MM 37: Topical session (Symposium MM): Big Data Analytics in Materials Science

Sessions: Big Data Analytics in Materials Science III and IV

Time: Thursday 15:00–18:45 Location: H43

Topical Talk MM 37.1 Thu 15:00 H43

Atomistic Machine Learning between Physics and Data — •Michele Ceriotti — EPFL, Lausanne, Switzerland

Statistical regression techniques have become very fashionable as a tool to predict the properties of systems at the atomic scale, sidestepping much of the computational cost of first-principles simulations and making it possible to perform simulations that require thorough statistical sampling without compromising on the accuracy of the electronic structure model. In this talk I will argue how data-driven modelling can be rooted in a mathematically rigorous and physically-motivated framework, and how this is beneficial to the accuracy and the transferability of the model. I will also highlight how machine learning - despite amounting essentially to data interpolation - can provide important physical insights on the behavior of complex systems, on the synthesizability and on the structure-property relations of materials. I will give examples concerning all sorts of atomistic systems, from semiconductors to molecular crystals, and properties as diverse as drug-protein interactions, dielectric response of aqueous systems and NMR chemical shielding in the solid state.

MM 37.2 Thu 15:30 H43

Atomic cluster expansion for accurate and transferable interatomic potentials — •Ralf Drautz — ICAMS, Ruhr-Universität Bochum, Germany

The atomic cluster expansion provides a complete descriptor of the local atomic environment. The effort for evaluating the atomic cluster expansion scales linearly with the number of neighbors, irrespective of the order of the expansion. This makes the atomic cluster expansion an attractive basis for the development of accurate and transferable interatomic potentials.

I will discuss the application of the atomic cluster expansion to small Cu clusters and demonstrate smooth convergence to meV accuracy. I will further show how an interatomic potential that is transferable from free atoms to bulk may be obtained by combining the atomic cluster expansion with classical interatomic potentials.

MM 37.3 Thu 15:45 H43

Simultaneous and Reinforced Learning of Materials Properties from Incomplete Databases with Multi-Task SISSO — •Emre Ahmetcik, Runhai Ouyang, Christian Carbogno, Matthias Scheffler, and Luca M. Ghiringhelli — Fritz-Haber-Institut der Max-Planck-Gesellschaft, 14195 Berlin-Dahlem, Germany

Identifying descriptors that capture the underlying mechanisms for different materials properties is a key challenge in data-driven materials science. Recently, the sure-independence screening and sparsifying operator (SISSO) [1] has been introduced and was successfully applied to a number of key materials-science problems [1-3]. SISSO is a compressed-sensing based methodology yielding predictive and insightful models identified from an enormous space (billions or more) of candidate analytical expressions. In this work, we have extended the methodology to a 'multi-task learning' approach, a powerful and nontrivial generalization which identifies a single descriptor capturing multiple target materials properties at the same time. This approach is specifically suited for a heterogeneous materials database with missing data, i.e., in which not all properties are reported for all materials. As showcases, we address the relative stability of octet-binary compounds for several crystal phases and the metal/insulator classification of binary materials distributed over many crystal-prototype classes.

[1] R. Ouyang *et. al.*, Phys. Rev. Mater. 2, 1-11 (2018).
[2] C. J. Bartel *et. al.*, Nat Commun. 9, 4168 (2018).
[3] C. J. Bartel *et. al.*, Sci. Adv., accepted (2018), arXiv:1801.07700.

MM 37.4 Thu 16:00 H43

Information-theoretic Feature Selection and its Applications in Materials Science — •Benjamin Regler, Matthias Scheffler, and Luca M. Ghiringhelli — Fritz Haber Institute of the Max Planck Society, Berlin, Germany

Feature selection is a technique for proposing subsets of relevant properties (features), along with a measure which scores the different subsets.

In this talk, we give an overview of feature selection methods applied to materials science problems and discuss how to identify relationships between fundamental properties at the atomistic scale and materials properties at the macroscopic scale. In particular, we focus on the complexity of machine-learning models and highlight the advantages of using a systematic feature selection prior to making predictions (i.e., building machine-learning models).

Moreover, we propose a parameter-free, deterministic information-theoretic feature-selection framework for identifying approximate functional relationships between properties of interest. Importantly, the framework detects redundant and irrelevant features by performing nonlinear correlation analysis.

As showcase, we apply our approach to crystal-structure and other properties prediction in a restricted class of materials such as functionalized or octet binaries.

We conclude that our approach reduces the complexity of machine-learning models, extracts the most informative set of features, and supplements the analyses and identification of relevant properties.

MM 37.5 Thu 16:15 H43

The speech of strangers in material science — •Juliana Schell[1,2], Peter Schaaf[3], Hans-Christian Hofsäss[4], and Doru C. Lupascu[2] — [1]European Organization for Nuclear Research (CERN), CH-1211 Geneva, Switzerland — [2]University of Duisburg-Essen, 45141 Essen, Germany — [3]TU Ilmenau, Gustav-Kirchhoff-Straße 5, 98693 Ilmenau, Germany — [4]Georg-August-Universität Göttingen, Friedrich-Hund-Platz 1, 37077 Göttingen, Germany

Differently from stable nuclei, radioactive isotopes like to talk about their local environment. Human vocal communication produces sound, radioactive isotopes emit particles and photons and can act as a probe once decaying embedded in a host material. Some of them emit two consecutive gamma rays and it is convenient to find the relative probability of detecting the second photon and obtain its angular correlation in respect to the first. Humans perform many different speech acts that vary according to their intention, while the precession of the angular correlation in time can vary with the magnetic dipole and the electric quadrupole interactions. During the past four decades, the solid-state physics programme at ISOLDE-CERN has interpreted the speech of these strangers with innovative material science research and optimized equipment. This presentation overviews the newest perturbed angular correlation results and measurement possibilities at ISOLDE-CERN [1].

[1] Juliana Schell, Peter Schaaf, and Doru C. Lupascu, Perturbed Angular Correlations at ISOLDE: a 40 years young technique, AIP Advances 7, 105017 (2017).

15 min. break

Topical Talk MM 37.6 Thu 16:45 H43

High-throughput with Particle Technology — •Lutz Mädler — University of Bremen, Faculty of Production Engineering, Badgasteiner Str. 1, 28359 Bremen, Germany — Leibniz Institute for Materials Engineering IWT, Badgasteiner Str. 3, 28359 Bremen, Germany

High-throughput screening is a well-established method for scientific experimentation in chemistry and biology. Examples are heterogeneous catalysts, drug developments and nanoparticle toxicology. These methods involve the synthesis of small sample volumes often in form of particles that are quickly tested. These tests are designed to quickly obtain easily accessible data (called descriptors) that are related with a predictor function to the desired properties. The descriptor-predictor-relation is found through mathematical modelling and calibration. One particle based high-throughput concept for the evaluation of potential toxicological hazards will be presented in more detail. Furthermore, a new concept is presented which transfers high-throughput screening to the exploration of new structural metals. The method comprises the synthesis of many small alloy samples in form of particles. These samples obtain a defined microstructure by fast or parallel thermal and mechanical treatments and are subsequently subjected to novel fast descriptor tests while a mathematical algorithm develops the predic-

tor function. The method presented here is a collaborative approach among many researchers and also involves sample routing and automation considerations, process modelling as well as big data methods.

Topical Talk MM 37.7 Thu 17:15 H43

Microstructure is the know-it-all - classification approaches with data mining and deep learning methods — •Frank Mücklich[1,2] and Dominik Britz[2,1] — [1]Universität des Saarlandes, Saarbrücken, Germany — [2]Material Engineering Center Saarland, Saarbrücken, Germany

The microstructure can be regarded as the *multi-scale archive* from which we can *read* the quantitative information about the microstructure formation processes and the prediction of the final material properties on each relevant scale. Recent advances in 3D tomography methods on the micro, nano and atomic scale allow to study the differences of microstructures with higher morphological and topological complexity. Classification strategies using Support Vector Machine and Deep Learning will be discussed using morphological and substructure parameters. Images are processed directly in the workflow after an adapted contrasting. The result might be simultaneously segmented and classified.

S. M. Azimi, D. Britz, M. Engstler, M. Fritz, and F. Mücklich, Advanced Steel Microstructural Classification by Deep Learning Methods, Scientific Reports (Nature) 8 (2018) 2128.

MM 37.8 Thu 17:45 H43

Electronic density-of-states fingerprints for finding similar materials — •Martin Kuban, Santiago Rigamonti, Markus Scheidgen, and Claudia Draxl — Humboldt-Universität zu Berlin

The recent development of large databases for computational materials science, like NOMAD [1], allows researchers to reuse data that was generated for different purposes. In this work, we make use of the data contained in NOMAD to find materials with similar properties. Similarity can be evaluated and quantified by comparing specialized representations of the materials properties, so-called *fingerprints*. We design a family of fingerprints derived from the electronic density-of-states (DOS), consisting of vectorial representations obtained from non-uniform scalings of the DOS. In contrast to previous works [2], our approach allows us to set the focus of searches for similar materials on special features of the DOS, as for instance the band gap, or the amount of states close to the Fermi level. We present examples for several materials ranging from metals to insulators. To demonstrate the usefulness and applicability of our approach, we have devised a *recommender system* for the NOMAD Encyclopedia.

[1] C. Draxl and M. Scheffler, *MRS Bulletin*, **43**, 676, (2018).
[2] O. Isayev *et al.*, *Chermistry of Materials* **27**, 735, (2015).

MM 37.9 Thu 18:00 H43

Accurate Thermal Conductivities of Complex, Strongly-Anharmonic Solids — •Florian Knoop, Thomas Purcell, Marcel Hülsberg, Matthias Scheffler, and Christian Carbogno — Fritz-Haber-Institut der Max-Planck-Gesellschaft, Berlin, Germany

Vibrations in modern functional materials, e.g., perovskites [1], are often governed by strongly anharmonic effects. Traditional perturbative *ab initio* approaches do not capture these effects and are thus not suited to describe the associated material properties, e.g., the lattice heat conductivity $\kappa(T)$. For these materials, $\kappa(T)$ can be accurately computed via the *ab initio* Green-Kubo (aiGK) method [2], in which anharmonic effects are fully accounted for via *ab initio* molecular dynamics (aiMD) simulations.

In our hierarchical high-throughput approach, we first compute the harmonic properties of thousands of compounds spanning over several classes of materials, including simple elemental semiconductors, but also complex technologically relevant materials such as ceramics and perovskites. From these materials, aiMD-inspired metrics are used to single out strongly anharmonic compounds, for which $\kappa(T)$ is then computed with the aiGK method. We discuss the challenges involved in this automatized approach, the reliability of our metrics, and the trends observed in materials space.

[1] A. van Roekeghem, *et al.*, *Phys. Rev. X.*, **6**, 041061 (2016)
[2] C. Carbogno, R. Ramprasad, and M. Scheffler, *Phys. Rev. Lett.*, **118**, 175901 (2017).

MM 37.10 Thu 18:15 H43

Surface Structure Search using Coarse Grained Modeling and Bayesian Linear Regression — •Lukas Hörmann, Andreas Jeindl, Alexander T. Egger, and Oliver T. Hofmann — Institute of Solid State Physics, Graz University of Technology, Austria

The key information about a monolayer of molecules on a substrate, aside from chemical composition, is arguably the polymorph it forms. First-principles prediction of such polymorphs is a major challenge, due to the large number of possible arrangements of molecules. To meet this challenge, we develop SAMPLE[1,2], which uses physically motivated coarse graining and statistical learning to explore the potential energy surface of commensurate organic monolayers on inorganic substrates.

We first determine adsorption geometries of isolated molecules on the substrate. By generating commensurate arrangements of these geometries, we compile a large number of possible polymorphs. Using experimental design theory, we select subsets of these polymorphs and calculate their adsorption energies using dispersion-corrected density functional theory. These subsets serve as training data for an energy model, based on molecular interactions. Using Bayesian linear regression, we determine the model parameters, yielding meaningful physical insight and allowing the prediction of adsorption energies for millions of possible polymorphs with high accuracy.

We demonstrate this on three complimentary systems: naphthalene on Cu(111), TCNE on Cu(111), and benzoquinone on Ag(111).

[1] Hörmann et al., arXiv:1811.11702 (2018)
[2] Scherbela et al., Phys. Rev. Materials 2, 043803 (2018)

MM 37.11 Thu 18:30 H43

Predicting Reaction Energetics with Machine Learning — •Sina Stocker, Johannes T. Margraf, and Karsten Reuter — Technische Universität München, Germany

Predictive-quality first-principles based microkinetic models are increasingly used to analyze (and subsequently optimize) reaction mechanisms in heterogeneous catalysis. In full rigor such models require the knowledge of all possible elementary reaction steps and their corresponding reaction barriers. Unfortunately, for complex catalytic processes, such as the generation of synthetic fuels out of syngas, the number of such steps becomes so large that an exhaustive first-principles calculation of all barriers becomes prohibitively expensive.

As a remedy, we explore the possibility of machine learning (ML) approaches to the prediction of the reaction energetics. An essential component in such data-driven approaches are efficient molecular representations (descriptors). We test a range of such representations that have been suggested to describe properties of closed-shell molecules and specifically assess their capabilities in describing open-shell systems and consequently reaction energetics. The obtained overall promising performance confirms the potential of ML approaches for a high throughput screening of elementary steps in large reaction networks.