Atomic energy mapping of neural network potential

Dongsun Yoo,* Kyuhyun Lee,* Wonseok Jeong,* Dongheon Lee, and Seungwu Han†

Department of Materials Science and Engineering

and Research Institute of Advanced Materials,

Seoul National University, Seoul 08826, Korea

Satoshi Watanabe

Department of Materials Engineering,

The University of Tokyo, Bunkyo, Tokyo 113-8656, Japan

(Dated: July 25, 2019)

* These three authors contributed equally.

† hansw@snu.ac.kr

I. SUPPLEMENTAL MATERIAL

A. Details on the training procedure

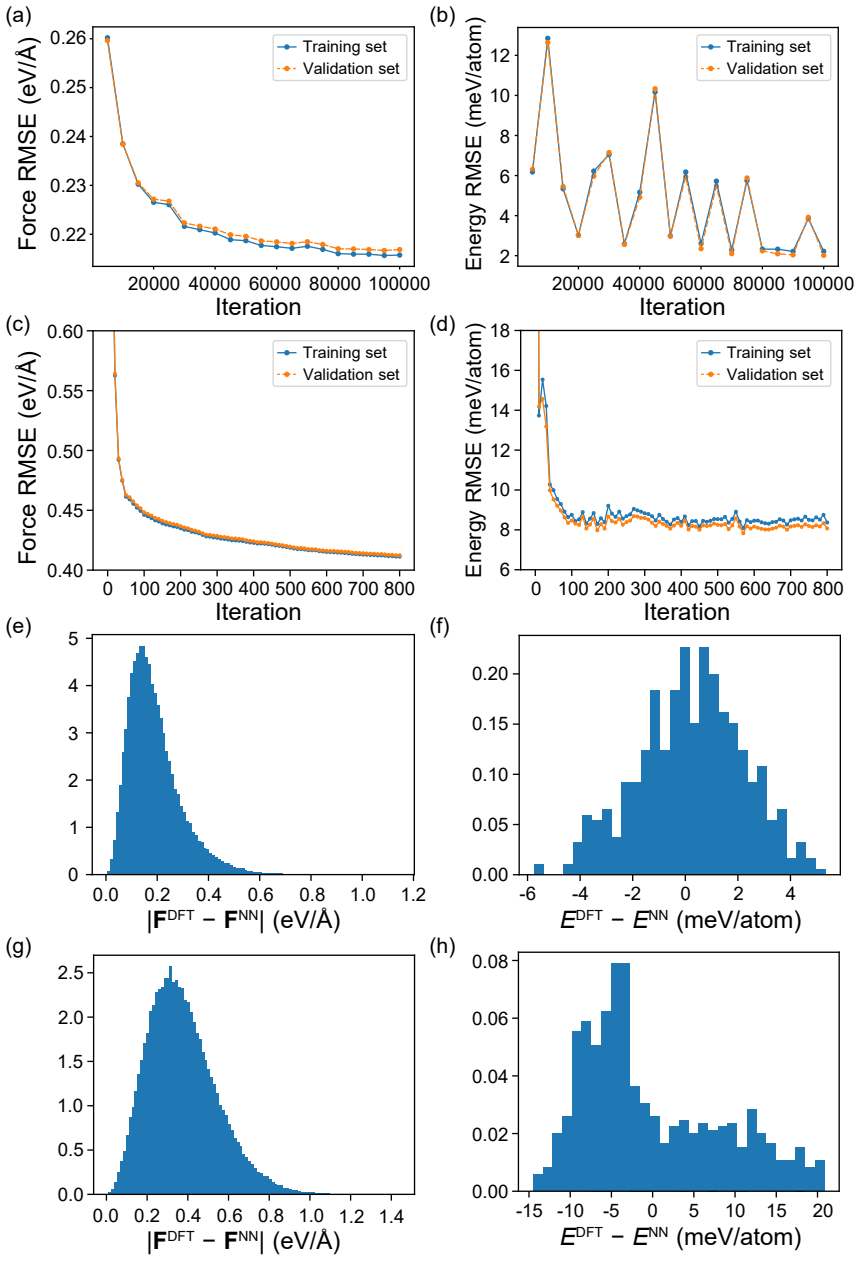

FIG. S1. Learning curves of force RMSE and energy RMSE for (a,b) Si slab model and (c,d) for Si nano-cluster model. The distribution of test set force error and total energy error for (e,f) Si slab model and (g,h) Si nano-cluster model.

NNPs in the Si crystal and slab examples are trained by Adam optimizer with batch

size of four. We applied exponential decay to learning rate (decays from 0.01 to 0.00017) since the energy RMSE fluctuates wildly at high learning rate. We apply L2 regularization with coefficient of 10^{-6}. The training procedure is terminated after 100,000 iterations. The settings for Si nano-cluster example are same except that full-batch L-BFGS optimizer is used and that the procedure is terminated after 800 epochs. The learning curve of Si slab and Si nanocluster examples are shown in Fig. S1. Energy and force RMSE are well converged and generalization error is small. Additionally, correlation plot between DFT and NNP is shown in Fig. S2.

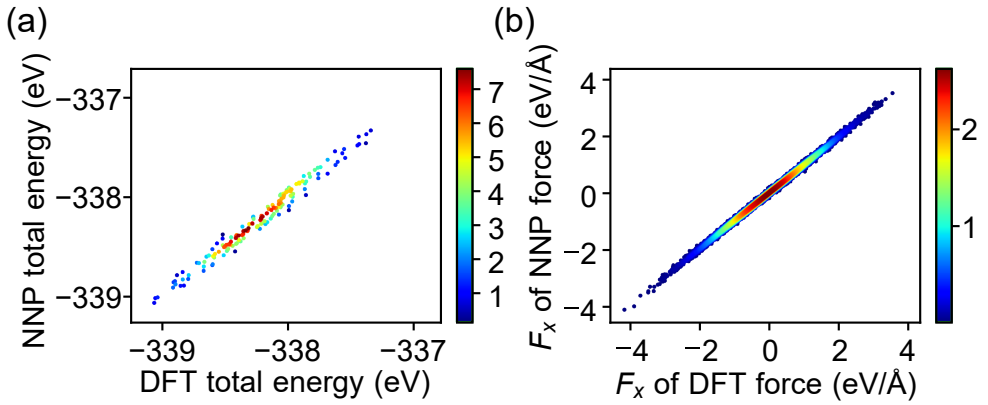

FIG. S2. The correlation plot between DFT and NNP for (a) total energy and (b) x-component of force (for validation set). Both energy and force show high correlation, which indicates that both energy and force are well fitted.

B. Convergence of RMSE against the size of training set

The size of the training set in the manuscript ranges from 350 to 832 structures. It is small compared to conventional training set used to develop NNP. However, as our target system is limited to small configuration space, and NNPs are trained with atomic forces as well (22,400-198,848 forces), the size is sufficient. Figure S3 shows the convergence of force RMSE for separate test set against the size of training set for Si slab model. At small training set size, overfitting is observed. However, it converges well with low generalization error at the size of 350 structures, which is the size used in the manuscript. Si crystal model is even simpler and nano-cluster model uses much larger training set as it contains diverse configurations.

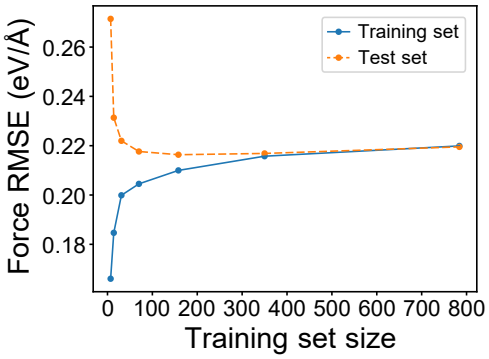

FIG. S3. Force RMSE vs. training set size (the number of structures in the training set).

C. Piecewise cubic spline example

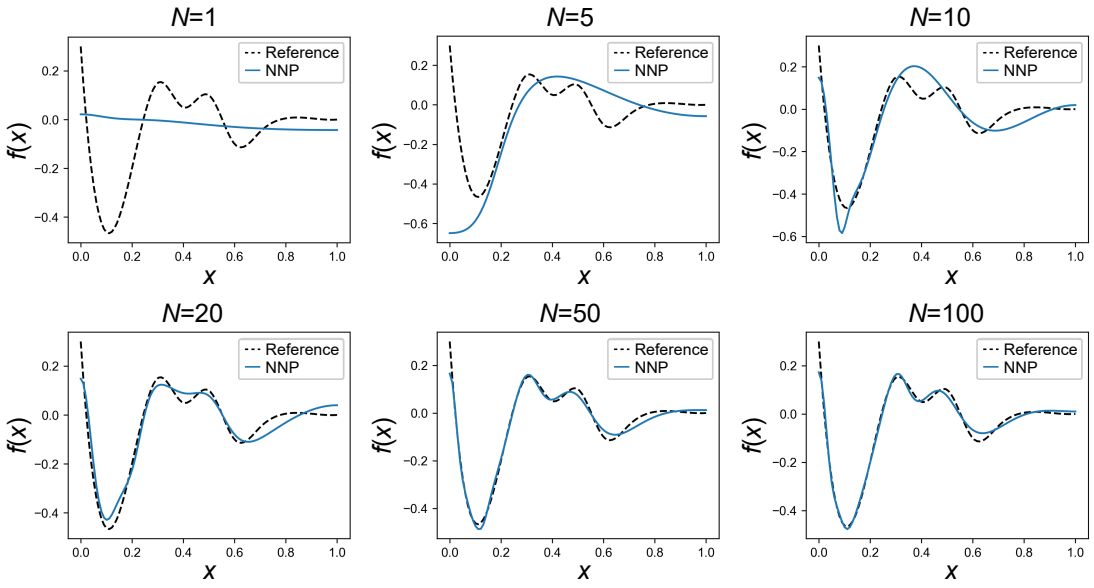

FIG. S4. $f(x)$ predicted with trained NN with varying size of training set (N).

We demonstrate with a simple mathematical model that the neural network (NN) is capable of inferring the underlying function when only sums of the function values are provided. We create a piecewise cubic spline $f(x)$ shown as a dashed line in Fig. S4. The size of the training set (N) varies from 1 to 100, and each sample is a sum of $f(x)$ at five random x's. A neural network of 8-30-30-1 structure is used, where x is encoded into 8 G_2 symmetry functions. Since only sum of $f(x)$ is trained, NN cannot predict $f(x)$ at small N, but at sufficiently large N, NN can infer the underlying $f(x)$ accurately.

D. Single-linkage clustering procedure

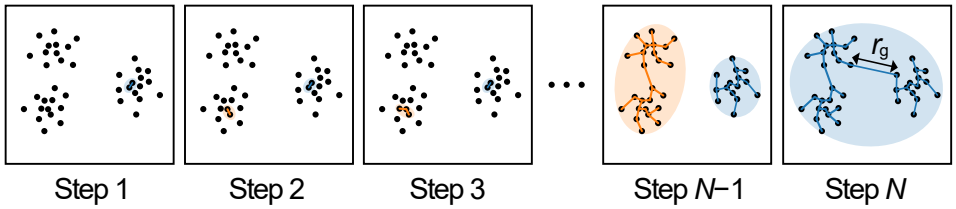

FIG. S5. Schematic illustration of single-linkage clustering procedure to obtain r_g.

The single-linkage clustering procedure described in the manuscript is illustrated in Fig. S5. At the start, every point is a cluster of size one. In the early steps, small clusters are merged and bigger clusters start to form. There remain a few large clusters and a number of tiny clusters in the latter steps. In the last step, two last big clusters are merged and r_g corresponds to the maximum distance between big clusters.