Electronic Supplementary Information

Accelerating Atomistic Simulations with Piecewise Machine Learned Ab Initio Potentials at Classical Force Field-like Cost

Yaolong Zhang, Ce Hu, and Bin Jiang*

Hefei National Laboratory for Physical Science at the Microscale, Key Laboratory of Surface and Interface Chemistry and Energy Catalysis of Anhui Higher Education Institutes, Department of Chemical Physics, University of Science and Technology of China, Hefei, Anhui 230026, China

*: corresponding author: bjiangch@ustc.edu.cn

I. Details on training neural networks

In the embedded atom neural network (EANN) approach[1] and its piecewise version (PEANN), the weights and biases of atomic NNs along with the atomic expansion coefficients were determined by minimizing the cost function defined by root-mean-errors between ab initio potential energies and atomic forces with respect to Cartesian coordinates and corresponding NN outputs[2],

$$S(\mathbf{w}) = \sum_{i=1}^{N_{data}} [(E_i^{NN} - E_i^{Ref})^2 + \eta \left| \vec{\mathbf{F}}_i^{NN} - \vec{\mathbf{F}}_i^{Ref} \right|^2] / N_{data} . \quad \text{(S1)}$$

Here, $\mathbf{w}$ is a collection of all adjustable parameters, N_{data} is the number of configurations in the training set. E_i^{NN}, E_i^{Ref}, $\vec{\mathbf{F}}_i^{NN}$ and $\vec{\mathbf{F}}_i^{Ref}$ are potential energies and atomic force vectors of *i*th configuration obtained by NN and reference ab initio calculations, respectively. Note that each E_i^{NN} and $\vec{\mathbf{F}}_i^{NN}$ are the sum of atomic NN outputs and for the atomic NN parameters are identical for same element. An efficient hybrid extreme machine learning Levenberg-Marquardt (ELM-LM) algorithm was employed to optimize these adjustable parameters[3].

For all systems discussed in this work, the NN structures consist of two hidden layers. To make a fair comparison, the number of neurons in each hidden layer, the number of descriptors, as well as the cutoff radius (r_c) were all kept the same for in EANN and PEANN potentials of each system. Table S1 gives such information for the four benchmark condensed phase systems, namely Cu, Ge, Mo, and water. Also compared in Table S1 are the computational costs of evaluating individually the inter-nuclear distances within r_c, the density-like structural descriptors, and NNs (plus the rest minor contributions), respectively. Other hyperparameters to determine the density-

like descriptors are listed in Tables S1-S5 for PEANN and Tables S6-S9 for EANN, for Cu, Ge, Mo, and water in sequence. Note that for the water system, we have applied the CUR matrix decomposition algorithm[4] to select the optimal descriptors that best represent the training set, as used by Ceriotti and coworkers to optimize the selection of Behler-Parrinello type atom centered symmetry functions[5].

II. Molecular dynamics simulations

To demonstrate the comparable performance of EANN and PEANN as other machine learning models, as done in Ref. 12, we have predicted material properties such as phonon dispersion curves, as well as energies and forces at unseen structures, taking Mo as an example (this is already the system with largest errors). To this end, in Fig. S1, we compare the DFT calculated phonon spectrum for 3×3×3 bulk Mo with the results predicted by all machine learning models discussed in the main text. In addition, following Zuo *et al.*[12], we have performed 250 ps NVT classical molecular dynamics (MD) simulations for a 3×3×3 bulk Mo with a 0.1 fs time step at 1300K maintained with Andersen thermostat, and extracted 40 snapshots with a time interval of 2.5 ps. It should be noted that none of the training and testing data were obtained from such high temperature MD simulations and the sampled configurations could be therefore very different and unknown, serving as good candidates for testing the extrapolability of machine learning potentials. The energy and force error distributions of our EANN and PEANN model are compared in Fig S2 with other machine learning models whose results are extracted from Ref. 12.

To validate the accuracy of our PEANN and EANN potentials for liquid water, we compare the O-O, O-H and H-H radial distribution functions (RDFs) of liquid water obtained by the classical molecular dynamics (MD) simulations using the TIP4P model in Ref. [6], Behler-Parrinello NN (BPNN) potential in Ref. 7, EANN and PEANN potentials in this work, respectively. MD simulations have been performed with 64 water molecules in a cubic box with its side length of 12.42 Å at temperature of 300 K. A total of 20 ps NVT MD simulations with a time step of 0.2 fs. The Andersen thermostat[8] was used for keeping the temperature in the simulations of the PEANN/EANN potentials in a modified VENUS code[9], while the Nose-Hoover Chains algorithm[10] was used for the TIP4P and BPNN models implemented with LAMMPS[11], respectively. It is found that the TIP4P force field requires a longer cutoff radius and long-range corrections to yield the correct description of RDFs as shown in Fig. 4, and our setup thus follows the original publication[6] where the short-range interactions were truncated at 7.75 Å.

References

1. Y. Zhang, C. Hu and B. Jiang, *J. Phys. Chem. Lett.*, 2019, **10**, 4962-4967.
2. J. Behler, *Int. J. Quant. Chem.*, 2015, **115**, 1032-1050.
3. Y.-l. Zhang, X.-y. Zhou and B. Jiang, *Chin. J. Chem. Phys.*, 2017, **30**, 727-734.
4. M. W. Mahoney and P. Drineas, *Proc. Nat. Acad. Sci. U.S.A.*, 2009, **106**, 697-702.
5. G. Imbalzano, A. Anelli, D. Giofré, S. Klees, J. Behler and M. Ceriotti, *J. Chem. Phys.*, 2018, **148**, 241730.
6. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926-935.
7. B. Cheng, E. A. Engel, J. Behler, C. Dellago and M. Ceriotti, *Proc. Nat. Acad. Sci. U.S.A.*, 2019, **116**, 1110-1115.
8. H. C. Andersen, *J. Chem. Phys.*, 1980, **72**, 2384-2393.
9. X. Hu, W. L. Hase and T. Pirraglia, *J. Comput. Chem.*, 1991, **12**, 1014-1024
10. G. J. Martyna, D. J. Tobias and M. L. Klein, *J. Chem. Phys.*, 1994, **101**, 4177-4189.
11. S. Plimpton, *J. Comput. Phys.*, 1995, **117**, 1-19.
12. Y. Zuo, C. Chen, X. Li, Z. Deng, Y. Chen, J. Behler, G. Csányi, A. V. Shapeev, A. P. Thompson, M. A. Wood and S. P. Ong, *J. Phys. Chem. A*, 2020, **124**, 731-745.

Table S1: NN structures denoted by the number of neurons in the input (descriptors), hidden, and output layers, the cutoff radii, as well as individual computational costs (μs/atom/CPU-core/step) of evaluating individually the inter-nuclear distances within r_c, the density-like structural descriptors, and NNs (plus the rest minor contributions), respectively. NN structures and cutoff radii are identical for EANN and PEANN.

System	NN structure	r_c (Å)	Individual costs: Distances/Descriptors/Others	
			PEANN	EANN
Cu	10×10×10×1	4.1	1.2/1.9/1.8	1.2/3.5/1.8
Ge	15×15×15×1	5.0	1.2/3.8/2.6	1.3/7.8/2.5
Mo	16×16×16×1	5.0	1.7/4.2/2.8	1.6/7.9/2.5
H_2O	33×20×20×1	6.3	3.3/23.9/6.5	3.3/66.3/8.0

Table S2: Hyperparameters of the piecewise descriptors of the PEANN Cu potential.

Numbering	L_{max}	r_{in} (Å)	r_{out} (Å)	α
1	0/1	0.50	1.70	12.83
2	0/1	1.10	2.30	6.04
3	0/1	1.70	2.90	2.85
4	0/1	2.30	3.50	1.35
5	0/1	2.90	4.10	0.63

Table S3: Hyperparameters of the piecewise descriptors of the PEANN Ge potential.

Numbering	L_{max}	r_{in} (Å)	r_{out} (Å)	α
1	0/1/2	1.00	2.09	9.35
2	0/1/2	1.73	2.82	3.76
3	0/1/2	2.45	3.54	1.51
4	0/1/2	3.18	4.27	0.61
5	0/1/2	3.91	5.00	0.24

Table S4: Hyperparameters of the piecewise descriptors of the PEANN Mo potential.

Numbering	L_{max}	r_{in} (Å)	r_{out} (Å)	α
1	0/1	1.50	2.00	1.60
2	0/1	0.43	2.43	7.49
3	0/1	0.86	2.86	3.66
4	0/1	1.29	3.29	0.62
5	0/1	1.71	3.71	0.19
6	0/1	3.04	4.14	0.02
7	0/1	3.17	4.57	0.02
8	0/1	3.60	5.00	0.01

Table S5: Hyperparameters of the piecewise descriptors of the PEANN bulk water potential.

Numbering	Central atom	L_{max}	r_{in} (Å)	r_{out} (Å)	α
1	O	0/1/2	-0.10	1.10	12.97
2	O	0/1/2	0.43	1.53	4.45
3	O	0/1/2	0.06	2.06	1.92
4	O	0/1/2	0.89	2.59	2.97
5	O	0/1/2	1.72	3.12	1.28×10^{-1}
6	O	0/1/2	1.95	3.65	1.13×10^{-1}
7	O	0/1/2	3.78	4.18	2.96×10^{-2}
8	O	0/1/2	3.31	4.71	1.42×10^{-2}
9	O	0/1/2	3.84	5.24	6.83×10^{-3}
10	O	0/1/2	4.37	5.77	3.27×10^{-3}
11	O	0/1/2	4.90	6.30	1.57×10^{-3}
1	H	0/1/2	-0.40	1.00	12.12
2	H	0/1/2	-0.47	1.57	3.84
3	H	0/1/2	0.06	2.06	1.84
4	H	0/1/2	0.59	2.59	1.16
5	H	0/1/2	1.72	3.12	1.29×10^{-1}
6	H	0/1/2	2.25	3.65	6.19×10^{-2}
7	H	0/1/2	2.78	4.18	2.97×10^{-2}
8	H	0/1/2	3.31	4.71	1.42×10^{-2}
9	H	0/1/2	3.84	5.24	6.83×10^{-3}
10	H	0/1/2	4.37	5.77	3.27×10^{-3}
11	H	0/1/2	4.90	6.30	1.57×10^{-3}

Table S6: Hyperparameters of the descriptors of the EANN Cu potential.

Numbering	L_{max}	r_s (Å)	α (Å^{-2})
1	0/1	0.00	0.22
2	0/1	0.95	0.22
3	0/1	1.90	0.22
4	0/1	2.85	0.22
5	0/1	3.80	0.22

Table S7: Hyperparameters of the descriptors of the EANN Ge potential.

Numbering	L_{max}	r_s(Å)	α (Å^{-2})
1	0/1/2	0.00	0.15
2	0/1/2	1.15	0.15
3	0/1/2	2.30	0.15
4	0/1/2	3.45	0.15
5	0/1/2	4.60	0.15

Table S8: Hyperparameters of the descriptors of the EANN Mo potential.

Numbering	L_{max}	r_s(Å)	α (Å^{-2})
1	0/1	0.00	0.43
2	0/1	0.68	0.43
3	0/1	1.36	0.43
4	0/1	2.04	0.43
5	0/1	2.72	0.43
6	0/1	3.40	0.43
7	0/1	4.08	0.43
8	0/1	4.76	0.43

Table S9: Hyperparameters of the descriptors of the EANN bulk water potential.

Numbering	Central atom	L_{max}	r_s(Å)	α (Å^{-2})
1	O	0/1/2	0.00	0.54
2	O	0/1/2	0.61	0.54
3	O	0/1/2	1.22	0.54
4	O	0/1/2	1.83	0.54
5	O	0/1/2	2.44	0.54
6	O	0/1/2	3.05	0.54
7	O	0/1/2	3.66	0.54
8	O	0/1/2	4.27	0.54
9	O	0/1/2	4.88	0.54
10	O	0/1/2	5.49	0.54
11	O	0/1/2	6.10	0.54
1	H	0/1/2	0.00	0.54
2	H	0/1/2	0.61	0.54
3	H	0/1/2	1.22	0.54
4	H	0/1/2	1.83	0.54
5	H	0/1/2	2.44	0.54
6	H	0/1/2	3.05	0.54
7	H	0/1/2	3.66	0.54
8	H	0/1/2	4.27	0.54
9	H	0/1/2	4.88	0.54
10	H	0/1/2	5.49	0.54
11	H	0/1/2	6.10	0.54

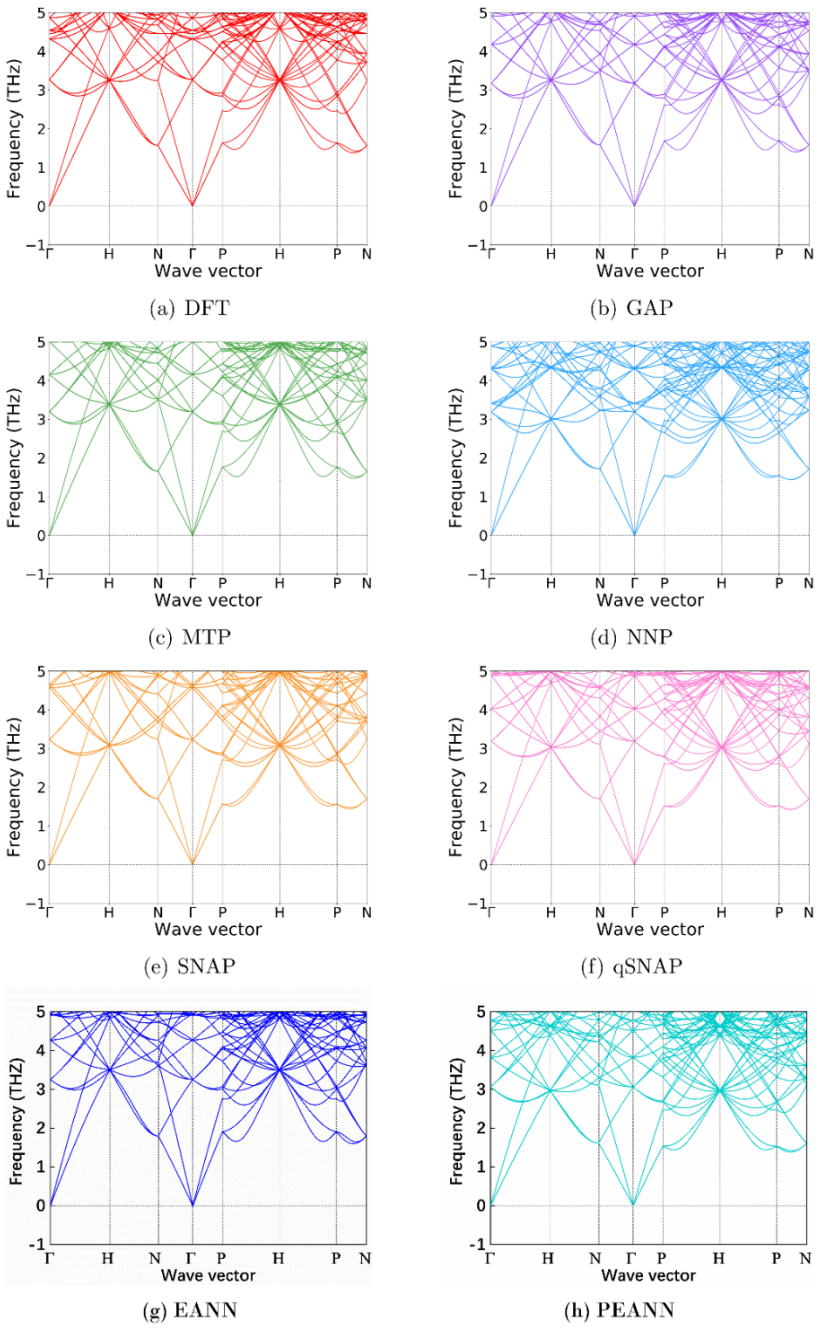

Fig. S1 Phonon spectra for 3×3×3 bulk Mo obtained by (a) DFT, (b) GAP, (c) MTP, (d) NNP, (e) SNAP, (f) QSNAP (g) EANN and (h) PEANN. Note: (a)~(f) were extracted from Ref. 12. The phonon spectra were plotted along the path through the Brillouin zone given by the high-symmetry points Γ-H-N-Γ-P-H-P-N.

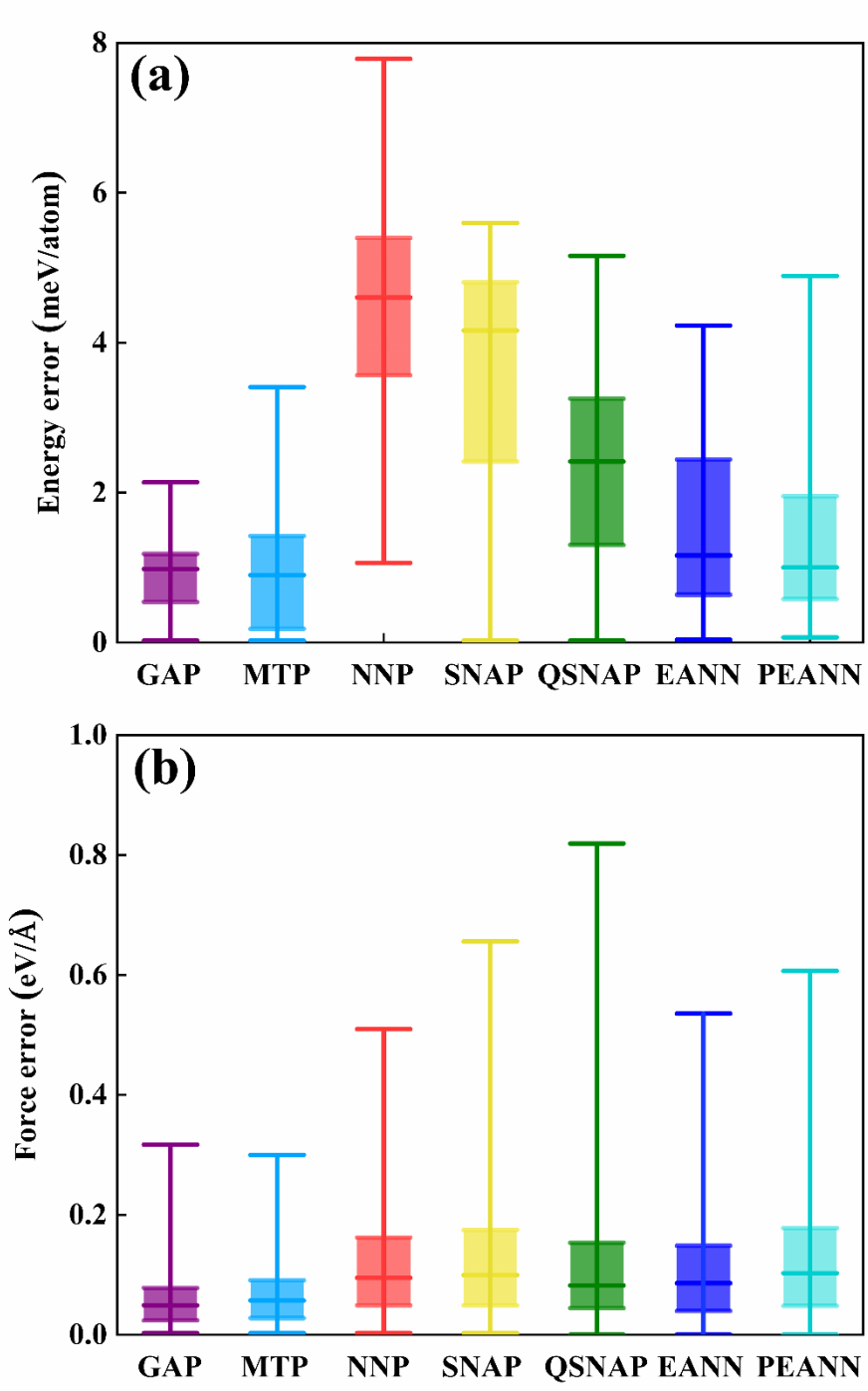

Fig. S2 Predicted error distributions of (a) energies and (b) atomic forces with 40 structures sampled from MD trajectories based on each ML model. The color filled areas are the interquartile range and the lines between them represent the median.

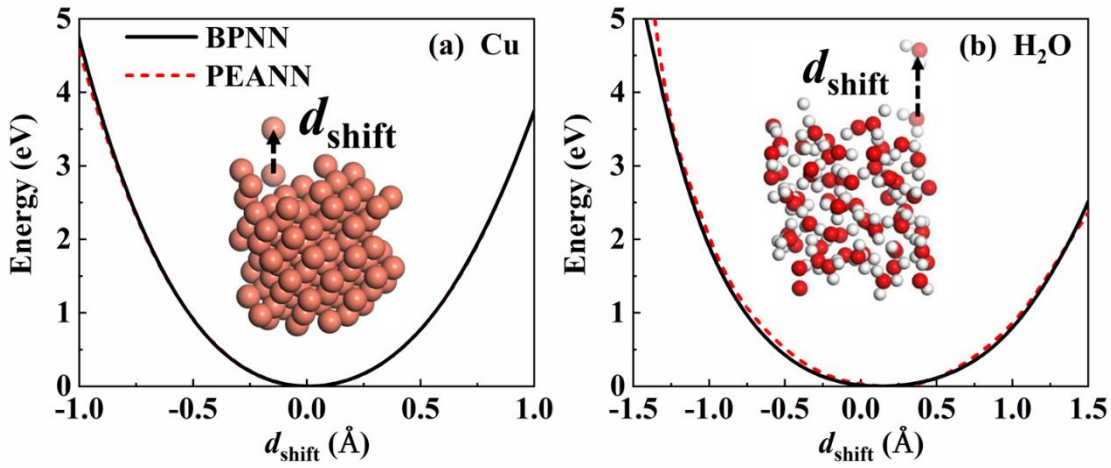

Fig. S3 Comparison of one-dimensional potential energy curves of BPNN and PEANN potentials as a function of the displacement of (a) a Cu atom and (b) a H_2O molecule relative to its original position in a random configuration of Cu and water bulk structures, respectively.