O 67: Frontiers of Electronic Structure Theory: Focus on Artificial Intelligence Applied to Real Materials 3

Time: Thursday 10:30–12:45 Location: S054

O 67.1 Thu 10:30 S054

Quantile Random Forest Model for Extrapolation to the Complete Basis Set Limit in Density Functional Theory Calculations — •Daniel Speckhard[1], Christian Carbogno[2], Sven Lubeck[2], Luca Ghiringhelli[2], Matthias Scheffler[2,1], and Claudia Draxl[1,2] — [1]Humboldt-Universität zu Berlin, Physics Department and IRIS Adlershof, Berlin, Germany — [2]The NOMAD Laboratory at the FHI-MPG and HU, Berlin, Germany

The precision of density-functional theory (DFT) calculations depends on a variety of computational parameters, the most critical being the basis-set size. With an infinitely large basis set, i.e., in the limit of a complete basis set (CBS), the result of the calculation is as precise as possible for the chosen exchange-correlation functional. Our aim in this work is to find a model that can extrapolate the result of an imprecise DFT calculation to the CBS limit. As a starting point, we use a dataset of 63 binary solids investigated with various basis-set sizes [1] with two all-electron DFT codes, `exciting` and *FHI-aims*, which use very different types of basis sets. A quantile random forest model is used to estimate the deviation of the total energy with respect to fully converged calculations as a function of the basis set size. The non-linear random forest model outperforms a previous approach that used a linear model. The quantile random forest model presented also provides prediction intervals which give the user an idea of the model's uncertainty.

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

O 67.2 Thu 10:45 S054

Symmetry and completeness in machine-learning models for atomistic simulations — •Sergey Pozdnyakov and Michele Ceriotti — EPFL, Switzerland

During the last decade, machine learning methods have drastically changed atomistic simulations. On the one hand, they scale linearly with the size of the system and thus, are significantly faster than the quantum mechanical calculations. On the other, they provide a functional form that is much more flexible than so-called classical force fields such as the Lennard Jones potential or embedded atom models. From one point of view, incorporating rotational symmetry is important for ML since it can make models more data-efficient and robust, but can also lead to incompleteness, limiting the ultimate accuracy of the model. I will discuss some examples of this and compare different types of models to show how one can find an optimal balance of the two effects.

O 67.3 Thu 11:00 S054

Fast, robust, interpretable machine-learning potentials — Stephen R. Xie[1,2], Richard G. Hennig[1], and •Matthias Rupp[3] — [1]University of Florida, Gainsville, USA — [2]KBR, NASA Ames Research Center, Mountain View, USA — [3]University of Konstanz, Germany

Machine-learning potentials (MLPs) are increasingly successful in all-atom dynamics simulations where they act as surrogate models for ab-initio electronic structure methods. MLPs often result in two to three orders of magnitude improvements in the number of simulated atoms or duration of simulated time, enabling new insights and applications. Current limitations include data inefficiency, instabilities ("holes" in high-dimensional MLPs [2]), and lack of interpretability.

To address this challenge, we combine effective two- and three-body potentials in a cubic B-spline basis with second order-regularized linear regression. The resulting "ultra-fast potentials" are data-efficient, physically interpretable, sufficiently accurate for applications, can be parametrized automatically, and are as fast as the fastest traditional empirical potentials. [1] We demonstrate these qualities in retrospective benchmarks and present the prediction of thermal conductivities via the Green-Kubo formalism as a first application.

[1] Stephen R. Xie, Matthias Rupp, Richard G. Hennig, Ultra-fast interpretable machine-learning potentials. arXiv:2110.00624, 2021 [2] Jeffrey Li, Chen Qu, Joel M. Bowman: Diffusion Monte Carlo with fictitious masses finds holes in potential energy surfaces, Mol. Phys. 119(17–18): e1976426, 2021.

O 67.4 Thu 11:15 S054

Improving the transferability of high-dimensional neural network potentials by low-order terms — •Alea Miako Tokita and Jörg Behler — Georg-August-Universität Göttingen, Institut für Physikalische Chemie, Theoretische Chemie, Tammannstraße 6, 37077 Göttingen, Germany

High-dimensional neural network potentials (HDNNPs) are able to provide accurate potential energy surfaces suitable for atomistic simulations of large systems. The key to this accuracy is the high flexibility of the atomic neural networks allowing to reproduce energies and forces from reference electronic structure calculations with very small errors. At the same time, this flexibility is limiting the transferability of HDNNPs to atomic configurations that are very different from the reference geometries. Here, we investigate possible improvements in transferability of HDNNPs by the explicit inclusion of low-order terms in the functional form of the potential. The performance is demonstrated for a series of molecular model systems.

O 67.5 Thu 11:30 S054

Predicting condensed-phase electron densities using machine learning — •Alan Lewis[1], Andrea Grisafi[2], Michele Ceriotti[2], and Mariana Rossi[1] — [1]MPI for Structure and Dynamics of Materials, Hamburg, Germany — [2]École Polytéchnique Fédèrale de Lausanne, Lausanne, Switzerland

The electron density is a fundamental quantity for understanding physical phenomena in materials, and is central to electronic structure theories such as density-functional theory. We present the SALTED machine learning method and demonstrate its ability to learn and predict the electronic densities of a range of materials from simple liquids and metals to hybrid organic-inorganic perovskites. This extends the framework presented in ACS Cent. Sci. 5, 57, 2019 to work with periodic boundary conditions and uses a resolution of the identity on a numeric atom-centered orbital basis to expand the all-electron periodic density. A Gaussian process regression model that makes use of local symmetry-adapted representations of the atomic structure is employed, making our method both data-efficient and highly transferable.[1] We also compare various methods of dealing with the non-orthogonality of the basis, accounting for correlations between pairs of off-centered density components, finding that the best compromise between accuracy and computational efficiency comes from approximating the density expansion coefficients by directly minimizing the loss function. The total energies derived from the densities obtained in this way present errors with respect to DFT of just 0.1 meV/atom.

[1] Lewis, Grisafi, Ceriotti, Rossi, JCTC 17, 11, 7203 (2021)

O 67.6 Thu 11:45 S054

Equivariant N-center representations for machine learning molecular Hamiltonians — •Jigyasa Nigam, Michael Willatt, and Michele Ceriotti — Laboratory of Computational Science and Modeling, Institute of Materials, Ecole Polytechnique Federale de Lausanne, 1015 Lausanne, Switzerland

Most of the widely used machine learning schemes that have been successful in predicting chemical and material properties rely on concise, symmetry-adapted descriptions of the underlying atomic structure. A class of these structural descriptions is built on hierarchical correlations of atom-centered densities(ACDC)[1]. These are subsequently used to model corresponding atomic properties or atomic contributions to a global observable. However, many quantum mechanical quantities, such as the effective single-particle Hamiltonian written on an atomic-orbital basis, are associated with multiple atom-centers. This effectively renders ACDCs inadequate to describe the additional degrees of freedom of such multicenter properties. We recently proposed an N-centered representation[2] that extends the ACDC framework to the case of targets that are simultaneously indexed by N atoms. I will demonstrate how devising a family of N-center representations opens avenues for new classes of machine learning models that are fully equivariant and describe their role in assisting electronic structure calculations.

[1] J. Nigam, S. Pozdnyakov, M. Ceriotti, JCP 153,121101, 2020
[2] J. Nigam, M. Willatt, M. Ceriotti, JCP 156, 014115, 2022

O 67.7 Thu 12:00 S054

Similarity-of-materials analysis for reusability and interoperability of data in materials databases — •Šimon Gabaj, Martin Kuban, Santiago Rigamonti, and Claudia Draxl — Humboldt-Universität zu Berlin, Zum Großen Windkanal 2, 12489 Berlin, Germany

Large computational materials databases, such as NOMAD [1-2], make it possible to reuse already generated materials data. All these data typically come from different sources and have been created for a different purpose, be it geometry optimization, electronic structure, or alike. To support interoperability and thus reusability of such data, we devise a data-analysis workflow making use of similarity fingerprints. First, we encode the electronic density of states (DOS) in a vectorial representation [3] to obtain a descriptor. Then, we employ the Tanimoto coefficient to compute the similarity between all pairs of calculations. We demonstrate our workflow with selected materials, chosen from the NOMAD database. In the ideal case, all calculations of the same material in the same geometry should be identical. This is, however, not observed. Using our approach, we can uncover correlations between the DOS similarity and methodology as well as computational parameters. This way, we can also identify parameters that are relevant for the convergence of results.

[1] Draxl, C., Scheffler, M., *MRS Bulletin*, **43**, 676, (2018)
[2] Draxl, C., Scheffler, M., *J. Phys. Mater.*, **2**, 036001, (2018)
[3] Kuban, M., *et al.*, to be published

O 67.8 Thu 12:15 S054

Supervised and unsupervised deep Learning of topological phase transitions from entanglement aspect for one- and two-dimensional chiral p-wave superconductors — •Ming-Chiang Chung — Max-Planck-Institut für Physik komplexer Systeme, Dresden, Germany — National Chung-Hsing University, Taichung, Taiwan

The one-dimensional or two-dimensional chiral p-wave superconductor proposed by Kitaev has long become a classic example for understanding topological phase transitions through various methods, such as examining the Berry phase, edge states of open chains, and, in particular, aspects from quantum entanglement of ground states. In order to understand the amount of information carried in the entanglement-related quantities, here we study topological phase transitions of the model with emphasis of using the deep learning approach. Using both supervised or unsupervised ways, we feed different quantities, including Majorana correlation matrices (MCMs), entanglement spectra (ES) or entanglement eigenvectors (EE) originating from Block correlation matrices, into the deep neural networks for training, and investigate which one could be the most useful input format in this approach. We find that ES is information that is too compressed compared to MCM or EE. MCM and EE can provide us abundant information to recognize not only the topological phase transitions in the model but also phases of matter with different U(1) gauges, which is not reachable by using ES only. We also build a procedure for using unsupervised learning to find the phase transition points. We have used this method for other models.

O 67.9 Thu 12:30 S054

Machine Learning the Square-Lattice Ising Model — •Burak Çivitcioğlu[1], Andreas Honecker[1], and Rudolf A. Römer[2] — [1]Laboratoire de Physique Theorique et Modelisation, CNRS UMR 8089, CY Cergy Paris Universit *e, Cergy-Pontoise, France — [2]Department of Physics, University of Warwick, Coventry, CV4 7AL, United Kingdom

Recently, machine-learning methods have been shown to be successful in identifying and classifying different phases of the square-lattice Ising model. We study the performance and limits of classification and regression models. In particular, we investigate how accurately the correlation length, energy and magnetisation can be recovered from a given configuration. We find that a supervised learning study of a regression model yields good predictions for magnetisation and energy, and acceptable predictions for the correlation length.