THE JOURNAL OF PHYSICAL CHEMISTRY C

Article

Cite This: J. Phys. Chem. C 2018, 122, 22790−22795

pubs.acs.org/JPCC

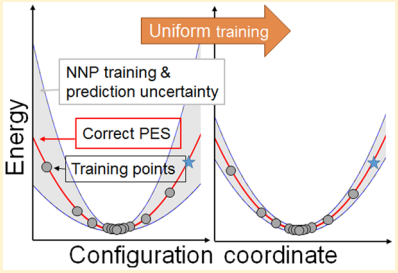
Toward Reliable and Transferable Machine Learning Potentials: Uniform Training by Overcoming Sampling Bias

Wonseok Jeong,[‡] Kyuhyun Lee,[‡] Dongsun Yoo, Dongheon Lee, and Seungwu Han*

Department of Materials Science and Engineering and Research Institute of Advanced Materials, Seoul National University, Seoul 08826, Korea

Supporting Information

ABSTRACT: The neural network interatomic potential (NNP) is anticipated to be a promising next-generation atomic potential for its self-learning capability and universal mathematical structure. While various examples demonstrate the usefulness of NNPs, we find that the NNP suffers from highly inhomogeneous feature-space sampling in the training set. As a result, underrepresented atomic configurations, often critical for simulations, cause large errors even though they are included in the training set. Using the Gaussian density function (GDF) that quantifies the sparsity of training points, we propose a weighting scheme that can effectively rectify the sampling bias. Various examples confirm that GDF weighting significantly improves the reliability and transferability of NNPs compared to the conventional training method, which is attributed to accurate mapping of atomic energies. By addressing a detrimental problem that is inherent in every machine learning potential, the present work will extend the application range of the machine learning potential.



By revealing atomic trajectories, classical molecular dynamics (MD) simulations have greatly advanced fundamental understanding of various chemical processes at the atomistic level. In classical MD, the chemical bonds are approximated by interatomic potentials that are parametrized by fitting key properties to experimental and ab initio data.[1] The functional form of the interatomic potentials reflects the underlying bonding nature such as ionic, covalent, and metallic characters. In many materials, however, the bonding nature is rather mixed, which makes it difficult to choose a proper function type. By describing reactive events with flexible functional forms, the reactive force field (ReaxFF) has significantly extended the application range of the classical MD simulations.[2] However, ReaxFF still relies on human intuition in introducing additional energy terms, which renders the potential generation a formidable task, especially for multicomponent systems.

Recently, the machine learning (ML) potential is gaining traction as a data-driven approach to generating interatomic potentials. In contrast to traditional interatomic potentials with preset analytic functions, the ML potentials assume general and flexible mathematical structures such as a neural network[3,4] or Gaussian process regression,[5,6] and their parameters are optimized through ML on extensive ab initio data. In particular, a good deal of interest is given to the high-dimensional neural network potential (NNP) suggested by Behler and Parrinello[4] that has been applied to a wide range of materials encompassing metals,[7,8] oxides,[9] semiconductors,[4,10] and molecular reactions.[11] Furthermore, several studies demonstrated the performance of NNPs in complex systems such as surface reactions,[12,13] phase transitions,[14] diffusion of Li in amorphous oxides,[15] proton transfer in aqueous solution,[16] and solid−liquid interfaces.[17,18]

In the high-dimensional NNP, the total energy is represented as a sum of atomic energies that depend on the local environment of each atom. The local atomic configuration is in turn described by the symmetry function vector **G** that encodes the radial and angular distribution of neighboring atoms within a certain cutoff radius.[4,19] The atomic energy is then obtained through the feed-forward artificial neural network with **G** as the input layer. That is to say, **G** corresponds to the feature space in the atomic neural network. The parameters in the NNP are optimized through learning ab initio [preferentially density functional theory (DFT)] energies and forces for specific structures (called the training set) that typically consist of crystalline structures with various deformations, surfaces, defects, and MD snapshots. Each atom in the training set represents a certain training point in the feature space **G**. The optimization algorithm minimizes the following loss function (Γ)

$$\Gamma = \Gamma_{\text{energy}} + \Gamma_{\text{force}} = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{E_i^{\text{DFT}} - E_i^{\text{NNP}}}{n_i}\right)^2 + \frac{\mu}{3M}\sum_{j=1}^{M}|\mathbf{F}_j^{\text{DFT}} - \mathbf{F}_j^{\text{NNP}}|^2 \quad (1)$$

where N is the total number of structures in the training set and n_i and $E_i^{\text{DFT(NNP)}}$ are the number of atoms and DFT (NNP) energy of the ith training structure respectively. In eq 1, j is the index for atoms in the whole training set whose total number is M, and $\mathbf{F}_j^{\text{DFT(NNP)}}$ is the DFT (NNP) force of the jth atom. In addition, μ determines the relative weight between

Received: August 19, 2018
Published: August 20, 2018

DOI: 10.1021/acs.jpcc.8b08063
J. Phys. Chem. C 2018, 122, 22790−22795

the energy loss function (Γ_{energy}) and the force loss function (Γ_{force}).

While the NNP is getting popular, the weakness and strength of the NNP are not fully understood at this moment, mainly because of its "black-box" nature. For example, it was recently noted that the prediction error of NNP increases in an uncontrollable way as the local environment deviates from $\{\mathbf{G}_j\}$ in the training set.[20,21] In the present work, we raise another critical issue in training NNPs, i.e., the distribution of $\{\mathbf{G}_j\}$ in the training set is highly inhomogeneous and biased. This results in unbalanced training, which significantly undermines accuracy and reliability of NNP. We propose an effective method that equalizes the learning level over $\{\mathbf{G}_j\}$ in the training set, thereby substantially improving the reliability as well as transferability of the NNP.

We first demonstrate the inhomogeneous distribution of $\{\mathbf{G}_j\}$ with an example of crystalline Si with one vacancy (see Figure 1a). The training set was generated by performing

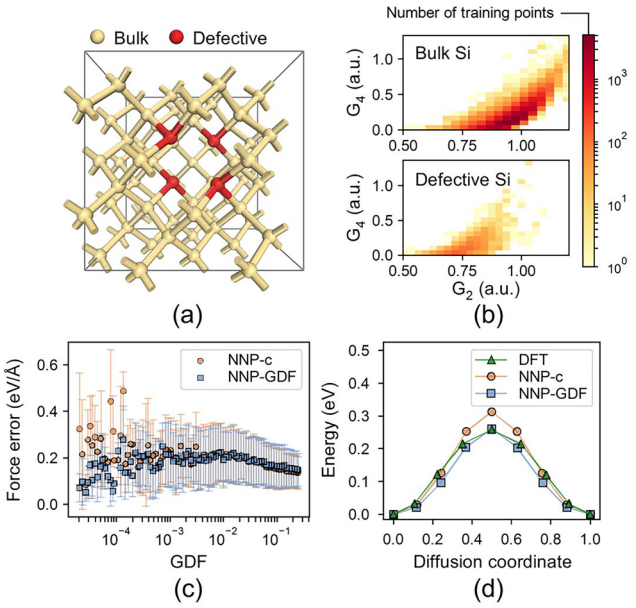


Figure 1. NNP training for crystalline Si with one vacancy. (a) Atomic structures used in training. The defective atoms surrounding the vacancy are marked in red. (b) Distribution of training points in the **G** space. G_2 and G_4 indicate a radial component and an angular component, respectively, that are selected out of 26-dimensional coordinates of **G**. The number of training points is enumerated on the 20 × 20 mesh and color-coded on the log scale. (c) GDF value versus force error for each training point. The data are interval-averaged along the GDF, and error bars are standard deviations. (d) Vacancy migration barrier calculated by the NEB method. The energy of the Si vacancy at equilibrium is set to zero.

DFT-MD simulations on the 63 atom supercell for 16 ps with the temperature varying from 500 to 1300 K and sampling the snapshots every 20 fs. (See the Supporting Information for details.) In total, 50 400 points were sampled in the feature space, and 90% of them were randomly selected and used for training the NNP. Figure 1b shows the frequency of $\mathbf{G}_j$ on the G_2−G_4 plane on the log scale, where G_2 and G_4 correspond to a radial component and an angular component, respectively.[19] (Other components also show similar distributions.) The distributions of four-fold-coordinated bulk Si atoms and three-fold-coordinated defective Si atoms neighboring the vacancy site (red atoms in Figure 1a) are displayed separately. It is striking in Figure 1b that the distribution is highly biased and concentrated within a narrow range of **G**. First, most **G** vectors belong to bulk Si atoms. This is because training structures contain far more bulk atoms than defective ones (59 versus 4). In addition, the sampling is concentrated around the equilibrium point, which is a result of the Boltzmann distribution.

Because the NNP learns on $\{\mathbf{G}_j\}$ in the training set, the inhomogeneous distribution enforces the NNP to be optimized toward specific configurations (bulk and equilibrium), sacrificing accuracy for underrepresented configurations (defect or off-equilibrium). In order to investigate this quantitatively, we define a Gaussian density function [GDF; $\rho(\mathbf{G})$] defined for an arbitrary **G** in the symmetry-function space as follows

$$\rho(\mathbf{G}) = \frac{1}{M}\sum_{j=1}^{M} \exp\left(-\frac{1}{2\sigma^2}\frac{|\mathbf{G}-\mathbf{G}_j|^2}{D}\right) \tag{2}$$

where σ is the Gaussian width and D is the dimension of the symmetry function vector (i.e., $\mathbf{G} \in \mathbb{R}^D$). Other notations are the same as those in eq 1. Equation 2 implies that the value of $\rho(\mathbf{G})$ ranges between 0 and 1; values close to 0 (or 1) mean that training points are scarce (or abundant) around the specific **G**. Figure S2 in the Supporting Information shows that the GDF well represents the actual distribution of **G**.

Figure 1c shows the remnant force error for each atom ($\Delta F_j = |\mathbf{F}_j^{DFT} - \mathbf{F}_j^{NNP}|$ after training the NNP with the conventional loss function in eq 1 (NNP-c hereafter). (See the Supporting Information for details of training the NNP.) For visual clarity, ΔF_j's are interval-averaged with respect to the GDF (see circles). It is seen that the force error increases for atoms with small GDF values and the root-mean-square error (RMSE) of defective Si is 0.27 eV/Å, much larger than 0.18 eV/Å for the bulk atoms. These force errors are for the training set and hence evidence the unbalanced training between bulk atoms and undersampled defective atoms. Figure 1d shows the energy barrier of vacancy migration calculated by the nudged-elastic-band (NEB) method.[22] Even though trajectories of the vacancy migration are included in the training set (the vacancy migrates several times at temperatures higher than 1000 K within the present simulation time), the NNP overestimates the energy barrier by 50 meV, which will significantly underestimate the diffusion coefficient of the vacancy.

We stress that the inhomogeneous distribution of $\{\mathbf{G}_j\}$ persists for any reasonable choice of training sets. For instance, nonbulk structures such as surfaces and defects are underrepresented in general because they should be modeled together with bulk atoms that outnumber the atoms under the interested environment. In addition, breaking and forming of bonds, critical in chemical reactions, are rare events and occur only a few times during long-time MD simulations.

In ref 23, the redundancy of atomic environments in the training set was also noted, and a solution was proposed on the basis of force-amplitude sampling. However, the method still risks omitting critical configurations. The sampling bias could also be overcome through efficient but deliberate sampling around an atypical configuration.[24−26] However, this approach has limitations because it is tricky to construct training models that consist of only specific configurations. In the below, we suggest a simple but very effective method that improves the

DOI: 10.1021/acs.jpcc.8b08063
J. Phys. Chem. C 2018, 122, 22790−22795

training uniformity for the given training set without additional sampling.

In order to alleviate the unbalanced training, it is necessary to enhance the influence of undersampled local environments. We first note that GDF defined in eq 2 can detect the scarcity of training points around a specific **G** point. In addition, while the DFT energy per se cannot be split into local atomic energies, the DFT forces are obtained for individual atoms. Exploiting these two facts, we modify the loss function as follows

$$\Gamma = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{E_i^{\mathrm{DFT}} - E_i^{\mathrm{NNP}}}{n_i}\right)^2 + \frac{\mu}{3M}\sum_{j=1}^{M}\Theta\left[\frac{1}{\rho(\mathbf{G}_j)}\right] |\mathbf{F}_j^{\mathrm{DFT}} - \mathbf{F}_j^{\mathrm{NNP}}|^2 \quad (3)$$

By choosing Θ in eq 3 as a monotonically increasing function, one can magnify the influence of $\mathbf{G}_j$'s with small GDF values on the loss function. Among various choices of Θ, we select a modified sigmoid function as it produced the best results in various cases

$$\Theta(x) = \frac{Ax}{1 + e^{-bx+c}} \quad (4)$$

where A is a normalizing constant that makes the average of Θ to be 1 and b and c are parameters that are fine-tuned for balanced training (chosen to be 150 and 1.0, respectively, in this work). In the below, we demonstrate how the modified loss function rectifies the unbalanced training.

Figure 1c shows that the NNP with GDF weighting (NNP-GDF hereafter) gives more uniform errors over the whole range of GDFs and the average force errors at **G**'s with GDF < 10^{-3} are less than 0.3 eV/Å. The RMSEs of defective and bulk atoms are 0.19 and 0.18 eV/Å, respectively, which are more even than those in NNP-c (see above). It is also noticeable that the average force error slightly increases for GDF > 10^{-1}. This supports that the GDF weighting effectively increases (decreases) the influence of underrepresented (overrepresented) **G** points. (The force errors on the test set also confirm a similar effect of GDF weighting.) In Figure 1d, the migration barrier by NNP-GDF agrees with the DFT result within 3 meV.

We also confirm benefits of the GDF weighting with the example of Si interstitials; the training set is generated by carrying out DFT-MD simulations with the 65 atom supercell including one interstitial atom at temperatures from 500 to 1300 K. Like in the vacancy example, force errors for interstitials are much larger than those for bulk atoms in NNP-c, but errors become more even when GDF weighting is applied. (See the Supporting Information for details.) To further check the accuracy of the NNP, we scan the potential energy surface (PES) of the Si interstitial around the equilibrium point (see Figure 2). The PES is obtained by displacing the interstitial atom on a spherical surface with a radius of 0.6 Å while other atoms are fixed. Comparing PESs from DFT, NNP-c, and NNP-GDF (Figure 2b–d, respectively), one can see that NNP-GDF gives a PES closer to the DFT result than NNP-c.

Furthermore, the GDF weighting improves the stability of NNP-MD simulations. For instance, MD simulations with a Si interstitial using NNP-c's trained in the above failed repeatedly within 1 ns for temperatures above 1200 K and resulted in unphysical structures (see Figure S4a in the Supporting Information). This can be understood as follows: the configurations with large vibrational amplitudes occur during MD at high temperatures but they are underrepresented due to the Boltzmann factor. This leads to large and unpredictable force errors. In contrast, MD with NNP-GDF run stably for a much longer time (see Figure S4b in the Supporting Information).

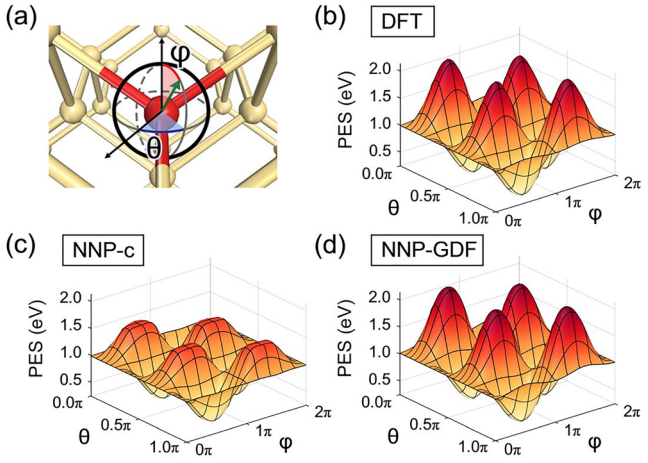


Figure 2. PES of the tetrahedral Si interstitial around the equilibrium point. (a) Angular coordinates of the displaced interstitial. The radial distance from the equilibrium point is 0.6 Å. (b–d) PESs as a function of angular coordinates calculated by DFT, NNP-c, and NNP-GDF, respectively. The energy is referenced to the value at equilibrium for each method.

Figure 3 schematically depicts the main idea of the present method using the diagram of energy versus configuration coordinates. The training points indicated by circles are concentrated near the energy minimum or equilibrium point. In the conventional training (Figure 3a), training and prediction uncertainties (shaded region) increase rapidly for underrepresented points. Through GDF weighting (Figure 3b), the uncertainty level becomes more even throughout the training range regardless of the sampling density. This also implies that the GDF weighting will effectively improve the transferability of NNP because new configurations lie outside of the training set (see stars in Figure 3) and NNP-GDF would give a prediction error smaller than NNP-c. This schematic idea can be confirmed with actual systems by estimating the prediction uncertainty from multiple NNPs trained on the

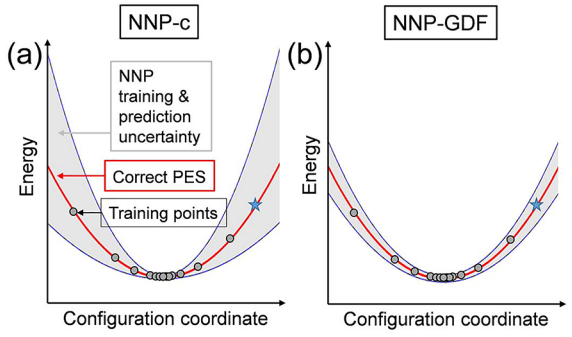


Figure 3. Schematic picture showing training and prediction uncertainty of the NNP when the sampling density is inhomogeneous. The training points are indicated by circles. (a) Conventional training. (b) Training with GDF weighting.

DOI: 10.1021/acs.jpcc.8b08063
J. Phys. Chem. C 2018, 122, 22790–22795

same data.[20,21] The detailed information on the example of Si is provided in Figure S7 in the Supporting Information.

For a fixed size of neutral network, it is unavoidable to sacrifice the accuracy of the PES in some part in order to improve the accuracy in another part. This means that the GDF weighting can undermine the accuracy of physical properties at the equilibrium point with high GDF values (the bottom region in Figure 3). For instance, Figures 1c and S3c show that the force error slightly increases for **G**'s with high GDF values. To check how the equilibrium property is actually affected by GDF weighting, we additionally compare equilibrium properties such as the lattice parameter, bulk modulus, phonon dispersion curve, and phonon DOS for crystalline Si between DFT, NNP-c, and NNP-GDF. The results are provided in Figures S5 and S6 and Table S2 of the Supporting Information. While the phonon dispersion curve from NNP-c is slightly more accurate than that from NNP-GDF, the overall accuracy of NNP-c and NNP-GDF are comparable for the equilibrium property.

As another test, we calculate formation energies of vacancy clusters when the training set includes only monovacancy structures. This will probe the transferability of the NNP to configurations that are slightly different from those in the training set. We first train NNP-c and NNP-GDF with the training set consisting of fcc Si structures that are prefect or include one vacancy. (In detail, snapshots are sampled from DFT-MD simulations with the temperature ramped from 500 to 1300 K and crystals with various deformations.) We then calculate the formation energy (E_{for}) as $E_{for} = E_{tot}(\text{defect}) - N \cdot E_{si}$, where $E_{tot}(\text{defect})$ is the total energy of the N-atom supercell with vacancies and E_{si} is the total energy per atom of the perfect fcc Si. The results are summarized in Table 1. Even

Table 1. Formation Energies of Vacancy Clusters in eV Calculated by DFT, NNP-c, and NNP-GDF[a]

	monovacancy	trivacancy	pentavacancy
DFT	3.59	7.14	9.87
NNP-c	3.34 (−6.8%)	7.83 (9.6%)	11.08 (12.2%)
NNP-GDF	3.45 (−4.0%)	7.09 (−0.6%)	10.19 (3.2%)

[a]The training set includes only monovacancy and crystalline structures. the 2×2×2 supercells are used for mono- and trivacancies, while the 3×3×3 supercells are used for the pentavacancy. The vacancy clusters are generated by removing Si atoms connected in the most compact way. The relative errors are indicated in parentheses.

for the monovacancy, NNP-GDF gives a smaller error than NNP-c even though monovacancy structures are included in the training set, implying that NNP-GDF learns on the vacancy property better than NNP-c. More importantly, Table 1 shows that the prediction error for larger vacancies increases for NNP-c (up to 12.2%) while those for NNP-GDF remain similar to that of the monovacancy. This demonstrates that NNP-GDF has better transferability than NNP-c. The enhanced transferability will allow for simplification of the training set.

As mentioned in the introduction, NNP represents the DFT total energy as a sum of atomic energies. Therefore, the accuracy of the NNP ultimately lies in its ability to correctly map out the atomic energies as a function of local environments. It is worthwhile to investigate how the atomic energy mapping is affected by GDF weighting. To this end, we train NNPs on total energies and forces obtained by the embedded-atom-method (EAM) potential for Ni[27] and directly compare atomic energies from the NNP with those from the EAM. The training set was generated for Ni vacancy structures (499 atom fcc supercells with temperatures from 0 to 400 K).

Figure 4a shows the correlation of the atomic energies between the EAM and NNP-c. While the overall agreements

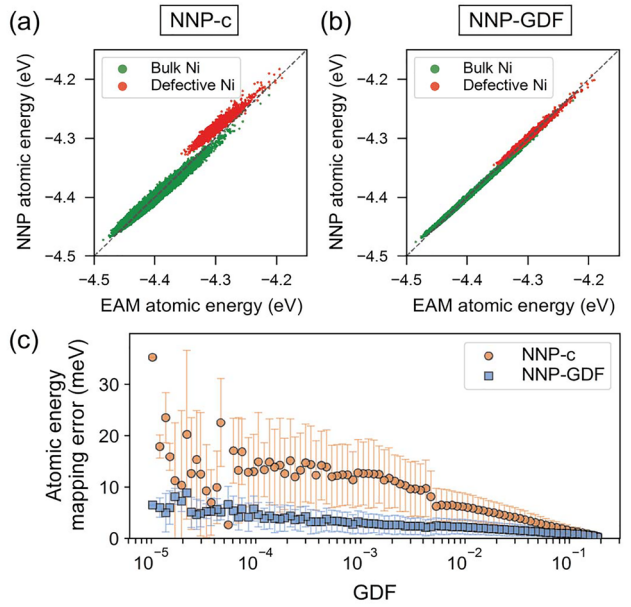


Figure 4. Correlation of atomic energies between the EAM and NNP. The training set consists of MD snapshots of crystalline Ni with one vacancy. (a) Result of conventionally trained NNPs. (b) Result of the GDF-weighted NNP. (c) GDF versus mapping errors that are interval-averaged. The error bar represents the standard deviation.

are good with a mean absolute error (MAE) of 3.3 meV, the mapping is relatively inaccurate for defective Ni atoms (16.8 and 2.9 meV for the defective and bulk Ni atoms, respectively). The mapping error with respect to the GDF in Figure 4c is consistent with this. On the other hand, NNP-GDF maps the atomic energy correctly for the whole range, as shown in Figure 4b,c. This can be explained as follows: the atomic force depends on the atomic energy function around the given configuration. Therefore, by fitting atomic forces evenly from underrepresented to overrepresented configurations, the GDF weighting constrains the atomic energy mapping to be consistent throughout all training points. The correct atomic energy mapping would be the fundamental reason underlying the better transferability of NNP-GDF. We note that the MAE of the total energy is 0.035 meV/atom for NNP-GDF, larger than 0.014 meV/atom for NNP-c. That is to say, NNP-GDF trades accuracy in the total energy for consistency in the atomic energy mapping.

We note that the data imbalance has been widely discussed within the ML community.[28] Suggested solutions are adding more features to the training set that has been undersampled (sometimes synthetic samples[29]) or leaving out some of the features that have been oversampled. However, it is not feasible to apply these methods to the present problem because the NNP predicts the atomic energy while it is trained through total energies that are the sum of atomic energies. This poses a

DOI: 10.1021/acs.jpcc.8b08063
J. Phys. Chem. C 2018, 122, 22790−22795

unique challenge in training the NNP, which motivated us to suggest the present method based on the GDF.

In a bigger picture, the GDF weighting contributes to establishing a close correspondence between what one wants the NNP to learn and what the NNP actually learns through the training procedure, which is at the heart of every ML potential. Clearly, the advantages of GDF weighting extend to other systems. For example, we carried out similar analysis on the Pd(111) surface with an oxygen adsorbate, which is a key step in various catalytic reactions,[12,30] and reconfirmed the merits of NNP-GDF (see the Supporting Information for details). We also note that when used together with the ensemble NNP approach in refs 20 and 21 the GDF weighting will significantly reduce the iteration number to achieve a converged set of NNPs. Lastly, inhomogeneous feature sampling is a general issue for any ML potentials that adopt local feature vectors as input. Therefore, the present method can be equally applied to other types of NN potentials such as deep tensor neural networks[31] and deep potential,[32] as well as other ML potentials.[5,6]

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpcc.8b08063.

DFT calculations and the neural network interatomic potential (NNP) training using a homemade package (SNU Interatomic Machine-learning PotentiaL packagE - ver. NN; SIMPLE-NN), comparison of the Gaussian density function (GDF) and actual density distribution in the feature space, and examples of Si interstitials and Pd(111)/O (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: hansw@snu.ac.kr.

ORCID

Wonseok Jeong: 0000-0001-8894-1857

Dongsun Yoo: 0000-0002-4889-8396

Author Contributions

‡W.J. and K.L. contributed equally.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the Technology Innovation Program (or Industrial Strategic Technology Development Program [10052925, Atomistic process and device modeling of sub-10 nm scale transistors]) funded By the Ministry of Trade, Industry & Energy (MOTIE, Korea) and Creative Materials Discovery Program through the National Research Foundation of Korea (NRF) funded by Ministry of Science and ICT (2017M3D1A1040689). The computations were performed at the KISTI supercomputing center (Grant No. KSC-2017-C3-0019)

■ REFERENCES

(1) Mendelev, M. I.; Han, S.; Srolovitz, D. J.; Ackland, G. J.; Sun, D. Y.; Asta, M. Development of new interatomic potentials appropriate for crystalline and liquid iron. *Philos. Mag.* **2003**, *83*, 3977−3994.

(2) Senftle, T. P.; Hong, S.; Islam, M. M.; Kylasa, S. B.; Zheng, Y.; Shin, Y. K.; Junkermeier, C.; Engel-Herbert, R.; Janik, M. J.; Aktulga.; et al. The ReaxFF reactive force-field: development, applications and future directions. *NPJ. Comput. Mater.* **2016**, *2*, 15011−15023.

(3) Blank, T. B.; Brown, S. D.; Calhoun, A. W.; Doren, D. J. Neural network models of potential energy surfaces. *J. Chem. Phys.* **1995**, *103*, 4129−4137.

(4) Behler, J.; Parrinello, M. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys. Rev. Lett.* **2007**, *98*, 146401.

(5) Bartók, A. P.; Payne, M. C.; Kondor, R.; Csányi, G. Gaussian approximation potentials: The accuracy of quantum mechanics, without the electrons. *Phys. Rev. Lett.* **2010**, *104*, 136403.

(6) Bartók, A. P.; Csányi, G. G aussian approximation potentials: A brief tutorial introduction. *Int. J. Quantum Chem.* **2015**, *115*, 1051.

(7) Artrith, N.; Kolpak, A. M. Understanding the composition and activity of electrocatalytic nanoalloys in aqueous solvents: A combination of DFT and accurate neural network potentials. *Nano Lett.* **2014**, *14*, 2670.

(8) Boes, J. R.; Kitchin, J. R. Modeling segregation on AuPd (111) surfaces with density functional theory and monte carlo simulations. *J. Phys. Chem. C* **2017**, *121*, 3479.

(9) Artrith, N.; Urban, A. An implementation of artificial neural-network potentials for atomistic materials simulations: Performance for TiO_2. *Comput. Mater. Sci.* **2016**, *114*, 135−150.

(10) Sosso, G. C.; Miceli, G.; Caravati, S.; Giberti, F.; Behler, J.; Bernasconi, M. Fast crystallization of the phase change compound GeTe by large-scale molecular dynamics simulations. *J. Phys. Chem. Lett.* **2013**, *4*, 4241−4246.

(11) Kolb, B.; Zhao, B.; Li, J.; Jiang, B.; Guo, H. P. Permutation invariant potential energy surfaces for polyatomic reactions using atomistic neural networks. *J. Chem. Phys.* **2016**, *144*, 224103.

(12) Boes, J. R.; Kitchin, J. R. Neural network predictions of oxygen interactions on a dynamic Pd surface. *Mol. Simul.* **2017**, *43*, 346−354.

(13) Ulissi, Z. W.; Tang, M. T.; Xiao, J.; Liu, X.; Torelli, D. A.; Karamad, M.; Cummins, K.; Hahn, C.; Lewis, N. S.; Jaramillo.; et al. Machine-Learning Methods Enable Exhaustive Searches for Active Bimetallic Facets and Reveal Active Site Motifs for CO_2 Reduction. *ACS Catal.* **2017**, *7*, 6600−6608.

(14) Khaliullin, R. Z.; Eshet, H.; Kühne, T. D.; Behler, J.; Parrinello, M. Nucleation mechanism for the direct graphite-to-diamond phase transition. *Nat. Mater.* **2011**, *10*, 693−697.

(15) Li, W.; Ando, Y.; Minamitani, E.; Watanabe, S. Study of Li atom diffusion in amorphous Li_3PO_4 with neural network potential. *J. Chem. Phys.* **2017**, *147*, 214106.

(16) Hellström, M.; Behler, J. Concentration-Dependent Proton Transfer Mechanisms in Aqueous NaOH Solutions: From Acceptor-Driven to Donor-Driven and Back. *J. Phys. Chem. Lett.* **2016**, *7*, 3302−3306.

(17) Natarajan, S. K.; Behler, J. Neural network molecular dynamics simulations of solid−liquid interfaces: water at low-index copper surfaces. *Phys. Chem. Chem. Phys.* **2016**, *18*, 28704−28705.

(18) Quaranta, V.; Hellström, M.; Behler, J. Proton-Transfer Mechanisms at the Water−ZnO Interface: The Role of Presolvation. *J. Phys. Chem. Lett.* **2017**, *8*, 1476−1483.

(19) Behler, J. Atom-centered symmetry functions for constructing high-dimensional neural network potentials. *J. Chem. Phys.* **2011**, *134*, 074106.

(20) Peterson, A. A.; Christensen, R.; Khorshidi, A. Addressing uncertainty in atomistic machine learning. *Phys. Chem. Chem. Phys.* **2017**, *19*, 10978−10985.

(21) Behler, J. First principles neural network potentials for reactive simulations of large molecular and condensed systems. *Angew. Chem., Int. Ed.* **2017**, *56*, 12828−12840.

(22) Sheppard, D.; Xiao, P.; Chemelewski, W.; Johnson, D. D.; Henkelman, G. A generalized solid-state nudged elastic band method. *J. Chem. Phys.* **2012**, *136*, 074103.

(23) Huan, T. D.; Batra, R.; Chapman, J.; Krishnan, S.; Chen, L.; Ramprasad, R. A universal strategy for the creation of machine learning-based atomistic force fields. *NPJ. Comput. Mater.* **2017**, *3*, 37.
(24) Pártay, L. B.; Bartók, A. P.; Csányi, G. Efficient sampling of atomic configurational spaces. *J. Phys. Chem. B* **2010**, *114*, 10502−10512.
(25) Toyoura, K.; Hirano, D.; Seko, A.; Shiga, M.; Kuwabara, A.; Karasuyama, M.; Shitara, K.; Takeuchi, I. Machine-learning-based selective sampling procedure for identifying the low-energy region in a potential energy surface: A case study on proton conduction in oxides. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2016**, *93*, 054112.
(26) Martoňák, R.; Laio, A.; Parrinello, M. Predicting Crystal Structures: The Parrinello-Rahman Method Revisited. *Phys. Rev. Lett.* **2003**, *90*, 075503.
(27) Foiles, S. M.; Baskes, M. I.; Daw, M. S. Embedded-atom-method functions for the fcc metals Cu, Ag, Au, Ni, Pd, Pt, and their alloys. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1986**, *33*, 7983.
(28) Haibo, H.; Garcia, E. A. Learning from Imbalanced Data. *IEEE Trans. on Knowl. and Data Eng.* **2009**, *21*, 1263.
(29) Chawla, N. V.; Bowyer, K. W.; Hall, L. O.; Kegelmeyer, W. P. SMOTE: Synthetic Minority Over-sampling Technique. *J. Artif. Int. Res.* **2002**, *16*, 321.
(30) Lu, Y.; Gasteiger, H. A.; Shao-Horn, Y. *J. Am. Chem. Soc.* **2011**, *133*, 19048−19051.
(31) Schütt, K. T.; Arbabzadah, F.; Chmiela, S.; Müller, K. R.; Tkatchenko, A. Catalytic activity trends of oxygen reduction reaction for nonaqueous Li-air batteries. *Nat. Commun.* **2017**, *8*, 13890.
(32) Han, J.; Zhang, L.; Car, R.; E, W. Deep Potential: A General Representation of a Many-Body Potential Energy Surface. arXiv:1707.01478v2.

DOI: 10.1021/acs.jpcc.8b08063
J. Phys. Chem. C 2018, 122, 22790−22795