A shortcut towards simulating millions of atoms using machine learning potentials

Linh La [1] **Yugam Aggarwal** [2] **Truyen Tran** [1] **Svetha Venkatesh** [1] **Sherif Abdulkader Tawfik** [1]

[1]*Applied Artificial Intelligence Initiative, Deakin University, Geelong, Victoria 3216, Australia* [2]*School of Information Technology, Deakin University, Burwood, Victoria 3125, Australia*. Correspondence to: Linh La t.la@research.deakin.edu.au, Sherif Abdulkader Tawfik s.abbas@deakin.edu.au.

1. Introduction

This work answers the following question: *how can machine learning potentials simulate millions of atoms without requiring thousands of GPUs?*

Ab initio molecular dynamics (AIMD)[1, 2] simulations have become the gold standard for investigating material behaviour under varying temperature and pressure conditions, enabling the study of diverse dynamical phenomena including catalysis, ionic diffusion, phase transitions, and interfacial reactions. However, the cubic scaling of conventional density functional theory (DFT) calculation, $O(N^3)$, renders AIMD prohibitively expensive for systems exceeding ~1000 atoms or timescales beyond tens of picoseconds, severely limiting its applicability to realistic materials processes. Machine learning interatomic potentials (MLIPs) [3, 4] are *transforming* the landscape of molecular dynamics (MD) by achieving near-DFT accuracy whilst achieving orders-of-magnitude speedups. MLIPs based on Graph Neural Networks (GNNs) [5, 6] are the most accurate ML models for predicting individual DFT-based targets such as the formation energy/atom, electronic bandgap, mechanical properties, and vibrational properties (MegNet [7] and coNGN [8]), as well as the atomic forces and lattice stress (CHGNet [9] M3GNET [10], MACE [11], and DIEP [12]), achieving broad chemical coverage by training on large-scale datasets comprising hundreds of thousands of diverse structures.

However, graph-based architectures that explicitly encode many-body interactions—particularly three-body (triplet) terms—scale unfavourably with system size and cutoff radius: molecular system with thousands of atoms must be encoded as graphs with millions of duplets (two-body interactions) and billions of triplets, making it hard to simulate realistic system with millions of atoms. Recent advances have demonstrated remarkable scalability: Guo et al. [13] simulated 13.5 million Cu atoms at 11.2 ns/day, while Li et al. [14] achieved 149 ns/day for 0.5 million atoms, both using the DeePMD-kit framework [15]. However, these impressive results rely on system-specific models trained on AIMD trajectories of particular compositions and structures, limiting their transferability. The development of "universal" or "foundational" MLIPs—models capable of accurately predicting energies, forces, and stresses for arbitrary chemical systems—remains a critical challenge for the field. Moreover, recent work by Fu et al. [16] demonstrates that low errors in energies and forces alone do not guarantee accurate molecular dynamics (MD) trajectories. They showed that models performing well on static property prediction can exhibit substantial drift in thermodynamic and structural properties during extended MD simulations, highlighting the critical importance of dynamics-based validation.

In this work, we address the scalability challenge by decoupling the role of three-body interactions during learning versus simulation. Aiming to mitigate the computational and memory bottlenecks inherent in large-scale molecular systems, pruning technique is utilized for the triplet. Firstly, we perform a *training pruning* experiment to determine if three-body terms are strictly necessary for predicting static properties. Secondly, we assess *inference pruning* by subjecting the models to molecular dynamics simulations, thereby testing their dynamical reliability beyond standard energy/force benchmarks. Our findings delineate the specific representation requirements for scalable potentials and underscore the necessity of dynamics-based validation.

2. Results

2.1 Regression on static properties

We investigated whether the computational cost of GNN could be reduced by pruning three-body interactions. Using the M3GNet model [10, 17] and the MatPES-R2SCAN-2025.v1 dataset, we trained networks on graphs where triplet connections were randomly reduced by 20% to 100% (relative to a 4 Å three-body cutoff). Crucially, the two-body graph structure (5 Å cutoff) was kept constant to isolate the effect of angular terms. Table 1 demonstrates that the resulting Mean Absolute Errors (MAEs) did not fluctuate significantly with pruning. These findings suggest that for the properties considered, the angular information is largely redundant, allowing for aggressive pruning without degradation in static accuracy.

Table 1: Mean Absolute Error of test set with model trained on pruned percentage of triplets.

Cutoff percentage	E_f, meV/atom	Bandgap, meV	log(G)	log(K)
0	27	285	0.092	0.054
20	28	287	0.093	0.055
40	28	291	0.094	0.056
60	28	291	0.094	0.056
80	28	288	0.092	0.055
100	27	288	0.096	0.056

2.2 Molecular Dynamics

To evaluate the dynamical reliability of the models under pruning, we conducted MD simulations on ten structures randomly selected from the dataset: Tl_2SeO_4, K_3NbO_8, Nb_3RhSe_6, Sc_2Te_3, UF_5, Cs_2Sb, $Ho_2Te_5O_{13}$, $ScMnGe_2$, $EuIn_4$. For each structure, we constructed a 5×5×5 supercell and performed three independent NPT ensemble simulations at 300 K for 10 ps with a 1 fs timestep. To quantify structural preservation, we compared the Radial Distribution Functions (RDFs) of the final configurations against the unpruned M3GNet baseline. Thermodynamic stability was confirmed by monitoring temperature and potential energy evolution; all systems maintained structural coherence without exhibiting unphysical bond distortions or steric clashes.

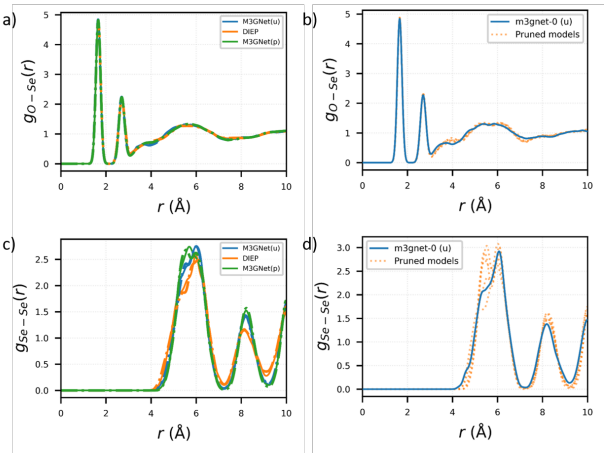

Fig. 1: Radial Distribution Functions (RDFs) of the O-Se and Se-Se pairs in Tl_2SeO_4. **(a, c)** Comparison of unpruned (continuous line) versus fully pruned (dashed line) structures during inference. **(b, d)** Comparison of RDFs across varying pruning percentages for models trained with pruned inputs. M3GNet-0 (u) denotes our retrained baseline model using full triplet interactions, whereas M3GNet(p) refers to the published model in [17].

Our findings regarding the impact of pruning are best exemplified by the O-Se and Se-Se interactions in Tl_2SeO_4:

- Inference Pruning: Pruning triplets during the MD inference phase preserved thermodynamic stability, with RDFs closely matching the full-graph baselines (Fig. 1a, 1c). This indicates that fully trained embeddings encode sufficient three-body information to maintain structural fidelity even when explicit triplet terms are removed. This is observed consistently across all ten studied structures and all pairs.
- Training Pruning: Conversely, models trained on pruned inputs exhibited significant structural deviations. While O-Se pairs remained stable (Fig. 1b), Se-Se pairs displayed a peak drift of ≈ 1 Å (Fig. 1d). We omit individual pruning percentages (20–100%) from the plots, as no consistent correlation was observed between the pruning ratio and the magnitude of structural drift. These structural instabilities were reproduced across the other analyzed systems and they vary between pairs of atoms.

These results suggest a decoupling of representation requirements: while triplet interactions are essential during training to learn the correct potential energy surface, the resulting model possesses a degree of redundancy that allows for efficient pruning during inference. While deep GNN architectures can theoretically pass information through message-passing layers using triplets, the explicit inclusion of three-body interactions in M3GNet acts as an inductive bias, allows the model to easily learn the potential energy surface. Once the potential energy model is fully trained, the explicit triplet terms become computationally unnecessary, suggesting that fully trained many-body models can be effectively distilled at inference time to improve scalability for large input structures, and indicates further potentially model compression.

3. Conclusion

We demonstrate a critical distinction between the representational capacity required to learn atomic forces versus that required to predict them during MD simulation. While three-body interactions are indispensable during training to capture the correct potential energy surface, they introduce redundant computational overhead during inference. By investigating on these phases, we achieve significant scalability gains without compromising thermodynamic stability or structural fidelity. These findings offer a pathway to universal, AIMD-level accuracy for million-atom systems without the need for system-specific retraining.

References

[1] Jürgen Hafner. Ab-initio simulations of materials using vasp: Density-functional theory and beyond. *Journal of computational chemistry*, 29(13):2044–2078, 2008.

[2] DA Andersson and Ben W Beeler. Ab initio molecular dynamics (aimd) simulations of nacl, ucl3 and nacl-ucl3 molten salts. *Journal of Nuclear Materials*, 568:153836, 2022.

[3] Bhupalee Kalita, Hatice Gokcan, and Olexandr Isayev. Machine learning interatomic potentials at the centennial crossroads of quantum mechanics. *Nature Computational Science*, pages 1–13, 2025.

[4] Xuan Zhang, Limei Wang, Jacob Helwig, Youzhi Luo, Cong Fu, Yaochen Xie, Meng Liu, Yuchao Lin, Zhao Xu, Keqiang Yan, Keir Adams, Maurice Weiler, Xiner Li, Tianfan Fu, Yucheng Wang, Alex Strasser, Haiyang Yu, YuQing Xie, Xiang Fu,

Shenglong Xu, Yi Liu, Yuanqi Du, Alexandra Saxton, Hongyi Ling, Hannah Lawrence, Hannes Stärk, Shurui Gui, Carl Edwards, Nicholas Gao, Adriana Ladera, Tailin Wu, Elyssa F. Hofgard, Aria Mansouri Tehrani, Rui Wang, Ameya Daigavane, Montgomery Bohde, Jerry Kurtin, Qian Huang, Tuong Phung, Minkai Xu, Chaitanya K. Joshi, Simon V. Mathis, Kamyar Azizzadenesheli, Ada Fang, Alán Aspuru-Guzik, Erik Bekkers, Michael Bronstein, Marinka Zitnik, Anima Anandkumar, Stefano Ermon, Pietro Liò, Rose Yu, Stephan Günnemann, Jure Leskovec, Heng Ji, Jimeng Sun, Regina Barzilay, Tommi Jaakkola, Connor W. Coley, Xiaoning Qian, Xiaofeng Qian, Tess Smidt, and Shuiwang Ji. Artificial intelligence for science in quantum, atomistic, and continuum systems. *Foundations and Trends in Machine Learning*, 18(4):385–912, 2025.

[5] TN Kipf. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.

[6] Patrick Reiser, Marlen Neubert, André Eberhard, Luca Torresi, Chen Zhou, Chen Shao, Houssam Metni, Clint van Hoesel, Henrik Schopmans, Timo Sommer, et al. Graph neural networks for materials science and chemistry. *Communications Materials*, 3(1):93, 2022.

[7] Chi Chen, Weike Ye, Yunxing Zuo, Chen Zheng, and Shyue Ping Ong. Graph networks as a universal machine learning framework for molecules and crystals. *Chemistry of Materials*, 31(9):3564–3572, 2019.

[8] Robin Ruff, Patrick Reiser, Jan Stühmer, and Pascal Friederich. Connectivity optimized nested graph networks for crystal structures, 2023.

[9] Bowen Deng, Peichen Zhong, KyuJung Jun, Janosh Riebesell, Kevin Han, Christopher J. Bartel, and Gerbrand Ceder. CHGNet as a pretrained universal neural network potential for charge-informed atomistic modelling. *Nat Mach Intell*, 5(9):1031–1041, September 2023.

[10] Chi Chen and Shyue Ping Ong. A universal graph deep learning interatomic potential for the periodic table. *Nat Comput Sci*, 2(11):718–728, November 2022.

[11] Ilyes Batatia, Philipp Benner, Yuan Chiang, Alin M Elena, Dávid P Kovács, Janosh Riebesell, Xavier R Advincula, Mark Asta, Matthew Avaylon, William J Baldwin, et al. A foundation model for atomistic materials chemistry. *The Journal of chemical physics*, 163(18), 2025.

[12] Sherif Abdulkader Tawfik, Tri Minh Nguyen, Salvy P. Russo, Truyen Tran, Sunil Gupta, and Svetha Venkatesh. Embedding material graphs using the electron-ion potential: application to material fracture. *Digital Discovery*, 3:2618 – 2627, 2024. Version Number: 3.

[13] Zhuoqiang Guo, Denghui Lu, Yujin Yan, Siyu Hu, Rongrong Liu, Guangming Tan, Ninghui Sun, Wanrun Jiang, Lijun Liu, Yixiao Chen, et al. Extending the limit of molecular dynamics with ab initio accuracy to 10 billion atoms. In *Proceedings of the 27th ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming*, pages 205–218, 2022.

[14] Jianxiong Li, Boyang Li, Zhuoqiang Guo, Mingzhen Li, Enji Li, Lijun Liu, Guojun Yuan, Zhan Wang, Guangming Tan, and Weile Jia. Scaling molecular dynamics with ab initio accuracy to 149 nanoseconds per day. In *SC24: International Conference for High Performance Computing, Networking, Storage and Analysis*, pages 1–15. IEEE, 2024.

[15] Jinzhe Zeng, Duo Zhang, Anyang Peng, Xiangyu Zhang, Sensen He, Yan Wang, Xinzijian Liu, Hangrui Bi, Yifan Li, Chun Cai, et al. Deepmd-kit v3: a multiple-backend framework for machine learning potentials. *Journal of Chemical Theory and Computation*, 21(9):4375–4385, 2025.

[16] Xiang Fu, Zhenghao Wu, Wujie Wang, Tian Xie, Sinan Keten, Rafael Gomez-Bombarelli, and Tommi Jaakkola. Forces are not enough: Benchmark and critical evaluation for machine learning force fields with molecular simulations. *Transactions on Machine Learning Research*, 2023.

[17] 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 preprint arXiv:2503.04070*, 2025.