization. By design, our approach is equally suitable for other crystalline thin-film materials.[1] Further, we discuss a high quality GIXD dataset with more than 1600 labeled features for performance evaluation of feature detection models in GIXD.

[1] V. Starostin et al. npj Comput Mater 8 (2022) 101