Metadynamics Sampling in Atomic Environment Space for Collecting Training Data for Machine Learning Potentials

Dongsun Yoo[1,2], Jisu Jung[1], Wonseok Jeong[1] and Seungwu Han[1]

[1] *Department of Materials Science and Engineering and Research Institute of Advanced Materials, Seoul National University, Seoul 08826, Korea.*
[2] *Present address: Samsung Display Co., Ltd., 1, Samsung-ro, Giheung-gu, Yongin-si, Gyeonggi-do 17133, Korea.*

wjdwltn928@snu.ac.kr

Machine learning potentials (MLPs) are garnering attention in atomistic simulation by providing the accuracy of quantum mechanical calculations at much lower costs. Due to its flexible functional form, MLPs shift the heart of interatomic potential development from building mathematical formula to collecting a proper training set. As an interpolation algorithm, MLPs can guarantee their reliability only within the training set. Therefore, the training set should cover all local atomic environments that may appear in the target simulation ideally. Usually, the training set is constructed from the crystal structure or molecular dynamics (MD) simulations. However, these conventional approaches govern by Boltzmann statistics have difficulty in sampling various configurations, so that practitioners handpick the essential configuration with expertise iteratively to improve MLPs. Recently, some researchers try to sample the diverse configuration with methods such as random structure search and entropy-maximization[1, 2], but their applications are still limited.

In this presentation, we suggest a novel sampling method based on metadynamics. The thermal barrier is overcome through the metadynamics accumulating bias potential along with the collective variables (CVs). The metadynamics using the coordinate in local atomic environment space, encoded atom-centered symmetry function vector, as collective variable drives the system into the unvisited configuration semi-automatically. We apply proposed metadynamics sampling to H:Pt (111), GeTe, and Si systems. According to the examples, a small number of metadynamics trajectories can collect enough diverse configurations to develop the reliable MLPs.

[1] V. L. Deringer, C. J. Pickard, & G. Csányi, Phys. Rev. Lett., **120**, 156001 (2020).
[2] M. Karabin & D. Perez, J. Chem. Phys., **153**, 094110 (2020).

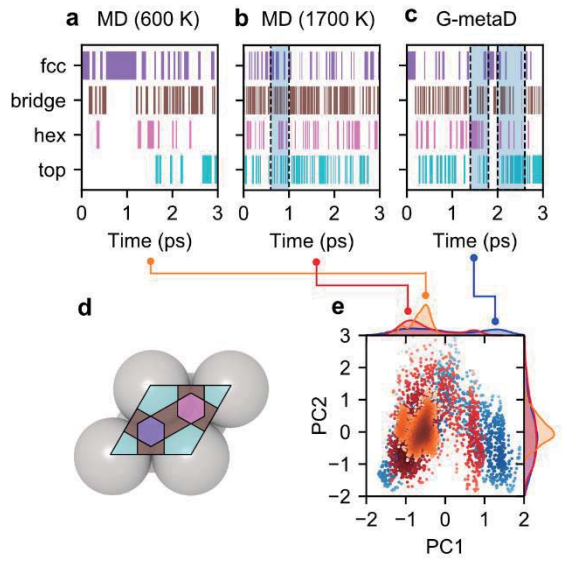

Fig.1: Sampled configurations during MD and Metadynamics. ***a-c*** *The classification of H sites along the time for MD at 600 K, 1700 K, and proposed metadynamics (G-metaD) at 600 K, respectively. The shaded area means that the H atom is on the sublayer. The color of sites is corresponded to* ***d****.* ***d*** *Characteristic area on Pt(111) surface.* ***e*** *Distribution of sampled points of H atom on principal-components axes.*

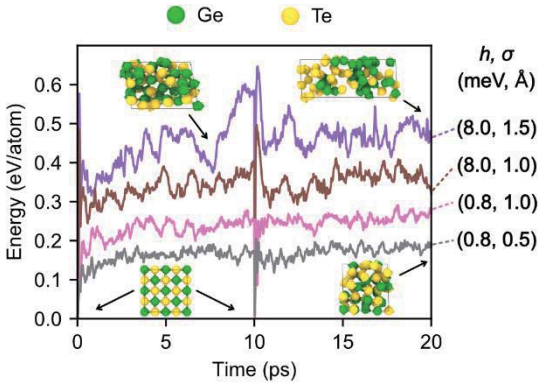

Fig.3:Time evolution of potential energy of four G-metaD along the time. The energy is referenced to that of the rock-salt structure. Structures during G-metaD are shown as inset figures.

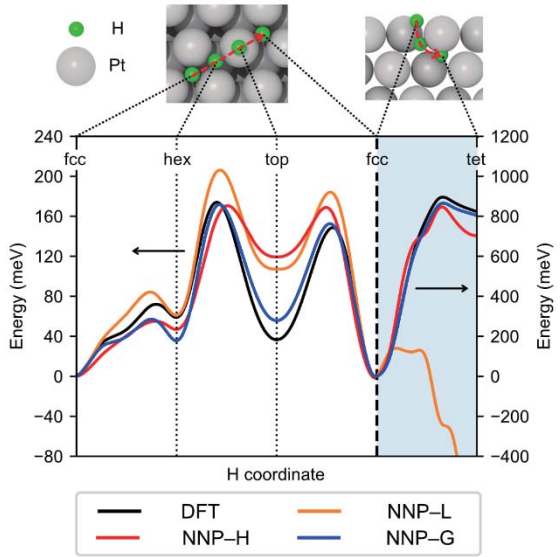

Fig.2: The minimum energy paths along fcc → hex → top → fcc → tet (subsurface tetrahedral site). NNP-L, NNP-H, and NNP-G stand for the NNP constructed from the MD 600K, MD 1700 K, and G-metaD, respectively.

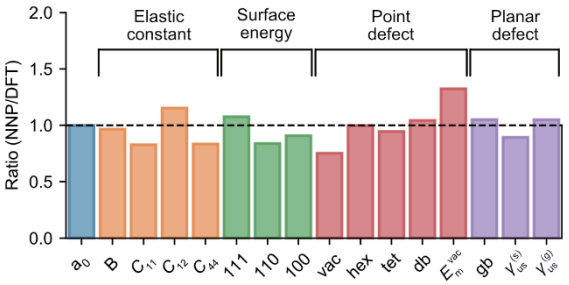

Fig.4: Ratios of NNP to DFT for static properties of Si in the diamond structure. Surface energies are calculated for (100)-(2×2),47 (110)-(1×1),48 and (111)-(3×3) reconstructions.49 Defect formation energies for the vacancy (vac) and interstitials (hexagonal (hex), tetrahedral (tetra), and dumbbell (db)). ΕEm vac is the migration energy of the vacancy. For extended defects, gb means the (112)Σ3 grain boundary. γγus (s) and γγus (g) are unstable stacking-fault energies on shuffle and glide planes of the diamond (111) plane.