Accelerating the construction of machine-learning interatomic potentials using surrogate models

Nongnuch Artrith*

Debye Institute for Nanomaterials Science, Utrecht University, 3584 CS Utrecht, Netherlands

**E-mail: n.artrith@uu.nl*

Machine-learning interatomic potentials trained on total energies and forces from first principles calculations enable linear-scaling atomistic simulations with an accuracy that is close to the reference method at a fraction of the computational cost. However, the construction of artificial neural network (ANN) potentials for complex materials, such as battery materials, may require huge first-principles databases to sample the structural and composition space with sufficient precision. In previous work, we showed that training ANN potentials on total energy and atomic force information based on an extrapolation of the total energy via Taylor expansion greatly reduces the computational cost of ANN potential construction [1]. Here, we discuss an extension and refinement of this methodology that further reduces the amount of required reference data. We propose a surrogate model based on Gaussian process regression (GPR) trained on energies and interatomic forces to accelerate the generation of reference data for ANN potential training. Within active learning schemes, our GPR surrogate model further facilitates the refinement of reference datasets with Bayesian sampling. The accelerated construction and improved performance of the GPR-aided ANN potential are demonstrated by applying it to the modeling of ethylene carbonate (EC) on Li metal surfaces, which is relevant for the formation of solid-electrolyte interphases in lithium-metal batteries.

Reference:

1. A.M. Cooper, J. Kästner, A. Urban, and N. Artrith, *npj Comput. Mater.* **6**, 54, (2020).