Electronic Supporting Information

EGraFFBench: Evaluation of Equivariant Graph Neural Network Force Fields for Atomistic Simulations

Vaibhav Bihani
Indian Institute of Technology Delhi,
Hauz Khas, New Delhi, India, 110016
vaibhav.bihani525@gmail.com

Sajid Mannan,
Indian Institute of Technology Delhi,
Hauz Khas, New Delhi, India, 110016
cez218288@civil.iitd.ac.in

Utkarsh Pratiush,
Indian Institute of Technology Delhi,
Hauz Khas, New Delhi, India, 110016
utkarshp1161@gmail.com

Tao Du,
Aalborg University, 9220 Aalborg, Denmark
dutao220@gmail.com

Zhimin Chen,
Aalborg University, 9220 Aalborg, Denmark
zhiminc@bio.aau.dk

Santiago Miret,
Intel Labs,
Santa Clara, CA, USA
santiago.miret@intel.com

Matthieu Micoulaut,
Sorbonne University,
Paris, France
matthieu.micoulaut@gmail.com

Morten M. Smedskjaer,
Aalborg University, 9220 Aalborg, Denmark
mos@bio.aau.dk

Sayan Ranu,
Indian Institute of Technology Delhi,
Hauz Khas, New Delhi, India 110016
sayanranu@cse.iitd.ac.in

N. M. Anoop Krishnan
Indian Institute of Technology Delhi,
Hauz Khas, New Delhi, India 110016
krishnan@iitd.ac.in

37th Conference on Neural Information Processing Systems (NeurIPS 2023).

A Appendix

A.1 Dataset details

Table S1 shows the details of which models have been evaluated on which datasets in the literature. We note that there have been no exhaustive analysis of all the models on even one dataset.

Dataset	# Atoms	# Atom types	NequIP	Allegro	BOTNet	MACE	Equiformer	TorchMDNet
MD17	9-21	2-3	✓	✓	-	-	✓	✓
LiPS	83	3	✓	-	-	-	-	-
3BPA	27	4	-	✓	✓	✓	-	-
AcAc	15	3	-	-	✓	✓	-	-
LiPS20	32-260	1-3	-	-	-	-	-	-
GeTe	200	2	-	-	-	-	-	-

Table S1: Datasets considered in the present work. The tick represents the datasets that have been evaluated on the respective EGraFF model in previous work. Note that none of the datasets have been evaluated and compared for all the models in the literature. LiPS20 and GeTe are two new datasets in the present work.

A.2 LiPS20

Material	Composition	Atom number	Number of configurations
β-$Li_3P_4S_4$	$Li_{24}P_8S_{32}$	64	1000
γ-$Li_3P_4S_4$	$Li_{48}P_{16}S_{64}$	128	1000
$Li_2P_2S_6$	$Li_{16}P_{16}S_{48}$	80	1000
Hexagonal Li_2PS_3	$Li_{32}P_{16}S_{48}$	96	1000
Orthorhombic Li_2PS_3	$Li_{32}P_{16}S_{48}$	96	1000
Li_2S	$Li_{64}S_{32}$	96	1000
Li_3P	$Li_{48}P_{16}$	64	1000
$Li_4P_2S_6$	$Li_{32}P_{16}S_{48}$	96	1000
$Li_7P_3S_{11}$	$Li_{28}P_{12}S_{44}$	84	1000
Li_7PS_6	$Li_{28}P_4S_{24}$	56	1000
$Li_{48}P_{16}S_{61}$	$Li_{48}P_{16}S_{61}$	125	1000
P_2S_5	P_8S_{20}	28	1000
P_4S_3	$P_{32}S_{24}$	56	1000
$67Li_2S - 33P_2S_5$	$Li_{82}P_{40}S_{138}$	260	1000
$70Li_2S - 30P_2S_5$	$Li_{82}P_{38}S_{133}$	253	1000
$75Li_2S - 25P_2S_5$	$Li_{91}P_{35}S_{129}$	255	1000
$80Li_2S - 20P_2S_5$	$Li_{92}P_{34}S_{128}$	254	1000
Li	Li_{54}	54	1000
P	P_{48}	48	1000
S	S_{32}	32	1000

Table S2: Different compositions in LiPS20 dataset

All the ab initio calculations were carried out at the DFT level (Kohn & Sham (1965)) using the Quickstep module of the CP2K package(Kühne et al. (2020)) with the hybrid Gaussian and plane wave method (GPW)(VandeVondele et al. (2005)). The basis functions are mapped onto a multi-grid system with the default number of four different grids with a plane-wave cutoff for the electronic density to be 500 Ry, and a relative cutoff of 50 Ry to ensure the computational accuracy. The AIMD trajectories at 3000 K were obtained in the NVT ensemble with a timestep of 0.5 fs for 2.5 ps. The temperature selection of 3000 K can enable the sampling of the melting process within the short time scale, which can be used for simulating both the crystal and glass structure afterward. The temperature was controlled using the Nosé–Hoover thermostat (Nosé (1984)). The exchange-correlation energy was calculated using the Perdew-Burke-Ernzerhof (PBE) approximation(Perdew et al. (1996)), and the dispersion interactions were handled by utilizing the empirical dispersion correction (D3) from Grimme (Grimme et al. (2010)). The pseudopotential GTH-PBE combined with the corresponding basis sets were employed to describe the valence electrons of Li (DZVP-MOLOPT-SR-GTH), P (TZVP-MOLOPT-GTH), and S (TZVP-MOLOPT-GTH), respectively(Goedecker et al.

(1996)). In addition to the dataset from the AIMD trajectories, the expanded dataset was realized by single energy calculation using the active machine learning method implemented in the DP-GEN package (Zhang et al. (2020)). The active machine learning scheme was carried out based on the glass structure of xLi2S-(100-x)P2S5 (x = 67, 70, 75, and 80) in order to strengthen the capability of the force field in reproducing the glass structures of different lithium thiophosphates. The training dataset consists following compositions, shuffled randomly: $Li, Li_2S, Li_{48}P_{16}S_{61}, P_4S_3, Li_7PS_6$. Crystal structures set included $beta - Li_3PS_4, Li_2PS_3 - hex, gamma - Li_3PS_4, Li_2PS_3 - orth$, and rest compositions were used as the test dataset.

A.3 Timestep and temperature details

Table S16 displays the temperature in Kelvin and the corresponding timestep in femtoseconds for various datasets utilized in the forward simulations. These values remain consistent with the original sampled datasets.

Dataset	Temperature(K)	Timestep(fs)
Acetylacetone	300, 600	1.0,0.5
3BPA	300, 600	1.0
MD17	500	0.5
LiPS	520	1.0
LiPS20	3000	1.0
GeTe	920	0.12

Table S3: Temperature (T) and Timestep(fs) for the forward simulation on different datasets

A.4 Radial distribution function

Figure S1 shows the reference and generated radial distribution functions(RDFs) for 3BPA, Acetylacetone, LiPS and GeTe. The generated RDFs are obtained after averaging over five simulations trajectories of 1000 steps.

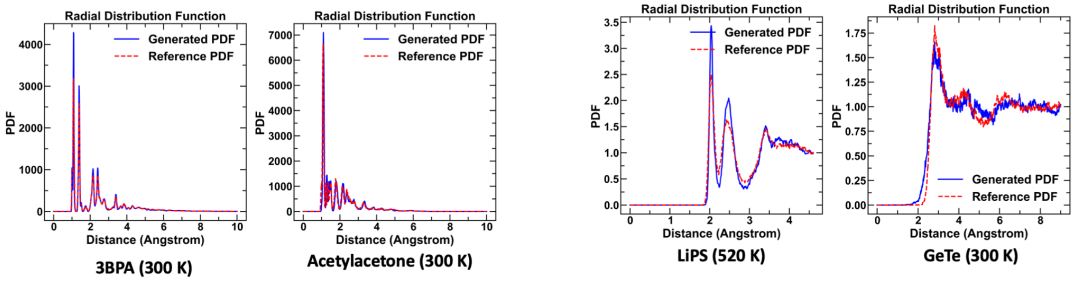

Figure S1: Pair distribution function(PDF) over the simulation trajectory. Reference PDF in red and generated PDF in blue represent ground truth and predicted RDFs. The values are computed as the average of five forward simulations for 1000 timesteps on each dataset with different initial conditions.

A.5 Mean Energy and force violation

Figure S2 shows the obtained geometric mean of energy and force violation errors for the trained models on all the datasets. We observe that the variation of energy error among the models is quite large for some datasets like MD17 and LiPS20, and very small for datasets like 3BPA and Acetylacetone.

A.6 Rollout Energy and force violation

The evolution of energy violation error, EV(t), and force violation error, FV(t), obtained as average over five forward simulations for different datasets are shown in Figure S3.

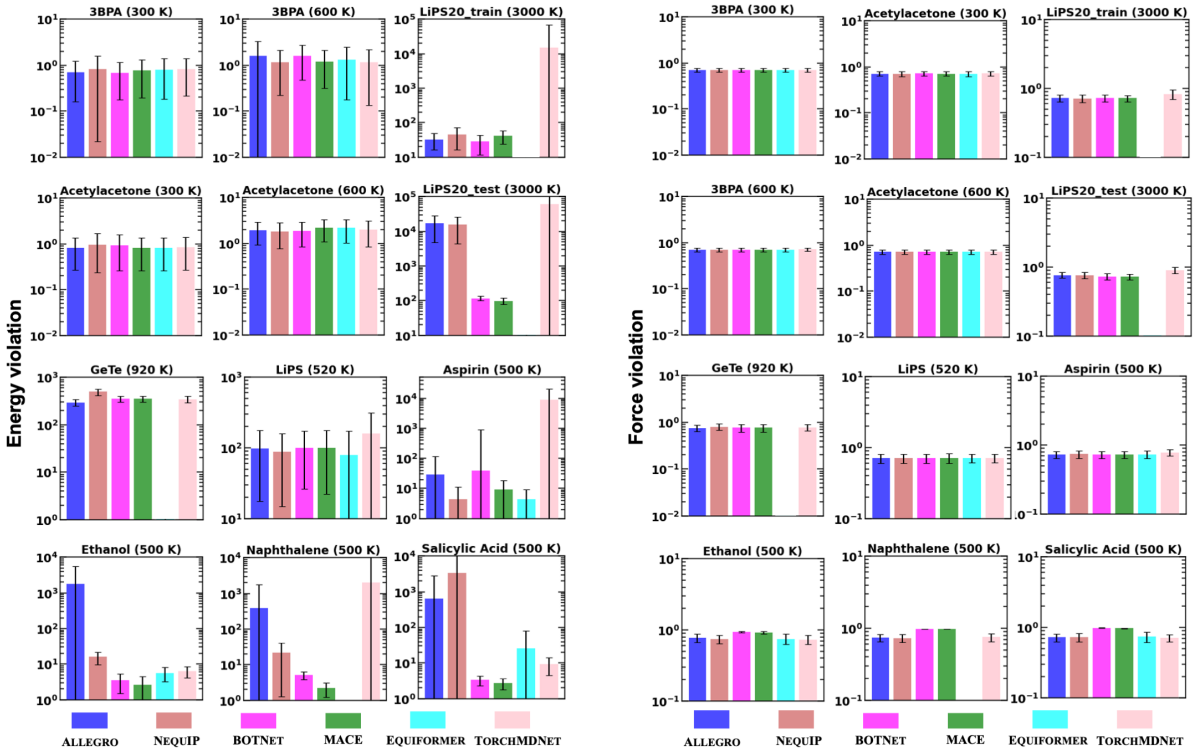

Figure S2: Geometric mean of energy ($\times 10^{-5}$) and force violation error over the simulation trajectory. The error bar shows a 95% confidence interval. The values are computed as the average of five forward simulations for 1000 timesteps on each dataset with different initial conditions.

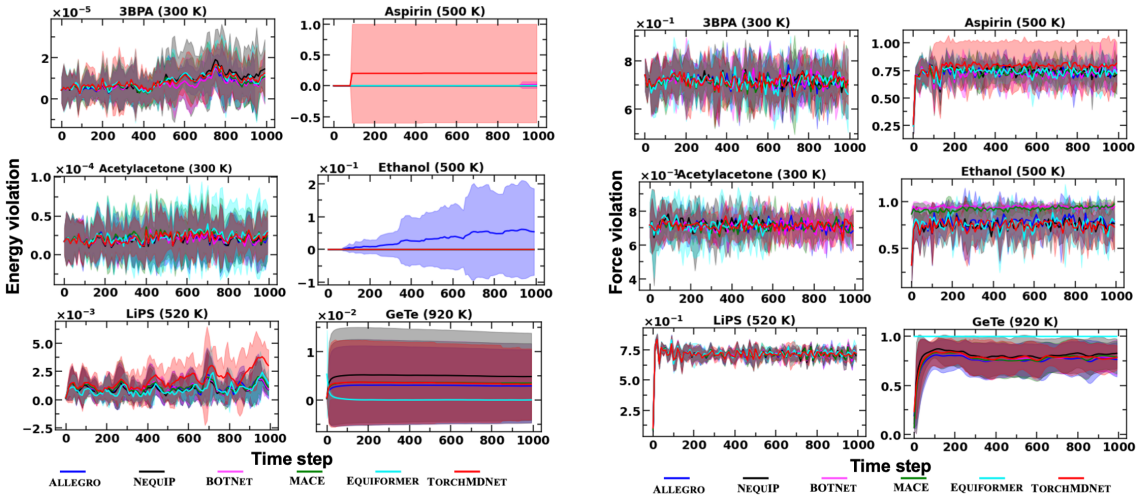

Figure S3: Energy ($\times 10^{-5}$) and force violation error over the simulation trajectory. The error bar shows a 95% confidence interval. The values are computed as the average of five forward simulations for 1000 timesteps on each dataset with different initial conditions.

A.7 Comparative Analysis

Figure S4 shows the comparative radial plots for different metrics for all the datasets. For better interpretability, we normalize all the metrics with respect to the its largest value in the dataset. Figure S5 shows the comparison of different pairs of related metrics for all the datasets and models.

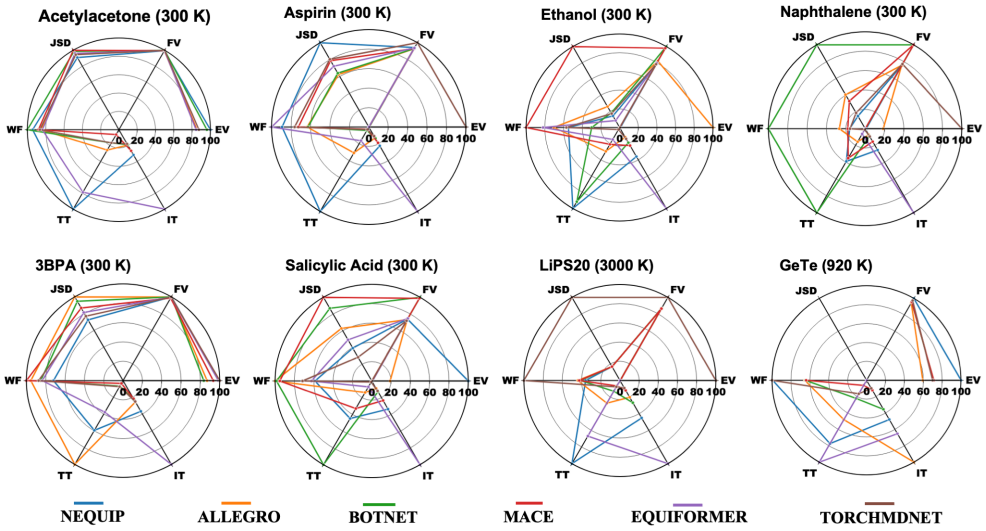

Figure S4: Comparative analysis of different metrics for all models across datasets. The color of the line indicates model identity. The values are normalized by dividing their respective maximum values and then multiplying it by 100.

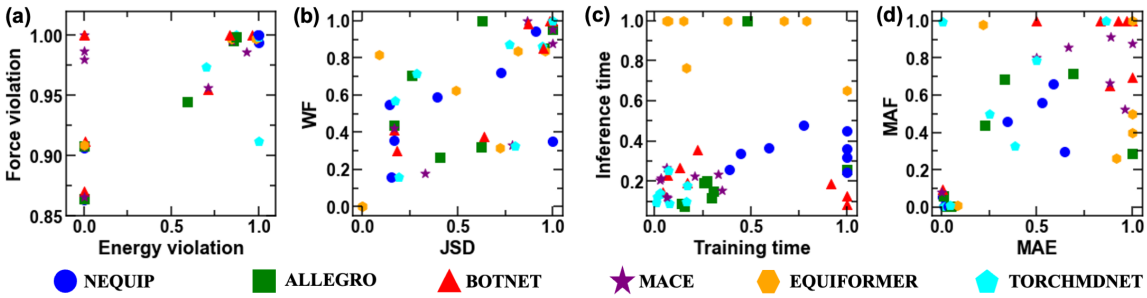

Figure S5: Comparision of (a) Energy violation and Force Violation,(b) JSD and WF, (c) Training time and Inference time, and (d) Mean absolute energy error(MAE) and Mean absolute force error (MAF), for all dataset. The values are normalized by the largest values to scale between 0 and 1.

A.8 Hardware details

All the models are trained using A100 80GB PCI GPUs, and inference performed using AMD EPYC 7282 16-Core Processor @ 2.80GHz with 1TB installed RAM. All the models uses PyTorch environment, with Atomic simulation environment (ASE) package for forward simulations. Specific versions details are given on the code repository.

A.9 Root mean square displacement plots

A.10 Hyperparameter details

The details of hyperparameters used for training each of the models are provided in the following tables. NEQUIP in Table S4, ALLEGRO in Table S5, BOTNET in Table S6, MACE in Table S7, EQUIFORMER in Table S8, ,TORCHMDNET in Table S9, DimeNET++ in Table S10,and PaiNN in Table S11,

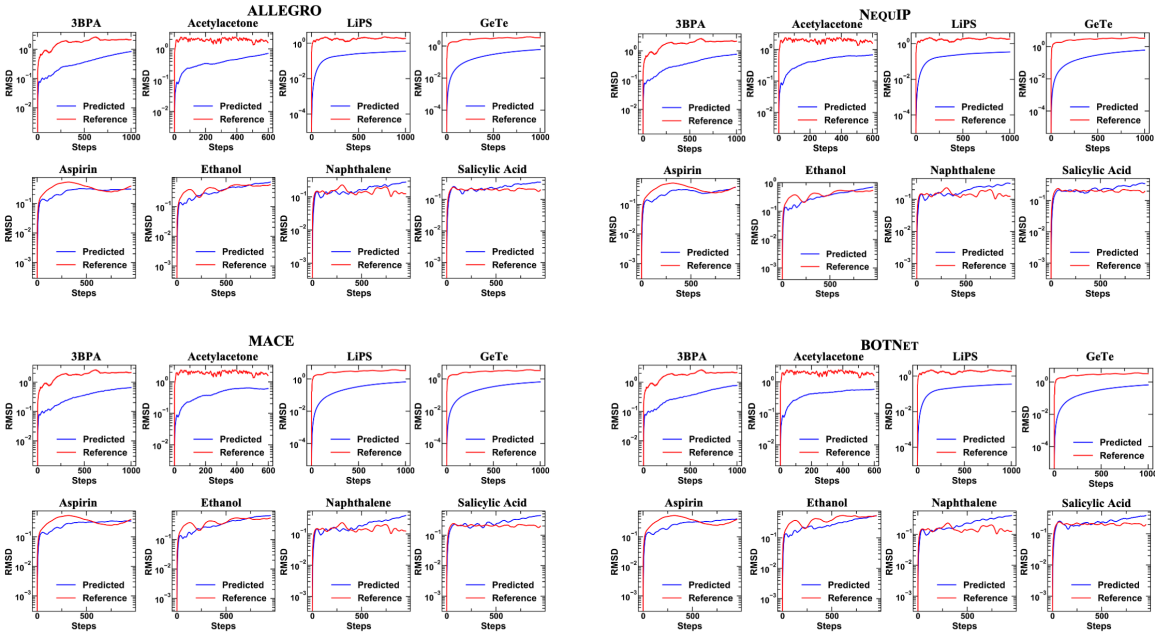

Figure S6: Root mean square displacement plots for models on all datasets. The values are computed as the average of five forward simulations for 1000 timesteps on each dataset with different initial conditions.

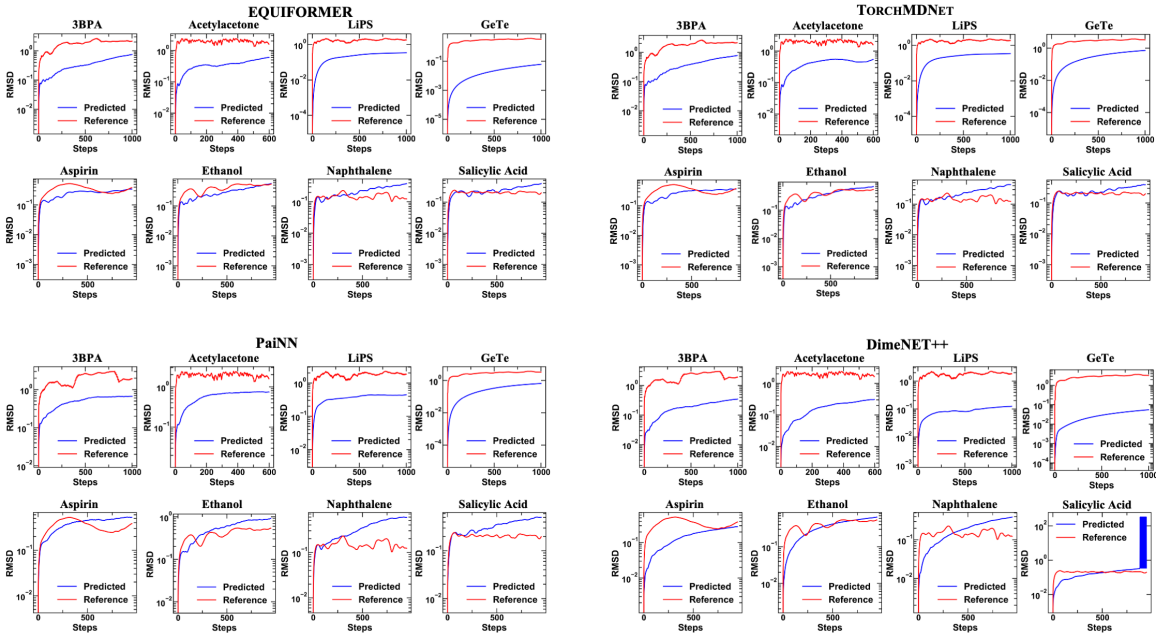

Figure S7: Root mean square displacement plots for all the models on all datasets. The values are computed as the average of five forward simulations for 1000 timesteps on each dataset with different initial conditions.

A.11 Literature comparison

A.11.1 Generalizability to unseen structures

The first task focuses on evaluating the models on an unseen small molecule structure. To this extent, we test the models, trained on four molecules of the MD17 dataset (aspirin, ethaenol, naphthalene,

Hyper-parameter	Value or description
R max	5.0
Number of Layers	6
L max	2
Number of Features	32
Nonlinearity Type	Gate
Nonlinearity Scalars (e)	Silu
Nonlinearity Scalars (o)	Tanh
Nonlinearity Gates (e)	Silu
Nonlinearity Gates (o)	Tanh
Number of Basis	8
BesselBasis Trainable	True
Polynomial Cutoff	6
Invariant Layers	3
Invariant Neurons	64
Learning Rate	0.005
Batch Size	1
EMA Decay	0.99
EMA Use Num Updates	True
Early Stopping Patiences (Validation Loss)	50
Early Stopping Lower Bounds (LR)	1.0e-6
Early Stopping Upper Bounds (Cumulative Wall)	5 days
Loss Coeffs (Forces)	1
Loss Coeffs (Total Energy)	1
Optimizer Name	Adam
LR Scheduler Name	ReduceLROnPlateau
LR Scheduler Patience	5
LR Scheduler Factor	0.8

Table S4: NEQUIP Hyperparameters

and salicylic acid), on the benzene molecule, an unseen molecule from the MD17 dataset. Note that the benzene molecule has a cyclic ring structure. Aspirin and Salicylic acid contain one ring, naphthalene is polycyclic with two rings, while ethanol has a chain structure with no rings. Table S13 shows the EV and FV and Table S14 shows the corresponding JSD and WF. We observe that all the models suffer very high errors in force and energy. EQUIFORMER trained on ethanol and salicylic acid exhibits unstable simulation after the first few steps. Interestingly, non-cyclic ethanol models perform better than aspirin and salicylic acid, although the latter structures are more similar to benzene. Similarly, the model trained on polycyclic Naphthalene performs better than other models. Altogether, we observe that despite having the same chemical elements, models trained on one small molecule do not generalize to an unseen molecule with a different structure.

A.12 Loss curves: PaiNN and DimeNET++

Parameter	Value
R Max	5.0
PolynomialCutoff	6
L Max	2
Num Layers	2
Env Embed Multiplicity	64
Embed Initial Edge	True
Two Body Latent MLP Dimensions	[128, 256, 512, 1024]
Two Body Latent MLP Nonlinearity	Silu
Latent MLP Latent Dimensions	[1024, 1024, 1024]
Latent MLP Nonlinearity	Silu
Latent Resnet	True
Edge Eng MLP Latent Dimensions	[128]
Edge Eng MLP Nonlinearity	None
Learning Rate	0.005
Batch Size	1
Max Epochs	10000
EMA Decay	0.99
Early Stopping Patiences(Validation loss)	50
Early Stopping Lower Bounds(LR)	1.0×10^{-6}
Early Stopping Upper Bounds(Cumulative wall)	5 days
Loss Coefficients(Forces)	1
Loss Coefficients(Total energy)	1
Optimizer Name	Adam
LR Scheduler Name	ReduceLROnPlateau
LR Scheduler Patience	5
LR Scheduler Factor	0.8

Table S5: ALLEGRO Hyperparameters

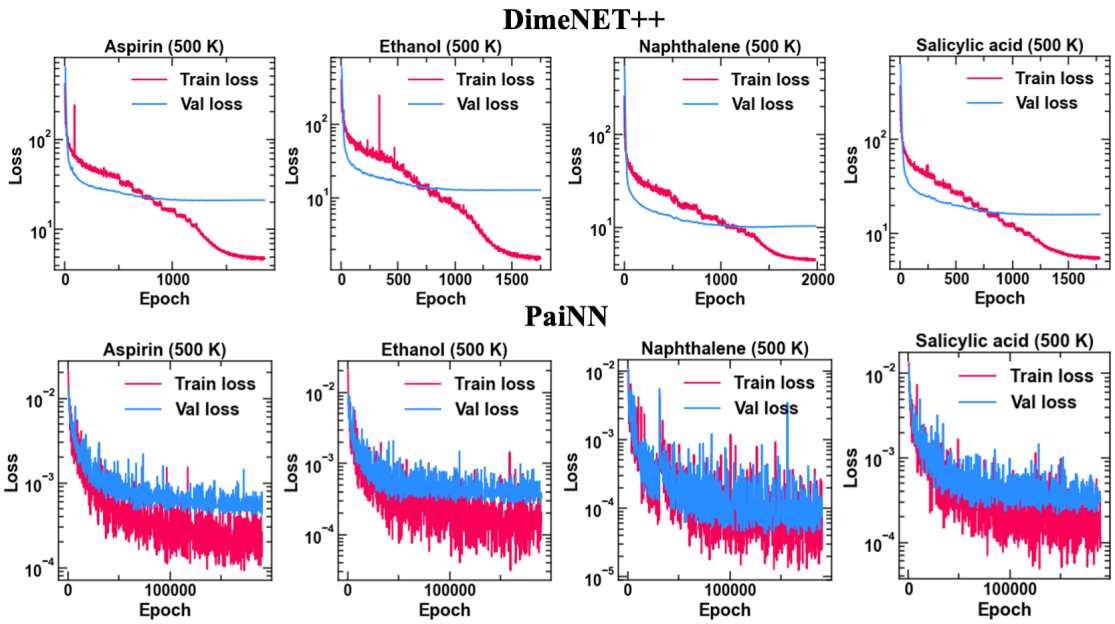

Figure S8: Loss curves for PaiNN and DimeNET++ models on MD17 molecules

Hyper-parameter	**Value or description**
R_{max}	5.0
Correlation order	1
Number of Radial basis	8
Numcber of Cutoff basis	5
L_{max}	3
Number of Interactions	5
MLP Irreps	16x0e
Hidden Irreps	16x0e+16x1o+16x2e
Gate	Silu
E_{0s}	{1:-13.663181292231226, 3:-216.78673811801755, 6:-1029.2809654211628, 7:-1484.1187695035828, 8:-2042.0330099956639, 15:-1537.0898574856286, 16:-1867.8202267974733}
Forces weight	10.0
SWA Forces Weight	1.0
Energy Weight	1.0
SWA Energy Weight	1000.0
Virials Weight	1.0
SWA Virials Weight	10.0
Config type Weights	{"Default":1.0}
optimizer	AMSGrad Adam
Batch Size	5
Validation Batch Size	5
Learning rate	0.01
SWA learning rate	0.001
Weight decay	5e-7
EMA	True
EMA Decay	0.99
Scheduler	ReduceLROnPlateau
LR factor	0.8
Scheduler patience	50
LR Scheduler gamma	0.9993
SWA	True
Max number of epochs	1500
Clip gradiants	10.0

Table S6: BOTNET Hyperparameters

Hyper-parameter	**Value or description**
R_{max}	5.0
Correlation order	3
Number of Radial basis	8
Numcber of Cutoff basis	5
L_{max}	3
Number of Interactions	2
MLP Irreps	16x0e
Hidden Irreps	16x0e+16x1o+16x2e
Gate	Silu
E_{0s}	{1:-13.663181292231226, 3:-216.78673811801755, 6:-1029.2809654211628, 7:-1484.1187695035828, 8:-2042.0330099956639, 15:-1537.0898574856286, 16:-1867.8202267974733}
Forces weight	10.0
SWA Forces Weight	1.0
Energy Weight	1.0
SWA Energy Weight	1000.0
Virials Weight	1.0
SWA Virials Weight	10.0
Config type Weights	{"Default":1.0}
optimizer	AMSGrad Adam
Batch Size	5
Validation Batch Size	5
Learning rate	0.01
SWA learning rate	0.001
Weight decay	5e-7
EMA Decay	0.99
Scheduler	ReduceLROnPlateau
LR factor	0.8
Scheduler patience	50
LR Scheduler gamma	0.9993
SWA	True
Max number of epochs	1500
Clip gradiants	10.0

Table S7: MACE Hyperparameters

Hyper-parameters	**Value or description**
Optimizer	AdamW
Learning rate scheduling	Cosine learning rate with linear warmup
Warmup epochs	10
Maximum learning rate	5×10^{-4}
Batch size	8
Number of epochs	5000
Weight decay	1×10^{-6}
Energy weight	1.0
Force weight	1.0
Dropout rate	0.0
Cutoff radius ($\mathring{A}$)	5
Number of radial basis	32
Hidden size of radial function	64
Number of hidden layers in radial function	2
Equiformer	
Number of Transformer blocks	6
Embedding dimension d_{embed}	$[(128, 0), (64, 1), (32, 2)]$
Spherical harmonics embedding dimension d_{sh}	$[(1, 0), (1, 1), (1, 2)]$
Number of attention heads h	4
Attention head dimension d_{head}	$[(32, 0), (16, 1), (8, 2)]$
Hidden dimension in feed forward networks d_{ffn}	$[(384, 0), (192, 1), (96, 2)]$
Output feature dimension d_{feature}	$[(512, 0)]$

Table S8: EQUIFORMER Hyperparameters

Hyper-parameter	Value or description
Activation	Silu
Aggregation	Add
Attention Activation	Silu
Batch Size	8
Radius Cutoff Lower	0.0
Radius Cutoff Upper	5.0
Derivative	True
Early Stopping Patience	300
EMA Alpha Force	1.0
EMA Alpha Energy	0.05
Embedding Dimension	128
Energy Weight	0.2
Force Weight	0.8
Inference Batch Size	64
Learning Rate	0.001
Learning Rate Factor	0.8
Minimum Learning Rate	1.0×10^{-7}
Learning Rate Patience	30
Learning Rate Warmup Steps	1000
Max Number of Neighbors	32
Max Z	100
Neighbor Embedding	True
Number of Epochs	5000
Number of Heads	8
Number of Layers	6
Number of Nodes	1
Number of Radial basis function	32
Number of Workers	6
Output Model	Scalar
Precision	32
Radial basis function Type	Expnorm
Reduce Operation	Add
Train Size	500
Weight Decay	0.0

Table S9: TORCHMDNET Hyperparameters

Hyper-parameters	Value
Hidden Channels	128
Output Embedding Channels	256
Interaction Embedding Size	64
Basis Embedding Size	8
Number of Blocks	4
Cutoff Distance	5.0
Envelope Exponent	5
Number of Radial Functions	6
Number of Spherical Functions	7
Number of Layers Before Skip	1
Number of Layers After Skip	2
Number of Output Layers	3
Regress Forces	True
Batch Size	1
Evaluation Batch Size	1
Number of Workers	4
Initial Learning Rate	0.001
Optimizer	Adam
Scheduler	ReduceLROnPlateau
Patience	5
Factor	0.8
Minimum Learning Rate	0.000001
Maximum Epochs	2000
Force Coefficient	1000
Energy Coefficient	1
Exponential Moving Average Decay	0.999
Gradient Clipping Threshold	10
Early Stopping Time	604800
Early Stopping Learning Rate	0.000001

Table S10: DimeNeT++ hyperparameters

Hyper-parameters	Value or desciption
Number of Atom Basis	128
Number of Interactions	3
Number of Radial Basis Functions	20
Cutoff Distance	5.0
Cutoff Network	'cosine'
Radial Basis Function	BesselBasis
Activation Function	silu
Maximum Atomic Number	100

Table S11: PaiNN hyperparameters

	NequIP		Allegro		BOTNet		MACE		Equiformer		TorchMDNet		PaiNN	DimeNET++
	E	F	E	F	E	F	E	F	E	F	E	F	F	F
Aspirin(Ours)	6.84	13.89	5.00	9.17	7.99	14.06	8.53	14.01	6.15	15.29	5.33	8.97	12.41	22.07
Aspirin(Liao & Smidt (2023))	5.7	8.0	-	-	-	-	-	-	5.3	7.2	5.3	11.0	-	-
Aspirin(Fu et al. (2023))	-	2.3	-	-	-	-	-	-	-	-	-	-	9.2	10.0
Aspirin(Thölke & Fabritiis (2022))	-	15.09	-	-	-	-	-	-	-	-	5.33	10.97	-	–
Ethanol(Ours)	2.67	7.49	2.34	5.01	2.60	6.80	2.36	3.19	2.66	9.73	2.67	5.93	11.81	17.19
Ethanol(Liao & Smidt (2023))	2.2	3.1	-	-	-	-	-	-	2.2	3.1	2.3	4.7	-	-
Ethanol(Fu et al. (2023))	-	1.3	-	-	-	-	-	-	-	-	-	-	5.0	4.2
Ethanol(Thölke & Fabritiis (2022))	-	9.02	-	-	-	-	-	-	-	-	2.25	4.73	-	-
Naphthalene(Ours)	5.70	6.20	5.14	2.64	6.67	6.07	6.26	1.98	3.88	7.01	2.55	4.03	4.07	19.65
NaphthaleneLiao & Smidt (2023)	4.9	1.7	-	-	-	-	-	-	3.7	2.1	3.7	2.6	-	-
Naphthalene(Fu et al. (2023))	-	1.10	-	-	-	-	-	-	-	-	-	-	3.8	5.7
Naphthalene(Thölke & Fabritiis (2022))	-	4.21	-	-	-	-	-	-	-	-	3.69	2.64	-	-
Salicylic Acid(Ours)	5.78	8.42	5.76	6.30	5.56	10.21	5.34	4.24	5.22	12.39	6.85	7.19	11.12	25.48
Salicylic acid(Liao & Smidt (2023))	4.6	3.9	-	-	-	-	-	-	4.5	4.1	4.0	5.6	-	-
Salicylic acid(Fu et al. (2023))	-	1.6	-	-	-	-	-	-	-	-	-	-	6.5	9.6
Salicylic acid(Thölke & Fabritiis (2022))	-	10.32	-	-	-	-	-	-	-	-	4.03	5.59	-	-
LiPS(Ours)	165.43	5.04	31.75	2.46	28.0	13.0	30.0	15.0	83.20	51.10	67.0	61.0	112.43	42.23
LiPS(Fu et al. (2023))	-	3.7	-	-	-	-	-	-	-	-	-	-	11.7	3.2

Table S12: Literature comparison

	NequIP		Allegro		BOTNet		MACE		Equiformer		TorchMDNet	
	E	F	E	F	E	F	E	F	E	F	E	F
Aspirin	22650	0.762	22676	0.765	21880	0.760	21881	0.766	47027.742	0.769	46864	0.765
	(11.622)	(0.060)	(0.311)	(0.070)	(6.874)	(0.061)	(12.11)	(0.061)	(3.88)	(0.065)	(184.678)	(0.058)
Ethanol	6154.4	0.740	6224.2	0.711	5860.5	0.935	5863.2	0.921	-	-	20262	0.712
	(0.402)	(0.056)	(12.501)	(0.040)	(0.325)	(0.016)	(0.338)	(0.022)	-	-	(19.401)	(0.052)
Naphthalene	4783.8	0.759	4799.7	0.743	4572.4	0.970	4572.1	0.959	24546	0.761	24440	0.777
	(16.411)	(0.067)	-	(0.070)	(0.32)	(0.008)	(0.324)	(0.012)	(6.069)	(0.061)	-	(0.057)
Salicylic acid	22840	0.766	22849	0.753	22055	0.982	2205	0.965	-	-	35947	0.769
	(0.308)	(0.067)	(0.314)	(0.076)	(0.309)	(0.005)	(0.310)	(0.007)	-	-	(2.907)	(0.057)

Table S13: EV (E) and FV (F) on the forward simulation of benzene molecule by the models trained on aspirin, ethanol, naphthalene, and salicylic acid.

	NequIP		Allegro		BOTNet		MACE		Equiformer		TorchMDNet	
	JSD	WF	JSD	WF	JSD	WF	JSD	WF	JSD	WF	JSD	WF
Aspirin	360854	73.801	573039	61.158	311916	62.842	473362	89.692	482522	75.081	494492	76.828
Ethanol	509375	63.321	1130600	51.601	1108865	57.181	1095829	41.878	-	-	1163851	65.746
Naphthalene	337082	65.799	339412	51.018	673988	21.228	821416	31.497	365549	65.117	475078	110.906
Salicylic acid	495068	70.401	525441	50.78	1308028	68.034	1340236	61.483	-	-	339296	71.778

Table S14: JSD and WF over simulation trajectory of benzene molecule using models trained on aspirin, ethanol, naphthalene, and salicylic acid.

Table S15: Code Versions Details

Models	Code Version
NequIP	"0.5.6"
Allegro	"0.2.0"
BOTNet	"0.2.0"
MACE	"0.2.0"
Equiformer	commit "a4360ad"
TorchMDNet	"0.2.6"

Dataset	Train	Validation	Test	Reference
Acetylacetone	500	650	650	Batatia et al. (2022)
3BPA	527	1669	2138	Batatia et al. (2022)
MD17	950	50	1000	Fu et al. (2023)
LiPS	19000	1000	5000	Fu et al. (2023)

Table S16: Train-Val-Test Split details

References

Ilyes Batatia, Simon Batzner, Dávid Péter Kovács, Albert Musaelian, Gregor N. C. Simm, Ralf Drautz, Christoph Ortner, Boris Kozinsky, and Gábor Csányi. The design space of e(3)-equivariant atom-centered interatomic potentials, 2022.

Xiang Fu, Zhenghao Wu, Wujie Wang, Tian Xie, Sinan Keten, Rafael Gomez-Bombarelli, and Tommi S. Jaakkola. Forces are not enough: Benchmark and critical evaluation for machine learning force fields with molecular simulations. *Transactions on Machine Learning Research*, 2023. ISSN 2835-8856. URL `https://openreview.net/forum?id=A8pqQipwkt`. Survey Certification.

Stefan Goedecker, Michael Teter, and Jürg Hutter. Separable dual-space gaussian pseudopotentials. *Physical Review B*, 54(3):1703, 1996.

Stefan Grimme, Jens Antony, Stephan Ehrlich, and Helge Krieg. A consistent and accurate ab initio parametrization of density functional dispersion correction (dft-d) for the 94 elements h-pu. *The Journal of chemical physics*, 132(15), 2010.

Walter Kohn and Lu Jeu Sham. Self-consistent equations including exchange and correlation effects. *Physical review*, 140(4A):A1133, 1965.

Thomas D Kühne, Marcella Iannuzzi, Mauro Del Ben, Vladimir V Rybkin, Patrick Seewald, Frederick Stein, Teodoro Laino, Rustam Z Khaliullin, Ole Schütt, Florian Schiffmann, et al. Cp2k: An electronic structure and molecular dynamics software package-quickstep: Efficient and accurate electronic structure calculations. *The Journal of Chemical Physics*, 152(19), 2020.

Yi-Lun Liao and Tess Smidt. Equiformer: Equivariant graph attention transformer for 3d atomistic graphs. In *International Conference on Learning Representations*, 2023. URL `https://openreview.net/forum?id=KwmPfARgOTD`.

Shūichi Nosé. A molecular dynamics method for simulations in the canonical ensemble. *Molecular physics*, 52(2):255–268, 1984.

John P Perdew, Kieron Burke, and Matthias Ernzerhof. Generalized gradient approximation made simple. *Physical review letters*, 77(18):3865, 1996.

Philipp Thölke and Gianni De Fabritiis. Equivariant transformers for neural network based molecular potentials. In *International Conference on Learning Representations*, 2022. URL `https://openreview.net/forum?id=zNHzqZ9wrRB`.

Joost VandeVondele, Matthias Krack, Fawzi Mohamed, Michele Parrinello, Thomas Chassaing, and Jürg Hutter. Quickstep: Fast and accurate density functional calculations using a mixed gaussian and plane waves approach. *Computer Physics Communications*, 167(2):103–128, 2005.

Yuzhi Zhang, Haidi Wang, Weijie Chen, Jinzhe Zeng, Linfeng Zhang, Han Wang, and E Weinan. Dp-gen: A concurrent learning platform for the generation of reliable deep learning based potential energy models. *Computer Physics Communications*, 253:107206, 2020.