When Do Machine Learning Models Learn Chemistry? From Electronic Structure to Latent Space Steering

Dr. Attila Cangi, Center for Advanced Systems Understanding (CASUS), Helmholtz-Zentrum Dresden-Rossendorf (HZDR)

Recent advances in machine learning are reshaping how we model and understand matter across scales—from electronic structure in materials to the chemical property spaces of molecules. In this talk, I present two complementary approaches that illustrate how data-driven models can both accelerate simulation and enable control over complex chemical systems.

First, I introduce the Materials Learning Algorithms (MALA) framework, a scalable machine learning approach for predicting electronic structure in large-scale atomistic simulations. By leveraging local descriptors of atomic environments, MALA efficiently approximates key electronic observables, including the local density of states, charge density, density of states, and total energy, enabling density functional theory-level insights at significantly reduced computational cost. Applications to diverse systems including boron clusters, liquid-solid aluminum interfaces, and extended defects in beryllium demonstrate its versatility and scalability.

In the second part, I shift focus from atomistic simulation to molecular generative modeling, examining how latent representations learned by transformer-based variational autoencoders can encode chemically meaningful structure. I address a central challenge in this domain: the risk that apparent structure-property relationships are driven by superficial sequence statistics rather than underlying chemistry. To disentangle these effects, I introduce a framework for identifying and validating latent directions associated with distinct molecular properties. By combining linear and nonlinear probing with decoded-molecule validation, we demonstrate robust and controllable steering of properties such as lipophilicity and polarity.

Short Curriculum Vitae

Attila Cangi is Head of the Department of Machine Learning for Materials Design at the Center for Advanced Systems Understanding (CASUS), Helmholtz-Zentrum Dresden-Rossendorf (HZDR). He leads a research team focused on applying artificial intelligence and machine learning to computational materials modeling. Their work centers on developing scalable methods that accelerate first-principles simulations of electronic and atomistic structures as well as advanced models for chemical property prediction, with the overarching goal of discovering and designing novel materials. Attila Cangi earned his Ph.D. in Chemistry from the University of California, Irvine. He previously held positions as a postdoctoral fellow at the Max Planck Institute of Microstructure Physics and as a staff scientist at Sandia National Laboratories.