

Supplementary Materials

Comparison of Different Machine-Learning Models for Prediction of Forces in Copper and Silicon Dioxide

Wenwen Li¹, Yasunobu Ando¹

¹ *Research Center for Computational Design of Advanced Functional Materials, National Institute of Advanced Industrial Science and Technology, Tsukuba, Ibaraki 305-8568, Japan*

List of Content

Section I Composition of reference database	2
Section II Test of cutoff distance R_c	3
Section III Genetical algorithm automatic selection protocol	4
Section IV ML force field under rotation operation.....	5
Section V ML force field for SiO ₂ with Coulomb interaction	6
Section VI Energy variation along NVT and NVE-MD	8
Section VII. Comparison of ML force field and high-dimensional NN potential.....	9
Section VIII Gaussian mixture model groups of Cu fingerprints.....	11
Section IX Parameters of structural fingerprints	12
Reference	16

Section I Composition of reference database

Table S1. The composition of reference database for Cu machine learning force field.

	Pristine Num. of Structures	Num. of Atoms	Vacancy Num. of Structures	Num. of Atoms	Interstitial Num. of Structures	Num. of Atoms
Fcc	3200	166400	3200	159632	3200	170594
surface (110)	1200	99200	1200	95554	-	-
surface (111)	920	94080	920	91007	-	-
amorphous	2800	115200	2800	110231	2800	118414
Total Structures: 22240						
Total Atoms: 1220312						

Table S2. The composition of reference database for SiO₂ machine learning force field.

	Pristine Num. of Structures	Num. of Atoms	Vacancy Num. of Structures	Num. of Atoms	Interstitial Num. of Structures	Num. of Atoms
quartz	9950	147600	4798	120557	1950	47996
cristobalite	7900	153600	3700	120691	1794	62433
quartz surface (110)	2500	54000	500	8492	-	-
cristobalite surface (111)	2500	72000	498	11434	-	-
amorphous	3000	63000	200	3983	100	2200
Total Structures: 39390						
Total Atoms: 867986						

Section II Test of cutoff distance R_c

The cutoff distance R_c of structural fingerprints is determined by the convergence test. The force prediction error usually reduces with the increasing of cutoff distance R_c . However, the decrease of error is not significant anymore when the R_c is large enough. So, we empirically decide the value of R_c as the position that convergence curve becoming flat. In this test, the linear regression model is build to get the force prediction error.

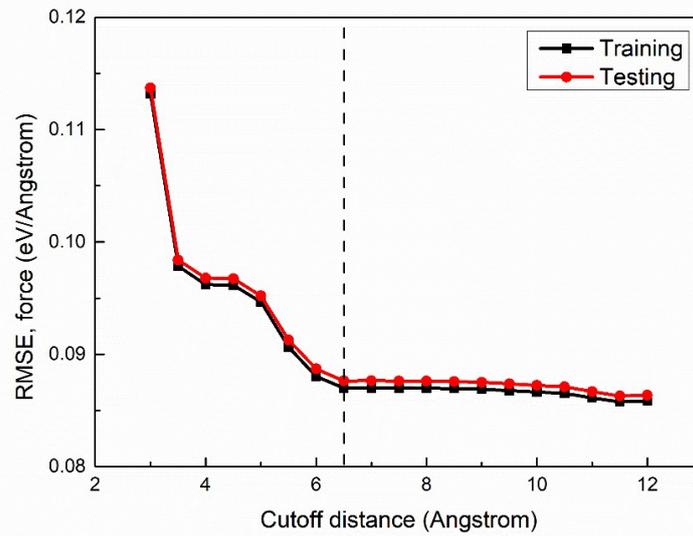


FIG. S1. The cutoff distance convergence test for Cu.

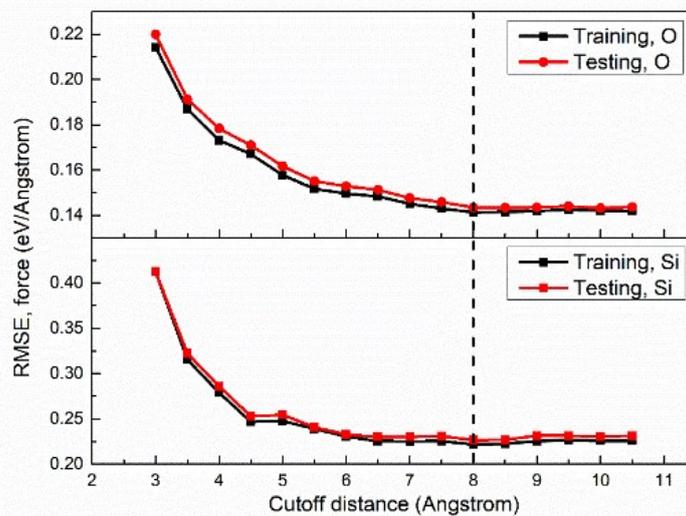


FIG. S2. The cutoff distance convergence test for SiO_2 .

Section III Genetical algorithm automatic selection protocol

Before the genetic algorithm selection, a large pool of fingerprint candidates is generated. In the genetic algorithm, one individual is corresponding to a set of fingerprints, which are selected from the candidate pool. The number of fingerprints in one individual, *i.e.* the length of gene, is fixed to N . The population size of any generation is fixed to M .

- Step 1. The initial generation was created by randomly selecting N fingerprints from the candidate pool.
- Step 2. The linear regression model was constructed with each individual set of fingerprints, and the δ_{RMS} of the test set is calculated. The fitness of everyone is represented as $[\max(\delta_{RMS}) - \delta_{RMS}]/[\max(\delta_{RMS}) - \min(\delta_{RMS})]$.
- Step 3. The individuals are randomly selected based on their fitness, and then, they went through crossover and mutation to create the new generation. Crossover means two individual exchange some of fingerprints. Mutation means substitute one fingerprints with another which randomly selected from candidate pool. The Mutation possibility is 20%.
- Step 4. Repeat the step 2 and 3 until the minimum $\delta_{RMS,F}$ is converged.

The Fig. S3 shows an example of how the GA automatic protocol select 8 fingerprints for Cu system (using set 4 as the candidate pool). The population size M is set to 144.

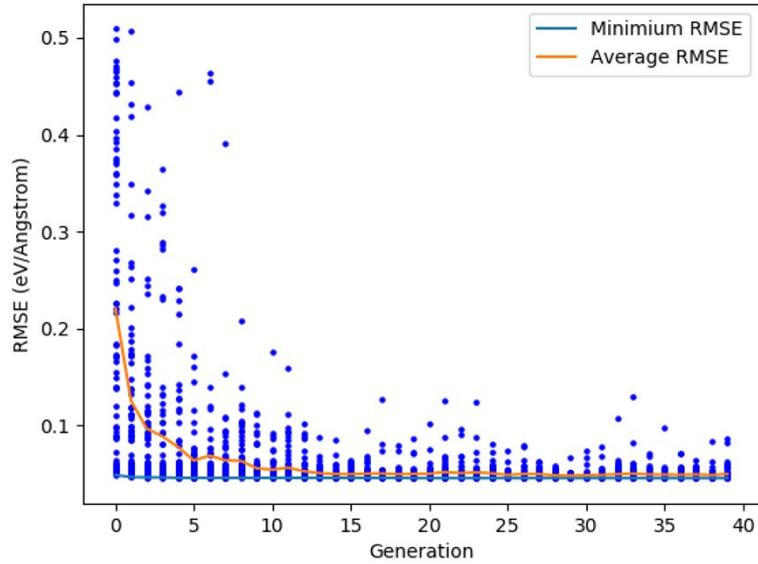


FIG. S3. The force prediction error δ_{RMS} of each individual along the genetic algorithm automatic selection protocol.

Section IV ML force field under rotation operation

When a structure going through certain rotation operation, for example rotating the structure along the z -axis for α degree, the position of each atom $\vec{R}_i = [x, y, z]^T$ changes to:

$$\vec{R}'_i = S \cdot \vec{R}_i \quad (\text{S4.1})$$

in which S is the rotation matrix:

$$S = \begin{bmatrix} \cos\alpha & -\sin\alpha & 0 \\ \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (\text{S4.2})$$

In addition, the force experienced by each atom $\vec{F}_i = [F_i^x, F_i^y, F_i^z]^T$ will also change to:

$$\vec{F}'_i = S \cdot \vec{F}_i \quad (\text{S4.3})$$

However, not all the ML force fields follow the equation S4.3. Next, we are going to prove that the linear model (LM) and neural network model 1 (NNM1) abide equation S4.3 but the neural network model 2 (NNM2) cannot.

First of all, a fingerprint matrix V_i , which contains n vectorial fingerprints of i^{th} atom along x , y and z directions, is introduced:

$$V_i = \begin{bmatrix} V_i^{1,x} & \dots & V_i^{n,x} \\ V_i^{1,y} & \dots & V_i^{n,y} \\ V_i^{1,z} & \dots & V_i^{n,z} \end{bmatrix} \quad (\text{S4.4})$$

According to the formula of radial and angular vectorial fingerprints (equation 3 and 4), we can easily demonstrate that the fingerprint matrix will change to:

$$V'_i = S \cdot V_i \quad (\text{S4.5})$$

after the rotation operation.

(I) Linear model

The linear model can be described as:

$$\vec{F}_i = V_i \cdot \vec{W} \quad (\text{S4.6})$$

in which W is the weight parameters of linear regression, $\vec{W} = [w_1 \dots w_n]^T$. When the structure going through the rotation operation, the force of i^{th} atom can be represented as:

$$\vec{F}'_i = V'_i \cdot \vec{W} = S \cdot V_i \cdot \vec{W} = S \cdot \vec{F}_i \quad (\text{S4.7})$$

(II) Neural network model 1

The neural network model type 1 utilized only linear activation functions and do not use bias for nodes. As a result, the model can be expressed as:

$$\vec{F}_i = V_i \cdot W^{01} \cdot W^{1,2} \dots W^{k-1,k} \quad (\text{S4.8})$$

in which $W^{k-1,k}$ is the matrix of weight parameters between the $(k-1)^{\text{th}}$ and k^{th} layers. Then the force after rotation changes to:

$$\vec{F}'_i = V'_i \cdot W^{01} \cdot W^{1,2} \dots W^{k-1,k} = S \cdot V_i \cdot W^{01} \cdot W^{1,2} \dots W^{k-1,k} = S \cdot \vec{F}_i \quad (\text{S4.9})$$

(III) Neural network model 2

The neural network model type 2 utilized hypertangent activation, and the model can be expressed as:

$$\vec{F}_i = \tanh\left(\tanh\left(\tanh\left(V_i \cdot W^{01}\right) \cdot W^{1,2}\right) \dots W^{k-1,k}\right) \quad (\text{S4.10})$$

Because of the employ of the non-linear activation, the equation S4.10 cannot deduce the equation S4.3.

Section V ML force field for SiO₂ with Coulomb interaction

The charge-explicit ML force field contains two linear regression models that predict the short- and long-range force respectively.

First of all, we should split the QM forces into short- and long-range parts in a reasonable way, since they are not directly accessible from the DFT calculation. A simple way is that we define the long-range part as the Coulomb force. The short-range part is total QM force minus the Coulomb force. The Coulomb force can be calculated with Coulomb law or Ewald summation based on the Bader charge analysis. However, such splitting method would increase the corrugation of short-range force (as mentioned in Ref¹), and increase the difficulty of fitting. As a result, a screening function is introduced, and the long-range force is defined as:

$$F_{long}^{\vec{}} = \sum_i \sum_{j \neq i} f_{screen}(R_{ij}) \times \frac{q_i q_j}{R_{ij}^3} \cdot \vec{R}_{ij} \quad (S5.1)$$

while the screening function f_{screen} is:

$$f_{screen}(R_{ij}) = \begin{cases} \frac{1}{2} \left[1 - \cos\left(\frac{\pi R_{ij}}{R_c}\right) \right] & \text{for } R_{ij} \leq R_c \\ 1 & \text{for } R_{ij} > R_c \end{cases} \quad (S5.2)$$

As a consequence, the short-range force is:

$$F_{short} = F_{QM} - F_{long} \quad (S5.3)$$

Then, the ML model for short-range force is constructed in the same way as the ordinary ML force field, *i.e.* using vectorial fingerprints as inputs and short-range forces as targets. In addition, another ML model is built to predict the environment-dependent atomic charges based on the chemical environment. The charge prediction model utilizes the Belher-Parrinello symmetry functions as the structural fingerprints and Bader charges as the target properties. The predicted long-range force is evaluated from equation S5.1 with the predicted atomic charge.

A part of the SiO₂ database (amorphous structure) was used to test the above method. As shown in Fig. S4, the δ_{RMS} of the atomic charge is 0.023e (O, 0.020; Si, 0.029). As consequence, the δ_{RMS} of the long-range force is 0.0042 eV/Å (O, 0.003; Si 0.007). On the other hand, the δ_{RMS} of short-range force is 0.146 eV/Å (O, 0.113; Si 0.213). The total force, as sum of short- and long-range forces, has the δ_{RMS} of 0.147 eV/Å (O, 0.113; Si 0.214). We found that the long-range forces that outside cutoff distance can be predicted quite accurately in this method, however, the short-range force prediction has been the main error source. As a comparison, the linear regression force field was constructed with the same set of reference data, and its δ_{RMS} is 0.150 eV/Å (O, 0.115; Si 0.218). Surprisingly, we found that ML model with coulomb interaction does not obviously accurate than that without coulomb interaction. It is also reported that the explicit consideration of charges does not improve the overall accuracy in the construction of high-dimensional NN potential for zinc oxide.

Furthermore, we calculated the Coulomb force inside 8 Å cutoff distances and compare it with the QM forces. The result is given in Fig. S5. We can not find the clear relationship between Coulomb forces and QM forces, which implies that the QM effects plays very important role in the force calculation, and only the accurate electrostatic model is not enough for the construction of multicomponent force fields.

In addition, to clarify the cutoff distance for the short-distance interaction, we test a serial of cutoff distance R_c varies from 3 to 8 Å using the same reference database, and the RMSEs of models (both with and without explicit Coulomb interaction) are listed in table S3. When R_c is 7 and 8 Å, explicit Coulomb force prediction do improve the accuracy, in compare with the ordinary force field. For $R_c = 7$ Å, the improvement is more significant. However, the charge-explicit FF performs worse than the ordinary one when $R_c < 7$ Å. To explain it, we plot the variance of one QM force component along a short MD trajectory, as well as the short-range force. As we can see, the corrugation of short force is increased substantially when the R_c is decreased from 8 to 3 Å, which certainly increased the difficulty of model fitting.

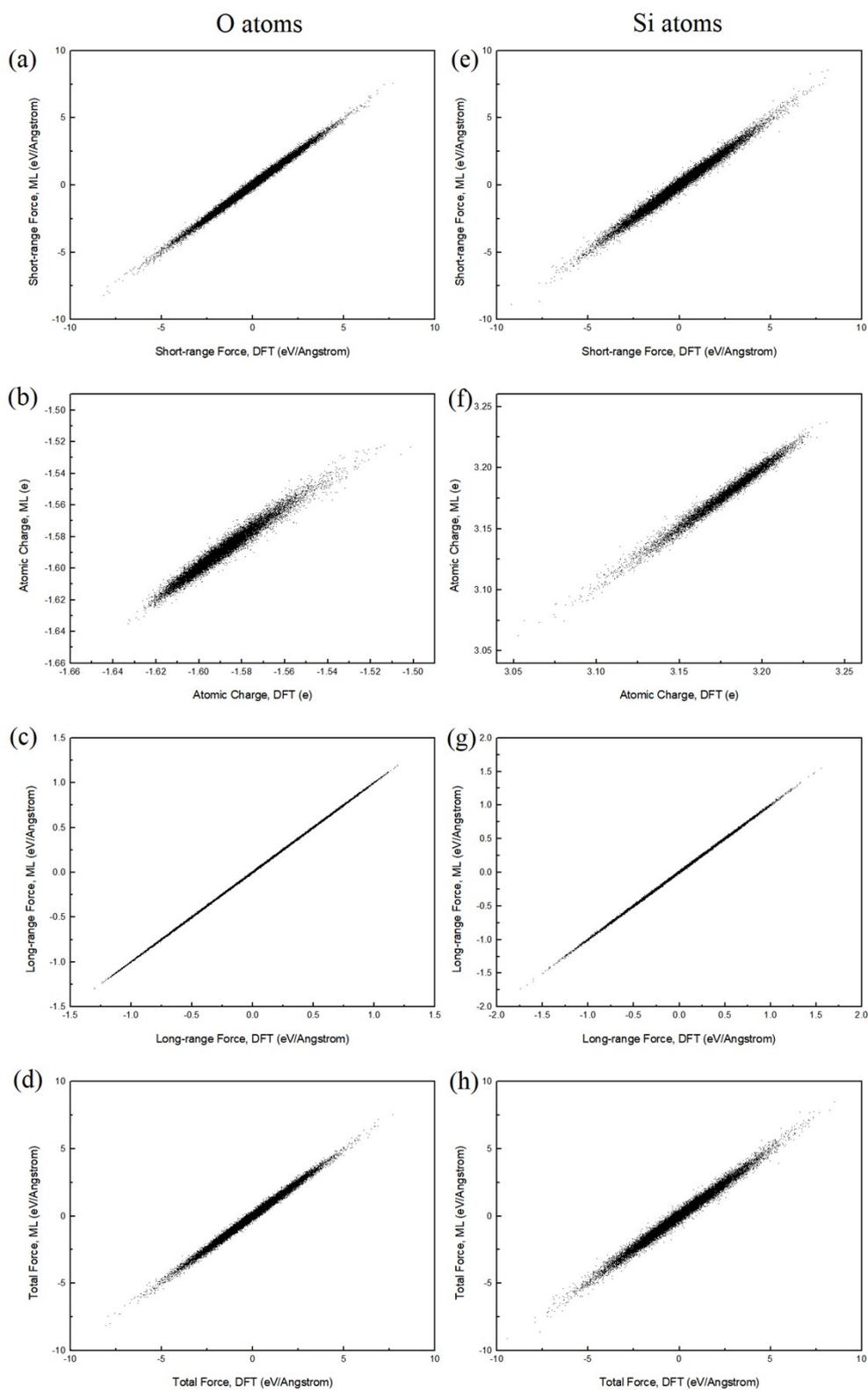


FIG. S4 The comparison of (a) short-range forces, (b) atomic charges, (c) long-range forces and (d) total forces of O atoms predicted with charge-explicit ML force field with the reference DFT results. The (e) – (h) show the corresponding plots for Si atoms.

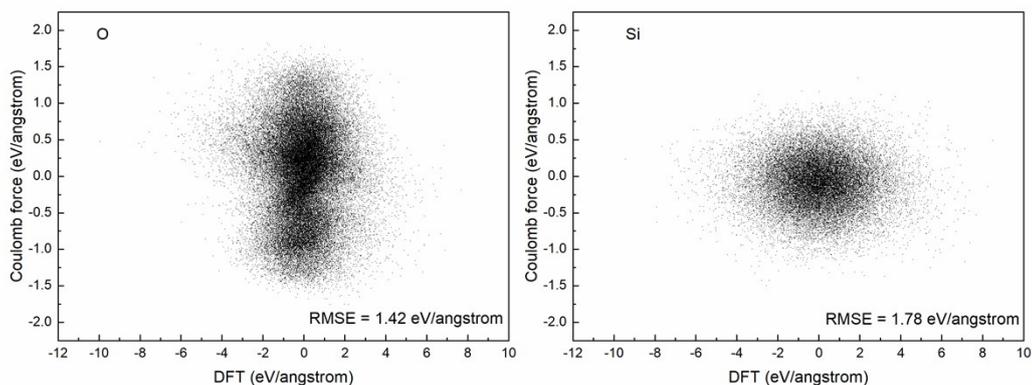


FIG S5. The comparison of Coulomb force calculated with the ML predicted charges and the DFT forces.

Table S3 The effect of cutoff distance on the charge-explicit ML force field.

Cutoff distance	Without Coulomb	With Coulomb		
	Total force	Total force	Short	Long
8.0	0.150	0.147	0.146	0.0042
7.0	0.157	0.150	0.149	0.0063
6.0	0.166	0.178	0.177	0.0091
5.0	0.178	0.223	0.222	0.0139
4.0	0.206	0.392	0.388	0.0223
3.0	0.293	1.083	1.066	0.0565

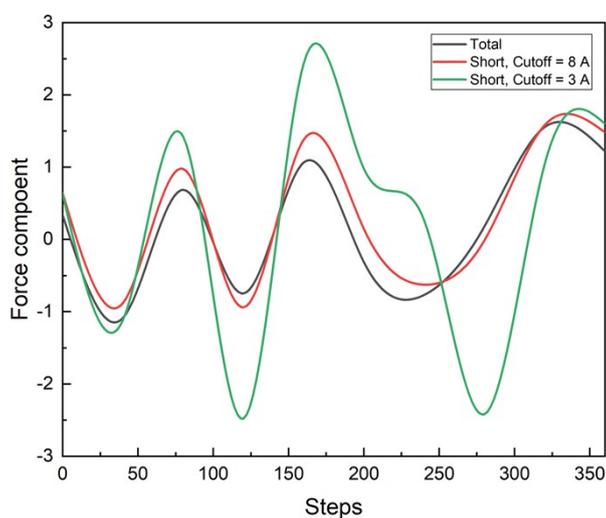


FIG. S6 The variation of one force component along a short MD trajectory. The QM force and the short-range forces inside 3 and 8 Å cutoff distances are shown. The corrugation of short force variation clearly increased when shorter cutoff distance applied.

Section VI Energy variation along NVT and NVE-MD

Additional test of force field is performed by utilizing it in the MD with NVE and NVT ensemble. In NVT-MD, the Nose-Hoover thermostat is employed to control the temperature. During the MD run, the evolution of energies, including kinetic and potential energies, are computed. The evolution of potential energy is calculated with:

$$E_{pot}(t + \Delta t) = E_{pot} + \Delta t \times \sum_i \sum_{\alpha} F_i^{\alpha} \cdot v_i^{\alpha} \quad (\text{S6.1})$$

in which, F_i^{α} and v_i^{α} are the force and velocity components of i^{th} atom along α ($\alpha=x, y$ or z) direction.

Figure S5 shows the variation of energies and temperature along the MD simulations. The MD are performed with both LM and LMM force fields. The fcc-Cu and quartz-SiO₂ are used as the initial structures. The initial temperatures of NVE-MDs are set as 500 K. From the Fig S5(a), we found that the total energies (sum of kinetic and potential energy) are basically constants along the MD runs. The insert figure gives the enlarged view of total energy variation. The fluctuation is very small, and basically within $\pm 1 \times 10^{-5}$ eV/atom. The Fig S5(b) shows the temperature variation along the NVT-MD with the Nose-Hoover thermostat. The temperature of system fluctuates around the setting temperature (500K) after the initial equilibration, which demonstrates that the thermostat is also applicable for the ML force field.

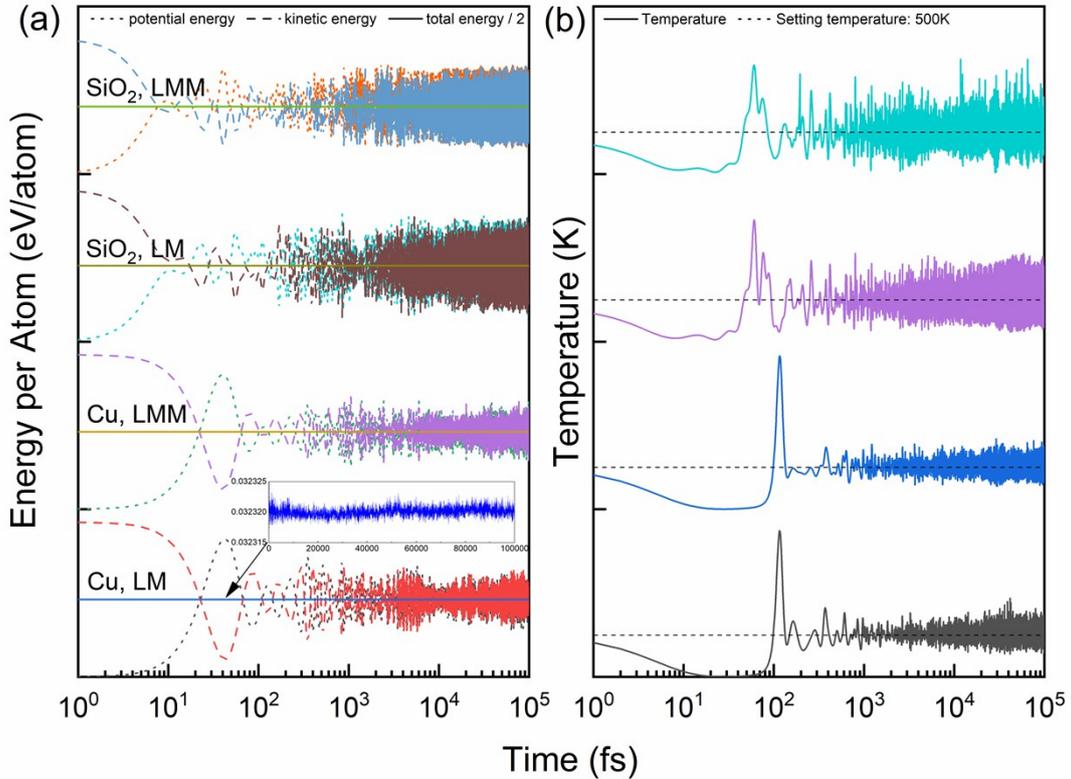


FIG. S7. (a) Potential energy, kinetic energy and total energy obtained for NVE MD. The insert gives the enlarged view of total energy evolution. (b) Temperature fluctuation of system along the NVT MD with Nose-Hoover thermostat.

Section VII. Comparison of ML force field and high-dimensional NN potential

The high-dimensional NN potentials (HD-NNP) are constructed used the same reference database as the ML force fields. The symmetry functions were chosen as the structural descriptors for NN potential. Same with the fingerprints, the cutoff distance r_c was set to 6.5 Å for Cu, and 8 Å for SiO₂. For convenience, we adopted the parameters of fingerprint set (8) of Cu, and parameters of set (5) of SiO₂. The detailed method the high-dimensional neural network construction can be seen in Ref. 2.

The NN potential for Cu has an architecture of 32-20-20-20-1, since the reduction of error is only margin if further increase the number of nodes. The δ_{RMS} of energy is 3.7 meV/atom for the training set and 4.0 meV/atom for the testing set. The $\delta_{RMS,F}$ is 0.11 eV/Å. Such accuracy is comparable with the NN potential of Cu built in Ref. (3), whose δ_{RMS} of energy and force is 3.9 meV/atom and 0.079 eV/Å, respectively. It is obvious that the ML force fields have higher force prediction accuracy than NN potential, in the case of Cu.

The architecture of 256-5-5-5-1 is adopted for SiO₂ NN potential, and its $\delta_{RMS,E}$ is 4.8 meV/atom for the training set and 5.0 meV/atom for the testing set. The $\delta_{RMS,F}$ is 0.23 eV/Å which is close to the linear regression force fields, but still not as accurate as the mixture model force fields.

The NN potentials were used in the various atomic simulation, as we described in the Section V. The results of RDF, ADF and barrier energy are also shown in the Fig. S6-8. As we can see, the results of NN potentials also have excellent agreement with that of DFT. However, the ML force fields, especially, the mixture model one, comparable or better than the NN potential across all regions. It can be attribution to the higher force prediction accuracy of ML force fields.

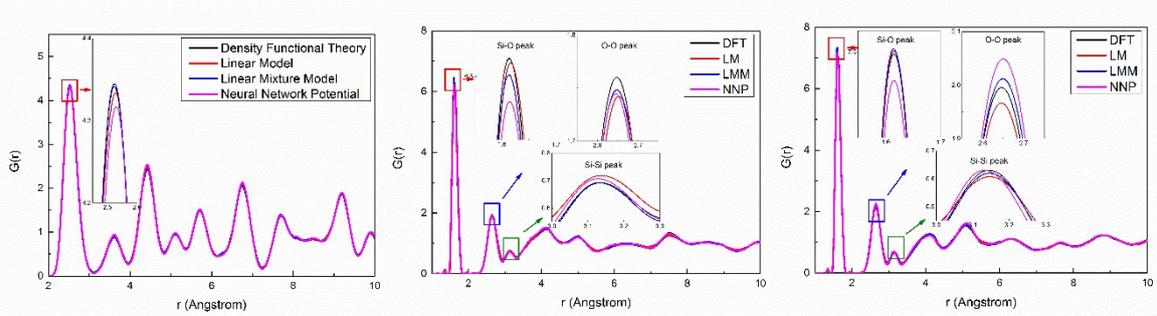


FIG. S8. The radial distribution function averaged over the 2.5 ps MD simulations based on density functional theory, linear regression / mixture model force fields and neural network potential. (a) fcc-Cu; (b) quartz-SiO₂; (c) cristobalite-SiO₂.

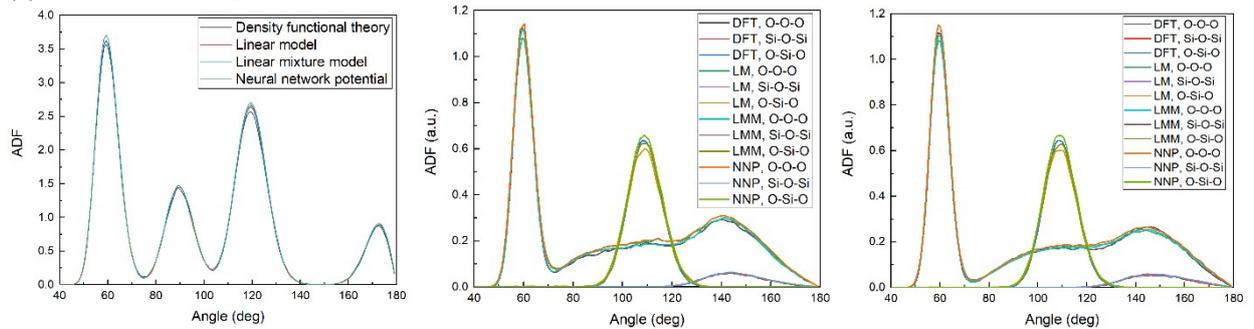


FIG. S9. The angular distribution function averaged over the 2.5 ps MD simulations based on density functional theory, linear regression / mixture model force fields and neural network potential. (a) fcc-Cu; (b) quartz-SiO₂; (c) cristobalite-SiO₂. The angles are determined with the maximum bond length of 3.0 Å.

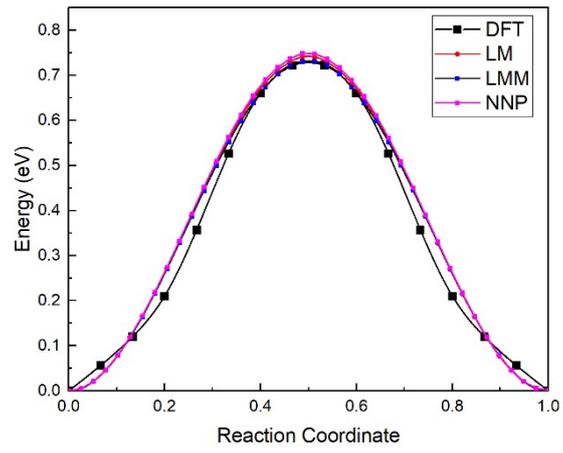


FIG. S10. The potential energy profile along the vacancy migration pathway in bulk Cu, which obtained in NEB calculation.

Section VIII Gaussian mixture model groups of Cu fingerprints

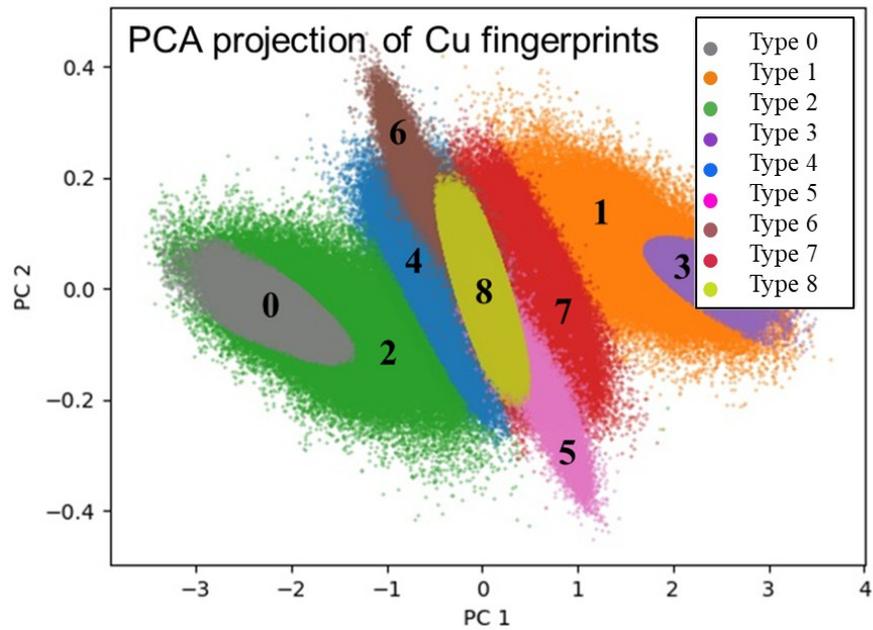


FIG S11. Structural fingerprints of Cu atoms in principal component space.

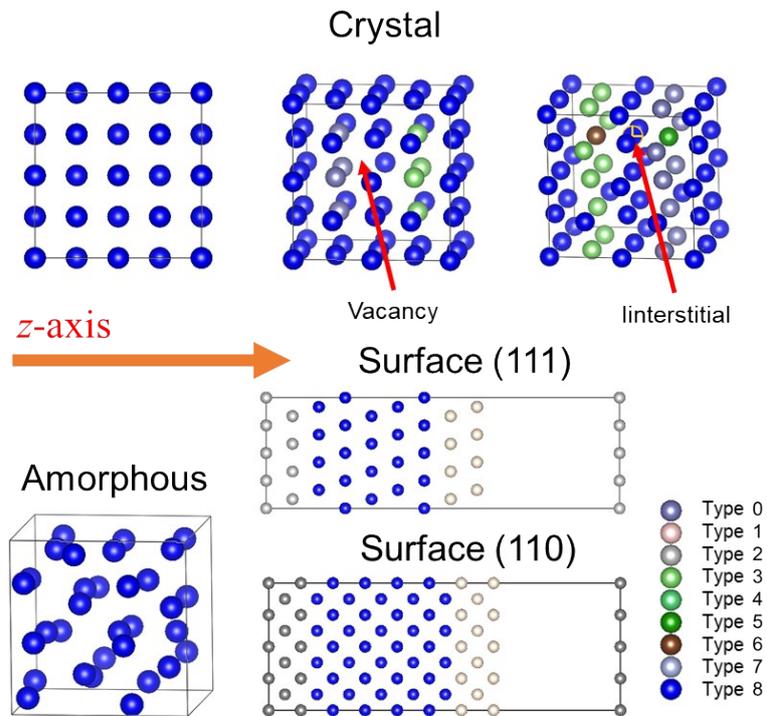


FIG S10. The local structure of 9 types of Si. The type 0 and 3 located around vacancy; type 4, 5, 6 and 7 located around interstitial; type 2 and 3 located on the surface; and type 8 is bulk atom.

Section IX Parameters of structural fingerprints

Table S4. The parameters of structural fingerprints of Cu, set 1

	type	η	R_s
1	radial	0.001	0.0
2	radial	0.03	0.0
3	radial	0.07	0.0
4	radial	0.14	0.0
5	radial	0.25	0.0
6	radial	0.40	0.0
7	radial	0.70	0.0
8	radial	1.40	0.0
9	radial	1.40	2.0
10	radial	1.40	2.5
11	radial	1.40	3.0
12	radial	1.40	3.5
13	radial	1.40	4.0
14	radial	1.40	4.5
15	radial	1.40	5.0
16	radial	1.40	5.5

Table S5. The parameters of structural fingerprints of Cu, set 2

	type	η	R_s	ζ	θ_s
1	radial	0.001	0.0		
2	radial	0.03	0.0		
3	radial	0.07	0.0		
4	radial	0.14	0.0		
5	radial	0.25	0.0		
6	radial	0.40	0.0		
7	radial	0.70	0.0		
8	radial	1.40	0.0		
9	angular	0.0003	0.0	1	0
10	angular	0.0003	0.0	1	$\pi/2$
11	angular	0.0003	0.0	1	π
12	angular	0.0003	0.0	1	$3\pi/2$
13	angular	0.0003	0.0	4	0
14	angular	0.0003	0.0	4	$\pi/2$
15	angular	0.0003	0.0	4	π
16	angular	0.0003	0.0	4	$3\pi/2$

Table S7. The parameters of structural fingerprints of Cu, set 3

type																
1~119	radial	η :	0.001	0.015	0.03	0.05	0.07	0.10	0.14	0.18	0.25	0.30	0.40	0.50	0.70	1.00
			1.40	2.00	3.50											
		R_s :	0	1.0	2.0	2.5	3.0	4.0	5.0							

Table S8. The parameters of structural fingerprints of Cu, set 4

type																
1~51	radial	η :	0.001	0.015	0.03	0.05	0.07	0.10	0.14	0.18	0.25	0.30	0.40	0.50	0.70	1.00
			1.40	2.00	3.50											
		R_s :	0	2.5	5.0											
52~117	angular	η :	0.001	0.015	0.03	0.05	0.07	0.10	0.14	0.18	0.25	0.30	0.40	0.50	0.70	1.00

	1.40	2.00	3.50
R_s :	0.0		
ζ :	0.0	0.5	1.0 1.5
θ_s :	1	2	4

Table S9. The parameters of structural fingerprints of Cu, set 5

	type	η	R_s
1	radial	0.5	1.0
2	radial	0.7	1.0
3	radial	1.0	1.0
4	radial	1.4	1.0
5	radial	0.14	2.0
6	radial	3.5	2.0
7	radial	1.4	2.5
8	radial	2.0	2.5
9	radial	3.5	3.0
10	radial	0.1	5.0
11	radial	0.14	5.0
12	radial	0.18	5.0
13	radial	1.0	5.0
14	radial	1.4	5.0
15	radial	2.0	5.0

Table S10. The parameters of structural fingerprints of Cu, set 6

	type	η	R_s
1	radial	0.7	0.0
2	radial	0.4	1.0
3	radial	3.5	1.0
4	radial	0.7	2.0
5	radial	2.0	2.0
6	radial	0.03	2.5
7	radial	0.05	4.0
8	radial	0.001	5.

Table S11. The parameters of structural fingerprints of Cu, set 6

	type	η	R_s
1	radial	0.7	0.0
2	radial	0.4	1.0
3	radial	3.5	1.0
4	radial	0.7	2.0
5	radial	2.0	2.0
6	radial	0.03	2.5
7	radial	0.05	4.0
8	radial	0.001	5.

Table S12. The parameters of structural fingerprints of Cu, set 7

	type	η	R_s	ζ	θ_s		type	η	R_s	ζ	θ_s
1	radial	0.07	0			44	angular	0.03	0	1	1
2	radial	0.1	0			45	angular	0.03	0	1	2
3	radial	0.14	0			46	angular	0.03	0	1	4
4	radial	0.18	0			47	angular	0.03	0	1.5	1
5	radial	0.7	0			48	angular	0.03	0	1.5	2

13	angular	0.001	0	0.5	4
14	angular	0.001	0	1	2
15	angular	0.001	0	1.5	2
16	angular	0.03	0	0	1
17	angular	0.03	0	0	4
18	angular	0.03	0	1	1
19	angular	0.03	0	1	2
20	angular	0.05	0	1.5	2
21	angular	0.05	0	1.5	4
22	angular	0.09	0	0	1
23	angular	0.09	0	0	2
24	angular	0.09	0	0.5	4
25	angular	0.09	0	1.5	1
26	angular	0.16	0	0	4
27	angular	0.16	0	1	4
28	angular	0.16	0	1.5	1
29	angular	0.16	0	1.5	4
30	angular	0.3	0	0	4
31	angular	0.3	0	1.5	2
32	angular	0.3	0	1.5	4

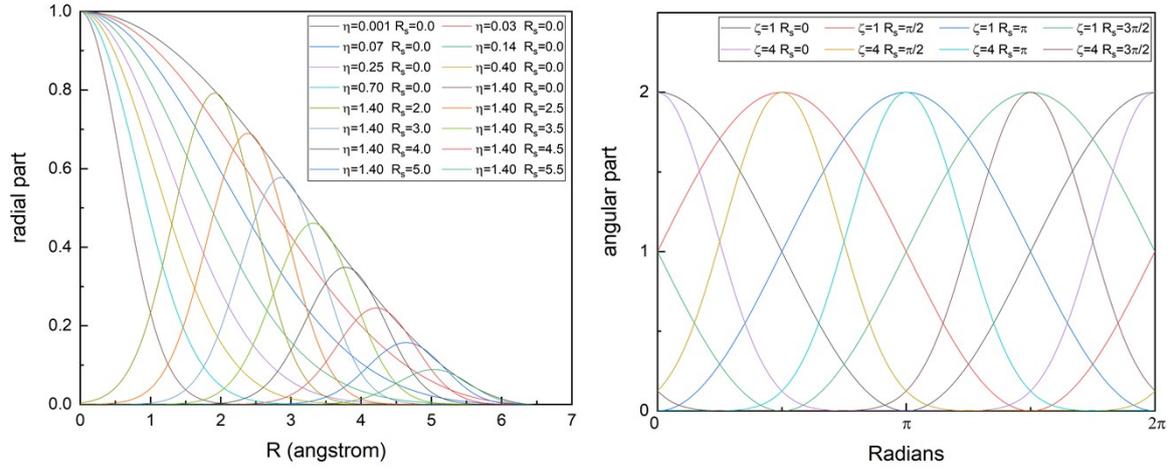


FIG S11. Examples of the fingerprints with different parameter sets. (a) radial part of radial fingerprints, $\exp[-\eta(r-R_s)^2]f_c(r)$, (b) the angular contributions of angular fingerprints $2^{1-\zeta}(1+\cos(\theta-\theta_s))^\zeta$.

Reference

- ¹ T. Morawietz, V. Sharma, and J. Behler, *J. Chem. Phys.* **136**, 064103 (2012).
- ² W. Li, Y. Ando, E. Minamitani, and S. Watanabe, *J. Chem. Phys.* **147**, 214106 (2017).
- ³ N. Artrith and J. Behler, *Phys. Rev. B* **85**, 045439 (2012).