Supplementary Materials: Beyond Scaling: Chemical Intuition as Emergent Ability of Universal Machine Learning Interatomic Potentials

Shinnosuke Hattori
Advanced Research Laboratory
Sony Group Corporation
Atsugi, Kanagawa, Japan
shinnosuke.hattori@sony.com

Kohei Shimamura
Department of Physics
Kumamoto University
Kumamoto, Japan

Ken-ichi Nomura
Collaboratory for Advanced
Computing and Simulation
University of Southern California
Los Angeles, CA, USA
knomura@usc.edu

Aiichiro Nakano
Collaboratory for Advanced
Computing and Simulation
University of Southern California
Los Angeles, CA, USA

Rajiv K. Kalia
Collaboratory for Advanced
Computing and Simulation
University of Southern California
Los Angeles, CA, USA

Priya Vashishta
Collaboratory for Advanced
Computing and Simulation
University of Southern California
Los Angeles, CA, USA

Methods

1 Training Settings

The hyperparameters used for training the Allegro [1] model are detailed below.

- **Model Architecture (tensor size: 64)**
  - Two-body latent MLP: Dimensions [64, 128, 256], SiLU nonlinearity.
  - Later latent MLP: Dimensions [256, 256, 256], SiLU nonlinearity.
  - Embedding MLP: Linear projection.
  - Final edge energy MLP: Single hidden layer (dimension 128), no nonlinearity.
- **Initialization:** Uniform distribution with unit variance for MLPs.
- **Cutoff:** Radial cutoff of 5.2 Å.
- **Optimizer:** Adam with learning rate of 1e-3.
- **Batch Size:** The batch size was varied between 8 and 128, adjusted according to the size of the training dataset.
- **Weight Decay:** Default weight decay was used (value of 0 for ‘torch.optim.Adam‘).
- **Epochs:** Training was performed until convergence was reached, determined by the learning rate scheduler monitoring the validation loss. A minimum of 100 epochs was executed for all training runs, including those using the largest SPICE 2 dataset.

Preprint. Under review.

- **Loss Function:** Combined Root Mean Square Error (RMSE) of per-atom energy, forces, and stress.
- **Loss Weights:** Energy : Force : Stress = 8 : 1 : 1.

2 Dataset

The molecule targeted dataset utilizes SPICE version 2.0 [2]. This significantly expanded dataset focuses on drug-like small molecules, peptides, and solvated amino acids relevant for bio-molecular simulations, now containing approximately 2 million conformations across nearly 114,000 molecules. Notably, version 2.0 extends the chemical space to include Boron (B) and Silicon (Si) containing molecules, alongside improved sampling of non-covalent interactions, while retaining forces and multipole moments calculated at the ωB97M-D3(BJ)/def2-TZVPPD level of theory. Crucially, both dataset types underwent an energy offset alignment pre-processing step to ensure direct energetic comparability despite their different origins and calculation levels. Further details regarding the composition provided in Table 1.

For broader generalization studies, particularly to assess the emergence of physical scales from non-molecular data, we also used data from MatPES [3], specifically its r²SCAN functional calculations; MatPES also contributed to our mixed-dataset training. For direct analysis of reactive properties and detailed bond characteristics within stable versus transition states, we further used structures from the Transition1x (T1x) dataset [4], for which energies were recomputed (as detailed below) to align with the SPICE level of theory. The T1x dataset was mainly used as a benchmark test set for evaluating activation barriers and reaction energies.

For T1x dataset structures (reactants, transition states (TS), and products), we recalculated energies and forces using Psi4 [5] at the SPICE dataset's theory level (ωB97M-D3(BJ)/def2-TZVPPD).

Table 1: Training datasets, MPtrj, MatPES, OFF (SPICE 1), and SPICE 2 are datasets with structures generated using PBE, r²SCAN, and ωB97M-D3(BJ) methods respectively. N_{data} is the total number of structures.

Name	SPICE 2	OFF (SPICE 1)	MatPES	MPtrj	N_{data}
Method	ωB97M-D3(BJ)	←	r²SCAN	PBE	–
Num. of structure	2.0 M	1.0 M	0.4 M	1.6 M	–
SPICE 2[2]	✓				2.0 M
MatPES[3]			✓		0.4 M
Hybrid	✓		✓		2.4 M
Allegro-FM[6]		✓		✓	2.6 M

Table 2: Analysis datasets

	Type	Abbr.	N_{data}
Transition1x[4]	All	T1x	30 k
	Reactant	T1x:r	10 k
	Transition state	T1x:ts	10 k
	Product	T1x:pr	10 k

3 BDE evaluation

We inferred the multiplicity for each bond from the GAFF types of the two atoms forming it. Any unclear automated assignments were manually corrected, and we added 0.5 to the multiplicity of aromatic ring bonds for our analysis. We then used these bond multiplicities, along with the element types of the atoms, to categorize bonds for our "Type-Decomposition" in the E3D analysis. This

categorization was particularly used when calculating and comparing BDE metrics like Δ_{BDE} and σ_{BDE}.

Table 3 lists these typical BDEs for a variety of common chemical bond types, along with their standard bond orders and the primary sources for these energy values. These reference BDEs, primarily from experimental measurements with some established theoretical values where experiments are unavailable (details in [7, 8]), are fundamental, long-recognized measures of chemical bond strength.

Table 3: Literature BDEs referenced from Lange's Handbook[7][a] and theoretical calculations[8][b].

Bond	Bond order	Energy (eV)	Ref. structure
C–H	1	4.27[a]	$H-CH_2CH_3$
N–H	1	4.14[a]	$H-(NCH_3)_2$
O–H	1	4.55[a]	$H-OCH_3$
C–C	1	3.83[a]	H_3C-CH_3
	1.5	5.22[b]	benzene
	2	5.97[b]	$-(HC)=(CH)-$
	3	7.71[b]	$HC\equiv CH$
C–N	1	3.45[a]	$(CH_3)-NH_2$
	1.5	4.93[a]	(interpolated)
	2	6.41[a]	$(CH_2)=NH$
	3	8.02[a]	$C\equiv N$
C–O	1	3.93[a]	$HO-CH_3$
	2	7.63[a]	$H_2C=O$
N–O	1	2.18[a]	$HO-NCH_3$
	1.5	3.59[a]	(interpolated)
	2	5.01[a]	$HN=O$
N–N	1	2.22[a]	$H_2N-NHC_6H_5$
	1.5	3.48[a]	(interpolated)
	2	4.75[a]	$HN=NH$
	2.5	7.30[a]	(interpolated)
	3	9.84[a]	$N\equiv N$

4 Cross-Domain Validation

Table 4: Cross-domain validation: RMSE performance on Si-O-H test subsets

Trained Dataset	RMSE on MatPES			RMSE on SPICE 2	
	E (meV/atom)	F_{Si} / F_{O} (meV/Å)	P (meV/Å^3)	E (meV/atom)	F_{Si} / F_{O} (meV/Å)
MatPES	71	**293 / 339**	19	104	1080 / 327
SPICE 2	579	1326 / 1022	37	22	149 / 67
Hybrid	**50**	585 / 394	**13**	**6**	**107 / 59**

Beyond our main work on carbon-based bonds, we also explored how accurately our models predict energies for chemical systems that are uncommon in the training data, particularly those involving Si–O–H bonds. We conducted these evaluations using general test sets split from our training data, which were different from the specific t1x reaction systems we focused on previously, as shown in Table 4. Usually, models trained only on a single type of dataset are expected to perform better when the test data is similar to their training data. Remarkably, however, our study showed that models trained with the hybrid dataset performed better in certain aspects—like predicting energies for both the MatPES and SPICE datasets and forces for the SPICE dataset—than models trained exclusively on either MatPES or SPICE data alone. This key finding strongly suggests that the hybrid dataset learned to generalize more effectively by being exposed to a wider range of chemical information. This improved generalization, in turn, allowed it to accurately understand and represent even infrequently seen chemical structures, such as Si–O–H bonds.

References

[1] Albert Musaelian, Simon Batzner, Anders Johansson, Lixin Sun, Cameron J Owen, Mordechai Kornbluth, and Boris Kozinsky. Learning local equivariant representations for large-scale atomistic dynamics. *Nat. Commun.*, 14(1):579, February 2023.

[2] Peter Eastman, Benjamin P Pritchard, John D Chodera, and Thomas E Markland. Nutmeg and SPICE: Models and data for biomolecular machine learning. *J. Chem. Theory Comput.*, 20(19):8583–8593, October 2024.

[3] Aaron D Kaplan, Runze Liu, Ji Qi, Tsz Wai Ko, Bowen Deng, Janosh Riebesell, Gerbrand Ceder, Kristin A Persson, and Shyue Ping Ong. A foundational potential energy surface dataset for materials. *arXiv [cond-mat.mtrl-sci]*, March 2025.

[4] Mathias Schreiner, Arghya Bhowmik, Tejs Vegge, Jonas Busk, and Ole Winther. Transition1x - a dataset for building generalizable reactive machine learning potentials. *Sci. Data*, 9(1):779, December 2022.

[5] Daniel G A Smith, Lori A Burns, Andrew C Simmonett, Robert M Parrish, Matthew C Schieber, Raimondas Galvelis, Peter Kraus, Holger Kruse, Roberto Di Remigio, Asem Alenaizan, Andrew M James, Susi Lehtola, Jonathon P Misiewicz, Maximilian Scheurer, Robert A Shaw, Jeffrey B Schriber, Yi Xie, Zachary L Glick, Dominic A Sirianni, Joseph Senan O'Brien, Jonathan M Waldrop, Ashutosh Kumar, Edward G Hohenstein, Benjamin P Pritchard, Bernard R Brooks, Henry F Schaefer, 3rd, Alexander Yu Sokolov, Konrad Patkowski, A Eugene DePrince, 3rd, Uğur Bozkaya, Rollin A King, Francesco A Evangelista, Justin M Turney, T Daniel Crawford, and C David Sherrill. Psi4 1.4: Open-source software for high-throughput quantum chemistry. *J. Chem. Phys.*, 152(18):184108, May 2020.

[6] K Nomura, Shinnosuke Hattori, Satoshi Ohmura, Ikumi Kanemasu, Kohei Shimamura, Nabankur Dasgupta, A Nakano, R Kalia, and P Vashishta. Allegro-FM: Towards equivariant foundation model for exascale molecular dynamics simulations. *The Journal of Physical Chemistry Letters*, 0(0):6637–6644, February 2025.

[7] James Speight. *Lange's handbook of chemistry, seventeenth edition*. McGraw-Hill Education, Columbus, OH, 17 edition, October 2016.

[8] Kai Exner And and P Schleyer. Theoretical bond energies: A critical evaluation. *Journal of Physical Chemistry A*, 105:3407–3416, March 2001.