

Supporting Information for “**Constructing and evaluating machine-learned interatomic potentials for Li-based disordered rocksalts**”

Vijay Choyal, Nidhish Sagar, and Gopalakrishnan Sai Gautam\*

Department of Materials Engineering, Indian Institute of Science, Bengaluru 560012,  
Karnataka, India

\*E-mail: [saigautamg@iisc.ac.in](mailto:saigautamg@iisc.ac.in)

**Table S1.** Optimized hyperparameters for all machine-learned interatomic potentials (MLIPs) considered in this work. The definitions of all hyperparameters are provided in the Methods section of the main text. AENET, GAP, SNAP, and MTP indicate atomic energy network, gaussian approximation potential, spectral neighbor analysis potential, and moment tensor potential, respectively.

MLIP	Hyperparameters	
MTP	$R_{min}$	automatic (or default)
	$R_{cut}$	5 Å
	$N_Q$	8
	$lev_{max}$	16
	Weight on total energies	1
	Weight on atomic forces	0.01
	Weight on stress tensors	0.001
SNAP	$R_{cut}$	5 Å
	$J_{max}$	6
GAP	$n_{max}$	4
	$l_{max}$	4
	$\zeta$	4.0
	$\sigma_{atom}$	0.5
	$R_c$	5 Å
AENET	$R_c$ (Radial)	8.0 Å
	$R_c$ (Angular)	6.5 Å
	Nodes per hidden layer	15
	Number of hidden layers	2
	$N$	16
	Number of epochs	3300

## Description of MLIPs

MTP represents the total energy of a configuration as a sum of  $\epsilon_i$ , which in turn is written as a linear expansion of basis functions,  $B_\alpha^i$ , with the linear coefficients ( $\xi_\alpha$ ) determined by fitting to the training set. The atomic forces can be derived from MTP by taking spatial derivatives on the basis functions.

$$\epsilon_i = \sum_{\alpha} \xi_{\alpha} B_{\alpha}^i \quad (\text{S1})$$

The  $B_\alpha^i$  are written using moment tensor descriptors ( $M_{\mu,\nu}^i$ ), which consist of radial ( $f_\mu$ ) and angular (outer products of position vectors of neighboring atoms,  $j$ , resulting in a tensor of rank  $\nu$ ) components.<sup>1</sup>  $r_{ij}$  is the distance between  $i$  and neighbor  $j$ .

$$M_{\mu,\nu}^i = \sum_j f_\mu(r_{ij}) \underbrace{r_{ij} \otimes \cdots \otimes r_{ij}}_{\nu \text{ times}} \quad (\text{S2})$$

$f_\mu$  is expanded using a Chebyshev polynomial basis of order  $N_Q$  between a minimum ( $R_{min}$ ) and  $R_{cut}$  distance from  $i$  to ensure smooth behavior of the MTP. The radial component has a set of radial parameters ( $c_{\mu,i,j}$ ), which are fit to training data alongside  $\xi_\alpha$ . Once the  $M_{\mu,\nu}^i$  are defined, each  $B_\alpha^i$  is written as a “contracted” moment, where contraction of a moment can yield a scalar quantity from underlying vectors or tensors via dot or Frobenius products, respectively. Specifically, moments can be contracted to different “levels”, depending on  $\mu$  and  $\nu$ , with  $B_\alpha$  (and eventually  $\epsilon_i$ ) written up to a maximum level of contraction ( $lev_{max}$ ). Thus,  $N_Q$ ,  $R_{min}$ ,  $R_{cut}$ , and  $lev_{max}$  are crucial hyperparameters in the construction of MTP.

MTP is usually trained simultaneously on total energies, atomic forces, and stress tensors, with the relative weights given to these quantities during training being hyperparameters as well. The training algorithm that is typically employed by MTP is either Broyden-Fletcher-Goldfarb-Shanno (BFGS, employed in this work) or linear regression.

**SNAP** describes the local atomic density  $\rho_i(r)$  around an atom of interest,  $i$ , in terms of  $\delta$  functions, as written below.  $r_j$  is the position of a neighboring atom,  $j$ .

$$\rho_i(r) = \delta(r) + \sum_{r_j < R_{cut}} f_c(r_j) w_j \delta(r - r_j) \quad (\text{S3})$$

The density function is subsequently mapped on to three angular coordinates (instead of the usual two), and expanded via 4D hyperspherical harmonics. Specifically, the distance coordinate ( $r$ ) is written in terms of a third angular coordinate,  $\theta_0 = \frac{\theta_{max}r}{R_{cut}}$ , in addition to the typical angular coordinates of a spherical system,  $\theta$  and  $\phi$ , resulting in distinct points on a unit 3-sphere. Expansion using hyperspherical harmonics results in bispectrum components that are scalar triple products of the harmonic expansion coefficients, which in turn characterize the strength of density correlations. The hyperparameter  $J_{max}$  determines the number of bispectrum components ( $K$ ) to be used during the expansion. The  $K$  bispectrum components, written arbitrarily as  $B_1^i, B_2^i, \dots, B_K^i$ , in turn contribute to the energy of an individual atom ( $\epsilon_i$ ) as a linear summation, weighted by  $\beta_K$  coefficients.

$$\epsilon_i = \beta_0 + \sum_{k=1}^K \beta_k (B_k^i - B_{k0}^i) \quad (\text{S4})$$

$\beta_k B_{k0}^i$  is the contribution of each bispectrum component to the energy of an isolated  $i$  atom, which results in  $\beta_0$  being the energy of an isolated atom as calculated by SNAP. Atomic forces can be derived from **Equation S4** by obtaining spatial derivatives of  $B_K^i$ . The total energy of a given structure is written as a sum of atomic energies. Finally, the  $\beta_k$  values are determined via linear regression with respect to training total energies and atomic forces. In the case of qSNAP, the formalism is identical to SNAP except for additionally including the contributions from all pairwise products of bispectrum components in **Equation S4**. Hence, the number of hyperparameters are identical for SNAP and qSNAP.

**GAP** uses a kernel based on the smooth overlap of atomic positions (SOAP) to describe the atomic density in the neighborhood of the reference atom in combination with Gaussian process regression to evaluate atomic energies and forces. The  $\rho_i$  at a distance  $R$  from a central atom  $i$  is described using a squared exponential function in SOAP, where  $R_{ij}$  is the distance between a neighbor  $j$  and  $i$  within the  $R_c$ , and  $\sigma_{atom}$  can be considered to represent the size of atoms, which controls the smoothness of the SOAP kernel.

$$\rho_i(R) = \sum_j f_c(R_c) \exp\left(-\frac{|R - R_{ij}|^2}{2\sigma_{atom}^2}\right) \quad (\text{S5})$$

$\epsilon_i$  can be represented as a weighted ( $w(r)$ ) integral of the atomic density functions, where the weights follow a Gaussian distribution. The total energy of a structure, is in turn, the summation of the atomic energies. The atomic forces can be obtained via partial derivatives of the total energy.

$$\epsilon_i = \int w(r) \rho_i(r) dr \quad (\text{S6})$$

In practice, the SOAP kernel is written as a numerical expansion (**Equation S7**), typically involving equispaced Gaussians as the orthonormal radial basis functions ( $g_n(r)$ ), and spherical harmonics as the angular basis functions ( $Y_{lm}(\hat{r})$ ).  $c_{nlm}^i$  are the expansion coefficients of the atom  $i$ , which are determined using Gaussian process regression against a training dataset of energies and/or forces by evaluating their covariance matrix with other atoms  $j$  in the structure. The hyperparameters  $n_{max}, l_{max}, \zeta$  determine the maximum powers of the radial and angular components in the expansion basis, and the power factor to accentuate the sensitivity of the SOAP kernel, respectively.

$$\rho_i(r) = \sum_{nlm} c_{nlm}^i g_n(r) Y_{lm}(\hat{r}) \quad (\text{S7})$$

**AENET** employs a feed forward neural network architecture, where multiple artificial neurons are arranged in layers, with each neuron connected only to neurons in adjacent layers via activation functions. Similar to MTP, SNAP, and GAP, AENET also describes the energy of a given configuration as the sum of atomic energies, with atomic forces calculated using spatial derivatives of the total energy. Importantly, AENET fingerprints the local environment of each atom  $i$  using radial and angular distribution functions ( $RDF_i$  and  $ADF_i$ , see **Equation S8**), which are expanded using Chebyshev polynomials. Notably, the  $RDF_i$  and  $ADF_i$  do not increase the computational complexity of the ANN with number of species in the structure as was the case in the classical Behler-Parinello ANNs.

$$RDF_i(r) = \sum_{R_{ij} < R_{cut}} \delta(r - R_{ij}) f_c(R_{ij}) w_{t_j} \quad (\text{S8})$$

$$ADF_i(\theta) = \sum_{R_{ij}, R_{ik} < R_{cut}} \delta(\theta - \theta_{ijk}) f_c(R_{ij}) f_c(R_{ik}) w_{t_j} w_{t_k}$$

The weights,  $w_{t_j}$  and  $w_{t_k}$  are assigned values such that there is one unique set of Chebyshev expansion coefficients describing the structural features and another set of expansion coefficients that describe atom types. Subsequently, the  $RDF_i$  and  $ADF_i$  constitute the input node to the ANN, with the output node being the atomic energy of species  $i$ . The weights and biases within the ANN are obtained based on the training set of energies, via the limited memory-BFGS algorithm.<sup>2,3</sup> The important hyperparameters for training the AENET potential include the order of the Chebyshev basis expansion used ( $N$ ), number of hidden layers and nodes per layer in the neural network, the cut-off radii for  $RDF_i$  and  $ADF_i$ , and the number of epochs (i.e., number of training iterations) employed during training.

**Table S2.** Hubbard  $U$  values and calculated reference atomic energies.

Element	Li	O	Mn	Ni*	Sc	Ti	V	Cr	Fe	Co	Cu
<b>Hubbard <math>U</math> (eV)</b>	0	0	3.9	6.2	0	0	3.25	3.7	5.3	3.32	0
<b>Energy (eV)</b>	-0.3	-1.7	-4.6	-3.0	-2.1	-2.4	-2.7	-4.5	-2.2	-1.1	-0.2
*Ni atom energy obtained from materials machine learning package (MAML) <sup>4</sup>											

**Table S3.** Compilation of density functional theory (DFT) and AENET calculated voltages for various layered, single-TM, LiTMO<sub>2</sub>-TMO<sub>2</sub> compositions (TM = Ti, V, Cr, Mn, Fe, Co, Ni, or Cu). Absolute and percentage errors made by AENET for each electrode composition is indicated, as are the mean absolute and mean percentage errors.

Composition	Voltage (V vs. Li)		Absolute error (V)	Percentage error (%)
	DFT	AENET		
LiTiO <sub>2</sub> -TiO <sub>2</sub>	1.82	2.13	0.31	17.01
LiVO <sub>2</sub> -VO <sub>2</sub>	3.32	3.43	0.12	3.59
LiCrO <sub>2</sub> -CrO <sub>2</sub>	4.16	3.89	0.27	6.42
LiMnO <sub>2</sub> -MnO <sub>2</sub>	2.89	3.13	0.24	8.27
LiFeO <sub>2</sub> -FeO <sub>2</sub>	3.94	4.63	0.69	17.54
LiCoO <sub>2</sub> -CoO <sub>2</sub>	3.81	4.27	0.46	12.12
LiNiO <sub>2</sub> -NiO <sub>2</sub>	4.64	4.38	0.26	5.56
LiCuO <sub>2</sub> -CuO <sub>2</sub>	3.97	4.37	0.39	9.92
<b>Mean error</b>			0.34	10.05

**Table S4.** Root mean square error (RMSE) and mean absolute error (MAE) in energies and forces among the train and test datasets using neural equivariant interatomic potentials (NequIP). The overall dataset size used is 2000, with the computational training time also indicated. The training was done on the same resources used for training the other MLIPs.

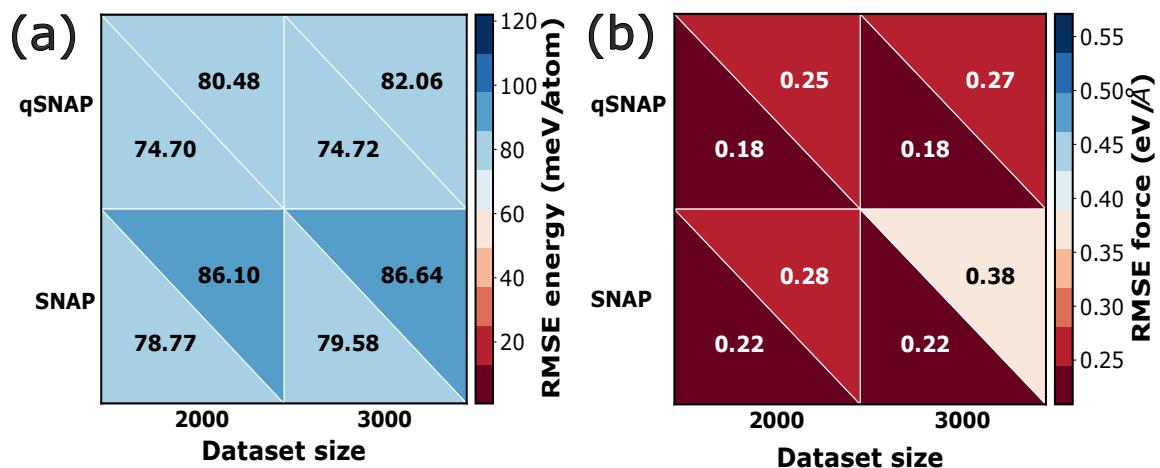
<b>Dataset size:</b> 2000		<b>Training time:</b> 279 hours		
<b>Dataset</b>	<b>Energy (RMSE) meV/atom</b>	<b>Energy (MAE) meV/atom</b>	<b>Force (RMSE) eV/Å</b>	<b>Force (MAE) eV/Å</b>
<b>Train</b>	48.658	31.766	0.084	0.054
<b>Test</b>	40.672	25.641	0.133	0.070

**Table S5.** Hyperparameters for NequIP used in this work. Detailed discussion of NequIP hyperparameters can be found in Ref.<sup>5</sup>.

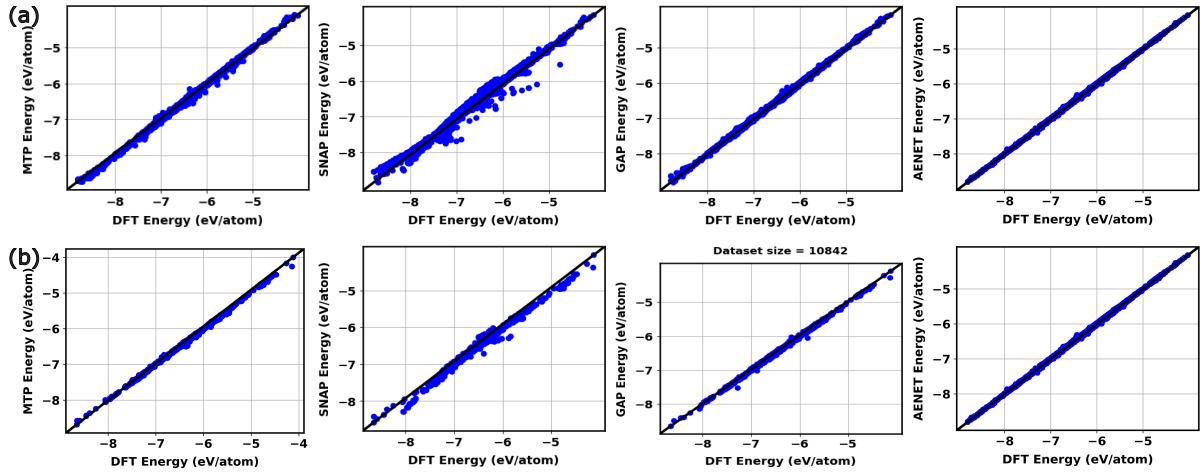
$R_c$ (Radial)	8.0 Å
$R_c$ (Angular)	6.5 Å
Nodes per hidden layer	15
Number of hidden layers	2
Polynomial Cutoff_p	6
Num_features	16
l_max	1
Learning_rate	0.0049

**Table S6.** Compilation of density functional theory (DFT) and MTP calculated voltages for various layered, single-TM, LiTMO<sub>2</sub>-TMO<sub>2</sub> compositions (TM = Ti, V, Cr, Mn, Fe, Co, Ni, or Cu). Absolute and percentage errors made by MTP for each electrode composition is indicated, as are the mean absolute and mean percentage errors.

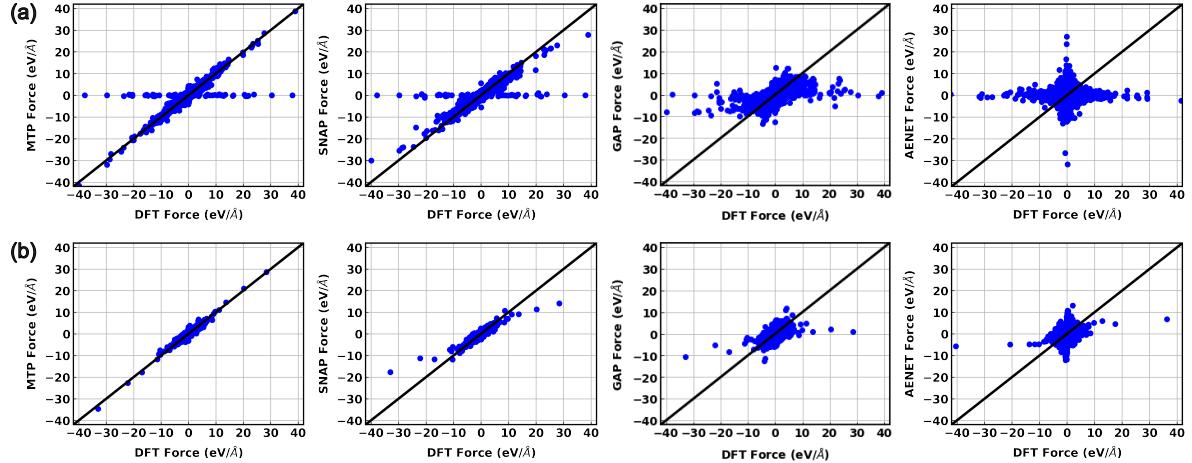
Composition	Voltage (V vs. Li)		Absolute error (V)	Percentage error (%)
	DFT	AENET		
LiTiO <sub>2</sub> -TiO <sub>2</sub>	1.82	2.30	0.48	26.35
LiVO <sub>2</sub> -VO <sub>2</sub>	3.32	2.90	0.42	12.67
LiCrO <sub>2</sub> -CrO <sub>2</sub>	4.16	3.73	0.43	10.34
LiMnO <sub>2</sub> -MnO <sub>2</sub>	2.89	3.38	0.49	17.10
LiFeO <sub>2</sub> -FeO <sub>2</sub>	3.94	4.07	0.13	3.21
LiCoO <sub>2</sub> -CoO <sub>2</sub>	3.81	3.45	0.36	9.37
LiNiO <sub>2</sub> -NiO <sub>2</sub>	4.64	4.27	0.37	7.97
LiCuO <sub>2</sub> -CuO <sub>2</sub>	3.97	3.56	0.41	10.44
<b>Mean error</b>			0.39	12.18



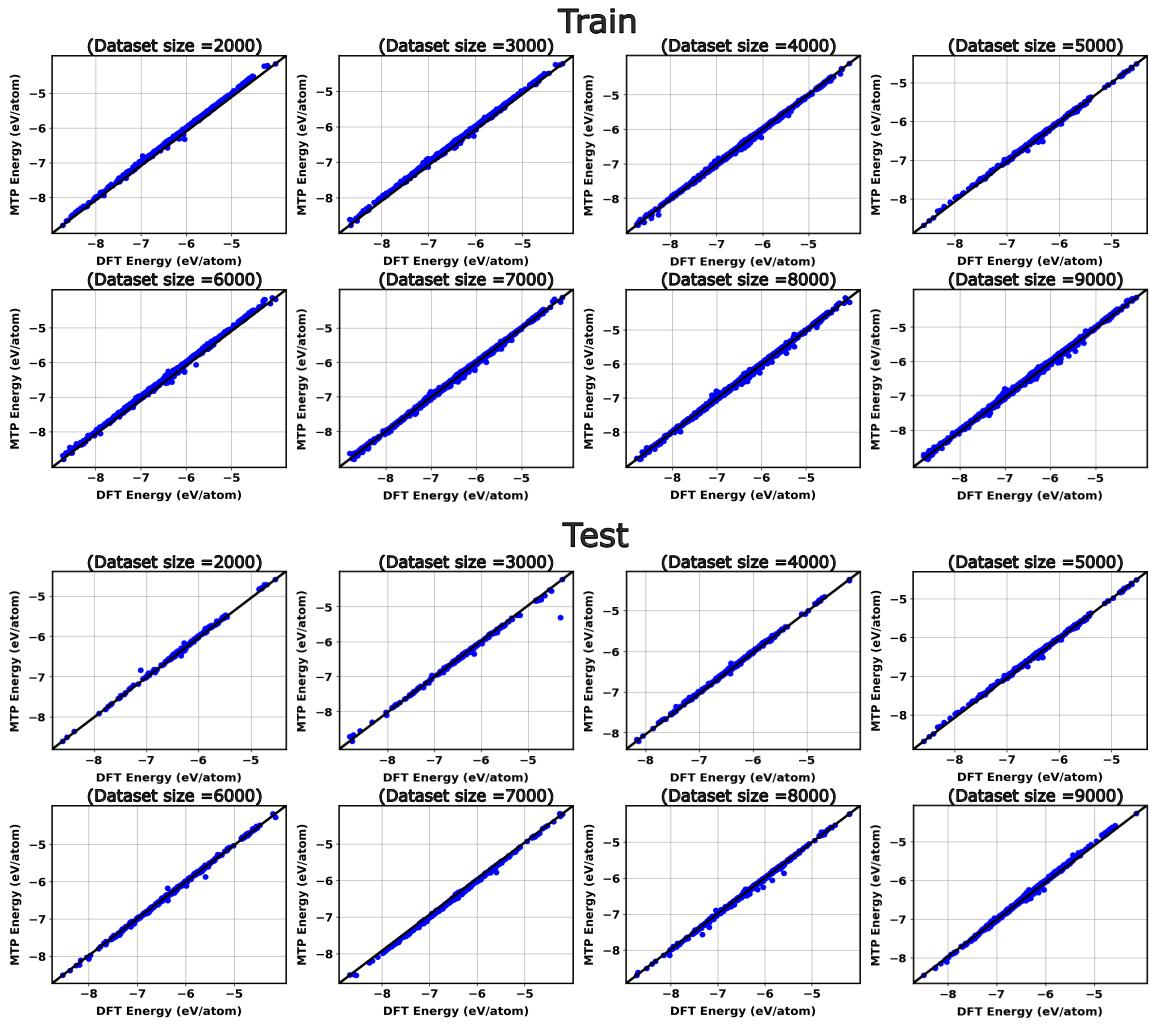
**Figure S1.** RMSE in (a) energies and (b) atomic forces using SNAP (lower rows) and qSNAP (upper rows), respectively, for dataset sizes of 2000 and 3000. The lower left and upper right triangles within each cell represent training and test errors, respectively.



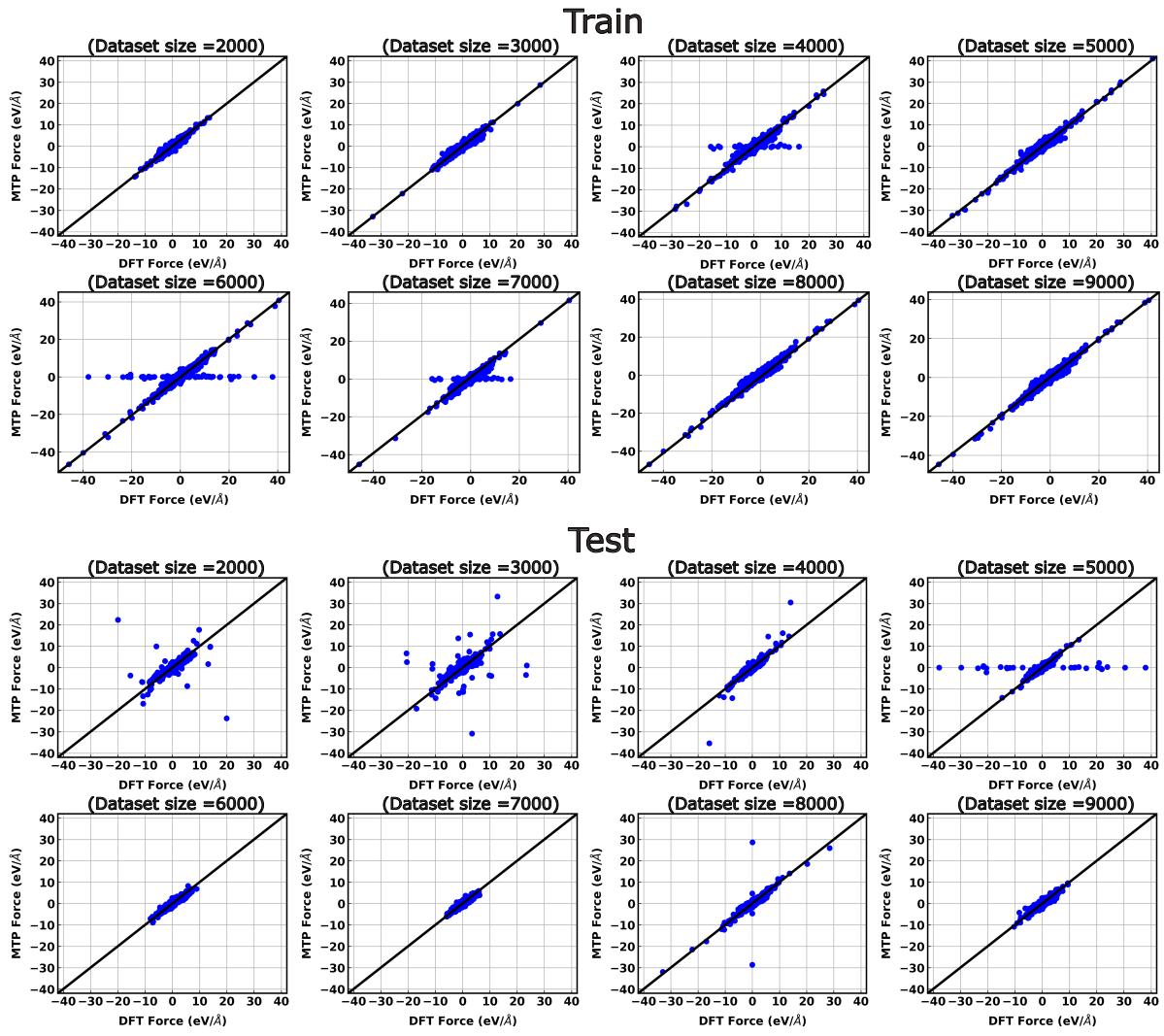
**Figure S2.** Parity plots of MLIP vs. density functional theory (DFT) total energies (in eV/atom) for MTP, SNAP, GAP, and AENET frameworks. Dataset size is 10842. Row (a) represents training dataset and row (b) indicates test dataset.



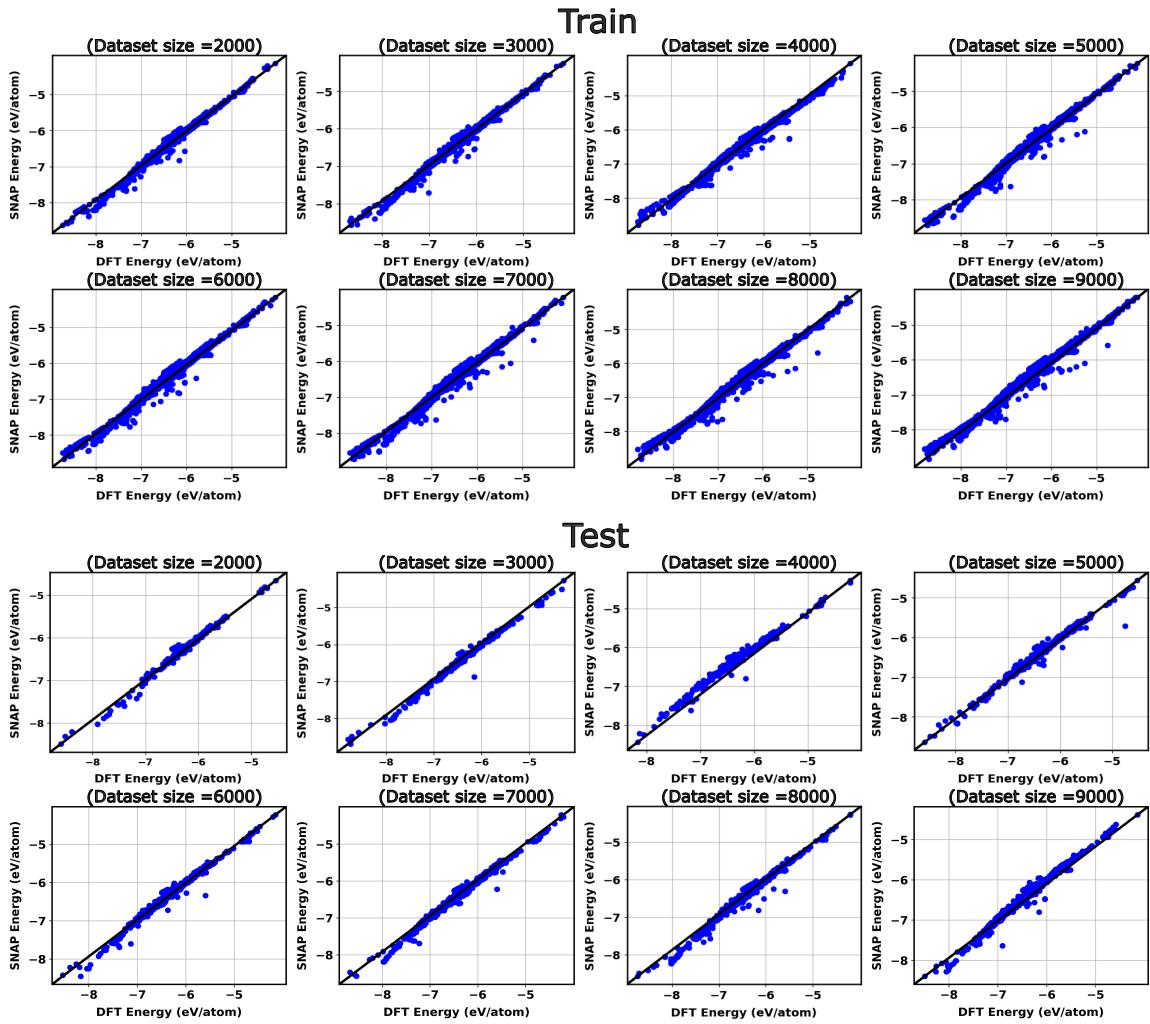
**Figure S3.** Parity plots of MLIP vs. DFT atomic forces (in eV/Å) for all MLIPs considered. Dataset size is 10842. Row (a) is training set and row (b) is test set.



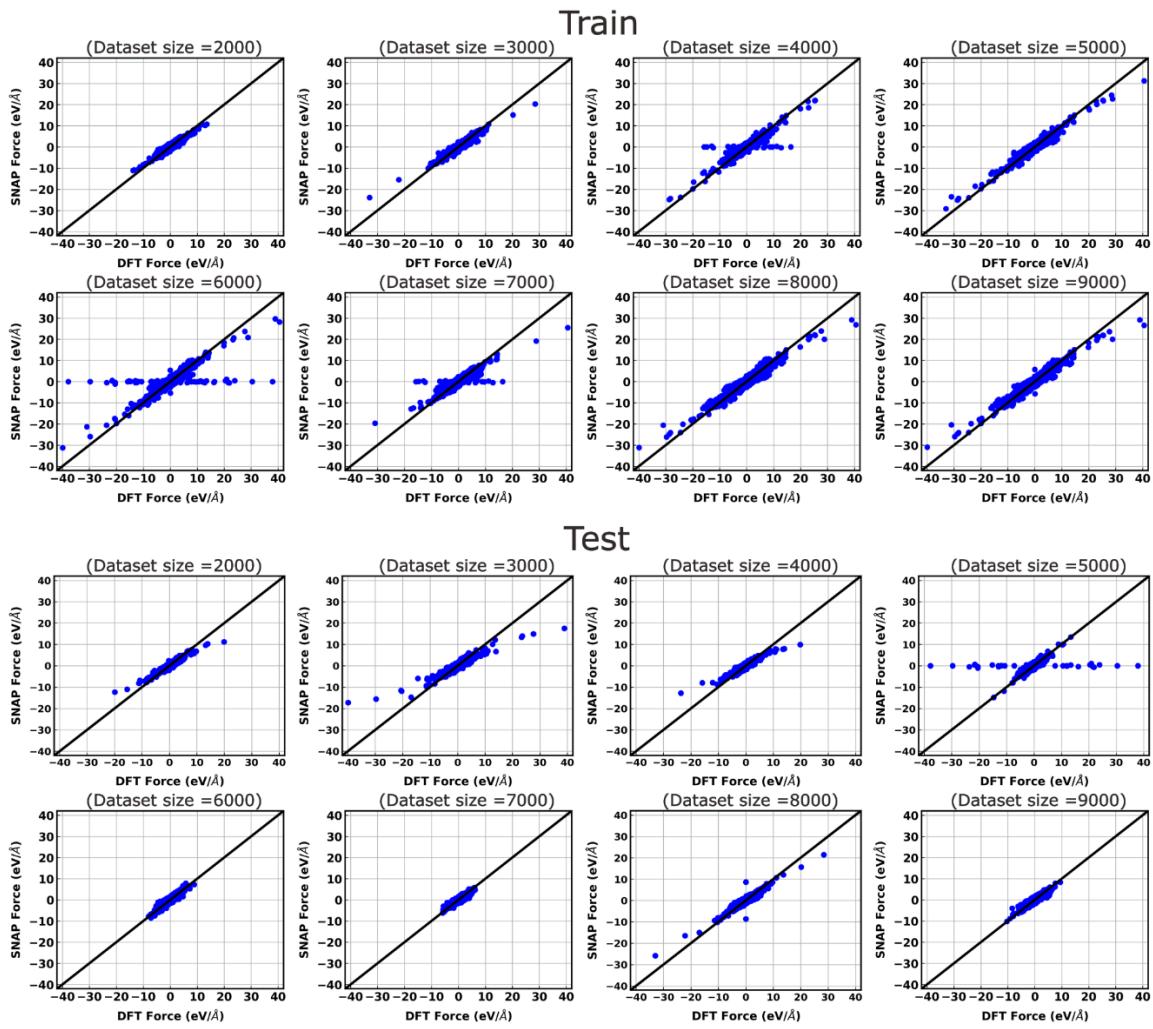
**Figure S4.** Parity plots of MTP vs. DFT energies (in eV/atom) for various dataset sizes, from 2000 to 9000. Top two rows correspond to training sets, while bottom two rows correspond to test sets.



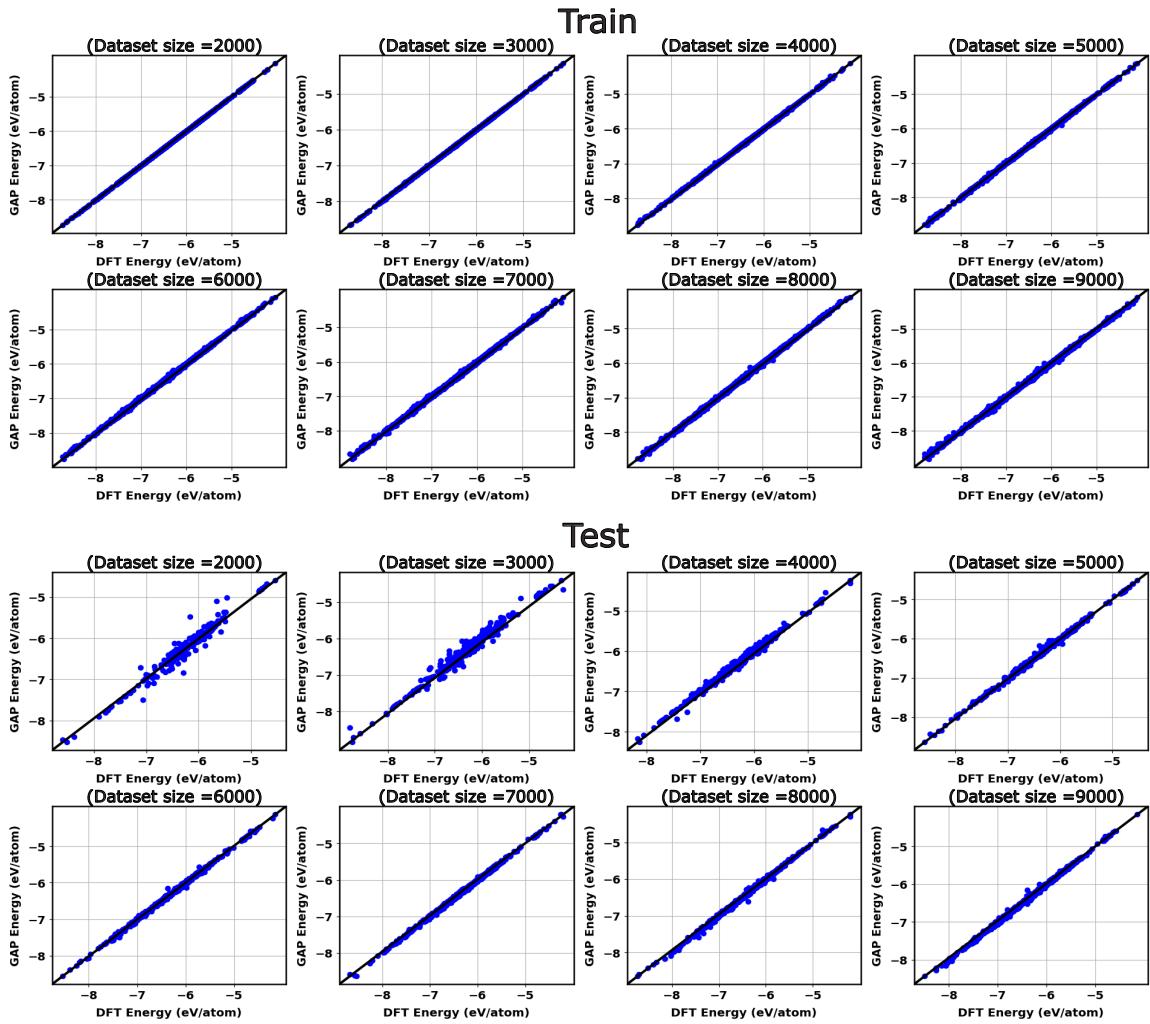
**Figure S5.** Parity plots of MTP vs. DFT atomic forces (in eV/Å) for different dataset sizes. Notations on the figure are identical to **Figure S4**.



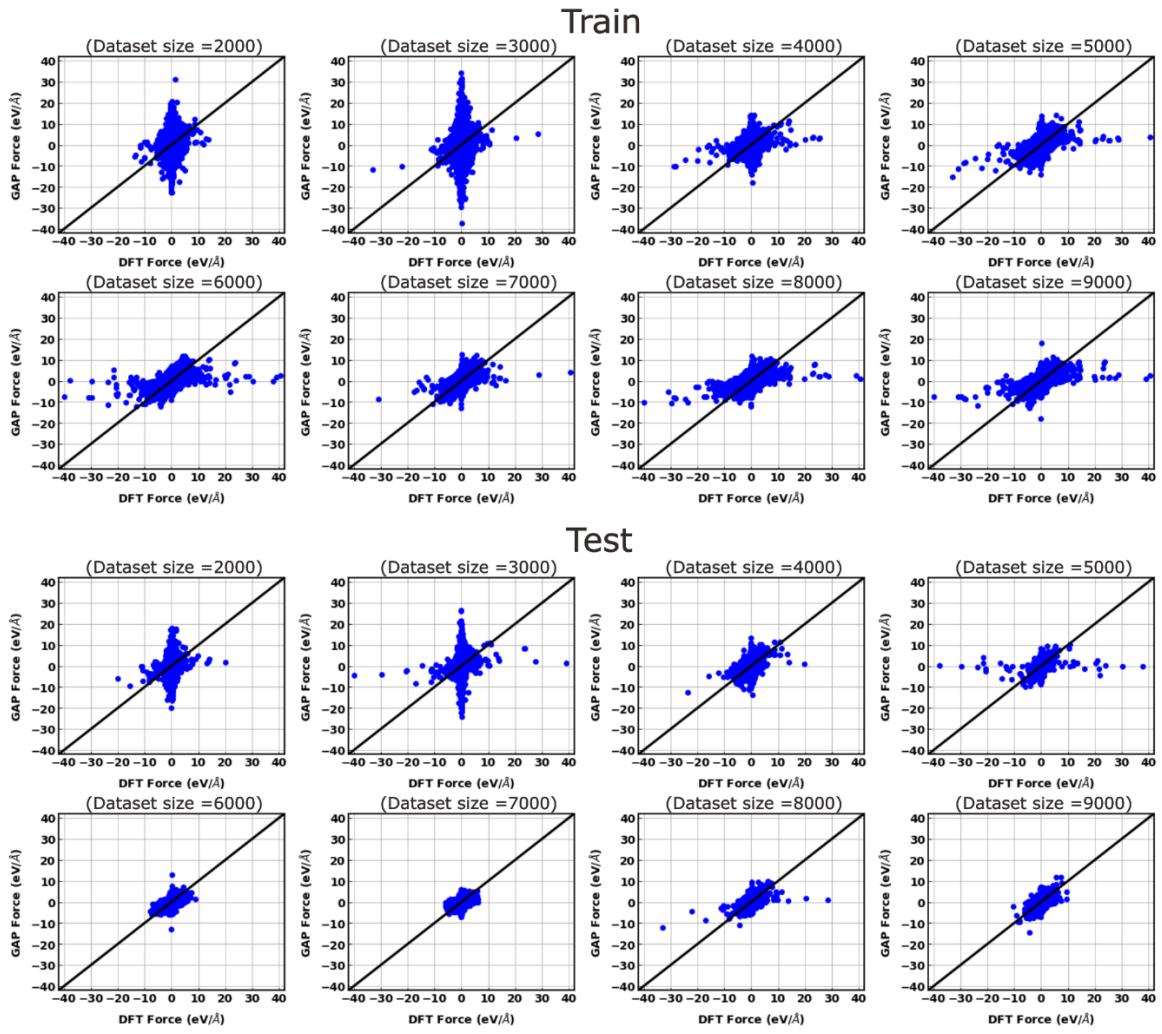
**Figure S6.** Parity plots of SNAP vs. DFT energies (in eV/atom) for different dataset sizes. Notations on the figure are identical to **Figure S4**.



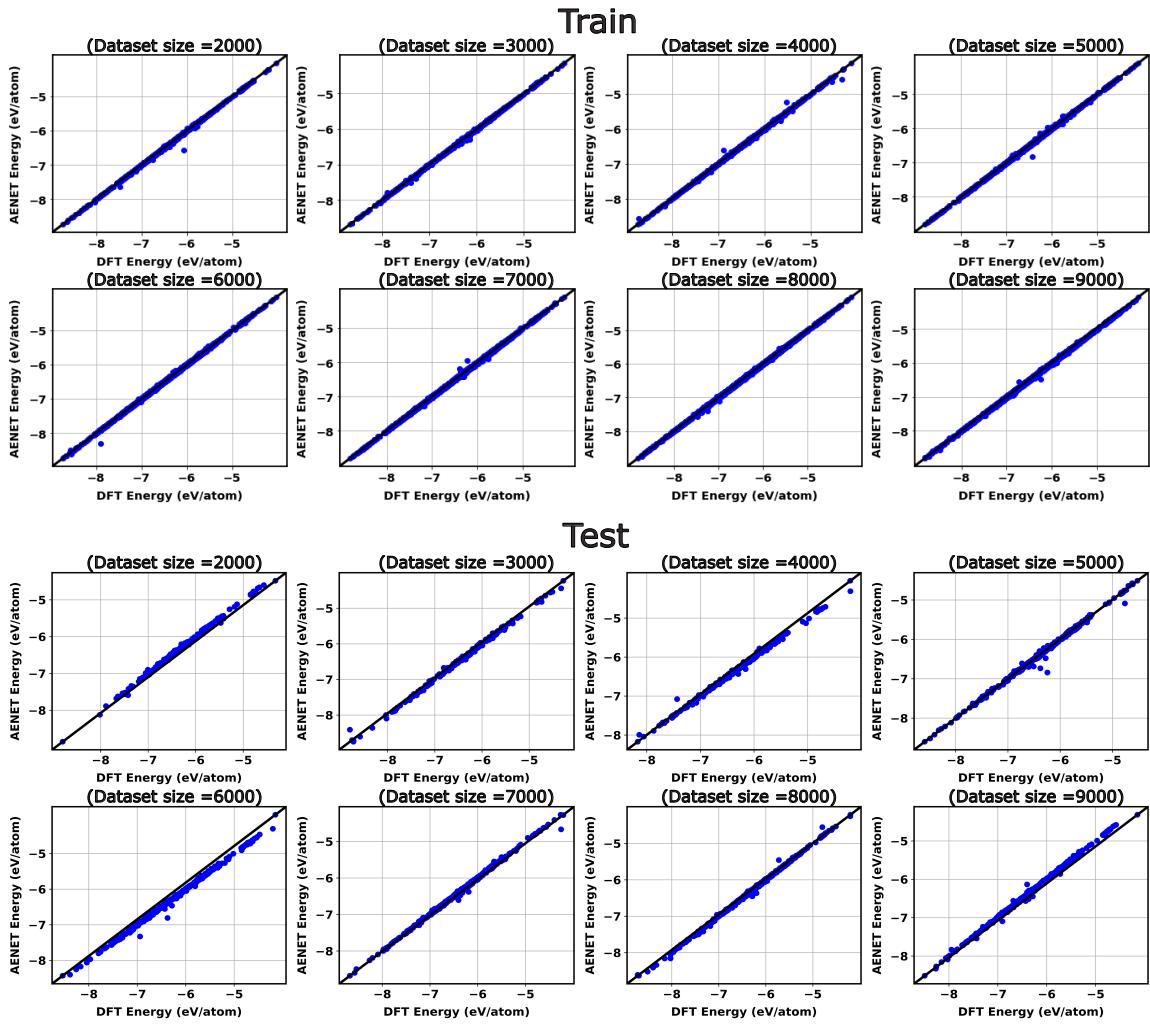
**Figure S7.** Parity plots of SNAP vs. DFT atomic forces (in eV/Å) for different dataset sizes. Notations on the figure are identical to **Figure S4**.



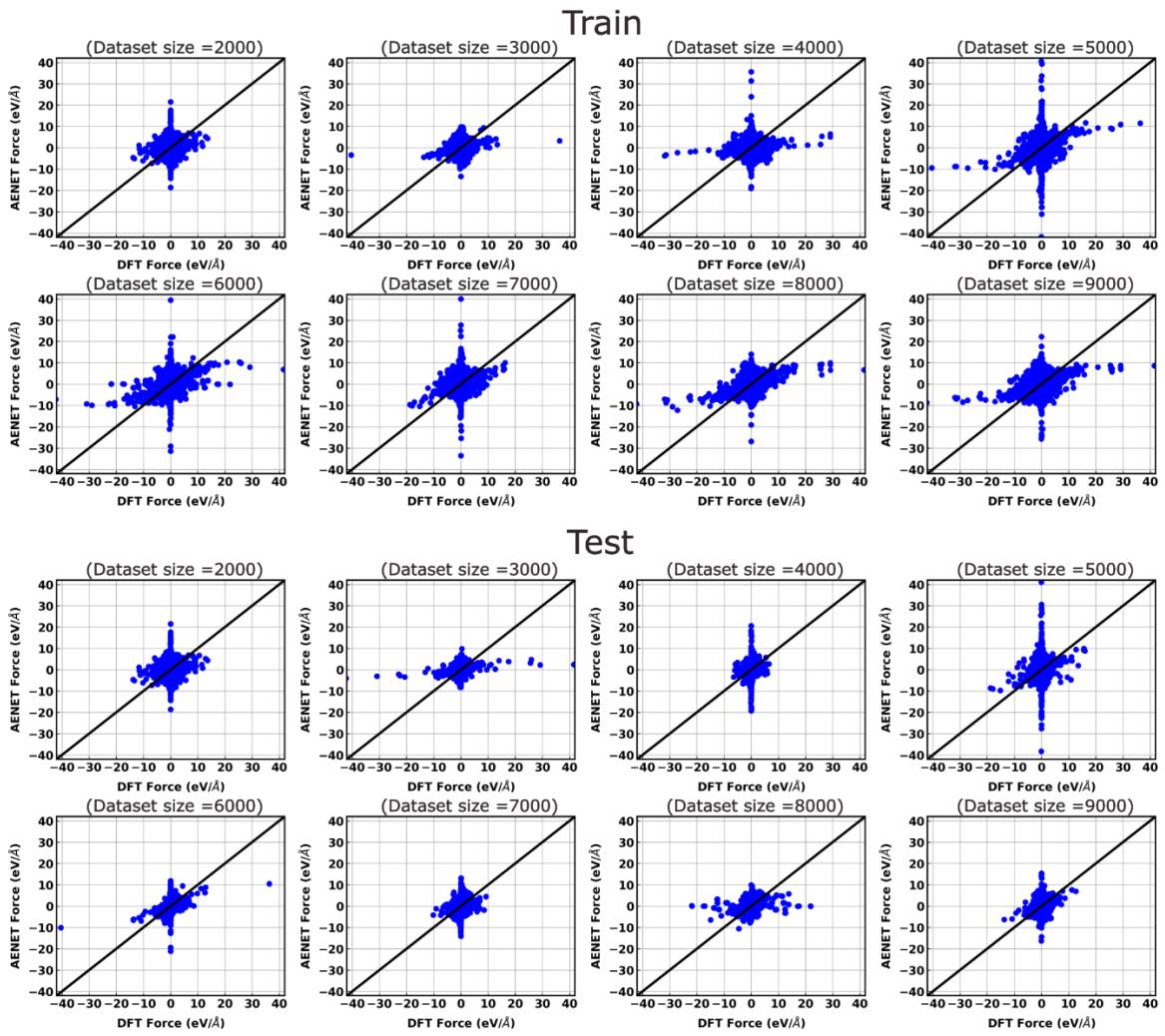
**Figure S8.** Parity plots of GAP vs. DFT energies (in eV/atom) for different dataset sizes. Notations on the figure are identical to **Figure S4**.



**Figure S9.** Parity plots of GAP vs. DFT atomic forces (in eV/Å) for different dataset sizes. Notations on the figure are identical to **Figure S4**.



**Figure S10.** Parity plots of AENET vs. DFT energies (in eV/atom) for different dataset sizes. Notations on the figure are identical to **Figure S4**.



**Figure S11.** Parity plots of AENET vs. DFT atomic forces (in eV/Å) for different dataset sizes. Notations on the figure are identical to **Figure S4**.

## References

- (1) Shapeev, A. V. Moment Tensor Potentials: A Class of Systematically Improvable Interatomic Potentials. *Model Simul* **2016**, *14* (3), 1153–1173.
- (2) Zhu, C.; Byrd, R. H. *Algorithm 778: L-BFGS-B: Fortran Subroutines for Large-Scale Bound-Constrained Optimization*; **1997**.
- (3) Szlachta, W. J.; Bartók, A. P.; Csányi, G. Accuracy and Transferability of Gaussian Approximation Potential Models for Tungsten. *Phys Rev B* **2014**, *90*, 104108.
- (4) Zuo, Y.; Chen, C.; Li, X.; Deng, Z.; Chen, Y.; Behler, J.; Csányi, G.; Shapeev, A. V.; Thompson, A. P.; Wood, M. A.; Ong, S. P. Performance and Cost Assessment of Machine Learning Interatomic Potentials. *J Phys Chem A* **2020**, *124* (4), 731–745.
- (5) Batzner, S.; Musaelian, A.; Sun, L.; Geiger, M.; Mailoa, J. P.; Kornbluth, M.; Molinari, N.; Smidt, T. E.; Kozinsky, B. E(3)-Equivariant Graph Neural Networks for Data-Efficient and Accurate Interatomic Potentials. *Nat Commun* **2022**, *13* (1), 2453.