How Unconstrained Machine-Learning Models Learn Physical Symmetries

M. Domina,* J. W. Abbott,* P. Pegolo,* F. Bigi, and M. Ceriotti†
Laboratory of Computational Science and Modeling, Institut des Matériaux, École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland

(Received 25 March 2026; accepted 15 July 2026; published 20 August 2026)

The requirement of generating predictions that exactly fulfill the fundamental symmetry of the corresponding physical quantities has profoundly shaped the development of machine-learning (ML) models for physical simulations. In many cases, models are built using constrained mathematical forms that ensure that symmetries are enforced exactly. However, unconstrained models that do not obey rotational symmetries are often found to have competitive performance, and to be able to *learn* to a high level of accuracy an approximate equivariant behavior with a simple data augmentation strategy. In this paper, we introduce rigorous metrics to measure the symmetry content of the learned representations in such models, and assess the accuracy with which the outputs fulfill the equivariant condition. We apply these metrics to two unconstrained, transformer-based models operating on decorated point clouds (a graph neural network for atomistic simulations and a PointNet-style architecture for particle physics) to investigate how symmetry information is processed across architectural layers and is learned during training. Based on these insights, we establish a rigorous framework for diagnosing spectral failure modes in ML models. Enabled by this analysis, we demonstrate that one can achieve superior stability and accuracy by strategically injecting the minimum required inductive biases, preserving the high expressivity and scalability of unconstrained architectures while guaranteeing physical fidelity.

DOI: 10.1103/cnsd-p8nc

I. INTRODUCTION

Symmetries are a cornerstone of modern physics. Their profound connection with conservation laws is enshrined in Noether's theorem, a principle that has long guided the formalization of empirical phenomena into theoretical frameworks and served as a foundation for the development of new theories [1,2]. Different fields of physical science are characterized by distinct symmetry groups, such as the orthogonal group O(3) in molecular mechanics [3], the Lorentz group SO(1, 3) in high-energy physics [4,5], and the special unitary groups SU(2^N) and SU(3) in quantum mechanics and chromodynamics, respectively [6–8]. Consequently, in the development of data-driven models for physics, incorporating physical symmetries has often been regarded as the most natural choice. This approach has led to the rapid and fruitful development of machine-learning (ML) approaches across the physical sciences, with atomistic simulations standing out as a very successful application domain [3,9–13]. However, ensuring that a model strictly preserves group equivariance, guaranteeing that outputs transform predictably under group actions on the input, imposes rigid architectural constraints that can be computationally expensive and limit the expressivity of models [14,15].

Conversely, mainstream computer science and machine-learning methods have evolved to maximize architectural expressivity and efficiency, letting domain-specific inductive biases be learned directly from the data. Recently, there has been growing interest in applying these "unconstrained" models to physics and chemistry, with notable examples ranging from biomolecular structure prediction, where AlphaFold 3 relaxed some of the strict equivariance constraints of previous versions [16], to the classification of plasma crystals [17], galaxy morphology [18], and particle traces [19] using PointNet-like architectures [20]. By relaxing strict equivariance constraints, such models can gain fitting power and computational efficiency, but must infer the relevant transformation laws from data, typically through augmentation over symmetry groups [21]. Recent works [22,23] have shown that output-level equivariance errors are usually small compared with the baseline model accuracy, and related analyses have shown that, in several molecular tasks, symmetry-agnostic models trained with rotational augmentation can rapidly reduce these errors during training [24].

In the field of atomistic simulations, unconstrained models are gaining traction as successful alternatives to explicitly equivariant models, providing fast, accurate, and transferable surrogate models for quantum mechanics. Machine-learning interatomic potentials (MLIPs), which predict potential energies, atomic forces, and cell stresses from atomic positions and chemical types, are one of the most active areas in this regard. For these models, however, approximate symmetry cannot be assessed only in terms of average test accuracy. A

*These authors contributed equally to this work.
†Contact author: michele.ceriotti@epfl.ch

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

rotation or inversion of the atomic structure should induce the corresponding transformation of vectorial and tensorial predictions, such as forces and stresses, and residual inconsistencies may affect simulations based on these predictions. Recent works [13,23,25,26] demonstrate that unconstrained models can match or outperform invariant and equivariant architectures on benchmarks, while offering superior scalability through operations that can be implemented efficiently in current accelerated architectures. The capability of unconstrained MLIPs extends beyond benchmarks, excelling in realistic scenarios such as complex materials science simulations [23] and high-throughput screening [27,28].

More recently, unconstrained architectures have been applied to tasks beyond invariant prediction (e.g., energy), extending to vectorial (e.g., forces [29,30]) and higher-rank tensorial targets [31–33]. These tasks are significantly more demanding, requiring the model to learn complex transformation laws via data augmentation. Unconstrained architectures appear capable of rising to these challenges, though potentially at the cost of longer training runs required to sample the group orbits that equivariant models encode by design [28].

Given the empirical evidence that these models *can* learn symmetry, natural questions arise: *When* and *how* are these symmetries learned, both across the model architecture and throughout the training process? And can this knowledge drive the design of better performing architectures? We introduce metrics to quantify how much the outputs of a model violate the equivariant conditions, as well as to partition the information encoded in the model in terms of the irreducible representations of the relevant groups, and use them to analyze quantitatively the flow of symmetry information across the architecture, and its evolution during training. We will then show that, indeed, looking inside the model "black box" and understanding the interplay of symmetry and data at every level of the architecture allows one to make informed decisions regarding which inductive biases are essential and which are superfluous, ultimately improving model performance.

II. QUANTIFYING SYMMETRY IN ML MODELS

We focus on a class of ML models that operate on decorated point clouds to predict physically relevant quantities [see Fig. 1(a)], even though similar considerations apply to more general settings, such as three-dimensional (3D) shape recognition or generative modeling. Common examples range from invariant quantities, such as potential energy in atomistic simulations, local pressure in fluid dynamics [34], or semantic class labels in vision models [35], to geometric tensors such as fluid velocity fields [34] and surface normals [36]. Treating the model as a generic smooth function f mapping an input x to an output y, physical consistency dictates that this output must transform predictably under the symmetry group G relevant to the system. When equivariance is hard coded into the architecture, these transformation laws are exact [see Fig. 1(b)]: The energy of a molecule is invariant under global rotation, while velocities rotate covariantly with the reference frame. Formally, a function f is equivariant with respect to a group G if $f(gx) = \rho(g)f(x)$ for all $g \in G$, where ρ is the representation of the group acting on the output space. For scalar targets (e.g., energy), $\rho(g)$ is the identity, recovering the

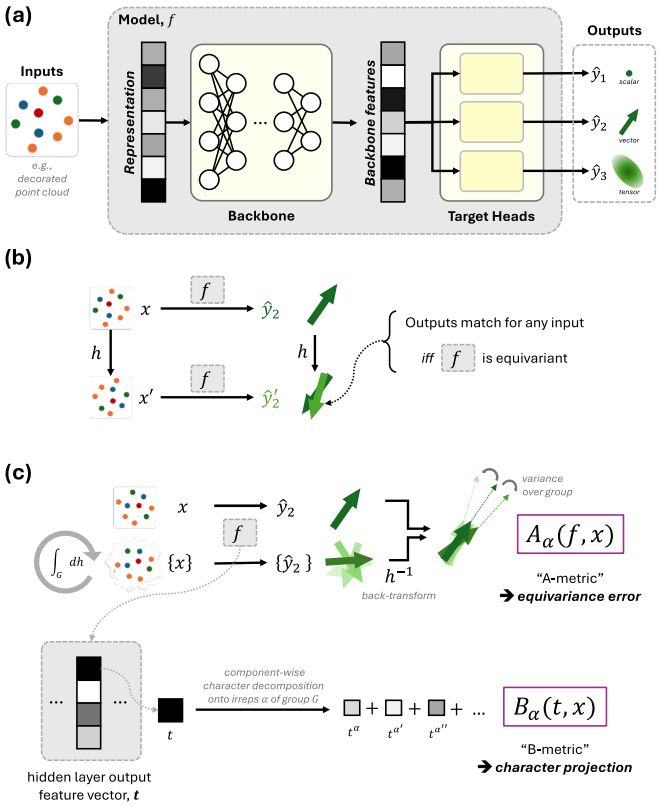

FIG. 1. Overview of the structure of a symmetry-aware ML model, the conditions of group equivariance, and the symmetry diagnostic metrics introduced in this work. (a) The ML model is represented by a generic smooth function, f, that predicts the physical properties (tensors of different rank), $\hat{y}$, of an input, e.g., a decorated point cloud, x. f can be a symmetry-preserving (i.e., equivariant) or unconstrained model. (b) Group equivariance is preserved if and only if the model predictions transform like the inputs under the action of the appropriate group symmetry operations. (c) The metrics A_α and B_α introduced in this work quantify the equivariance error of model predictions and the group symmetry content of internal features. For a set of inputs given by Haar integration of x over the group, the *equivariance error*, A_α, is given by the standard deviation of the back-transformed model predictions, while the *character projections*, B_α, give the group symmetry decomposition of model features from arbitrary layers.

definition of invariance: $f(gx) = f(x)$. Generally, if the output belongs to an irreducible representation (irrep) α of dimension d_α, the group action is mediated by the matrix $\rho_\alpha(g)$.

When equivariance is learned from data rather than enforced, it is inherently approximate. To quantify the fidelity of a model f to the transformation rules of a representation α, we use the *equivariance error* A_α:

$$A_\alpha(f,x) = \sqrt{\left\langle \left\| f(hx) - \langle \rho_\alpha(g^{-1}) f(ghx) \rangle_{g\in G} \right\|_2^2 \right\rangle_{h\in G}}. \quad (1)$$

Here, angled brackets $\langle\cdot\rangle$ denote the Haar average over the group with respect to the transformations of the input for any continuous group, or a group average for discrete groups. Note that the term $\rho_\alpha(g^{-1})f(gx)$ represents the output transformed back to the reference frame of x; for a perfectly equivariant function, this quantity is constant [equal to $f(x)$]. There-

fore, A_α effectively measures the standard deviation of the reprojected outputs over the group orbit [Fig. 1(c)]. This metric has already been introduced in Ref. [37] over a carefully designed subset of group elements, later used in Ref. [22] to estimate the equivariance error of forces predicted by the Point-Edge Transformer (PET) MLIP model, and in Ref. [38] to quantify approximate equivariance in unconstrained models trained with a penalty term enforcing equivariance. Reference [24] uses the same group-averaged variance to measure how much of the validation error comes from imperfect output equivariance, and follows this quantity during training. Here, we pair this output-level diagnostic with the character projections introduced below to resolve the irreducible symmetry content of intermediate representations. In the Supplemental Material [39], we provide proofs that this is a proper metric of equivariance as it vanishes if and only if f is exactly equivariant with respect to α. Practically, as shown in the Supplemental Material [39], this metric can be written as

$$A_\alpha(f,x) = \sqrt{\langle \|f\|_2^2 \rangle_G - \| \langle \rho_\alpha(g^{-1}) f(gx) \rangle_{g\in G} \|_2^2}, \quad (2)$$

which can be computed more efficiently, as it only requires a single group average and avoids the evaluation of the model f over compositions of group elements.

Output equivariance, however, is only the endpoint of the learning process. Two models with comparable values of A_α can rely on very different internal symmetry spectra, and these differences can determine whether a given architecture can efficiently learn targets with different angular order or parity. While the final output is optimized to be equivariant, the internal features of an unconstrained model are not bound to any specific symmetry constraints. Hidden layers may mix components of different equivariant characters while still effectively propagating information. To probe the symmetry content within the latent space, we introduce a spectral decomposition of the feature norm. Analogously to the power spectrum of a signal, we decompose the group-averaged norm of any internal feature t [e.g., the output of any hidden layer deep inside the model; see Fig. 1(c)] into contributions from different irreps:

$$\langle \|t\|_2^2 \rangle_G = \sum_\alpha B_\alpha(t,x). \quad (3)$$

Essentially, we look for a quantity that tells the fraction of t that transforms as the irrep α (more formally, the squared norm of the character-filtered components associated with the irrep α, consistent with the Peter-Weyl decomposition [42]). The quantity of interest here is the *character projection* B_α:

$$B_\alpha(t,x) = d_\alpha^2 \langle \| \langle \chi_\alpha(h^{-1}) t(hgx) \rangle_{h\in G} \|_2^2 \rangle_{g\in G}, \quad (4)$$

where $\chi_\alpha(g) = \mathrm{Tr}\, \rho_\alpha(g)$ are the group characters [43]. $B_\alpha(t,x)$ is effectively defined in terms of group convolutions [44–46]. For convenience, we will use normalized character projections, $B_\alpha(t,x)/\langle \|t\|_2^2 \rangle_G$, to simplify comparisons between different features and different layers of the architecture. Details about this quantity can be found in the Supplemental Material [39], including an efficient evaluation scheme to minimize the required number of calls to the model function. We remark that other approaches [47] can give complementary information to the metrics introduced here [48].

Both A_α and B_α are independent of the choice of reference frame for the input x (see the Supplemental Material [39]), which makes the two metrics intrinsic properties of the model and the input, rather than artifacts of a particular choice of orientation. Together, they provide complementary information on the symmetry of a model, with A_α being appropriate to measure how accurately the outputs of the model obey the desired symmetry, and B_α being able to assess the representation characters that contaminate the outputs, and more broadly to identify the *spectral content* of the model's internal features.

Although in this work we focus on O(3), the construction of A_α and B_α extends naturally to any compact group, including finite groups, by replacing the group average, irreducible representations, and characters with those of the symmetry group of interest. For compact groups such as SU(3), this generalization is therefore formal and direct. In contrast, for noncompact groups such as the Lorentz group SO(1, 3), a normalized Haar average over the full group is not available, so practical counterparts of the present diagnostics would require restricted or task-dependent sampling measures over the group orbit.

III. UNCONSTRAINED MODELS FOR ATOMISTIC SIMULATIONS

Surrogate models that target the outputs of quantum mechanical calculations of atomic-scale structures are an ideal starting point to demonstrate the conceptual and heuristic value of a symmetry analysis based on the introduced metrics. These models' inputs are usually atomic positions and chemical species, formally described as *decorated point clouds*. The point cloud edges are equivariant under any combination of global rotations, inversions, and translations, making the relevant group here the Euclidean group, E(3). Translational invariance is easily enforced by using interatomic distances rather than absolute positions, as is typical for graph-based models, so we can just focus on the orthogonal group O(3).

Physical properties targeted by these models naturally transform under (selected) irreps of O(3), and therefore unconstrained models must learn the appropriate transformation laws via data augmentation. We indicate the irreducible representations of the O(3) group by (λ, σ), where $\lambda = 0, 1, 2, \ldots$ is the *angular order*, and $\sigma = \pm 1$ the *parity* under inversion, which discriminates between the proper subspace ($\sigma = +1$) and its pseudocomplement ($\sigma = -1$). For example, scalars are proper invariants $(0, +1)$, vectors belong to $(1, +1)$, and pseudovectors to $(1, -1)$. In the Supplemental Material [39], we provide a detailed discussion of the explicit form of the equivariance error and the character projections for this group. In practice, the required group averages are computed using a product integration grid (i.e., a Lebedev and Laikov grid [49] on the sphere, plus a trapezoidal grid over the remaining Euler angle [43,50]), which affords a computationally efficient and exact evaluation of the metrics introduced earlier (see details and proofs in the Supplemental Material [39]).

We focus our analysis on the PET architecture, a transformer-based graph neural network (GNN) that takes both edge distances and vectors of an atomic-point cloud as inputs [13]. It has demonstrated excellent performance as an MLIP in both standard benchmarks [13,28] and materials

science applications [23,51,52], and it serves as a representative example of a growing family of transformer-based architectures in this domain [26,53]. Data augmentation during training helps the model learn the O(3) invariance of the potential energy surface (PES), leading to negligible equivariance errors compared to baseline accuracy, and to stable dynamics in the condensed phase [22].

We study the PET architecture in steps of increasing complexity to highlight different aspects of diagnostics and design. We start in Sec. III A by analyzing a universal MLIP trained on the MAD-1.5 dataset [54], targeting the potential energy [scalar, $(0,+1)$ irrep], atomic forces [vectors, $(1,+1)$ irrep], and cell stress [symmetric, rank-2 Cartesian tensor, $(0,+1)$ and $(2,+1)$ irreps]. Then, in Sec. III B, we track the spectral decomposition of angular information across various layers of the architecture, from random initialization through to the fully trained model. With the insights gained here, in Sec. III C, we propose a simple *post hoc* processing of the readout weights that further purifies the symmetry content of the outputs, leading to lower equivariance errors. Finally, in Secs. III D and III E, we investigate how the inductive biases built into PET affect learning dynamics for physical targets beyond the PES, and demonstrate how a symmetry analysis can inform simple but effective architectural modifications to improve learning dynamics and model accuracy.

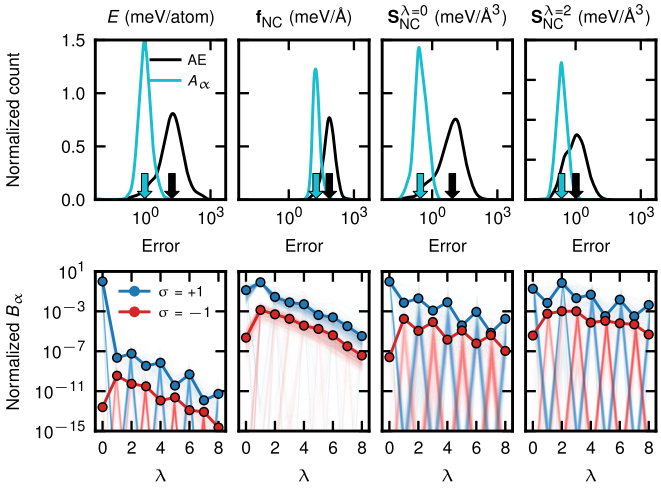

FIG. 2. Equivariance diagnostics for a PET MLIP. Top: distributions of the AE and equivariance error, A_α, for energy E, nonconservative forces $\mathbf{f}_{\mathrm{NC}}$, and nonconservative stress $\mathbf{S}_{\mathrm{NC}}$ (split by λ channel). The arrows on the x axis indicate the distribution medians. Bottom: normalized character projections, B_α, for the corresponding quantities as a function of the probed angular momentum channel λ. Solid lines and markers are averaged over 500 randomly sampled test structures, while faint lines show the individual structure projections.

A. Learning the potential energy surface

We trained a PET MLIP on the 1.5 version [54] of the Massive Atomic Diversity (MAD) dataset [55]. The model is trained from scratch for 2000 epochs to properly examine its symmetry content at random initialization and track its evolution during training. We first evaluate the trained model on energy, forces, and stress of a subset of the MAD-1.5 test split. The top panel of Fig. 2 shows a comparison of the resulting distribution of equivariance errors with that of the absolute model errors. In all cases, the distribution of equivariance errors is shifted toward smaller values than that of the absolute errors (AEs), with the median of their ratios being 10%, 31%, and 26% for energy, nonconservative forces, and nonconservative stress, respectively. We focus on forces and stresses predicted directly as outputs of the network to investigate the symmetry behavior of PET for higher-order irreps because, when they are computed as derivatives of the energy, their tensorial nature arises from the differential operators and is therefore directly tied to the invariant behavior of the energy.

Given that the outputs are not *exactly* equivariant, we can use the normalized character projections to investigate the nature of the errors [Fig. 2, bottom panel, showing $B_\alpha(y,x)$ for 500 randomly sampled test structures]. As expected, the scalar channel heavily dominates the energy predictions, with negligible contributions from higher angular momentum or pseudosubspace channels [56]. Forces manifest as vectors with active $\lambda=1$ channels, while the stress tensor (a symmetric Cartesian tensor of rank 2) activates both $\lambda=0$ and $\lambda=2$ proper channels [57,58]. The second important point to note is that pseudochannels are always less active than proper ones. This is dramatically evident in the case of energy, while the difference is less marked for forces and stress. In particular, the pseudoscalar $(0,-1)$ channel carries very low signal intensity in all cases, and does not follow the exponentially decaying trend seen for the higher-angular order terms of the pseudosubspace ($\lambda>0, \sigma=-1$).

One can get more insights by monitoring how the characters evolve during a training run. To this end, we introduce character projection heatmaps [Fig. 3(a)] to visualize the evolution of B_α for different angular momentum and parity channels, starting from the untrained, randomly initialized model all the way to the fully trained model. Dark colors indicate smaller components, and bright colors indicate larger components for each given channel. At any given vertical slice (i.e., epoch), the heatmaps visualize the spectral decomposition analogous to those seen in the bottom panels of Fig. 2.

For each of the outputs of the PET model along training, we monitor the test-set model error, equivariance error, and character projection heatmaps [Fig. 3(b)]. One can see that the randomly initialized model has a strong bias toward scalar $(0,+1)$ character and has near-zero spectral content for $\lambda>2$ and/or $\sigma=-1$. Then, we observe a sudden transition that is most evident for the nonconservative force and $\lambda=2$ nonconservative stress predictions: The scalar character drops, and the $(1,+1)$ and $(2,+1)$ characters become dominant, respectively, accompanied by a sudden jump in high λ and $\sigma=-1$ components. During later epochs, as both the target root-mean-squared-error (RMSE) and equivariance errors decrease, the character decomposition becomes more sharply peaked around the desired irreps, without other sudden transitions apart from the appearance of the $\lambda=2$ stress component at epoch ∼200. The scalar stress component is the only target whose equivariance error increases during training from its initial value: The reason is that at the beginning of the training the scalar stress predictions are close to zero, which is trivially invariant. The late onset of learning for the nonconservative

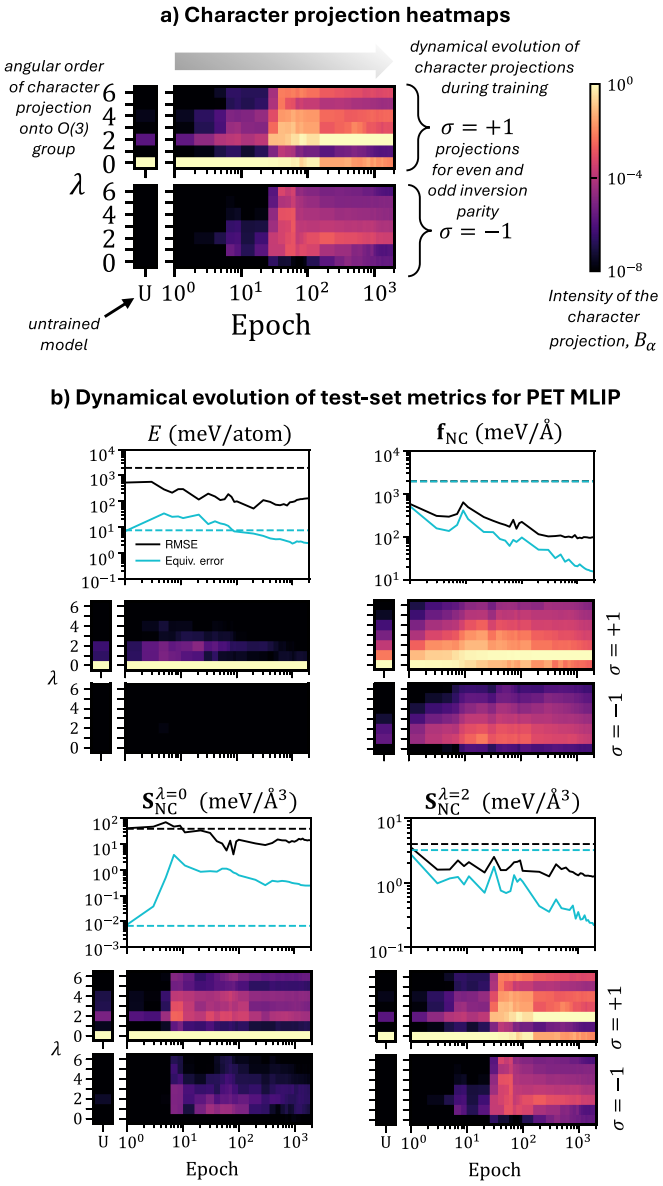

FIG. 3. (a) Character projection heatmaps report on the magnitude of the character projection B_α for a quantity as a function of the character α of the relevant group and along successive epochs of a training run. In this case, the characters are the (λ, σ) irreps of the O(3) group. (b) Training curves and character heatmaps for energy, E, nonconservative forces, $\mathbf{f}_{NC}$, and the two irreducible spherical components of the nonconservative stress, $\mathbf{S}_{NC}$. For the training curves, the dotted line represents the initial RMSE (black) or equivariance error (blue) of the untrained model.

stress implies that the model is not fully converged at the end of the run, with consequences that will be apparent in Sec. III C.

B. Looking into the black box

The character decomposition analysis of PET demonstrates that there is a lot of structure in the symmetry-breaking errors and that the model performs especially well on invariant targets. To understand *why*, and to improve PET's ability to learn targets that are not invariant, we have to look deeper at *how* it learns symmetry. To do so, we explore the information content of PET features on the two aforementioned axes: across layers of the architecture and throughout the training process.

Figure 4 introduces a high-level overview of the PET MLIP architecture. For the outputs of several important hidden layers of the model trained in the previous section, we perform a character decomposition analysis, with the resulting heatmaps annotating the architecture. The atom-centered environments of a molecule or material form the input to the architecture as decorated point clouds [Fig. 4(a)]. Internal representations of these environments are generated by the backbone PET architecture. This is comprised of a series of message-passing GNN layers, each of which contains several modules [Fig. 4(b)]. The resulting backbone features (BBF) contain rich geometric information common to all model outputs. These are transformed into last-layer features (LLF) by target-specific nonlinear heads, which are mapped to model outputs by linear readout layers. At the coarsest level, the internal representations constructed by the backbone architecture evolve with rich dynamics during training, developing high λ as well as pseudotensorial ($\sigma = -1$) characters, which are especially strong for the edge features.

In analyzing the finer details of the backbone architecture, we can see how the character trends we observed for the model outputs develop. For a given local atomic environment in the batch, the chemical nature of the central atom and its neighbors are embedded as learnable scalar tokens [center embeddings (CE) and neighbor embeddings (NE)], with geometry embeddings (GE) representing the local 3D structure—specifically, the distances r_{ij} and interatomic vectors $\mathbf{r}_{ij}$ to each neighbor [Fig. 4(b)]. These GE have pure scalar and vectorial nature, which does not change during training (heatmaps A). A multilayer perceptron converts the GE into fixed-size edge embeddings (EE). The nonlinearity generates higher-λ characters, but still lacks pseudotensorial character (heatmaps B).

The transformer blocks, at the heart of the PET architecture, are capable of generating rich geometric features thanks to the attention mechanism, which mixes information from multiple edges and populates the pseudotensorial channels (heatmaps C–F). This is especially true for edge features, though the signal still remains weak. After two transformer layers, the first GNN layer is complete, and at this point the model has been able to combine geometric information within a single neighborhood. Even at initialization, with random weights, the outputs of the GNN block are dominated by the scalar character, but contain significant higher-λ components (heatmaps E and F). The message-passing step combines geometric information with edge messages from the previous step. As such, the pseudotensorial activation gained in the first GNN layer propagates to the next and is therefore visible in the new EE (heatmap G), while the outputs of the second GNN layer (heatmaps H) show no other notable changes relative to those of the first GNN layer. The final BBF are obtained after the message passing block of the second GNN layer. The edge BBF in general carry much richer angular information than the node BBF.

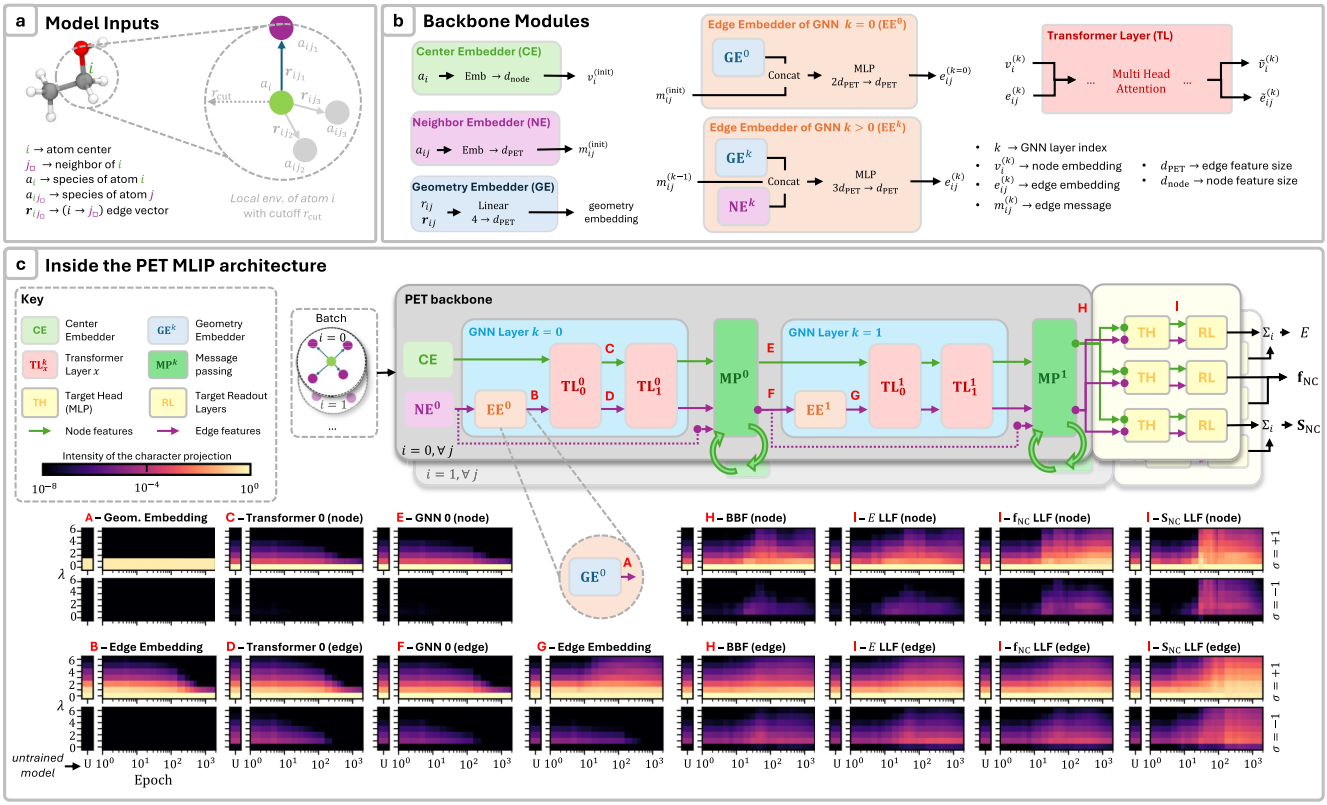

FIG. 4. Overview of a PET MLIP architecture and dynamical evolution of the internal features. (a) Local atomic environments of atoms in a molecule/material are represented as decorated point clouds. The atomic species of the central atom (a_i) and its neighbors (a_{ij}), together with the edge vectors ($\mathbf{r}_{ij}$) and their magnitudes (r_{ij}), form the inputs to PET. (b) Backbone modules: Embedding modules map the inputs to the latent space via center, neighbor, and geometry embedders. These feed into one edge embedder per GNN layer, which then enters the transformer layers. (c) Full architecture: The complete model pipeline leading to predictions for energy E, nonconservative forces $\mathbf{f}_{\mathrm{NC}}$, and nonconservative stress $\mathbf{S}_{\mathrm{NC}}$. The following architecture hyperparameters were chosen: number of GNN layers $n_{\mathrm{GNN}} = 2$, number of transformer layers $n_{\mathrm{TL}} = 2$, cutoff $r_{\mathrm{cut}} = 4.5$ Å, edge feature size $d_{\mathrm{PET}} = 128$, node feature size $d_{\mathrm{node}} = 256$, and LLF size $d_{\mathrm{head}} = 128$. The surrounding heatmaps, labeled with red letters (A–I), correspond to specific points in the architecture. They display the intensity of the normalized character projections, $B_\alpha / \langle \|t\|_2^2 \rangle_G$, as a function of the training epoch (x axis, log scale) and the probed λ channel (y axis). Within each heatmap group, the top and bottom panels represent the σ channels, respectively (as detailed on the right of heatmaps I). The thin, isolated column on the far left of each heatmap represents the untrained model, denoted by "U" on the x axis.

The BBF are then split between several target-specific nonlinear heads and mapped to the LLF (heatmaps I). These are the last nonlinear operations in the network. In the randomly initialized model, the LLF have only low-order character, but at later stages they develop a richer geometric structure, that persists until the end of the training. This can again be seen particularly for edges for each target's LLF. Between targets, the symmetry content of the LLF features differs considerably: For instance, the energy LLF have greater suppression of high-angular order features compared to the forces and stress. Finally, the LLF are mapped to predictions via the linear readout layers. These can only modulate the magnitude of the existing components, and act as a filter that determines the final characters and equivariance errors of the predictions. Edge predictions are aggregated over neighbor atoms and, if the output is a per-structure quantity (energy or stress), both node and edge predictions are aggregated over central atoms. By summing node and edge outputs, the final predictions as seen previously in Fig. 3 are obtained.

This detailed analysis provides a compelling picture of how PET learns to be (approximately) equivariant. It also gives indications that could be useful for the design of rigorously equivariant models, because the character decomposition learned by the unconstrained architecture suggests what types of symmetry channels are needed to efficiently learn energy, forces and stresses. It shows that, with the GE we use, PET is strongly biased toward low-λ components, and that even if the network has the expressive power to generate higher-order character features, it uses this power sparingly, converging to intermediate representations dominated by low λ, $\sigma = +1$ character. This is consistent with the observation that equivariant networks can often yield reasonably accurate predictions with internal representations capped at $\lambda_{\max} = 2$ [59]. The presence of high-order components in the edge representations suggests, however, that a higher resolution is important to process geometric message-passing information, so it might be useful to experiment with equivariant models that invest

some computational budget into including higher-λ terms, at least for some of the GNN layers.

C. Symmetry purification of the readout

In the previous section, we observed that the final readout layers must filter out undesired symmetry content from the LLF. This is achieved through symmetry augmentation with random rotations during training, resulting in a model with equivariance error that is a small fraction of the model error. Still, the LLF tokens are contaminated by a significant fraction of irreps not needed to describe the target subspace. We propose here a simple protocol to purify the linear readout, which can be formulated as a regularized regression problem. If $\phi(x)$ are the LLF for the input x, the linear readout can be expressed as $y(x) = \theta^T \phi(x)$. By computing the LLF over a group orbit [or a grid, in the case of continuous groups such as O(3)], we can define two loss terms,

$$L_\mu = \left\langle \left\| \rho(g^{-1})\theta^T \phi(gx) - y \right\|_2^2 \right\rangle_{g\in G},$$
$$L_\sigma = \left\langle \left\| \theta^T \phi(gx) \right\|_2^2 \right\rangle_{g\in G} - \left\| \langle \rho(g^{-1})\theta^T \phi(gx) \rangle_{g\in G} \right\|_2^2. \quad (5)$$

The first describes the model mean-square error averaged over the group, while the second corresponds to the squared equivariance error A^2. The combined loss $L = L_\mu + \gamma L_\sigma$, where γ controls the relative weighting of the equivariance penalty, can be cast into an explicit least-squares problem and solved for the readout weights θ. All the terms needed to determine θ can be computed with a single sweep over the training set. The derivation of a closed form for the optimal weights that minimize this loss is reported in the Supplemental Material [39].

For the PET model, we find that this procedure yields a modest improvement of the equivariance error for most of the outputs, confirming that on-the-fly augmentation performed during conventional model training succeeds in minimizing A (Fig. 5). The exception is the scalar component of the nonconservative stress. In this case, both the RMSE and the equivariance error A improve substantially, by up to about 5% and more than 10%, respectively, relative to the base model, with the largest improvements occurring at different values of γ. We hypothesize that this indicates that the training did not fully optimize the readout for this target. For the other targets, we can observe that by tuning the weight of L_σ, the equivariance error can be reduced by up to 5% (or even more for the symmetric stress) by sacrificing less than 1% in the RMSE relative to the target. The inclusion of an explicit equivariance penalty during primary training, as in Ref. [38] would also influence the optimization of the backbone weights, but requires multiple model evaluations per structure, increasing the computational cost of training. Instead, the equivariant readout optimization we propose here has a cost equivalent to running symmetrized inference only once on each structure in the dataset, and can be applied routinely as a way to validate, and possibly improve, the symmetry properties of unconstrained models.

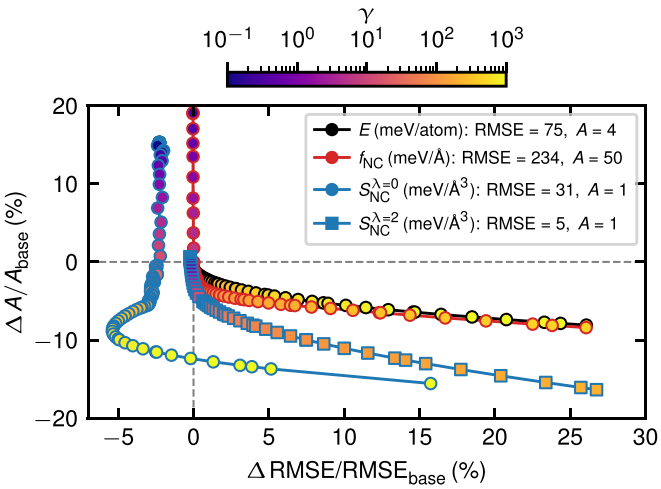

FIG. 5. Model RMSE and equivariance error (A) for the energy E, nonconservative forces $\mathbf{f}_{\rm NC}$, and $\lambda = 0, 2$ components of the nonconservative stress $\mathbf{S}_{\rm NC}$ for a universal PET model (trained and tested on MAD-1.5) whose energy, nonconservative force, and nonconservative stress readout layers have been retrained with a loss $L = L_\mu + \gamma L_\sigma$ combining model error and an equivariance error penalty. The marker colors correspond to the weighting of the equivariance penalty, γ. The percentage values, given relative to the RMSE and A of the standard model, are reported in the legend.

D. Stress testing the geometric expressivity of the model

The fact that the internal representations are dominated by low-order characters raises the question of whether PET can learn targets with a symmetry that is not naturally expressed by the initial random weights, which we are going to investigate next. The clearest signal from the symmetry analysis of the model features is that $\sigma = -1$ irreps, and pseudoscalars, in particular, have negligible character throughout the PET architecture. To determine whether this impacts the representation power of the model, we define a purely geometric pseudoscalar target Q as the sum of triple products of all triplets of bonds in a molecular system:

$$Q = \sum_{i<j<k<l} (\mathbf{r}_i - \mathbf{r}_l) \cdot [(\mathbf{r}_j - \mathbf{r}_l) \times (\mathbf{r}_k - \mathbf{r}_j)], \quad (6)$$

and generate a toy dataset of 1000 rattled conformers of bromochlorofluoromethane (CHBrClF) differing only in initial orientation and small random coordinate distortions. We trained PET from scratch using an 80:10:10 split.

The top panels of Fig. 6 show the character projection heatmaps for the internal features (EE, BBF, LLF) and predictions (Q) over the training epochs, while the bottom panel reports the RMSE and equivariance error (A_α). The network exhibits a striking two-phase learning behavior along the training, similar to the so-called grokking phenomenon often observed in other models [60]—albeit in this case on both training and validation sets. From this point of view, this observation suggests that the ability of the network to generate high-order symmetry terms starting from simple vectorial descriptors might be interpreted as a phase transition in the parameter space of the transformer [61]. During the initial phase (up to ∼120 epochs, red dashed line), only

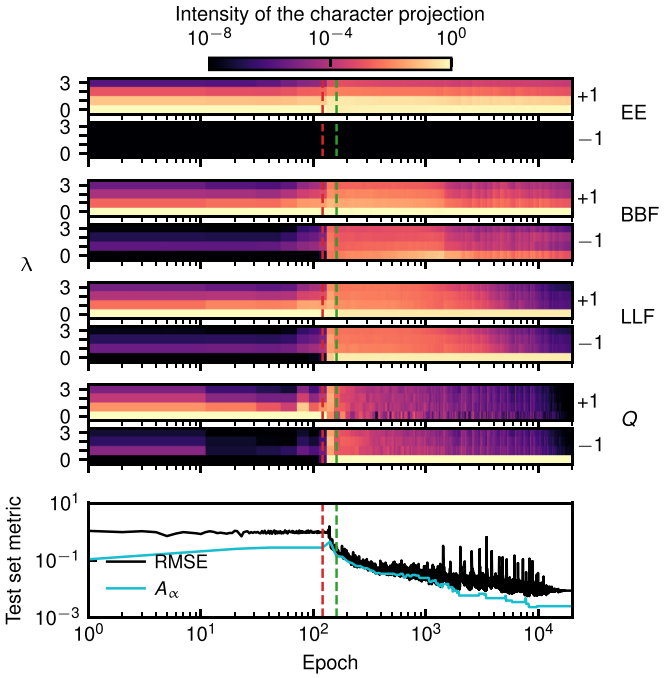

FIG. 6. Equivariance diagnostic for the geometric pseudoscalar training. Top: character projection heatmaps for a PET model across architectural layers over the course of training. Bottom: test-set RMSE and average equivariance error. The region between the red and the green dashed lines marks the onset of rapid learning. This phase is driven by a strong activation of the pseudoscalar channel across all capable layers, accompanied by a broader, weaker activation of higher-order tensorial ($\sigma = +1$) and pseudotensorial ($\sigma = -1$) channels (indicated on the right).

proper ($\sigma = +1$) channels are active, leaving pseudofeatures virtually dormant. Because the proper (and especially the purely invariant) channels lack the necessary symmetry to express Q, both the test RMSE and A_α stagnate. A dramatic shift occurs between the red and green dashed lines, where pseudotensorial channels suddenly activate across the BBF, LLF, and output (Q). This internal transition triggers a steep and continuous drop in both RMSE and A_α. As expected, the EE features show no pseudotensorial activation because at this stage—just before the transformer—the model processes edges individually and cannot yet combine them. Constructing a pseudotensor requires mixing at least two vectors (a "second-order" effect), while forming a pseudoscalar requires combining at least three (a "third-order" effect; see the Supplemental Material [39] for details). The network must synthesize these pseudorepresentations from strictly proper inputs within the attention layer, leading to a delayed onset of learning. Because pseudochannels are initially weak, their gradient signals are suppressed and their construction requires coordinated higher-order interactions across multiple features, making optimization intrinsically slow until a threshold is reached where these components can self-amplify. Given that Q is a purely geometric target, somewhat trivially defined from molecular geometry, it is remarkable that even in this case the learning task is so difficult.

This diagnostic shines light on how architectural design restricts the model's ability to build features with a specific group symmetry and how this may impact learning dynamics. Underrepresentation of the pseudosubspace carries implications beyond artificial tasks, potentially impacting real-world atomistic applications, such as learning of NMR chemical shielding tensors [32] (which require pseudovectorial irreps), circular dichroism [62], or targeting Hamiltonian and density matrices that contain many components with $\sigma = -1$ character [63–66].

E. Increasing the geometric expressivity of PET

Next, we challenge the PET model to learn a high-λ target. For this, we start with the real-space electron densities, $\rho(\mathbf{r})$ (defined on a set of grid points, $\mathbf{r}$) for the spin singlet, charge neutral, organic (covering 11 elements) subset of the QCML molecular dataset [67] computed with FHI-aims [68]. The densities are then projected onto an atom-centered auxiliary basis set (a product of spherical harmonics and radial functions), $\varphi_{in\lambda\mu}$, to yield the *electron density projections*, $\mathbf{w}$:

$$\mathrm{w}_b = \langle \varphi_b(\mathbf{r}) | \rho(\mathbf{r}) \rangle, \tag{7}$$

where $b \equiv \{in\lambda\mu\}$ is a composite index labeling the atom i, radial channel n, angular order λ, and component μ. The auxiliary basis set is chosen to be expressive enough to expand the electron density with low basis-set expansion error as $\rho(\mathbf{r}) \approx \sum_b c_b \varphi_b(\mathbf{r})$, with $c_{b'} = \sum_b S^{-1}_{bb'} \mathrm{w}_b$ being the expansion coefficients and S the auxiliary basis overlap matrix. The electron density projections form a challenging target derived from a quantum mechanical observable that is central in atomistic modeling and, crucially, decompose onto a series of proper-tensorial irreps up to some maximum angular order, $\lambda^{\mathrm{target}}_{\mathrm{max}} = 8$ in our case. By isolating and targeting only the highest-λ channel, we can observe clearly how the absence of high-angular order features at initialization affects learning in a realistic scenario.

Figure 7(a) tracks the architecture's attempt to learn the $(8, +1)$ components of the density projections. As shown by the entirely flat RMSE curve over 2000 epochs, standard PET fails to learn this target. Because the network only embeds edge distances and vectors, the initial GE strictly activate at $\lambda = \{0, 1\}$. While the nonlinear multilayer perceptron generates some higher-λ activation in the EE, it is insufficient. Lacking meaningful $\lambda = 8$ information in its features, the model collapses to a trivial solution: It predicts zeros to minimize the loss, as evidenced by the declining equivariance error and a vanishing norm ratio between predictions and targets.

We propose to address this deficiency in the architecture by a simple and inexpensive modification to the edge geometry embeddings. We replace standard edge distances and vectors with a solid spherical harmonics (SSHs) expansion of the edge vectors up to a maximum angular order, $\lambda^{\mathrm{emb}}_{\mathrm{max}}$. SSHs naturally represent both angular structure and radial information as each carries a well-defined angular momentum channel by construction, and can be evaluated in an autodifferentiable manner with negligible overhead [69]. In fact, due to the functional form of the SSHs, $r^{\lambda}_{ij} Y^{\mu}_{\lambda}(\mathbf{r}_{ij})$, the GE of standard PET is almost equivalent in information content to that of an (embedded) SSH expansion capped at $\lambda^{\mathrm{emb}}_{\mathrm{max}} = 1$. By increasing this cap, we provide the model

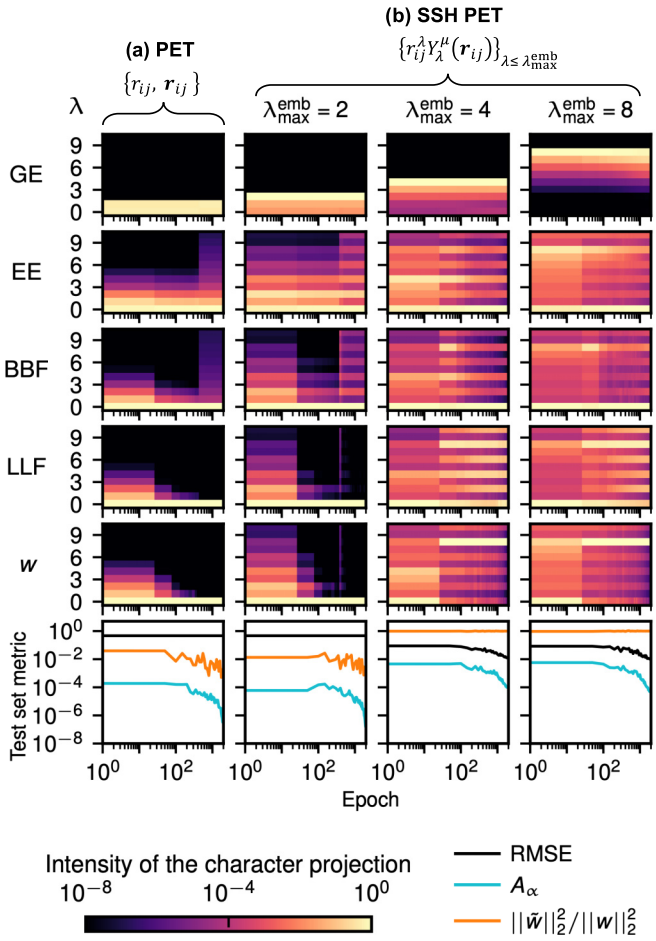

FIG. 7. Impact of higher-angular order inductive biases on learning high-angular order targets. Standard PET (first column) and modified variants—which replace standard embeddings with solid spherical harmonics up to $\lambda_{\max}^{\mathrm{emb}} \in \{2, 4, 8\}$—are trained on the $\lambda_{\mathrm{target}} = 8$ channel of the electron density projections (w). Top: proper-tensorial ($\sigma = +1$) heatmaps tracking the angular order ($\lambda \in \{0, \ldots, 10\}$) evolution of features across key network layers during training. Colors denote the normalized intensity of the character projection. Bottom: validation metrics per epoch, including prediction RMSE (black), equivariance error (cyan), and the predicted-to-target norm ratio (orange).

with a vastly superior starting point to describe high-λ quantities.

To investigate the effect of this modified inductive bias, we trained from scratch the same PET architecture as in Fig. 7(a), modifying only the GE layer to use SSH embeddings of increasing maximum angular order $\lambda_{\max}^{\mathrm{emb}} \in \{2, 4, 8\}$. As shown in Fig. 7(b), explicitly increasing the input's angular information drastically alters the learning dynamics. While $\lambda_{\max}^{\mathrm{emb}} = 2$ behaves similarly to standard PET and fails to learn, models initialized with $\lambda_{\max}^{\mathrm{emb}} = 4$ and 8 succeed. For these models, $\lambda = 8$ character is visible in B_α right from the start, the RMSE steadily declines, and the norm ratio stabilizes around 1, proving the model is genuinely learning the target rather than collapsing to 0. By the end of the training, the internal features possess rich, high-order angular information, successfully adapting the implicit inductive biases of the architecture and the geometric nature of the physical target.

IV. BEYOND ATOMISTIC SIMULATIONS

Even though we focus our analysis on the application of unconstrained ML models to atomistic simulations, there are many domains of physics that rely on models that do not obey exactly all the relevant symmetries. As a brief but instructive example, we consider the classification of particle trajectories in liquid argon time projection chambers (LArTPCs) neutrino detection experiments. We investigate the symmetry properties of the point-based liquid argon masked autoencoder (PoLAr-MAE) method [19], which relies on an unconstrained architecture similar to the general-purpose PointNet [20] and is built upon the publicly available PILArNet dataset [70]. Particle trajectories are represented as a point cloud, with each point corresponding to a detector interaction, decorated by the energy of the detected particle. Based on this input, the model classifies points along a particle trajectory into four classes: electron shower, track, Michel electron, δ-ray (cf. Fig. 8). PoLAr-MAE shows overall excellent classification accuracy, except for a few fine-grained structures. We observe that, even though the classification is usually invariant to rigid rotations of the reference frame (the model is trained with random rotational augmentation, similar to PET), there are a few track segments for which the class becomes dependent on orientation, especially for the subtoken structures where the classification is unreliable in the first place.

In Fig. 8, we show a summary of the architecture of the model and in-depth diagnostics of the internal representation. In the figure, we also show the pointwise equivariance error of the class probabilities for a representative event (the same depicted in Fig. 9). A pointwise character decomposition of the internal descriptors for this same event is reported in the Supplemental Material [39]. One sees clearly that the uncertainty in the classification is reflected in the equivariance errors for the four channels predicted by PoLAr-MAE. The presence of a substantial nonscalar behavior is already apparent from the character decomposition of the segmentation outputs (heatmap F) with a dominant scalar nature and exponentially decaying λ characters. The pseudocharacters are all much smaller as it was the case for PET and also in this case the pseudoscalar contribution is minuscule.

Looking at the internal layers, we see that the position embeddings A have predominantly vectorial character (i.e., the position embedder does not generate high-order characters), whereas the local group tokens, which are processed by a mini-PointNet unit, display a richer structure, with a substantial $\lambda = 2$ component (heatmap B). The transformer encoder blocks (heatmaps C–E) do not display remarkable trends: The dominant character is scalar, and successive blocks progressively dampen the higher-λ characters. However, the convolutional segmenters increase the contamination from nonscalar characters. Overall, PoLAr-MAE shows trends that are broadly similar to those observed for node features in PET, missing however the richer, high-λ characters of PET's edge features, suggesting that introducing explicit edge representations might increase the expressive power of the model. Large equivariance errors are associated with instabilities

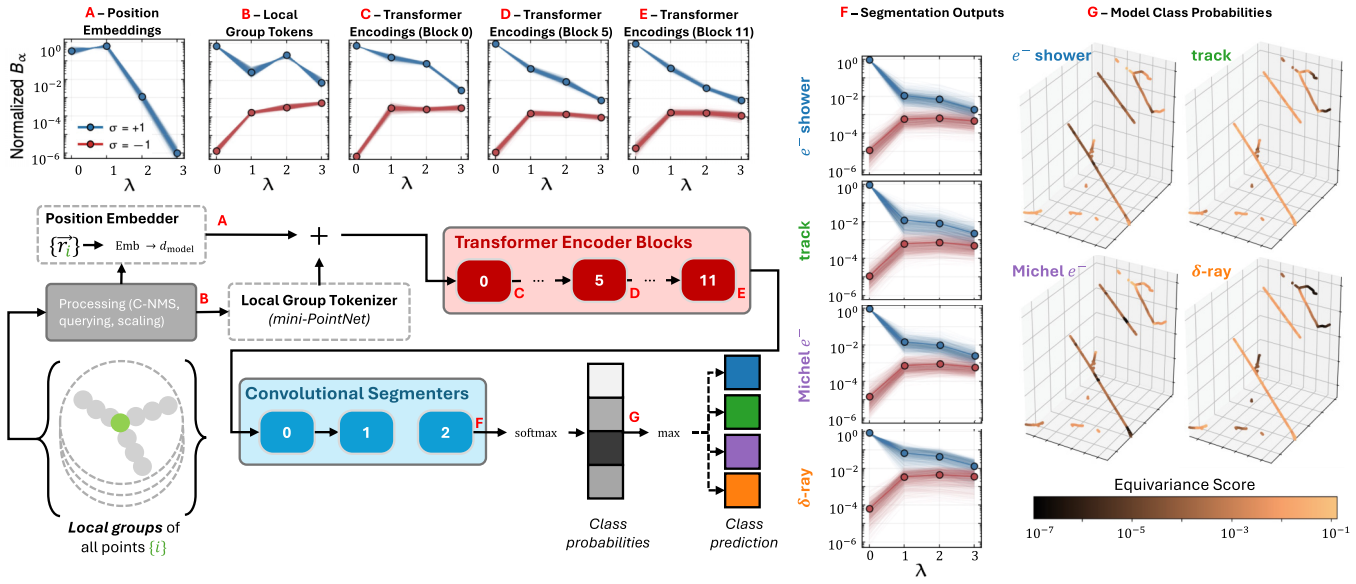

FIG. 8. Schematic representation of the PoLAr-MAE architecture, with the character decomposition of the internal tokens and the segmentation outputs plotted for 1000 analyzed events. For representation purposes, we show the decomposition after aggregation over the points of each event. Equivariance errors for the model class probabilities are shown for a representative event (the same as in Fig. 9), showing how values of A_α correlate either with branching points or trajectory segments for which the classification is unstable with respect to rotations or reflections of the reference frame.

of the classification. While introducing inference-time rotational augmentation could reduce the errors, a more elegant and computationally efficient solution would be to apply the readout purification protocol introduced in Sec. III C to the final convolutional segmenters. By explicitly penalizing the equivariance error of the classification head over the O(3) group, one could enforce rotational invariance for these challenging subtoken structures without incurring inference-time overhead.

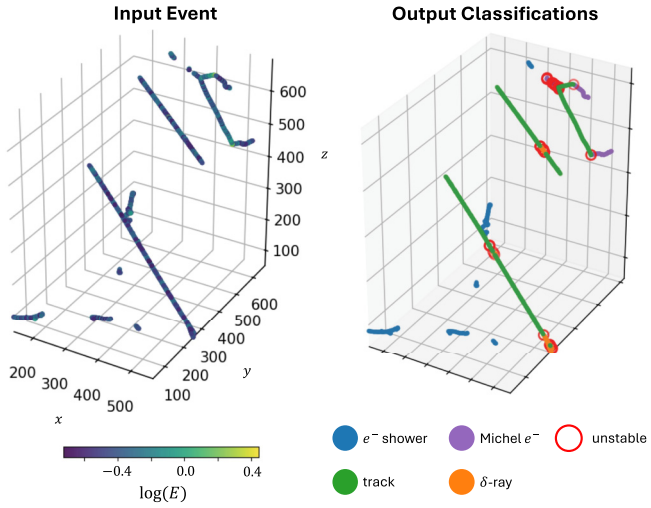

FIG. 9. A representative input and output of the PoLAr-MAE model. For a given input event (left panel), e.g., clusters of particle tracks (point clouds) in a liquid argon detector decorated with particle energies, the model makes pointwise classifications (right panel) of the discrete event class. Points surrounded by a red circle change classification under at least one sampled rotation or reflection. Axis labels are given in terms of voxel indexes.

V. CONCLUSIONS

Many fields of application, from image recognition to natural language processing, have come to accept the *bitter lesson* that incorporating domain-field knowledge into the design of machine-learning models does not always pay off, in the long run, against architectures designed to be simple, fast, and able to scale to large datasets. For a relatively long time, physics appeared to be the exception, with considerable effort being devoted, with excellent results, to the design of symmetry-adapted models that are bound to fulfill invariant or covariant behavior with respect to symmetry operations by design. The recent success of unconstrained architectures, however, shifts the focus from enforcing all symmetries by construction to understanding which parts of the symmetry structure can be learned reliably, and which should be imposed through the architecture, the objective, or targeted descriptors. Because of the foundational role played by symmetry, this prompts the questions of *how*, and *how accurately*, models without these built-in constraints come to learn symmetry. We develop an analysis framework that allows one to follow the character of the outputs as well as the internal representation of the model across the architecture and during a training run, and that is useful across domains, as we demonstrate by applying it to rotational symmetry in both atomic-scale simulations and the classification of particle trajectories from LArTPC experiments. In both cases ML models can learn rotational symmetry to a high accuracy, with symmetry errors being much smaller than model error. In fact, the median ratios of the two types of error are typically between 10% and 30%.

We propose an inexpensive purification step for the linear readout of the model. We find that the PET-based atomistic model we consider is already trained to an excellent level using random augmentation, and that the improvements from this simple postprocessing step are modest except for the scalar component of the stress for which the equivariance error can be reduced substantially. The group characters of the internal representations of the model display highly nontrivial dynamics during training. Initially, low-angular-momentum λ character dominates across the network, which we trace in part to the initial encoding of the geometry in terms of scalars and vectors. Higher-λ terms appear soon for the internal representation, especially for the tokens associated with the interatomic edges, and only toward the end of the run the target properties achieve an accurate approximation of the desired symmetries. The stress, which being a symmetric Cartesian tensor contains a $\lambda = 2$ component, is learned only in this later stage. This observation prompted us to investigate the performance of the model when learning pseudoscalar and high-λ target properties, associated with characters that are poorly represented in the network. We find that indeed the standard PET architecture struggles to learn these symmetries—a deficiency that we address by injecting higher-λ descriptors at the level of the initial structural encoding, demonstrating the heuristic power of our symmetry analysis.

The analysis of the particle trace classification model is broadly similar. The final features show a high degree of rotational invariance, and equivariance error correlates strongly with instabilities in the classification—indicating the importance of training the model to a high level of symmetry. Internal features show a predominantly scalar character throughout the model, which appears beneficial at least when performing a task that is based on invariant outputs.

A character analysis such as the one we introduce helps to select between different architectures and architectural hyperparameters, and to diagnose difficulties in learning symmetry behavior. Some of the insights we extract might also be relevant for models that are equivariant by design. The fast decay of learned representation characters in the PET architecture validates the common choice of limiting the angular momentum order to $\lambda = 1$ or $\lambda = 2$ in the internal representation of equivariant MLIPs. When using the same architectures for high-λ targets, however, it might not be enough to increase the order at the level of the heads, as we observe that the lack of some components early in the network negatively affects the accuracy of outputs of the corresponding symmetry, despite the fact that the nonlinear heads we use can in principle generate terms of arbitrary order.

The tension between general-purpose, flexible, scalable architectures and architectures that incorporate exact physical priors is unlikely to be resolved soon. In the meantime, the analysis framework introduced here provides a rigorous approach to quantify the symmetry behavior of ML models, yielding actionable insights that improve the accuracy and training dynamics of unconstrained architectures across the physical sciences. As demonstrated by our architectural modifications for learning the ground-state electron density, understanding a model's failure modes allows researchers to achieve stability and accuracy by injecting the minimum required inductive biases, without sacrificing the expressivity and scalability of unconstrained architectures.

ACKNOWLEDGMENTS

The authors would like to acknowledge support by the following grants: the Swiss National Science Foundation (Grant No. 200020_214879 to M.D. and M.C.), the ERC Horizon 2020 Grant No. 101001890-FIAMMA (to J.W.A. and M.C.), the Swiss Platform for Advanced Scientific Computing (PASC) and the Swiss AI Initiative (2025 Fellowship Program) (to F.B.), and a Sinergia grant of the SNSF (Grant No. CRSII5_202296 to P.P.). Computation for this work relied on resources from the Swiss National Supercomputing Centre (CSCS) under the lp133 and lp161 projects, and the EPFL HPC platform (SCITAS).

DATA AVAILABILITY

All software components used in this study are open source and freely available. The PYTHON package used to implement and train our ML models is metatrain, available on GitHub in Ref. [71]. The complete set of data and workflows required to reproduce all figures in this manuscript is provided in an online data repository [72] hosted on the Materials Cloud platform [73].

[1] E. Noether, Invariante variationsprobleme, Nachr. Ges. Wiss. Goettingen, Math.-Phys. Kl. **1918**, 235 (1918).

[2] D. J. Gross, The role of symmetry in fundamental physics, Proc. Natl. Acad. Sci. USA **93**, 14256 (1996).

[3] A. P. Bartók, M. C. Payne, R. Kondor, and G. Csányi, Gaussian approximation potentials: The accuracy of quantum mechanics, without the electrons, Phys. Rev. Lett. **104**, 136403 (2010).

[4] A. Bogatskiy, T. Hoffman, D. W. Miller, and J. T. Offermann, Pelican: Permutation equivariant and Lorentz invariant or covariant aggregator network for particle physics, arXiv:2211.00454.

[5] C. Li, H. Qu, S. Qian, Q. Meng, S. Gong, J. Zhang, T.-Y. Liu, and Q. Li, Does Lorentz-symmetric design boost network performance in jet physics? Phys. Rev. D **109**, 056003 (2024).

[6] M. Favoni, A. Ipp, D. I. Müller, and D. Schuh, Lattice gauge equivariant convolutional neural networks, Phys. Rev. Lett. **128**, 032003 (2022).

[7] G. Aarts, K. Fukushima, T. Hatsuda, A. Ipp, S. Shi, L. Wang, and K. Zhou, Physics-driven learning for inverse problems in quantum chromodynamics, Nat. Rev. Phys. **7**, 154 (2025).

[8] I. Batatia, M. Geiger, J. Munoz, T. Smidt, L. Silberman, and C. Ortner, A general framework for equivariant neural networks on reductive Lie groups, in *Advances in Neural Information Processing Systems*, edited by A. Oh, T. Naumann, A. Globerson,

K. Saenko, M. Hardt, and S. Levine (NeurIPS/Curran Associates, Inc., Red Hook, NY, 2023), Vol. 36, pp. 55260–55284.

[9] J. Behler and M. Parrinello, Generalized neural-network representation of high-dimensional potential-energy surfaces, Phys. Rev. Lett. **98**, 146401 (2007).

[10] A. V. Shapeev, Moment tensor potentials: A class of systematically improvable interatomic potentials, Multiscale Model. Simul. **14**, 1153 (2016).

[11] R. Drautz, Atomic cluster expansion for accurate and transferable interatomic potentials, Phys. Rev. B **99**, 014104 (2019).

[12] I. Batatia, D. Kovács, G. Simm, C. Ortner, and G. Csányi, MACE: Higher order equivariant message passing neural networks for fast and accurate force fields, in *Advances in Neural Information Processing Systems*, edited by S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh (Neural Information Processing Systems Foundation, Inc., La Jolla, CA, 2022), Vol. 35, pp. 11423–11436.

[13] S. Pozdnyakov and M. Ceriotti, Smooth, exact rotational symmetrization for deep learning on point clouds, in *Advances in Neural Information Processing Systems* (NeurIPS/Curran Associates, Inc., Red Hook, NY, 2023), Vol. 36, pp. 79469–79501.

[14] S. N. Pozdnyakov, M. J. Willatt, A. P. Bartók, C. Ortner, G. Csányi, and M. Ceriotti, Incompleteness of atomic structure representations, Phys. Rev. Lett. **125**, 166001 (2020).

[15] S. N. Pozdnyakov and M. Ceriotti, Incompleteness of graph neural networks for points clouds in three dimensions, Mach. Learn.: Sci. Technol. **3**, 045020 (2022).

[16] J. Abramson, J. Adler, J. Dunger, R. Evans, T. Green, A. Pritzel, O. Ronneberger, L. Willmore, A. J. Ballard, J. Bambrick, *et al.*, Accurate structure prediction of biomolecular interactions with AlphaFold 3, Nature (London) **630**, 493 (2024).

[17] N. Dormagen, M. Klein, A. S. Schmitz, L. Wimmer, M. H. Thoma, and M. Schwarz, Time-resolved classification of plasma crystals in a direct current discharge using an advanced graph neural network, Mach. Learn.: Sci. Technol. **6**, 045016 (2025).

[18] S. Anagnostidis, A. Thomsen, T. Kacprzak, T. Tröster, L. Biggio, A. Refregier, and T. Hofmann, Cosmology from galaxy redshift surveys with pointnet, arXiv:2211.12346.

[19] S. Young, Y.-J. Jwa, and K. Terao, Particle trajectory representation learning with masked point modeling, Mach. Learn.: Sci. Technol. **7**, 025023 (2026).

[20] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, Pointnet: Deep learning on point sets for 3d classification and segmentation, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, Piscataway, NJ, 2017), pp. 652–660.

[21] S. Chen, E. Dobriban, and J. H. Lee, A group-theoretic framework for data augmentation, J. Mach. Learn. Res. **21**, 1 (2020).

[22] M. F. Langer, S. N. Pozdnyakov, and M. Ceriotti, Probing the effects of broken symmetries in machine learning, Mach. Learn.: Sci. Technol. **5**, 04LT01 (2024).

[23] A. Mazitov, F. Bigi, M. Kellner, P. Pegolo, D. Tisi, G. Fraux, S. Pozdnyakov, P. Loche, and M. Ceriotti, PET-MAD as a lightweight universal interatomic potential for advanced materials modeling, Nat. Commun. **16**, 10653 (2025).

[24] M. W. Shen, E. Nowara, M. Maser, and K. Cho, Training dynamics of learning 3D-rotational equivariance, Trans. Mach. Learn. Res. (2025), j2C certification.

[25] B. Rhodes, S. Vandenhaute, V. Šimkus, J. Gin, J. Godwin, T. Duignan, and M. Neumann, Orb-v3: Atomistic simulation at scale, arXiv:2504.06231.

[26] T. Kreiman, Y. Bai, F. Atieh, E. Weaver, E. Qu, and A. S. Krishnapriyan, Transformers discover molecular structure without graph priors, arXiv:2510.02259.

[27] J. Riebesell, R. E. A. Goodall, P. Benner, Y. Chiang, B. Deng, G. Ceder, M. Asta, A. A. Lee, A. Jain, and K. A. Persson, A framework to evaluate machine learning crystal stability predictions, Nat. Mach. Intell. **7**, 836 (2025).

[28] F. Bigi, P. Pegolo, A. Mazitov, J. Schmidt, and M. Ceriotti, Pushing the limits of unconstrained machine-learned interatomic potentials, Mach. Learn.: Sci. Technol. **7**, 035051 (2026).

[29] F. Bigi, M. F. Langer, and M. Ceriotti, The dark side of the forces: Assessing non-conservative force models for atomistic machine learning, in *Proceedings of the 42nd International Conference on Machine Learning*, edited by A. Singh, M. Fazel, D. Hsu, S. Lacoste-Julien, F. Berkenkamp, T. Maharaj, K. Wagstaff, and J. Zhu, Proceedings of Machine Learning Research Vol. 267 (PMLR, 2025), pp. 4384–4414.

[30] M. Neumann, J. Gin, B. Rhodes, S. Bennett, Z. Li, H. Choubisa, A. Hussey, and J. Godwin, Orb: A fast, scalable neural network potential, arXiv:2410.22570.

[31] A. A. Duval, V. Schmidt, A. Hernández-García, S. Miret, F. D. Malliaros, Y. Bengio, and D. Rolnick, FAENet: Frame averaging equivariant GNN for materials modeling, in *Proceedings of the 40th International Conference on Machine Learning*, edited by A. Krause, E. Brunskill, K. Cho, B. Engelhardt, S. Sabato, and J. Scarlett, Proceedings of Machine Learning Research Vol. 202 (PMLR, 2023), pp. 9013–9033.

[32] M. Kellner, J. B. Holmes, R. Rodriguez-Madrid, F. Viscosi, Y. Zhang, L. Emsley, and M. Ceriotti, A deep learning model for chemical shieldings in molecular organic solids including anisotropy, J. Phys. Chem. Lett. **16**, 8714 (2025).

[33] H. Hua, C. Liang, D. Pan, I. King, S. Liu, K. Tsuda, and W. Lin, Scalable dielectric tensor predictions for inorganic materials using equivariant graph neural networks, arXiv:2601.04755.

[34] M. Lino, S. Fotiadis, A. A. Bharath, and C. D. Cantwell, Multi-scale rotation-equivariant graph neural networks for unsteady Eulerian fluid dynamics, Phys. Fluids **34**, 087110 (2022).

[35] N. Thomas, T. Smidt, S. Kearnes, L. Yang, L. Li, K. Kohlhoff, and P. Riley, Tensor field networks: Rotation- and translation-equivariant neural networks for 3d point clouds, arXiv:1802.08219.

[36] F. Fuchs, D. Worrall, V. Fischer, and M. Welling, Se(3)-transformers: 3d roto-translation equivariant attention networks, in *Advances in Neural Information Processing Systems*, edited by H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin (NeurIPS/Curran Associates, Inc., Red Hook, NY, 2020), Vol. 33, pp. 1970–1981.

[37] O. Puny, M. Atzmon, H. Ben-Hamu, I. Misra, A. Grover, E. J. Smith, and Y. Lipman, Frame averaging for invariant and equivariant network design, in *International Conference on Learning Representations* (ICLR, 2022).

[38] A. A. Elhag, T. K. Rusch, F. D. Giovanni, and M. Bronstein, Relaxed equivariance via multitask learning, arXiv:2410.17878.

[39] See Supplemental Material at http://link.aps.org/supplemental/10.1103/cnsd-p8nc for detailed derivations and proofs of

the equivariance metrics and character decompositions, the O(3) integration scheme and readout optimization, and additional analyses of PET and PoLAr-MAE, which includes Refs. [40,41].

[40] J. N. Goldberg, A. J. Macfarlane, E. T. Newman, F. Rohrlich, and E. C. G. Sudarshan, Spin-*s* spherical harmonics and ð, J. Math. Phys. **8**, 2155 (1967).

[41] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, in *Proceedings of the 31st International Conference on Neural Information Processing Systems* (Curran Associates, Inc., Red Hook, NY, 2017), Vol. 30, pp. 6000–6010.

[42] G. B. Folland, *A Course in Abstract Harmonic Analysis*, 2nd ed (CRC Press, New York, 2015).

[43] D. A. Varshalovich, A. N. Moskalev, and V. K. Khersonskii, *Quantum Theory of Angular Momentum* (World Scientific, Singapore, 1988).

[44] J. E. Gerken, J. Aronsson, O. Carlsson, H. Linander, F. Ohlsson, C. Petersson, and D. Persson, Geometric deep learning and equivariant neural networks, Artif. Intell. Rev. **56**, 14605 (2023).

[45] T. Cohen and M. Welling, Group equivariant convolutional networks, in *Proceedings of the 33rd International Conference on Machine Learning*, edited by M. F. Balcan and K. Q. Weinberger, Proceedings of Machine Learning Research Vol. 48 (PMLR, New York, NY, USA, 2016), pp. 2990–2999.

[46] C. Esteves, C. Allen-Blanchette, X. Zhou, and K. Daniilidis, Polar transformer networks, in *International Conference on Learning Representations* (ICLR, 2018) (accepted as poster).

[47] N. Gruver, M. Finzi, M. Goldblum, and A. G. Wilson, The Lie derivative for measuring learned equivariance, in *International Conference on Learning Representations* (ICLR, 2023).

[48] We note that Lie-derivative-based diagnostics quantify equivariance error and, through the chain rule, admit a natural layerwise decomposition of how this error is introduced across the architecture [47]. In contrast, the Haar-averaged quantities introduced here directly resolve the symmetry content of intermediate features and outputs in terms of irreducible representations of the full group. The two approaches are therefore complementary: one identifies where equivariance is broken, while the other reveals which symmetry channels are actually expressed. In the present context, this representation-resolved perspective is especially powerful, because it makes the emergence, suppression, and transfer of specific symmetry components directly visible throughout the model and along training.

[49] V. I. Lebedev and D. N. Laikov, A quadrature formula for the sphere of the 131st algebraic order of accuracy, Dokl. Math. **59**, 477 (1999).

[50] N. Shimizu and Y. Tsunoda, *SO*(3) quadratures in angular-momentum projection, Comput. Phys. Commun. **283**, 108583 (2023).

[51] L. Fadillah, L. Braks, J. Oh, M. Liu, H. Türk, D. Tisi, M. Mensi, M. Ceriotti, J. W. Choi, and A. Coskun, Molecular surface engineering of sulfide electrolytes with enhanced humidity tolerance for robust lithium metal all-solid-state batteries, Adv. Mater. **38**, e15013 (2026).

[52] H. Türk, D. Tisi, and M. Ceriotti, Reconstructions and dynamics of β-lithium thiophosphate surfaces, PRX Energy **4**, 033010 (2025).

[53] M. Eissler, T. Korjakow, S. Ganscha, O. T. Unke, K.-R. Müller, and S. Gugler, How simple can you go? An off-the-shelf transformer approach to molecular dynamics, J. Chem. Phys. **164**, 094308 (2026).

[54] C. Malosso, F. Bigi, P. Pegolo, J. W. Abbott, P. Loche, M. Rossi, M. Ceriotti, and A. Mazitov, High-quality, high-information datasets for universal atomistic machine learning, arXiv:2603.02089.

[55] A. Mazitov, S. Chorna, G. Fraux, M. Bercx, G. Pizzi, S. De, and M. Ceriotti, Massive Atomic Diversity: A compact universal dataset for atomistic machine learning, Sci. Data **12**, 1857 (2025).

[56] The potential energy has a large "compositional baseline" built as a sum of species-dependent learned atomic energies. We have removed this trivially invariant term from the energy predictions before the computation of B_α.

[57] A. Grisafi, D. M. Wilkins, G. Csányi, and M. Ceriotti, Symmetry-adapted machine learning for tensorial properties of atomistic systems, Phys. Rev. Lett. **120**, 036002 (2018).

[58] M. Domina, F. Bigi, P. Pegolo, and M. Ceriotti, Representing spherical tensors with scalar-based machine-learning models, J. Chem. Phys. **163**, 164114 (2025).

[59] S. Batzner, A. Musaelian, L. Sun, M. Geiger, J. P. Mailoa, M. Kornbluth, N. Molinari, T. E. Smidt, and B. Kozinsky, E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials, Nat. Commun. **13**, 2453 (2022).

[60] A. Power, Y. Burda, H. Edwards, I. Babuschkin, and V. Misra, Grokking: Generalization beyond overfitting on small algorithmic datasets, arXiv:2201.02177.

[61] Z. Liu, O. Kitouni, N. Nolte, E. J. Michaud, M. Tegmark, and M. Williams, Towards understanding grokking: An effective theory of representation learning, arXiv:2205.10343.

[62] L. Zhao, J. Zhang, Y. Zhang, S. Ye, G. Zhang, X. Chen, B. Jiang, and J. Jiang, Accurate machine learning prediction of protein circular dichroism spectra with embedded density descriptors, JACS Au **1**, 2377 (2021).

[63] J. Nigam, M. J. Willatt, and M. Ceriotti, Equivariant representations for molecular Hamiltonians and *N*-center atomic-scale properties, J. Chem. Phys. **156**, 014115 (2022).

[64] H. Li, Z. Wang, N. Zou, M. Ye, R. Xu, X. Gong, W. Duan, and Y. Xu, Deep-learning density functional theory Hamiltonian for efficient *ab initio* electronic-structure calculation, Nat. Comput. Sci. **2**, 367 (2022).

[65] D. Suman, J. Nigam, S. Saade, P. Pegolo, H. Türk, X. Zhang, G. K.-L. Chan, and M. Ceriotti, Exploring the design space of machine learning models for quantum chemistry with a fully differentiable framework, J. Chem. Theory Comput. **21**, 6505 (2025).

[66] P. Febrer, P. B. Jørgensen, M. Pruneda, A. García, P. Ordejón, and A. Bhowmik, Graph2Mat: Universal graph to matrix conversion for electron density prediction, Mach. Learn.: Sci. Technol. **6**, 025013 (2025).

[67] S. Ganscha, O. T. Unke, D. Ahlin, H. Maennel, S. Kashubin, and K.-R. Müller, The QCML dataset, quantum chemistry reference data from 33.5M DFT and 14.7B semi-empirical calculations, Sci. Data **12**, 406 (2025).

[68] J. W. Abbott, C. M. Acosta, A. Akkoush, A. Ambrosetti, V. Atalla, A. Bagrets, J. Behler, D. Berger, H. Bertschi, B. Bieniek, *et al.*, Roadmap on advancements of the FHI-aims software package, Electronic Structure, doi:10.1088/2516-1075/ae8067.

[69] F. Bigi, G. Fraux, N. J. Browning, and M. Ceriotti, Fast evaluation of spherical harmonics with sphericart, J. Chem. Phys. **159**, 064802 (2023).

[70] C. Adams, K. Terao, and T. Wongjirad, Pilarnet: Public dataset for particle imaging liquid argon detectors in high energy physics, arXiv:2006.01993.

[71] F. Bigi, J. W. Abbott, P. Loche, *et al.*, Metatensor and metatomic: Foundational libraries for interoperable atomistic machine learning, J. Chem. Phys. **164**, 064113 (2026).

[72] M. Domina, J. W. Abbott, P. Pegolo, F. Bigi, and M. Ceriotti, How unconstrained machine-learning models learn physical symmetries, Materials Cloud Archive **2026.144** (2026).

[73] 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, Materials Cloud, a platform for open computational science, Sci. Data **7**, 299 (2020).