O 83: Frontiers of Electronic Structure Theory: Focus on Artificial Intelligence Applied to Real Materials 4

Time: Friday 10:30–13:00   Location: S054

O 83.1 Fri 10:30 S054

Alchemical machine learning for high entropy alloys —
•Nataliya Lopanitsyna, Guillaume Fraux, and Michele Ceriotti — École Polytechnique Fédérale de Lausanne, Switzerland

High entropy alloys (HEAs) are a class of metallic materials composed of five or more principal elements. Interest in HEAs has grown over the last decades due to their exceptional structural and mechanical properties. HEAs are particularly challenging for atomistic modeling. Machine-learning (ML) models have emerged as a promising alternative to inaccurate empirical forcefields and very demanding first-principles simulations, with the ability to deliver the accuracy of first principle methods with lower computational resources. However, the complexity of ML models grows exponentially with the number of different elements due to the unfavourable scaling of their associated feature space sizes, limiting the chemical diversity of the systems tackled thus far. To address the problems arising from the high feature space dimensionality, first, we propose a chemical embedding compression scheme to reduce the dimensionality of the feature space required for multi-component systems, based on the framework of Willatt et al [Phys. Chem. Chem. Phys., 2018], and implemented in PyTorch. Second, we generate a dataset of several thousands configurations, assembled from 25 d-block elements, which aims to represent cross-elemental interactions, evaluating their energies and forces at the DFT level. We demonstrate the effectiveness of the alchemical ML model in learning the energetics of this extremely diverse dataset, and provide showcase calculations of the properties of some realistic HEA compositions.

O 83.2 Fri 10:45 S054

Stacking the odds: Distribution-biased generative deep learning for molecular design —
•Joe Gilkes[1,2], Julia Westermayr[1], Rhyan Barrett[3], and Reinhard J. Maurer[1] — [1]Department of Chemistry, University of Warwick, UK — [2]HetSys CDT, University of Warwick, UK — [3]Warwick Mathematics Institute, UK

Organic electronics applications pose a number of often competing requirements on molecular design that are hard to satisfy by conventional synthesis. Devices such as organic light-emitting diodes (OLEDs) must exhibit closely aligned optoelectronic properties, yet their component molecules must be easily synthesisable and stable. The odds of finding suitable molecules when drawing random samples from chemical space are still too low for targeted design of candidate systems for OLED devices. We develop an automated molecular design approach based on iterative biasing of a generative deep learning model. In successive iterations, the output of this model is filtered with a deep learning surrogate model of electronic structure and then used to retrain the generative model with a bias. This enables us to create models that are progressively biased towards, e.g., higher ionisation potentials, or smaller fundamental gaps. We also demonstrate how we can bias towards multiple properties simultaneously by filtering our results with the SCScore model for synthetic complexity. This creates more synthetically viable molecules while still meeting optoelectronic requirements. Our approach efficiently creates novel molecules with tuned optoelectronic properties. Clustering analysis reveals trends in bonding patterns which can be utilised in molecular design.

O 83.3 Fri 11:00 S054

Machine learning TCP phases with domain knowledge of the interatomic bond —
•Mariano Forti, Alesya Burakovskaya, Ralf Drautz, and Thomas Hammerschmidt — ICAMS, Ruhr-Universität Bochum, Universitätsstr. 150, 44801 Bochum, Germany.

The understanding of the precipitation of topological close packed (TCP) phases in single-crystal superalloys is of central importance for the design of these materials for high-temperature applications. However, the structural complexity of these intermetallic compounds and the chemical complexity of the superalloys with typically N=5-10 elements hampers the exhaustive sampling of chemical space by density-functional theory (DFT) calculations. For example, the computation of the convex hull of the R phase with 11 inequivalent lattice sites would require N^{11} DFT calculations in an N-component system. We overcome this computational limitation by combining machine learning (ML) techniques with descriptors of the local atomic environment of the TCP phases. We present descriptors that are derived from bond order potential (BOP) theory which retain domain knowledge of the interatomic interaction from tight-binding Hamiltonians. We demonstrate that these descriptors enable us to predict the structural stability of TCP phases with simple regression algorithms. We apply this methodology to several systems with experimental evidence of R phase formation.

O 83.4 Fri 11:15 S054

Ab initio random structure search of organic molecules at substrates —
•Dmitrii Maksimov[1,2] and Mariana Rossi[1,2] — [1]Fritz Haber Institute of the Max Planck Society, Berlin, Germany — [2]Max Planck Institute for the Structure and Dynamics of Matter, Hamburg, Germany

Finding stable structures of molecular adsorbates (in isolation or forming layers) from calculations is challenging, exacerbated when the adsorbates are flexible. To make matters more complicated, in these situations, it is often difficult to find good and cheap potentials of such complex interfaces that allow a thorough and reliable global search of the structural space. To make this problem tractable with *ab initio* potentials, we present a random global geometry optimization package that can explicitly take into account the internal degrees of freedom of molecules, their position and orientation with respect to fixed surroundings, as well as periodic boundary conditions [1]. Electronic structure calculations and local geometry optimizations are performed through a connection to the ASE software [2], making it possible to interface this algorithm with various codes. To increase the efficiency of geometry optimizations, we introduce a framework to construct initial approximate Hessians for BFGS algorithms that are specially tailored to accelerate the relaxation of van der Waals bonded structures and handle large structural changes. We showcase the algorithm for the adsorption of di-L-alanine at Cu(110). [1] https://github.com/sabia-group/gensec [2] Larsen et. al., J. Phys.: Condens. Matter **29**, 273002 (2017).

O 83.5 Fri 11:30 S054

Active learning and element-embedding approach in neural networks for infinite-layer versus perovskite oxides —
Armin Sahinovic and •Benjamin Geisler — Fakultät für Physik, Universität Duisburg-Essen

The observation of superconductivity in $NdNiO_2$ films on $SrTiO_3(001)$ by Li *et al.* [1] has sparked considerable interest in the materials class of infinite-layer oxides. Here we combine first-principles simulations and active learning of neural networks to explore formation energies of oxygen vacancy layers, lattice parameters, and their statistical correlations in infinite-layer versus perovskite oxides across the periodic table, and place the superconducting nickelate and cuprate families in a comprehensive context. Neural networks accurately predict these observables, which act as a fingerprint of the complex reduction reaction, using only a fraction of the data for training. Unbiased by external knowledge, element embedding autonomously identifies chemical similarities between the individual elements in line with human knowledge. Active learning renders the training highly efficient, based on the physical concepts of entropy and information, and provides systematic accuracy control [2]. We recently applied this concept also to nitrides and fluorides [3]. This exemplifies how AI may assist on the quantum scale in discovering novel materials with optimized properties.

[1] D. Li *et al.*, Nature **572**, 624 (2019)
[2] A. Sahinovic and B. Geisler, PR Research **3**, L042022 (2021)
[3] A. Sahinovic and B. Geisler, J. Phys.: Condens. Matter **34**, 214003 (2022)

O 83.6 Fri 11:45 S054

Indirect learning interatomic potential models for accelerated materials simulations —
•Joe D. Morrow and Volker L. Deringer — Department of Chemistry, Inorganic Chemistry Laboratory, University of Oxford, Oxford OX1 3QR, United Kingdom

Machine learning (ML) based interatomic potentials are emerging tools for materials simulations but require a trade-off between accuracy and speed. We show how one ML potential can be used to train another: we use an existing, accurate, but more computationally expensive model to generate reference data (labels and locations) for a series of much

faster "indirectly-learned" potentials. Extensive reference datasets can be easily generated without the need for quantum-mechanical reference computations at the indirect learning stage, and we find that the additional data significantly improve the predictions of fast potentials with less flexible functional forms.

We apply the technique to disordered silicon, including a simulation of vitrification and polycrystalline grain formation under pressure with a system size of a million atoms. When comparing indirectly learned potentials to models learned directly from a DFT-labelled database, the latter make unphysical predictions for large systems (10^5 atoms) that are not apparent in smaller simulations ($\leq 10^4$ atoms). This emphasises the importance of carefully validating ML potentials *chemically*, not only via numerical error measures. Our work provides conceptual insight into the machine learning of interatomic potential models, and it suggests a route toward accelerated simulations of nanostructured materials.

O 83.7   Fri 12:00   S054
Predicting hot electrons free energies from ground-state data — •Chiheb Ben Mahmoud, Federico Grasselli, and Michele Ceriotti — EPFL, Lausanne, Switzerland

Machine-learning potentials, while extremely successful in describing the stability of condensed phases, are usually trained on ground-state electronic-structure calculations depending exclusively on the atomic positions and ignoring the electronic temperature. Hence, they are limited in their ability to describe hot electrons. We introduce a rigorous framework to calculate the finite-temperature electron free energy based exclusively on ground-state total energy and electronic density of states, while allowing to sample on-the-fly the electronic free energy at any temperature [1]. Our physically-motivated approach facilitates modeling material properties in extreme conditions with a fraction of the usual cost. We demonstrate it by computing the equation of state and heat capacity of hydrogen in planetary conditions. This approach demonstrates the impact of a universal model describing structural and electronic properties inexpensively and its ability to enable more accurate and predictive materials modeling and design.

[1]: C Ben Mahmoud, F Grasselli, M Ceriotti*- arXiv preprint arXiv:2205.05591, 2022

O 83.8   Fri 12:15   S054
Machine Learning the RPA density-density response function — •Mario Zauchner, Johannes Lischner, and Andrew Horsfield — Imperial College London, London, United Kingdom

Clusters and nanoparticles are used in a variety of scientific and industrial applications, including optoelectronics, photocatalysis, single electron transistors and medical imaging, among others. Electronic excitations often play a key role in these applications, but theoretical techniques for calculating excited-state properties of materials, such as the first-principles GW/Bethe-Salpeter method, are typically limited to very small systems. A key bottleneck of such excited-state calculations of clusters and nanoparticles is the determination of the static density-density response function, which is often calculated using a sum-over-states technique. In this talk, we present a technique to decompose the density-density response function into atomic contributions. This can be achieved by exploiting the locality of the density-density response function in non-metallic systems. These atomic contributions can then be used to train a machine-learning model us-

ing a set of structural features with the same rotational symmetry as the atomic response functions, thus allowing direct prediction of the density-density response function using only structural information.

O 83.9   Fri 12:30   S054
MD-based Raman Spectra using Machine Learning — •Manuel Grumet[1], Karin S. Thalmann[1], Tomáš Bučko[2,3], and David A. Egger[1] — [1]Department of Physics, Technical University of Munich, Garching, Germany — [2]Comenius University in Bratislava, Slovakia — [3]Slovak Academy of Sciences, Slovakia

Theoretical calculations of Raman spectra based on molecular dynamics (MD) trajectories allow to directly incorporate both anharmonic and temperature-dependent effects and thus yield more realistic spectra compared to a phonon-based approach [1]. The spectra can be calculated from the Fourier-transformed velocity correlation function of the polarizability tensor α. However, this requires evaluating α for a large number of MD configurations along each trajectory, which has high computational cost if done by ab-initio methods.

We therefore use kernel-based machine learning (ML) methods with density-based descriptors [2, 3] to predict α based on atomic positions. Ab-initio calculations are then only needed for obtaining a training data set, reducing the computational cost significantly. We use a number of test systems, including both solids and small molecules, to test and optimize several different variants of this approach and compare the achieved prediction performances. We also test transferability of the trained models to trajectories at different temperatures.

[1] M. Thomas et al., Phys. Chem. Chem. Phys. **15**, 6608 (2013)
[2] A. P. Bartók et al., Phys. Rev. B **87**, 184115 (2013)
[3] A. Grisafi et al., Phys. Rev. Lett. **120**, 036002 (2018)

O 83.10   Fri 12:45   S054
Thermal Transport via Green-Kubo Method and Message-Passing Neural-Network Potentials — Marcel F. Langer[1,2], Florian Knoop[2,3], Christian Carbogno[2], Matthias Scheffler[2], and •Matthias Rupp[2,4] — [1]TU Berlin, Germany — [2]The NOMAD Laboratory, FHI-MPG & HU Berlin, Germany — [3]Theoretical Physics Division, Linköping U, Sweden — [4]Konstanz U, Germany

Accurate, precise, and efficient computational access to thermal conductivities of materials is relevant for scientific understanding and industrial applications. The Green-Kubo method with first-principles calculations enables the determination of thermal conductivities, even for strongly anharmonic materials [1]. However, the high computational cost of long dynamics simulations of large supercells required for convergence limits applicability for large-scale, high-throughput materials discovery. Machine-learning potentials can reduce this cost [2].

Message passing neural networks (MPNNs) are a promising, but for this task yet untested, class of models due to their relational inductive bias, implicit long-range nature, and ability to incorporate directional information. We adapt the heat flux definition for MPNNs, investigate the impact of equivariance, present a systematic account of their convergence behavior and performance, and compare them to a simpler baseline model.

[1]: C. Carbogno, R. Ramprasad, and M. Scheffler, Phys. Rev. Lett. 118 175901 (2017) [2]: P. Korotaev et al., Phys. Rev. B 100 144308 (2019); C. Mangold et al., J. Appl. Phys. 127, 244901 (2020); C. Verdi et al., NPJ Computer. Mat. 7 156 (2021)