Machine Learning for Atomistic Modeling and Sampling of Material Interfaces and Lattices

by

Xiaochen Du

Submitted to the Department of Chemical Engineering and
the Center for Computational Science and Engineering
on May 12, 2026 in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY IN CHEMICAL ENGINEERING AND COMPUTATION

ABSTRACT

The atomic-scale arrangement of material surfaces, interfaces, and bulk lattices governs catalytic activity, electrochemical stability, and electronic function. The combinatorial space of possible structures is too large for exhaustive first-principles enumeration, and expert-curated candidates routinely miss thermodynamically stable phases under finite temperature, varying chemical potential, or applied electrical potential. Machine learning offers two complementary tools: fast surrogate energy models and generative samplers that propose atomic configurations directly. This thesis develops both tools to discover stable surfaces and bulk lattices.

We first introduce Virtual Surface Site Relaxation–Monte Carlo (VSSR-MC), an algorithm that couples Monte Carlo sampling over candidate adsorption sites with a neural-network potential. Applied to semiconductor and perovskite oxide surfaces under vacuum, VSSR-MC recovers known reconstructions and reveals previously unreported low-energy terminations. We then extend the method to electrochemical interfaces by modeling surface, bulk, and electrolyte thermodynamics as functions of pH and applied potential. On $LaMnO_3(001)$, the resulting surface Pourbaix diagrams uncover catalytically relevant terminations that handpicked searches had missed.

Next, for bulk lattice thermodynamics, where the bottleneck is free-energy estimation rather than energy evaluation, we develop any-order autoregressive models (AO-ARMs). AO-ARMs support flexible partial-lattice conditioning, yield exact likelihoods, and provide an out-painting procedure that transfers small-lattice models to larger supercells without retraining. Transformer-based AO-ARMs capture the critical behavior of the two-dimensional Ising model and the ordered intermetallic phases of copper-gold alloys, with free-energy deviations below 5 meV per atom on the extrapolated supercells.

Finally, to bypass Monte Carlo sampling for surfaces altogether, we train a generative model that jointly proposes adsorbate positions and chemical identities. On an oxide-on-metal benchmark, our variable-composition model recovers low-energy surface configurations and scales cleanly with dataset size.

Together, these methods expand what can be discovered at material interfaces and in bulk lattices, providing a route toward predictive materials design under realistic operating conditions.