The Journal of Chemical Physics

RESEARCH ARTICLE | FEBRUARY 13 2026

Resolving the body-order paradox of machine learning interatomic potentials

Special Collection: Festschrift in honor of Christoph Dellago: Exploring Paths and Barriers in Statistical Mechanics

Sanggyu Chong ; Tong Jiang ; Michelangelo Domina ; Filippo Bigi ; Federico Grasselli ; Joonho Lee ; Michele Ceriotti

J. Chem. Phys. 164, 064121 (2026)
https://doi.org/10.1063/5.0303302

AIP Publishing

The Journal of Chemical Physics ARTICLE pubs.aip.org/aip/jcp

Resolving the body-order paradox of machine learning interatomic potentials

Cite as: J. Chem. Phys. 164, 064121 (2026); doi: 10.1063/5.0303302
Submitted: 20 September 2025 • Accepted: 20 January 2026 •
Published Online: 13 February 2026

Sanggyu Chong,[1,a)] Tong Jiang,[2] Michelangelo Domina,[1] Filippo Bigi,[1] Federico Grasselli,[1,b)] Joonho Lee,[2] and Michele Ceriotti[1]

AFFILIATIONS

[1] Laboratory of Computational Science and Modeling, Institute of Materials, École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland
[2] Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts 02138, USA

Note: This paper is part of the Special Topic Festschrift in Honor of Christoph Dellago: Exploring Paths and Barriers in Statistical Mechanics.
[a)] **Author to whom correspondence should be addressed:** sanggyu.chong@epfl.ch
[b)] **Present address:** Dipartimento di Scienze Fisiche, Informatiche e Matematiche, Università degli Studi di Modena e Reggio Emilia, 41125 Modena, Italy.

ABSTRACT

In many cases, the predictions of machine learning interatomic potentials (MLIPs) can be interpreted as a sum of body-ordered contributions, which is explicit when the model is directly built on neighbor density correlation descriptors and is implicit when the model captures the correlations through the non-linear functions of low body-order terms. In both cases, the "effective body-orderedness" of MLIPs remains largely unexplained: how do the models decompose the total energy into body-ordered contributions, and how does their body-orderedness affect the accuracy and learning behavior? In answering these questions, we first discuss the complexities in imposing the many-body expansion on *ab initio* calculations at the atomic limit. Next, we train a curated set of MLIPs on datasets of hydrogen clusters and reveal the inherent tendency of the ML models to deduce their own, effective body-order trends, which are dependent on the model type and dataset makeup. Finally, we present different trends in the convergence of the body-orders and generalizability of the models, providing useful insights into the development of future MLIPs.

I. INTRODUCTION

The many-body expansion (MBE) expresses the global observable of a chemical system as a sum of contributions from interactions at different body-orders, where the "bodies" are taken to be atoms, molecules, or larger fragments in the system. The MBE enables the interpretation of complex interactions in terms of simpler body-ordered contributions and can provide reasonable approximations to the global quantity, especially when the contributions decay rapidly with the number of bodies involved. The MBE has given rise to many local or fragment-based quantum chemistry methods[1–9] that offer favorable scaling for large systems, as well as force fields that can be used for atomic scale simulations at large length and time scales.[10–12]

Machine learning interatomic potentials (MLIPs)[13–17] enable *ab initio*-quality atomistic simulations with linear system size scaling and low prefactors, further extending the accessible length and time scales of the simulations. MLIPs commonly adopt a locality ansatz based on the nearsightedness principle[18,19] and represent the local environments of chemical systems in terms of body-ordered correlations between the central atom and its neighbors.[20–22] This allows MLIPs to effectively capture the complex many-body interactions while retaining favorable scalability and transferability.

Here, one could draw a parallel between the MBE interpretation of physical observables and the correlation-based atomic representations of common MLIPs. Within this parallel, a paradox also emerges: how do the MLIPs make accurate predictions with a limited set of atomic body-order correlations, despite the MBE being exact

only in the limit of all body-ordered contributions in the system (with their count approaching infinity in bulk systems)? This highlights a deeper gap in our understanding of the "body-orderedness" of MLIPs and its implications for their performance and learning behavior, especially for recent graph neural network (NN)-based models where the body-ordered correlations are implicit and their contributions are non-separable.

In this study, we systematically analyze the body-orderedness of NN-based MLIPs and the resulting learning behavior. First, we revisit the trends in the MBE of *ab initio* calculations, which commonly serve as a reference for machine learning interatomic potential (MLIP) training, at the atomic limit. Next, we train a curated set of MLIPs on datasets of hydrogen clusters to investigate how the models infer body-order trends of hydrogen[23,24] and how they compare to the reference. We further explore how the trends vary between the different MLIPs and dataset makeup and discuss how they may affect model accuracy and learning behavior. Finally, we consider the relationship between the body-ordered energetics and the out-of-distribution accuracy or generalizability of MLIPs.

II. THEORY AND METHODS

A. Parallelism between the MBE and the body-orderedness of MLIPs

In applying the MBE at the atomic limit, the total energy E_A of system A with N atoms is expressed as

$$E_A(\mathbf{r}_1,\mathbf{r}_2,\ldots,\mathbf{r}_N) = \sum_i V^{(1)} + \sum_{i<j} V^{(2)}(\mathbf{r}_i,\mathbf{r}_j) + \sum_{i<j<k} V^{(3)}(\mathbf{r}_i,\mathbf{r}_j,\mathbf{r}_k) + \cdots + V^{(N)}(\mathbf{r}_1,\mathbf{r}_2,\ldots,\mathbf{r}_N), \tag{1}$$

for atoms $i,j,k \ldots \in A$ and their coordinates $\mathbf{r}$. The summations run over the canonically complete set of lower body-ordered "sub-clusters" that can be identified within A. The mth body-ordered energy contribution $V^{(m)}$ from a sub-cluster with m atoms can be expressed in terms of its total energy $E^{(m)}$ minus all the lower body-order terms,

$$V^{(m)}(\mathbf{r}_1,\mathbf{r}_2,\ldots,\mathbf{r}_m) = E^{(m)}(\mathbf{r}_1,\mathbf{r}_2,\ldots,\mathbf{r}_m) - \sum_{1\le k<m} \sum_{i'<j'<\ldots<k'} V^{(k)}(\mathbf{r}_{i'},\mathbf{r}_{j'},\ldots,\mathbf{r}_{k'}). \tag{2}$$

In the second term, the inner sum is the same as those in Eq. (1), and the outer sum runs over all body-orders k lower than m. Here, we interpret $V^{(1)}$ to be E_1, the energy of an isolated atom. For $m = 3$,

$$V^{(3)}(\mathbf{r}_1,\mathbf{r}_2,\mathbf{r}_3) = E^{(3)}(\mathbf{r}_1,\mathbf{r}_2,\mathbf{r}_3) - 3E_1 - V^{(2)}(\mathbf{r}_1,\mathbf{r}_2) - V^{(2)}(\mathbf{r}_1,\mathbf{r}_3) - V^{(2)}(\mathbf{r}_2,\mathbf{r}_3). \tag{3}$$

The number of canonically complete sub-clusters for a given m is $N!/(m!(N-m)!)$ and is the largest for $m \approx N/2$. The body-ordered (negative) *force* contribution $\partial V^{(m)}/\partial \mathbf{r}_i$ can be further derived from Eq. (2) as follows:

$$\frac{\partial V^{(m)}}{\partial \mathbf{r}_i} = \frac{\partial E^{(m)}}{\partial \mathbf{r}_i} - \sum_{2\le k<m} \sum_{i''<j''<\ldots<k''} \frac{\partial V^{(k)}(\mathbf{r}_{i''},\mathbf{r}_{j''},\ldots,\mathbf{r}_{k''})}{\partial \mathbf{r}_i}. \tag{4}$$

$\partial V^{(m)}/\partial \mathbf{r}_i$ is a local quantity for atom i, and the inner summation runs over the neighboring atoms of i.

The locality ansatz of MLIPs leads to the following expression for the predicted total energy $\tilde{E}_A$:

$$\tilde{E}_A = \sum_{i\in A} \varepsilon_i = \sum_{i\in A} \varepsilon_i(\{\mathbf{r}_{ij}\}_{j\in A_i}). \tag{5}$$

ε_i is the "local" energy associated with atom i, and it is a function of the local geometry described via $\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j$ for all j in A_i, the local environment of atom i. The MLIPs differ in the functional form used to predict ε_i, which usually involves combining multiple $\mathbf{r}_{ij}$ terms to describe higher body-order correlations.[21,25] Among the models, there exists a subset[15,16] where features that formally depend on a fixed number of neighbors are used as the polynomial basis for a linear expansion,

$$\varepsilon_i = \sum_\nu \boldsymbol{\phi}_\nu(\{\mathbf{r}_{ij}\}_{j\in A_i}) \cdot \mathbf{w}_\nu. \tag{6}$$

Here, ν is the "correlation order" involving multiple neighbor atoms j (then, body order equals $\nu+1$). The expression tells us that the predicted energy can be explicitly decomposed into body-ordered contributions. Despite the apparent parallel with the MBE [Eq. (1)], previous studies[26–28] have shown that there does not exist a 1-to-1 correspondence between the two because $\boldsymbol{\phi}_\nu$ also contains terms associated with lower body-order correlations, unless they are explicitly elimintated[27] (see Appendix A).

In this study, we focus on the NN-based MLIPs, which have shown greater accuracies for much larger portions of the chemical space,[29–32] yet their body-orderedness has never been considered in detail. For the NN-based models, ε_i often involves nonlinearities that make it challenging to derive an analytical connection to the MBE. One of such models is the Behler–Parrinello NN (BPNN),[13] where the local descriptors of 2- and 3-body correlations are taken as inputs to a fully connected feed-forward network that predicts ε_i, which are then summed across the system to obtain the total energy. Here, we specifically consider SOAP-BPNN, which adopts the Smooth Overlap of Atomic Position (SOAP)[33] as the local descriptor of choice. Starting from the 3-body SOAP descriptors, SOAP-BPNN *implicitly* reaches higher body-orders through multiple NN layers and nonlinear activation functions. Since SOAP is an incomplete descriptor,[34] the resulting SOAP-BPNN descriptor also exhibits incompleteness at higher body-orders.

More recent MLIPs incorporate message-passing NNs[35] and transformers[36] in their architectures. One example that we consider is MACE,[37] an equivariant message-passing NN-based MLIP grounded in the atomic cluster expansion (ACE) formalism.[16] In MACE, each layer describes the body-order correlations of the given atomic environment up to $\nu = 3$. Equivariant message passing effectively introduces a new one-particle basis, raising ν by 1.[25] With two of such layers, MACE *explicitly* achieves a maximum body-order of 13.[38] Finally, we consider the point-edge

transformer (PET),[39] a transformer-based model that does not enforce exact rotational symmetry. In PET, the `softmax` function used in the attention mechanism, as well as the activation functions in the feedforward blocks, *implicitly* leads to a theoretically infinite body-order.

Even though there is no explicit decomposition of the predictions of these nonlinear NN-based MLIPs into body-ordered contributions, the connection is quite strong, with low-order correlations being used as inputs to internal operations that ultimately raise the body-order perceived by the model. It is, therefore, interesting to compute, *empirically* through Eqs. (2) and (4), how the MBE of the trained models compares to that of the reference electronic structure methods, how it depends on the training details, and how it affects the transferability of the models.

B. Hydrogen cluster sampling and calculation details

In this study, we construct and utilize the datasets of hydrogen octamers (8-mers) for MLIP training and analysis. Given the simple electronic structure of hydrogen, this ensures that the analysis can be focused on the body-order-specific effects as much as possible and reduces the cost of calculations that probe the body-ordered energetics beyond the density functional theory (DFT) level (see Sec. III).

The 8-mers are sampled from the bulk hydrogen datasets of Cheng *et al.*[23] We construct two distinct datasets: a "high density" (high ρ) dataset, in which the clusters are sampled from 100 configurations with the highest density (average ρ = 1.34 g/cm^3), and a "low density" (low ρ) dataset, where the clusters are sampled from 100 configurations with the lowest density (average ρ = 0.461 g/cm^3). In the high ρ dataset, each 8-mer is sampled by choosing a hydrogen at random from a given configuration and then taking its seven nearest neighbors. In the low ρ dataset, sampling is performed by taking a random hydrogen atom and its closest neighbor and then adding three nearest *pairs* of atoms to complete the 8-mer. Example clusters are shown in Fig. 1. Such a difference in the sampling protocol ensures the sampling of distinct chemical trends seen in the corresponding bulk phases, where high ρ systems are atomic, covalently bound, and metallic, and low ρ systems are molecular, bound by non-covalent interactions, and insulating. From each dataset, we randomly select 500 clusters, for which we also create the accompanying datasets containing canonically complete sets of their sub-clusters.

To obtain the energies and forces for MLIP training, DFT calculations are performed using FHI-aims.[40] The PBE exchange–correlation functional[41] is employed with the "tight" species-default basis set, and Gaussian smearing with σ = 0.025 eV is used to determine the occupations. To assess the body-ordered energetics of hydrogen clusters at a much higher level of theory, we also perform spin-adapted density matrix renormalization group (DMRG) calculations[42] using BLOCK2[43] on a few representative clusters, using one- and two-electron integrals generated by PYSCF.[44,45] During method investigation, coupled-cluster with single, double, and triple excitations (CCSDT) calculations were also carried out in Q-CHEM[46] for internal checks. Further details of the cluster sampling protocol and the reference calculations are given in the supplementary material.

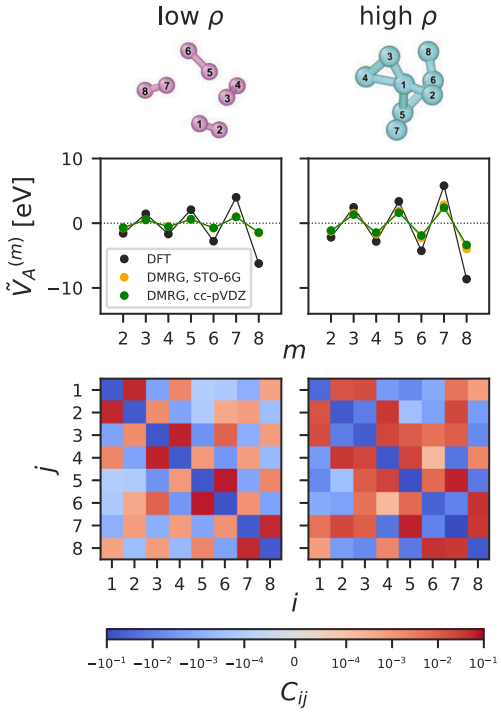

FIG. 1. Body-ordered energetics and spin–spin correlations for a sample low ρ 8-mer (left) and a sample high ρ 8-mer (right). The atomic configurations of the sampled clusters are presented in the first row. Plots of the body-ordered energetics probed by $\tilde{V}_A^{(m)}$ of Eq. (7) are presented in the second row for the DFT and DMRG calculations. The spin–spin correlation results in the third row are obtained from DMRG calculations and are plotted in a symmetric log scale with a linearity threshold of 10^{-4}. Atomic indices used in these plots are shown in the first row.

C. MLIP training details

We consider three NN-based models widely varying in their architectures: SOAP-BPNN, MACE, and PET. In all cases, models are trained via stochastic gradient descent with the Adam optimizer[47] on the mean squared loss of energies and forces. Train, validation, and test splits of the datasets are kept consistent between the models. We set a consistent cutoff radius of 5.5 Å in all three models, which corresponds to the value used in the MLIPs developed in the previous work and encompasses the entire cluster.[23] In SOAP-BPNN and MACE, the baseline energy, i.e., the energy of the isolated atom as perceived by the model, is fixed to the reference isolated atom energy. In PET, the existence of central tokens and their incorporation into the attention mechanism prevent the model from having a strict, pre-defined baseline. Instead, we take a data-driven approach and include an isolated atom reference configuration in the training sets for PET, allowing the model to learn an isolated atom energy to a good accuracy (<0.01 eV from the reference value). We note that any deviation in the learned isolated atom energy from the reference may give PET a small advantage in the learning tasks. We have seen, however, that allowing the MLIPs to adjust freely their baseline energy does not significantly impact our observations (see the supplementary material). The ratio between the energy and force losses is fixed at 1:1 for SOAP-BPNN and PET. For MACE,

we adopt its default training routine, which is to use a 1:100 ratio between energy and force losses in the first stage and 1000:100 in the second stage. The rest of the training details and hyperparameters are set to the defaults of the models. Full sets of the inputs and hyperparameters are provided in our Materials Cloud repository.[48]

III. BODY-ORDERED ENERGETICS OF HYDROGEN CLUSTERS

Before any MLIP analysis, we first evaluate the body-ordered energetics of hydrogen 8-mers in the reference *ab initio* calculations. In many cases, the MBE formalism is applied at the *molecular* level,[49–51] where $V^{(m)}$ is expected to converge quickly to zero with increasing m. An archetypal system is water,[52–61] for which many-body force fields have shown good accuracies with limited body-orders, up to four, including also a polarizable model baseline. Even for water, however, body-ordered energetics at the DFT theory-level can converge slowly and show large oscillations when ions are introduced.[62,63] In the context of MLIPs and their body-orders, where body-ordered correlations or other geometric features are computed between atoms, the consistent scale at which to apply the MBE formalism is also that of individual *atoms*. At the atomic level, where covalent and metallic interactions dominate, even more complex and non-trivial trends have been previously observed for mercury, sodium, silicon, and gold.[64,65]

Here, we quantify the atomic MBE convergence for the hydrogen 8-mers used in this study. For all body-orders from 2 to 8, we compute

$$\tilde{V}_A^{(m)} = \frac{\sum_{i<j<\ldots<m} V^{(m)}(\mathbf{r}_i, \mathbf{r}_j, \ldots, \mathbf{r}_m)}{\frac{N!}{m!(N-m)!} \cdot m}. \tag{7}$$

For each m, energy contributions from the canonically complete set of m-mers are summed and then normalized by their total count and m to yield the average body-ordered energy contribution, per atom. This allows for a "fair" comparison between the body-ordered contributions without any effects associated with the number of sub-clusters, which is largest when $m \approx N/2$. The resulting trends for sample 8-mers from low ρ and high ρ datasets are shown in Fig. 1.

In both cases, an oscillatory and divergent trend is observed in the DFT body-ordered energetics (black markers). A negative energy contribution is first observed at $m = 2$ and then a larger positive contribution at $m = 3$, and the sign of $\tilde{V}_A^{(m)}$ continues to alternate with the magnitude increasing with m. The anticipated chemical trend between the low ρ and high ρ 8-mers is manifested as a difference in the magnitudes of $\tilde{V}_A^{(m)}$ across all m, with larger contributions observed for the high ρ 8-mers over the low ρ 8-mers. Similar trends persist with the exact correlation treatment using DMRG, revealing that the apparent trend is not a mere consequence of the approximate nature of DFT[63] (see Appendix B for further consideration of the DFT self-interaction error).

Previously, long-range many-body interactions have been observed in 1D hydrogen chains of up to 50 atoms,[66] where strong antiferromagnetic (AFM) spin–spin correlations are present. Here, we also analyze spin–spin correlations C_{ij} for representative 8-mers in three spatial dimensions, where $C_{ij} = \langle \hat{n}_{i\uparrow}\hat{n}_{j\downarrow}\rangle - \langle \hat{n}_{i\uparrow}\rangle\langle \hat{n}_{j\downarrow}\rangle$. Figure 1 shows that both low and high ρ 8-mers display substantial spin–spin correlations across all atomic pairs, thereby contributing to the apparent non-convergence of the body-ordered energetics. The correlation patterns reveal distinct behaviors between the two density regimes: in the low ρ 8-mer, strong correlations are present between bonded atom pairs, and non-negligible correlations also persist between the intermolecular pairs. The high ρ 8-mer exhibits stronger and more delocalized spin correlations across all pairs within the cluster.

These results demonstrate that the "true" body-ordered energetics of hydrogen 8-mers are inherently oscillatory and non-converging. We note that the contrast between our results and the mathematically proven exponential convergence of body-orders[67] arises from the choice of E_1 in the MBE. While our expansion references itself to the "vacuum" (i.e., each sub-cluster is considered

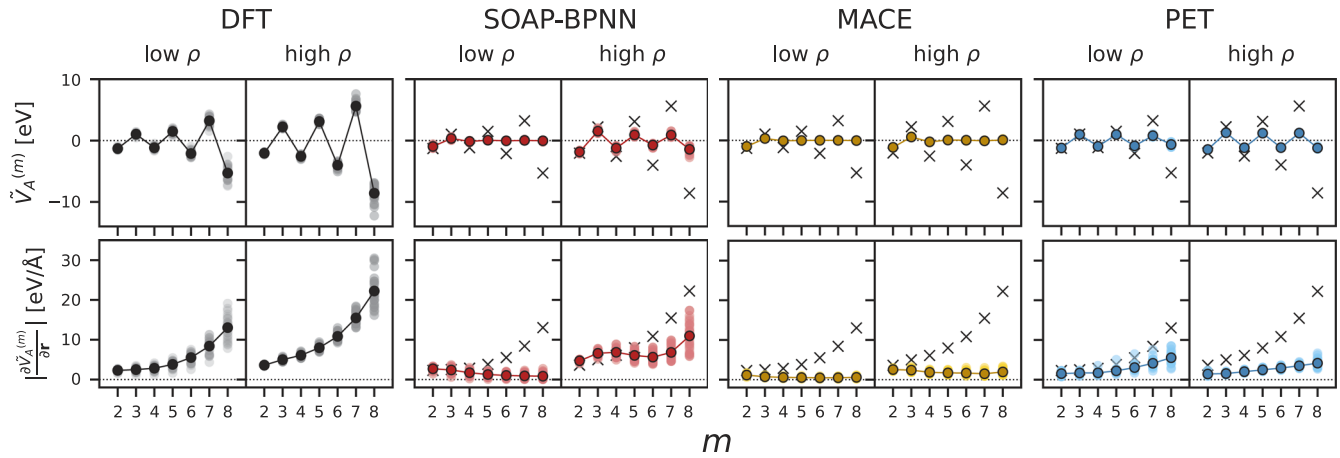

FIG. 2. Body-order trends of 25 low ρ and 25 high ρ hydrogen 8-mers computed with DFT and MLIPs, as probed by $\tilde{V}_A^{(m)}$ of Eq. (7) and $|\partial \tilde{V}_A^{(m)}/\partial \mathbf{r}|$ of Eq. (8). The raw distribution of values is shown in lighter markers, and the mean values across the samples are shown with a darker outlined marker. The DFT mean values are presented as crosses in the plots for the MLIPs.

in complete isolation), considerations in Ref. 67 assume full awareness of the *entire* local environment even for the sub-clusters, which is more consistent with the stated goal of investigating the convergence of MLIPs, even though it leaves the definition of E_1 somewhat vague. In Appendix C, we further unravel the dependence of the convergence trends on the baseline choice and rationalize that for energy-stable atomic clusters such as the hydrogen 8-mers of our study, the observed oscillatory trend in the vacuum-referenced MBE is reasonable under the choice of E_1 as the isolated atom energy.

In light of the strong dependence of the body-ordered energetics on the choice of the baseline energy E_1, it might be more appropriate to evaluate the convergence of the expansion by looking at the magnitude of the *forces* that are independent of the choice of E_1. We compute the following quantity:

$$\left|\frac{\partial \tilde{V}_A^{(m)}}{\partial \mathbf{r}}\right| = \frac{\sum_{A^{(m)}\in A}\sqrt{\frac{1}{m}\sum_i \left|\frac{\partial V^{(m)}}{\partial \mathbf{r}_i}\right|^2}}{\frac{N!}{m!(N-m)!}}. \tag{8}$$

The numerator is the sum of the root mean square (RMS) of the norms of $\partial V^{(m)}/\partial \mathbf{r}_i$ for the canonically complete set of m-mers, where the RMS of norms is computed over the atoms of a given m-mer. The denominator performs a normalization analogous to Eq. (7). As shown in the first panel of Fig. 2, for DFT, the magnitude of the BO forces does not converge by $m = 8$, indicating that the slow convergence is not only a consequence of the choice of a vacuum reference for the expansion, but also reflects the high degree of electronic correlations for many of the sub-clusters.

IV. EFFECTIVE BODY-ORDEREDNESS OF MLIPS

We now analyze the "effective" body-orderedness of MLIPs by training SOAP-BPNN, MACE, and PET on an hydrogen 8-mer dataset and computing $\tilde{V}_A^{(m)}$ and $|\partial \tilde{V}_A^{(m)}/\partial \mathbf{r}|$ from Eqs. (7) and (8). The dataset includes both low ρ and high ρ 8-mers in a 1:1 ratio. 10 000 8-mers split into 8:1:1 proportions are used as the training, validation, and test sets, with stratification between the ρ regimes. All resulting models show energy and force RMSEs below 0.025 eV/atom (6.2% RMSE) and 0.375 eV/Å (15.6% RMSE, see Table S1). The analysis is performed on 25 low ρ and 25 high ρ 8-mers in the test set, for which the DFT reference values are available.

Figure 2 reveals that all MLIPs deviate from the DFT body-order trends in both energies and forces and infer their own, effective body-orders for the hydrogen 8-mers, with much smaller magnitudes across all m. In SOAP-BPNN, body-orders of the low ρ 8-mers are fast-converging with m, whereas those of the high ρ 8-mers exhibit an oscillatory, slow-converging behavior in $\tilde{V}_A^{(m)}$ and an overall increasing trend in $|\partial \tilde{V}_A^{(m)}/\partial \mathbf{r}|$. The spread of values across individual samples is also much narrower for low ρ 8-mers compared to high ρ 8-mers that reach a standard deviation (σ) of 0.481 eV and 3.704 eV/Å in energies and forces for $m = 8$.

In MACE, the body-order trends of $\tilde{V}_A^{(m)}$ are fast-converging for both densities with the significant contributions limited to $m \leq 4$, and $|\partial \tilde{V}_A^{(m)}/\partial \mathbf{r}|$ also exhibits the lowest values across all m. The body-ordered contributions of both energies and forces are larger in high ρ 8-mers than in low ρ 8-mers. The spread across individual samples is the narrowest among all MLIPs, with an average σ of 0.060 eV and 0.434 eV/Å observed for high ρ 8-mers. This suggests that the effective body-ordering of MACE is applicable across many samples with minimal variation. In PET, $\tilde{V}_A^{(m)}$ shows an oscillating, non-converging trend and $|\partial \tilde{V}_A^{(m)}/\partial \mathbf{r}|$ increases with m for both low and high ρ 8-mers. While the magnitudes of $\tilde{V}_A^{(m)}$ are higher for high ρ 8-mers over low ρ 8-mers, those of $|\partial \tilde{V}_A^{(m)}/\partial \mathbf{r}|$ exhibit the reverse trend, which is the opposite of the reference. PET shows moderate spread of values compared to the other two models, where the largest σ values of 0.239 eV and 1.475 eV/Å are observed for low ρ 8-mers.

In Fig. S1, we present the "per-m" RMSEs of the three MLIPs on the *sub-clusters* of the test set 8-mers, assessing the accuracy of the effective body-ordered energetics of the MLIPs. All three models exhibit similarly large RMSEs on average for both energies (0.325 eV/atom) and forces (1.21 eV/Å) for all $m < 8$. This corroborates that the intuited body-ordered energetics are only effective and significantly far from the DFT reference. In the supplementary material, we also explore the body-order trends of the MLIPs when E_1 is no longer fixed to the isolated atom energy. We learn that all MLIPs retain similar trends with notably smaller contributions across all body-orders, with the exception of PET exhibiting slightly larger $|\partial \tilde{V}_A^{(m)}/\partial \mathbf{r}|$ values for both low ρ and high ρ 8-mers.

V. EXPLICIT RESOLUTION OF REFERENCE BODY-ORDERS

Next, we re-train the MLIPs on augmented datasets that aim to resolve their body-orders to the DFT reference. Canonically complete sets of m-mers, $2 \leq m < 8$, for 200 low ρ and 200 high ρ 8-mers (98 400 new structures in total), are added to the training set. The same is done for the validation and test sets, with 25 low ρ and 25 high ρ 8-mers each. Multiple models are trained at different augmentation proportions (inclusion of all 98 400 m-mers is 1), and in doing so, stratification is performed so that the complete set of canonical sub-clusters for a given 8-atom configuration is included all at once.

Figure 3 shows that the DFT body-order trends are quickly captured by both MACE and PET as the m-mers are added to the training set. In fact, for both ρ 8-mers, near-complete resolution is achieved for $m \leq 6$ at 0.01 added m-mer proportion, and further increase in the proportion quickly resolves the $6 \leq m \leq 8$ contributions, faster for PET than MACE. This proves that while the MACE and PET architectures have sufficient flexibility to learn the reference body-order trends, without explicit resolution of the body-orders, their inherent tendencies result in trends that are drastically different from the reference. In SOAP-BPNN, $m \leq 4$ contributions are also resolved quickly at 0.01 added m-mer proportion, akin to the other two models. SOAP-BPNN, however, struggles to capture the higher body-order contributions at all added m-mer proportions, although it still gradually approaches the reference values. Better resolution is achieved for high ρ 8-mers than for low ρ 8-mers. In all MLIPs, the resolution takes place consistently faster for lower m, which must be related to how the dimensionality of the potential energy surfaces of m-mers exponentially increases with m.

Figure 4 shows the energy and force RMSEs for the 8-mers and their fragments. The RMSEs for the m-mers decrease monotonically for all MLIPs as more and more sub-clusters are augmented to the dataset, as expected. For the 8-atom structures, SOAP-BPNN and MACE exhibit a compromise in the accuracy of the full

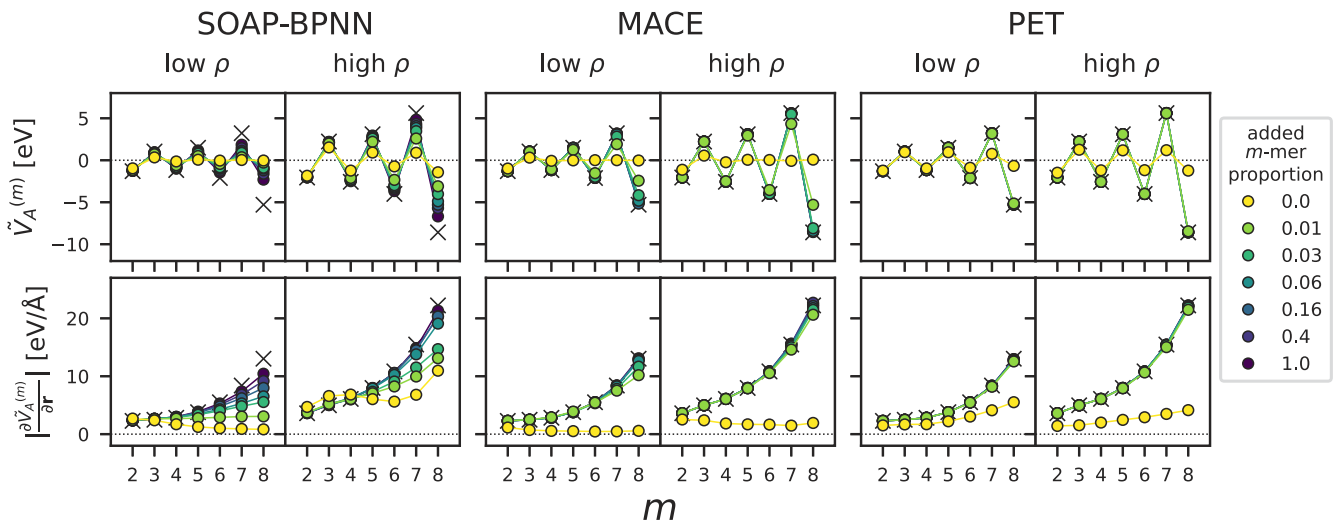

FIG. 3. Changes in the body-order trends of MLIPs when explicit resolution to the DFT reference is attempted. Mean $\tilde{V}_A^{(m)}$ and $|\partial \tilde{V}_A^{(m)}/\partial \mathbf{r}|$ values across 25 low ρ and 25 high ρ test set hydrogen 8-mers are shown. Multiple plots are made for different m-mer proportions, where the proportion corresponds to the number of m-mers added over all available m-mers for body-order resolution. In each panel, reference mean $|\partial \tilde{V}_A^{(m)}/\partial \mathbf{r}|$ values from DFT are marked with crosses.

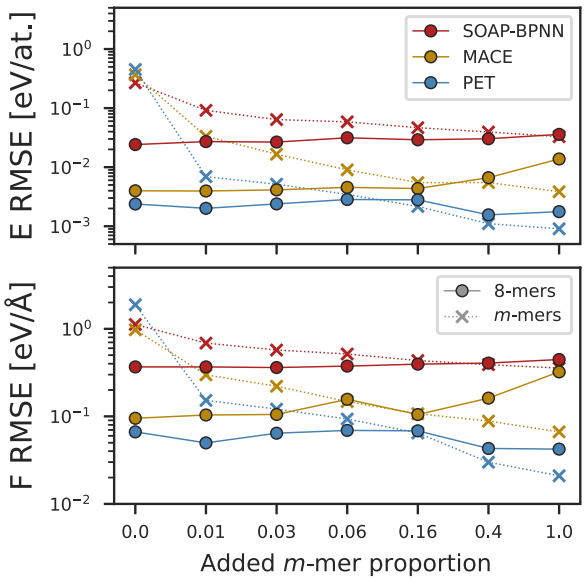

FIG. 4. Energy and force RMSEs of the MLIPs computed for the full structure 8-mers (filled circles) and the sub-cluster m-mers (crosses) under different degrees of body-order resolution achieved with the addition of m-mers to the training set.

configurations with the addition of sub-cluster m-mers in the training dataset. The RMSEs generally increase as more m-mers are incorporated, with MACE exhibiting a more pronounced increase—by a factor of 3.4 in both the energy and force RMSEs upon full augmentation. In PET, the RMSEs eventually decrease to values that are lower than those before any m-mers are introduced to the dataset. Among the considered MLIPs, PET is the only architecture for which learning the reference body-ordered energetics further improves the accuracy on the full structures.

VI. BODY-ORDERED INTERPRETATION OF MLIP LEARNING DYNAMICS

Having established that in the absence of explicit fragments in the training set, the MLIPs infer their own body-order trends, we now investigate how these trends depend on the composition of the training dataset. We first train the models on the sub-sampled versions of the original training set from Sec. IV in order to probe the learning dynamics of the MLIPs in the context of body-orders. We use dataset proportions ranging from 0.01 to 0.4 and keep the validation and test sets fixed. The learning curves are shared in Fig. S3. We also repeat the exercise for datasets exclusively composed of either 10 000 low ρ 8-mers or 10 000 high ρ 8-mers. The resulting model accuracies for the latter two cases are shared in Tables S2 and S3.

Figure 5 presents the body-order trends for the MLIPs trained on different proportions of the original 8-mer dataset of Sec. IV. In SOAP-BPNN, the model intuits an overall converging body-order trend at 0.01 dataset proportion for both density regimes, while assigning larger magnitudes for the higher ρ 8-mers. As the dataset expands, the initial, converging trend is largely retained in the low ρ 8-mers, with only a slight increase in the $|\partial \tilde{V}_A^{(m)}/\partial \mathbf{r}|$ values for the larger body-orders. For the high ρ 8-mers, the model steadily increases the contributions for $3 \leq m \leq 8$ until reaching the final observed trend at the full dataset size. Note that the $m = 2$ contribution remains "pinned" in both ρ 8-mers.

MACE exhibits a fast-converging body-order trend at the smallest dataset proportion. In $\tilde{V}_A^{(m)}$, this initial profile is kept constant with minor fluctuations for both low ρ and high ρ 8-mers under all dataset proportions. Even in $|\partial \tilde{V}_A^{(m)}/\partial \mathbf{r}|$, the initial trend is largely kept constant in the low ρ 8-mers for all dataset sizes and is

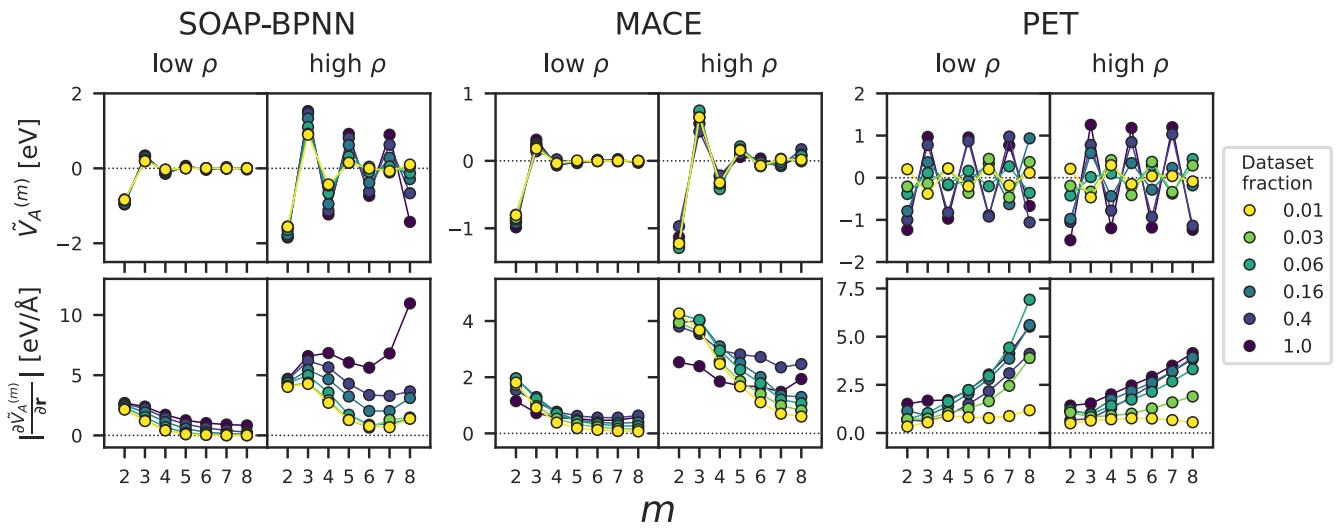

FIG. 5. Body-order trends of the MLIPs trained on different fractions of the original training dataset. Mean $\tilde{V}_A^{(m)}$ and $|\partial \tilde{V}_A^{(m)}/\partial \mathbf{r}|$ values over 25 low ρ 8-mers and 25 high ρ 8-mers are separately plotted. Note that the *y*-axis ranges are tailored for each model to focus on their individual learning behavior.

retained up to 0.16 dataset proportion in the high ρ 8-mers. Past this proportion, the high ρ 8-mers start to exhibit relatively larger deviations in $|\partial \tilde{V}_A^{(m)}/\partial \mathbf{r}|$ from the initial trend. Contrary to the other two MLIPs, PET does not display any converging trend and exhibits non-negligible contributions across all m even at the smallest dataset proportion. As the dataset is further expanded, PET freely adapts with no discernible trend in $\tilde{V}_A^{(m)}$ and quickly increases the contributions from higher m in $|\partial \tilde{V}_A^{(m)}/\partial \mathbf{r}|$, in stark contrast to the behavior of other two models.

Most notably, we observe the tendency of MACE to prioritize the use of lower body-orders in its learning, which is also faintly present in SOAP-BPNN and absent in PET. Similar behaviors persist when the models are exclusively trained on the low ρ 8-mers or the high ρ 8-mers (see Figs. S4–S7). MACE consistently prefers a fast-converging body-order trend, while SOAP-BPNN exhibits a converging trend for the low-density-only case and does not for the high-density-only case, and PET continues to show entirely arbitrary trends, especially in $\tilde{V}_A^{(m)}$. We attribute this distinct trend of MACE to the *over-representation* of lower body-orders (see Appendix D), which encourages the model to prioritize the use of lower body-ordered in formation in the learning task. We conjecture that the over-representation of lower body-orders restricts the use of the higher body-ordered features or descriptors to its full capacity, resulting in the limited capability of MACE to optimize its body-orders as the dataset is further expanded.

One may wonder how much these trends depend on the details of an architecture and the choice of hyperparameters. As we show in the supplementary material, increasing the number of channels or correlation order ν of MACE modulates the body-order trend but does not change the overall fast-decaying tendency. Using nonlinear interaction blocks[68] in MACE alleviates the model from converging trends and induces an oscillatory, diverging behavior in $\tilde{V}_A^{(m)}$ more similar to PET. Changing the token size for PET leads to changes in the body-order terms but not in their qualitative behavior, which remains oscillatory and without a clear relation to that of the underlying DFT reference.

VII. EXTRAPOLATIVE PERFORMANCE OF MLIPS

One common interpretation of the MBE is that once sufficient convergence is reached in the body-ordered energetics, the expansion can be generalized to any applicable system with good accuracy. To assess this, we evaluate the extrapolative performance of the MLIPs, which have shown varying trends of convergence in their effective body-ordered energetics with respect to m. The energy and force RMSEs are computed for a dataset of out-of-distribution 8-mers that have been further sampled from the intermediate density regime of the Cheng *et al.*[23] bulk hydrogen dataset, between the two density extrema from which the original 8-mers were sampled. The intermediate density 8-mers are organized into quintiles, where their original bulk hydrogen configurations exhibit densities of 0.720, 0.876, 0.999, 1.14, and 1.25 g/cm^3, respectively. The results are presented in Fig. 6.

When the models are trained on both low ρ and high ρ 8-mers, errors below 0.1 eV/atom are also observed in all three MLIPs for the out-of-distribution intermediate density 8-mers. The lowest RMSEs are consistently observed for PET, which achieves average RMSEs of 0.0061 eV/atom for the energies and 0.148 eV/Å for the forces across all intermediate density quintiles. The next best performance is observed for MACE, with an energy RMSE of 0.0228 eV/atom for the first quintile, below 0.02 eV/atom for all other quintiles, and an average force RMSE of 0.239 eV/Å across all quintiles. SOAP-BPNN exhibits average RMSEs of 0.0454 eV/atom and 0.674 eV/Å for the energies and forces, respectively, which are still well within an order of magnitude from the RMSEs observed for the original test set.

When the models are trained on the low ρ 8-mers only, the out-of-distribution performance becomes worse for all three MLIPs. In

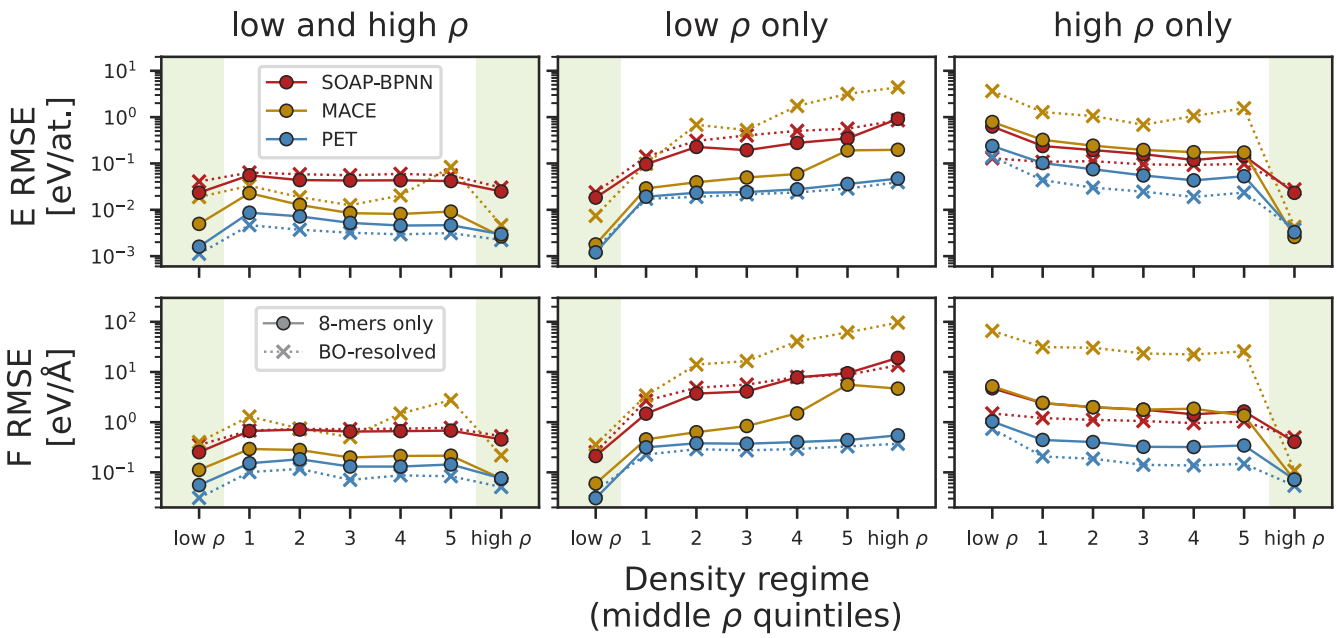

FIG. 6. Energy (top) and force (bottom) RMSEs of the MLIPs computed across the entire density range. *x* axis denotes the different density regimes for which the RMSEs have been computed. The two extrema are the original low ρ and high ρ regimes used in model training, and the five enumerated middle ticks correspond to the quintiles of the out-of-distribution intermediate density regimes. Each column corresponds to a different training dataset combination, which is shaded in light green. Circles show the RMSEs for the models trained on the 8-mers only, and the crosses show the RMSEs for the models trained on the 8-mers and their sub-clusters.

fact, an overall increasing trend in the RMSEs is observed from the first intermediate density quintile (closest to low ρ) all the way to the high ρ 8-mers. For MACE, RMSEs for the fifth quintile and high ρ are much worse than those of the other quintiles, exhibiting up to two orders of magnitude difference compared to the low ρ 8-mers. For PET, the RMSEs are relatively more constant across the intermediate density quintiles and high ρ, with average values of 0.0296 eV/atom and 0.409 eV/Å for the energies and forces. The analogous reverse trend is observed when the models are trained on the high ρ 8-mers only, where the RMSEs for the intermediate ρ and low ρ 8-mers become significantly higher than before and show an increasing trend from the fifth quintile (closest to high ρ) to the first quintile and low ρ.

Figure 6 also presents the RMSEs for the intermediate ρ 8-mers when the body-order trends of the models are explicitly resolved to the DFT reference (Sec. V). MACE exhibits performance degradation in all cases, and the degradation is far more pronounced when trained on the low or high ρ 8-mers only. SOAP-BPNN shows reduced RMSEs for the case of high ρ 8-mers only and slight degradation in the other two cases. PET consistently displays slightly lower RMSEs for the out-of-distribution 8-mers when trained on the body-order-resolved datasets. This showcases the flexibility of PET to learn simultaneously the energetics of target 8-mers as well as their sub-clusters and then use the extra information from the sub-clusters to achieve further improvements in the RMSEs. In the supplementary material, we show that these trends are generally robust to changes in the hyperparameters of MACE and PET, except when nonlinear interaction blocks are used in MACE, which mitigates the previously observed performance degradation with body-order resolution to the DFT reference.

Altogether, the generalizability of the MLIPs does not correlate clearly with any specific convergence trends in their body-orders. The explicit resolution of the body-orders also does not bring forth dramatic improvements in the extrapolative performance of the MLIPs. This hints at the absence of any practical benefit in enforcing the models to infer fast converging body-orders or directly learn the reference body-ordered energetics. If anything, such complexities can limit the learning capacity and add strain to the training exercise, potentially degrading both in- and out-of-distribution performance.

VIII. CONCLUSION AND OUTLOOK

In this study, we have carefully analyzed the behavior of three different MLIPs in terms of their "body-orderedness," examining the body-ordered energy and force trends for hydrogen clusters extracted from bulk simulations at different densities, comparing between DFT and the MLIPs trained on a number of different dataset makeups. In the reference DFT calculations, as seen in many other systems, we first observed that the MBE of energy shows a non-converging, oscillatory behavior for both "molecular" (low density) and "atomic" (high density) hydrogen clusters. The effect cannot simply be dismissed as an artifact of DFT, as the trend is also reproduced in state-of-the-art DMRG calculations. Even though the oscillatory behavior can be explained in terms of the choice of the isolated atom energy as the baseline, higher body-order terms are also large for force-based metrics, which are insensitive to the choice

of the baseline. An analysis of the electronic structure of the fragments points to the strong spin correlations as the origin of high body-order terms.

When trained exclusively on 8-mers extracted from realistic bulk structures, the MLIPs all learn an effective MBE that is far from the reference—without any adverse effect on their in-domain accuracy. While the effective MBE is largely arbitrary for all models, MACE tends to prioritize the use of lower body-orders for a fast-converging trend. When the sub-clusters are incorporated into the training set to explicitly resolve the body-orders to the reference, MACE and PET quickly converge to the reference body-ordered energetics, but the relatively low descriptive power of SOAP-BPNN limits its accuracy on the m-mers and hence its ability to learn the reference MBE. We also observe that explicit body-order resolution degrades the accuracy on the full structures of interest for MACE, while it does not for PET. Contrary to what one might expect, the fast decay of the effective MBE does not translate into more robust extrapolative behavior. Explicitly resolving the body-orders does not improve the extrapolative performance for SOAP-BPNN, degrades it for MACE, and improves it slightly for PET.

While our experiments focus on one comparatively simple system, they suggest that there is little value in targeting explicitly the MBE, or in designing models that implicitly favor learning a fast-decaying effective body-ordered decomposition. On the contrary, it appears that an unconstrained architecture such as PET that does not build upon a hierarchical expansion of the neighbor density correlations but simply aims to achieve a highly expressive approximation of the target demonstrates consistently the best performance, for both in- and out-of-distribution tasks.

The "paradox" of the MBE is resolved by recognizing that models trained on reasonably stable structures do not have to reproduce the MBE of the target. The large deviation between the true and empirical trends is a simple consequence of the fact that, for a model trained on those reasonable structures, the fragments that appear in the decomposition are highly distorted and amount to extrapolative predictions. We also speculate that the tendency of MACE to privilege a fast-decaying effective body-ordered energy decomposition may be a consequence of the "contamination" of high-order correlations with low-body-order components. We have shown in the supplementary material that the newly proposed nonlinear interaction blocks of MACE can alleviate this effect to an extent.[68] Given the strategy to "purify" the body-ordered components of the closely related atomic cluster expansion,[27] it could be interesting to investigate the behavior of a "purified MACE" architecture, to verify our hypothesis and observe if there are any consequences for accuracy and transferability. Overall, our observations suggest that, despite being an attractive approximation framework, and despite the strong mathematical connection to many widely used MLIP frameworks, the body-order decomposition is not especially useful as a guiding principle to design MLIPs.

SUPPLEMENTARY MATERIAL

See the supplementary material for further details of the *ab initio* calculations, RMSE values, and results of additional experiments with variations in the dataset, baseline energy E_1, and ML model hyperparameters.

ACKNOWLEDGMENTS

The authors thank Alexander Goscinski, Arslan Mazitov, and Gabor Csányi for useful discussions. S.C., M.D., and M.C. acknowledge support from an SNSF (Project ID 200020_214879). F.B. and M.C. were supported by a project of the Platform for Advanced Scientific Computing (PASC). F.G. received funding from the ERC CoG grant FIAMMA (Grant ID 101001890). Calculations relied on computer time from the EPFL HPC platform SCITAS and from the CSCS (Project ID lp26).

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Sanggyu Chong: Conceptualization (equal); Data curation (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Project administration (equal); Validation (equal); Visualization (equal); Writing – original draft (equal); Writing – review & editing (equal). **Tong Jiang**: Data curation (equal); Formal analysis (equal); Investigation (equal); Project administration (equal); Validation (equal); Writing – original draft (equal); Writing – review & editing (equal). **Michelangelo Domina**: Formal analysis (equal); Investigation (equal); Writing – original draft (equal); Writing – review & editing (equal). **Filippo Bigi**: Conceptualization (equal); Formal analysis (equal); Investigation (equal); Writing – original draft (equal); Writing – review & editing (equal). **Federico Grasselli**: Conceptualization (equal); Formal analysis (equal); Investigation (equal); Writing – review & editing (equal). **Joonho Lee**: Formal analysis (equal); Investigation (equal); Supervision (equal); Writing – review & editing (equal). **Michele Ceriotti**: Conceptualization (equal); Formal analysis (equal); Funding acquisition (equal); Investigation (equal); Methodology (equal); Project administration (equal); Supervision (equal); Writing – original draft (equal); Writing – review & editing (equal).

DATA AVAILABILITY

The data that support the findings of this study are openly available in Materials Cloud at https://doi.org/10.24435/materialscloud:q7-da, Ref. 48.

APPENDIX A: LINEAR BODY-ORDERED MODELS

In linear and kernel-based MLIPs with the locality ansatz, the descriptors are often built to describe atom-centered body-order correlations. One example is the Gaussian Approximation Potential (GAP)[14] model utilizing SOAP descriptors,[33] which is a 3-body descriptor. In some special cases (e.g., moment-tensor potential[15] and ACE[16]), $\mathbf{x}_i$ of Eq. (5) can be expressed in terms of multiple body-ordered components,

$$\mathbf{x}_i = \bigoplus_{\nu=1}^{\nu_{\max}} \sum_{j_1,\ldots,j_\nu \neq i} \varphi^\nu(\mathbf{r}_{ij_1},\ldots,\mathbf{r}_{ij_\nu}) f_c(r_{ij_1},\ldots,r_{ij_\nu}). \tag{A1}$$

Local environments are described at successively increasing body-orders (here expressed in terms of ν) by functions φ^ν that describe

the body-order $\nu+1$, and these descriptors are concatenated together to obtain $\mathbf{x}_i$. Since there exist separate "blocks" corresponding to the different body-orders, the same separation can be applied to the trained weights of the model, which results in Eq. (6).

In such models, body-ordered descriptors of Eq. (A1) often contain self-interacting terms, i.e., contributions where $j_1 = j_2$. In ACE, allowing for such self-correlations, which is sometimes referred to as the "density trick," is what guarantees favorable scaling with the number of atoms in the environments. Chong *et al.*[26] have shown that the presence of such self-interactions leads to a model learning behavior where the apparent correspondence between the summands of Eqs. (1) and (6) cannot be captured by the model. More recently, Ho *et al.*[27] proposed a purification operator that removes the self-correlation contributions from the ACE descriptors, which allows the model to recover the above-mentioned correspondence in some specific cases (see Sec. V A of Chong *et al.*[26]).

APPENDIX B: BODY-ORDERED ENERGETICS OF 8-MER BEFORE AND AFTER FRAGMENTATION

To further verify that the slowly converging body-ordered energetics observed from the quantum chemical calculations are not stemming from DFT artifacts, namely the self-interaction error, we consider the "fragmentation" of an 8-mer into two 4-mer fragments and the changes in the body-ordered energetics thereof at different theory levels. The 8-mer of interest is constructed by first arranging four hydrogen atoms into a square within the *xy*-plane with a bond length of 1.4 Å, then replicating this configuration, rotating it by 45°, and offsetting the replica in the *z*-direction by a distance that also results in similar nearest-neighbor bond lengths. Taking this as the pre-fragmentation configuration, we then generate the post-fragmentation configuration by further increasing the offset distance between the two replicas by 50 Å.

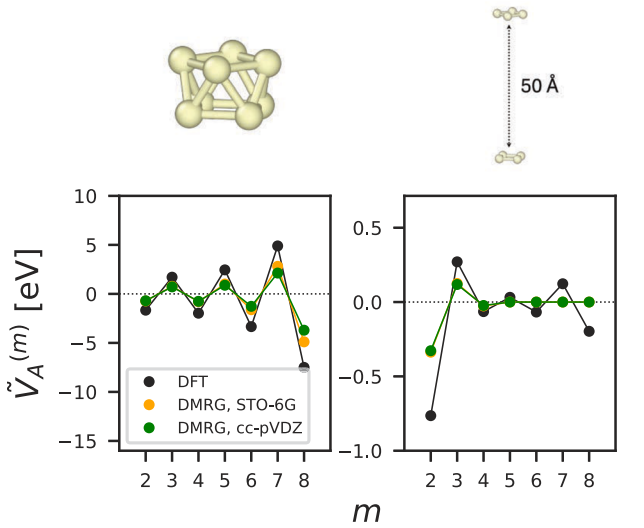

FIG. 7. Body-ordered energetics of an 8-mer before (left) and after (right) fragmentation into two 4-mer fragments separated by 50 Å. Calculations are done with DFT as well as DMRG with STO-6G and cc-pVDZ basis sets. Corresponding atomic configurations are shown on top of the plots. Note that the *y*-axis of the two plots is presented in different scales.

In Fig. 7, a similar oscillatory trend is observed across all calculations for the 8-mer before fragmentation. When the cluster is separated into fragments, both DMRG calculations show a strict convergence of $\tilde{V}_A^{(m)}$ to 0 for $m > 4$, which is the physically reasonable trend. On the contrary, DFT results contain unphysical, residual contributions for $m > 4$ that originate from the self-interaction error. These results prove that the prevalent oscillatory body-ordered energetics capture the true, physical MBE trends at the atomic level for covalent systems and are not a consequence of the DFT self-interaction error.

APPENDIX C: DEPENDENCE OF BODY-ORDER CONVERGENCE ON BASELINE ENERGY

The alternating behavior of the body-ordered energies might appear strange but can easily be explained as follows. Begin by noting that $V^{(m)}$ can be expressed as the cohesive energy of the m-mer cluster (i.e., $E^{(m)} - m \cdot E_1$) minus the sub-cluster contributions. Assume that this term is proportional to m, $E^{(m)} - m \cdot E_1 \approx m\epsilon$ (which is true in the asymptotic limit, but not necessarily for smaller clusters). Then, it is easy to see that

$$\begin{aligned} V^{(2)} &= E^{(2)} - 2E_1 \approx 2\epsilon, \\ V^{(3)} &= E^{(3)} - 3E_1 - \sum_{i<j} V^{(2)}(\mathbf{r}_i, \mathbf{r}_j) \approx 3\epsilon - 6\epsilon = -3\epsilon, \\ V^{(4)} &= E^{(4)} - 4E_1 - \sum_{i<j} V^{(2)}(\mathbf{r}_i, \mathbf{r}_j) \\ &\quad - \sum_{i<j<k} V^{(3)}(\mathbf{r}_i, \mathbf{r}_j, \mathbf{r}_k) \approx 4\epsilon - 12\epsilon + 12\epsilon = 4\epsilon, \end{aligned} \tag{C1}$$

which corresponds to the alternating, diverging trend. The fact that this trend continues can be proved by induction,

$$V^{(m)} = E^{(m)} - mE_1 - \sum_{k=2}^{m-1} V^{(k)} \frac{m!}{k!(m-k)!}, \tag{C2}$$

assuming that up to $(m-1)$th body-order $V^{(k)} \approx (-1)^k k\epsilon$, one sees that the summation evaluates to $m(1-(-1)^m)\epsilon$ so that indeed $V^{(m)} = (-1)^m m\epsilon$.

These considerations justify the oscillatory behavior observed for the body-ordered energies and indicate that adjusting E_1 to an effective value that zeros out ϵ would facilitate the convergence of the expansion—consistent with the derivation in Ref. 67 that assumes the definition of a system-specific, effective energy reference.

APPENDIX D: OVER-REPRESENTATION OF LOW BODY-ORDERS IN MACE

Here, we discuss two "channels" that lead to the over-representation of the low body-order contributions in MACE. First, consider the initial features of MACE, which are defined in terms of ACE based on the density trick, as discussed in Sec. II A. Since this implies the presence of self-interaction terms in the formalism,[16,21] all higher body-order features also contain effectively lower body-order contributions, which make the descriptors deviate from the

strictly canonical expression of the body orders. That is, in the summation of Eq. (A1), there exist terms where $j_k = j_l$. Recognizing their presence, we can re-express the descriptors by decomposing their contributions, as done in the following for the example of $\nu = 2$,

$$\sum_{j_1, j_2 \neq i} \varphi^2(\mathbf{r}_{ij_1}, \mathbf{r}_{ij_2}) f_c(r_{ij_1}, r_{ij_2}) = x^2_{i,\text{self}} + x^2_{i,\text{pure}}, \tag{D1}$$

where the self-interacting terms are the one such defined from the same atom $j_1 = j_2$, namely

$$x^2_{i,\text{self}} := \sum_j \varphi^2(\mathbf{r}_{ij}, \mathbf{r}_{ij}) f_c(r_{ij}, r_{ij}), \tag{D2}$$

while $x^2_{i,\text{pure}}$ contains all the remaining pure $\nu = 2$ terms such that $j_1 \neq j_2$. Because the summation in $x^2_{i,\text{self}}$ runs over only one index, the self-interaction term is a contribution with an effective lower body-order. Extending this argument to every increasing ν, this proves that the basic ACE features, grounded in the density trick, imply that every ν-term contains all the lower body-orders. As such, also every feature in MACE contains an over-representation of lower orders. Moreover, as shown in Ref. 69 (see Table 2), expansion terms containing self-interaction give rise to an ill-conditioned representation. Therefore, one can infer that the terms for lower body-orders are of a magnitude that is comparable to that of the pure ones.

The second channel is provided by the update function, containing also a residual connection, used in the MACE architecture as defined in Ref. 37, which both contribute to the over-representation of lower body-orders in the effective descriptors. The update function is defined as [from Ref. 37, Eq. (12)]

$$\mathbf{h}^{(t+1)}_{i,L} = U(\mathbf{m}^{(t)}_{i,L}, \mathbf{h}^{(t)}_{i,L}) := \mathbf{U}^{(t)}_{kL} \cdot \mathbf{m}^{(t)}_{i,L} + \mathbf{W}^{(t)}_{z_i,kL} \cdot \mathbf{h}^{(t)}_{i,L}, \tag{D3}$$

where U is the update function, $\mathbf{h}^{(t)}_{i,L}$ are the features of the model after the tth step of message passing, $\mathbf{m}^{(t)}_{i,L}$ are the messages at the same steps, containing higher body order terms, and $\mathbf{U}^{(t)}_{kL}$ and $\mathbf{W}^{(t)}_{z_i,kL}$ are learnable weights. At the core of the computation of the messages, there are the same ACE contractions with self-interactions discussed above, and thus, they have an over-representation of lower body-orders. Moreover, in defining the new features, the skip connections allow us to utilize also the previous ones, which themselves contain the over-representation of lower body-orders.

REFERENCES

[1] W. Yang, Phys. Rev. Lett. **66**, 1438 (1991).
[2] S. Saebo and P. Pulay, Annu. Rev. Phys. Chem. **44**, 213 (1993).
[3] W. Yang and T.-S. Lee, J. Chem. Phys. **103**, 5674 (1995).
[4] R. Baer and M. Head-Gordon, Phys. Rev. Lett. **79**, 3962 (1997).
[5] K. Kitaura, E. Ikeo, T. Asada, T. Nakano, and M. Uebayasi, Chem. Phys. Lett. **313**, 701 (1999).
[6] E. E. Dahlke and D. G. Truhlar, J. Chem. Theor. Comput. **3**, 46 (2007).
[7] I. V. Solovyev, J. Phys. Condens. Matter **20**, 293201 (2008).
[8] R. M. Richard and J. M. Herbert, J. Chem. Phys. **137**, 064113 (2012).
[9] S. R. Pruitt, C. Bertoni, K. R. Brorsen, and M. S. Gordon, Accounts Chem. Res. **47**, 2786 (2014).
[10] A. K. Rappe, C. J. Casewit, K. S. Colwell, W. A. Goddard III, and W. M. Skiff, J. Am. Chem. Soc. **114**, 10024 (1992).
[11] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case, J. Comput. Chem. **25**, 1157 (2004).
[12] K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov, and A. D. Mackerell, Jr., J. Comput. Chem. **31**, 671 (2010).
[13] J. Behler and M. Parrinello, Phys. Rev. Lett. **98**, 146401 (2007).
[14] A. P. Bartók, M. C. Payne, R. Kondor, and G. Csányi, Phys. Rev. Lett. **104**, 136403 (2010).
[15] A. V. Shapeev, Multiscale Model. Simul. **14**, 1153 (2016).
[16] R. Drautz, Phys. Rev. B **99**, 014104 (2019).
[17] J. Behler, Chem. Rev. **121**, 10037 (2021) pMID: 33779150.
[18] W. Kohn, Phys. Rev. Lett. **76**, 3168 (1996).
[19] E. Prodan and W. Kohn, Proc. Natl. Acad. Sci. **102**, 11635 (2005).
[20] M. J. Willatt, F. Musil, and M. Ceriotti, Phys. Chem. Chem. Phys. **20**, 29661 (2018).
[21] F. Musil, A. Grisafi, A. P. Bartók, C. Ortner, G. Csányi, and M. Ceriotti, Chem. Rev. **121**, 9759 (2021).
[22] A. Glielmo, B. E. Husic, A. Rodriguez, C. Clementi, F. Noé, and A. Laio, Chem. Rev. **121**, 9722 (2021).
[23] B. Cheng, G. Mazzola, C. J. Pickard, and M. Ceriotti, Nature **585**, 217 (2020).
[24] T. Bischoff, B. Jäckl, and M. Rupp, "Hydrogen under pressure as a benchmark for machine-learning interatomic potentials," arXiv:2409.13390 (2024) [cond-mat.mtrl-sci].
[25] J. Nigam, S. Pozdnyakov, G. Fraux, and M. Ceriotti, J. Chem. Phys. **156**, 204115 (2022).
[26] S. Chong, F. Bigi, F. Grasselli, P. Loche, M. Kellner, and M. Ceriotti, Faraday Discuss. **256**, 322 (2025).
[27] C. H. Ho, T. S. Gutleb, and C. Ortner, J. Comput. Phys. **515**, 113271 (2024).
[28] A. M. Tan, F. Pellegrini, and S. de Gironcoli, J. Phys. Chem. A **129**, 7229 (2025).
[29] I. Batatia, P. Benner, Y. Chiang, A. M. Elena, D. P. Kovács, J. Riebesell, X. R. Advincula, M. Asta, M. Avaylon, W. J. Baldwin, F. Berger, N. Bernstein, A. Bhowmik, F. Bigi, S. M. Blau, V. Cărare, M. Ceriotti, S. Chong, J. P. Darby, S. De, F. D. Pia, V. L. Deringer, R. Elijošius, Z. El-Machachi, F. Falcioni, E. Fako, A. C. Ferrari, J. L. A. Gardner, M. J. Gawkowski, A. Genreith-Schriever, J. George, R. E. A. Goodall, C. P. Grey, P. Grigorev, S. Han, W. Handley, H. H. Heenen, K. Hermansson, C. H. Ho, C. Holm, J. Jaafar, S. Hofmann, K. S. Jakob, H. Jung, V. Kapil, A. D. Kaplan, N. Karimitari, J. R. Kermode, N. Kroupa, J. Kullgren, M. C. Kuner, D. Kuryla, G. Liepuoniute, C. Lin, J. T. Margraf, I.-B. Magdău, A. Michaelides, J. H. Moore, A. A. Naik, S. P. Niblett, S. W. Norwood, N. O'Neill, C. Ortner, K. A. Persson, K. Reuter, A. S. Rosen, L. A. M. Rosset, L. L. Schaaf, C. Schran, B. X. Shi, E. Sivonxay, T. K. Stenczel, V. Svahn, C. Sutton, T. D. Swinburne, J. Tilly, C. van der Oord, E. Varga-Umbrich, T. Vegge, M. Vondrák, Y. Wang, W. C. Witt, T. Wolf, F. Zills, and G. Csányi, "A foundation model for atomistic materials chemistry," J. Chem. Phys. **163**, 184110 (2025).
[30] Y. Park, J. Kim, S. Hwang, and S. Han, J. Chem. Theor. Comput. **20**, 4857 (2024).
[31] A. Mazitov, F. Bigi, M. Kellner, P. Pegolo, D. Tisi, G. Fraux, S. Pozdnyakov, P. Loche, and M. Ceriotti, "PET-MAD as a universal interatomic potential for advanced materials modeling," Nat. Commun. **16**, 10653 (2025).
[32] B. Rhodes, S. Vandenhaute, V. Šimkus, J. Gin, J. Godwin, T. Duignan, and M. Neumann, "Orb-v3: Atomistic simulation at scale," arXiv:2504.06231 (2025) [cond-mat.mtrl-sci].
[33] A. P. Bartók, R. Kondor, and G. Csányi, Phys. Rev. B **87**, 184115 (2013).
[34] S. N. Pozdnyakov, M. J. Willatt, A. P. Bartók, C. Ortner, G. Csányi, and M. Ceriotti, Phys. Rev. Lett. **125**, 166001 (2020).
[35] J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl, "Neural message passing for quantum chemistry," arXiv:1704.01212 (2017) [cs.LG] .
[36] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," arXiv:1706.03762 (2023) [cs.CL] .
[37] I. Batatia, D. P. Kovacs, G. Simm, C. Ortner, and G. Csanyi, Adv. Neural Inf. Process. Syst. **35**, 11423 (2022).
[38] D. P. Kovács, I. Batatia, E. S. Arany, and G. Csányi, J. Chem. Phys. **159**, 044118 (2023).
[39] S. Pozdnyakov and M. Ceriotti, Adv. Neural Inf. Process. Syst. **36**, 79469 (2023).

[40]V. Blum, R. Gehrke, F. Hanke, P. Havu, V. Havu, X. Ren, K. Reuter, and M. Scheffler, Comput. Phys. Commun. **180**, 2175 (2009).
[41]J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).
[42]S. Sharma and G. K.-L. Chan, J. Chem. Phys. **136**, 124121 (2012).
[43]H. Zhai, H. R. Larsson, S. Lee, Z.-H. Cui, T. Zhu, C. Sun, L. Peng, R. Peng, K. Liao, J. Tölle *et al.*, J. Chem. Phys. **159**, 234801 (2023).
[44]Q. Sun, T. C. Berkelbach, N. S. Blunt, G. H. Booth, S. Guo, Z. Li, J. Liu, J. D. McClain, E. R. Sayfutyarova, S. Sharma *et al.*, WIREs Comput. Mol. Biosci. **8**, e1340 (2018).
[45]Q. Sun, X. Zhang, S. Banerjee, P. Bao, M. Barbry, N. S. Blunt, N. A. Bogdanov, G. H. Booth, J. Chen, Z.-H. Cui *et al.*, J. Chem. Phys. **153**, 024109 (2020).
[46]E. Epifanovsky, A. T. Gilbert, X. Feng, J. Lee, Y. Mao, N. Mardirossian, P. Pokhilko, A. F. White, M. P. Coons, A. L. Dempwolff *et al.*, J. Chem. Phys. **155**, 084801 (2021).
[47]D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," arXiv:1412.6980 (2017) [cs.LG].
[48]L. Talirz, S. Kumbhar, E. Passaro, A. V. Yakutovich, V. Granata, F. Gargiulo, M. Borelli, M. Uhrin, S. P. Huber, S. Zoupanos, C. S. Adorf, C. W. Andersen, O. Schütt, C. A. Pignedoli, D. Passerone, J. VandeVondele, T. C. Schulthess, B. Smit, G. Pizzi, and N. Marzari, Sci. Data **7**, 299 (2020).
[49]M. Riera, A. Hirales, R. Ghosh, and F. Paesani, J. Phys. Chem. A **124**, 11207 (2020).
[50]M. Riera, E. P. Yeh, and F. Paesani, J. Chem. Theor. Comput. **16**, 2246 (2020).
[51]E. F. Bull-Vulpe, M. Riera, S. L. Bore, and F. Paesani, J. Chem. Theor. Comput. **19**, 4494 (2023).
[52]U. Góra, R. Podeszwa, W. Cencek, and K. Szalewicz, J. Chem. Phys. **135**, 224102 (2011).
[53]V. Babin, G. R. Medders, and F. Paesani, J. Phys. Chem. Lett. **3**, 3765 (2012).
[54]G. R. Medders and F. Paesani, J. Chem. Theor. Comput. **9**, 4844 (2013).
[55]G. R. Medders, V. Babin, and F. Paesani, J. Chem. Theor. Comput. **9**, 1103 (2013).
[56]R. M. Richard, K. U. Lao, and J. M. Herbert, Accounts Chem. Res. **47**, 2828 (2014).
[57]P. Pinski and G. Csányi, J. Chem. Theory Comput. **10**, 68 (2014).
[58]K. U. Lao, K.-Y. Liu, R. M. Richard, and J. M. Herbert, J. Chem. Phys. **144**, 164105 (2016).
[59]F. Paesani, Accounts Chem. Res. **49**, 1844 (2016).
[60]J. P. Heindel and S. S. Xantheas, J. Chem. Theor. Comput. **17**, 7341 (2021).
[61]Y. Zhai, A. Caruso, S. L. Bore, Z. Luo, and F. Paesani, J. Chem. Phys. **158**, 084111 (2023).
[62]J. F. Ouyang, M. W. Cvitkovic, and R. P. A. Bettens, J. Chem. Theor. Comput. **10**, 3699 (2014).
[63]D. R. Broderick and J. M. Herbert, Chem. Sci. **15**, 19893 (2024).
[64]B. Paulus, K. Rosciszewski, N. Gaston, P. Schwerdtfeger, and H. Stoll, Phys. Rev. B **70**, 165106 (2004).
[65]A. Hermann, R. P. Krawczyk, M. Lein, P. Schwerdtfeger, I. P. Hamilton, and J. J. P. Stewart, Phys. Rev. **76**, 013202 (2007).
[66]M. Motta, C. Genovese, F. Ma, Z.-H. Cui, R. Sawaya, G. K.-L. Chan, N. Chepiga, P. Helms, C. Jiménez-Hoyos, A. J. Millis, U. Ray, E. Ronca, H. Shi, S. Sorella, E. M. Stoudenmire, S. R. White, and S. Zhang, "Simons collaboration on the many-electron problem)," Phys. Rev. X **10**, 031058 (2020).
[67]J. Thomas, H. Chen, and C. Ortner, Arch. Ration. Mech. Anal. **246**, 1 (2022).
[68]I. Batatia, C. Lin, J. Hart, E. Kasoar, A. M. Elena, S. W. Norwood, T. Wolf, and G. Csányi, "Cross learning between electronic structure theories for unifying molecular, surface, and inorganic crystal foundation force fields," arXiv:2510.25380 (2025) [physics.chem-ph].
[69]G. Dusson, M. Bachmayr, G. Csányi, R. Drautz, S. Etter, C. van der Oord, and C. Ortner, J. Comput. Phys. **454**, 110946 (2022).