Toward reliable and transferable machine learning potentials: uniform training by overcoming sampling bias

Wonseok Jeong‡[1], *Kyuhyun Lee*‡[1], *Dongsun Yoo*[1], *Dongheon Lee*[1], *& Seungwu Han*[*,1].

[1]Department of Materials Science and Engineering and Research Institute of Advanced Materials, Seoul National University, Seoul 08826, Korea.

‡ These authors contributed equally.

AUTHOR INFORMATION

Corresponding Author

*Email: hansw@snu.ac.kr

Computational Procedures

We use the Vienna Ab initio Simulation Package (VASP) to produce density-functional-theory (DFT) reference data.[S1] The projector augmented wave (PAW) pseudopotentials[S2] with the generalized gradient approximation[S3] were employed for electronic structure calculations. The energy cutoffs for the plane-wave basis are 200 and 400 eV for Si and PdO, respectively. For Si, the 2×2×2 supercells are used with the Brillouin zone integration on the regular mesh of 3×3×3. For Pd(111)/O, 2×2 Pd(111) surface slab with 4 atomic layers is used to construct training set with the k-point sampling of 2×2×1.

The training of NNP and NNP-MD simulations are carried out by an in-house package named SNU Interatomic Machine-learning PotentiaL packagE - ver. NN; SIMPLE-NN. SIMPLE-NN incorporates google TensorFlow[S4] as a NN training engine and LAMMPS for MD simulations.[S5] Behler-Parinello symmetry functions are implemented in the package for input feature vector generation from the DFT results. NNPs are optimized with L-BFGS-B algorithm.[S6] The network structures are 26-10-10-1 and 70-20-20-1 for Si and Pd(111)/O, respectively. From the DFT reference data set, 10% are randomly selected and left out of the training for the use of validation set.

For efficiency, the NNPs are first optimized using only Γ_{energy}, and then trained with $\Gamma_{\text{energy}} + \Gamma_{\text{force}}$. The scaling coefficient of 0.003 (μ in Eq. (1) in the main text) is used. We trained the network until the stopping criterions has reached. The stopping criterions of NNP training is as follows:

$$\frac{\Gamma^k - \Gamma^{k+1}}{\max(|\Gamma^k|, |\Gamma^{k+1}|, 1)} \leq 10^{-11}$$

$$\max(|\, proj_i \nabla\Gamma|_{i=1,\ldots,n}) \leq 10^{-9}$$

where k, $proj_i\nabla$ and n are the index of the training iteration, ith component of projected gradient, and the index of the projected gradient component, respectively.

Fig S1 shows the decrease of the log(Γ) as a function of iterations.

The energy and force RMSEs are summarized in Table 1.

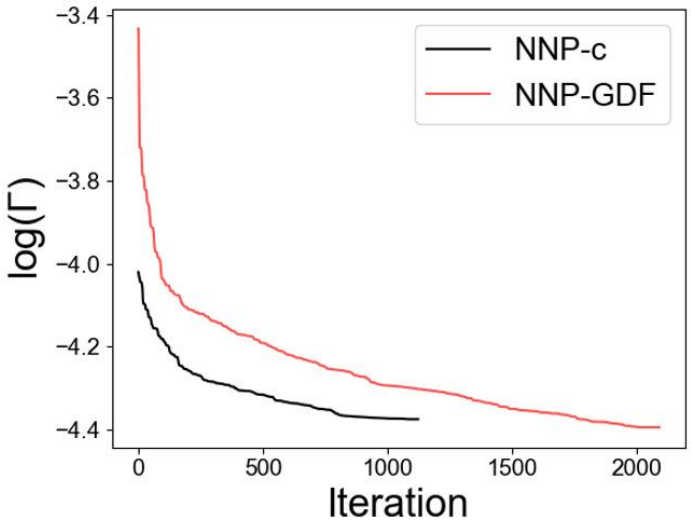

Figure S1. The loss function during the Si vacancy training iterations.

Training configurations		Energy RMSE (meV/atom)		Force RMSE (eV/Å)	
		Training set	Validation set	Training set	Validation set
Si vacancy	NNP-c	2.25	2.25	0.18	0.18
	NNP-GDF	2.53	2.76	0.17	0.19
Si interstitial	NNP-c	2.83	3.17	0.19	0.20
	NNP-GDF	2.91	2.55	0.24	0.20
Pd(111)/O	NNP-c	1.41	1.39	0.11	0.11
	NNP-GDF	1.53	1.42	0.13	0.14

Table S1. The energy and force RMSEs of NNPs. RMSEs tend to slightly increase for NNP-GDF. We note that the main objective of GDF weighting is leveling the learning level rather than reducing overall RMSE.

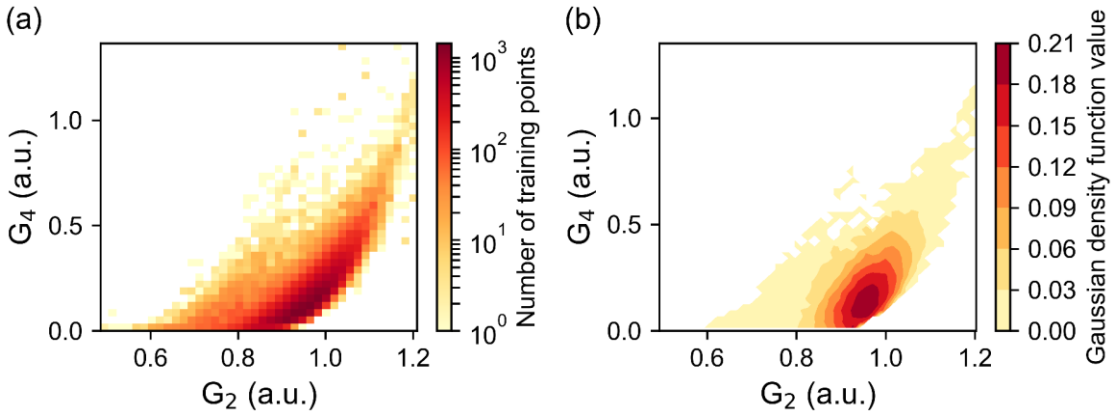

Figure S2. The distribution of 50,400 training points in **G**-space. G_2 and G_4 indicate a radial component and an angular component, respectively, that are selected out of 26-dimension coordinates of **G**. (a) The number of training points are enumerated within 40×40 mesh grid and color-coded in the log scale. (b) The training points color-coded by GDF value.

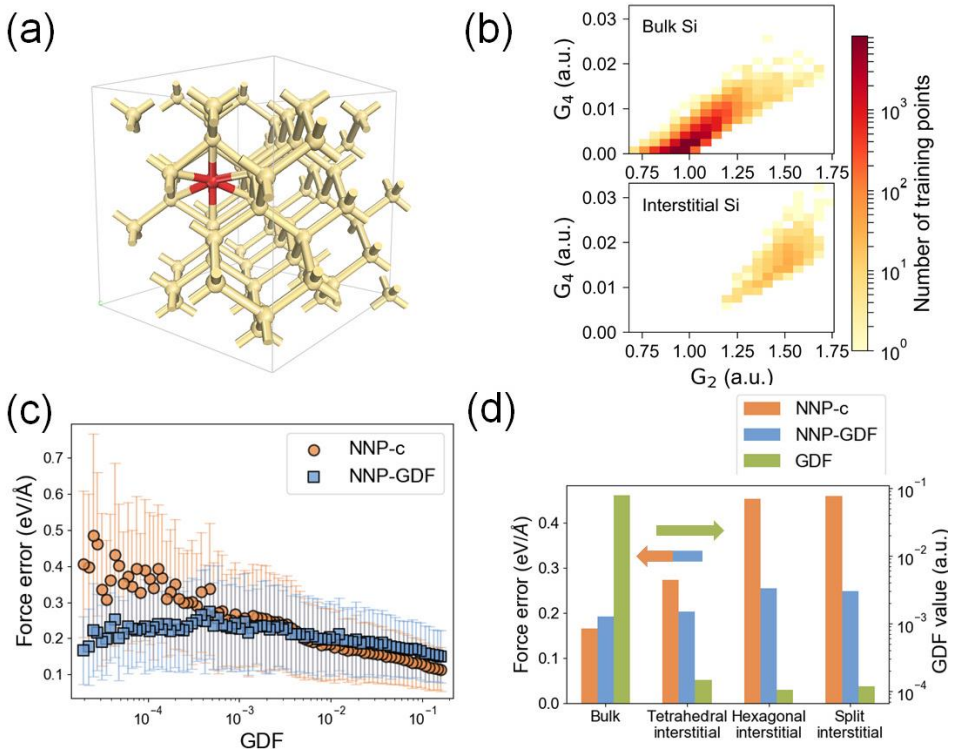

Figure S3. The NNP training for crystalline Si with one interstitial. (a) The atomic structures used in training. The interstitial atom is marked in red. (b) The distribution of 52,000 training

points in the **G** space. G_2 and G_4 indicate a radial component and an angular component, respectively, that are selected out of 26-dimensional coordinates of **G**. The number of training points are enumerated on the 20×20 mesh and color-coded in the log scale. (c) The GDF value versus force error for each training point. The results with NNP trained with the conventional method (NNP-c) are compared with those with the GDF weighting (NNP-GDF). The data are interval-averaged along GDF and error bars represent the standard deviation. (d) The root mean square force error and GDF values by the Si interstitial character. The force error is compared between NNP-c and NNP-GDF.

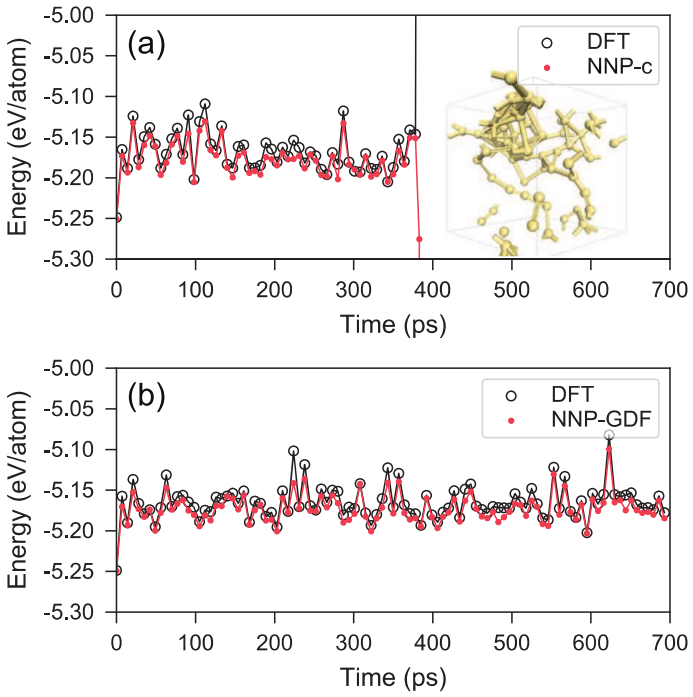

Figure S4. The energy per atom along the MD trajectory of Si with an interstitial defect at 1200 K. The trajectory is sampled by interval of 7 ps and the interatomic potentials are (a) NNP-c and (b) NNP-GDF. The red filled dots represent NNP energies and black empty circles represents DFT energies for the same structure. The inset picture is the structure after 400 ps of simulation.

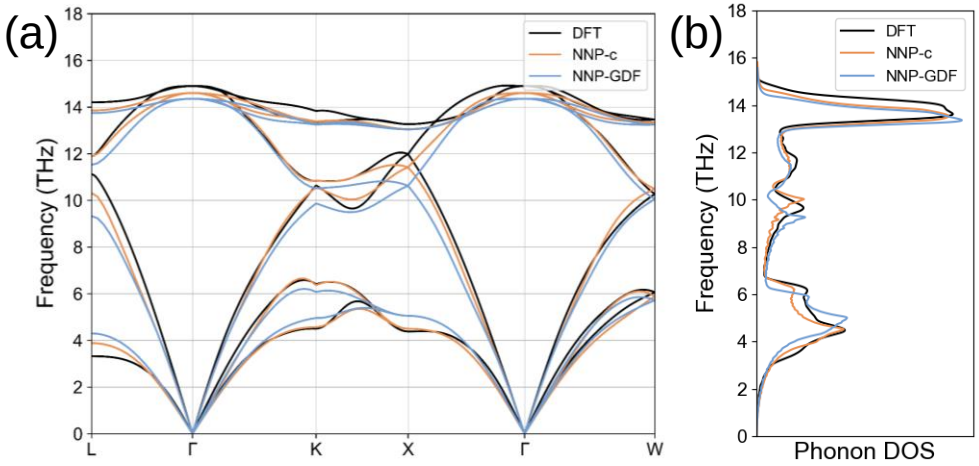

Figure S5. (a) Phonon dispersion curves for crystalline Si calculated by DFT, NNP-c and NNP-GDF. (b) Corresponding phonon DOS. The NNPs are trained with 500-1300 K MD snapshots, isotropic volume changed cells, volume-conserving uniaxial and monoclinic strained cells of Si perfect crystal, and 500-1300 K MD snapshots of Si crystalline with one vacancy.

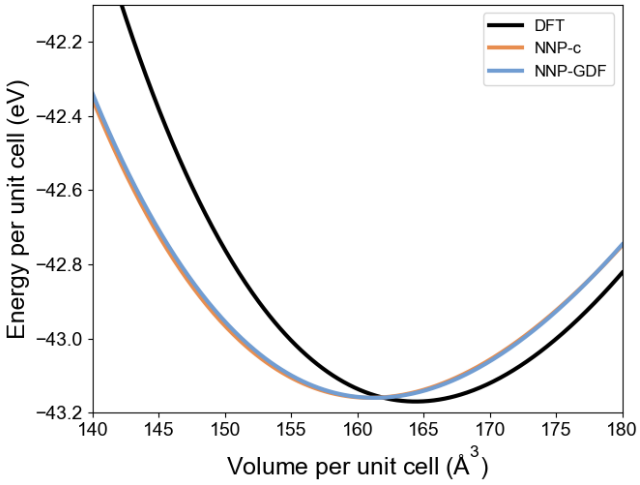

Figure S6. The equation of state for crystalline Si calculated by DFT, NNP-c, and NNP-GDF. The training condition is the same as in Fig. S5.

	DFT	NNP-c	NNP-GDF
Lattice constant ($Å^3$)	5.48	5.44	5.44
Bulk modulus (GPa)	88.2	72.3	72.5

Table S2. The lattice constant and bulk modulus for crystalline Si by DFT, NNP-c and NNP-GDF. They are obtained by fitting the equation of state in Figure S6 to the Birch-Murnaghan formula.

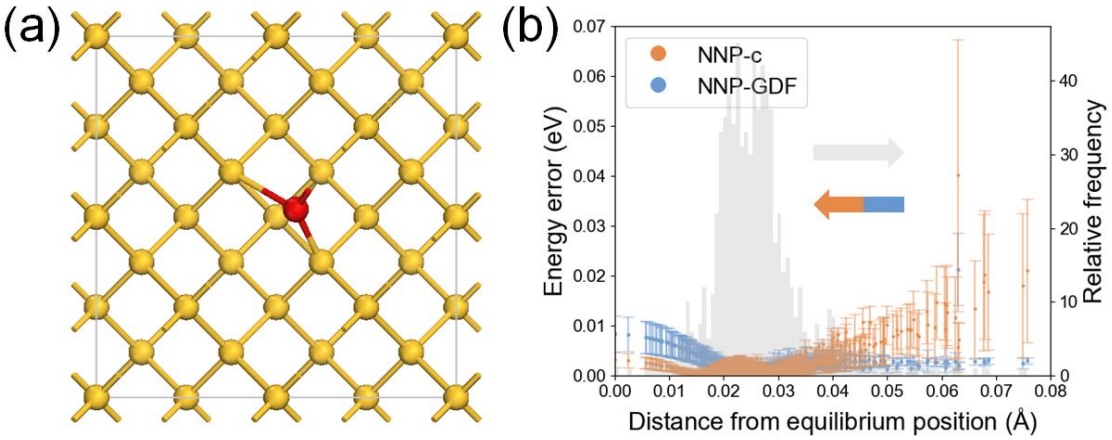

Figure S7. The prediction uncertainty estimated by training 9 NNPs with the same data. (a) Training structures made of Si. The temperature of the red Si atom is set to 3500 K while other atoms are fixed at the equilibrium position. (b) The energy errors of 9 NNPs with respect to the reference DFT energies are plotted as a function of the distance from equilibrium. Each dot is the average value and the vertical bar represents the standard deviation among 9 NNPs, which corresponds to the prediction uncertainty. The gray histogram shows the relative frequency within the training set. It is seen that the prediction uncertainty in NNP-c increases rapidly as the distance from the equilibrium increases and the sampling density decreases. In contrast, NNP-GDF maintains similar levels of prediction uncertainty.

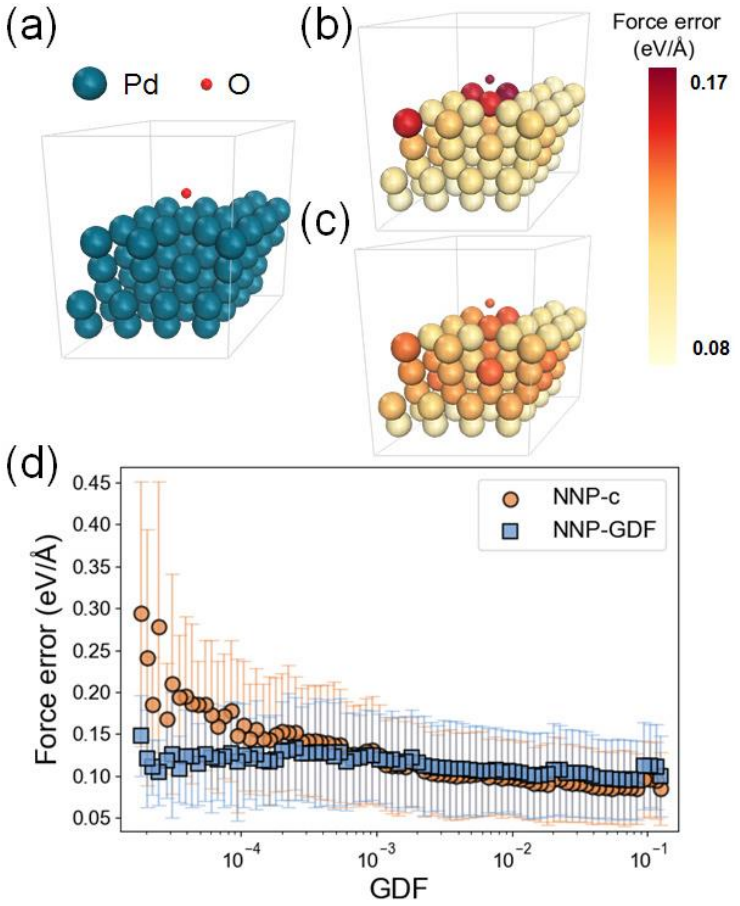

Figure S8. (a) The structures of Pd(111) surface with one oxygen adsorbate. The big (small) spheres represent the Pd (O) atoms. (b) The force error given by NNP-c is averaged in the atom-wise manner and color-coded on each atom. It is seen that the force error is higher for O and neighboring Pd atoms. (c) A similar figure with NNP-GDF. The force error more even than in (b). (d) The GDF value versus force error for each training point. The data are interval-averaged along GDF and error bars represent the standard deviation.

References

(S1) Kresse, G.; Hafner, Ab initio molecular dynamics for liquid metals. *J. Phys. Rev. B* **1993**, 47, 558.

(S2) Blöchl, P. E. Projector augmented-wave method. *Phys. Rev. B* **1994**, 50, 17953.

(S3) Perdew, J. P.; Burke K.; Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **1996**, 77, 3865.

(S4) Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G. S.; Davis, A.; Dean, J.; Devin, M. tensorflow: Large-scale machine learning on heterogeneous systems, software available from tensorflow.org.

(S5) Plimpton, S. Fast Parallel Algorithms for Short-Range Molecular Dynamics. *J. Comput. Phys.* **1995**, 117, 1-19.

(S6) Byrd, R. H.; Lu, P.; Nocedal, J.; Zhu, C. A limited memory algorithm for bound constrained optimization. *SIAM J. Sci. Comput.* **1995**, 16, 1190-1208.