Platonic representation of foundation machine learning interatomic potentials

Zhenzhu Li[1,2] Aron Walsh[1]

[1]Department of Materials, Imperial College London, SW72AZ, London, UK [2]Imperial Global Singapore, 138402, Singapore

Correspondence to: Zhenzhu Li zhenzhu.li@imperial.ac.uk

1. Introduction

Machine learning interatomic potentials (MLIPs) offer a powerful tool to accelerate the inference of energy (E), forces (F), and stresses (σ) in atomistic structures. The development of MLIPs has evolved from early data-fitting approaches using crystal graphs to recent architectural designs incorporating atom-centred expansions with high body-order, message passing and equivariance[1-7]. In this work, we demonstrate that foundation MLIPs converge toward a shared, architecture independent latent geometry, termed the Platonic representation.

2. Main findings

We introduce an anchor based projection framework to unify the embeddings of various foundation models, enabling interoperability that is otherwise mathematically inaccessible. Using this unified space, we apply cross-model optimal transport analysis to quantify the geometric distance between pre-trained potentials. Furthermore, we develop an embedding arithmetic scheme that consistently represents materials and reactions across different architectures. Finally, we suggest that deviations from this Platonic geometry can potentially serve as a diagnostic tool, effectively highlighting training divergence and physical inconsistencies.

3. Methodology

Aligning incompatible representations We chose seven foundation MLIPs representing distinct architectures, datasets, and approaches to equivariance and energy conservation. These include three MACE-MP-0 variants (Large, Medium, and Small) trained on the Materials Project Trajectory Dataset (MPtrj); two OMat24-based models (MACE-omat and Seven-omat); and two Orb-v3 models (Orb-v3-con-omat and Orb-v3-dir-omat), which differ in their treatment of force conservation. For each model, we extracted 282,847 atomic embeddings across 27,136 structures from the MP-20 dataset. We applied principal component analysis (PCA) to project these embeddings into a two-dimensional space, where the first two principal components (PCA1 and PCA2) capture the directions of greatest variance. While we also examined nonlinear visualisation techniques such as uniform manifold approximation and projection (UMAP) , but they distort global geometry and reconfigure relative distances. We found that original representations act as arbitrary coordinate systems learned by each model architecture.

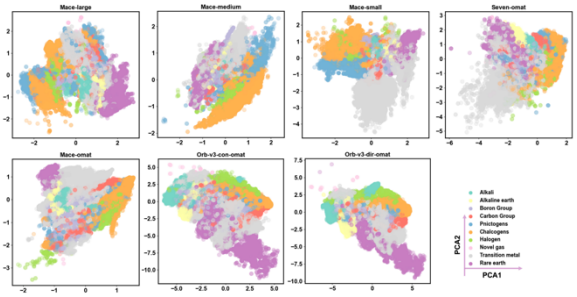

Fig. 1: Model-specific embeddings are incompatible before transformation. 2DPCA projections of atomic embeddings from seven foundation MLIPs reveal distinct variance directions and element clustering patterns. Although all models are trained to predict the same physical quantities for overlapping material sets, they learn embeddings in incompatible coordinate systems.

Convergence to a shared chemical geometry For alignment, we first project the model-specific embeddings into a unified latent space. We achieve this using the relative representation strategy proposed by Moschella et al.[8], establishing a common coordinate frame independent of the original architecture. We use cosine similarity instead of distance functions (Euclidean L2 or Manhattan L1) as it provides scale invariance—critical when comparing embeddings with varying norms. We select a set of K anchor vectors, $\{a_1, a_2, \ldots, a_K\}$, from the embedding manifold. The Platonic transformation, $T(e_i)$, projects an embedding (e_i) into the anchor-defined space based on its cosine similarity to these reference points. The resulting vector $z_i \in R^K$ constitutes the embedding in the unified Platonic space.

As shown in Fig. 2, anchors act as stable reference points defining a shared coordinate system. Once the principal geometric relationships are pinned to these anchors, the

remaining embeddings naturally align, reflecting the same underlying physical drivers regardless of the model origin. Furthermore, the evolution of the embedding topology with increasing K reveals a hierarchical organisation within the chemical space. While a small number of anchors captures the coarse global structure, increasing the anchor density resolves finer structural details.

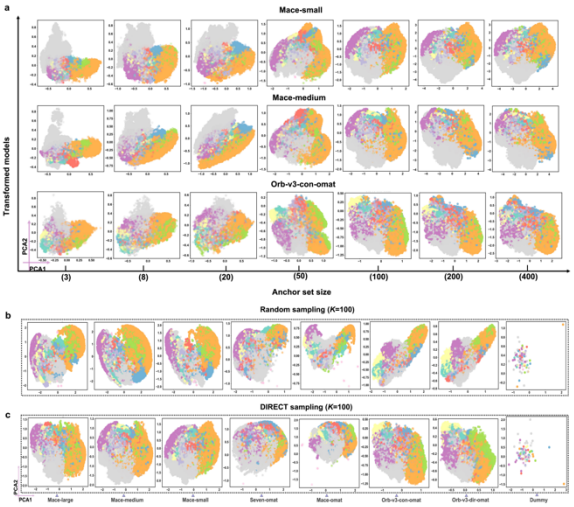

Fig. 2: Variation of transformed representations with anchor set size and sampling strategy. (a) Transformed representations as a function of anchor set size (K = 3 to 400). (b) 2D PCA projections of converged representations using 100 randomly sampled anchors. (c) Projections using 100 DIRECT-sampled anchors. Despite architectural diversity, all models transformed with DIRECT sampling show substantial alignment. Non-equivariant models (Orb-v3) exhibit systematic skewness. The Dummy model (untrained, random weights) displays no chemical structure, confirming that alignment reflects learned physical knowledge. Colourmap follows the labelling in Fig. 1.

3.1 Related work

Edamadaka et al. reported universally converging representations of matter across scientific foundation models using statistical analysis, and Chorna et al. compared the latent features of models focusing on the dissimilarity[9-10]. These two works appear almost simultaneously with our work, indicating a global research attention on the convergence of neural networks for materials. This work provided a special angle to unify the representations, presenting the converged picture and also applying the Platonic representation in model stitching and cross-model arithmetic.

3.2 Algebraic consistency and model stitching

A potential utility of a Platonic representation is embedding arithmetic across different models. By mapping into a unified coordinate system, vector operations can be performed. We evaluate this through three case studies: a complex oxide ($Na_3MnCoNiO_6$), a symmetry sensitive polymorphic pair of TiO_2 , and a solid-state reaction to synthesise $BaTiO_3$.

Table 1: Cross-model embedding arithmetic. Vector norms (l) and cosine similarities (c-sim, relative to MACE-large) for: (i) $Na_3MnCoNiO_6$ material embeddings; (ii) TiO_2 polymorph differences; and (iii) $BaTiO_3$ formation reaction vectors (standard vs. zero-shot stitched). i-sim denotes intra-model similarity between polymorphs.

Model	z_{Mater} (l, c-sim)	z_{Morph} (l_1, l_2, i-sim, c-sim)	z_{React} (l, c-sim)	$z_{React-stitch}$ (l, c-sim)
MACE-large	1.53, 1.00	1.51, 1.59, 0.97, 1.00	1.26, 1.00	1.39, 1.00
MACE-medium	1.05, 0.84	1.21, 1.43, 0.95, 0.46	1.08, 0.69	1.69, 0.92
MACE-small	1.14, 0.87	1.62, 1.53, 0.97, 0.40	0.99, 0.85	0.99, 0.88
Seven-omat	4.50, 0.79	5.29, 5.50, 1.00, 0.14	5.58, 0.77	3.72, 0.38
MACE-omat	5.44, 0.79	5.88, 6.15, 1.00, 0.47	5.74, 0.75	4.33, 0.43
Orb-v3-con	0.68, 0.54	1.56, 1.83, 0.99, 0.36	1.05, 0.48	2.28, 0.82
Orb-v3-dir	1.28, 0.79	2.01, 2.30, 0.99, 0.39	1.64, 0.73	1.90, 0.67

For example, We define the reaction embedding as $z_{React} = z_{products} - z_{reactants}$. For the formation of BaTiO3 from its binaries, we observe consistency across models (~> 0.7 c-sim, except the Orb-v3-con model). Embedding compatibility can support zero-shot model stitching. This allows us to algebraically substitute the product state representation of one model with that of another, treating them as compatible vectors within the shared geometry. We constructed a hybrid reaction embedding, $z_{React-stitch}$, by pairing reactant embeddings from MACE-large with product embeddings from other models. As detailed in Table 1, inter-model compatibility is high. MACE-MP-0 variants show strong agreement (> 0.88). Surprisingly, Orb-v3-con-omat exhibits higher stitchability with MACE-large (0.82) than the other OMat24 trained models. This demonstrates that models trained on non-overlapping datasets (MPtrj vs. OMat24) can be algebraically combined to yield geometrically reasonable embeddings, opening potential routes for modular reuse of pre-trained foundation potentials.

Notes

Preprint: Zhenzhu Li and Aron Walsh, https://arxiv.org/abs/2512.05349.

Acknowledgments

We thank Lars Schaaf and Kinga Mastej for useful discussions and suggestions related to

embedding analysis and the chemical consequences. We are grateful to the UK Materials and Molecular Modelling Hub for computational resources, which is partially funded by EPSRC (EP/T022213/1, EP/W032260/1 and EP/P020194/1). We thank the EPSRC for support via the AI for Chemistry: AIchemy hub (EPSRC grant EP/Y028775/1 and EP/Y028759/1).

References

[1] Chen, C.; Ong, S. P. A universal graph deep learning interatomic potential for the periodic table. Nature Computational Science 2022, 2, 718–728.

[2] Kov´acs, D. P.; Oord, C. v. d.; Kucera, J.; Allen, A. E. A.; Cole, D. J.; Ortner, C.;Csányi, G. Linear Atomic Cluster Expansion Force Fields for Organic Molecules: Beyond RMSE. Journal of Chemical Theory and Computation 2021, 17, 7696–7711.

[3] Batatia, I.; Batzner, S.; Kov´acs, D. P.; Musaelian, A.; Simm, G. N. C.; Drautz, R.; Ortner, C.; Kozinsky, B.; Cs´anyi, G. The design space of E(3)-equivariant atom-centred interatomic potentials. Nature Machine Intelligence 2025, 7, 56–67.

[4]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. Nature Communications 2022, 13, 2453.

[5] Park, Y.; Kim, J.; Hwang, S.; Han, S. Scalable Parallel Algorithm for Graph Neural Network Interatomic Potentials in Molecular Dynamics Simulations. Journal of Chemical Theory and Computation 2024, 20, 4857–4868.

[6] Rhodes, B.; Vandenhaute, S.; ˇSimkus, V.; Gin, J.; Godwin, J.; Duignan, T.; Neumann, M. Orb-v3: atomistic simulation at scale. 2025; https://arxiv.org/abs/2504.06231.

[7] Bigi, F.; Langer, M.; Ceriotti, M. The dark side of the forces: assessing non-conservative force models for atomistic machine learning. 2025; https://arxiv.org/abs/2412.11569.

[8] Moschella, L.; Maiorca, V.; Fumero, M.; Norelli, A.; Locatello, F.; Rodola, E. Relative representations enable zero-shot latent space communication. ICLR. 2023.

[9] Chorna, S.; Tisi, D.; Malosso, C.; How, W. B.; Ceriotti, M.; Chong, S. Comparing the latent features of universal machine-learning interatomic potentials. 2026; https://arxiv.org/abs/2512.05717.

[10]Edamadaka, S.; Yang, S.; Li, J.; G´omez-Bombarelli, R. Universally Converging Representations of Matter Across Scientific Foundation Models. 2025; https://arxiv.org/abs/2512.03750.