SYNM 1: From Physics and Big Data to the Design of Novel Materials

Time: Monday 15:00–17:45 Location: H1

Invited Talk SYNM 1.1 Mon 15:00 H1

How to tackle the "I" in FAIR? — •Claudia Draxl — Physics Department, Humboldt-Universität Berlin, Germany

Veracity, one of the 4V challenges of Big Data, is an issue for the FAIRness of materials-science results. This concerns in particular, the interoperability. On the computational side, the precision of different DFT codes was thoroughly investigated [1]. Later, it was demonstrated how ultimate precision for molecules and solids can been reached [2] and how different methodology impacts the results [3]. Also code-specific uncertainties coming from numerical settings were assessed [4]. More recently, we demonstrated [5] how a spectral fingerprint, combined with a similarity metric, can be used for establishing quantitative relationships between materials data. Besides the identification of electronically similar materials, this approach can be used for assessing uncertainty in data that potentially come from different sources. Various examples highlight how to quantify differences between measured optical spectra or the impact of methodology and computational parameters on calculated spectral properties. Moreover, combining our fingerprint with a clustering approach allows us to explore materials spaces in view of finding (un)expected trends or patterns [6].

[1] K. Lejaeghere et al., Science 351, aad3000 (2016). [2] A. Gulans et al., PRB 97, 161105(R) (2018). [3] A. Gulans and C. Draxl, https://arxiv.org/abs/2204.02751 [4] C. Carbogno, et al., npj Comput. Mater. 8, 69 (2022). [5] M. Kuban,et al., MRS Bulletin Impact (2022); https://arxiv.org/abs/2204.04056 [6] M. Kuban et al., https://arxiv.org/abs/2201.02187

Invited Talk SYNM 1.2 Mon 15:30 H1

Beyond the average error: machine learning for the discovery of novel materials — •Mario Boley[1], Simon Teshuva[1], Felix Luong[1], Lucas Foppa[2], and Matthias Scheffler[2] — [1]Monash University — [2]Fritz Haber Institute of the Max Planck Society

Machine learning models promise to radically accelerate the discovery of novel functional materials by rapidly screening huge candidate spaces for materials with rare combinations of properties. While current models allow for accurate property prediction on average, e.g., indicated by the root mean squared error, such measures are only loosely connected to materials discovery. They do not capture that discovery is a process, where models are repeatedly retrained with new data. Moreover, they depend on some fixed sampling distribution, which is irrelevant when actively exploring candidates and distorted by the overwhelming mass of mediocre materials. Finally, they do not reflect model uncertainty, which is a crucial parameter for effective active learning strategies. Here, based on the example of searching for stable and stress resistant double perovskites under a bandgap constraint, we evaluate models in terms of their optimisation regret—a function of consecutive improvements of the target property by newly discovered materials. In particular, we employ non-parametric regression models (Gaussian processes and random forests) within the "expected improvement" search strategy. Starting from an initial dataset of 815 computed double perovskite properties, we are able to discover materials with improved target values already within the first 20 newly acquired data points.

Invited Talk SYNM 1.3 Mon 16:00 H1

The Phase Diagram of All Inorganic Materials — •Chris Wolverton — Northwestern University, Evanston, IL USA

One of the holy grails of materials science, unlocking structure-property relationships, has largely been pursued via bottom-up investigations of how the arrangement of atoms and interatomic bonding in a material determine its macroscopic behavior. Here we consider a complementary approach, a top-down study of the organizational structure of networks of materials, based on the interaction between materials themselves. We demonstrate the utility of applying network theory to materials science in two applications: First, we unravel the complete *phase stability network of all inorganic materials* as a densely-connected complex network of 21,000 thermodynamically stable compounds (nodes) interlinked by 41 million tie-lines (edges) defining their two-phase equilibria, as computed by high-throughput density functional theory. Using the connectivity of nodes in this phase stability network, we derive a rational, data-driven metric for material reactivity, the *nobility index*, and quantitatively identify the noblest materials in nature. Second, we apply network theory to the problem of synthesizability of inorganic materials, a grand challenge for accelerating their discovery using computations. We use machine-learning of our network to predict the likelihood that hypothetical, computer generated materials will be amenable to successful experimental synthesis.

15 min. break

Invited Talk SYNM 1.4 Mon 16:45 H1

Automated data-driven upscaling of transport properties in materials — •Danny Perez[1] and Thomas Swinburne[2] — [1]Los Alamos National Laboratory, Los Alamos, USA — [2]CRNS/CINaM, Marseille, France

Transport properties of complex defects are crucial factors that control the performance of many material systems, e.g., the radiation tolerance of materials for nuclear fusion or fission applications. Characterizing the transport of complex defects is however notoriously tedious and time-consuming, especially as the defects grow, leading to a combinatorial explosion in the number of possible conformations and local transition pathways. I will present a large-scale data-driven approach to automatically obtain reduced-order models of defect evolution, transport coefficients, as well as effective continuum transport equations, from large number of short molecular dynamics (MD) simulations. The optimal MD simulations to carry out are identified on-the-fly using a Bayesian uncertainty quantification framework and automatically executed on a massively-parallel task-execution infrastructure. We show how this microscopic information can be systematically and efficiently upscaled into meso and macro-scale representations that can inform microstructure evolution models.

Invited Talk SYNM 1.5 Mon 17:15 H1

Data-driven understanding of concentrated electrolytes — •Alpha Lee — Department of Physics, University of Cambridge, Cambridge, UK

Electrolytes play an important role in a plethora of applications ranging from energy storage to biomaterials. Notwithstanding this, the structure and dynamics of concentrated electrolytes remain enigmatic. In this talk, I will show how machine learning can unravel hidden patterns in simulations of electrolytes, helping us understand the structure of concentrated electrolytes and mechanisms of ion transport. In the first part of my talk, I will illustrate how debates such as extent of "ion pairing" in concentrated electrolytes can be addressed using methods in unsupervised machine learning and Bayesian hypothesis testing. In the second part of my talk, I will discuss how machine learning help relate local ionic structure to molar ionic conductivity. This furnishes microscopic insights on what are the drivers of conductivity as a function of ion concentrations. More broadly, I will discuss the role that machine learning can play in not only delivering predictive models, but also serving as an "intuition pump" to understand complex soft matter systems.