WALLPAPERNET: A GENERAL BLUEPRINT FOR WALLPAPER-GROUP-EQUIVARIANT GRAPH NEURAL NETWORKS

Anonymous authors
Paper under double-blind review

ABSTRACT

Equivariant graph neural networks are typically designed for continuous Euclidean symmetry groups such as $E(3)$ or $SE(3)$, where arbitrary global translations are symmetries and can be removed using relative coordinates. In many settings, however, a graph is defined relative to an underlying periodic structure, so its symmetry is no longer the full Euclidean group. In two dimensions, these symmetries are the 17 wallpaper groups. This structure creates an opportunity: rather than explicitly representing the entire periodic system as a large graph, one can represent only the smaller graph of interest while incorporating the symmetry of the surrounding periodic structure into the model. We present a general blueprint for wallpaper-group-equivariant message passing by separating the symmetry problem into two components: point-group equivariance and lattice-relative translational symmetry. We demonstrate how it can be applied to a molecular adsorption problem on a periodic graphene substrate, providing high accuracy at a fraction of the cost of full-system message passing.

1 INTRODUCTION

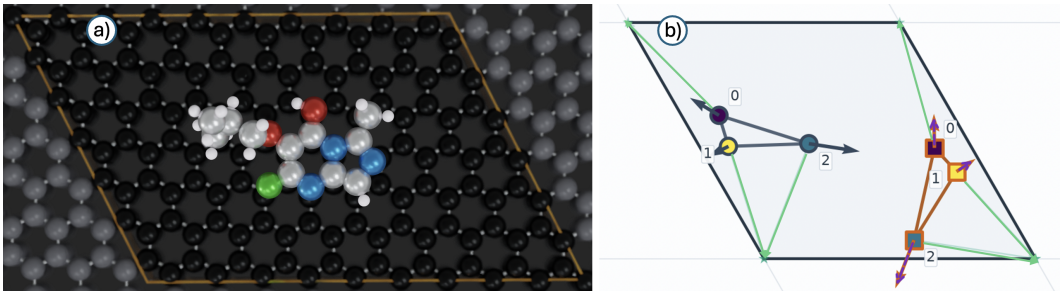


Figure 1: **Wallpaper-equivariant molecular graphs on periodic substrates.** (a) Atomic simulation cell for molecular adsorption on graphene. The black atoms form the frozen graphene support and define a periodic substrate with $p6m$ symmetry, while the colored atoms belong to the adsorbate. (b) A three-node graph shown before and after a $p6m$ wallpaper-group action. The green arrows denote Wyckoff-anchor vectors connecting graph nodes to symmetry-related reference anchors in the periodic cell. These vectors encode the adsorbate registry relative to the substrate and transform covariantly with the graph.

Message-passing neural networks provide a standard language for learning on molecular and atomistic graphs (Gilmer et al., 2017). Modern geometric graph networks exploit Euclidean symmetry to improve data efficiency and guarantee physically meaningful transformation laws. A major family of methods uses invariant scalar messages and reconstructs covariant coordinates or vectors from invariant scalar coefficients, as in $E(n)$-equivariant graph neural networks (EGNNs) (Satorras et al., 2021). A second family propagates representation-valued features, including vectors and higher-order tensors, as in Tensor Field Networks (Thomas et al., 2018), SE(3)-Transformers (Fuchs et al., 2020),

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

SEGNNs (Brandstetter et al., 2022), e3nn (Geiger & Smidt, 2022), and equivariant interatomic potentials such as NequIP (Batzner et al., 2022).

These methods are naturally suited to isolated systems for which an arbitrary rigid translation is physically irrelevant. Relative edge vectors $r_{ij} = x_i - x_j$ remove the common translation exactly. For a graph interacting with a periodic substrate, however, an arbitrary in-plane translation is generally *not* a symmetry: translating the graph relative to the lattice changes its adsorption registry or local environment. The relevant transformations are the affine isometries of a crystallographic plane group, i.e., one of the 17 wallpaper groups tabulated in the International Tables for Crystallography (Aroyo, 2016).

This distinction creates two separate symmetry problems. First, rotations and reflections must act correctly on graph features. Second, the model must retain the graph's position relative to the periodic lattice modulo the discrete translation lattice. Relative edges provide covariant geometric inputs for the first problem, but do not encode the relative position inside the cell required for the second. In crystallography, the relative position of a molecule to the cell surface is often described as *registry*, and we adopt this terminology for an arbitrary graph position with respect to the periodic lattice.

Wallpaper symmetry naturally separates into a finite point-group action and a periodic lattice registry. We handle the former with invariant or representation-valued message passing and the latter with Wyckoff-anchor features.

Figure 1 summarizes the physical and graph representations used throughout this work. Panel (a) shows a molecular adsorption system on a periodic graphene substrate, while panel (b) illustrates a three-node graph before and after a $p6m$ wallpaper-group action.

Our main contributions are:

1. We introduce a general framework for wallpaper-group-equivariant message passing on compact graphs coupled to periodic two-dimensional environments. The framework separates finite point-group equivariance from lattice registry, encoded using Wyckoff-anchor geometry and translation-suborbit member information for nonsymmorphic groups.
2. We develop `we3nn`, a finite-group equivariant neural-network library for cyclic and dihedral point groups, supporting equivariant linear and bilinear maps, regular representations, and representation-valued message passing.
3. We benchmark invariant and representation-valued realizations of the framework on synthetic scalar and vector prediction tasks spanning all 17 wallpaper groups, including a comparison with full-cell baselines.
4. We apply the framework to a wallpaper-equivariant interatomic potential for molecular adsorption on a fixed $p6m$ graphene substrate, demonstrating its use in a realistic periodic atomistic setting.

Related work. Our architectures build on equivariant geometric GNNs (Satorras et al., 2021; Thomas et al., 2018; Geiger & Smidt, 2022; Schütt et al., 2021; Batzner et al., 2022) and finite-group equivariant constructions (Cohen & Welling, 2016; Finzi et al., 2021; Cesa et al., 2022). Unlike crystal graph models (Xie & Grossman, 2018; Kaba & Ravanbakhsh, 2022; Ruff et al., 2024), we treat the substrate as a fixed periodic background and supply its registry to an adsorbate-only graph. An extended discussion is given in Appendix A.

2 Problem formulation and crystallographic symmetry

2.1 A graph coupled to a periodic background

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a geometric graph with node features h_i and positions $x_i \in \mathbb{R}^3$ with in-plane parts $x_{i,xy} \in \mathbb{R}^2$. The graph is coupled to a periodic background in the xy plane, which is not represented as an explicit large graph. Instead, the model acts on the small graph $\mathcal{G}$ together with symmetry-aware geometric features encoding its registry relative to the periodic environment.

Let G denote the wallpaper group of the periodic background. An operation $g \in G$ acts on the in-plane coordinates as

$$gx = Q_g x + \tau_g, \qquad Q_g \in O(2), \qquad \tau_g \in \mathbb{R}^2. \tag{1}$$

The linear part Q_g is a rotation or reflection. The translation τ_g may also contain the fractional shift of a glide reflection.

We write $X = \{(h_i, x_i)\}_{i=1}^N$ for the graph with N nodes embedded in three-dimensional space. A wallpaper operation g acts on the in-plane coordinates while leaving the out-of-plane coordinate unchanged,

$$x_i \longmapsto \widetilde{Q}_g x_i + \widetilde{\tau}_g, \qquad \widetilde{Q}_g = \begin{pmatrix} Q_g & 0 \\ 0 & 1 \end{pmatrix}, \qquad \widetilde{\tau}_g = \begin{pmatrix} \tau_g \\ 0 \end{pmatrix}. \tag{2}$$

We denote the transformed graph by gX.

For a symmetry-compatible graph construction, g induces a permutation π_g of the graph nodes and edges. For invariant scalar predictions such as an energy E, and covariant vector predictions such as forces $F_i \in \mathbb{R}^3$, the target transformation laws are

$$E(gX) = E(X), \tag{3}$$

and

$$F_{\pi_g(i)}(gX) = \widetilde{Q}_g F_i(X). \tag{4}$$

2.2 Translation subgroup and point group

For a planar translation vector $t \in \mathbb{R}^2$, let $T_t(x) = x + t$. The primitive Bravais lattice (Equation (72)) consists of the translations that preserve the periodic background:

$$\Lambda = \{t \in \mathbb{R}^2 : T_t \in G\}. \tag{5}$$

Thus, Λ describes the exact translational repetitions of the periodic background.

The linear-part map

$$\rho : G \to O(2), \qquad \rho(g) = Q_g, \tag{6}$$

preserves group composition and maps precisely the pure translations in G to the identity matrix. Its kernel is therefore the translation subgroup, which is consequently normal in G. The finite crystallographic point group is

$$P := \rho(G) \cong G/\Lambda. \tag{7}$$

Equivalently, wallpaper symmetry has the short exact sequence

$$1 \longrightarrow \Lambda \longrightarrow G \longrightarrow P \longrightarrow 1, \tag{8}$$

where Λ is the normal subgroup of lattice translations and the quotient map sends each wallpaper operation to its point-group class in $P \simeq G/\Lambda$. Two wallpaper operations belong to the same quotient class exactly when they differ by a lattice translation. For a symmorphic wallpaper group this extension splits,

$$G \simeq \Lambda \rtimes P, \tag{9}$$

whereas for a nonsymmorphic wallpaper group it does not. This distinction will determine whether the selected Wyckoff-anchor orbit requires an additional translation-suborbit member encoding.

We obtain generators, Wyckoff positions, and standard settings from the International Tables and the Bilbao Crystallographic Server (Aroyo et al., 2006). The table of wallpaper groups is reproduced in Appendix Table 3. The distinction between symmorphic and nonsymmorphic crystallographic groups is standard (Aroyo, 2016).

When the periodic background is represented explicitly, a joint Euclidean transformation of the graph and the cell preserves their relative geometry. For a common translation $x_i \mapsto x_i + t$, the relative edge vector satisfies $(x_i + t) - (x_j + t) = r_{ij}$. This is the usual mechanism by which Euclidean geometric GNNs remove arbitrary translations (Duval et al., 2023). In the periodic problem, however, $\{r_{ij}\}_{(i,j)\in\mathcal{E}}$ is unchanged if every graph node is shifted by a non-lattice vector, even though the graph has moved to a different position inside the periodic cell. Relative graph geometry is therefore insufficient to uniquely represent the graph position in the periodic environment.

3 WALLPAPER-GROUP-EQUIVARIANT MESSAGE PASSING

WallpaperNet combines point-group equivariance (Section 3.1) with a Wyckoff-anchor (WA) encoding of the lattice registry (Section 3.2), extended by translation-suborbit members for nonsymmorphic groups (Section 3.3). Section 3.4 shows that the combination is equivariant under the full wallpaper group.

3.1 POINT-GROUP EQUIVARIANCE

We consider two realizations of point-group equivariance. The first maps geometric inputs to invariant scalars and uses an ordinary multilayer perceptron (MLP). The second retains representation-valued features and uses equivariant linear and bilinear maps.

3.1.1 INVARIANT-MESSAGE SCALARIZATION

The planar crystallographic point groups are cyclic or dihedral groups C_n and D_n. The construction below applies for arbitrary $n \in \mathbb{N}$. For wallpaper groups, crystallographic restriction limits the occurring rotation orders to $n \in \{1, 2, 3, 4, 6\}$.

For an in-plane vector $v = (v_x, v_y)$, define

$$z(v) = v_x + \mathrm{i}v_y, \qquad \hat{z}(v) = \frac{z(v)}{\|v\|} \tag{10}$$

for $v \neq 0$. For C_n, a complete angular invariant for a single nonzero vector can be represented by

$$\psi_{C_n}(v) = (\mathrm{Re}[\hat{z}(v)^n], \mathrm{Im}[\hat{z}(v)^n], \|v\|)\,. \tag{11}$$

For D_n, let α specify the orientation of one reflection axis. A reflection-invariant encoding is

$$\psi_{D_n}(v) = \left(\mathrm{Re}\left[e^{-\mathrm{i}n\alpha}\hat{z}(v)^n\right], \|v\|\right). \tag{12}$$

These features are invariant under the corresponding cyclic or dihedral point-group action. We write ψ_P for the encoding of the point group P at hand. The derivation and the treatment of $v = 0$ are given in Appendix B.1.

For an oriented edge $j \to i$, let $s_{ij} \in \mathbb{R}^{d_s}$ denote the concatenation of the invariant message inputs: invariant node features, distance and radial features, $\psi_P(r_{ij,xy})$, $\psi_P(v_{\mathrm{WA},i})$ for the WA vector $v_{\mathrm{WA},i}$ of node i (Section 3.2), and optional joint invariants (Appendix B.1). With the message MLP $M : \mathbb{R}^{d_s} \to \mathbb{R}^{d_m}$ and the coefficient head $A : \mathbb{R}^{d_m} \to \mathbb{R}$, we obtain an invariant message $m_{ij} = M(s_{ij})$ and an invariant scalar coefficient $w_{ij} = A(m_{ij})$. As in EGNN-style models (Satorras et al., 2021), the in-plane vector output is reconstructed from the neighbors $\mathcal{N}(i)$ of node i as

$$F_{i,xy} = \sum_{j \in \mathcal{N}(i)} w_{ij} r_{ij,xy}. \tag{13}$$

Since w_{ij} is invariant and $r_{ij,xy} \mapsto Q_g r_{ij,xy}$,

$$F_{\pi_g(i),xy}(gX) = Q_g F_{i,xy}(X). \tag{14}$$

Thus an invariant-message network can produce a covariant vector output without propagating nontrivial hidden representations.

3.1.2 REPRESENTATION-VALUED MESSAGE PASSING

Alternatively, features may retain nontrivial point-group transformation laws. Each feature space carries a real orthogonal representation of the point group P. Scalars, planar vectors, regular fields, and anchor-member features carry their respective trivial, standard, regular, and permutation representations.

The common building blocks are intertwiner spaces. For real representation spaces (U, ρ_U) and (V, ρ_V) of the finite point group P, define

$$\mathrm{Hom}_P(U, V) = \{W \in \mathrm{Hom}_{\mathbb{R}}(U, V) : W\rho_U(p) = \rho_V(p)W \quad \forall p \in P\}\,. \tag{15}$$

Here we use ρ_U, ρ_V to denote feature representations of P, distinct from the linear-part map $\rho : G \to O(2)$ introduced in Section 2.2. Bilinear interactions of inputs $x \in U_1$ and $y \in U_2$ take the form $T(x, y) = C(x \otimes y)$ with $C \in \mathrm{Hom}_P(U_1 \otimes U_2, V)$. Thus linear and bilinear interactions are parameterized by bases of the appropriate intertwiner spaces. In `we3nn`, these bases are fixed and their coefficients are learned (Appendices B.1 and B.2).

For nonlinear processing, representation-valued features are mapped to regular fields, on which the group acts by permuting group-indexed coordinates. One regular field is a function $u : P \to \mathbb{R}$, with one scalar coordinate per group element, and the regular representation $\rho_{\mathrm{reg}}(p) = \lambda(p)$ acts as

$$[\lambda(p)u](k) = u(p^{-1}k), \tag{16}$$

where the element p acts, while k indexes a group coordinate. A scalar activation $\sigma : \mathbb{R} \to \mathbb{R}$, applied componentwise, commutes with this permutation action and therefore preserves equivariance. The result is then mapped to the desired output type (conventions in Appendix B.1).

3.2 Wyckoff-anchor encoding of periodic registry

Selection of one anchor orbit. We select one crystallographic anchor orbit for each wallpaper group. Let $A^0 = \{q_1, \ldots, q_m\}$ contain the representatives of this orbit modulo primitive translations. Its periodic lift is

$$\mathcal{A} = A^0 + \Lambda = \{q_\alpha + t : \alpha = 1, \ldots, m,\ t \in \Lambda\}. \tag{17}$$

The orbit is G-closed, $g\mathcal{A} = \mathcal{A}$ for all $g \in G$.

For the selected symmorphic cases, the orbit has one translation-suborbit member per primitive cell, $m = 1$, whereas the selected nonsymmorphic cases pg, pmg, pgg, and $p4g$ require $m = 2$ and are treated in Section 3.3. The complete choices are listed in Appendix B.11, Table 3.

Nearest-anchor vector. For a point $x \in \mathbb{R}^2$, define

$$a(x) \in \operatorname*{argmin}_{a \in \mathcal{A}} \|a - x\|. \tag{18}$$

Whenever the minimizer is unique (see Appendix B.7), define

$$v_{\mathrm{WA}}(x) = a(x) - x. \tag{19}$$

Proposition 1 (WA covariance)**.** *If $\mathcal{A}$ is closed under G and the nearest anchor is unique, then*

$$a(gx) = ga(x) \tag{20}$$

and

$$v_{\mathrm{WA}}(gx) = Q_g v_{\mathrm{WA}}(x). \tag{21}$$

The proof is in Appendix B.8. For $m = 1$, $\mathcal{A} = q_1 + \Lambda$ and

$$x \equiv q_1 - v_{\mathrm{WA}}(x) \pmod{\Lambda}. \tag{22}$$

Thus the WA vector retains complete position information modulo primitive translations, except at nearest-anchor ties.

3.3 Nonsymmorphic translation-suborbit encoding

The vector covariance in Equation (21) does not by itself retain which representative of a multi-member anchor orbit was selected. This matters in nonsymmorphic groups because different translation-suborbit members are permuted by wallpaper operations.

Member permutation representation. Write $\mathcal{A} = \coprod_{\alpha=1}^{m} (q_\alpha + \Lambda)$. Let $\alpha(x)$ denote the translation-suborbit member containing the nearest anchor, and for node i define

$$v_i = v_{\mathrm{WA}}(x_{i,xy}), \qquad \ell_i = \mathrm{onehot}(\alpha(x_{i,xy})) \in \mathbb{R}^m. \tag{23}$$

For $g \in G$, let $p_g := g\Lambda \in G/\Lambda \simeq P$. The induced member permutation representation $\rho_{\mathrm{mem}} : P \to O(m)$ is defined by

$$\rho_{\mathrm{mem}}(p_g)e_\alpha = e_\beta \quad \Longleftrightarrow \quad gq_\alpha \equiv q_\beta \pmod{\Lambda}, \tag{24}$$

where e_α is the α-th standard basis vector of $\mathbb{R}^m$. Its image consists of permutation matrices. Hence (Appendix B.10)

$$v_i \mapsto Q_g v_i, \qquad \ell_i \mapsto \rho_{\mathrm{mem}}(p_g)\ell_i. \tag{25}$$

Deterministic joint orbit lift. For $p \in P$, let $Q_p \in O(2)$ denote its standard planar matrix representation, so that $Q_{p_g} = Q_g$. The joint orbit lift of node i is

$$\bar{X}_{i,p} = \left[Q_p v_i,\ \rho_{\mathrm{mem}}(p)\ell_i,\ \left(\rho_{\mathrm{mem}}(p)\ell_i\right) \otimes (Q_p v_i)\right]. \tag{26}$$

The tensor-product term is implemented as member-gated WA components. It explicitly preserves the association between a translation-suborbit member and its transformed WA geometry, rather than requiring a generic MLP to recover this multiplicative relation from independent inputs. Each group block contains $C_{\mathrm{lift}} = 2 + 3m$ channels.

The raw lift transforms by the right-regular law $\bar{X}_{\pi_g(i),p}(gX) = \bar{X}_{i,pp_g}(X)$. The fixed inversion $X^{\mathrm{reg}}_{i,k} = \bar{X}_{i,k^{-1}}$ of the group index converts it to the left-regular convention used by `we3nn` (Appendix B.9), and X^{reg}_i transforms as

$$X^{\mathrm{reg}}_{\pi_g(i),k}(gX) = X^{\mathrm{reg}}_{i,p_g^{-1}k}(X), \qquad k \in P. \tag{27}$$

Thus, X^{reg}_i consists of C_{lift} copies of the regular representation.

Learned regular encoder. The deterministic registry fields are processed by a regular-to-regular equivariant network

$$f_{\mathrm{reg}} : C_{\mathrm{lift}}\rho_{\mathrm{reg}} \longrightarrow d_{\mathrm{reg}}\rho_{\mathrm{reg}}, \tag{28}$$

composed of equivariant regular-to-regular linear layers and pointwise nonlinearities. The resulting learned registry feature is

$$Z_i = f_{\mathrm{reg}}(X^{\mathrm{reg}}_i) \in \mathbb{R}^{d_{\mathrm{reg}}|P|}. \tag{29}$$

Because each learned linear map is an intertwiner and pointwise nonlinearities commute with the regular action, Z_i transforms in d_{reg} copies of the regular representation.

Invariant and equivariant uses. For the invariant-message architecture, we project the learned regular feature onto the trivial representation using the Reynolds average,

$$L_i = \frac{1}{|P|}\sum_{k\in P} Z_{i,k}. \tag{30}$$

Since the group action only permutes the regular group index, $L_{\pi_g(i)}(gX) = L_i(X)$ for all $g \in G$. The components of L_i are therefore invariant scalar features and may be passed directly to the invariant-message MLP. For the fully equivariant architecture, the group axis is retained, and Z_i is passed directly to the equivariant message-passing network as d_{reg} copies of the regular representation. Retaining the group axis preserves access to both trivial and nontrivial regular components. A small identifiability experiments in Appendix D isolate the information supplied by the WA encoding. For the nonsymmorphic groups pg, pmg, pgg, and $p4g$, they show that the translation-suborbit member representation can distinguish configurations.

3.4 Full wallpaper equivariance

Let G be the wallpaper group with primitive translation lattice Λ, and let $p_g := g\Lambda \in G/\Lambda \simeq P$. All feature representations below are representations of P and are evaluated at p_g, whereas g acts affinely on positions. Let π_g denote the corresponding node reindexing, with $\pi_g = \mathrm{id}$ when node indices are retained. Let $h^{(\ell)}_i \in V_\ell$ denote the feature of node i after message-passing layer ℓ, where V_ℓ carries the point-group representation $\rho_\ell : P \to O(V_\ell)$. For an oriented edge $j \to i$, let $z_{ij} \in V_{\mathrm{in}}$ denote the complete representation-valued edge input and $m_{ij} \in V_m$ the resulting message, with corresponding representations ρ_{in} and ρ_m.

Theorem 1 (Wallpaper equivariance of representation-valued graph fields)**.** *Assume that the graph construction and geometric encodings are compatible with the wallpaper action, including the prescribed anchor-member actions and unique nearest anchors wherever WA encoding is used. Assume that the initial node fields transform in their assigned point-group representations, and that all message functions, node updates, and readouts are equivariant for these representations. Aggregation is a sum, possibly with invariant scalar weights. Then every node field $h^{(\ell)}_i$ produced by the network satisfies*

$$h^{(\ell)}_{\pi_g(i)}(gX) = \rho_\ell(p_g)h^{(\ell)}_i(X), \qquad g \in G. \tag{31}$$

The proof follows by induction over message-passing layers. The encoded geometry transforms in its assigned representation, equivariant message maps preserve this law, and symmetry-compatible sum aggregation commutes with node reindexing. The complete proof is given in Appendix B.12. The theorem covers both message-passing constructions: the invariant-message model is recovered by assigning trivial representations to its hidden features and combining invariant scalar coefficients with covariant geometric vectors. A permutation-invariant graph readout into the trivial representation therefore yields an invariant scalar prediction, as required by Equation (3).

4 EXPERIMENTS

In this section, we study the performance on two different benchmarks. First, we create a synthetic toy wallpaper benchmark across all 17 wallpaper groups and compare WallpaperNet on it with full-cell baseline models. Second, we apply the resulting findings to construct a WallpaperNet-based machine-learning interatomic potential (MLIP) for an actual $p6m$ problem, which we justify with DFT computations.

4.1 SYNTHETIC BENCHMARK AND LEARNING TARGETS

To test whether WallpaperNet can reach the accuracy of full-cell atomistic models, we introduce a synthetic benchmark covering all 17 wallpaper groups with 1,000 configurations each: a three- to five-atom carbon fragment above a fixed, symmetry-constrained carbon support, with varying conformation, height, and lateral registry. Reference energies and forces are computed with GFN2-xTB on an explicit (3×3) support cluster (Appendix E). We predict the total energy and the atomic forces, either with a per-node equivariant force head or as the negative energy gradient.

4.1.1 MODEL CHOICES

We test seven WallpaperNet architectures of increasing complexity, from the distance-based invariant WP-Inv and its extension WP-InvExt with covariant edge and Wyckoff-anchor directions, through the regular-representation models WP-Equi, WP-VecEqui, WP-Reg, and WP-RegTP, to WP-O3, which couples three-dimensional $O(3)$ tensor features with wallpaper-group-equivariant processing. The baselines are three full-cell models without wallpaper symmetry: an invariant EGNN-style GNN and e3nn models with maximum angular degree $\ell_{\max} = 1$ and 2. All models share the optimizer and learning-rate schedule and have approximately matched parameter counts (Appendix G.9 and Table 25). Table 1 summarizes the results averaged over all 17 groups, and Appendix H reports the individual tables for every group.

4.1.2 DISCUSSION

A per-atom energy error below approximately 43 meV/atom is commonly associated with chemical accuracy (Röcken & Zavadlav, 2025; Wu et al., 2025; Ihalage & Hao, 2022). Here, the mean group-wise energy standard deviation of 17.52 meV/atom is a more informative reference, and because the systems are far from equilibrium, we compare forces through the force-component R^2 and the relative force L_2 error $\|f^{\mathrm{pred}} - f\|_2 / \|f\|_2$ over all flattened molecular force components. WP-O3 is the strongest WallpaperNet architecture and slightly outperforms both full-cell e3nn baselines for energy (4.67 versus 5.09 and $5.42\,\mathrm{meV\,atom^{-1}}$) and for all combined-training force metrics (18.6% versus approximately 20.6–20.7% relative force error). For direct vector-force training, the e3nn baselines are slightly better (11.9% versus 12.3%). For every architecture, the direct force head is more accurate than the conservative energy gradient. Overall, WallpaperNet reaches accuracy comparable to full-cell models with less message passing per system. Further discussion is given in Appendix H.

4.2 WALLPAPERNET FOR MOLECULAR ADSORPTION ON GRAPHENE

We next test whether the registry-aware construction can learn a realistic interatomic potential for molecular adsorption on graphene, which is relevant to sensing, catalysis, electronics, and energy storage (Kong et al., 2014; Hong et al., 2013; Ersan et al., 2017) and is commonly modeled with the graphene atoms held fixed (Chen et al., 2017; Londero et al., 2012; Cunha et al., 2018; Song

Table 1: Toy-model performance averaged over all 17 wallpaper groups. For each seed, metrics are first averaged over groups. Entries report the mean across three seeds with the sample standard deviation as a subscript. C/V denotes combined/vector-head training. Energy MAE is taken from combined training. Energy MAE is in meV atom^{-1} and force-component MAE in meV Å^{-1}. The mean group-wise dataset energy SD is 17.52 meV atom^{-1}. Bold marks the best mean for each metric and training mode.

	Energy MAE ↓	Rel. force L_2 (%) ↓		Force comp. MAE ↓		Force comp. R^2 ↑	
Model		C	V	C	V	C	V
WallpaperNet models							
WP-Inv	$9.42_{\pm 0.09}$	$31.5_{\pm 0.6}$	$14.1_{\pm 0.4}$	$714.4_{\pm 7.7}$	$298.6_{\pm 8.4}$	$0.900_{\pm 0.004}$	$0.980_{\pm 0.001}$
WP-InvExt	$6.31_{\pm 0.26}$	$24.7_{\pm 0.9}$	$14.2_{\pm 0.3}$	$541.7_{\pm 14.4}$	$301.8_{\pm 5.5}$	$0.937_{\pm 0.004}$	$0.979_{\pm 0.001}$
WP-Equi	$11.16_{\pm 0.78}$	$34.8_{\pm 2.7}$	$17.7_{\pm 0.8}$	$781.2_{\pm 57.3}$	$386.1_{\pm 14.5}$	$0.874_{\pm 0.024}$	$0.967_{\pm 0.003}$
WP-VecEqui	$10.93_{\pm 1.05}$	$32.9_{\pm 0.4}$	$24.3_{\pm 0.9}$	$739.7_{\pm 11.7}$	$518.3_{\pm 16.9}$	$0.890_{\pm 0.002}$	$0.938_{\pm 0.004}$
WP-Reg	$10.53_{\pm 0.28}$	$32.9_{\pm 0.7}$	$19.1_{\pm 0.8}$	$737.6_{\pm 17.6}$	$394.7_{\pm 16.6}$	$0.891_{\pm 0.005}$	$0.963_{\pm 0.003}$
WP-RegTP	$10.03_{\pm 0.32}$	$30.8_{\pm 0.5}$	$25.6_{\pm 1.1}$	$706.2_{\pm 5.5}$	$580.0_{\pm 20.6}$	$0.904_{\pm 0.003}$	$0.931_{\pm 0.007}$
WP-O3	$\mathbf{4.67}_{\pm 0.22}$	$\mathbf{18.6}_{\pm 0.1}$	$12.3_{\pm 0.5}$	$\mathbf{418.9}_{\pm 0.9}$	$264.5_{\pm 8.0}$	$\mathbf{0.965}_{\pm 0.000}$	$0.984_{\pm 0.002}$
Full-cell baselines							
Invariant GNN	$7.17_{\pm 0.62}$	$26.9_{\pm 1.4}$	$17.5_{\pm 2.3}$	$594.3_{\pm 31.2}$	$373.3_{\pm 51.1}$	$0.926_{\pm 0.008}$	$0.956_{\pm 0.019}$
e3nn ($\ell_{\max} = 1$)	$5.09_{\pm 0.25}$	$20.6_{\pm 0.7}$	$11.9_{\pm 0.7}$	$456.7_{\pm 12.5}$	$\mathbf{251.6}_{\pm 19.0}$	$0.957_{\pm 0.003}$	$\mathbf{0.986}_{\pm 0.002}$
e3nn ($\ell_{\max} = 2$)	$5.42_{\pm 0.12}$	$20.7_{\pm 0.5}$	$\mathbf{11.9}_{\pm 0.5}$	$466.2_{\pm 12.4}$	$254.2_{\pm 15.3}$	$0.957_{\pm 0.002}$	$0.985_{\pm 0.002}$

et al., 2024; Carlson et al., 2024). Graphene has wallpaper group $p6m$, with in-plane point group $C_{6v} \cong D_6$.

4.2.1 Dataset and experimental setup

The dataset contains $15{,}089$ adsorption systems, each with one organic molecule on frozen, three-layer AAA-stacked graphene, relaxed with GFN2-xTB (Bannwarth et al., 2019). The WallpaperNet graph contains only the adsorbate atoms, on average 24.93 per system versus 386.8 in the full system. The data are split 80/10/10, and we train for three objectives: (1) the full-system energy with energy-derived adsorbate forces as an auxiliary term, (2) adsorbate forces as the negative energy gradient with a small energy term, and (3) adsorbate forces from a direct equivariant head. Besides WP-Inv, WP-Equi, WP-Reg, WP-RegTP, and WP-O3, we test WP-InvEq and WP-VecInv, which transfer the design of WP-InvExt, against the three full-cell baselines scaled to approximately one million parameters. Details are given in Appendices F, G, and G.10.

4.2.2 Discussion

Table 2 shows that WallpaperNet learns an accurate interatomic potential for molecular adsorption on graphene without representing the substrate explicitly. In energy-dominated training, all WallpaperNet models remain far below the chemical-accuracy threshold of $43\,\mathrm{meV\,atom^{-1}}$, and the simplest invariant model, WP-Inv, gives the lowest energy MAE of $1.73\,\mathrm{meV\,atom^{-1}}$, below both full-cell e3nn baselines. The ordering changes for forces: with a direct vector head, WP-O3 gives the lowest error of all ten models, 19.22% relative force error and $31.3\,\mathrm{meV\,Å^{-1}}$ component MAE, compared with approximately 25% for the full-cell e3nn baselines. After rebalancing the energy and force terms, WP-O3 recovers almost the same accuracy conservatively (19.78%), whereas several other architectures do not, showing that an equivariant force output alone does not guarantee an energy surface with accurate derivatives. These results are obtained on graphs with approximately $15\times$ fewer nodes and $45\times$ fewer edges than the full-cell representation and with lower training memory. The full discussion, including the loss weighting and the memory and timing benchmarks, is given in Appendix I.1.

Table 2: Test-set performance of the seven wallpaper-symmetry models and three full-cell baselines. Energy metrics are taken from the combined experiments, while force metrics are reported for C and V, where C denotes force-weighted conservative training and V denotes direct vector-head training. Entries report the mean over three random seeds with the sample standard deviation as a subscript. Energy MAE is in meV atom^{-1} and force-component MAE in meV Å^{-1}. Lower is better for the error metrics, and higher is better for R^2. The zero-force score and the complete results are given in Table 44.

	Energy MAE ↓	Rel. force L_2 (%) ↓		Force comp. MAE ↓		Force comp. R^2 ↑	
Model		C	V	C	V	C	V
Wallpaper-symmetry models							
WP-Inv	$\mathbf{1.73}_{\pm 0.02}$	$80.79_{\pm 1.35}$	$41.35_{\pm 7.79}$	$139.9_{\pm 0.6}$	$75.0_{\pm 16.3}$	$0.35_{\pm 0.02}$	$0.82_{\pm 0.07}$
WP-InvEq	$4.75_{\pm 5.09}$	$78.68_{\pm 1.15}$	$25.67_{\pm 0.63}$	$137.1_{\pm 2.0}$	$43.5_{\pm 1.0}$	$0.38_{\pm 0.02}$	$0.93_{\pm 0.00}$
WP-Equi	$1.90_{\pm 0.17}$	$80.59_{\pm 0.94}$	$27.95_{\pm 0.80}$	$140.7_{\pm 1.4}$	$48.8_{\pm 0.9}$	$0.35_{\pm 0.02}$	$0.92_{\pm 0.00}$
WP-VecInv	$2.24_{\pm 0.06}$	$62.49_{\pm 15.95}$	$26.62_{\pm 0.93}$	$109.5_{\pm 25.4}$	$44.8_{\pm 0.8}$	$0.59_{\pm 0.21}$	$0.93_{\pm 0.01}$
WP-Reg	$2.09_{\pm 0.12}$	$30.70_{\pm 1.90}$	$27.15_{\pm 1.07}$	$54.0_{\pm 2.7}$	$45.7_{\pm 1.8}$	$0.91_{\pm 0.01}$	$0.93_{\pm 0.01}$
WP-RegTP	$1.86_{\pm 0.24}$	$30.49_{\pm 1.33}$	$26.21_{\pm 1.14}$	$52.7_{\pm 2.3}$	$44.6_{\pm 1.1}$	$0.91_{\pm 0.01}$	$0.93_{\pm 0.01}$
WP-O3	$1.80_{\pm 0.08}$	$\mathbf{19.78}_{\pm 0.70}$	$\mathbf{19.22}_{\pm 0.75}$	$\mathbf{32.9}_{\pm 0.9}$	$\mathbf{31.3}_{\pm 0.2}$	$\mathbf{0.96}_{\pm 0.00}$	$\mathbf{0.96}_{\pm 0.00}$
Full-cell baselines							
Invariant GNN	$2.58_{\pm 0.23}$	$67.51_{\pm 1.03}$	$47.22_{\pm 6.46}$	$119.9_{\pm 2.5}$	$83.6_{\pm 12.2}$	$0.54_{\pm 0.01}$	$0.77_{\pm 0.06}$
e3nn ($\ell_{\max} = 1$)	$1.92_{\pm 0.07}$	$25.89_{\pm 0.89}$	$25.31_{\pm 3.50}$	$44.7_{\pm 2.1}$	$43.9_{\pm 7.2}$	$0.93_{\pm 0.00}$	$0.94_{\pm 0.02}$
e3nn ($\ell_{\max} = 2$)	$2.04_{\pm 0.22}$	$21.43_{\pm 0.75}$	$25.55_{\pm 5.82}$	$36.2_{\pm 1.2}$	$45.0_{\pm 11.1}$	$0.95_{\pm 0.00}$	$0.93_{\pm 0.03}$

5 Conclusion

We introduced WallpaperNet, a general blueprint for wallpaper-group-equivariant graph neural networks. It separates the symmetry of a graph embedded in a periodic structure into point-group equivariance and a lattice-relative registry, which is encoded by a Wyckoff-anchor vector and, for nonsymmorphic groups, a translation-suborbit member label. We showed how to use the blueprint by building seven architectures and evaluating them on a synthetic benchmark covering all 17 wallpaper groups, where the strongest model, WP-O3, is on par with full-cell e3nn baselines. We then applied it to atoms by learning an interatomic potential for molecular adsorption on graphene. Because the substrate enters through its symmetry rather than as explicit nodes, the message-passing graph has roughly $15\times$ fewer nodes and $45\times$ fewer edges. Every WallpaperNet model therefore needs less training memory than the full-cell baselines, all of them predict energies well below chemical accuracy, and in direct-force training WP-O3 both reaches the lowest force error of all models and trains faster than every full-cell baseline.

Limitations and future work. WallpaperNet assumes a fixed periodic support with a known symmetry group and is trained for that support. Standard MLIPs, in contrast, apply to arbitrary atomic systems, and pretrained universal potentials require no system-specific training. At the same time, crystals offer many more fixed geometric settings than the 17 wallpaper groups, e.g., the 230 space groups. Extending the blueprint to these settings is a promising direction for future work.

AI use statement

In this work, we used generative AI tools to assist with research execution, software development, literature discovery, and the writing and presentation of the paper.

Generative AI was used during the development of `we3nn`, including assistance with implementing methods and writing supporting software. All AI-assisted implementation was reviewed by a human author. Where possible, the correctness of the implementation was additionally assessed by comparing its behavior and results with the established e3nn and escnn libraries. Generative AI was also used to create and modify auxiliary scripts and tools, including scripts for extracting data from experiment logs and for submitting computational jobs to a cluster.

We used AI-based chat tools for literature discovery by asking them to identify potentially relevant related work. The resulting references were subsequently checked, and the relevant papers were read and assessed by a human author rather than relying on AI-generated summaries or descriptions.

Generative AI was also used to provide feedback on experimental design and execution. Suggestions produced by the AI were reviewed by the main author before being incorporated into the research process. In addition, we used AI tools to inspect equations and notation for potential inconsistencies or collisions in notation. We also used AI to convert equations from LaTeX to Python code, which was then reviewed and tested by a human author. We also used AI to describe the code in a human-readable format, which was then reviewed and edited by a human author.

For preparation of the manuscript, generative AI was used for text editing, formatting, and stylistic feedback. Generative AI was used to draft passages from author-provided outlines, to correct grammatical errors and typographical mistakes, and to assess whether passages were understandable to readers from different research backgrounds. We also used AI tools to obtain suggestions about which material could potentially be moved to the appendix.

AI was used to generate synthetic data sets, in the sense that it was used to write code, that after execution, generated synthetic data sets. But we never used AI output as synthetic data directly.

All AI-assisted research, code, literature suggestions, and manuscript edits were reviewed by the main author. The authors take responsibility for the final content of this work, including all text, claims, code, results, and other artifacts produced with the assistance of generative AI.

AI was not used to develop theoretical models or conceptual frameworks, formulate mathematical claims, or provide critical ingredients for proving mathematical claims. AI was not used for assistance with translation, cleaning or reformat datasets and it was not used for independent data analysis. We did not use generative AI to generate any of the research ideas.

Reproducibility statement

All finite-group equivariant layers are built with our `we3nn` library, which is available in anonymized form at `https://anonymous.4open.science/r/we3nn_private-693D/` and described in Appendix B.2. The WallpaperNet architectures and the full-cell baselines are described in Appendix G.4, and their training objectives and optimization settings for the synthetic and the molecule–graphene experiments in Appendices G.8–G.10. The construction of the synthetic benchmark and of the molecule–graphene dataset is described in Appendices E and F, the Wyckoff-anchor identifiability experiment in Appendix D, and the per-group and per-seed results in Appendices H and I. In case of acceptance, we will publicly release `we3nn`, our datasets, and our trained models together with their checkpoints. To make it easy to connect the reported findings with the released code, the manuscript retains the original names of layers, implementations, and models.

References

M. I. Aroyo, J. M. Perez-Mato, C. Capillas, E. Kroumova, S. Ivantchev, G. Madariaga, A. Kirov, and H. Wondratschek. Bilbao crystallographic server I: Databases and crystallographic computing programs. *Zeitschrift für Kristallographie*, 221(1):15–27, 2006.

Mois I. Aroyo (ed.). *International Tables for Crystallography, Volume a: Space-group Symmetry*. International Union of Crystallography, second online edition edition, 2016. doi: 10.1107/97809553602060000114.

Christoph Bannwarth, Sebastian Ehlert, and Stefan Grimme. GFN2-xTB - An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions. *Journal of Chemical Theory and Computation*, 15(3):1652–1671, March 2019. ISSN 1549-9626. doi: 10.1021/acs.jctc.8b01176.

Simon Batzner, Albert Musaelian, Lixin Sun, Mario Geiger, Jonathan P. Mailoa, Mordechai Kornbluth, Nicola Molinari, Tess E. Smidt, and Boris Kozinsky. E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nature Communications*, 13(1):2453, May 2022. ISSN 2041-1723. doi: 10.1038/s41467-022-29939-5.

Johannes Brandstetter, Rob Hesselink, Elise van der Pol, Erik J. Bekkers, and Max Welling. Geometric and physical quantities improve E(3) equivariant message passing. In *International Conference on Learning Representations*, 2022.

Eike Caldeweyher, Sebastian Ehlert, Andreas Hansen, Hagen Neugebauer, Sebastian Spicher, Christoph Bannwarth, and Stefan Grimme. A generally applicable atomic-charge dependent London dispersion correction. *The Journal of Chemical Physics*, 150(15):154122, April 2019. ISSN 0021-9606, 1089-7690. doi: 10.1063/1.5090222.

Shane R. Carlson, Otto Schullian, Maximilian R. Becker, and Roland R. Netz. Modeling Water Interactions with Graphene and Graphite via Force Fields Consistent with Experimental Contact Angles. *The Journal of Physical Chemistry Letters*, 15(24):6325–6333, June 2024. ISSN 1948-7185, 1948-7185. doi: 10.1021/acs.jpclett.4c01143.

Gabriele Cesa, Leon Lang, and Maurice Weiler. A program to build E(N)-equivariant steerable CNNs. In *International Conference on Learning Representations (ICLR)*, 2022.

Tsan-Yao Chen, Yanhui Zhang, Liang-Ching Hsu, Alice Hu, Yu Zhuang, Chia-Ming Fan, Cheng-Yu Wang, Tsui-Yun Chung, Cheng-Si Tsao, and Haw-Yeu Chuang. Crystal shape controlled H2 storage rate in nanoporous carbon composite with ultra-fine Pt nanoparticle. *Scientific Reports*, 7 (1):42438, February 2017. ISSN 2045-2322. doi: 10.1038/srep42438.

Taco Cohen and Max Welling. Group equivariant convolutional networks. In *International Conference on Machine Learning*, pp. 2990–2999. PMLR, 2016.

Eunice Cunha, Maria Fernanda Proença, Maria Goreti Pereira, Maria José Fernandes, Robert J. Young, Karol Strutyński, Manuel Melle-Franco, Mariam Gonzalez-Debs, Paulo E. Lopes, and Maria da Conceição Paiva. Water Dispersible Few-Layer Graphene Stabilized by a Novel Pyrene Derivative at Micromolar Concentration. *Nanomaterials*, 8(9):675, August 2018. ISSN 2079-4991. doi: 10.3390/nano8090675.

Alexandre Duval, Simon V. Mathis, Chaitanya K. Joshi, Victor Schmidt, Santiago Miret, Fragkiskos D. Malliaros, Taco Cohen, Pietro Liò, Yoshua Bengio, and Michael Bronstein. A Hitchhiker's Guide to Geometric GNNs for 3D Atomic Systems. https://arxiv.org/abs/2312.07511, 2023.

Gamze Ersan, Onur G. Apul, Francois Perreault, and Tanju Karanfil. Adsorption of organic contaminants by graphene nanosheets: A review. *Water Research*, 126:385–398, December 2017. ISSN 00431354. doi: 10.1016/j.watres.2017.08.010.

Evgraf S. Fedorov. Symmetry in the plane. *Proceedings of the Imperial St. Petersburg Mineralogical Society*, 28:345–390, 1891.

Rostislav Fedorov and Ganna Gryn'ova. Unlocking the Potential: Predicting Redox Behavior of Organic Molecules, from Linear Fits to Neural Networks. *Journal of Chemical Theory and Computation*, 19(15):4796–4814, August 2023. ISSN 1549-9618, 1549-9626. doi: 10.1021/acs.jctc.3c00355.

Rostislav Fedorov, Anastasiia Nihei, and Ganna Gryn'ova. Multi-Solvent Graph Neural Network for Reduction Potential Prediction Across the Chemical Space. *Journal of Chemical Information and Modeling*, 66(2):847–854, January 2026. doi: 10.1021/acs.jcim.5c01450.

Marc Finzi, Max Welling, and Andrew Gordon Wilson. A Practical Method for Constructing Equivariant Multilayer Perceptrons for Arbitrary Matrix Groups, April 2021.

Fabian B. Fuchs, Daniel E. Worrall, Volker Fischer, and Max Welling. SE(3)-transformers: 3D roto-translation equivariant attention networks. In *Advances in Neural Information Processing Systems*, volume 33, 2020.

Mario Geiger and Tess Smidt. E3nn: Euclidean neural networks. *arXiv preprint arXiv:2207.09453*, 2022.

Justin Gilmer, Samuel S. Schoenholz, Patrick F. Riley, Oriol Vinyals, and George E. Dahl. Neural message passing for quantum chemistry. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pp. 1263–1272. PMLR, 2017.

Ask Hjorth Larsen, Jens Jørgen Mortensen, Jakob Blomqvist, Ivano E Castelli, Rune Christensen, Marcin Dułak, Jesper Friis, Michael N Groves, Bjørk Hammer, Cory Hargus, Eric D Hermes, Paul C Jennings, Peter Bjerre Jensen, James Kermode, John R Kitchin, Esben Leonhard Kolsbjerg, Joseph Kubal, Kristen Kaasbjerg, Steen Lysgaard, Jón Bergmann Maronsson, Tristan Maxson, Thomas Olsen, Lars Pastewka, Andrew Peterson, Carsten Rostgaard, Jakob Schiøtz, Ole Schütt, Mikkel Strange, Kristian S Thygesen, Tejs Vegge, Lasse Vilhelmsen, Michael Walter, Zhenhua Zeng, and Karsten W Jacobsen. The atomic simulation environment—a Python library for working with atoms. *Journal of Physics: Condensed Matter*, 29(27):273002, July 2017. ISSN 0953-8984, 1361-648X. doi: 10.1088/1361-648X/aa680e.

Guo Hong, Qi-Hui Wu, Jianguo Ren, Chundong Wang, Wenjun Zhang, and Shuit-Tong Lee. Recent progress in organic molecule/graphene interfaces. *Nano Today*, 8(4):388–402, August 2013. ISSN 17480132. doi: 10.1016/j.nantod.2013.07.003.

B. Hourahine, B. Aradi, V. Blum, F. Bonafé, A. Buccheri, C. Camacho, C. Cevallos, M. Y. Deshaye, T. Dumitrică, A. Dominguez, S. Ehlert, M. Elstner, T. van der Heide, J. Hermann, S. Irle, J. J. Kranz, C. Köhler, T. Kowalczyk, T. Kubař, I. S. Lee, V. Lutsker, R. J. Maurer, S. K. Min, I. Mitchell, C. Negre, T. A. Niehaus, A. M. N. Niklasson, A. J. Page, A. Pecchia, G. Penazzi, M. P. Persson, J. Řezáč, C. G. Sánchez, M. Sternberg, M. Stöhr, F. Stuckenberg, A. Tkatchenko, V. W.-z. Yu, and T. Frauenheim. DFTB+, a software package for efficient approximate density functional theory based atomistic simulations. *The Journal of Chemical Physics*, 152(12):124101, March 2020. ISSN 0021-9606, 1089-7690. doi: 10.1063/1.5143190.

Achintha Ihalage and Yang Hao. Formula Graph Self-Attention Network for Representation-Domain Independent Materials Discovery. *Advanced Science*, 9(18):2200164, June 2022. ISSN 2198-3844, 2198-3844. doi: 10.1002/advs.202200164.

Oumar Kaba and Siamak Ravanbakhsh. Equivariant networks for crystal structures. In *Advances in Neural Information Processing Systems*, volume 35, 2022.

Lingmei Kong, Axel Enders, Talat S Rahman, and Peter A Dowben. Molecular adsorption on graphene. *Journal of Physics: Condensed Matter*, 26(44):443001, November 2014. ISSN 0953-8984, 1361-648X. doi: 10.1088/0953-8984/26/44/443001.

Greg Landrum, Paolo Tosco, Brian Kelley, Ricardo Rodriguez, David Cosgrove, Riccardo Vianello, sriniker, Peter Gedeck, Gareth Jones, Eisuke Kawashima, NadineSchneider, Dan Nealschneider, Andrew Dalke, Matt Swain, Brian Cole, Samo Turk, Aleksandr Savelev, tadhurst- cdd, Alain Vaucher, Maciej Wójcikowski, Ichiru Take, Rachel Walker, Vincent F. Scalfani, Hussein Faara, Kazuya Ujihara, Daniel Probst, Juuso Lehtivarjo, guillaume godin, Axel Pahl, and Niels Maeder. Rdkit/rdkit: 2025_03_2 (Q1 2025) Release. Zenodo, April 2025.

Daniel B. Litvin and Thomas R. Wike. *Character Tables and Compatibility Relations of The Eighty Layer Groups and Seventeen Plane Groups*. Springer US, Boston, MA, 1991. ISBN 978-1-4612-7744-6 978-1-4613-0495-1. doi: 10.1007/978-1-4613-0495-1.

Chong Liu, Yuxin Li, Motoshi Takao, Takashi Toyao, Zen Maeno, Takashi Kamachi, Yoyo Hinuma, Ichigaku Takigawa, and Ken-ichi Shimizu. Frontier Molecular Orbital Based Analysis of Solid–Adsorbate Interactions over Group 13 Metal Oxide Surfaces. *The Journal of Physical Chemistry C*, 124(28):15355–15365, July 2020. ISSN 1932-7447, 1932-7455. doi: 10.1021/acs.jpcc.0c04480.

Elisa Londero, Emma K Karlson, Marcus Landahl, Dimitri Ostrovskii, Jonatan D Rydberg, and Elsebeth Schröder. Desorption of n-alkanes from graphene: A van der Waals density functional study. *Journal of Physics: Condensed Matter*, 24(42):424212, October 2012. ISSN 0953-8984, 1361-648X. doi: 10.1088/0953-8984/24/42/424212.

Ilya Loshchilov and Frank Hutter. Decoupled Weight Decay Regularization, January 2019.

Johann Lüder, Mun Ho Cheow, and Sergei Manzhos. Understanding doping strategies in the design of organic electrode materials for Li and Na ion batteries: An electronic structure perspective. *Physical Chemistry Chemical Physics*, 19(20):13195–13209, 2017. ISSN 1463-9076, 1463-9084. doi: 10.1039/C7CP01554B.

Diego Mateo, Jose Luis Cerrillo, Sara Durini, and Jorge Gascon. Fundamentals and applications of photo-thermal catalysis. *Chemical Society Reviews*, 50(3):2173–2210, 2021. ISSN 0306-0012, 1460-4744. doi: 10.1039/D0CS00357C.

Andrew J. Medford, Aleksandra Vojvodic, Jens S. Hummelshøj, Johannes Voss, Frank Abild-Pedersen, Felix Studt, Thomas Bligaard, Anders Nilsson, and Jens K. Nørskov. From the Sabatier principle to a predictive theory of transition-metal heterogeneous catalysis. *Journal of Catalysis*, 328:36–42, August 2015. ISSN 00219517. doi: 10.1016/j.jcat.2014.12.033.

Jens Jørgen Mortensen, Ask Hjorth Larsen, Mikael Kuisma, Aleksei V. Ivanov, Alireza Taghizadeh, Andrew Peterson, Anubhab Haldar, Asmus Ougaard Dohn, Christian Schäfer, Elvar Örn Jónsson, Eric D. Hermes, Fredrik Andreas Nilsson, Georg Kastlunger, Gianluca Levi, Hannes Jónsson, Hannu Häkkinen, Jakub Fojt, Jiban Kangsabanik, Joachim Sødequist, Jouko Lehtomäki, Julian Heske, Jussi Enkovaara, Kirsten Trøstrup Winther, Marcin Dulak, Marko M. Melander, Martin Ovesen, Martti Louhivuori, Michael Walter, Morten Gjerding, Olga Lopez-Acevedo, Paul Erhart, Robert Warmbier, Rolf Würdemann, Sami Kaappa, Simone Latini, Tara Maria Boland, Thomas Bligaard, Thorbjørn Skovhus, Toma Susi, Tristan Maxson, Tuomas Rossi, Xi Chen, Yorick Leonard A. Schmerwitz, Jakob Schiøtz, Thomas Olsen, Karsten Wedel Jacobsen, and Kristian Sommer Thygesen. GPAW: An open Python package for electronic structure calculations. *The Journal of Chemical Physics*, 160(9):092503, March 2024. ISSN 0021-9606, 1089-7690. doi: 10.1063/5.0182685.

John P. Perdew, Kieron Burke, and Matthias Ernzerhof. Generalized Gradient Approximation Made Simple. *Physical Review Letters*, 77(18):3865–3868, October 1996. ISSN 0031-9007, 1079-7114. doi: 10.1103/physrevlett.77.3865.

G. Pólya. XII. Über die Analogie der Kristallsymmetrie in der Ebene. *Zeitschrift für Kristallographie - Crystalline Materials*, 60(1-6):278–282, November 1924. ISSN 2196-7105, 2194-4946. doi: 10.1524/zkri.1924.60.1.278.

Sebastien Röcken and Julija Zavadlav. Enhancing Machine Learning Potentials through Transfer Learning across Chemical Elements. *Journal of Chemical Information and Modeling*, 65(14): 7406–7414, July 2025. ISSN 1549-9596, 1549-960X. doi: 10.1021/acs.jcim.5c00293.

Robin Ruff, Patrick Reiser, Jan Stühmer, and Pascal Friederich. Connectivity optimized nested line graph networks for crystal structures. *Digital Discovery*, 3(3):594–601, 2024. ISSN 2635-098X. doi: 10.1039/D4DD00018H.

Víctor Garcia Satorras, Emiel Hoogeboom, and Max Welling. E(n) equivariant graph neural networks. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 9323–9332. PMLR, 2021.

Kristof T. Schütt, Oliver T. Unke, and Michael Gastegger. Equivariant message passing for the prediction of tensorial properties and molecular spectra, 2021.

Leonhard L. Scott and Jean-Pierre Serre. *Linear Representations of Finite Groups*. Number 42 in Graduate Texts in Mathematics. Springer London, Limited, New York Heidelberg Berlin, 2012. ISBN 978-1-4684-9458-7.

Young-Joon Song, Charlotte Gallenkamp, Genís Lleopart, Vera Krewald, and Roser Valentí. Influence of graphene on the electronic and magnetic properties of an iron(<scp>iii</scp>) porphyrin chloride complex. *Physical Chemistry Chemical Physics*, 26(41):26370–26376, 2024. ISSN 1463-9076, 1463-9084. doi: 10.1039/D4CP01551G.

John C. Thomas, Anirudh Raju Natarajan, and Anton Van Der Ven. Comparing crystal structures with symmetry and geometry. *npj Computational Materials*, 7(1):164, October 2021. ISSN 2057-3960. doi: 10.1038/s41524-021-00627-0.

Nathaniel Thomas, Tess Smidt, Steven Kearnes, Lusann Yang, Li Li, Kai Kohlhoff, and Patrick Riley. Tensor field networks: Rotation- and translation-equivariant neural networks for 3D point clouds. *arXiv preprint arXiv:1802.08219*, 2018.

Zhihong Wu, Lei Zhou, Pengfei Hou, Yuyan Liu, Ruixing Wang, Taoli Guo, and Jin-Cheng Liu. A Machine Learning Interatomic Potential Data Set and Model for Catalysis with Local Fine-Tuning to Chemical Accuracy. *JACS Au*, 5(12):6151–6161, December 2025. ISSN 2691-3704, 2691-3704. doi: 10.1021/jacsau.5c01112.

Tian Xie and Jeffrey C. Grossman. Crystal Graph Convolutional Neural Networks for an Accurate and Interpretable Prediction of Material Properties. *Physical Review Letters*, 120(14):145301, April 2018. ISSN 0031-9007, 1079-7114. doi: 10.1103/PhysRevLett.120.145301.

Yang-Xin Yu. Can all nitrogen-doped defects improve the performance of graphene anode materials for lithium-ion batteries? *Physical Chemistry Chemical Physics*, 15(39):16819, 2013. ISSN 1463-9076, 1463-9084. doi: 10.1039/c3cp51689j.

Yang-Xin Yu. Binding Energy and Work Function of Organic Electrode Materials Phenanthraquinone, Pyromellitic Dianhydride and Their Derivatives Adsorbed on Graphene. *ACS Applied Materials & Interfaces*, 6(18):16267–16275, September 2014a. ISSN 1944-8244, 1944-8252. doi: 10.1021/am504452a.

Yang-Xin Yu. A dispersion-corrected DFT study on adsorption of battery active materials anthraquinone and its derivatives on monolayer graphene and h-BN. *J. Mater. Chem. A*, 2(23): 8910–8917, 2014b. ISSN 2050-7488, 2050-7496. doi: 10.1039/C4TA00103F.

APPENDIX CONTENTS

A Extended related work **16**

B Theory and implementation details **16**

B.1 Point-group-equivariant message passing: derivations and conventions 16
B.2 `we3nn`: finite-group equivariant linear and bilinear maps 17
B.3 Equivariant tensor products 22
B.4 Tensor products involving regular representations 24
B.5 Internal and externally supplied tensor-product weights 26
B.6 Bravais lattice 26
B.7 Voronoi boundaries 26
B.8 WA covariance proof 26
B.9 Translation-suborbit regular-representation implementation details 26
B.10 Action on translation-suborbit members 27
B.11 Selected anchor orbits 28
B.12 Proof of full wallpaper equivariance 28

C Wallpaper-group action visualizations **29**

D Wyckoff-anchor identifiability across all wallpaper groups **53**

E Synthetic wallpaper-group benchmark **57**

F Dataset construction and simulation details **61**

F.1 Motivation for the electronic-structure level 61
F.2 Adsorbate preparation and simulation-cell construction 62
F.3 Validation against periodic DFT 62
F.4 Dataset statistics and force distribution 62

G Architectures for real graphene–adsorbate and synthetic experiments **63**

G.1 Common notation and graph construction 63
G.2 Training scenarios 65
G.3 Registered and loss-connected parameter counts 66
G.4 Architecture descriptions 68
G.5 Symmetry summary 76
G.6 Experimental comparability and interpretation 77
G.7 Training metrics 78
G.8 Toy-model training objectives 78
G.9 Toy-model optimization and training conditions 79
G.10 Real-system optimization and training conditions 79

H Toy benchmark results for all wallpaper groups **80**

I Molecule–graphene system results **92**

I.1 Discussion of the real-system results 92
I.2 Per-seed results and force-weighted conservative objective 94
I.3 Memory footprint of molecule–graphene systems 95
I.4 One-batch maximum utilization speed tests 96

A Extended related work

Equivariant geometric learning. Satorras et al. (2021) construct EGNN vector updates from invariant scalar messages and relative displacement vectors. Thomas et al. (2018) instead introduce representation-valued features and spherical-harmonic filters coupled through equivariant tensor products. Geiger & Smidt (2022) provide the `e3nn` framework for constructing such $E(3)$-equivariant operations. Schütt et al. (2021) propagate scalar and vector features through equivariant message passing, while Batzner et al. (2022) apply higher-order $E(3)$-equivariant interactions to interatomic potentials for molecular and materials systems. These approaches provide building blocks for our scalar- and representation-valued architectures. Their Euclidean equivariance should, however, be distinguished from the registry problem considered here: transforming an adsorbate and substrate jointly preserves their relative configuration, whereas translating the adsorbate relative to a fixed periodic substrate changes its registry.

Finite-group equivariant constructions. Cohen & Welling (2016) introduce group-equivariant convolutional networks that represent transformations through group-indexed features and shared filters. Finzi et al. (2021) construct equivariant layers by solving representation constraints, while Cesa et al. (2022) provide a general Wigner–Eckart construction of steerable kernel spaces. Our `we3nn` implementation builds on these representation-theoretic principles to parameterize finite planar-group linear and bilinear maps. Regular fields support componentwise nonlinearities, while intertwiner spaces specify the allowed interactions. The general mathematical equivariance criterion is therefore established. Our focus is on its extension to graphs and its use with cell-registry encodings.

Crystalline and surface models. Xie & Grossman (2018) represent crystalline environments as atomic graphs for property prediction in CGCNN. Kaba & Ravanbakhsh (2022) explicitly incorporate crystalline symmetry into generalized message-passing operations, and Ruff et al. (2024) exploit asymmetric-unit graph representations to reduce the graph size for crystalline materials. These are close precedents for crystallographic inductive biases and symmetry-based graph reduction. In contrast to explicit substrate-atom representations, our formulation treats the substrate as a fixed periodic background and supplies its registry to an adsorbate-only graph through Wyckoff-anchor geometry and equivariant member labels. The emphasis is on retaining this registry, including nonsymmorphic member permutations, without constructing substrate nodes in the message-passing graph.

Wallpaper-group symmetry. The 17 wallpaper, or crystallographic plane, groups were established in foundational classifications by Fedorov (1891) and Pólya (1924). These groups combine lattice translations with compatible rotations, reflections, and glide reflections and form the two-dimensional analogue of the crystallographic space groups. Standard crystallographic tabulations specify their symmetry operations, general and special positions, and Wyckoff positions (Aroyo et al., 2006; Aroyo, 2016). Representation-theoretic data for the plane groups have likewise been tabulated systematically (Litvin & Wike, 1991). Our construction uses this structure as an inductive bias: Wyckoff-related geometry identifies the registry of a graph relative to the periodic background, while finite-group representations determine how the corresponding features transform. Additionally, our blueprint is applicable to GNNs on graphs with wallpaper symmetry, which has applications beyond chemistry and materials science.

B Theory and implementation details

B.1 Point-group-equivariant message passing: derivations and conventions

Cyclic and dihedral scalarization. For ψ_{C_n} in Equation (11): under a generator rotation by $2\pi/n$, the phase acquires $e^{2\pi \mathrm{i}/n}$ and its nth power is unchanged. In ψ_{D_n}, Equation (12), the imaginary part is omitted because it changes sign under reflection. At $v = 0$, continuous polynomial versions using $z(v)^n$ may be used instead of normalized angular features. Besides $\psi_P(r_{ij,xy})$ and $\psi_P(v_{\mathrm{WA},i})$, the invariant message input may contain joint invariants such as dot and (if reflection is absent) out-of-plane cross products. The radial bases, the joint invariants used by each architecture, and the input dimensions are given in Appendix G.

Scalar-weighted vector reconstruction. Although each edge term $w_{ij}r_{ij,xy}$ is parallel to its own edge, the aggregated vector $F_{i,xy}$ is generally not parallel to any individual edge. Two non-collinear in-plane neighbors already span $\mathbb{R}^2$.

Equivariant linear and bilinear maps. Irreducible representations are representations with no nonzero proper invariant subspace. Direct sums of irreducible representations describe general finite-group feature types. Bilinear interactions are described by the same construction as Equation (15) with a tensor-product input. Writing $\rho_{12}(p) = \rho_1(p) \otimes \rho_2(p)$, their linearizations belong to

$$\begin{aligned}\mathrm{Hom}_P(U_1 \otimes U_2, V) = \big\{C \in \mathrm{Hom}_{\mathbb{R}}(U_1 \otimes U_2, V) : \\ C\rho_{12}(p) = \rho_V(p)C \quad \forall p \in P\big\}.\end{aligned} \tag{32}$$

Specifically, $T(x, y) = C(x \otimes y)$ is bilinear and equivariant precisely when C belongs to this space.

We parameterize affine linear layers and bilinear interactions as

$$\begin{aligned}L_\theta(x) = W_\theta x + b, \qquad W_\theta = \sum_a \theta_a B_a, \\ T_\eta(x, y) = C_\eta(x \otimes y), \quad C_\eta = \sum_c \eta_c C_c,\end{aligned} \tag{33}$$

where $\{B_a\}$ and $\{C_c\}$ are fixed bases of the corresponding intertwiner spaces for the selected representation blocks. Biases satisfy $b \in V^P = \{v \in V : \rho_V(p)v = v\ \forall p \in P\}$. Multiplicity-channel connectivity is specified separately. Tensor-product coefficients may also be generated from invariant contextual features, for example, a radial MLP may generate tensor-product coefficients from $\|r_{ij}\|$, while preserving equivariance of the complete operation.

Construction of the symmetry bases. Both spaces are instances of $\mathrm{Hom}_P(D, V)$, with $D = U$ for linear maps and $D = U_1 \otimes U_2$ for bilinear interactions. Their bases can be obtained by Reynolds averaging followed by numerical basis extraction (Thomas et al., 2021), by solving the linear intertwiner constraints as a null-space problem (Finzi et al., 2021), or through an irreducible-decomposition construction combining Clebsch–Gordan coefficients and the required real endomorphism bases (Cesa et al., 2022). These are alternative constructions of the same target intertwiner space, not different equivariance requirements. For blocks involving regular representations, `we3nn` exploits analytic formulas for these spaces, as detailed in Appendix B.2.

Regular fields. For $H \in \mathbb{N}$ copies of the regular field of Equation (16), let $\mathcal{H} = \mathbb{R}[P]^{\oplus H}$, with

$$\rho_{\mathcal{H}}(p) = I_H \otimes \lambda(p), \tag{34}$$

where I_H is the identity on the copy index. Flattening uses copy-major order, with the group coordinate varying fastest. Invariant gates have one value per regular copy and are shared over its group coordinates.

B.2 WE3NN: FINITE-GROUP EQUIVARIANT LINEAR AND BILINEAR MAPS

Notation. Appendices B.2–B.5 describe `we3nn` as a standalone library. Here G denotes a finite cyclic or dihedral group, the point group P of the main text, with elements g, h, q, and ρ denotes a generic feature representation, written ρ_U, ρ_V in the main text. These symbols do not refer to the wallpaper group or the linear-part map. The indices λ, μ, ν label irreducible representations V_λ with matrices ρ_λ. They are unrelated to the left-regular action λ of the main text, and the regular representation is written ρ_{reg}. Dimensions of intertwiner spaces are denoted K, and p indexes their basis elements.

To support representation-valued GNNs for finite cyclic and dihedral groups, we developed `we3nn`. The library constructs exact equivariant linear and bilinear maps between finite-dimensional representations of

$$G \in \{C_n, D_n\}.$$

Although the library also exposes a generic finite-group interface, the discussion below is restricted to its built-in C_n and D_n representations. The central design principle is to parameterize only the vector space of maps that commute with the prescribed group actions.

Motivation. `we3nn` is motivated by the gap between `escnn` and `e3nn`: it brings the flexible point-group equivariance concepts developed in `escnn` to graph and message-passing neural networks, compatible with `e3nn`. Many useful abstractions and features already exist in `escnn` and can be adapted, and `we3nn` deliberately borrows much of its terminology, notation, and API style. Graph networks, however, also require additional primitives, most notably bilinear tensor products. While the interface is inspired by `escnn`, the core implementation is different: rather than relying on explicit Clebsch–Gordan decompositions and Wigner–Eckart constructions, `we3nn` builds equivariant maps directly from finite-group representation constraints and intertwiners.

Representations and feature types. Let V be a finite-dimensional real vector space carrying a representation

$$\rho : G \to \mathrm{GL}(V),$$

Where $\mathrm{GL}(V)$ is the group of invertible linear maps on V. A feature $x \in V$ transforms according to

$$x \longmapsto \rho(g)x.$$

Feature types may contain several representations,

$$V = \bigoplus_{\lambda \in \mathrm{Irr}(G)} m_\lambda V_\lambda,$$

where $\mathrm{Irr}(G)$ is the set of irreducible representations, and m_λ denotes the number of copies, or multiplicity, of representation V_λ. The learnable operators are constructed separately for each compatible pair or triple of representation types. Equal representation occurrences share the same symmetry basis, while independent trainable coefficients mix their multiplicity channels.

B.2.1 Equivariant linear maps

Consider a linear map

$$y = Wx, \qquad W : V_{\mathrm{in}} \to V_{\mathrm{out}},$$

where the input and output transform according to ρ_{in} and ρ_{out}. Equivariance requires

$$\rho_{\mathrm{out}}(g)W = W\rho_{\mathrm{in}}(g) \qquad \forall g \in G. \tag{35}$$

The admissible matrices therefore form the intertwiner space

$$\mathrm{Hom}_G(V_{\mathrm{in}}, V_{\mathrm{out}}) = \{W : \rho_{\mathrm{out}}(g)W = W\rho_{\mathrm{in}}(g) \quad \forall g \in G\}. \tag{36}$$

Let

$$\{B_p\}_{p=1}^{K}$$

be a basis of this space, where

$$K = \dim \mathrm{Hom}_G(V_{\mathrm{in}}, V_{\mathrm{out}}),$$

then `WE_Linear` parameterizes the learned matrix as

$$W = \sum_{p=1}^{K} \theta_p B_p. \tag{37}$$

The basis matrices B_p are fixed by symmetry and only the reduced coefficients θ_p are trainable. Consequently, Equation (35) holds for every value of the learned parameters.

When multiple copies of the input V_λ and output representations V_μ are present, the coefficients become

$$\theta^{\mu\lambda}_{rcp},$$

where c indexes input multiplicity and r output multiplicity:

$$W^{\mu\lambda}_{rc} = \sum_{p=1}^{K_{\mu\lambda}} \theta^{\mu\lambda}_{rcp} B^{\mu\lambda}_{p}. \tag{38}$$

Generic representation-to-representation linear map. When neither representation is regular, the default structural backend constructs the intertwiner basis through finite-group averaging. Starting from the elementary matrix basis $E_{oi} \in \mathbb{R}^{d_{\text{out}} \times d_{\text{in}}}$, whose (o, i) entry is 1 and all other entries are 0, each matrix is projected with the Reynolds operator

$$\Pi_{\text{lin}}(A) = \frac{1}{|G|} \sum_{g \in G} \rho_{\text{out}}(g) A \rho_{\text{in}}(g)^{\mathsf{T}}. \tag{39}$$

The representations used here are orthogonal, so $\rho(g)^{-1} = \rho(g)^{\mathsf{T}}$. Hence $\Pi_{\text{lin}}(A)$ is the group average of the natural action $A \mapsto \rho_{\text{out}}(g) A \rho_{\text{in}}(g)^{-1}$ on linear maps. The projected elementary matrices span the complete intertwiner space. The implementation flattens each matrix to a vector, applies a singular-value decomposition (SVD) to remove linear dependencies, and reshapes the retained right-singular vectors into a Frobenius-orthonormal basis $\{B_p\}_{p=1}^{K}$. The optional generic backend can equivalently obtain the same space by solving the linear null-space constraints in Equation (35).

For irreducible representations, the statement can be sharpened. If V_λ and V_μ are inequivalent irreps, then according to Schur's lemma (Scott & Serre, 2012),

$$\text{Hom}_G(V_\lambda, V_\mu) = \{0\}.$$

If they are equivalent, the intertwiner space is isomorphic to the real endomorphism algebra

$$\text{End}_G(V_\lambda) = \text{Hom}_G(V_\lambda, V_\lambda).$$

Over $\mathbb{R}$, this algebra need not be one-dimensional. In particular, a non-one-dimensional real irrep of C_n is of complex type and has a two-dimensional real endomorphism algebra. In a standard rotation basis, these two directions may be represented by the identity matrix and the 90° rotation matrix. The implementation therefore constructs the complete real intertwiner space instead of assuming one scalar degree of freedom for every equal-irrep pair.

B.2.2 Irrep-to-regular linear map

Let V_λ be a representation of dimension d_λ, and let

$$\mathbb{R}[G] = \text{span}\{\delta_h : h \in G\}$$

denote the regular representation. For each $h \in G$, let δ_h denotes the basis vector with the group element h in the h-th coordinate and zeros elsewhere. The left-regular action of G on $\mathbb{R}[G]$ is defined by

$$\rho_{\text{reg}}(g)\delta_h = \delta_{gh}. \tag{40}$$

Here $g, \in G$ denotes the group element acting on the feature, $h, q \in G$ indexes the original regular coordinate, and the transformed regular coordinate respectively. Consequently, if $f = \sum_{h \in G} f_h \delta_h \in \mathbb{R}[G]$, then its transformed coordinates obey

$$[\rho_{\text{reg}}(g) f]_q = f_{g^{-1}q}, \qquad q \in G.$$

For an equivariant map

$$W : V_\lambda \longrightarrow \mathbb{R}[G],$$

one has

$$\dim_{\mathbb{R}} \mathrm{Hom}_G(V_\lambda, \mathbb{R}[G]) = d_\lambda. \tag{41}$$

Thus a single V_λ-to-regular block has d_λ independent symmetry-allowed directions.

Let $\{\mathbf{e}_i\}_{i=1}^{d_\lambda}$ be the coordinate basis of V_λ, where i indexes coordinates within the representation. The analytic Frobenius-orthonormal basis used by `we3nn` for $\mathrm{Hom}_G(V_\lambda, \mathbb{R}[G])$ is

$$(B_\alpha)_{h,i} = \frac{1}{\sqrt{|G|}} [\rho_\lambda(h)]_{i\alpha}, \qquad \alpha, i \in \{1, \dots, d_\lambda\}, \quad h \in G. \tag{42}$$

Here B_α is the α-th basis intertwiner, α labels the d_λ independent equivariant maps, h labels a coordinate of the regular output, and i labels a coordinate of the input representation V_λ.

The learned equivariant map is a linear combination of these fixed basis intertwiners,

$$W = \sum_{\alpha=1}^{d_\lambda} \theta_\alpha B_\alpha, \tag{43}$$

where

$$\theta = (\theta_1, \dots, \theta_{d_\lambda})^\mathsf{T} \in \mathbb{R}^{d_\lambda}$$

contains the learned scalar coefficients.

For an input

$$x = (x_1, \dots, x_{d_\lambda})^\mathsf{T} \in V_\lambda,$$

the output $f = Wx \in \mathbb{R}[G]$ contains one scalar coordinate f_h for each group element $h \in G$. Its h-th coordinate is

$$f_h = \sum_{i=1}^{d_\lambda} W_{hi} x_i \tag{44}$$

$$= \frac{1}{\sqrt{|G|}} \sum_{\alpha=1}^{d_\lambda} \sum_{i=1}^{d_\lambda} \theta_\alpha [\rho_\lambda(h)]_{i\alpha} x_i \tag{45}$$

$$= \frac{1}{\sqrt{|G|}} \theta^\mathsf{T} \rho_\lambda(h)^\mathsf{T} x. \tag{46}$$

Equivalently,

$$f_h = \frac{1}{\sqrt{|G|}} [\rho_\lambda(h)\theta]^\mathsf{T} x. \tag{47}$$

Thus, the same learned probe θ is transformed by every group element h, and the corresponding transformed probe $\rho_\lambda(h)\theta$ is compared with the input x through an inner product. The resulting scalar is stored in the regular coordinate indexed by h.

For the standard two-dimensional C_6 vector representation, for example, the output contains six coordinates,

$$(f_0, \dots, f_5),$$

and each f_h measures the inner product of the input vector with a different C_6-rotated copy of the learned probe θ.

B.2.3 Regular-to-irrep linear map

Consider the reverse map

$$W : \mathbb{R}[G] \longrightarrow V_\lambda.$$

Again,

$$\dim_{\mathbb{R}} \mathrm{Hom}_G(\mathbb{R}[G], V_\lambda) = d_\lambda. \tag{48}$$

An equivariant map is completely determined by the image of the identity basis vector δ_e. Indeed, if $W\delta_e = v \in V_\lambda$, then for any $h \in G$,

$$W\delta_h = W\rho_{\mathrm{reg}}(h)\delta_e \tag{49}$$

$$= \rho_\lambda(h) W \delta_e \tag{50}$$

$$= \rho_\lambda(h) v. \tag{51}$$

The analytic orthonormal basis is

$$(B_\alpha)_{i,h} = \frac{1}{\sqrt{|G|}}[\rho_\lambda(h)]_{i\alpha}, \qquad \alpha, i \in \{1, \dots, d_\lambda\}, \quad h \in G. \tag{52}$$

For a regular input $f = \sum_{h\in G} f_h \delta_h$ and learned coefficient vector $\theta = (\theta_1, \dots, \theta_{d_\lambda})^\mathsf{T}$, the output is

$$z = \frac{1}{\sqrt{|G|}} \sum_{h\in G} f_h \rho_\lambda(h)\theta. \tag{53}$$

In components,

$$z_i = \frac{1}{\sqrt{|G|}} \sum_{h\in G} \sum_{\alpha=1}^{d_\lambda} f_h [\rho_\lambda(h)]_{i\alpha}\theta_\alpha.$$

With this orthonormal convention, $\theta = \sqrt{|G|}\, W\delta_e$. Thus the layer synthesizes an ordinary representation-valued feature by combining group-indexed regular coordinates with symmetry-related copies of a learned template.

B.2.4 Regular-to-regular linear map

Let

$$W : \mathbb{R}[G] \longrightarrow \mathbb{R}[G],$$

and write the input and output as $f = \sum_{h\in G} f_h \delta_h$ and $z = \sum_{q\in G} z_q \delta_q$, respectively. Equivariance under the left-regular action implies

$$W_{q,h} = \frac{1}{\sqrt{|G|}} \theta_{h^{-1}q}, \qquad q, h \in G, \tag{54}$$

where $W_{q,h}$ is the matrix entry from input coordinate h to output coordinate q, and θ_η is one learned scalar for each relative group element $\eta = h^{-1}q \in G$. Equivalently,

$$z_q = \frac{1}{\sqrt{|G|}} \sum_{h\in G} \theta_{h^{-1}q} f_h.$$

This is a finite-group convolution in the indexing convention induced by the library's left-regular action.

An unconstrained regular-to-regular matrix contains $|G|^2$ coefficients, whereas the equivariant map contains only $|G|$ independent coefficients. With U regular input copies and M regular output copies,

$$\mathbb{R}[G]^{\oplus U} \longrightarrow \mathbb{R}[G]^{\oplus M},$$

the operation is

$$z_{m,q} = \frac{1}{\sqrt{|G|}} \sum_{u=1}^{U} \sum_{h\in G} \theta_{mu,\, h^{-1}q} f_{u,h},$$

where u indexes an input copy, m indexes an output copy, and the learnable tensor has shape

$$M \times U \times |G|. \tag{55}$$

The factor $1/\sqrt{|G|}$ in the analytic regular-representation formulas makes the symmetry bases orthonormal in the Euclidean/Frobenius inner product. It is a normalization convention rather than an additional symmetry constraint, and it can be absorbed into the learned coefficients.

B.2.5 Equivariant bias and nonlinearities

For an affine map

$$y = Wx + b,$$

equivariance additionally requires

$$\rho_{\text{out}}(g)b = b \qquad \text{for every } g \in G. \tag{56}$$

The allowed bias therefore belongs to the invariant subspace

$$V_{\text{out}}^G = \{v \in V_{\text{out}} : \rho_{\text{out}}(g)v = v \text{ for every } g \in G\} \cong \text{Hom}_G(\mathbb{R}_{\text{triv}}, V_{\text{out}}),$$

where $\mathbb{R}_{\text{triv}}$ is the one-dimensional trivial representation, for which $\rho_{\text{triv}}(g) = 1$. A nontrivial vector irrep has no invariant vector and therefore admits no arbitrary vector bias. A regular field has a one-dimensional invariant subspace, corresponding to a bias that is constant over all group coordinates.

The regular representation is a permutation representation. More generally, if a representation acts by permuting its coordinates, then any scalar nonlinearity $\sigma : \mathbb{R} \to \mathbb{R}$ applied identically to every coordinate is equivariant. For a regular feature f,

$$\sigma(\rho_{\text{reg}}(g)f) = \rho_{\text{reg}}(g)\sigma(f), \tag{57}$$

where $\sigma(f)$ denotes coordinatewise application. This permits ordinary nonlinearities such as ReLU between regular-to-regular equivariant linear layers.

B.3 Equivariant tensor products

Whereas equivariant linear layer parameterizes equivariant linear maps, the Tensor Product and Fully Connected TensorProduct parameterize equivariant bilinear maps.

Let

$$x \in V_\lambda, \qquad y \in V_\mu,$$

where V_λ and V_μ have dimensions d_λ and d_μ, and transform as

$$x \longmapsto \rho_\lambda(g)x, \qquad y \longmapsto \rho_\mu(g)y.$$

Their raw tensor product belongs to $V_\lambda \otimes V_\mu$ and transforms according to

$$\rho_{\lambda\otimes\mu}(g) = \rho_\lambda(g) \otimes \rho_\mu(g). \tag{58}$$

Suppose the desired output belongs to V_ν, with dimension d_ν. An equivariant bilinear map

$$B : V_\lambda \times V_\mu \longrightarrow V_\nu$$

is equivalent to a linear intertwiner

$$C : V_\lambda \otimes V_\mu \longrightarrow V_\nu.$$

Thus all equivariant bilinear maps form the real vector space

$$\text{Hom}_G(V_\lambda \otimes V_\mu, V_\nu). \tag{59}$$

Let

$$K_{\nu;\lambda\mu} = \dim_{\mathbb{R}} \text{Hom}_G(V_\lambda \otimes V_\mu, V_\nu),$$

and let $\{C_p\}_{p=1}^{K_{\nu;\lambda\mu}}$ be a basis of this space. A basis tensor has entries

$$(C_p)_{oij}, \qquad o = 1, \dots, d_\nu, \quad i = 1, \dots, d_\lambda, \quad j = 1, \dots, d_\mu.$$

For trainable scalar coefficients w_p, the bilinear map is

$$B(x, y)_o = \sum_{p=1}^{K_{\nu;\lambda\mu}} w_p \sum_{i=1}^{d_\lambda} \sum_{j=1}^{d_\mu} (C_p)_{oij} x_i y_j. \tag{60}$$

The tensors C_p are fixed by group symmetry, while the coefficients w_p are learned or supplied externally.

B.3.1 GENERIC IRREP TENSOR PRODUCT

When the left, right, and output representations are non-regular irreps, constructs the complete real coupling space

$$\mathrm{Hom}_G(V_\lambda \otimes V_\mu, V_\nu).$$

View a rank-three tensor $T \in \mathbb{R}^{d_\nu \times d_\lambda \times d_\mu}$ as a matrix from $V_\lambda \otimes V_\mu$ to V_ν. The tensor-product Reynolds projection is

$$\Pi_{\mathrm{TP}}(T) = \frac{1}{|G|} \sum_{g \in G} \rho_\nu(g)\, T \left(\rho_\lambda(g) \otimes \rho_\mu(g)\right)^{\mathsf{T}}. \tag{61}$$

In coordinate form,

$$[\Pi_{\mathrm{TP}}(T)]_{oij} = \frac{1}{|G|} \sum_{g \in G} \sum_{o'=1}^{d_\nu} \sum_{i'=1}^{d_\lambda} \sum_{j'=1}^{d_\mu} [\rho_\nu(g)]_{oo'} [\rho_\lambda(g)]_{ii'} [\rho_\mu(g)]_{jj'} T_{o'i'j'}.$$

The implementation applies this projection to the elementary rank-three tensor basis, flattens the projected tensors, and uses an SVD to retain all linearly independent directions. The resulting tensors

$$C_p \in \mathbb{R}^{d_\nu \times d_\lambda \times d_\mu}$$

are orthonormal in the entrywise Frobenius inner product and span the complete real equivariant coupling space.

For real irreps, the number $K_{\nu;\lambda\mu}$ is the real Hom-space dimension, not necessarily the number of copies of V_ν in the tensor-product decomposition. More precisely, for an irreducible output,

$$K_{\nu;\lambda\mu} = N_{\nu;\lambda\mu} \dim_{\mathbb{R}} \mathrm{End}_G(V_\nu),$$

where $N_{\nu;\lambda\mu}$ is the representation-copy multiplicity of V_ν inside $V_\lambda \otimes V_\mu$. For a two-dimensional complex-type real irrep of C_n, the endomorphism factor is two. Therefore one representation copy gives two independent real neural-network coupling weights. The tensor-product modules use this complete real basis.

Now suppose the left input contains U copies of V_λ, the right input contains V copies of V_μ, and the output contains M copies of V_ν:

$$V_\lambda^{\oplus U}, \qquad V_\mu^{\oplus V}, \qquad V_\nu^{\oplus M}.$$

For the default fully connected connection mode `uvw`, the layer learns scalars

$$w_{muvp},$$

where

$$m = 1, \ldots, M, \quad u = 1, \ldots, U, \quad v = 1, \ldots, V, \quad p = 1, \ldots, K_{\nu;\lambda\mu}.$$

It computes

$$z_{m,o} = \sum_{u=1}^{U} \sum_{v=1}^{V} \sum_{p=1}^{K_{\nu;\lambda\mu}} w_{muvp} \sum_{i=1}^{d_\lambda} \sum_{j=1}^{d_\mu} (C_p)_{oij} x_{u,i} y_{v,j}. \tag{62}$$

Here $x_{u,i}$ is coordinate i of left copy u, $y_{v,j}$ is coordinate j of right copy v, and $z_{m,o}$ is coordinate o of output copy m. The trainable tensor for this representation triple has shape

$$M \times U \times V \times K_{\nu;\lambda\mu}.$$

The connection mode `uvu` identifies the output multiplicity index with the left-input multiplicity index. It therefore requires $M = U$ for the selected representation block and learns w_{uvp}, with shape

$$U \times V \times K_{\nu;\lambda\mu}.$$

The operation is

$$z_{u,o} = \sum_{v=1}^{V} \sum_{p=1}^{K_{\nu;\lambda\mu}} w_{uvp} \sum_{i=1}^{d_\lambda} \sum_{j=1}^{d_\mu} (C_p)_{oij} x_{u,i} y_{v,j}. \tag{63}$$

There is no sum over u: output copy u is connected only to left-input copy u, while all right-input copies v may still be mixed. In the automatic compiler, the pairing is by occurrence order within the relevant multiplicity chunks.

B.4 TENSOR PRODUCTS INVOLVING REGULAR REPRESENTATIONS

When a regular representation occurs among the input or output representations, `we3nn` uses analytic orbit constructions instead of forming the potentially large generic rank-three coupling space. If the output is regular, the regular-output construction is selected even when one or both inputs are also regular. Otherwise, a regular left or right input is handled by the corresponding analytic input construction.

B.4.1 ARBITRARY REPRESENTATIONS TO REGULAR OUTPUT

Consider a single-copy bilinear map

$$B : V_\lambda \times V_\mu \longrightarrow \mathbb{R}[G].$$

Let

$$A \in \mathbb{R}^{d_\lambda \times d_\mu}$$

be an arbitrary learned coefficient matrix. Its entry A_{ij} couples coordinate i of V_λ to coordinate j of V_μ in the reference frame. For each $q \in G$, define the inverse-transformed input coordinates

$$x^{(q)} = \rho_\lambda(q)^\mathsf{T} x = \rho_\lambda(q^{-1})x, \qquad y^{(q)} = \rho_\mu(q)^\mathsf{T} y = \rho_\mu(q^{-1})y.$$

The regular output $z = \sum_{q\in G} z_q \delta_q$ is

$$z_q = \frac{1}{\sqrt{|G|}} \sum_{i=1}^{d_\lambda} \sum_{j=1}^{d_\mu} A_{ij} \left[\rho_\lambda(q)^\mathsf{T} x\right]_i \left[\rho_\mu(q)^\mathsf{T} y\right]_j . \tag{64}$$

Thus the coordinate z_q stores the same learned reference-frame bilinear form evaluated after both inputs have been moved to the frame associated with q. The map is completely determined by this arbitrary bilinear form at one reference group coordinate, and

$$\dim_{\mathbb{R}} \mathrm{Hom}_G(V_\lambda \otimes V_\mu, \mathbb{R}[G]) = d_\lambda d_\mu. \tag{65}$$

Accordingly, the matrix entries A_{ij} are a reshaping of the $K = d_\lambda d_\mu$ independent real coupling coefficients.

With U, V, and M left, right, and regular-output copies, respectively, the `uvw` coefficient tensor can be written A_{muvij} and has shape

$$M \times U \times V \times d_\lambda \times d_\mu.$$

The implementation may store the last two axes as the analytic coupling shape rather than flattening them to a single index p.

B.4.2 REGULAR LEFT INPUT

Consider

$$B : \mathbb{R}[G] \times V_\mu \longrightarrow V_\nu.$$

Write the regular input as $f = \sum_{q\in G} f_q \delta_q$, let $y \in V_\mu$, and let

$$A : V_\mu \longrightarrow V_\nu, \qquad A \in \mathbb{R}^{d_\nu \times d_\mu},$$

be an arbitrary learned reference-frame linear map. The equivariant bilinear operation is

$$z = \frac{1}{\sqrt{|G|}} \sum_{q\in G} f_q \, \rho_\nu(q) A \rho_\mu(q)^\mathsf{T} y. \tag{66}$$

In coordinates, this is

$$z_o = \frac{1}{\sqrt{|G|}} \sum_{q\in G} \sum_{i=1}^{d_\nu} \sum_{j=1}^{d_\mu} f_q [\rho_\nu(q)]_{oi} A_{ij} [\rho_\mu(q)^\mathsf{T} y]_j.$$

The regular scalar f_q selects the contribution from group coordinate q. The non-regular input is transformed into the corresponding frame, the reference map A is applied, and the result is transformed into the output frame. The coupling-space dimension is

$$\dim_{\mathbb{R}} \mathrm{Hom}_G(\mathbb{R}[G] \otimes V_\mu, V_\nu) = d_\nu d_\mu. \tag{67}$$

B.4.3 REGULAR RIGHT INPUT

The right-regular case is analogous. For

$$B : V_\lambda \times \mathbb{R}[G] \longrightarrow V_\nu,$$

write the regular right input as $f = \sum_{q\in G} f_q \delta_q$, let $x \in V_\lambda$, and let

$$A : V_\lambda \longrightarrow V_\nu, \qquad A \in \mathbb{R}^{d_\nu \times d_\lambda}.$$

Then

$$z = \frac{1}{\sqrt{|G|}} \sum_{q\in G} f_q \, \rho_\nu(q) A \rho_\lambda(q)^{\mathsf{T}} x, \tag{68}$$

or, in components,

$$z_o = \frac{1}{\sqrt{|G|}} \sum_{q\in G} \sum_{i=1}^{d_\nu} \sum_{j=1}^{d_\lambda} f_q [\rho_\nu(q)]_{oi} A_{ij} [\rho_\lambda(q)^{\mathsf{T}} x]_j.$$

The coupling-space dimension is

$$\dim_{\mathbb{R}} \mathrm{Hom}_G(V_\lambda \otimes \mathbb{R}[G], V_\nu) = d_\nu d_\lambda. \tag{69}$$

B.4.4 REGULAR–REGULAR TO REGULAR TENSOR PRODUCT

An important special case is

$$\mathbb{R}[G] \otimes \mathbb{R}[G] \longrightarrow \mathbb{R}[G].$$

Since the output is regular, this case is an instance of Equation (64). For a single left input copy, single right input copy, and single output copy, the learned reference bilinear form is

$$A \in \mathbb{R}^{|G|\times|G|}$$

and contains $|G|^2$ independent coefficients.

Let x_a and y_b be the input coordinates associated with $a, b \in G$. Then, for every output group coordinate $q \in G$,

$$z_q = \frac{1}{\sqrt{|G|}} \sum_{a\in G} \sum_{b\in G} A_{ab} x_{qa} y_{qb}, \tag{70}$$

where qa and qb denote group multiplication. Thus the same learned interaction between relative group coordinates a and b is evaluated around every output coordinate q.

For the cyclic group C_n, identify group elements with integers modulo n. Group multiplication then becomes addition modulo n, and

$$z_k = \frac{1}{\sqrt{n}} \sum_{a=0}^{n-1} \sum_{b=0}^{n-1} A_{ab} x_{k+a} y_{k+b}, \tag{71}$$

where $k + a$ and $k + b$ are reduced modulo n.

Coordinatewise multiplication,

$$z_k = x_k y_k,$$

is contained in this family. Under the orthonormal $1/\sqrt{n}$ convention used above, it is obtained by setting

$$A_{00} = \sqrt{n} \qquad \text{and} \qquad A_{ab} = 0 \text{ for } (a, b) \neq (0, 0).$$

If $A_{00} = 1$, the result is instead $z_k = n^{-1/2} x_k y_k$. The normalization does not affect expressivity because it can be absorbed into the learned coefficient. The full tensor product is more expressive than coordinatewise multiplication because it can couple arbitrary pairs of relative group orientations.

B.5 Internal and externally supplied tensor-product weights

The coupling tensors define the symmetry-allowed bilinear basis, but their scalar coefficients may be supplied in two ways.

With internal weights, the tensor-product module owns trainable parameters that are shared across all samples. Different representation paths and multiplicity combinations receive independent learned coefficients, but those coefficients do not depend on the numerical value of a particular input sample.

Alternatively, the tensor product can receive externally generated weights, for example, an invariant contextual feature. For geometric neural networks, a radial MLP may generate tensor-product coefficients from $\|r_{ij}\|$. Since the supplied coefficients are invariant scalars, input-dependent weighting does not alter the equivariance of the coupling basis.

B.6 Bravais lattice

A two-dimensional Bravais lattice is generated by two linearly independent primitive lattice vectors a_1 and a_2:

$$\Lambda = \{n_1 a_1 + n_2 a_2 \ : \ n_1, n_2 \in \mathbb{Z}\}\,. \tag{72}$$

B.7 Voronoi boundaries

The minimizer is unique except on the Voronoi boundaries of $\mathcal{A}$, which form a measure-zero subset of the plane. Because of the floating-point representation of coordinates, the Voronoi boundaries are not expected to be encountered in practice.

B.8 WA covariance proof

Proof. Because g is an isometry,

$$\|ga - gx\| = \|a - x\|. \tag{73}$$

Since g is a bijection of $\mathcal{A}$, the unique nearest anchor to gx is $ga(x)$. Therefore

$$\begin{aligned} v_{\mathrm{WA}}(gx) &= a(gx) - gx && (74)\\ &= ga(x) - gx && (75)\\ &= (Q_g a(x) + \tau_g) - (Q_g x + \tau_g) && (76)\\ &= Q_g(a(x) - x) && (77)\\ &= Q_g v_{\mathrm{WA}}(x). && (78) \end{aligned}$$

The translational part τ_g, including the fractional translation in a glide reflection, cancels exactly. □

B.9 Translation-suborbit regular-representation implementation details

The deterministic orbit lift is first constructed using a right-regular convention. We denote this raw lifted tensor by $\bar{X}_i$. Under a wallpaper operation $g \in G$, with corresponding point-group element $p_g = g\Lambda \in P$, it satisfies

$$\bar{X}_{\pi_g(i),p}(gX) = \bar{X}_{i,pp_g}(X), \qquad p \in P. \tag{79}$$

Thus, transforming the input by g induces right multiplication by p_g on the group index.

The regular representation used by `we3nn` follows the left-regular convention

$$[\lambda(p_g)u](k) = u(p_g^{-1}k). \tag{80}$$

We therefore apply the fixed inversion reindexing

$$X^{\mathrm{reg}}_{i,k} = \bar{X}_{i,k^{-1}}, \qquad k \in P. \tag{81}$$

Using Equation (79), the transformed field obeys

$$\begin{aligned} X^{\mathrm{reg}}_{\pi_g(i),k}(gX) &= \bar{X}_{\pi_g(i),k^{-1}}(gX) && (82)\\ &= \bar{X}_{i,k^{-1}p_g}(X) && (83)\\ &= X^{\mathrm{reg}}_{i,p_g^{-1}k}(X). && (84) \end{aligned}$$

Hence, inversion of the group index converts the deterministic orbit lift from the right-regular convention to the left-regular convention used by `we3nn`, so that X_i^{reg} can be passed directly to regular-representation layers.

Dimensions of the lift. For a single selected WA family with m member channels, each group block of the lift in Equation (26) contains

$$C_{\mathrm{lift}} = 2 + m + 2m = 2 + 3m \tag{85}$$

channels. For the selected nonsymmorphic anchor orbits, $m = 2$, and therefore

$$C_{\mathrm{lift}} = 8. \tag{86}$$

Equivariant use of the registry feature. When the group axis of Z_i is retained, the registry feature can keep both its trivial component and the nontrivial components that encode how the WA geometry is associated with the permuting translation-suborbit member.

B.10 Action on translation-suborbit members

The decomposition

$$\mathcal{A} = \coprod_{\alpha=1}^{m} (q_\alpha + \Lambda) \tag{87}$$

is chosen such that the representatives q_α are distinct modulo the Bravais lattice. Every wallpaper operation maps each translation suborbit to another translation suborbit in the same anchor family. Indeed, let $g(x) = Q_g x + \tau_g$. Since the linear part of a wallpaper operation preserves the Bravais lattice,

$$Q_g \Lambda = \Lambda. \tag{88}$$

It follows that

$$\begin{aligned} g\big(q_\alpha + \Lambda\big) &= \big\{ g\big(q_\alpha + t\big) \,\big|\, t \in \Lambda \big\} && (89)\\ &= \{ g q_\alpha + Q_g t \,|\, t \in \Lambda \} && (90)\\ &= g q_\alpha + \Lambda. && (91) \end{aligned}$$

Because $\mathcal{A}$ is G-invariant, $g q_\alpha \in \mathcal{A}$. Hence, there exists a unique index $\beta \in \{1, \dots, m\}$ such that

$$g q_\alpha \equiv q_\beta \pmod{\Lambda}. \tag{92}$$

We write $\sigma_g(\alpha) = \beta$. Since g is invertible, σ_g is a permutation of the translation-suborbit indices, and

$$\sigma_{g_1 g_2} = \sigma_{g_1} \circ \sigma_{g_2}. \tag{93}$$

A lattice translation $t \in \Lambda$ maps every suborbit $q_\alpha + \Lambda$ to itself, so σ_t is the identity and σ_g depends only on the coset $p_g = g\Lambda$. The permutation matrices of σ_g are therefore exactly the member permutation representation of Equation (24),

$$\rho_{\mathrm{mem}}(p_g) e_\alpha = e_{\sigma_g(\alpha)}. \tag{94}$$

If $\alpha(x)$ denotes the translation-suborbit member of the unique nearest anchor to x, nearest-anchor equivariance $a(gx) = g a(x)$ implies

$$\alpha(gx) = \sigma_g\big(\alpha(x)\big). \tag{95}$$

Consequently, the member one-hot vector $\ell(x) = \mathrm{onehot}(\alpha(x))$ transforms according to

$$\ell(gx) = \rho_{\mathrm{mem}}(p_g)\, \ell(x). \tag{96}$$

For node i, with $\ell_i = \ell(x_{i,xy})$ and $v_i = v_{\mathrm{WA}}(x_{i,xy})$, this gives $\ell_{\pi_g(i)}(gX) = \rho_{\mathrm{mem}}(p_g)\ell_i(X)$, and Equation (21) gives $v_{\pi_g(i)}(gX) = Q_g v_i(X)$. Thus, the member label is not generally an invariant scalar feature. Rather, it transforms under the permutation representation induced by the wallpaper-group action on the translation suborbits.

B.11 SELECTED ANCHOR ORBITS

For each symmorphic wallpaper group, we choose a point whose site-symmetry point group equals the full point group. Such an orbit has one translation-suborbit member per primitive cell, $m = 1$. For the four nonsymmorphic wallpaper groups pg, pmg, pgg, and $p4g$, the selected minimum-multiplicity orbit has $m = 2$ in the implementation described in Section 3.3. Complete visualizations of the group actions and the associated WA-vector transformations are provided in Appendix C.

Table 3 records the fixed anchor-orbit choices used by the simulator. Coordinates are fractional coordinates in the basis and origin convention used by the implementation. Labels such as C2-A, C3-A, and C4-A are implementation labels rather than universal Wyckoff symbols.

The column "N of anchor-orbit choices" reports how many inequivalent anchor orbits have the same highest site-symmetry type as the chosen anchor. For Wyckoff positions with free coordinates, continuously many such orbits exist. In those cases, the table gives the number and dimension of the corresponding Wyckoff families. These choices are not used simultaneously. The final column states whether the wallpaper group is symmorphic or nonsymmorphic.

Table 3: Selected Wyckoff-anchor orbits. The member multiplicity m is counted modulo primitive translations. The fifth column reports the number of inequivalent anchor-orbit choices. The subscript prim marks representatives given modulo primitive translations: the conventional centered cells of cm and cmm contain a centering-related copy that is equivalent modulo primitive translations.

Group	Selected anchor	Representatives A^0	m	N of anchor-orbit choices	Symmorphicity
$p1$	lattice anchor	$\{q_1 = (0,0)\}$	1	continuum: one two-parameter family	symm.
$p2$	C2-A	$\{q_1 = (\frac{1}{2},\frac{1}{2})\}$	1	4 discrete C2 orbits	symm.
pm	lattice/ mirror anchor	$\{q_1 = (0,0)\}$	1	continuum: two one-parameter mirror families	symm.
pg	minimum orbit	$\{q_1 = (0,0),\ q_2 = (0,\frac{1}{2})\}$	2	continuum: one two-parameter family	non-symm.
cm	lattice/ mirror anchor	$\{q_1 = (0,0)\}_{\mathrm{prim}}$	1	continuum: one one-parameter mirror family	symm.
pmm	C2-A	$\{q_1 = (\frac{1}{2},\frac{1}{2})\}$	1	4 discrete C2v orbits	symm.
pmg	C2-A	$\{q_1 = (\frac{1}{2},\frac{1}{2}),\ q_2 = (0,\frac{1}{2})\}$	2	2 discrete C2 orbits	non-symm.
pgg	C2-A	$\{q_1 = (0,0),\ q_2 = (\frac{1}{2},\frac{1}{2})\}$	2	2 discrete C2 orbits	non-symm.
cmm	C2-A	$\{q_1 = (0,0)\}_{\mathrm{prim}}$	1	2 discrete C2v orbits	symm.
$p4$	C4-A	$\{q_1 = (\frac{1}{2},\frac{1}{2})\}$	1	2 discrete C4 orbits	symm.
$p4m$	C4-A	$\{q_1 = (\frac{1}{2},\frac{1}{2})\}$	1	2 discrete C4v orbits	symm.
$p4g$	C4	$\{q_1 = (\frac{1}{2},\frac{1}{2}),\ q_2 = (0,0)\}$	2	1 discrete C4 orbit	non-symm.
$p3$	C3-A	$\{q_1 = (\frac{1}{3},\frac{2}{3})\}$	1	3 discrete C3 orbits	symm.
$p3m1$	C3-A	$\{q_1 = (\frac{1}{3},\frac{2}{3})\}$	1	3 discrete C3v orbits	symm.
$p31m$	C3-B	$\{q_1 = (0,0)\}$	1	1 discrete C3v orbit	symm.
$p6$	C6	$\{q_1 = (0,0)\}$	1	1 discrete C6 orbit	symm.
$p6m$	C6	$\{q_1 = (0,0)\}$	1	1 discrete C6v orbit	symm.

The number of anchor-orbit choices in the fifth column is tied to the highest site-symmetry type selected for the anchor. It does not count orbits of the same multiplicity with a different site-symmetry type. The table records the fixed implementation choice of the WallpaperNet framework, rather than every possible crystallographic origin convention.

B.12 PROOF OF FULL WALLPAPER EQUIVARIANCE

We prove Theorem 1 in the notation of Section 3.4.

Proof. Suppose the incoming fields satisfy their assigned transformation laws. Together with the geometric encodings, they form an edge input obeying

$$z_{\pi_g(i)\pi_g(j)}(gX) = \rho_{\text{in}}(p_g) z_{ij}(X). \tag{97}$$

Equivariance of the message function gives

$$m_{\pi_g(i)\pi_g(j)}(gX) = \rho_m(p_g) m_{ij}(X). \tag{98}$$

Because the graph construction is compatible with the wallpaper action, sum aggregation commutes with the induced node reindexing. Therefore, for invariant scalar aggregation weights satisfying

$$a_{\pi_g(i)\pi_g(j)}(gX) = a_{ij}(X), \tag{99}$$

we have

$$\sum_j a_{\pi_g(i)\pi_g(j)}(gX)\, m_{\pi_g(i)\pi_g(j)}(gX) = \sum_j a_{ij}(X)\, \rho_m(p_g) m_{ij}(X) \tag{100}$$

$$= \rho_m(p_g) \sum_j a_{ij}(X) m_{ij}(X). \tag{101}$$

Thus the aggregated message transforms in the representation ρ_m. The equivariant node update therefore preserves the prescribed transformation law at the next layer. Initialization supplies the base case, so induction proves Equation (31) for every layer and equivariant readout. □

In an invariant-message model, hidden scalar fields carry the trivial representation. Its planar scalar-weighted vector construction is equivariant because scalar multiplication is the canonical intertwiner

$$\mu \in \text{Hom}_P\left(\mathbb{R}_{\text{triv}} \otimes V_{\text{vec}}, V_{\text{vec}}\right), \qquad \mu(a \otimes v) = av. \tag{102}$$

Here $\mathbb{R}_{\text{triv}}$ is the one-dimensional trivial representation and $V_{\text{vec}} = \mathbb{R}^2$ carries the standard planar representation,

$$\rho_{\text{vec}}(p) = Q_p. \tag{103}$$

C WALLPAPER-GROUP ACTION VISUALIZATIONS

Tables 4–20 visualize the selected WA orbit and the action of the symmetry operations for all 17 wallpaper groups. Each table includes its own legend, and arrows show the transformation of the molecular graph and the corresponding WA vectors.

Table 4: Wallpaper group `p1`, WA is pointing to lattice anchor at $q_1 = (0, 0)$

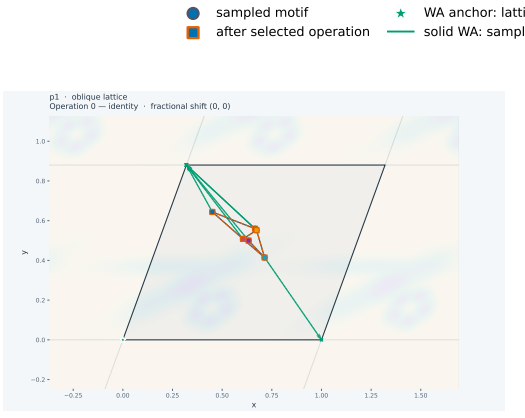


Table 5: Wallpaper group p2, WA is pointing to C2-A at $q_1 = (\frac{1}{2}, \frac{1}{2})$

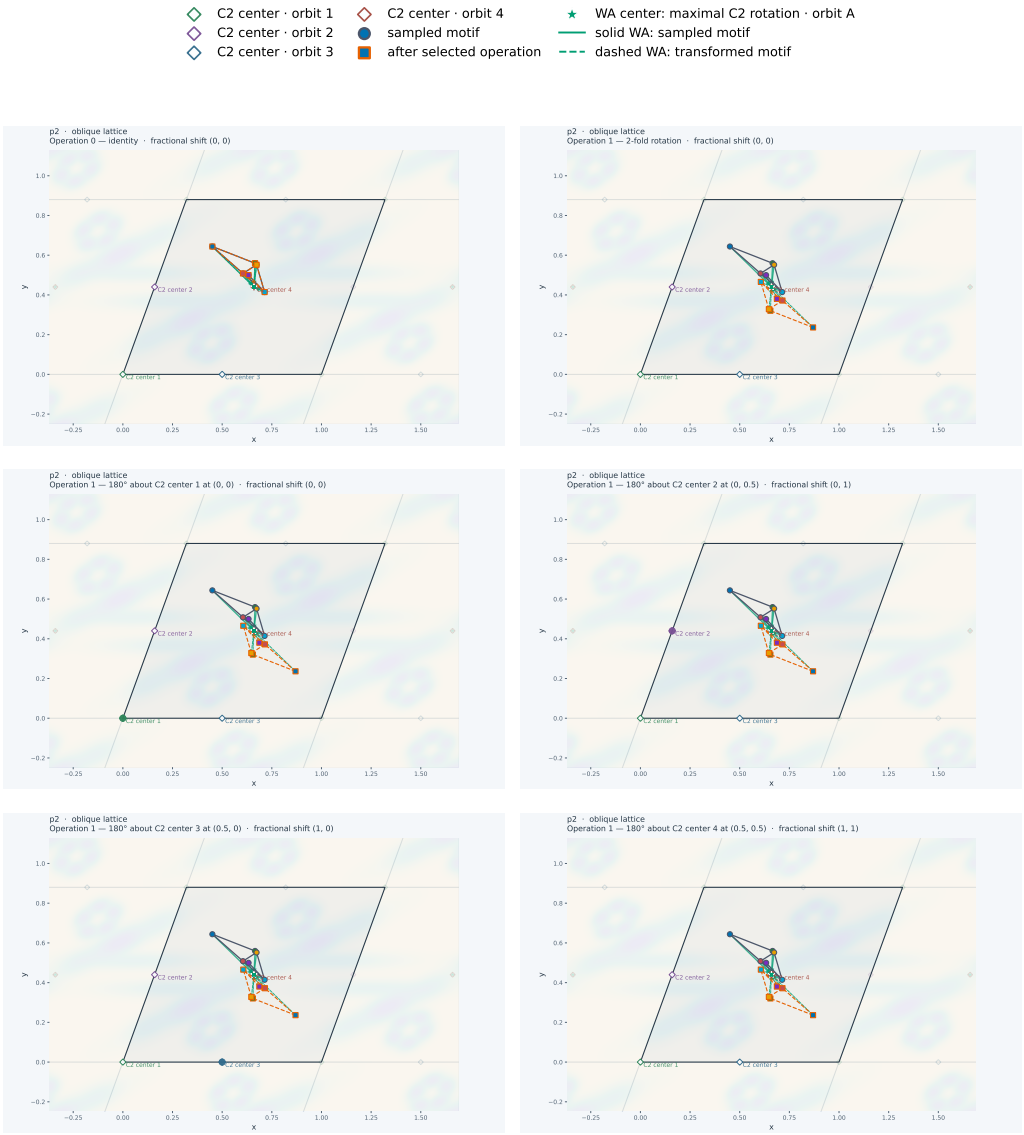


Table 6: Wallpaper group pm, WA is pointing to lattice anchor at $q_1 = (0, 0)$

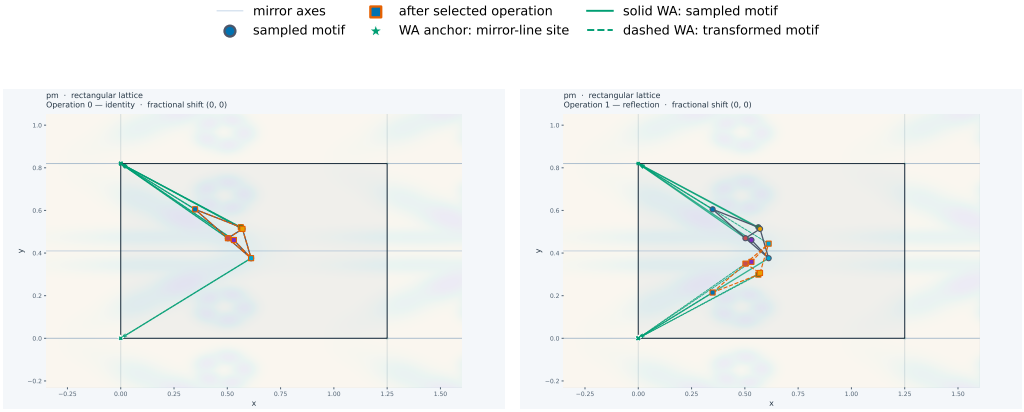


Table 7: Wallpaper group pg, WA is pointing to lattice anchor at $q_1 = (0, 0),\ q_2 = (0, \frac{1}{2})$

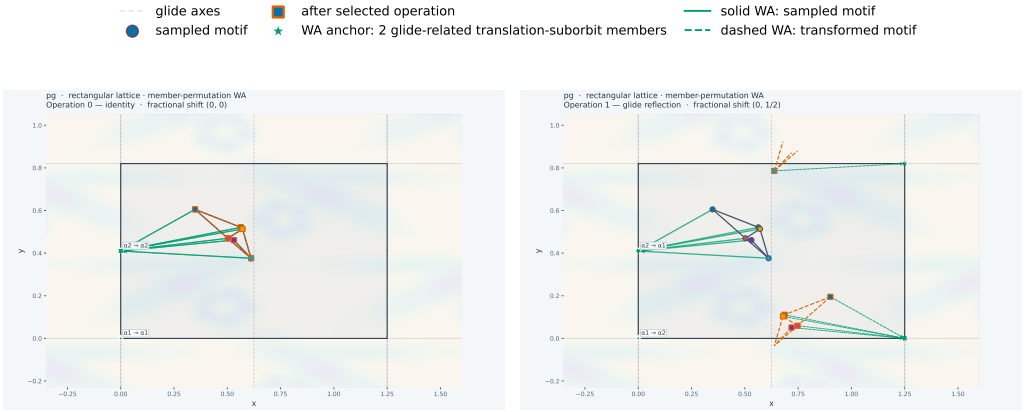


Table 8: Wallpaper group cm, WA is pointing to lattice anchor at $q_1 = (0, 0)$, centered conventional cell shown, with the WA member multiplicity counted modulo primitive translations

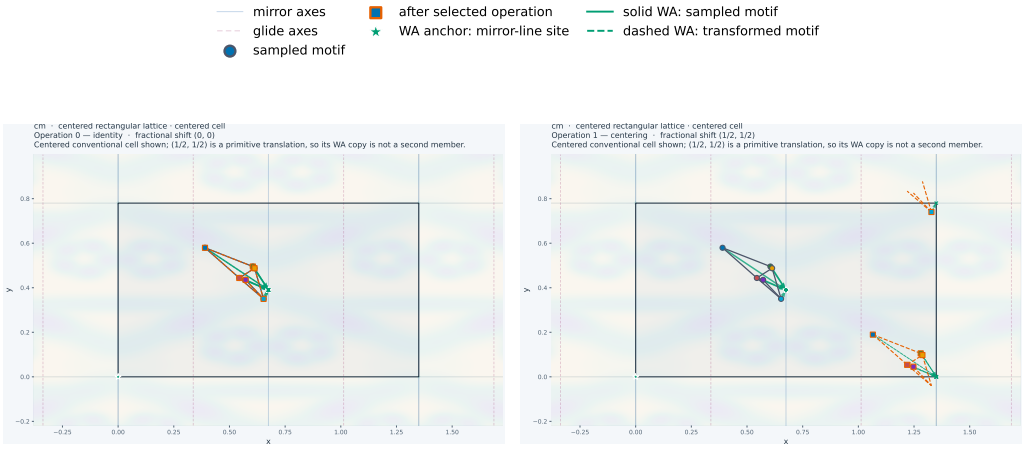


Table 8: Wallpaper group cm (continued)

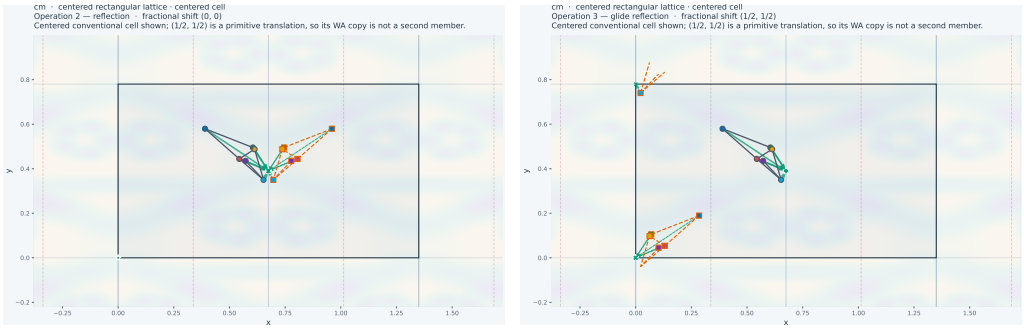


Table 9: Wallpaper group pmm, WA is pointing to C2-A at $q_1 = (\frac{1}{2}, \frac{1}{2})$

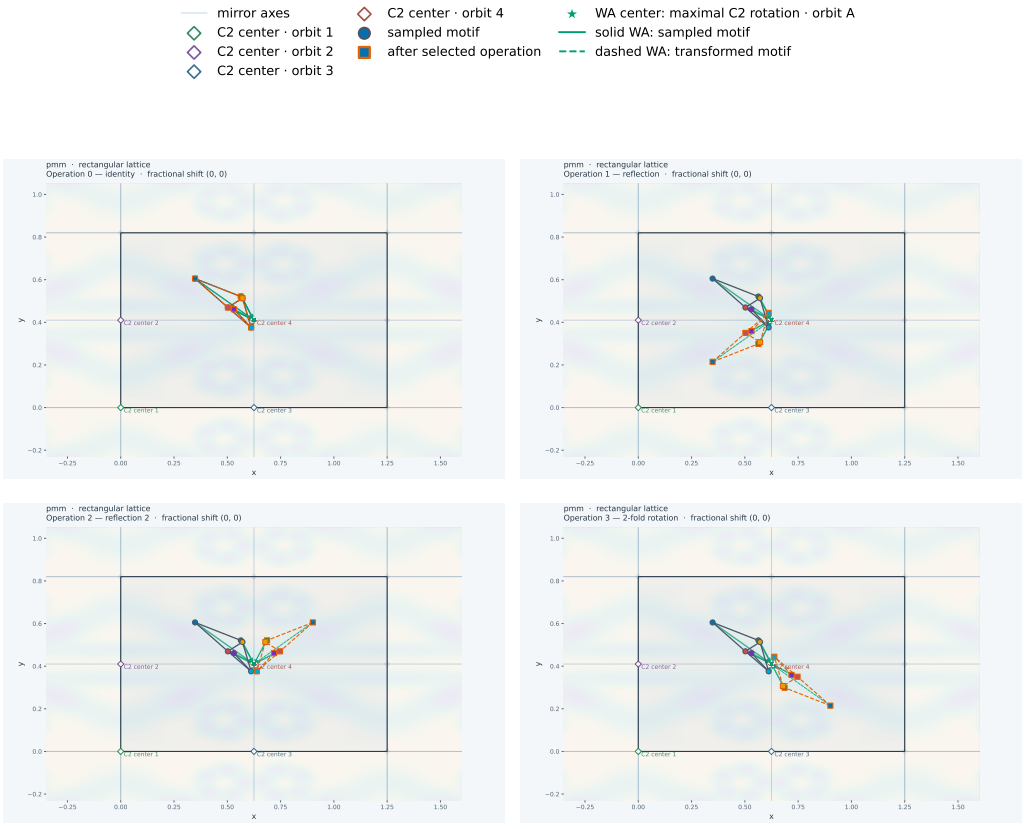


Table 9: Wallpaper group pmm (continued)

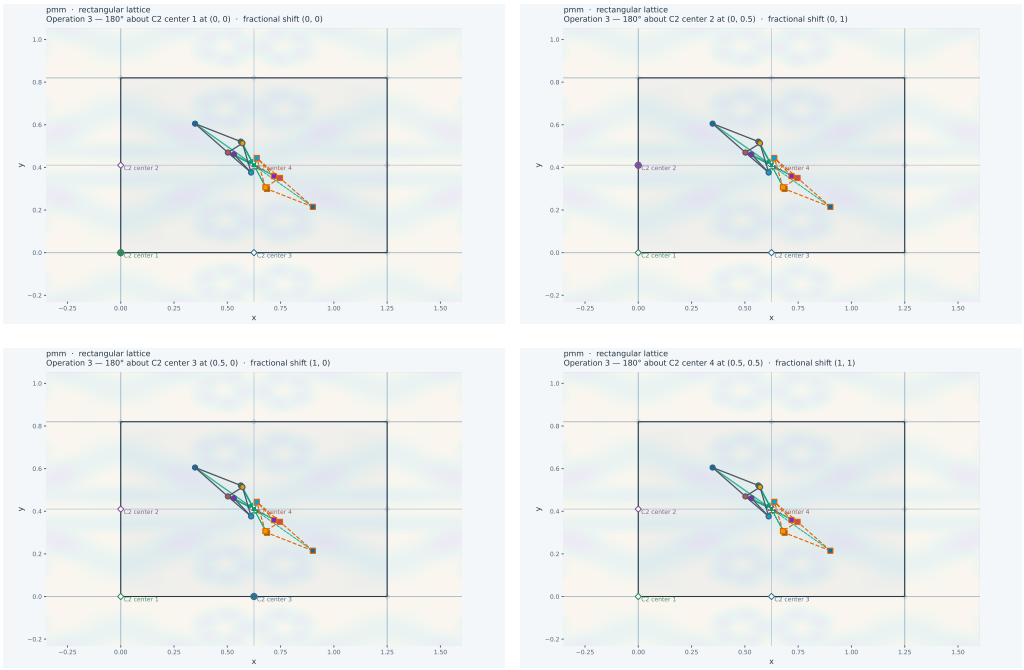


Table 10: Wallpaper group pmg, WA is pointing to C2-A at $q_1 = (\frac{1}{2}, \frac{1}{2}),\ q_2 = (0, \frac{1}{2})$

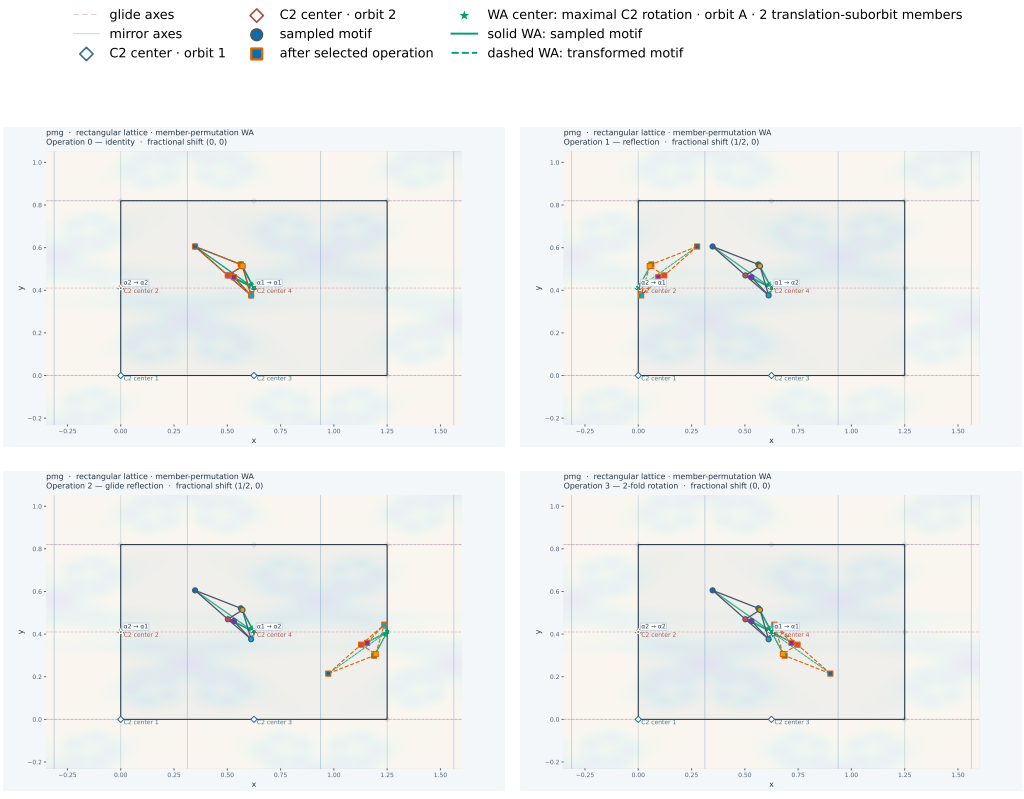


Table 10: Wallpaper group pmg (continued)

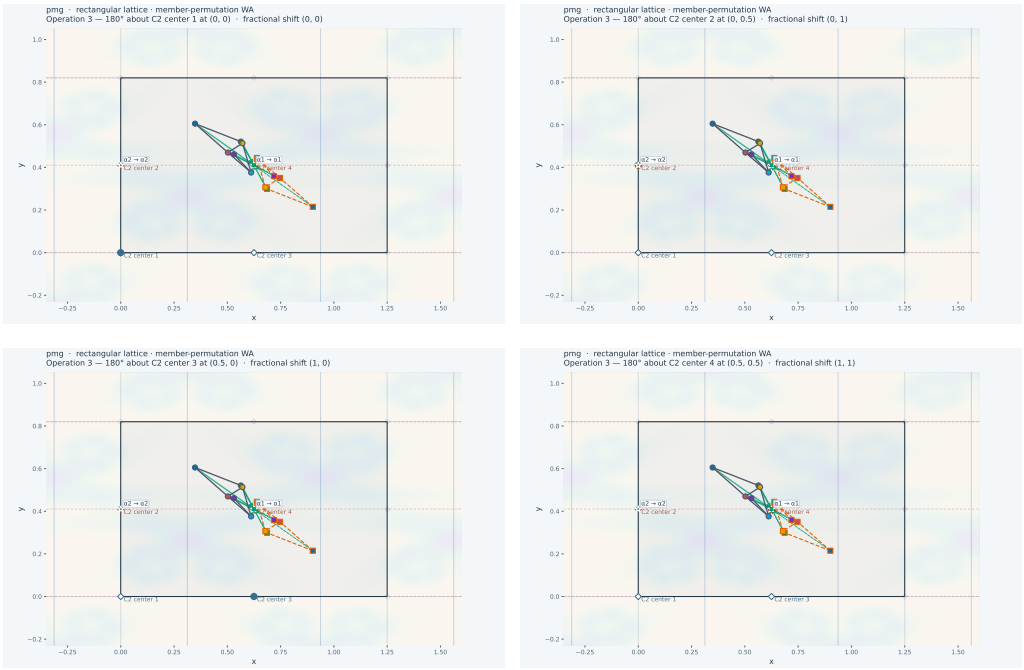


Table 11: Wallpaper group pgg, WA is pointing to C2-A at $q_1 = (\frac{1}{2}, \frac{1}{2}),\ q_2 = (0, 0)$

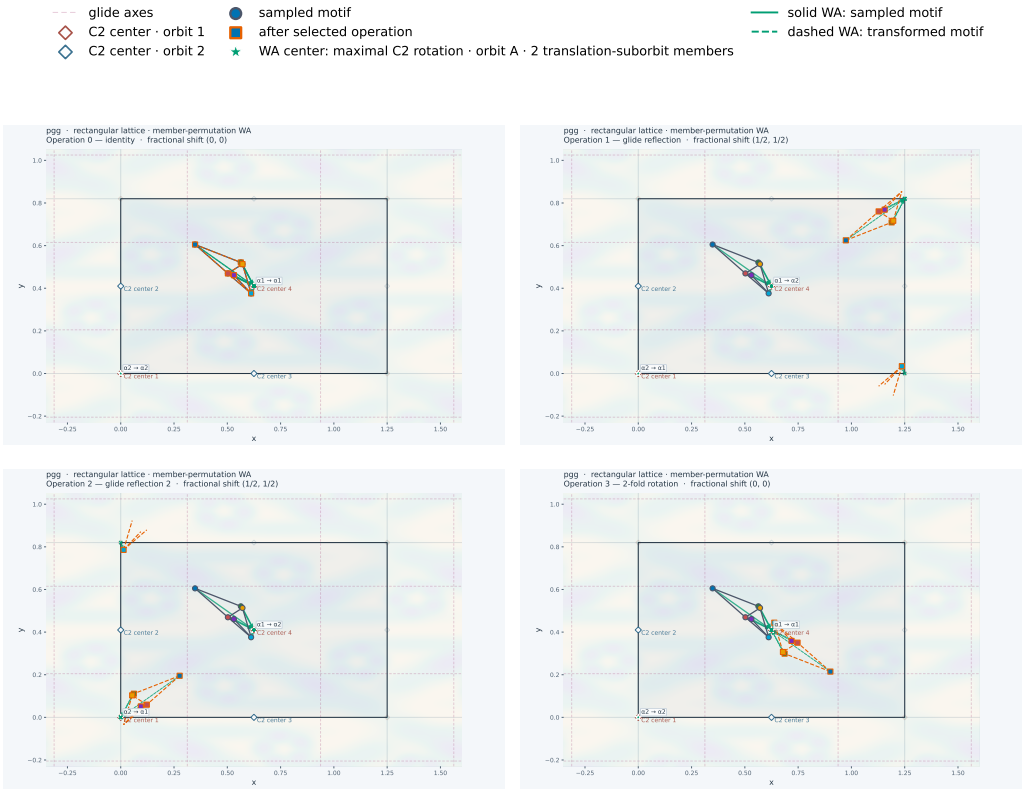


Table 11: Wallpaper group `pgg` (continued)

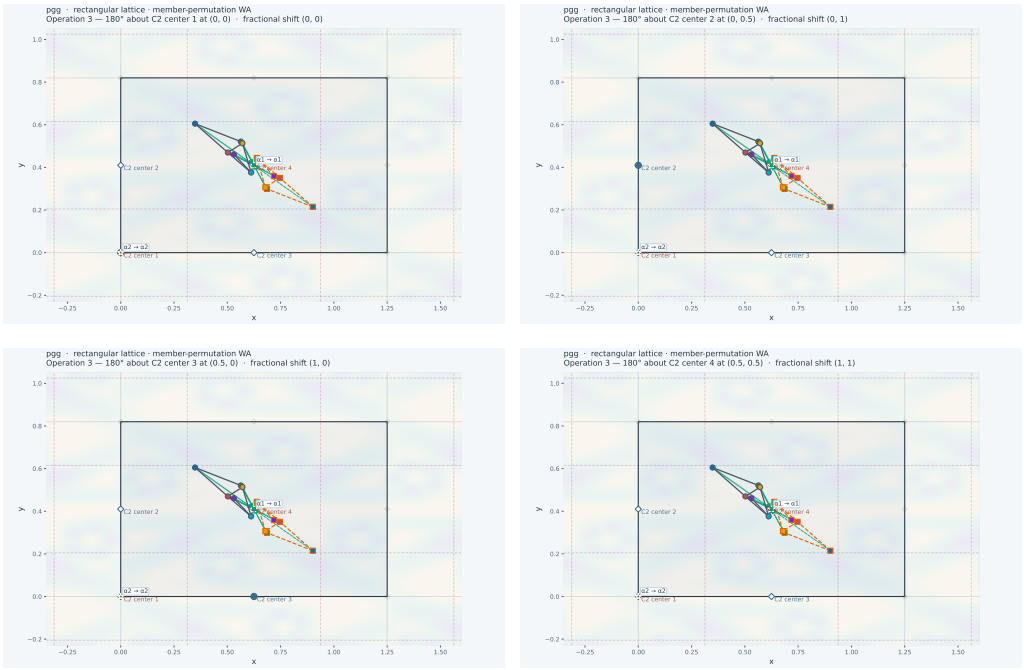


Table 12: Wallpaper group `cmm`, WA is pointing to C2-A at $q_1 = (0, 0)$, centered conventional cell shown, with the WA member multiplicity counted modulo primitive translations

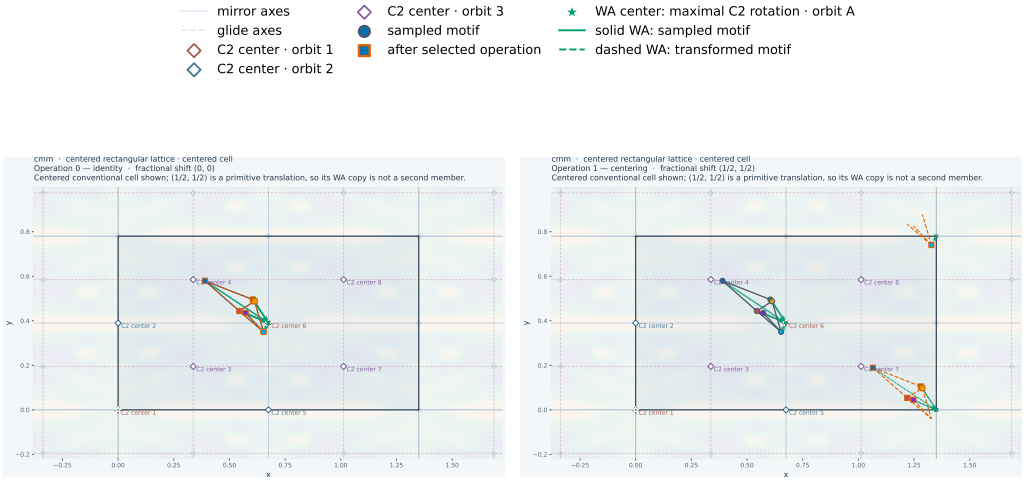


Table 12: Wallpaper group cmm (continued)

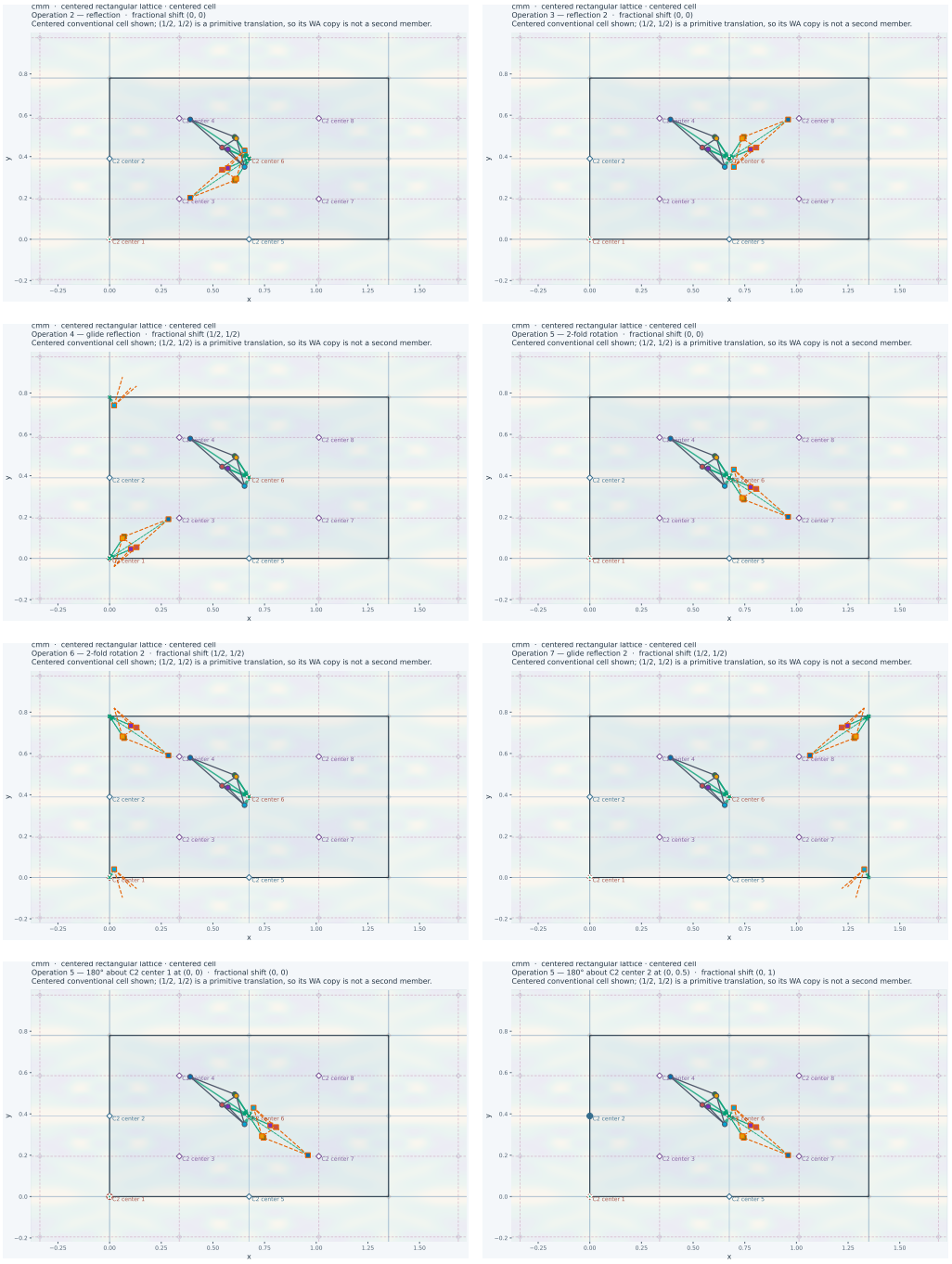


Table 12: Wallpaper group `cmm` (continued)

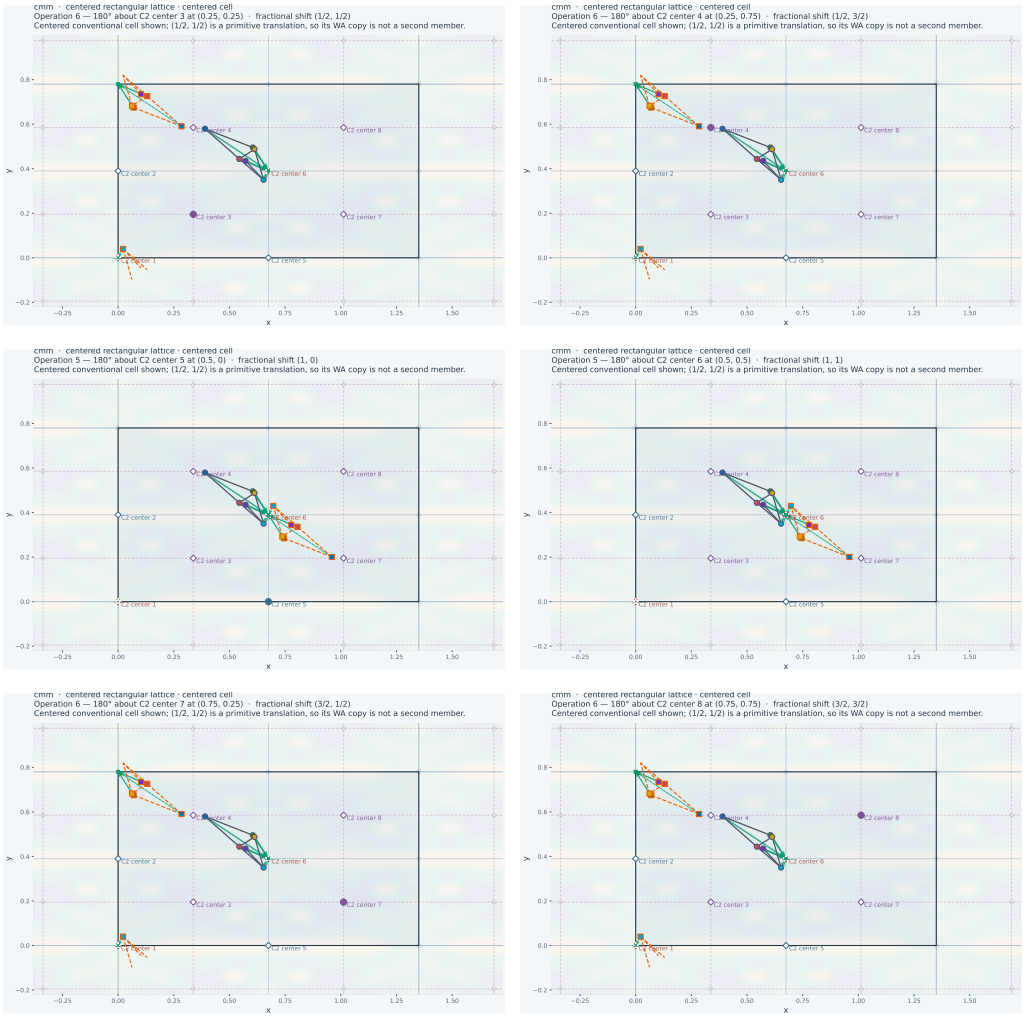


Table 13: Wallpaper group `p4`, WA is pointing to C4-A at $q_1 = (\frac{1}{2}, \frac{1}{2})$

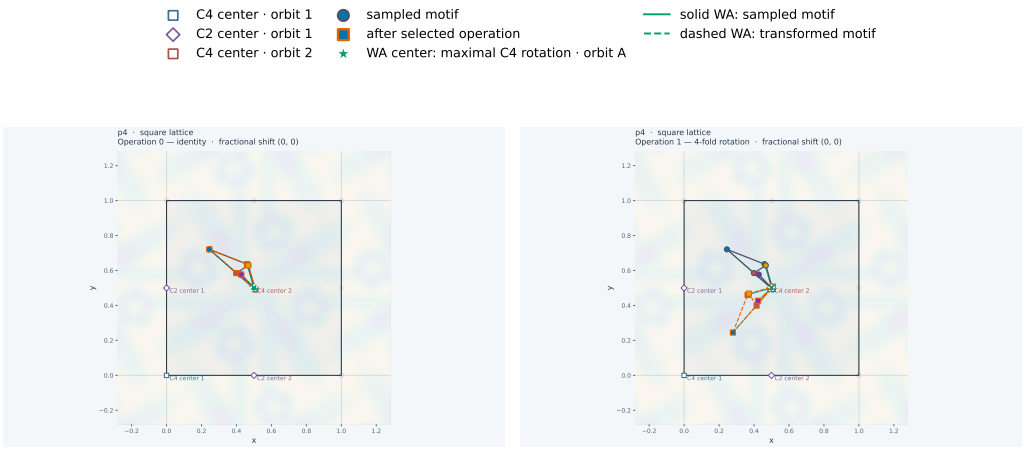
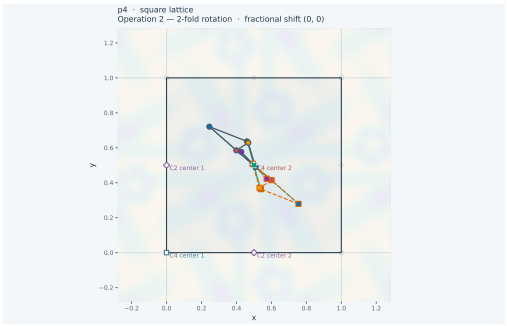
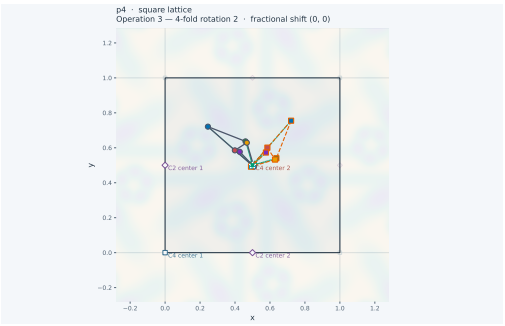
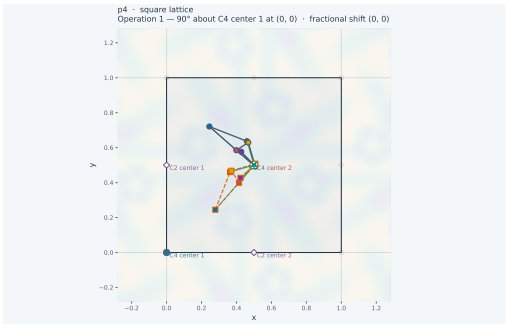
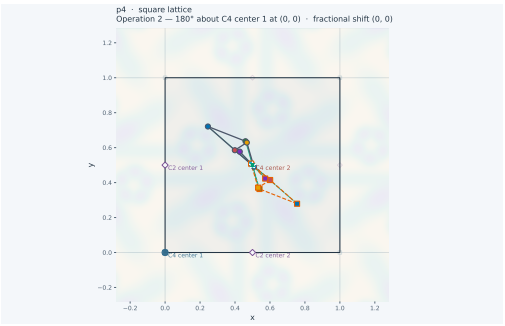


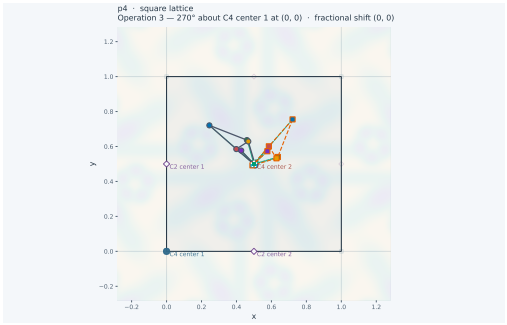
Table 13: Wallpaper group `p4` (continued)

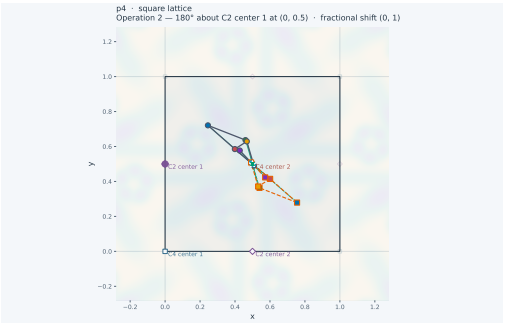


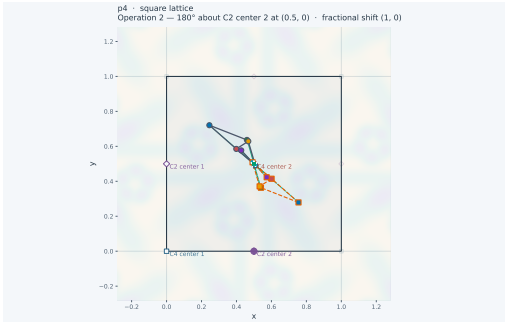












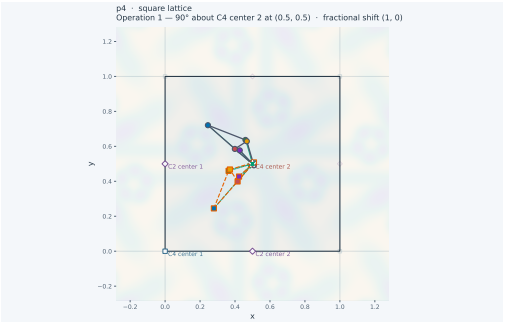


Table 13: Wallpaper group `p4` (continued)

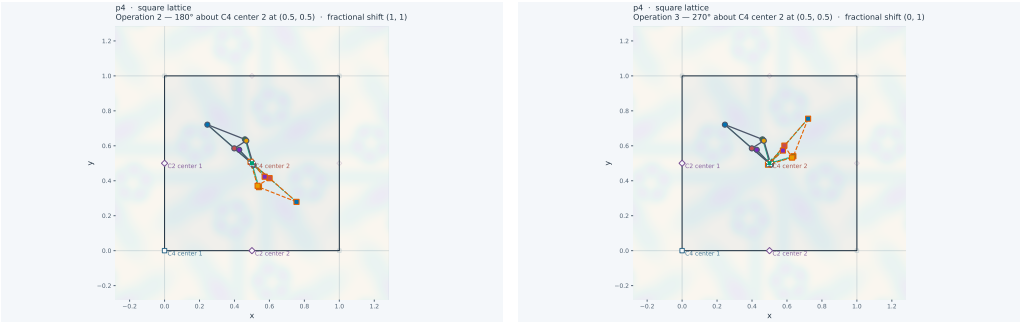


Table 14: Wallpaper group `p4m`, WA is pointing to C4-A at $q_1 = (\frac{1}{2}, \frac{1}{2})$

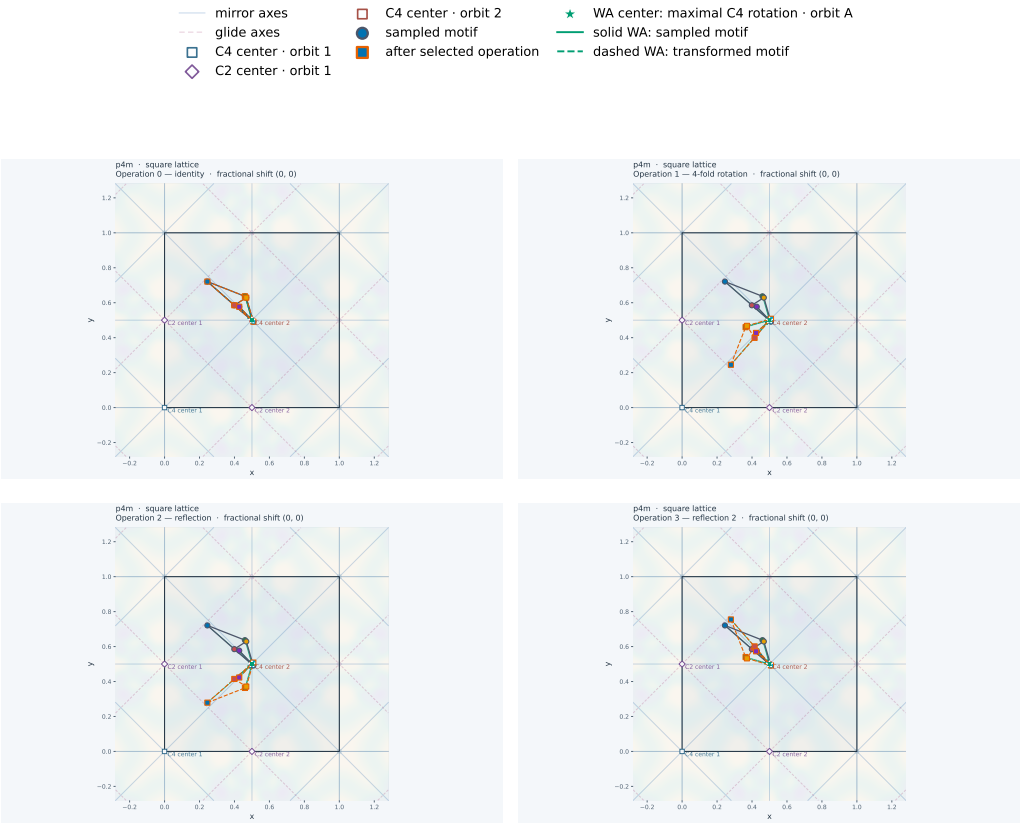


Table 14: Wallpaper group p4m (continued)

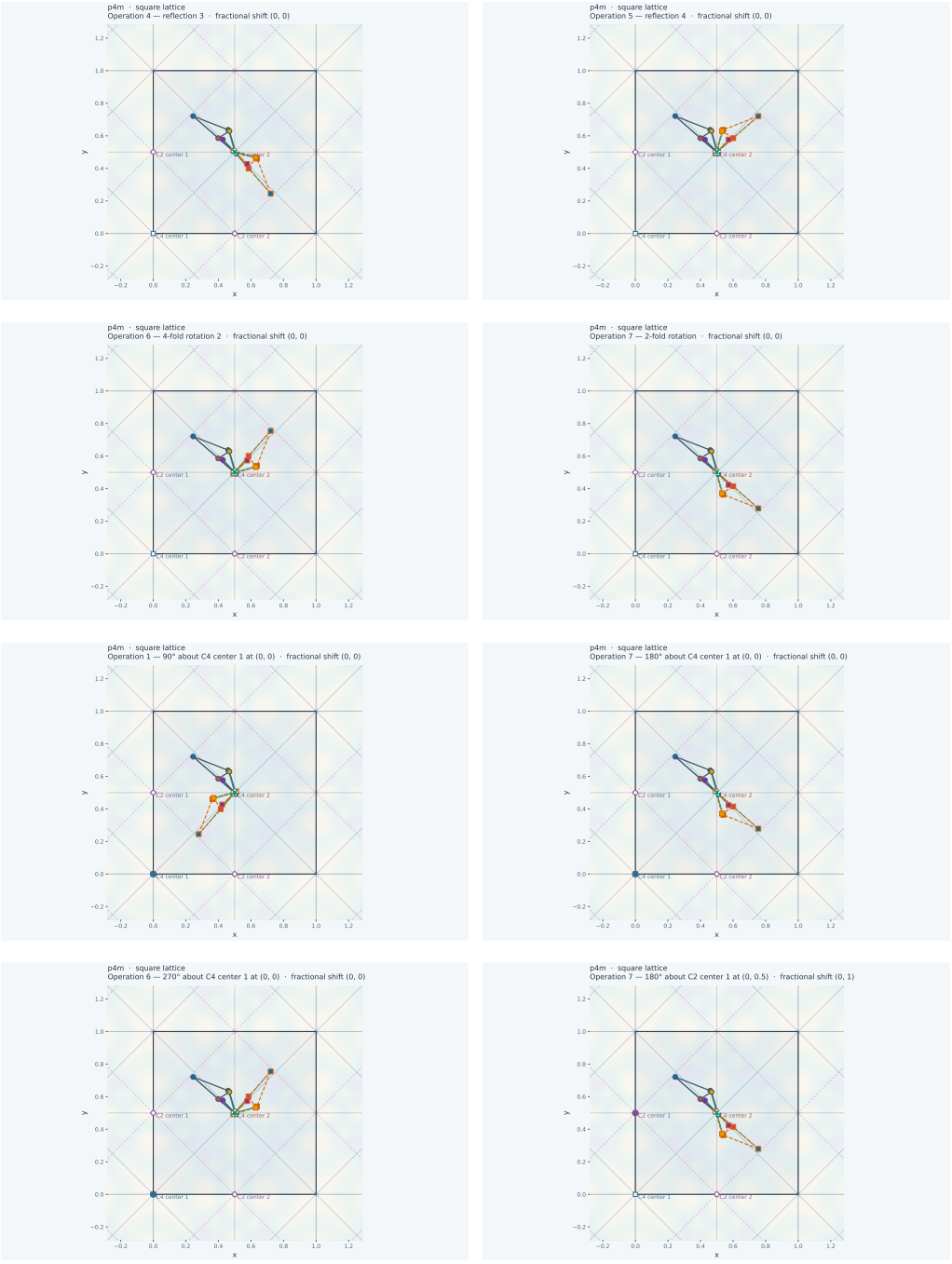


Table 14: Wallpaper group `p4m` (continued)

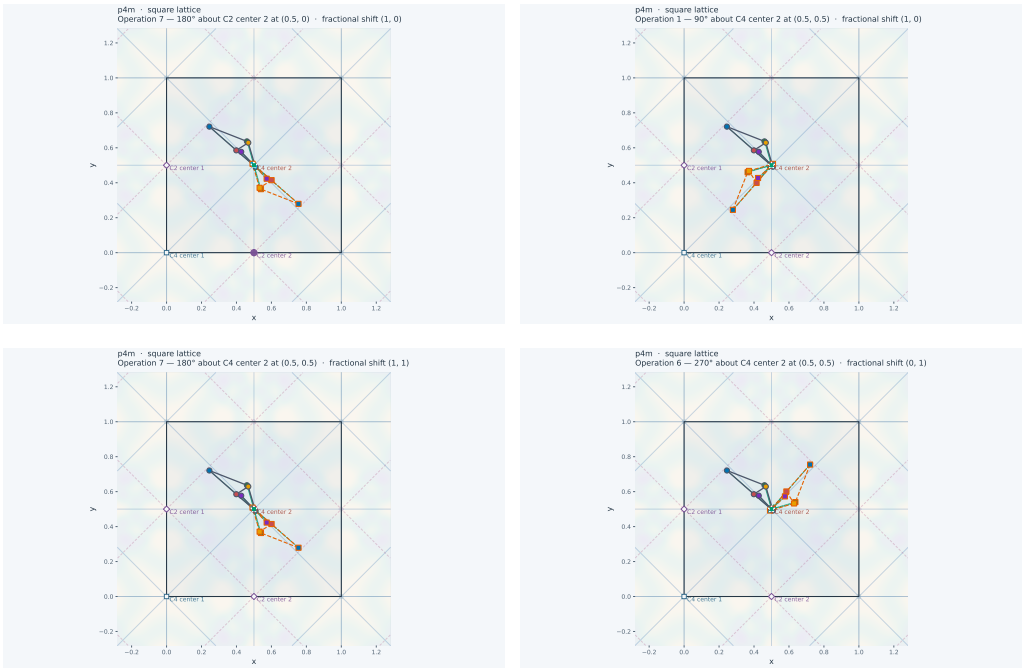


Table 15: Wallpaper group `p4g`, WA is pointing to C4 at $q_1 = (\frac{1}{2}, \frac{1}{2}),\ q_2 = (0, 0)$

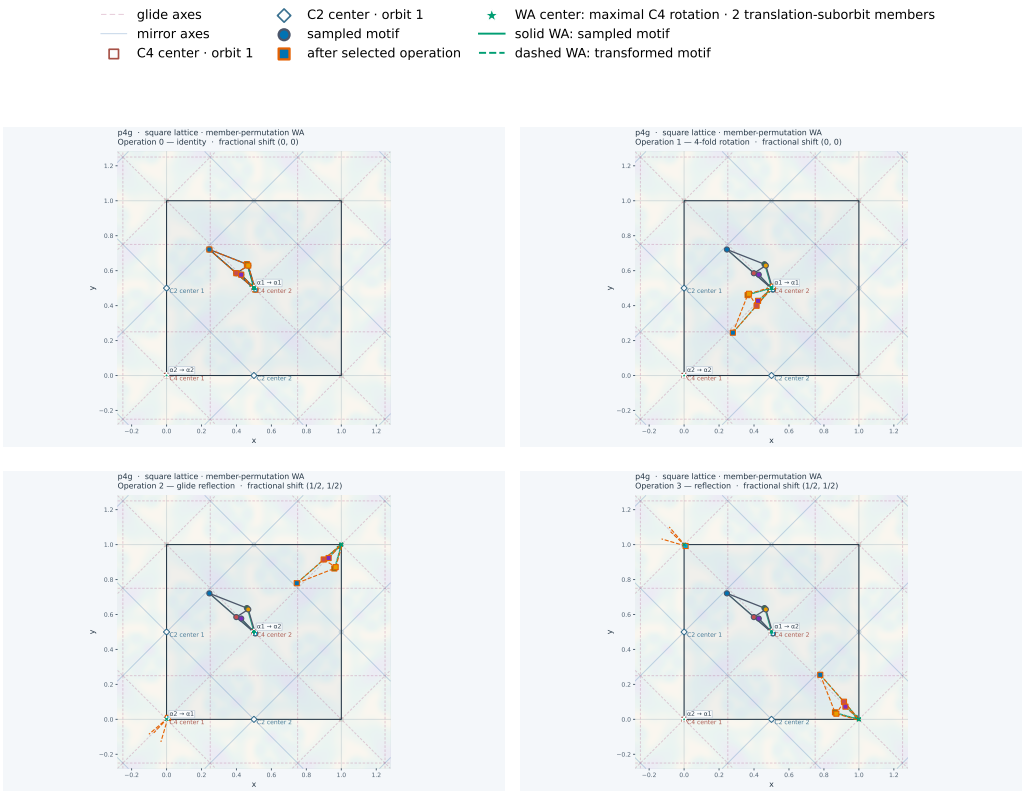


Table 15: Wallpaper group p4g (continued)

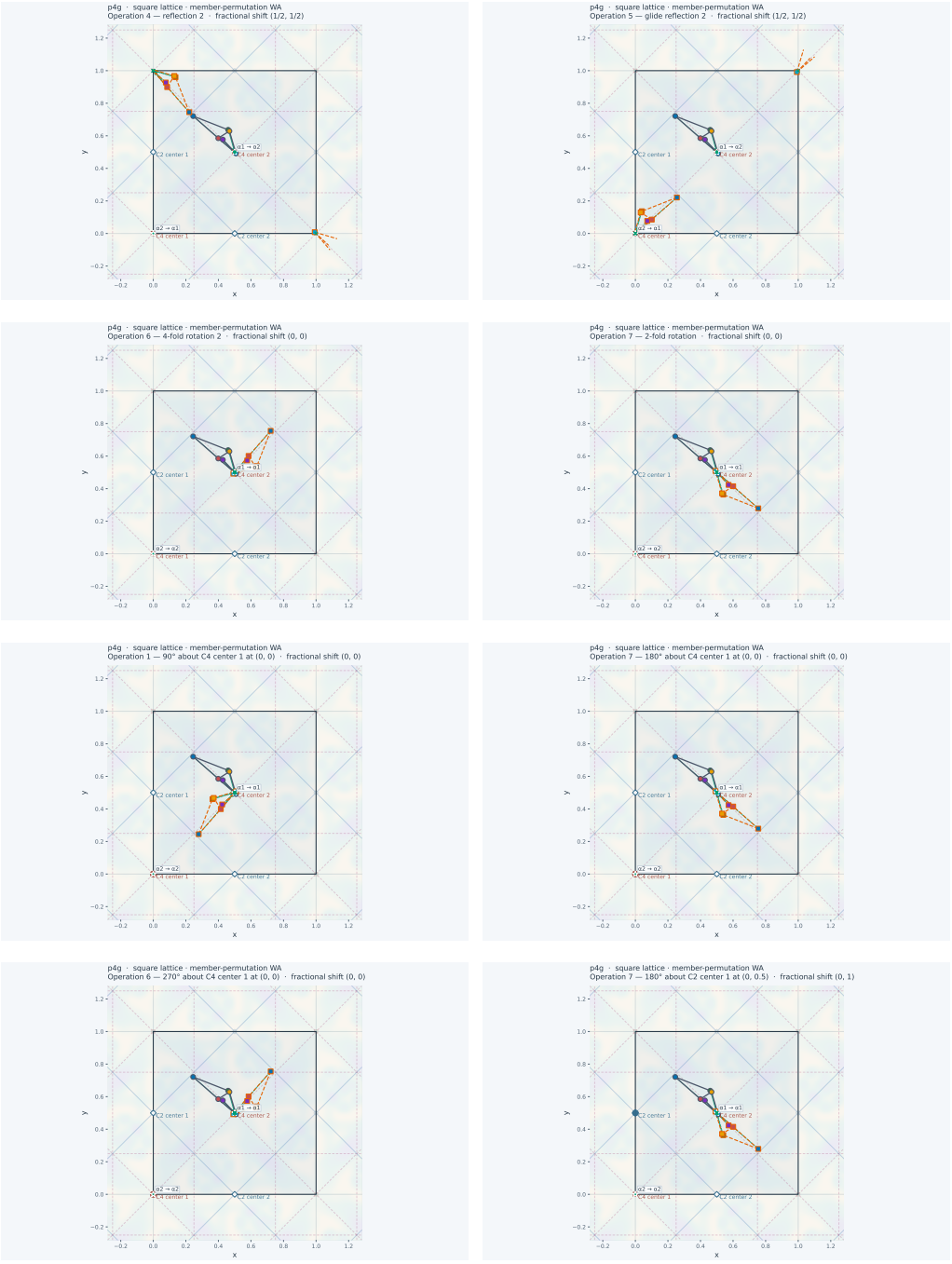


Table 15: Wallpaper group p4g (continued)

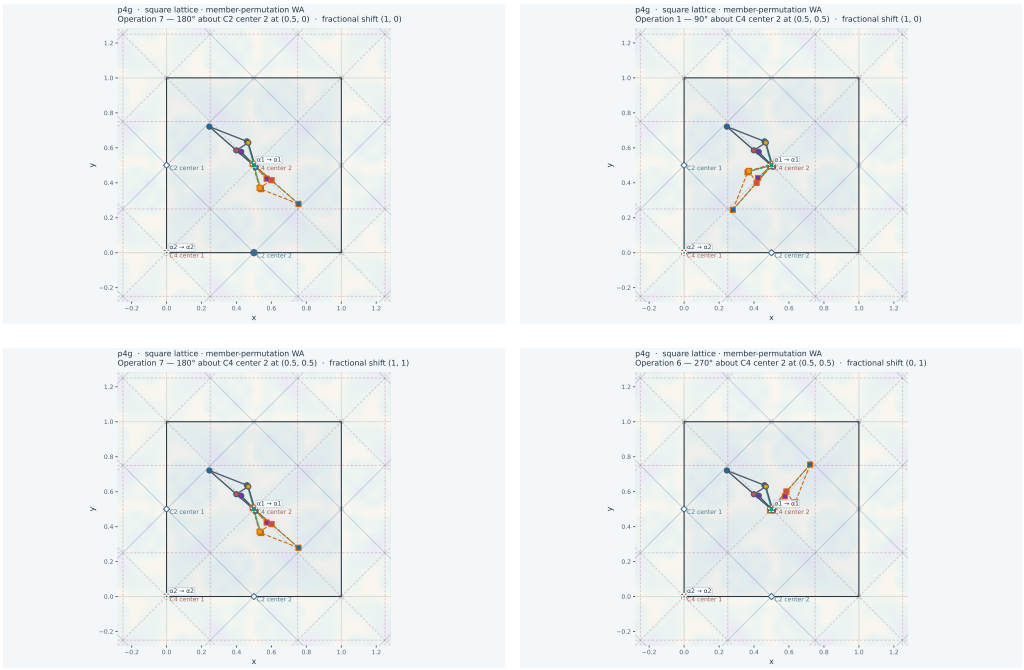


Table 16: Wallpaper group p3, WA is pointing to C3-A at $q_1 = (\frac{1}{3}, \frac{2}{3})$

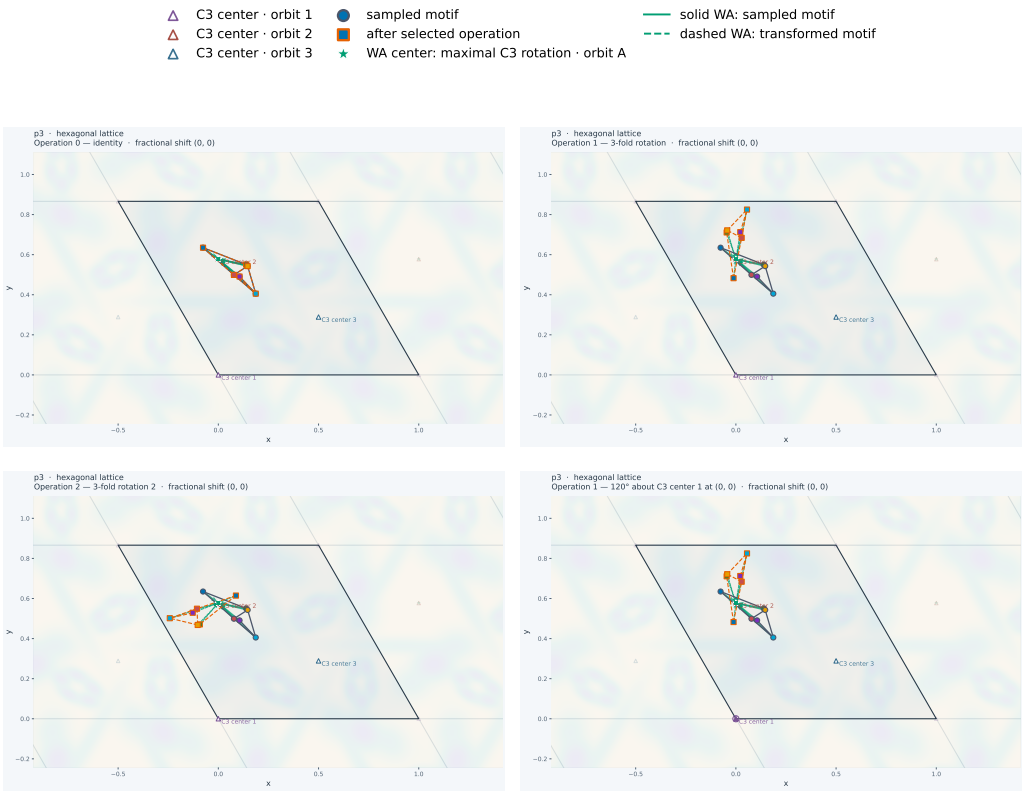
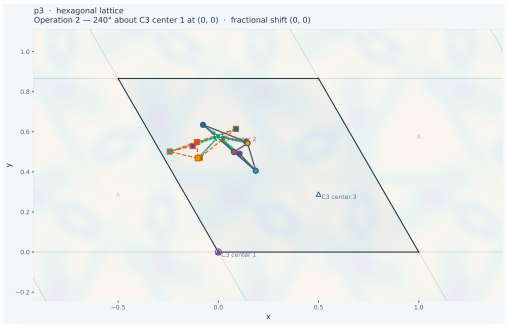
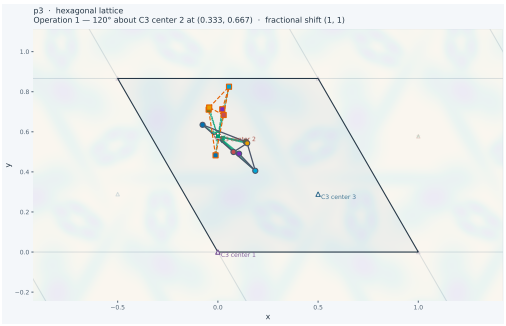
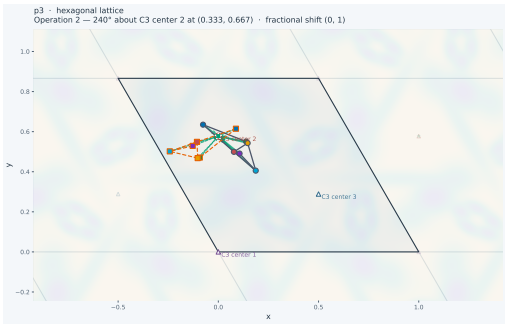
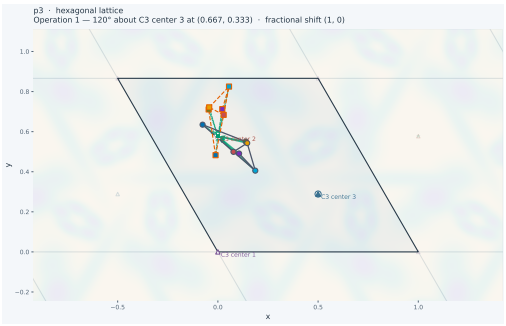


Table 16: Wallpaper group p3 (continued)









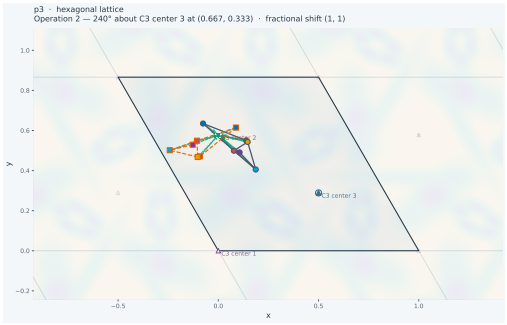


Table 17: Wallpaper group p3m1, WA is pointing to C3-A at $q_1 = (\frac{1}{3}, \frac{2}{3})$

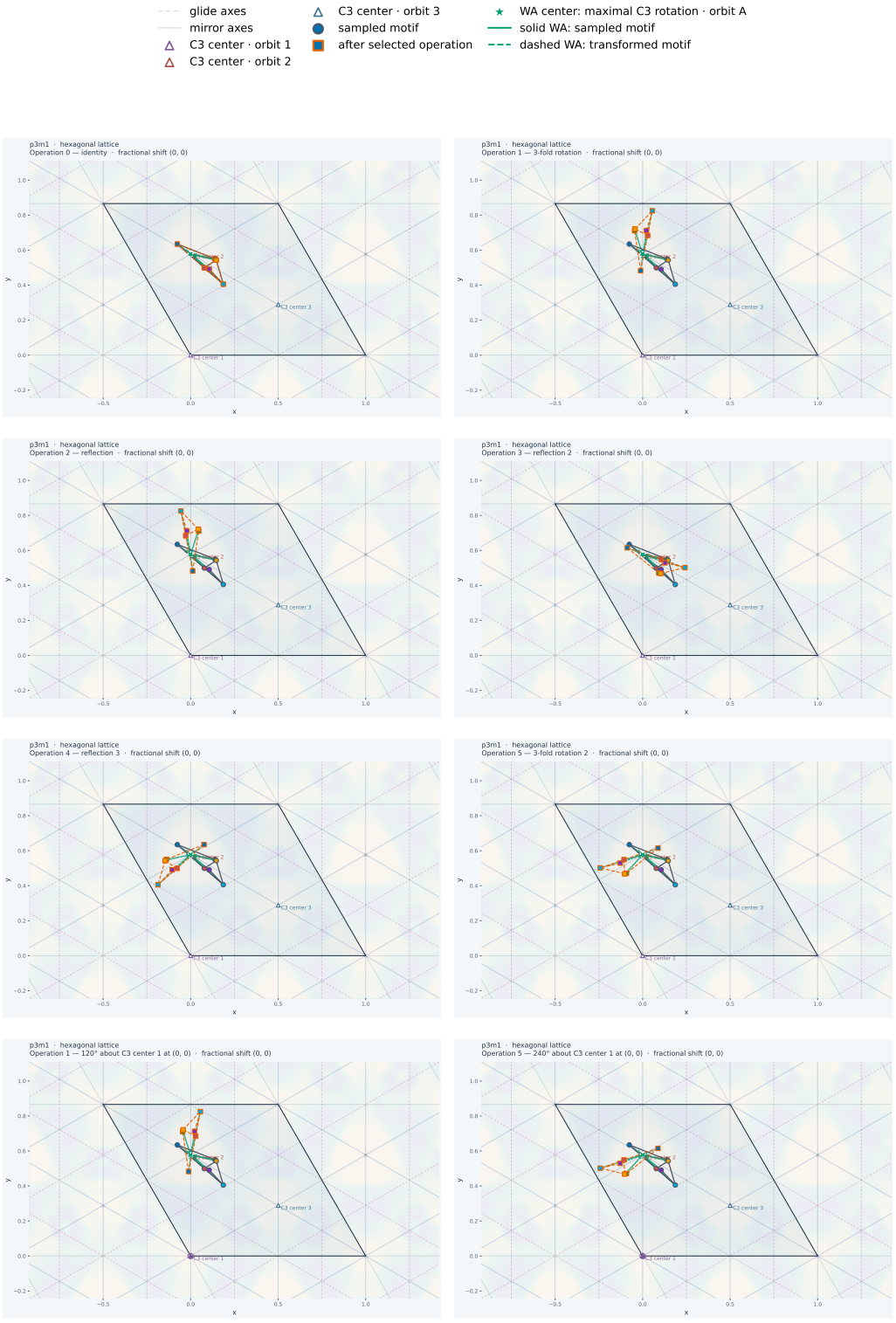


Table 17: Wallpaper group `p3m1` (continued)

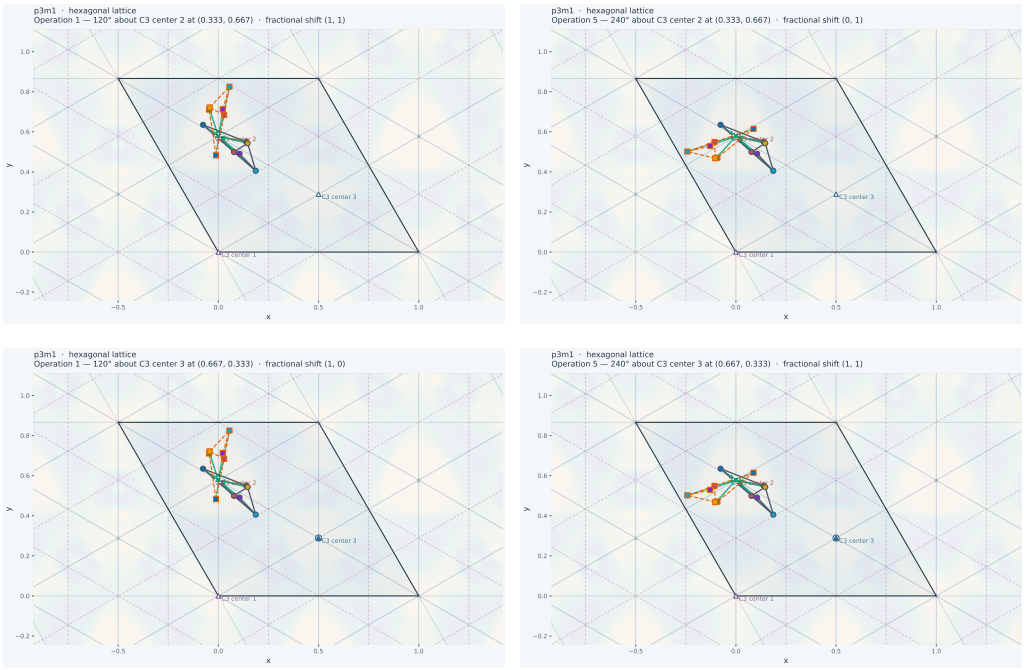


Table 18: Wallpaper group `p31m`, WA is pointing to C3-B at $q_1 = (0, 0)$

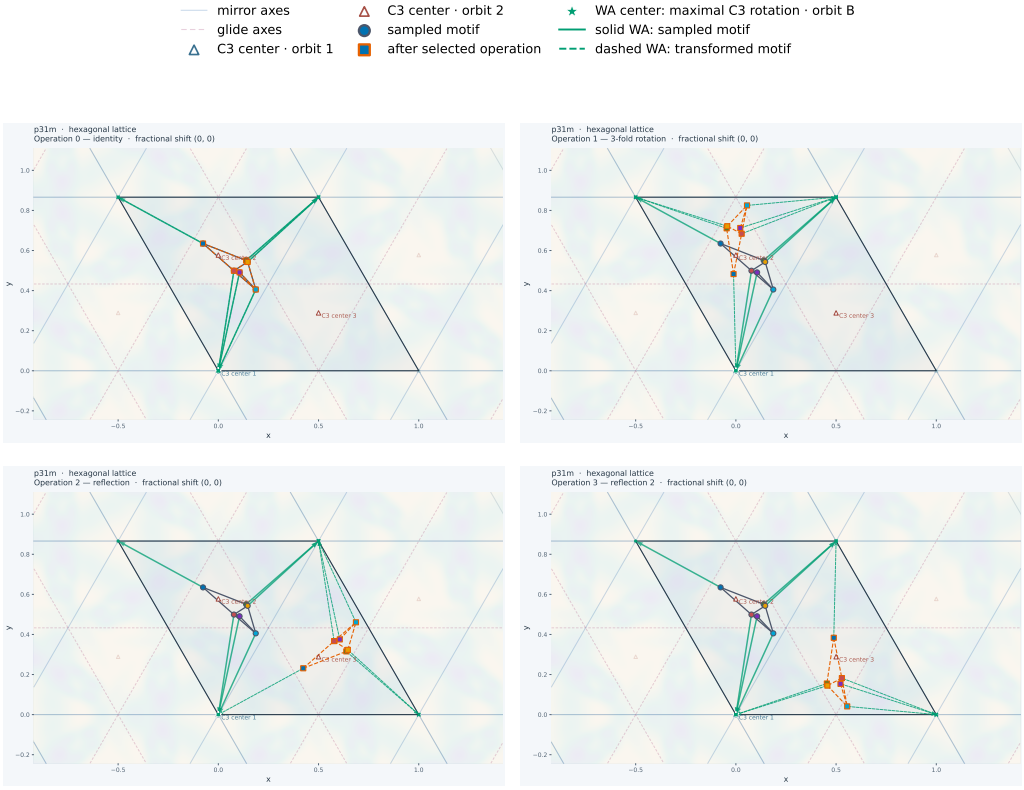


Table 18: Wallpaper group `p31m` (continued)

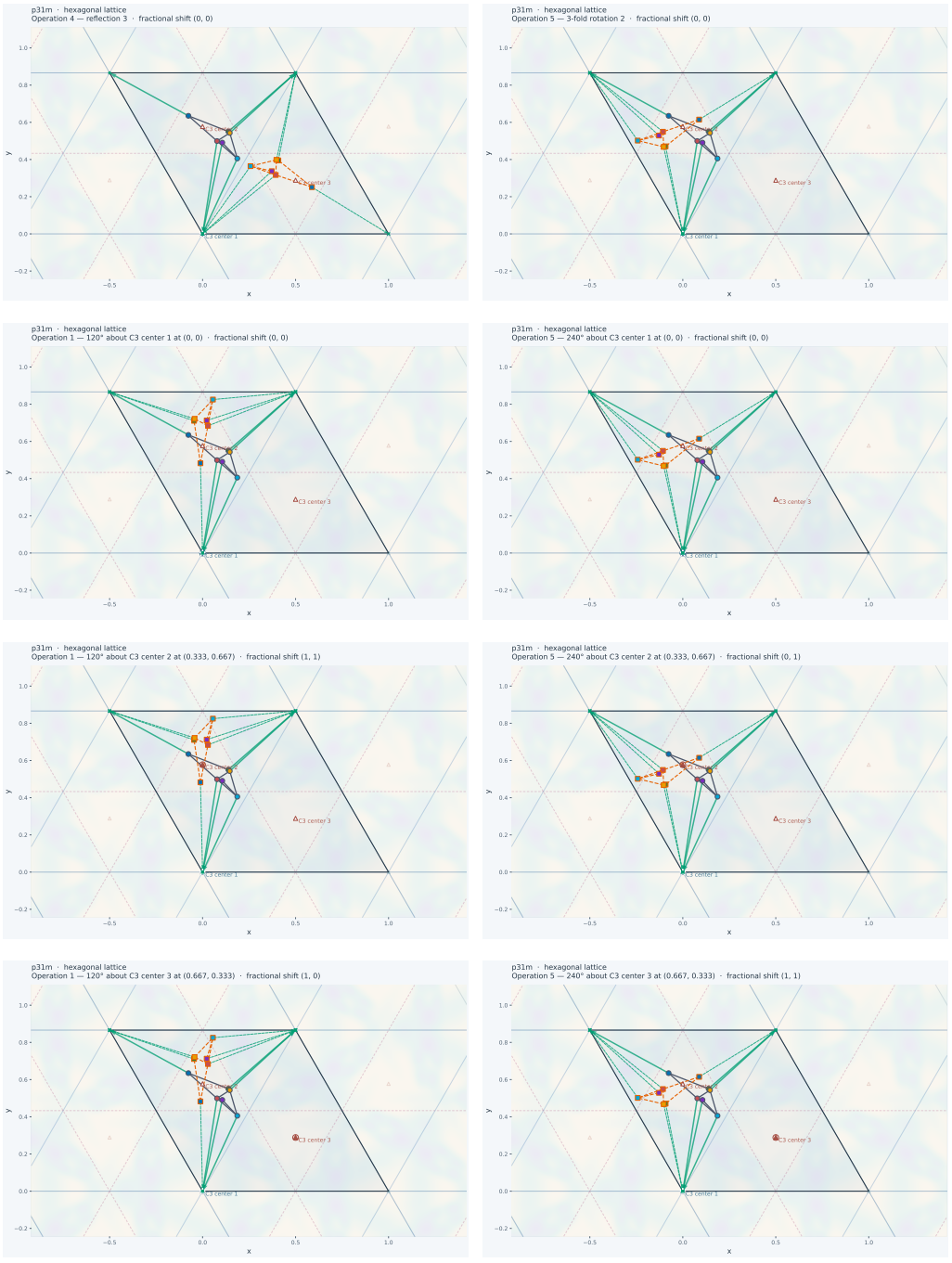


Table 19: Wallpaper group p6, WA is pointing to C6 at $q_1 = (0, 0)$

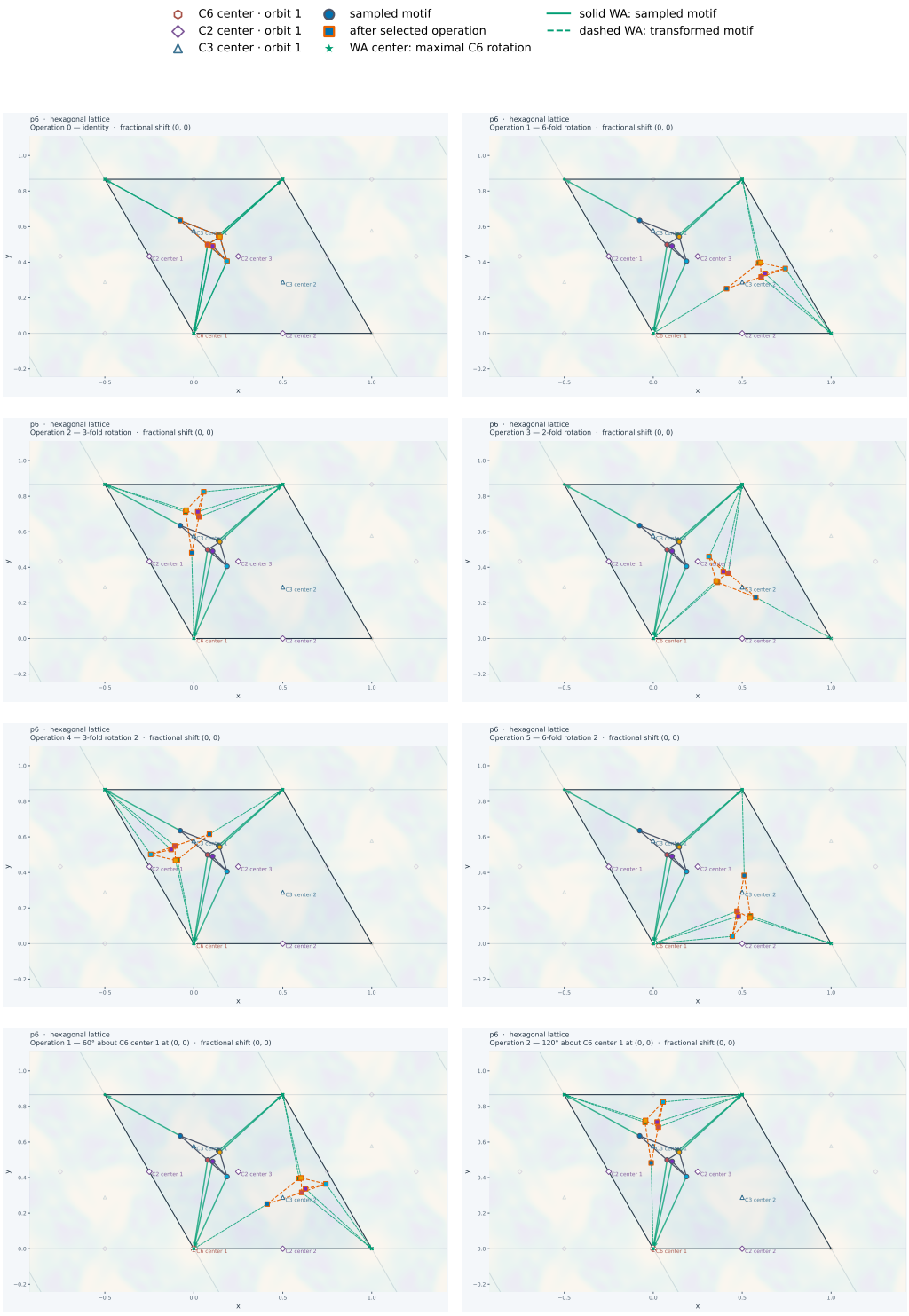
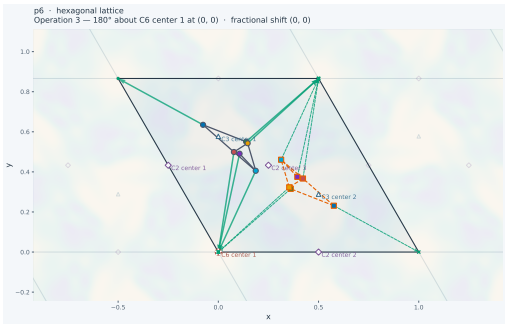
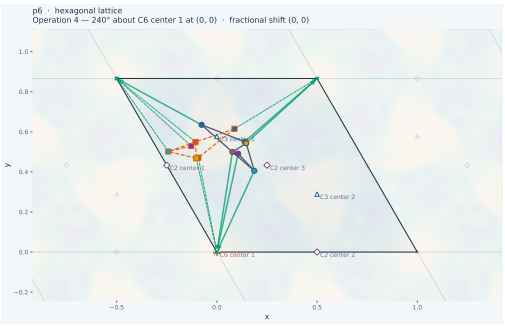
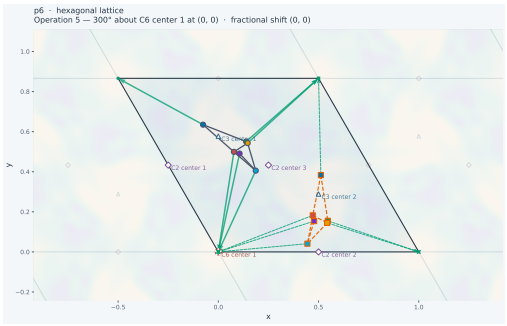
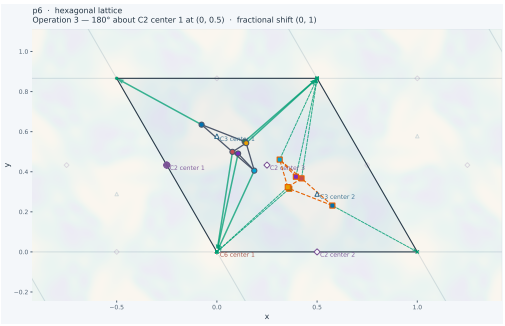


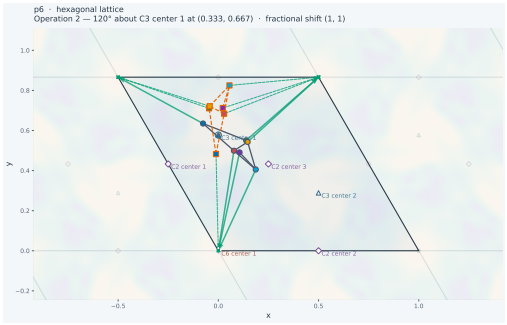
Table 19: Wallpaper group p6 (continued)

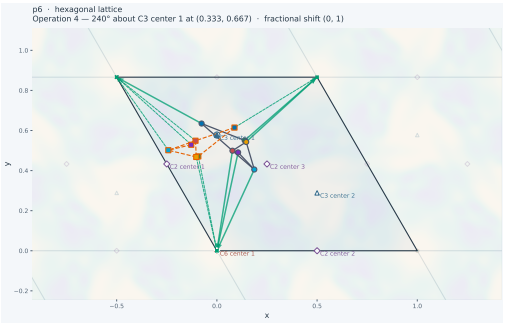


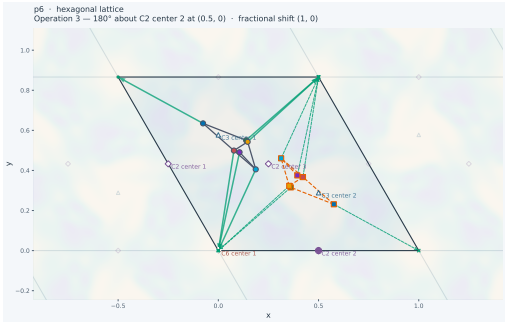












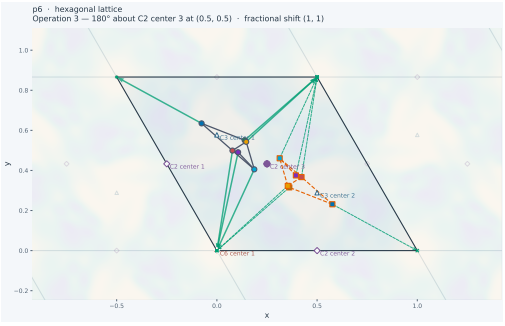


Table 19: Wallpaper group p6 (continued)

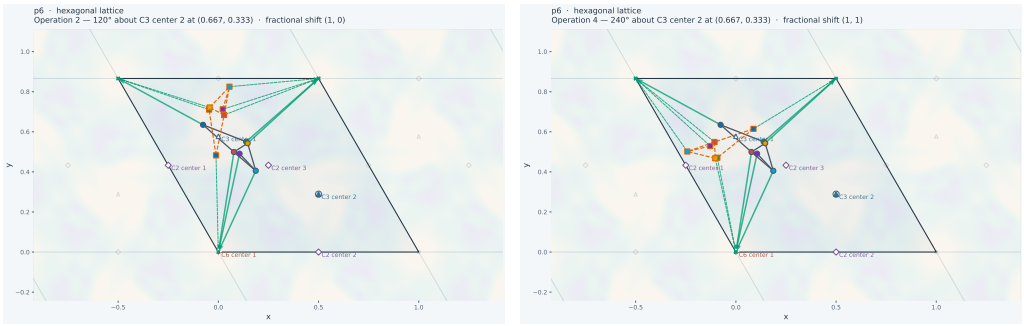


Table 20: Wallpaper group p6m, WA is pointing to C6 at $q_1 = (0, 0)$

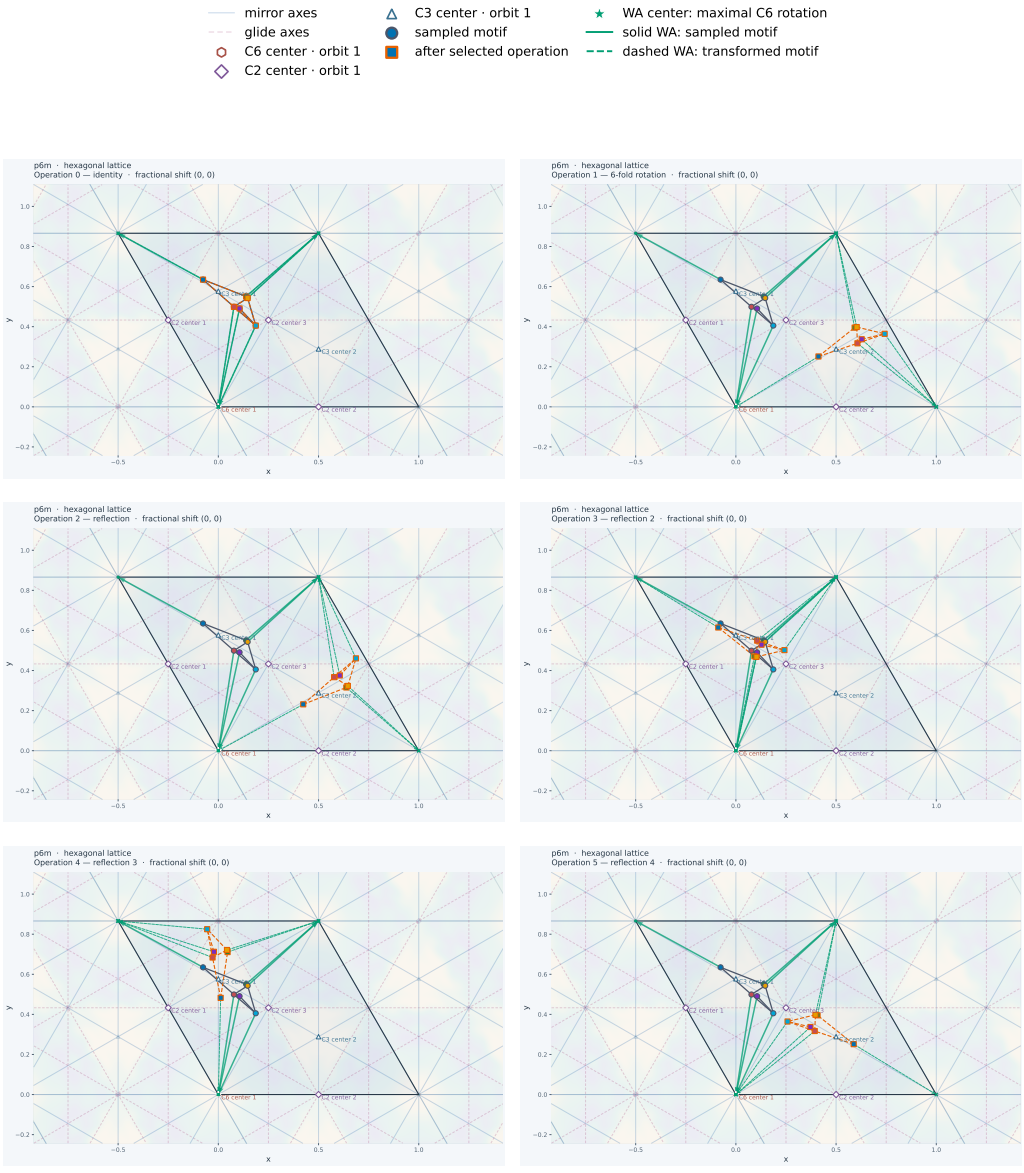


Table 20: Wallpaper group p6m (continued)

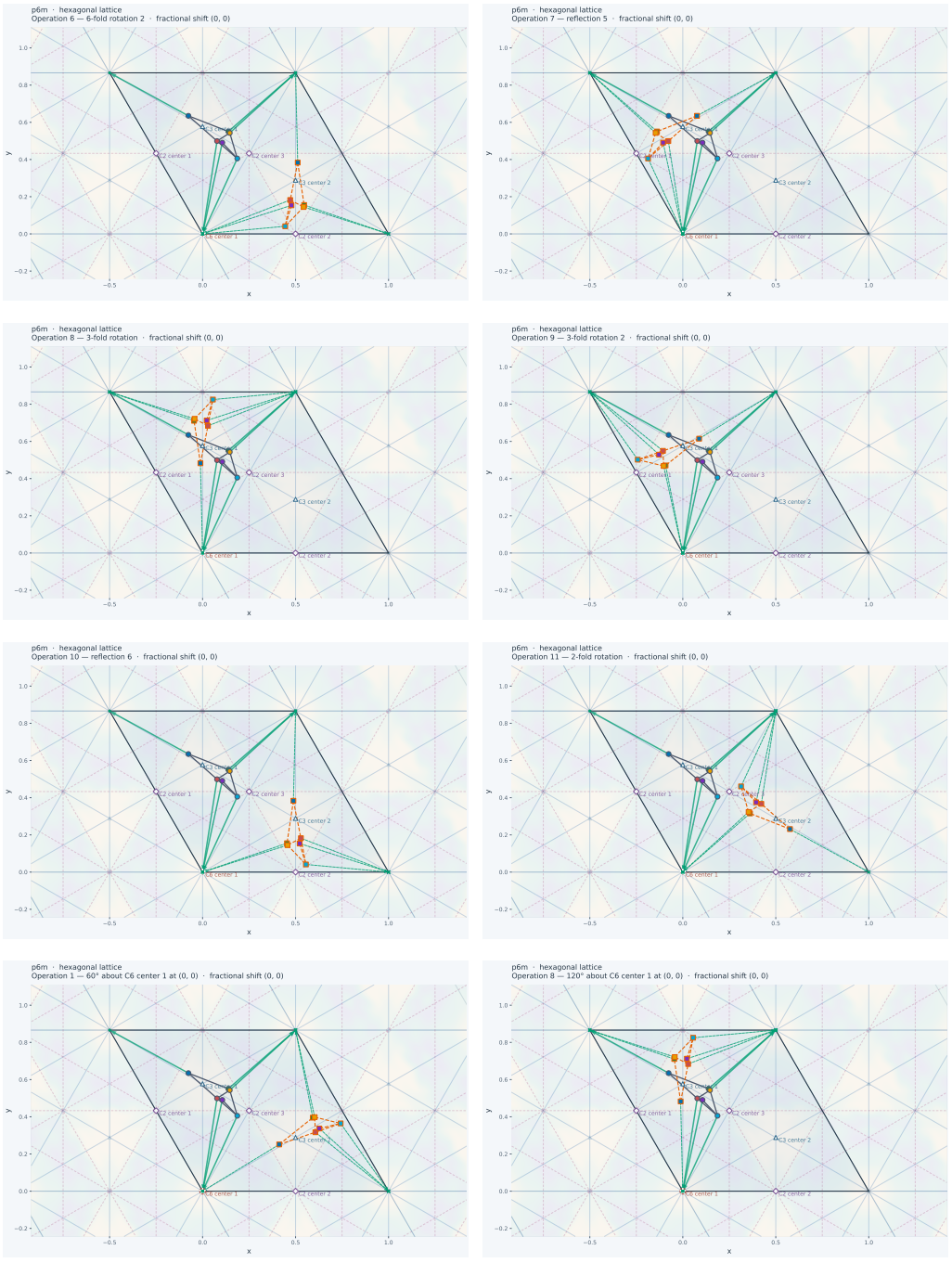


Table 20: Wallpaper group p6m (continued)

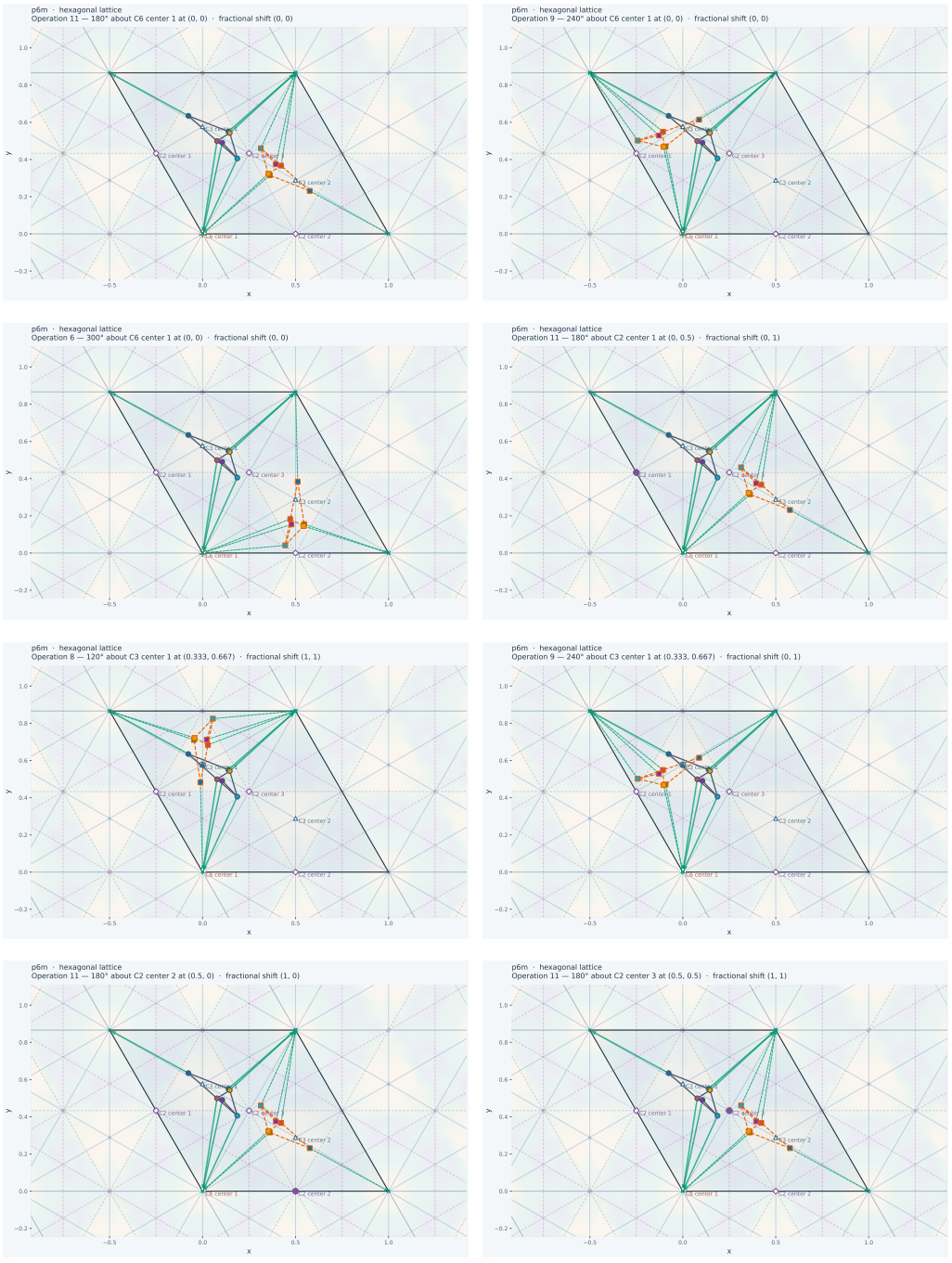
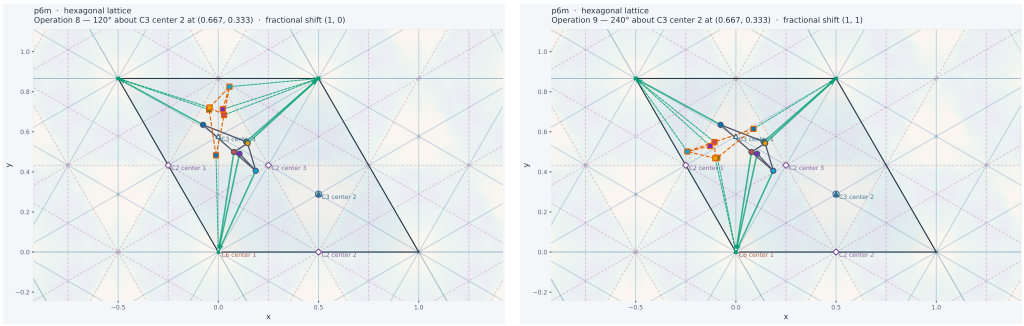


Table 20: Wallpaper group p6m (continued)



D WYCKOFF-ANCHOR IDENTIFIABILITY ACROSS ALL WALLPAPER GROUPS

Wallpaper-equivariant message passing based only on relative edge geometry cannot, in general, identify the absolute registry of a graph within a periodic cell. We therefore construct a controlled experiment to isolate the information supplied by the Wyckoff-anchor, to numerically show its importance.

The experiment compares two parameter-matched models. The *WA-aware* model receives the graph, its relative edge vectors, and the WA encoding. The *WA-free* control uses exactly the same equivariant architecture and representation types, but every WA vector and, for nonsymmorphic groups, every translation-suborbit member field is replaced by zero. The comparison therefore isolates the information supplied by the WA representation rather than differences in architecture or parameter count.

This experiment is an identifiability test rather than an atomistic simulation. The graph geometry is fixed, the only variation between samples is its position within the periodic cell, and the prediction target is a known analytic wallpaper-periodic scalar field.

Fixed graph and registry sampling. All 17 wallpaper groups are considered. Each sample contains the same asymmetric four-node graph embedded in two dimensions. We treat fractional and Cartesian positions as column vectors. In fractional coordinates, the uncentered node offsets are

$$p_1 = (-0.040, -0.025), \quad p_2 = (0.038, -0.018),$$
$$p_3 = (0.018, 0.043), \qquad p_4 = (-0.016, 0.011),$$

and their mean is subtracted to obtain centered offsets q_i.

For a sampled graph center c, node i is positioned at

$$f_i = c + q_i, \qquad x_i = C_G f_i,$$

where the columns of $B_G \in \mathbb{R}^{2\times 2}$ are the conventional-cell lattice vectors for wallpaper group G, and $C_G = 5B_G$ is the cell used in the experiment. The scaling changes the Cartesian length scale but not the fractional symmetry operations.

The graph center is restricted to positions for which all four nodes remain inside the central cell, so no coordinate wrapping occurs. The graph is complete and directed without self-edges, giving 12 directed edges. All nodes use the same learned initial embedding and therefore carry no distinguishing node-type information.

Samples differ only by rigid translation of the complete graph. Consequently,

$$r_{ij} = x_i - x_j$$

is identical for every registry position. A graph network that receives only these relative edges therefore obtains exactly the same geometric input for configurations that can nevertheless have different target values. This creates the input collision that the experiment is designed to test.

Each wallpaper group contains 320 training registries and 256 independently sampled test registries. The embedded graph itself is identical in both splits, so the experiment measures generalization to unseen positions within the wallpaper cell rather than generalization to unseen graph geometries.

Analytic wallpaper-periodic target. We define a known analytic scalar target over the fractional coordinates of the graph. Starting from an integer reciprocal mode $k = (k_u, k_v)$, the seed field is

$$h_k(f) = \cos\big(2\pi k^{\mathsf{T}} f\big) + 0.37 \sin\big(2\pi k^{\mathsf{T}} f\big) .$$

Thus, before symmetrization, the target is a simple sinusoidal Fourier field over position in the periodic cell.

Let $\Lambda_G^{\mathrm{cell}} = C_G \mathbb{Z}^2$ denote the displayed conventional-cell translation lattice and let $\bar{G} = G/\Lambda_G^{\mathrm{cell}}$ be the corresponding finite affine quotient. For each representative $g \in \bar{G}$, the action in fractional coordinates is

$$g(f) = S_g f + s_g \pmod{\mathbb{Z}^2}, \qquad g \in \bar{G},$$

where $S_g = C_G^{-1} Q_g C_G$ is the fractional-coordinate linear part and s_g is the fractional translation. This quotient is taken modulo the displayed conventional-cell translations, rather than the primitive lattice Λ used in Section 2. It therefore also retains any centering translation inside the displayed cell. The corresponding wallpaper-invariant node field is obtained by a finite Reynolds projection,

$$V_{G,k}(f) = \frac{1}{|\bar{G}|} \sum_{g \in \bar{G}} h_k(S_g f + s_g).$$

The scalar target for the complete graph is then obtained by averaging this field over its nodes,

$$E_G(X) = \frac{1}{N} \sum_{i=1}^{N} V_{G,k_G}(f_i), \qquad N = 4.$$

By closure of the affine quotient and periodicity of h_k,

$$E_G(gX) = E_G(X), \qquad g \in G,$$

so the resulting target is invariant under the prescribed wallpaper group. It is also invariant to permutations of the graph nodes because the node contributions are averaged.

The reciprocal mode k_G is chosen as a low-order mode that survives the corresponding wallpaper-group projection. We use $k_G = (1, 0)$ for $p1, p2, pm, pmm, p4, p4m, p3, p3m1, p31m, p6$, and $p6m$. For cm, cmm, pg, and pmg, we use $k_G = (1, 1)$, and for pgg and $p4g$, we use $k_G = (2, 1)$.

In addition to the scalar target, we define an equivariant vector target from its exact coordinate gradient,

$$F_i = -\nabla_{x_i} E_G(X).$$

These vector targets are therefore deterministic derivatives of the same analytic scalar field rather than independently generated labels. The learned model is treated analogously: its predicted vector field is obtained by differentiating its predicted scalar output, and no separate direct vector head is used.

Repeated target landscapes across wallpaper groups. The analytic target is constructed to be invariant under a prescribed wallpaper group, but it is not intended to uniquely identify that group. In particular, Reynolds projection guarantees (G)-invariance but does not require (G) to be the full symmetry group of the resulting scalar field. Because we use a single low-order reciprocal-lattice mode, the selected mode can already possess additional symmetries, and the projection for different wallpaper groups can therefore produce identical or visually equivalent periodic landscapes. This behavior is expected and does not affect the purpose of the experiment. The comparison is performed independently within each wallpaper group and tests whether the network can identify the registry of a rigidly translated graph. It does not test whether the analytic target itself distinguishes the 17 wallpaper groups.

Equivariant model. Both variants use the same single-message-passing-layer, finite-point-group-equivariant `WELinear` architecture. Every node starts from the same learned 24-dimensional scalar embedding. For each directed edge $j \to i$, the WA-aware input contains the receiver and sender scalar embeddings, the squared edge distance, the planar relative edge vector, the planar WA vector, and, for nonsymmorphic groups, the translation-suborbit member representation. Because the four-node graph is complete and directed without self-edges, each node receives three messages, giving 12 directed edge messages per configuration. The aggregated 24-dimensional message is concatenated

with the 24-dimensional node state and passed through a $48 \to 24 \to 24$ residual node MLP. The final node readout has widths $24 \to 24 \to 1$, and the four scalar node outputs are mean-pooled to obtain the graph-level prediction.

The typed edge representation is mapped by an equivariant linear layer to 16 copies of the regular representation of the corresponding finite point group. This is followed by a representation-compatible ELU nonlinearity and a second equivariant linear map to 24 trivial scalar channels,

$$z_{ij} \longrightarrow 16\,\rho_{\mathrm{reg}} \longrightarrow 16\,\rho_{\mathrm{reg}} \longrightarrow m_{ij} \in 24\,\rho_{\mathrm{triv}}.$$

The scalar edge messages are summed at the receiver and used in a residual node update,

$$h_i' = h_i + \mathrm{MLP}_{\mathrm{node}}\left([h_i, \sum_{j \neq i} m_{ij}]\right).$$

A scalar readout is subsequently applied to each node and mean-pooled over the graph.

The WA-free control uses the same architecture, representation types, widths, and number of trainable parameters within each wallpaper group. Its only modification is that WA features are replaced by 0s.

Optimization and evaluation. Only training-set statistics are used to normalize the scalar target,

$$\widetilde{E} = \frac{E - \mu_E}{\sigma_E}, \qquad \widetilde{F} = \frac{F}{\sigma_E}.$$

The force normalization follows directly from differentiating the normalized energy. In particular,

$$\widehat{\widetilde{F}}_i = -\nabla_{x_i} \widehat{\widetilde{E}}, \qquad \widehat{F}_i = \sigma_E \widehat{\widetilde{F}}_i.$$

The training objective is

$$\mathcal{L} = \mathrm{MSE}(\widehat{\widetilde{E}}, \widetilde{E}) + 0.2\ \mathrm{MSE}(\widehat{\widetilde{F}}, \widetilde{F}).$$

Both models are trained for 500 epochs using AdamW with batch size 64, an initial learning rate of 3×10^{-3}, weight decay 10^{-6}, cosine learning-rate decay to 5% of the initial value, and global gradient clipping at 10. No validation split or early stopping is used, and the final epoch is evaluated.

We report scalar-target MAE, RMSE, and R^2, together with force-component MAE and a dataset-level relative force L_2 error. If s indexes test samples, the latter is

$$\varepsilon_{F,L_2}^{\mathrm{rel}} = \frac{\left(\sum_s \sum_{i=1}^N \|\widehat{F}_i^{(s)} - F_i^{(s)}\|_2^2\right)^{1/2}}{\left(\sum_s \sum_{i=1}^N \|F_i^{(s)}\|_2^2\right)^{1/2}}.$$

Results. The numerical results are reported in Table 21, and the corresponding scalar fields and force profiles are shown in Figure 2. The WA-aware model successfully identifies the registry-dependent target for every wallpaper group. Its test scalar-target R^2 is at least 0.9897 for all 17 groups. The energy RMSE ranges from 1.65×10^{-3} to 4.94×10^{-2}, while the relative force L_2 error ranges from 2.75×10^{-2} to 3.48×10^{-1}, depending on the group.

In contrast, the WA-free control exhibits the behavior expected from its exact input collisions. Since all translated configurations provide the same relative graph geometry, its optimal scalar prediction is approximately a constant corresponding to the training-set mean. Its test R^2 is consequently approximately zero or slightly negative. A registry-independent scalar prediction also has approximately zero coordinate gradient, yielding relative force L_2 errors close to one for every wallpaper group.

Interpretation and limitations. This experiment establishes an information limitation for graph representations based only on relative coordinates. If a rigid translation leaves every model input unchanged while changing the prediction target, no deterministic model operating on those inputs can represent the target. Finite point-group equivariance alone does not provide the missing translation-registry information.

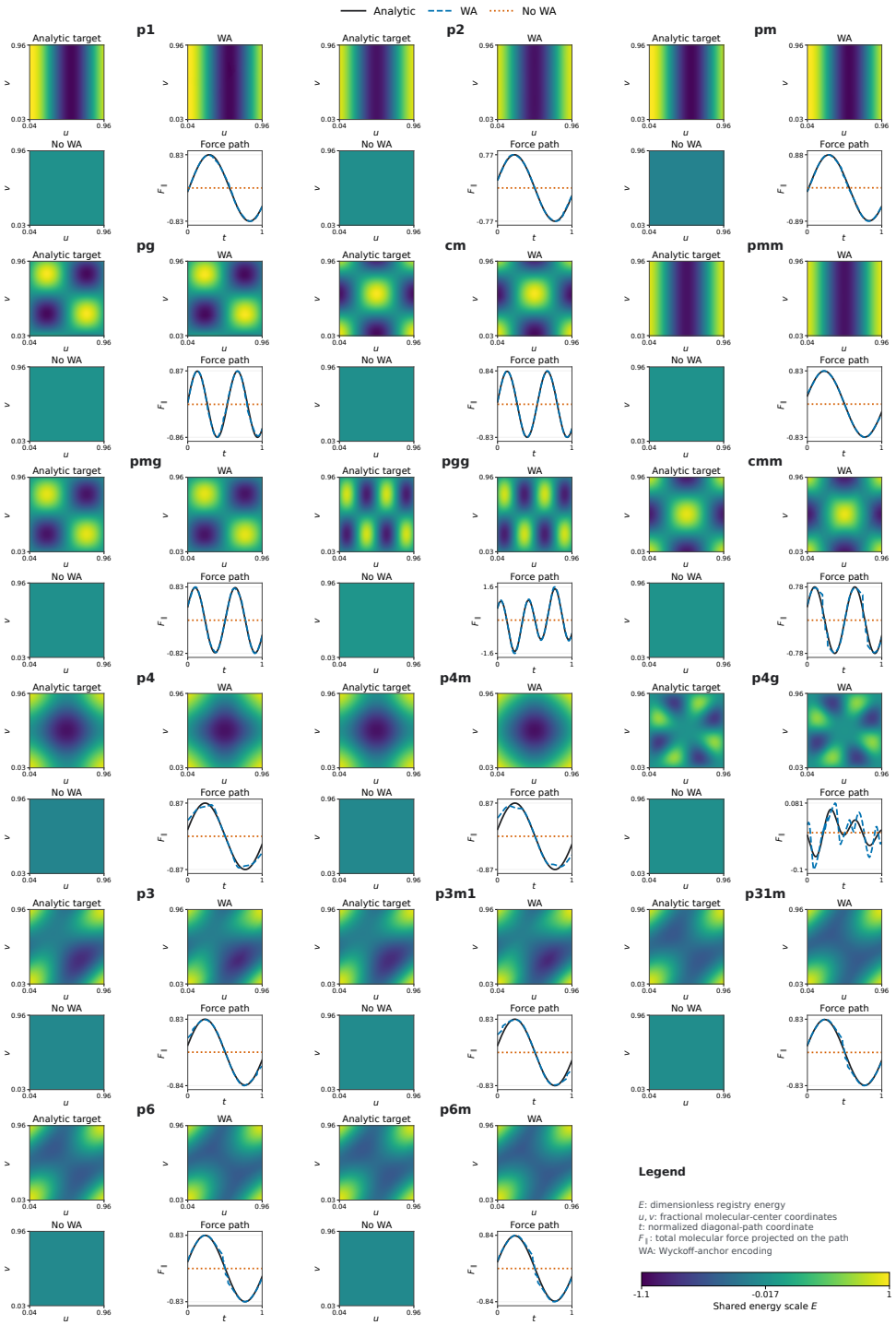


Figure 2: Wyckoff-anchor identifiability results for all 17 wallpaper groups. For each group, the panels compare the analytic periodic target, the WA-aware `WELinear` prediction, the parameter-matched WA-free prediction, and the force along the indicated reference path. The WA-aware model recovers both the registry-dependent scalar field and its gradient, whereas the WA-free control collapses to an approximately constant scalar prediction and a near-zero force.

Table 21: Held-out errors for the Wyckoff-anchor identifiability experiment. All metrics are evaluated on 256 independently sampled test translations per wallpaper group. Energy and force errors are reported in the arbitrary units of the analytic target. Lower is better for all error metrics, and higher is better for energy R^2. "WA" denotes `WELinear` with the complete WA encoding, and "No WA" denotes its parameter-matched WA-free control.

Group	Model	Parameters	Energy MAE	Energy RMSE	Energy R^2	Force-component MAE	Relative force L_2
$p1$	WA	3697	0.002613	0.003625	0.999974	0.004610	0.04473
	No WA	3697	0.634311	0.707385	−0.008664	0.152430	1.00000
$p2$	WA	3697	0.002398	0.003072	0.999979	0.004576	0.03931
	No WA	3697	0.593303	0.665364	−0.005738	0.143588	1.00000
pm	WA	3697	0.002929	0.003475	0.999974	0.003460	0.04237
	No WA	3697	0.607179	0.694349	−0.027733	0.092653	0.99999
pg	WA	3761	0.005303	0.006673	0.999846	0.011865	0.10526
	No WA	3761	0.450136	0.537128	−0.001104	0.137822	1.00000
cm	WA	3697	0.004503	0.005898	0.999868	0.008377	0.07944
	No WA	3697	0.419240	0.513149	−0.000015	0.136741	1.00000
pmm	WA	3697	0.001333	0.001654	0.999993	0.002161	0.02753
	No WA	3697	0.582893	0.660050	−0.056104	0.085241	1.00000
pmg	WA	3825	0.004778	0.005853	0.999880	0.009671	0.08614
	No WA	3825	0.464532	0.535519	−0.001146	0.134341	1.00000
pgg	WA	3825	0.005683	0.006936	0.999819	0.014312	0.08448
	No WA	3825	0.445993	0.515654	−0.001492	0.181650	0.99997
cmm	WA	3697	0.014763	0.018893	0.998436	0.022645	0.20213
	No WA	3697	0.380032	0.477817	−0.000370	0.126929	1.00000
$p4$	WA	3697	0.015689	0.019680	0.998291	0.017312	0.20629
	No WA	3697	0.391711	0.479993	−0.016825	0.105393	0.99998
$p4m$	WA	3697	0.031993	0.049409	0.989661	0.026882	0.34761
	No WA	3697	0.405150	0.486931	−0.004129	0.105298	1.00000
$p4g$	WA	3953	0.010970	0.013299	0.998741	0.031948	0.22845
	No WA	3953	0.304266	0.375349	−0.002736	0.143413	0.99999
$p3$	WA	3697	0.010445	0.013420	0.998990	0.013996	0.18085
	No WA	3697	0.328657	0.423490	−0.005963	0.089013	0.99999
$p3m1$	WA	3697	0.014300	0.017517	0.998098	0.018869	0.23138
	No WA	3697	0.311802	0.401951	−0.001774	0.090252	1.00000
$p31m$	WA	3697	0.009437	0.013730	0.998681	0.009836	0.14281
	No WA	3697	0.314269	0.378082	−0.000078	0.085820	1.00000
$p6$	WA	3697	0.011164	0.016016	0.998123	0.010694	0.16201
	No WA	3697	0.304423	0.370006	−0.001695	0.083255	0.99999
$p6m$	WA	3697	0.009741	0.014438	0.998181	0.011095	0.16585
	No WA	3697	0.278382	0.338782	−0.001671	0.083724	1.00000

The WA representation resolves this ambiguity by specifying the graph's position relative to symmetry-defined anchors of the wallpaper group. For the constructed nonsymmorphic cases, this continuous displacement must additionally be supplemented by translation-suborbit membership.

The experiment is intentionally restricted to representational identifiability. It uses one fixed four-node graph, an analytically constructed wallpaper-periodic target, one train/test realization, and one optimization run per wallpaper group. It therefore does not test generalization to new graph geometries or establish predictive performance for a particular physical system. Its purpose is to illustrate, in isolation, that the WA representation supplies registry information that cannot be recovered from the relative geometry of a rigid graph alone.

E SYNTHETIC WALLPAPER-GROUP BENCHMARK

Although wallpaper symmetries arise in any two-dimensional periodic structure, crystallographic systems provide one of their most direct physical applications. The benchmark is deliberately

Table 22: Realized support geometries for the 17 wallpaper groups. Support counts are given as unit-cell / explicit 3×3-cluster counts. The bond interval reports the periodic support bonds selected during construction. The coordination histogram is given as $\{$coordination : number of sites$\}$ for one conventional cell.

Group	Lattice	Point-group order	Support atoms unit cell / 3×3	$\lvert\mathbf{a}\rvert$ [Å]	$\lvert\mathbf{b}\rvert$ [Å]	γ	Support-bond range [Å]	Selected WA family
$p1$	oblique	1	3 / 27	2.265	3.485	90°	1.336–1.558	lattice
$p2$	oblique	2	3 / 27	3.701	1.851	90°	1.318–1.542	max-rotation-A
pm	rectangular	2	3 / 27	2.773	2.798	90°	1.386–1.474	lattice
pg	rectangular	2	4 / 36	4.867	2.060	90°	1.311–1.594	lattice
cm	centered rectangular	2	3 / 54	5.935	2.743	90°	1.372–1.488	lattice
pmm	rectangular	4	3 / 27	2.847	2.873	90°	1.423–1.437	max-rotation-A
pmg	rectangular	4	4 / 36	2.735	3.588	90°	1.367–1.495	max-rotation-A
pgg	rectangular	4	4 / 36	3.887	2.550	90°	1.363–1.430	max-rotation-A
cmm	centered rectangular	4	3 / 54	3.064	5.311	90°	1.328–1.532	max-rotation-A
$p4$	square	4	3 / 27	2.860	2.860	90°	1.430–1.430	max-rotation-A
$p4m$	square	8	3 / 27	2.860	2.860	90°	1.430–1.430	max-rotation-A
$p4g$	square	8	4 / 36	2.860	2.860	90°	1.430–1.430	max-rotation
$p3$	hexagonal	3	3 / 27	2.748	2.748	120°	1.430–1.430	max-rotation-A
$p3m1$	hexagonal	6	3 / 27	2.860	2.860	120°	1.430–1.430	max-rotation-A
$p31m$	hexagonal	6	3 / 27	2.625	2.625	120°	1.430–1.430	max-rotation-B
$p6$	hexagonal	6	3 / 27	2.860	2.860	120°	1.430–1.430	max-rotation
$p6m$	hexagonal	12	3 / 27	2.860	2.860	120°	1.430–1.430	max-rotation

designed as a toy problem: it removes much of the chemical complexity of realistic materials while retaining the geometric ingredients needed to study the effect of wallpaper symmetry in a setting that resembles a surface–adsorbate system. The support has on average 3 atoms per unit cell.

We construct a synthetic atomistic benchmark to isolate the effect of wallpaper-group symmetry in a controlled surface–adsorbate setting. The dataset contains all 17 crystallographic wallpaper groups, with 1,000 independently generated molecular configurations for each group, resulting in 17,000 systems in total. Every atom is carbon. Each system consists of a wallpaper-symmetric support and a small, intentionally non-symmetric carbon fragment containing three, four, or five atoms.

For each wallpaper group, a single support realization and cell are generated and subsequently held fixed across all 1,000 samples. Molecular geometry, atom count, vertical position, and lateral registry are sampled independently for each example.

Wallpaper-symmetric support generation. Support structures are constructed directly from the affine actions of the corresponding wallpaper group. Candidate seed positions are enumerated on rational grids with denominators 6, 8, and 12 and expanded under all group operations. The resulting symmetry-closed orbits are deduplicated before candidate support structures are assembled.

The target support size is three atoms per primitive cell. When the symmetry-orbit multiplicities do not permit exactly three atoms, nearby atom counts are considered. This occurs for pg, pmg, pgg, and $p4g$, for which the realized support contains four atoms per primitive cell. For the centered groups cm and cmm, three atoms per primitive cell correspond to six atoms in the displayed conventional cell.

For each candidate support, the lattice metric is selected to produce carbon-like local distances while respecting the lattice constraints imposed by the wallpaper group. Square and hexagonal groups retain their required metric, whereas rectangular and oblique groups are searched over a set of aspect ratios. Candidate periodic neighbor shells are evaluated over translations in $\{-1, 0, 1\}^2$. Each support atom is required to have at least two neighbors. The cell is then uniformly rescaled such that the median selected support bond is $1.43\,$Å, and all selected bonds are required to lie between 1.28 and $1.60\,$Å. Finally, the explicit 3×3 support must form a connected graph at a $1.60\,$Å cutoff. All support atoms are placed in a common plane at $z = -1.0\,$Å.

Table 22 summarizes the realized support geometry for all wallpaper groups.

Although 17 wallpaper groups are represented, there are 15 geometrically distinct realized supports. In particular, $p3m1$, $p6$, and $p6m$ share the same support coordinates and cell metric, but differ in the symmetry operations associated with the structure.

Molecular fragment generation. The molecular fragment contains

$$N_{\mathrm{mol}} \in \{3, 4, 5\} \tag{104}$$

carbon atoms, with the atom count sampled independently for every example. A connected path backbone is generated sequentially in the surface plane. Given the direction of the previous backbone edge, the direction of the next edge is changed by

$$\Delta\theta \sim \mathcal{U}(0.55, 2.45)\ \mathrm{rad}, \tag{105}$$

with the sign of the turn sampled independently. Thus, the magnitude of the proposed directional change covers approximately $31.5°$–$140.4°$. Equivalently, when expressed as the conventional internal angle between successive backbone bonds, the proposal range is approximately $39.6°$–$148.5°$.

The projected backbone distance is sampled independently as

$$d^{xy}_{i,i+1} \sim \mathcal{U}(1.38, 1.46)\,\text{Å}. \tag{106}$$

The angle and distance intervals above are proposal distributions. A proposed fragment is retained only if it satisfies additional geometric constraints. Specifically,

1. every pair of molecular atoms must be separated by at least 1.22 Å in the generated plane,
2. the difference between the largest and smallest molecular pair distances must exceed 0.12 Å, suppressing highly degenerate regular point sets,
3. a radius graph with cutoff 6.0 Å must give every atom degree at least two, and
4. every atom must have two neighbors whose normalized two-dimensional cross product exceeds 0.20, preventing locally collinear configurations.

Up to 2,048 candidate conformations are attempted. After acceptance, the planar molecular coordinates are translated such that their centroid is at the origin.

Because these rejection criteria act after sampling, the realized bond-angle distribution is not exactly uniform over the proposal interval.

Molecular registry and height. After generating the internal molecular geometry, the fragment is placed relative to the support. In fractional coordinates of the conventional cell, the molecular centroid is sampled uniformly from

$$(u_c, v_c) \sim \mathcal{U}\big([0.44, 0.56]^2\big)\,. \tag{107}$$

The molecule is therefore sampled near the center of the conventional cell. Individual molecular atoms are not required to remain inside the central cell. In the realized dataset, 7,613 systems (44.78%) have every molecular atom inside the central cell, whereas 9,387 systems (55.22%) have at least one molecular atom outside it. Thus, centering the molecular centroid near the middle of the cell does not constrain the complete fragment to a single cell.

Each molecular atom receives an independent vertical coordinate

$$z_i = 1.80 + \epsilon_i\ \text{Å}, \qquad \epsilon_i \sim \mathcal{U}(-0.16, 0.16), \tag{108}$$

giving

$$z_i \in [1.64, 1.96]\ \text{Å}. \tag{109}$$

Since the support lies at $z = -1.0$ Å, the purely vertical molecule–support separation lies between 2.64 and 2.96 Å before accounting for lateral displacement.

The conventional support cell is tiled 3×3, and the molecule is translated by $\mathbf{a} + \mathbf{b}$ so that its centroid lies above the central tile.

Realized molecular geometry. The realized dataset is consistent with the prescribed proposal distributions and rejection criteria. Projected backbone distances range from 1.380 to $1.460\,\text{Å}$, with a mean and median of $1.420\,\text{Å}$. Independent vertical displacements increase the corresponding full three-dimensional backbone distances slightly, yielding a range of 1.380–$1.492\,\text{Å}$ and a mean of $1.426\,\text{Å}$.

Considering all molecular atom pairs, the realized distances range from 1.221 to $5.407\,\text{Å}$, with a median of $1.473\,\text{Å}$. Consequently, every pair of atoms in the C_3–C_5 fragments lies within the $6\,\text{Å}$ molecular graph cutoff used in the experiments. The resulting radius graphs are therefore complete graphs K_3, K_4, or K_5.

The closest support atom to each molecular atom lies between 2.642 and $3.261\,\text{Å}$, with a mean of $2.879\,\text{Å}$. The absolute vertical difference across molecular backbone bonds has a mean of $0.106\,\text{Å}$, a 95th percentile of $0.248\,\text{Å}$, and a maximum of $0.318\,\text{Å}$.

Across all 17,000 samples, the three molecular sizes occur with approximately equal frequency: 5,627 systems contain C_3, 5,764 contain C_4, and 5,609 contain C_5.

Reference energies and forces. Each generated configuration is labeled using GFN2-xTB as implemented in xTB version 6.7.1. All systems contain only carbon atoms and are treated as neutral, closed-shell systems. No `-chrg` argument is supplied, so xTB uses its default total charge of zero, while `-uhf 0` explicitly sets the number of unpaired electrons to zero. The reference energy is the final total energy parsed from a single-point calculation. The calculation uses the default electronic temperature of $300\,\text{K}$, and the resulting energy is stored in Hartree and converted to electronvolts. The calculation logs report the default accuracy parameter of 1.0, a limit of 250 SCF iterations, an SCF convergence threshold of $10^{-6}\,\text{Hartree}$, and a wavefunction convergence threshold of $10^{-3}\,e$.

Atomic forces are obtained from a separate gradient calculation. When available, this calculation is initialized from the converged restart file produced by the single-point calculation. The final Cartesian gradient is read in $\text{Hartree}/a_0$, and the force on atom i is defined as the negative of the gradient of the energy. The conversion to the force units used for learning is $1\,\text{Hartree}/a_0 = 51.42206709\,\text{eV}/\text{Å}$.

Force labels are generated and retained for every atom, including the support atoms, although a downstream model evaluation may select only the molecular forces. Because the energy and gradient are produced by separate xTB processes, the stored energy is not taken from the process that writes the gradient. Both calculations nevertheless use the same geometry, method, charge, spin state, and electronic temperature. For an inspected example, their energies differ by approximately $4 \times 10^{-10}\,\text{Hartree}$, which is numerically negligible but not bitwise zero.

If the initial single-point calculation fails because the SCF procedure does not converge, one recovery attempt is made. The electronic structure is first converged at an electronic temperature of $1000\,\text{K}$, and the resulting restart density is then reused for a second single-point calculation at $300\,\text{K}$. Only the energy recomputed at $300\,\text{K}$ is retained. This recovery procedure was required for 10 of the 17,000 configurations, and all 17,000 configurations have complete reference results.

The relatively large forces in the dataset reflect the intentionally non-equilibrium sampled geometries. The calculations are performed on the explicitly included finite 3×3 support cluster. Although the serialized extended-XYZ structures contain a simulation cell and in-plane periodic-boundary metadata for subsequent graph construction, no periodic lattice option is passed to xTB.

Finite-cluster limitation. Because the reference labels are computed for this explicit, nonperiodic 3×3 support cluster rather than for an infinite periodic support, they are only approximately invariant and equivariant under the wallpaper group. Symmetry-equivalent placements of the fragment can lie at different distances from the cluster boundary, so their reference energies and forces can differ slightly, whereas the WallpaperNet models treat them as equivalent by construction. We reduce this finite-size effect by placing the molecular centroid above the central cell of the cluster, as described above. One full support cell then separates the central cell from the cluster edge in every in-plane direction, which keeps the fragment away from the boundary. Figure 3 shows an example five-atom fragment in the synthetic $p4m$ environment. The remaining boundary contribution acts as a residual symmetry-breaking perturbation of the targets.

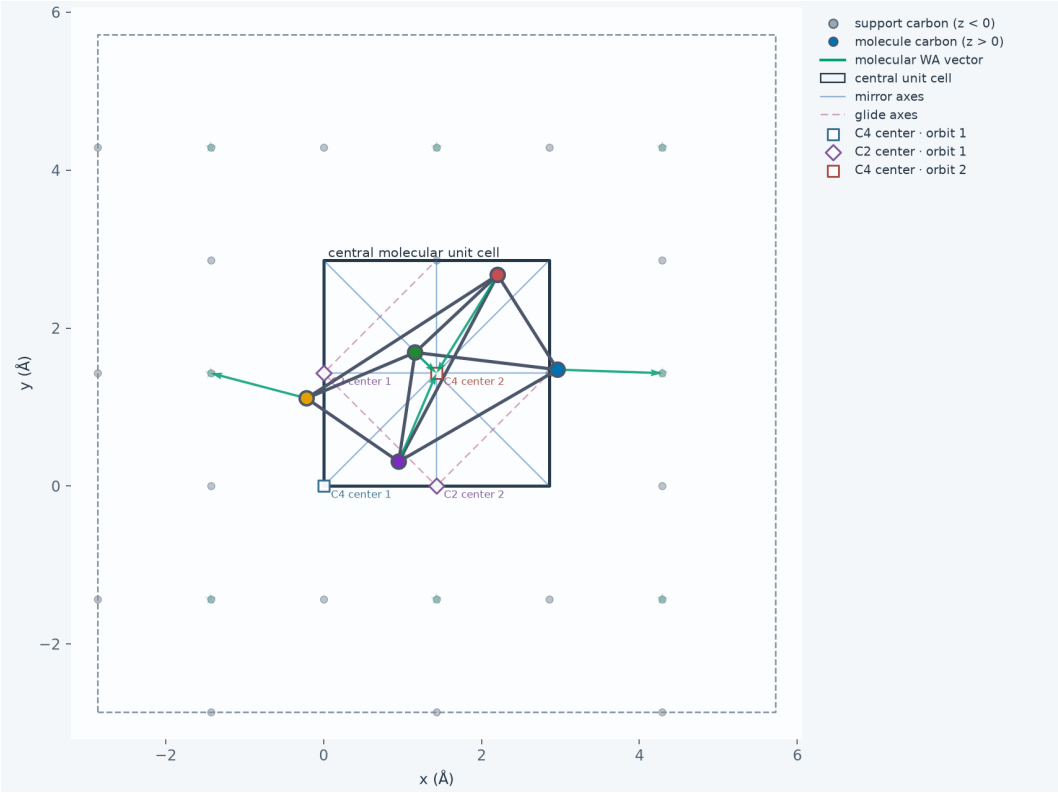


Figure 3: Top view of a synthetic $p4m$ system with a five-atom carbon fragment. Gray points are the support atoms of the finite 3×3 cluster, with three atoms per primitive cell, and the dashed line marks the cluster boundary. The fragment is placed above the central unit cell, shown as a solid square. Colored nodes and dark lines show the molecular graph, and green arrows show the molecular WA vectors. Mirror and glide axes and the C_4 and C_2 rotation centers of $p4m$ are also shown.

F Dataset construction and simulation details

F.1 Motivation for the electronic-structure level

The graphene adsorption dataset was designed to provide a large and chemically diverse set of adsorption geometries, energies, and forces at a computational cost compatible with training an interatomic potential. Because the dataset contains thousands of adsorbates and, for each adsorbate, an entire geometry-optimization trajectory, direct density-functional-theory (DFT) calculations for all configurations would be prohibitively expensive. We therefore used the semiempirical GFN2-xTB method for dataset generation.

The suitability of this approximation was assessed on a chemically diverse benchmark set of 300 organic molecules. The benchmark molecules were selected from the underlying molecular database using molecular fingerprints and Tanimoto distances, choosing mutually dissimilar structures in order to span the available chemical space. The GFN2-xTB protocol was selected based on comparison with periodic PBE-D4/DZP calculations (Perdew et al., 1996; Caldeweyher et al., 2019). Ordinary least-squares regressions of the GFN2-xTB results on the corresponding PBE-D4/DZP reference values yielded $R^2 = 0.849$ and a residual MAE of $0.058\,\mathrm{eV}$ for the binding energies ($N = 300$), and $R^2 = 0.895$ and a residual MAE of $0.196\,\mathrm{eV}$ for the unoccupied LUMO-like adsorbate-projected peak positions ($N = 299$). The adsorption energy governs the stability of the adsorbate–graphene complex, whereas the LUMO frontier energy governs the propensity for interfacial electron transfer and is therefore important for molecular reduction, charge-transfer processes, and catalytic activation (Lüder et al., 2017; Fedorov & Gryn'ova, 2023; Yu, 2014b;a; 2013; Medford et al., 2015; Mateo et al., 2021; Liu et al., 2020).

F.2 Adsorbate preparation and simulation-cell construction

The adsorption structures were generated automatically to permit high-throughput treatment of a chemically diverse molecular dataset. For each molecule, aromatic-ring planes were identified from sets of three ring atoms. Their surface-normal vectors were oriented consistently and summed to obtain an overall molecular normal. The molecule was subsequently rotated such that this normal was aligned with the surface-normal (z) direction. When necessary, the molecule was additionally reflected with respect to the xy plane to prevent bulky substituents from initially pointing into the graphene surface.

The lateral dimensions of the graphene supercell were chosen separately for each adsorbate. A molecular footprint was first constructed with an approximately 2 Å margin around the molecule, and the graphene unit cell was repeated until a suitable hexagonal supercell was obtained. The surface coverage was kept approximately comparable between systems by targeting a ratio of molecular footprint area to graphene area of approximately 0.4. The adsorbate was initially positioned 3 Å above the graphene surface, measured from its lowest atom. Along the surface-normal direction, at least 15 Å of vacuum was included above the highest adsorbate atom.

The substrate consisted of three graphene layers. The number of layers was chosen based on a convergence test of the adsorption energy: increasing the substrate beyond three layers produced only minor changes in the calculated binding energy. All graphene atoms were kept fixed throughout the subsequent geometry optimizations and property calculations.

F.3 Validation against periodic DFT

The accuracy required from the semiempirical reference level was assessed using a benchmark set of 300 chemically diverse adsorbates. Rather than selecting this set randomly, molecular structures were encoded using Morgan fingerprints and compared using Tanimoto distances. Molecules were selected to maximize their mutual dissimilarity, providing a benchmark intended to span the chemical diversity of the full molecular library.

Binding energies were computed using the energies of the complete graphene-adsorbate system, the graphene substrate, and the isolated adsorbate in its adsorption geometry, with identical supercell dimensions and boundary conditions. We use the positive binding-energy convention

$$E_{\mathrm{B}} = E_{\mathrm{graphene}} + E_{\mathrm{adsorbate}} - E_{\mathrm{graphene+adsorbate}}, \tag{110}$$

where $E_{\mathrm{graphene+adsorbate}}$ is the energy of the relaxed combined system, E_{graphene} is the energy of the isolated graphene substrate, and $E_{\mathrm{adsorbate}}$ is the energy of the isolated adsorbate in its adsorption geometry. With this convention, stronger favorable adsorption corresponds to a larger positive E_{B}. The original benchmark compared GFN2-xTB against periodic PBE-D4/DZP calculations.

As a complementary electronic-structure test, projected densities of states were calculated for the adsorbate and substrate. A $5 \times 5 \times 1$ Monkhorst-Pack mesh was used for these calculations, coarser meshes were found to spuriously open a gap in graphene. Energies were referenced to the Fermi energy, and the first maximum of the unoccupied adsorbate-projected density of states was used as a practical definition of the adsorbate LUMO.

Taken together, the energetic and frontier-state benchmarks indicate that GFN2-xTB retains the relevant trends of the periodic DFT description at a fraction of its computational cost. This trade-off is essential for generating training data for many thousands of graphene-adsorbate systems and was therefore used as the reference level for the machine-learning dataset.

Molecular preprocessing, periodic-cell construction, and geometry handling used ASE and RDKit (Hjorth Larsen et al., 2017; Landrum et al., 2025). The initial molecular geometries were drawn from the chemical space described by Fedorov et al. (2026). The adsorbates were relaxed with GFN2-xTB in DFTB+ (Hourahine et al., 2020), and the DFT computation was carried out in GPAW (Mortensen et al., 2024).

F.4 Dataset statistics and force distribution

For the WallpaperNet models, the graph representation contains only the molecular adsorbate atoms. The graphene atoms are excluded from message passing, although their number is retained as

metadata. Each system contributes one graph record. Energies are reported in eV per system and forces in $\mathrm{eV\,\AA^{-1}}$.

Table 23 summarizes the complete dataset. All reported standard deviations are population standard deviations.

Table 23: Composition and force-magnitude statistics for the complete graphene adsorption dataset.

Quantity	N	Mean	SD	Min.	Median	Max.
Molecular atoms/system	15,089	24.926	3.744	8	25	44
Graphene atoms/system	15,089	361.900	50.753	150	378	660
Full-cell atoms/system	15,089	386.825	51.593	158	400	699
Force norm ($\mathrm{eV\,\AA^{-1}}$)	376,104	0.360	0.290	0.0005	0.271	4.545

The total system energies span from $-40{,}589.4$ to $-9{,}197.05\,\mathrm{eV}$, with a median of $-23{,}471.2\,\mathrm{eV}$. The 5th and 95th percentiles are $-26{,}726.2$ and $-18{,}242.2\,\mathrm{eV}$, respectively. The force norm has a mean of $0.360\,\mathrm{eV\,\AA^{-1}}$, a median of $0.271\,\mathrm{eV\,\AA^{-1}}$, and a 5th–95th percentile interval of 0.063–$0.940\,\mathrm{eV\,\AA^{-1}}$. The mean signed in-plane components are effectively zero, while the mean surface-normal component is $-0.00956\,\mathrm{eV\,\AA^{-1}}$. These force statistics include one vector per molecular atom and exclude the frozen graphene atoms.

G Architectures for real graphene–adsorbate and synthetic experiments

The synthetic experiments evaluate two training scenarios, whereas the real-system study evaluates three. The construction is presented first in the generic form used by the synthetic (toy) wallpaper-group experiments and then specialized to the real graphene–adsorbate data. The WP aliases correspond to the implementation names as follows: WP-Inv is `egnn_like`, WP-InvEq is `mixed_egnn_like`, WP-Equi is `wa_linear`, WP-VecInv is `mix_equivariant_advanced`, WP-Reg is `wa_advanced_sr`, WP-RegTP is `wa_linear_mod`, and WP-O3 is `wa_advanced_o3_feedback`.

Seven architectures operate on the adsorbate/molecule graph only. Three architectures (`full_cell_egnn`, `full_cell_e3nn_lmax1`, and `full_cell_e3nn_lmax2`) operate on the complete adsorbate–graphene cell. For every architecture, forces are supervised and evaluated only on adsorbate atoms. Graphene atoms therefore contribute context in full-cell models, but do not contribute entries to the force loss.

The corresponding toy model names are, in the same order, `molecule_invariant`, `molecule_mixed_conservative` (combined mode) or `molecule_mixed` (direct-force mode), `p_molecule_full_equivariant`, `wa_advanced`, `wa_advanced_sr`, `wa_linear_mod`, `wa_advanced_o3_feedback`, `full_graph_invariant`, `full_graph_e3nn_lmax1_tensor_product`, and `full_graph_e3nn_tensor_product`. The toy and real-system models share the same architectural blueprint, but are not always algebraically identical.

G.1 Common notation and graph construction

Node positions are $x_i \in \mathbb{R}^3$. The real experiments use a cutoff of 6.0 and at most 128 neighbors per target. For an edge, define

$$\begin{aligned} r_{ij} &= x_i - x_j, \\ d_{ij} &= \|r_{ij}\|, \\ \widehat{r}_{ij} &= r_{ij} / \max(d_{ij}, \varepsilon), \\ r_{ij,xy} &= (r_{ij,x}, r_{ij,y}), \\ r_{ij,z} &= (r_{ij})_z. \end{aligned}$$

Here $j \to i$ denotes a directed edge, $\mathcal{N}(i)$ is the set of senders connected to receiver i, $\deg(i) = |\mathcal{N}(i)|$, and $\varepsilon = 10^{-8}$ prevents division by zero. We embed a planar WA vector in three dimensions as $\overline{w}_i = (w_{i,x}, w_{i,y}, 0)$. Indices i and j always label receiver and sender nodes, respectively. k labels a feature channel, p labels a message path, and ℓ labels a message-passing layer. Whether a symbol denotes a scalar, vector, feature array, or representation-valued field is specified by its definition and ambient space.

Each adsorbate node also has a two-dimensional Wyckoff-anchor vector in the surface plane. To simplify indexing in this section, we write w_i for the Wyckoff-anchor vector $v_{\mathrm{WA},i}$ defined earlier. It encodes the node's displacement relative to its associated crystallographic anchor. WA vectors are used by the molecule-only wallpaper models and are ignored by the full-cell baselines.

The toy implementation is group-generic. A node has a single WA vector w_i and, for nonsymmorphic wallpaper groups, a learned regular-representation encoding of the translation-coset/member label. The real data use the fixed $p6m$ point group (D_6 point symmetry), one two-dimensional WA vector per adsorbate atom, and no nonsymmorphic member feature.

Initial scalar node features. In the real-system models, all Wallpaper architectures begin with the same three scalar node attributes:

$$q_i = [Z_i,\, m_i,\, z_i],$$

where Z_i is the atomic number, m_i is the atomic mass, and z_i is the z coordinate of the node. (It's important to note that for the full-cell architectures, the z coordinate is not used as an additional node feature.) For the full-cell architectures, the z coordinate is not used in the message passing. The full-cell architectures don't receive the z coordinate as input. A learned affine map L_{in} produces the initial scalar hidden state of width H:

$$h_i^{(0)} = L_{\mathrm{in}}(q_i) \in \mathbb{R}^H.$$

All models use four message-passing layers and ReLU in the ordinary scalar MLPs unless a layer explicitly uses a symmetry-aware SiLU or ELU activation.

The toy models instead use a simplified schema: a linear map of Z_i, an MLP of z_i, and, for full-cell models, a linear map of the binary node-type indicator.

Radial basis. Every real-system architecture expands d_{ij} in 16 smooth finite-support radial basis functions on $[0, 6]$ as implemented in `e3nn`. The basis is multiplied by $\sqrt{16}$. Denoting this expansion by $B(d_{ij})$, scalar edge networks receive both $\|r_{ij}\|^2$ and $B(d_{ij})$, except for the e3nn UVU models, whose tensor-product weights are generated from $B(d_{ij})$.

The toy models use the same cutoff and 16 channels, but their common radial map, denoted by $B_{\mathrm{toy}}(d_{ij}) \in \mathbb{R}^{16}$, is a Gaussian basis multiplied by a cosine cutoff.

Compact D_6 invariants. For a planar vector $v = (v_x, v_y)$, let $\zeta(v) = (v_x + \mathrm{i}v_y)/\max(\|v\|, \varepsilon)$, where $\mathrm{i}^2 = -1$. The compact D_6 feature map used by the scalar D_6 layers is

$$\phi_{D_6}(v) = \left[\mathrm{Re}(\zeta(v)^6),\, \mathrm{Re}(\zeta(v)^6),\, \|v\|\right].$$

The repeated first component is intentional in the current implementation: a reflection changes $\mathrm{Im}(\zeta^6)$, so only $\mathrm{Re}(\zeta^6)$ is invariant under D_6. Repeating it retains a three-channel interface without introducing a reflection-odd scalar. The clamped denominator gives $\zeta(0) = 0$.

For models with joint edge–WA features enabled, the D_6-invariant contraction

$$J_{D_6}(r_{ij,xy}, w_i) = [r_{ij,xy} \cdot w_i,\, r_{ij,xy} \cdot w_i].$$

is appended. The duplicate is again intentional and keeps two joint channels. Unlike the C_6 version, no signed planar cross product is used because it is odd under reflection.

Aggregation and residual updates. For compactness, define the two aggregators used below:

$$\mathcal{A}_i^{\mathrm{sum}}(u) = \sum_{j \in \mathcal{N}(i)} u_{ij}, \qquad \mathcal{A}_i^{\mathrm{dn}}(u) = \frac{1}{\sqrt{\max(1, \deg i)}} \sum_{j \in \mathcal{N}(i)} u_{ij}.$$

The scalar and simple equivariant layers use $\mathcal{A}^{\mathrm{sum}}$. The advanced regular-state WA layers use $\mathcal{A}^{\mathrm{dn}}$.

Scalar hidden updates use a learned residual projection W_{res} so that input and output widths may differ. In the real-system matrix all four layers have equal input and output width H.

In later equations, 1 denotes the trivial representation, V the two-dimensional planar vector representation, and $\mathrm{Reg}(G)$ the regular representation of the finite point group G, where G denotes the wallpaper point group relevant to the current model. The notation $n\,\rho$ means n copies of representation ρ. $\mathcal{L}_{\rho \to \rho'}$ denotes a learned equivariant linear map, $\mathrm{TP}_{\mathrm{UVU}}$ a factorized equivariant tensor product, $\mathrm{InvNorm}$ the vector of smooth norms of non-scalar irrep fields, and σ_{Reg} a representation-compatible nonlinearity. ϕ_G denotes the invariant single-vector feature map for group G, and J_G the corresponding invariant joint feature of an edge and a WA vector. Brackets denote feature concatenation and $\odot$ denotes channel-wise multiplication. Symbols such as U, M, E, F, G, and L with subscripts denote independently learned maps unless stated otherwise. The symbol S in a map subscript denotes the persistent-state representation of the architecture under discussion. A prime marks the output of one message-passing layer, and a tilde marks an intermediate quantity. For $O(3)$ irreducible representations, ℓ is the angular degree and the suffixes e and o denote even and odd inversion parity, respectively. For example, $1o$ is a polar vector and $1e$ is an axial vector.

G.2 Training scenarios

Shared conservative formulation. The final scalar node states are sum-pooled over all graph nodes:

$$g = \sum_i h_i^{(4)}.$$

The pooled vector is concatenated with ten global quantities. With N_{sup} the number of support atoms, $a, b, c \in \mathbb{R}^3$ the cell vectors, and vec denoting concatenation of Cartesian components, define

$$c_{\mathrm{cell}} = [N_{\mathrm{sup}}, \mathrm{vec}(a), \mathrm{vec}(b), \mathrm{vec}(c)] \in \mathbb{R}^{10}, \qquad E_{\mathrm{pred}} = E_{\mathrm{ref}} + M_{\mathrm{graph}}([g, c_{\mathrm{cell}}]).$$

Here E_{ref} is the appropriate sum of atomic reference energies and M_{graph} is a four-layer MLP with widths $H + 10 \to H \to H \to H \to 1$. The toy implementation deliberately matches this topology: it supplies the synthetic support-atom count and cell, and adds the reference energy of the complete synthetic system even when message passing is restricted to the adsorbate. For molecule-only real-system models, the prediction instead adds the adsorbate atom-reference energies and the separately supplied sum of the atom-reference energies of the graphene. For full-cell real models, it adds the atom-reference energies of all included atoms. Thus every conservative run predicts the full system's total energy.

Forces are obtained from the predicted total energy:

$$F_i^{\mathrm{pred}} = -\frac{\partial E_{\mathrm{pred}}}{\partial x_i}.$$

Training therefore differentiates through this coordinate gradient. Let E and F_i denote the reference energy and force, respectively, and let $\mathcal{M}$ be the set of adsorbate atoms. The graph-energy L1 loss, component-wise force MSE, and weighted objective are

$$L_E = |E_{\mathrm{pred}} - E|, \qquad L_F = \frac{1}{3|\mathcal{M}|} \sum_{i \in \mathcal{M}} \|F_i^{\mathrm{pred}} - F_i\|_2^2, \qquad \mathcal{L} = \lambda_E L_E + \lambda_F L_F.$$

Here $\lambda_E, \lambda_F \geq 0$ are the energy- and force-loss weights. The force loss is masked to molecule atoms even for full-cell graphs. Both conservative scenarios enforce energy conservation by construction, up to ordinary numerical differentiation error.

Each message-passing layer also returns a direct force proposal because the same layer class is shared with force-head training. That returned proposal is discarded in combined mode and is not part of the conservative force prediction. Internally, some hybrid architectures can still use equivariant vector intermediates to update their scalar or persistent states.

Toy conservative training. The toy experiment has one conservative joint energy–force condition, with $(\lambda_E, \lambda_F) = (1, 1)$. Its other condition is the direct force head described below.

Real-system conservative training. The real-system experiment has two conservative conditions. The standard combined condition uses $(\lambda_E, \lambda_F) = (1, 1)$ and is the condition used to assess prediction of the full total energy. The force-weighted conservative condition keeps the identical energy model and gradient-derived force definition but weights the energy term by a coefficient determined from the training split, $\lambda_E \approx 1.22 \times 10^{-6}$ with $\lambda_F = 1$ (Appendix I). The coefficient matches the numerical scales of the energy-L1 and force-MSE terms and accounts for the ratio of adsorbate to full-system atoms, so both terms contribute to the objective.

Direct force model. The direct-head scenario, used in both the toy and real-system studies, does not predict or supervise energy. Each of the four message-passing layers emits an equivariant node-vector proposal $f_i^{(\ell)}$, and the network output is their sum:

$$F_i^{\text{pred}} = \sum_{\ell=1}^{4} f_i^{(\ell)}.$$

The objective is component-wise MSE over the x, y, and z components of adsorbate atoms. Full-cell models still propagate messages through graphene nodes, but graphene force entries are removed by `mol_mask` before loss and metric computation. Direct forces transform according to the representation implemented by each layer, but are not constrained to be the gradient of a scalar energy and are therefore not guaranteed to be conservative.

The graph-energy readout remains instantiated by the shared base model but is not called in this scenario.

G.3 Registered and loss-connected parameter counts

Here, registered means every parameter with `requires_grad=True`. A parameter is counted as *active* for a training mode when its gradient is not `None` after backpropagation from that mode's objective. The models' capacity was selected to place every model near one million registered trainable parameters for the generated real-system configurations and around 150,000 registered trainable parameters for the generated toy-model configurations, to achieve a fair comparison between the models.

For a conservative model, forces are obtained as the negative gradient of the energy with respect to the molecular atom positions. The real-system standard combined loss, $(\lambda_E, \lambda_F) = (1, 1)$, and the force-weighted conservative loss, $(\lambda_E, \lambda_F) \approx (1.22 \times 10^{-6}, 1)$, have the same active-parameter count: their energy coefficient is nonzero and both use the same energy graph. The direct-head count instead follows only the explicitly predicted force. The parameter count for every model for real system and toy models for a p6m wallpaper group are given in Table 24 and Table 25 respectively.

Real-system models. These counts use the ten generated real-system architectures. The registered counts reproduce the values stored in their configurations. The conservative column applies to both real conservative scenarios described above. The abbreviations C and V denote conservative combined training and direct-force-head training, respectively.

Table 24: Parameter counts for the real-system models.

Architecture	MP implementation	Scope	H	Registered	Active C	Active V
WP-Inv	`D6_invariant`	molecule only	144	997,929	829,729	829,592
WP-InvEq	`D6_p_mix_equivariant`	molecule only	138	996,373	839,869	822,423
WP-Equi	`D6_p_full_equivariant`	molecule only	116	992,501	668,393	930,208
WP-VecInv	`D6_p_mix_equivariant_advanced`	molecule only	104	1,004,661	936,420	764,210
WP-Reg	`D6_wa_advanced_sr`	molecule only	103	999,746	663,073	636,638
WP-RegTP	`D6_wa_linear_mod`	molecule only	94	995,897	994,933	827,733
WP-O3	`D6_wa_advanced_o3_feedback`	molecule only	70	994,449	836,836	816,169
Full cell invariant GNN	`mp_full_invariant`	full cell	158	991,613	990,977	789,214
Full cell e3nn $\ell_{\max} = 1$	`mp_full_equivariant_tensor_state_ll_uvu`	full cell	120	997,429	943,729	750,732
Full cell e3nn $\ell_{\max} = 2$	`mp_full_equivariant_tensor_state_uvu`	full cell	112	1,000,457	946,177	762,424

The difference between registered and active parameters is intentional. Conservative training does not use parameters exclusive to a returned direct force proposal. Direct-head training does not call the graph-energy readout. Parameters shared by message passing, persistent states, and feedback paths are counted whenever the selected loss remains connected to them.

Toy-model architectures for $p6m$. The toy architecture reproduce the core ideas of the real-system architectures, but with simplified implementations in most cases. The only difference is WP-InvExt and real system WP-InvEq, toy WP-VecEqui and real WP-VecInv have different implementations. The toy architectures are group dependent, so a single group must be fixed before quoting one width and parameter count. This table uses the complete $p6m$ configuration matrix, the seed affects data splitting and initialization but not architecture size. The abbreviations C and V denote conservative combined training and direct-force-head training, respectively.

Where the paired configurations use the same implementation, width, or registered count, that value is shown once. A slash gives C/V. In particular, WP-InvExt and full-cell e3nn $\ell_{\max} = 1$ use independently calibrated model variants or widths.

Table 25: Parameter counts for the $p6m$ toy models.

Architecture	Toy implementation (C/V where different)	Scope	H (C/V)	Registered (C/V)	Active C	Active V
WP-Inv	`molecule_invariant`	molecule only	58	149,997	149,525	122,098
WP-InvExt	`molecule_mixed_conservative/molecule_mixed`	molecule only	40/58	151,925/150,233	148,644	122,334
WP-Equi	`p_molecule_full_equivariant`	molecule only	46	149,413	123,649	130,598
WP-VecEqui	`wa_advanced`	molecule only	50	143,885	143,028	120,902
WP-Reg	`wa_advanced_sr`	molecule only	27	149,335	140,347	139,706
WP-RegTP	`wa_linear_mod`	molecule only	44	149,917	149,733	113,707
WP-O3	`wa_advanced_o3_feedback`	molecule only	22	148,957	148,655	143,297
Full cell invariant GNN	`full_graph_invariant`	full cell	58	148,137	147,901	120,238
Full cell e3nn $\ell_{\max} = 1$	`full_graph_e3nn_lmax1_tensor_product`	full cell	20/29	152,765/195,359	150,861	184,895
Full cell e3nn $\ell_{\max} = 2$	`full_graph_e3nn_tensor_product`	full cell	22	144,109	142,620	137,563

G.4 ARCHITECTURE DESCRIPTIONS

Overview. The toy benchmark uses seven WallpaperNet architectures of increasing complexity. WP-Inv is a distance-based model with angular invariants and an EGNN-style vector head for force prediction. WP-InvExt introduces more geometry features while remaining scalar only: it extends invariant scalar messages with covariant edge and Wyckoff-anchor directions, with an additional vector state used during the conservative training objective. WP-Equi replaces invariant message construction with a nonlinear wallpaper-group-equivariant edge network whose hidden fields transform as regular representations. WP-VecEqui couples edge and Wyckoff-anchor regular fields through a single explicit tensor product and uses the resulting field to update a persistent regular-representation node state. WP-Reg extends this approach through separately gated regular-representation paths, including receiver- and sender-anchor fields and state-conditioned tensor-product interactions. WP-RegTP explicitly couples direct, edge, and Wyckoff-anchor regular fields through successive tensor products. WP-O3 combines three-dimensional $O(3)$ tensor features with wallpaper-group-equivariant processing through bidirectional feedback. The full-cell invariant GNN applies invariant EGNN-style message passing to the complete molecule–support graph. The full-cell e3nn $\ell_{\max} = 1$ model uses $O(3)$-equivariant tensor products with scalar and vector features, and the $\ell_{\max} = 2$ model additionally propagates rank-two features to capture quadrupolar angular structure.

For the real-system experiments, we reuse the WP-Inv, WP-Equi, WP-Reg, WP-RegTP, and WP-O3 designs. Because WP-InvExt outperformed most explicitly equivariant toy models despite relying primarily on scalar message passing, its design choices are transferred into two real-system variants. WP-InvEq keeps essentially the WP-Inv invariant core but sends the scalar and force outputs through explicit D_6-equivariant nonlinear regular-representation post-maps and has no persistent non-scalar state. WP-VecInv instead keeps invariant scalar messages while introducing persistent polar vectors whose invariant contractions influence subsequent scalar messages. Its direct force is formed from edge/WA covariant bases and then passed through a D_6-equivariant output network. The toy WP-VecEqui architecture is omitted because it was the least accurate toy model and substantially overlaps with the more effective WP-Reg and WP-RegTP designs.

WP-Inv: scalar D_6-invariant molecule model. This is the simplest molecule-only reference. It follows the scalar-state pattern of an EGNN while augmenting the scalar edge message with D_6 invariants and WA information. For each edge, define the invariant input and message

$$u_{ij}^{D_6} = [h_i, h_j, \|r_{ij}\|^2, B(d_{ij}), \phi_{D_6}(w_i), \phi_{D_6}(r_{ij,xy}), r_{ij,z}, J_{D_6}(r_{ij,xy}, w_i)],$$
$$m_{ij} = M_{\text{inv}}^{D_6}(u_{ij}^{D_6}) \in \mathbb{R}^H.$$

The joint term is enabled in the real-system configuration. The message MLP has three affine layers of width H with two ReLUs. Messages are summed at the receiver, and a second three-layer MLP updates the scalar state:

$$h_i' = W_{\text{res}} h_i + U_{\text{node}}([h_i, \mathcal{A}_i^{\text{sum}}(m)]) .$$

There is no persistent non-scalar hidden state. The direct planar force is an EGNN-like scalar-weighted edge vector, while the vertical component is an independent scalar MLP:

$$f_{ij,xy} = \alpha(m_{ij}) r_{ij,xy},$$
$$f_{ij,z} = \beta(m_{ij}),$$
$$f_i = \mathcal{A}_i^{\text{sum}}([f_{ij,xy}, f_{ij,z}]).$$

Here $M_{\text{inv}}^{D_6}$ and U_{node} are scalar MLPs, while $\alpha : \mathbb{R}^H \to \mathbb{R}$ and $\beta : \mathbb{R}^H \to \mathbb{R}$ are learned scalar readouts.

The scalar stream is D_6 invariant, and multiplication of the planar vector by an invariant coefficient gives a D_6-covariant planar output. The z component is a scalar under in-plane D_6 operations.

The toy `molecule_invariant` uses the same message-sum-node-update template, with ϕ_G and J_G constructed for the selected wallpaper group. It additionally can incorporate a nonsymmorphic member encoding.

WP-InvExt: invariant messages extended by covariant WA bases. This is a toy-only architecture that uses `molecule_mixed_conservative` for conservative training and `molecule_mixed` for direct-force training. The name indicates that an invariant scalar core is extended by covariant edge and Wyckoff-anchor directions. The core idea of this implementation was for the model to stay essentially invariant, but use more provided geometric information. Since the conservative training does not use a force head constructed from the edge and WA directions, we added a persistent vector-state hidden state for the node.

For a selected wallpaper group G, both variants form

$$u_{ij}^G = [h_i, h_j, \|r_{ij}\|^2, B_{\text{toy}}(d_{ij}), \phi_G(r_{ij,xy}), \phi_G(w_i), r_{ij,z}, J_G(r_{ij,xy}, w_i), \rho_i],$$
$$m_{ij} = M_{\text{inv}}^G(u_{ij}^G) \in \mathbb{R}^H,$$

where ρ_i is present only for nonsymmorphic groups and encodes the selected translation-suborbit member. The direct `molecule_mixed` variant has only scalar node states:

$$h_i' = W_{\text{res}} h_i + U_{\text{inv}}([h_i, \mathcal{A}_i^{\text{sum}}(m)])\,.$$

Its planar force spans two covariant bases rather than only the molecular edge:

$$\begin{aligned} f_{ij,xy} &= a_{ij}^{(r)} r_{ij,xy} + a_{ij}^{(w)} w_i, \\ f_{ij,z} &= b_{ij}, \qquad\qquad [a_{ij}^{(r)}, a_{ij}^{(w)}, b_{ij}] = F_{\text{inv}}(m_{ij}). \\ f_i &= \mathcal{A}_i^{\text{sum}}([f_{ij,xy}, f_{ij,z}]), \end{aligned}$$

Invariant coefficients multiplying covariant bases give a G-equivariant planar force without introducing a regular-representation hidden field. The conservative `molecule_mixed_conservative` variant adds H persistent polar three-vectors $V_i \in \mathbb{R}^{H\times 3}$. For every channel k, it constructs

$$\begin{aligned} c_{ij,k} = [&\|V_{i,k}\|^2, \|V_{j,k}\|^2, V_{i,k}\cdot V_{j,k}, V_{i,k}\cdot \widehat{r}_{ij}, V_{j,k}\cdot \widehat{r}_{ij}, \\ &V_{i,k}\cdot \overline{w}_i, V_{j,k}\cdot \overline{w}_i], \qquad s_{ij} = S_{\text{inv}}(\{c_{ij,k}\}_{k=1}^H) \in \mathbb{R}^H. \end{aligned}$$

All entries of $c_{ij,k}$ are invariant scalars. The scalar node update is

$$h_i' = W_{\text{res}} h_i + U_{\text{mix}}([h_i, \mathcal{A}_i^{\text{sum}}(m), \mathcal{A}_i^{\text{sum}}(s)])\,.$$

Here S_{inv} is an invariant MLP that maps the complete set of channel-wise contractions to the length-H message s_{ij}. An invariant coefficient map G_V produces $[(a_{ij,k})_{k=1}^H, (b_{ij,k})_{k=1}^H] = G_V(m_{ij})$. These coefficients update the polar memory as

$$V_{i,k}' = V_{i,k} + \gamma \sum_{j\in\mathcal{N}(i)} [a_{ij,k} r_{ij} + b_{ij,k}\overline{w}_i]\,.$$

Here γ is a learned scalar gain. This variant returns a zero direct-force proposal. Its force is defined only by differentiating the scalar energy. Since this design is similar to WP-Inv and but it's design improved the toy results notably. We therefore used its invariant-core idea in the real system experiments, resulting in the WP-InvEq and WP-VecInv architectures, together with equivariant output maps.

WP-InvEq: invariant core with D_6-equivariant output maps. This model uses the same compact invariant edge message and scalar node update as WP-Inv, including the enabled joint edge–WA dot features. Its defining difference is the addition of symmetry-aware `we3nn` maps after the ordinary scalar/EGNN computation.

Writing the invariant message as m_{ij}, the provisional update and force are

$$\widetilde{h}_i = W_{\text{res}} h_i + U([h_i, \mathcal{A}_i^{\text{sum}}(m)])\,, \qquad \widetilde{f}_i = \mathcal{A}_i^{\text{sum}}\left([\alpha(m_{ij}) r_{ij,xy},\ \beta(m_{ij})]\right).$$

The H scalar channels are interpreted as H copies of the D_6 trivial representation. They are mapped to copies of the D_6 regular representation, passed through a representation-compatible

ReLU, and mapped back to trivial scalars. Likewise, the provisional force $[f_x, f_y, f_z]$ is interpreted as the direct sum of the planar standard representation and one trivial z scalar, mapped through a regular-representation hidden field, activated, and mapped back to vector plus scalar output.

$$h_i' = \mathcal{L}_{n_h \mathrm{Reg}(D_6) \to 1^H} \left(\sigma_{\mathrm{Reg}} \left(\mathcal{L}_{1^H \to n_h \mathrm{Reg}(D_6)} (\widetilde{h}_i) \right) \right),$$
$$f_i = \mathcal{L}_{n_f \mathrm{Reg}(D_6) \to (V \oplus 1)} \left(\sigma_{\mathrm{Reg}} \left(\mathcal{L}_{(V \oplus 1) \to n_f \mathrm{Reg}(D_6)} (\widetilde{f}_i) \right) \right).$$

The integers n_h and n_f are the hidden regular-field multiplicities of the scalar and force post-maps, respectively.

Consequently, most feature extraction remains an inexpensive invariant MLP, while the final scalar and vector transformations are explicitly D_6 equivariant. No non-scalar state persists from one layer to the next.

WP-VecInv: invariant messages with persistent polar vectors. WP-VecInv is the alias for the `D6_p_mix_equivariant_advanced` model. It uses a persistent bank of polar vectors, while operating on scalar coefficients.

Its ordinary invariant message is

$$m_{ij} = M_{\mathrm{inv}}(u_{ij}^{D_6}),$$

with the same D_6 input $u_{ij}^{D_6}$ defined for WP-Inv. Let $V_i \in \mathbb{R}^{H \times 3}$ be the persistent vector state. The second message uses the seven contractions

$$c_{ij,k} = [\|V_{i,k}\|^2, \|V_{j,k}\|^2, V_{i,k} \cdot V_{j,k}, V_{i,k} \cdot \widehat{r}_{ij}, V_{j,k} \cdot \widehat{r}_{ij}, V_{i,k} \cdot \overline{w}_i, V_{j,k} \cdot \overline{w}_i], \qquad s_{ij} = S_{\mathrm{inv}}(\{c_{ij,k}\}_{k=1}^H) \in \mathbb{R}^H.$$

Message aggregation and the scalar node update are

$$h_i' = W_{\mathrm{res}} h_i + U_{\mathrm{mix}}([h_i, \mathcal{A}_i^{\mathrm{sum}}(m), \mathcal{A}_i^{\mathrm{sum}}(s)]).$$

An invariant map G_V produces the two length-H coefficient vectors $[(a_{ij,k})_{k=1}^H, (b_{ij,k})_{k=1}^H] = G_V(m_{ij})$. The persistent vector update is

$$V_{i,k}' = V_{i,k} + \gamma \sum_{j \in \mathcal{N}(i)} [a_{ij,k} r_{ij} + b_{ij,k} \overline{w}_i], \qquad \gamma(0) = 0.1.$$

Here γ is a learned scalar gain initialized to 0.1. Because $a_{ij,k}$ and $b_{ij,k}$ are invariant scalars, this update is covariant and preserves directional information between message-passing layers.

For the direct head, the learned invariant readout $F_{\mathrm{inv}}(m_{ij}) = [a_{ij}^{(r)}, a_{ij}^{(w)}, \beta_{ij}]$ forms the provisional output

$$\widetilde{f}_i = \mathcal{A}_i^{\mathrm{sum}} \left([a_{ij}^{(r)} r_{ij,xy} + a_{ij}^{(w)} w_i,\ \beta_{ij}] \right).$$

A final real-system `we3nn` map enforces the typed D_6 output:

$$f_i = \mathcal{L}_{\mathrm{Reg} \to (V \oplus 1)} \left[\sigma_{\mathrm{Reg}} \left(\mathcal{L}_{(V \oplus 1) \to \mathrm{Reg}} (\widetilde{f}_i) \right) \right].$$

In conservative training, f_i is discarded but the vector state remains active because its invariant contractions modify h_i' and hence the energy. In direct training, the force proposals are summed across layers. Unlike WP-InvEq, WP-VecInv does not pass the scalar node output through a regular-representation post-map.

WP-VecEqui: persistent regular fields with WA–edge tensor products. WP-VecEqui is the toy only architecture, denoted as `wa_advanced` model. Its persistent state is a finite-group regular field

$$S_i \in n_S \, \mathrm{Reg}(G),$$

and its edge computation remains equivariant before being projected to scalar or vector outputs. Here n_S is the persistent-state multiplicity. The WA encoder receives the WA vector, its circular harmonics, its squared magnitude, and, for nonsymmorphic groups, the regular member encoding:

$$W_{ij} = E_{\mathrm{WA}}^G \left(w_i, \mathcal{Y}_G(w_i), \|w_i\|^2, \rho_i \right) \in n_R \, \mathrm{Reg}(G).$$

Here $\mathcal{Y}_G$ is the group-adapted circular-harmonic feature map, E^G_{WA} is a learned equivariant encoder, and n_R is the regular-field multiplicity used by the edge paths. The edge encoder receives typed planar geometry together with invariant chemistry/radial/vertical features and the receiver/sender persistent states:

$$R_{ij} = E^G_{\mathrm{edge}}\left(h_i, h_j, r_{ij,xy}, \mathcal{Y}_G(r_{ij,xy}), \|r_{ij}\|^2, B_{\mathrm{toy}}(d_{ij}), r_{ij,z}, S_i, S_j\right) \in n_R\,\mathrm{Reg}(G).$$

The map E^G_{edge} is the corresponding learned equivariant edge encoder. The two regular fields interact through an explicit UVU tensor product:

$$C_{ij} = \sigma_{\mathrm{Reg}}(\mathrm{TP}_{\mathrm{UVU}}(R_{ij}, W_{ij})) \in n_R\,\mathrm{Reg}(G).$$

Three equivariant paths contribute to the scalar edge message:

$$z_{ij} = \mathcal{L}_{W\to 1^H}(W_{ij}) + \mathcal{L}_{R\to 1^H}(R_{ij}) + \gamma_C \mathcal{L}_{C\to 1^H}(C_{ij}).$$

The field $z_{ij} \in \mathbb{R}^H$ is the scalar edge message. The subscripts W, R, and C identify independent projections of the WA, edge, and coupled fields. The scalar γ_C is a learned coupling gain. Degree-normalized aggregation is followed by an ordinary scalar node update:

$$h_i' = W_{\mathrm{res}} h_i + U\left([h_i, \mathcal{A}_i^{\mathrm{dn}}(z)]\right).$$

The persistent regular state receives a residual equivariant update:

$$\begin{aligned}\bar{R}_i &= \mathcal{A}_i^{\mathrm{dn}}\left(W_{ij} + R_{ij} + \gamma_C C_{ij}\right),\\ S_i' &= S_i + \eta_S\,\sigma_S\left(\mathcal{L}_{\mathrm{Reg}\to S}(\bar{R}_i)\right), \qquad \eta_S(0) = 0.1.\end{aligned}$$

Here η_S is a learned residual-state gain initialized to 0.1, and σ_S is the representation-compatible state activation. Finally, regular-to-vector and regular-to-trivial projections produce the direct force:

$$\begin{aligned}f_{ij,xy} &= \mathcal{L}_{W\to V}(W_{ij}) + \mathcal{L}_{R\to V}(R_{ij}) + \gamma_F \mathcal{L}_{C\to V}(C_{ij}),\\ f_{ij,z} &= \mathcal{L}_{W\to 1}(W_{ij}) + \mathcal{L}_{R\to 1}(R_{ij}) + \gamma_F \mathcal{L}_{C\to 1}(C_{ij}),\\ f_i &= \mathcal{A}_i^{\mathrm{dn}}([f_{ij,xy}, f_{ij,z}]).\end{aligned}$$

The scalar γ_F is a learned force-coupling gain.

This architecture shares many design choices with WP-Reg and WP-RegTP, while performing consistently worse, so we did not use it in the real-system experiments.

WP-Equi: fully D_6-equivariant edge-message model. WP-Equi moves the edge computation itself into `we3nn`. Its input field is a direct sum of:

- trivial representations for h_i, h_j, $\|r_{ij}\|^2$, the 16 radial features, $r_{ij,z}$, and the two enabled joint edge–WA contractions.
- one D_6 standard representation for w_i.
- one D_6 standard representation for $r_{ij,xy}$.

This typed edge field passes through four D_6-equivariant linear maps. The first three are followed by equivariant ReLU activations, and the hidden fields use copies of the D_6 regular representation. Separate equivariant projections produce H trivial scalar message channels, one planar standard vector, and one trivial z scalar. Scalar messages are aggregated with a residual scalar node projection, vector outputs are summed at the receiver.

More explicitly, if q_{ij} is the typed direct sum listed above,

$$\begin{aligned}R_{ij} &= \mathcal{M}_{D_6}(q_{ij}) \in n_{\mathrm{reg}}\mathrm{Reg}(D_6),\\ m_{ij} &= \mathcal{L}_{\mathrm{Reg}\to 1^H}(R_{ij}),\\ h_i' &= W_{\mathrm{res}} h_i + \sum_{j\in\mathcal{N}(i)} m_{ij},\\ f_i &= \mathcal{A}_i^{\mathrm{sum}}([\mathcal{L}_{\mathrm{Reg}\to V}(R_{ij}), \mathcal{L}_{\mathrm{Reg}\to 1}(R_{ij})]).\end{aligned}$$

Here $\mathcal{M}_{D_6}$ denotes the four-layer equivariant edge network and n_{reg} its hidden regular-representation multiplicity. In contrast to WP-Inv, there is no separate nonlinear receiver MLP: the nonlinear computation occurs before aggregation in the typed edge network. Unlike the advanced WA models, WP-Equi has no persistent regular or tensor state. Its advantage over WP-InvEq is that WA and planar edge vectors enter a fully typed D_6-equivariant message network rather than first being reduced to compact scalar invariants. Toy `p_molecule_full_equivariant` applies the same idea for an arbitrary wallpaper point group, with all WA/member fields and periodic geometry. Its scalar receiver update additionally uses a nonlinear node MLP, $h_i' = W_{\mathrm{res}} h_i + U([h_i, \mathcal{A}_i^{\mathrm{sum}}(m)])$, whereas the real D_6 implementation adds the projected scalar message directly as written above. The input and radial featurization also differ as described in the initial-feature and radial-basis paragraphs above.

WP-Reg: regular-representation WA state with gated paths. WP-Reg uses explicit D_6 regular representations for edge messages and for a persistent node state $S_i \in n_S \mathrm{Reg}(D_6)$, where n_S is its multiplicity.

Circular harmonics of planar edge and WA vectors are evaluated using a smooth zero-vector prescription. Nontrivial angular components are multiplied by $\|v\|^2/(\|v\|^2 + 10^{-4})$, which suppresses singular angular derivatives near the origin. Separate equivariant encoders map receiver WA, sender WA, and edge features to regular fields. UVU tensor products form receiver-edge, sender-edge, and state-conditioned couplings.

Six regular-message paths are retained: receiver WA, sender WA, encoded edge, lifted state, receiver coupling, and sender/state coupling. An invariant MLP produces channel-wise sigmoid gates for these paths. The gates are scaled by two and initialized to one. The gated paths are summed with a $1/\sqrt{6}$ normalization.

In equations, let

$$\begin{aligned} R_i &= E_{\mathrm{WA}}(w_i), \quad R_j = E_{\mathrm{WA}}(w_j), \quad E_{ij} = E_{\mathrm{edge}}(h_i, h_j, r_{ij}, B(d_{ij}), S_i, S_j), \\ S_{ij} &= \mathcal{L}_i(S_i) + \mathcal{L}_j(S_j), \\ C_i &= \mathrm{TP}_{\mathrm{UVU}}(E_{ij}, R_i), \quad C_j = \mathrm{TP}_{\mathrm{UVU}}(E_{ij}, R_j), \\ C_{S,ij} &= \mathrm{TP}_{\mathrm{UVU}}(C_i + C_j, S_{ij}). \end{aligned}$$

Here E_{WA} and E_{edge} are equivariant encoders, $\mathcal{L}_i$ and $\mathcal{L}_j$ are learned state lifts, and all quantities in this display are regular fields. If $P_{ij}^{(p)} \in \{R_i, R_j, E_{ij}, S_{ij}, C_i, C_j + C_{S,ij}\}$ and the invariant context produces gates $g_{ij}^{(p)}$, the edge message is

$$M_{ij} = \frac{1}{\sqrt{6}} \sum_{p=1}^{6} 2\sigma(g_{ij}^{(p)}) \odot P_{ij}^{(p)}.$$

The index p enumerates the six paths, σ is the logistic sigmoid, and each gate has the channel multiplicity of its associated regular field.

The regular message is projected to trivial scalar channels for the node update. The persistent regular state receives a residual update with learned per-state-channel gains initialized to 0.1. Direct planar force prediction uses three covariant bases (edge, receiver WA, sender WA), invariant basis weights, and a final D_6-equivariant linear readout. The vertical component combines the regular message with an invariant edge-z basis and terminates in a trivial representation. Let $\delta_{ij} \in \mathbb{R}^H$ denote the direct invariant scalar edge path and let γ_S be the learned vector of residual-state gains, initialized component-wise to 0.1. Then

$$\begin{aligned} h_i' &= W_{\mathrm{res}} h_i + U\big([h_i, \mathcal{A}_i^{\mathrm{dn}}(\mathcal{L}_{\mathrm{Reg}\to 1^H}(M_{ij}) + \delta_{ij})]\big), \\ S_i' &= S_i + \gamma_S \odot \mathcal{L}_{\mathrm{Reg}\to S}\big(\mathcal{A}_i^{\mathrm{dn}}(M_{ij})\big), \\ f_i &= \mathcal{A}_i^{\mathrm{dn}}\left(\mathcal{F}(M_{ij}, r_{ij,xy}, w_i, w_j, r_{ij,z})\right). \end{aligned}$$

Here $\mathcal{F}$ denotes the learned typed readout that returns a planar standard vector and a trivial z scalar.

The real-system configuration leaves the optional compact joint edge–WA invariant off. edge–WA interactions are instead represented by the explicit equivariant coupling paths.

Toy `wa_advanced_sr` implements the same six-path, gated regular-state construction for each selected wallpaper group. It generalizes the representation types, number of WA vectors, and optional nonsymmorphic member field, it also uses periodic edges and the toy radial and node encodings.

WP-RegTP: regular fields with explicit tensor products. This is a denser finite-group architecture. The model constructs three principal regular fields:

1. A direct field from scalar invariants, receiver/sender persistent states, $r_{ij,xy}$, and w_i.
2. An edge field from the planar edge vector and smooth circular harmonics.
3. A WA field from the WA vector and smooth circular harmonics.

An explicit D_6 tensor product couples the edge and WA fields. A second tensor product couples this result to the lifted persistent state. The direct, edge, WA, edge–WA, and state-coupled paths are concatenated and mixed by two equivariant regular-to-regular layers with SiLU activations.

Let D_{ij}, E_{ij}, and W_i denote the direct, edge, and WA regular fields, and let S_i be the persistent regular state. The complete message is

$$\begin{aligned} C_{ij} &= \mathrm{TP}_{D_6}(E_{ij}, W_i), \\ Q_{ij} &= \mathrm{TP}_{D_6}(D_{ij} + C_{ij}, \mathcal{L}_S(S_i + S_j)), \\ M_{ij} &= \mathrm{Mix}_{D_6}([D_{ij}, E_{ij}, W_i, C_{ij}, Q_{ij}]). \end{aligned}$$

Here $\mathcal{L}_S$ lifts the receiver and sender states into the tensor-product field, TP_{D_6} is the learned D_6-equivariant tensor product, and Mix_{D_6} is the subsequent equivariant channel mixing network.

The scalar update has a strong invariant baseline. It combines the old scalar state, a direct invariant edge aggregation, and invariant norms obtained by decomposing the regular field into D_6 irreducible representations. A separate regular-derived scalar correction is multiplied by a learned gain initialized to zero. This initialization makes early optimization behave like the stable invariant baseline and allows the symmetry-rich correction to enter gradually. Let $\delta_{ij} \in \mathbb{R}^H$ be the direct invariant scalar edge path and define its receiver aggregation $\delta_i = \mathcal{A}_i^{\mathrm{sum}}(\delta)$. Then

$$\begin{aligned} \bar{M}_i &= \mathcal{A}_i^{\mathrm{dn}}(M_{ij}), \\ h_i^{\mathrm{base}} &= W_{\mathrm{base}} h_i \\ &\quad + U_{\mathrm{base}}\big([h_i, \delta_i, E_{\mathrm{norm}}(\mathrm{InvNorm}(\bar{M}_i))]\big), \\ h_i' &= h_i^{\mathrm{base}} + \eta\, U_{\mathrm{reg}}\big([h_i, \mathcal{A}_i^{\mathrm{dn}}(\mathcal{L}_{\mathrm{Reg}\to 1^H} M_{ij})]\big), \\ \eta(0) &= 0, \qquad S_i' = \mathcal{L}_{\mathrm{Reg}\to S}(\bar{M}_i), \\ f_i &= \mathcal{A}_i^{\mathrm{dn}}\left(\mathcal{L}_{\mathrm{Reg}\to(V\oplus 1)}(M_{ij})\right). \end{aligned}$$

The maps W_{base}, U_{base}, E_{norm}, and U_{reg} are learned scalar maps. The scalar η is the regular correction gain, whose initialization is denoted by $\eta(0) = 0$. $\mathcal{L}_{\mathrm{Reg}\to S}$ replaces rather than residually updates the persistent state.

The persistent state is replaced after every message. Planar and vertical force components are equivariant projections of the regular edge message. Compact D_6 edge and WA features and the duplicated edge–WA dot product are always included by this implementation, regardless of the external joint-feature flag. Toy `wa_linear_mod` uses this same direct field–geometry field–WA field–tensor-product design and the same zero-initialized regular correction, but instantiates it for arbitrary wallpaper groups with the toy periodic graph and group-dependent registry inputs.

WP-O3: coupled $O(3)$ and D_6 processing. This hybrid model combines a full $O(3)$-equivariant UVU tensor-state layer with the WP-Reg D_6 wallpaper layer. Its purpose is to preserve general 3D geometric interactions while injecting the known planar D_6 structure and WA conditioning.

The $O(3)$ branch uses $1o$, $1e$, and $2e$ families, the D_6 wallpaper sublayer uses regular hidden fields and regular state fields. A separate three-component polar feedback vector is retained between layers.

Within one layer:

1. The previous D_6 feedback vector is embedded into the $O(3)$ tensor state with a learned equivariant map and gain initialized to 0.02.
2. The $O(3)$ UVU layer updates scalar and tensor states and proposes an $O(3)$ polar force.
3. The $O(3)$ tensor irreps and force are restricted to the planar D_6 subgroup and equivariantly encoded as the initial condition for the wallpaper state.
4. The D_6 advanced scalar/regular layer performs WA-conditioned message passing.
5. The difference between the resulting wallpaper state and its condition is projected to a new polar feedback vector for the next layer.
6. D_6-equivariant maps combine the two scalar outputs and the two force proposals.

Writing the incoming state as (T_i, p_i), where T_i is the $O(3)$ tensor state and p_i the polar feedback vector, one layer can be summarized as

$$\begin{aligned}
\widetilde{T}_i &= T_i + \gamma_{\mathrm{fb}} \mathcal{L}_{1o\to T}(p_i), \\
(h_i^{O3}, T_i', f_i^{O3}) &= \mathrm{UVU}_{O(3)}(h_i, \widetilde{T}_i, \{r_{ij}\}_j), \\
C_i &= \mathrm{Cond}_{D_6}\left(\mathrm{Res}_{O(3)\downarrow D_6}[T_i', f_i^{O3}]\right), \\
(h_i^W, S_i^W, f_i^W) &= \mathrm{WPReg}(h_i^{O3}, \{r_{ij}, w_i\}_j, C_i), \\
p_i' &= \mathcal{L}_{S\to 1o}(S_i^W - C_i), \\
h_i' &= \mathcal{L}_{\mathrm{scal}}[h_i^{O3}, h_i^W], \qquad f_i' = \mathcal{L}_{\mathrm{polar}}[f_i^{O3}, f_i^W].
\end{aligned}$$

Here γ_{fb} is the learned feedback gain initialized to 0.02. $\mathcal{L}_{1o\to T}$ embeds a polar vector into the $O(3)$ tensor-state type. $\mathrm{UVU}_{O(3)}$ denotes the complete $O(3)$ message-passing sublayer. The map $\mathrm{Res}_{O(3)\downarrow D_6}$ restricts $O(3)$ irreps to the planar subgroup. The map Cond_{D_6} encodes the restricted field as the initial wallpaper-state condition C_i. The maps $\mathcal{L}_{S\to 1o}$, $\mathcal{L}_{\mathrm{scal}}$, and $\mathcal{L}_{\mathrm{polar}}$ are learned equivariant output maps. The wallpaper condition C_i and state S_i^W have the same representation type, so their difference is well defined. The $O(3)$ operation in the second line is itself edge message passing. With $a_{ij} = Y_{\ell\le 2}(\widehat{r}_{ij})$,

$$q_{ij} = \mathrm{TP}_{\mathrm{UVU}}\Big(P[h_j, \widetilde{T}_j], a_{ij}, W(B(d_{ij}))\Big), \qquad q_i = \mathcal{A}_i^{\mathrm{sum}}(q_{ij}).$$

Here $Y_{\ell\le 2}$ is the spherical-harmonic edge embedding, P is a learned equivariant projection to the tensor-product input type, and W is the radial MLP that generates tensor-product weights. The scalar and tensor parts of q_i produce the $O(3)$ state update shown above.

The model therefore realizes bidirectional interaction: 3D $O(3)$ features condition D_6 processing, and a D_6 polar summary feeds back into the next $O(3)$ layer.

The toy architecture `wa_advanced_o3_feedback` has the same $O(3)$-to-wallpaper conditioning and wallpaper-to-$O(3)$ feedback topology. Its $O(3)$ branch, however, uses a fully connected tensor product rather than the real model's factorized UVU product. It is also wallpaper-group generic, periodic, and can include registry fields.

Scalar full-cell GNN baseline with EGNN-style force head. This model operates on all adsorbate and graphene atoms and deliberately uses neither D_6 features nor WA encoding. Its edge message is

$$m_{ij} = \mathrm{MLP}\big([h_i, h_j, \|r_{ij}\|^2, B(d_{ij})]\big),$$

followed by summed aggregation and the residual scalar update

$$h_i' = W_{\mathrm{res}} h_i + U([h_i, \mathcal{A}_i^{\mathrm{sum}}(m)]).$$

The direct vector proposal is the standard EGNN form

$$f_i = \sum_{j\in\mathcal{N}(i)} \alpha(m_{ij}) r_{ij}.$$

Because r_{ij} is a full 3D relative vector and α is invariant, the proposal is $O(3)$ equivariant. The scalar energy stream is invariant to global rotations, reflections, and translations (apart from the separately concatenated global cell representation in the common readout). Toy `full_graph_invariant` has the same EGNN-style edge message, scalar receiver update, and edge-weighted 3D force. It differs in the common toy input, radial, and periodic-edge construction. Here $\alpha : \mathbb{R}^H \to \mathbb{R}$ is a learned invariant scalar readout.

E3nn full-cell model with $\ell_{\max} = 1$ geometric features. This full-cell e3nn model carries scalar $0e$ channels and persistent $1o$ and $1e$ representations. Edge attributes are normalized spherical harmonics through $\ell = 1$, i.e., $0e + 1o$. Distances are represented by the common 16-function smooth radial basis.

To control memory and arithmetic cost, wide node states are projected once per node to a narrow tensor-product state. All symmetry-allowed products between this state and the edge spherical harmonics are built with UVU tensor products. In UVU mode the output multiplicity is tied to the input multiplicity, avoiding the cubic multiplicity scaling of a fully connected UVW product. A radial MLP produces edge-specific tensor-product weights.

The narrow equivariant edge products are summed before a wide equivariant channel-mixing map. Scalar updates use the old scalar state, scalar messages, and smooth norms of tensor messages. Tensor updates use sigmoid gates generated from scalar features and a residual equivariant self-map. A separate UVU path selects a polar $1o$ output for the direct force proposal.

Let T_i denote the concatenated $1o + 1e$ state, $a_{ij} = Y_{\ell \le 1}(\widehat{r}_{ij})$, and $n_j = P[h_j, T_j]$ the narrow projected node field. Here $Y_{\ell \le 1}$ is the normalized spherical-harmonic embedding and P is a learned equivariant projection. The message and force radial networks W_m and W_F generate the weights for their respective tensor products. The message and receiver updates are

$$
\begin{aligned}
q_{ij} &= \mathrm{TP}_{\mathrm{UVU}}(n_j, a_{ij}, W_m(B(d_{ij}))), \\
q_i &= \mathcal{A}_i^{\mathrm{sum}}(q_{ij}), \qquad (q_i^0, q_i^T) = \mathrm{Mix}(q_i), \\
h_i' &= L_0\big(L_{\mathrm{self}}^0 h_i + U([h_i, q_i^0, \mathrm{InvNorm}(q_i^T)])\big), \\
T_i' &= L_T\big(L_{\mathrm{self}}^T T_i + \sigma(G([h_i, q_i^0])) \odot q_i^T\big), \\
f_i &= L_F \mathcal{A}_i^{\mathrm{sum}}(\mathrm{TP}_{\mathrm{UVU}}(n_j, a_{ij}, W_F(B(d_{ij})))).
\end{aligned}
$$

In this display, Mix is the post-aggregation equivariant channel mixer. The fields q_i^0 and q_i^T are its scalar and non-scalar outputs. The maps L_0, L_T, and L_F are learned equivariant output maps. The maps L_{self}^0 and L_{self}^T are residual self-maps. Here U is the scalar update MLP, G is the scalar gate network, σ is the logistic sigmoid, and $\odot$ applies the resulting gates channel-wise to q_i^T. The last projection selects a polar $1o$ vector. The UVU intermediate is aggregated before the wide channel-mixing map, which is algebraically valid because that map is linear and avoids constructing a wide field per edge.

Including both polar $1o$ and axial $1e$ vector fields lets the model distinguish their parity under inversion. With $\ell_{\max} = 1$ edge harmonics, higher-order directional structure must emerge through stacked products and persistent node states rather than entering directly as $\ell = 2$ edge features. Toy `full_graph_e3nn_lmax1_tensor_product` uses the same $0e + 1o$ edge harmonics, persistent $1o + 1e$ state families, gated tensor update, and polar force type, but its edge tensor product is fully connected rather than the real model's narrow UVU factorization.

E3nn full-cell model with $\ell_{\max} = 2$ geometric features. This architecture has the same factorized UVU design as the $\ell_{\max} = 1$ model but extends both edge geometry and persistent state through $\ell = 2$. It uses $0e$, $1o$, $1e$, and $2e$ components for the node state. Edge attributes contain spherical harmonics $0e + 1o + 2e$.

Radial weights, pre-aggregation UVU products, post-aggregation equivariant channel mixing, smooth tensor norms, scalar gates, and the polar $1o$ force path otherwise follow the $\ell_{\max} = 1$ architecture.

With T_i now containing $1o + 1e + 2e$ and $a_{ij} = Y_{\ell\leq 2}(\widehat{r}_{ij})$, the symbols P, W_m, W_F, Mix, L_0, L_T, L_F, L^0_{self}, L^T_{self}, U, and G retain the definitions given for the $\ell_{\max} = 1$ model. Its equations are

$$
\begin{aligned}
q_{ij} &= \mathrm{TP}_{\mathrm{UVU}}(P[h_j, T_j], a_{ij}, W_m(B(d_{ij}))), \quad (q_i^0, q_i^T) = \mathrm{Mix}(\mathcal{A}_i^{\mathrm{sum}}(q_{ij})), \\
h'_i &= L_0\big(L^0_{\text{self}} h_i + U([h_i, q_i^0, \mathrm{InvNorm}(q_i^T)])\big)\,, \\
T'_i &= L_T\big(L^T_{\text{self}} T_i + \sigma(G([h_i, q_i^0])) \odot q_i^T\big)\,, \\
f_i &= L_F \mathcal{A}_i^{\mathrm{sum}}(\mathrm{TP}_{\mathrm{UVU}}(P[h_j, T_j], a_{ij}, W_F(B(d_{ij}))))\,.
\end{aligned}
$$

The $2e$ channels represent rank-2 symmetric traceless information and permit quadrupolar angular structure to be communicated directly in one layer. This increases angular expressivity and state memory relative to $\ell_{\max} = 1$, while the smaller scalar width keeps the total registered parameter count near one million. Toy `full_graph_e3nn_tensor_product` carries the corresponding $1o + 1e + 2e$ state and $\ell \leq 2$ edge harmonics, but again uses a fully connected tensor product rather than the real narrow UVU schedule. Its geometry and initial features follow the common toy construction, whereas the real model uses the full cached cell and the real feature/radial definitions.

G.5 Symmetry summary

Table 26 summarizes how each architecture handles the symmetry. Here G denotes the selected wallpaper point group. For the real-system molecule-only models, $G = D_6$. "Persistent state" means a non-scalar node state passed between message-passing layers, rather than a temporary equivariant edge field. The final column describes the explicit force proposal used in direct-force-head training. In conservative training this proposal is discarded and forces are obtained as $F = -\nabla_x E$.

Table 26: Symmetry summary of the WallpaperNet architectures and the full-cell baselines. For each model, the columns give the message-passing graph, the experiments in which it is evaluated, the scalar-message and update mechanism, the persistent equivariant node state, and the explicit force proposal used in direct-force-head training.

Architecture	Graph	Evaluated on	Scalar-message and update mechanism	Persistent equivariant node state	Direct-force construction
Molecule-only architectures					
WP-Inv	molecule	toy and real	Invariant MLP messages built from compact G-invariants, invariant receiver update	none	Invariant coefficient multiplying the planar edge vector, together with a trivial-representation z scalar
WP-InvExt	molecule	toy only	Invariant scalar MLP, the conservative variant additionally feeds invariant contractions of its vector memory into the scalar update	conservative: H polar three-vectors, direct: none	direct variant: invariant coefficients multiplying the planar edge and WA vectors, plus a trivial z scalar, conservative variant: no direct proposal
WP-InvEq	molecule	real only	Invariant D_6 core followed by an explicit $1 \to \mathrm{Reg} \to 1$ scalar post-map	none	Provisional edge-weighted force followed by $(V \oplus 1) \to \mathrm{Reg} \to (V \oplus 1)$ equivariant post-maps
WP-Equi	molecule	toy and real	Fully typed G-equivariant edge network, trivial-representation projections provide scalar messages	none	Standard-vector V and trivial-z projections of the equivariant edge field
WP-VecInv	molecule	real only	Invariant D_6 messages augmented by invariant contractions of persistent polar vectors	H polar three-vectors with a residual covariant update	Invariant edge/WA basis coefficients followed by a $(V \oplus 1) \to \mathrm{Reg} \to (V \oplus 1)$ output network
WP-VecEqui	molecule	toy only	Separate regular edge and WA fields coupled by an explicit UVU tensor product, trivial projections update the scalar stream	residually updated G-regular state	Regular-to-standard-vector and regular-to-trivial projections of the WA, edge, and coupled fields

Table 26: Symmetry summary of the WallpaperNet architectures and the full-cell baselines (continued).

Architecture	Graph	Evaluated on	Scalar-message and update mechanism	Persistent equivariant node state	Direct-force construction
WP-Reg	molecule	toy and real	Six gated regular-field paths with explicit UVU couplings, followed by trivial projections and an invariant scalar update	residually updated G-regular state	Equivariant readout using the planar edge, receiver-WA, and sender-WA covariant bases, plus a trivial z output
WP-RegTP	molecule	toy and real	Direct, edge, and WA regular fields coupled by tensor products, irrep norms and a zero-initialized regular correction feed the scalar update	G-regular state replaced by a projection of the aggregated message	Equivariant G-regular-to-$(V \oplus 1)$ projection of the edge message
WP-O3	molecule	toy and real	Coupled $O(3)$ tensor and wallpaper-group regular pathways, with bidirectional conditioning and scalar fusion	$O(3)$ $1o + 1e + 2e$ tensor state plus a wallpaper-derived polar feedback vector	Equivariant fusion of the $O(3)$ polar proposal and the wallpaper-group force proposal
Full-cell baselines					
Full-cell invariant GNN	full cell	toy and real	$O(3)$-invariant scalar EGNN messages and an invariant receiver update	none	Invariant coefficient multiplying the full three-dimensional edge vector
Full-cell e3nn ($\ell_{\max} = 1$)	full cell	toy and real	$O(3)$-equivariant tensor products with $0e + 1o$ edge harmonics, $0e$ projections and invariant tensor norms update the scalar stream	persistent $1o + 1e$ state	Equivariant projection to a polar $1o$ vector
Full-cell e3nn ($\ell_{\max} = 2$)	full cell	toy and real	$O(3)$-equivariant tensor products with $0e + 1o + 2e$ edge harmonics, $0e$ projections and invariant tensor norms update the scalar stream	persistent $1o + 1e + 2e$ state	Equivariant projection to a polar $1o$ vector

G.6 Experimental comparability and interpretation

All models use four message-passing layers, the same initial atomic attributes, the same radial cutoff and basis size, and approximately one million registered parameters. Nevertheless, the comparison intentionally includes two distinct information regimes. Molecule-only models receive WA vectors and an explicit graphene reference energy but do not propagate through graphene atoms. Full-cell models receive all graphene coordinates and species directly and do not use WA encoding or a separate graphene graph reference. Their graph sizes and edge counts are therefore much larger.

The three real-system scenarios answer different questions. Standard combined training jointly tests full-total-energy prediction and conservative forces. Because the real system has a varying number of atoms in each cell, its total energy varies substantially across samples. We therefore add a force-weighted conservative experiment in which the energy loss is weighted by a data-derived coefficient of approximately 1.22×10^{-6} relative to a unit force weight (Appendix I), which balances the energy and force terms while retaining $F = -\nabla E$. Force-head training tests direct equivariant vector regression, which avoids second derivatives during training and can be substantially faster, but does not enforce a globally integrable energy/force field.

Finally, the graph-level energy readout concatenates raw Cartesian cell-vector components embedded as raw features. Those components are frame-aligned in the present dataset and held fixed in the present dataset convention. Since every cell has its own size, the total energy of the cell is dependent on the cell size as well. Therefore the cell vectors are included in the energy readout to allow the model to learn the energy dependence on the cell size.

G.7 TRAINING METRICS

Let $\mathcal{M}_{\text{eval}}$ denote the set of all molecular atoms in the evaluated split, $F_i \in \mathbb{R}^3$ the reference force on atom i, and F_i^{pred} its predicted force. Support-atom forces are excluded from all toy force metrics.

The reported relative force L_2 error is the norm over all molecular atoms and Cartesian components:

$$L_{2,\text{rel}}^F = \frac{\left[\sum_{i\in\mathcal{M}_{\text{eval}}} \left\|F_i^{\text{pred}} - F_i\right\|_2^2\right]^{1/2}}{\left[\sum_{i\in\mathcal{M}_{\text{eval}}} \|F_i\|_2^2\right]^{1/2}}.$$

Equivalently, after flattening all force components into the vectors f^{pred} and f, respectively,

$$L_{2,\text{rel}}^F = \frac{\|f^{\text{pred}} - f\|_2}{\|f\|_2}.$$

For K scalar targets y_k and predictions y_k^{pred}, the coefficient of determination is

$$R^2 = 1 - \frac{\sum_{k=1}^{K}(y_k^{\text{pred}} - y_k)^2}{\sum_{k=1}^{K}(y_k - \overline{y})^2}, \qquad \overline{y} = \frac{1}{K}\sum_{k=1}^{K} y_k.$$

For the reported force-component R^2, the index k runs over every x-, y-, and z-component of every molecular atom in the evaluated split. For an energy R^2, k runs over systems, using either the total energy or the specified atom-normalized energy.

The zero-force score compares the model with the baseline $F_i^{\text{pred}} = 0$:

$$S_{\text{zero}} = 1 - \frac{\sum_{i\in\mathcal{M}_{\text{eval}}} \left\|F_i^{\text{pred}} - F_i\right\|_2^2}{\sum_{i\in\mathcal{M}_{\text{eval}}} \|F_i\|_2^2}.$$

Consequently,

$$S_{\text{zero}} = 1 - \left(L_{2,\text{rel}}^F\right)^2.$$

A perfect prediction gives $S_{\text{zero}} = 1$, predicting zero force gives $S_{\text{zero}} = 0$, and a prediction worse than the zero-force baseline gives $S_{\text{zero}} < 0$.

G.8 TOY-MODEL TRAINING OBJECTIVES

Let $\mathcal{B}$ be a mini-batch of $B = |\mathcal{B}|$ systems. The energy loss is the mean absolute error of the unnormalized full-system total energies:

$$L_E = \frac{1}{B}\sum_{b\in\mathcal{B}} \left|E_b^{\text{pred}} - E_b\right|.$$

Let $\mathcal{M}_{\mathcal{B}}$ be the set of molecular atoms in the mini-batch. The force loss is the mean squared error over molecular atoms and Cartesian components:

$$L_F = \frac{1}{3|\mathcal{M}_{\mathcal{B}}|}\sum_{i\in\mathcal{M}_{\mathcal{B}}}\sum_{\alpha\in\{x,y,z\}} \left(F_{i\alpha}^{\text{pred}} - F_{i\alpha}\right)^2 = \frac{1}{3|\mathcal{M}_{\mathcal{B}}|}\sum_{i\in\mathcal{M}_{\mathcal{B}}} \left\|F_i^{\text{pred}} - F_i\right\|_2^2.$$

In conservative combined training, forces are obtained from the predicted total energy:

$$F_i^{\text{pred}} = -\frac{\partial}{\partial x_i}\sum_{b\in\mathcal{B}} E_b^{\text{pred}}.$$

Because the systems in a batch are disconnected, this is equivalent to differentiating the energy of the system containing atom i. Periodic edge displacements and WA vectors are reconnected to the atomic coordinates before taking this derivative.

The conservative combined objective is

$$\mathcal{L}_{\text{combined}} = \lambda_E L_E + \lambda_F L_F, \qquad \lambda_E = \lambda_F = 1,$$

and therefore

$$\mathcal{L}_{\text{combined}} = L_E + L_F.$$

In direct-force-head training, the four layer-wise equivariant force proposals are summed, as defined in the direct-force scenario above:

$$F_i^{\text{pred}} = \sum_{\ell=1}^{4} f_i^{(\ell)}.$$

The energy term is deliberately omitted. The objective evaluated by the training code is

$$\mathcal{L}_{\text{head}} = \lambda_F L_F = L_F.$$

Thus the force-head experiments neither optimize nor report an energy loss.

G.9 Toy-model optimization and training conditions

Every toy model was trained for a maximum of 1000 epochs. No early stopping was used, so an uninterrupted run executes all 1000 epochs. The checkpoint with the lowest validation loss was retained and used for the reported training, validation, and test metrics.

Optimization used AdamW (Loshchilov & Hutter, 2019) with initial learning rate

$$\eta_0 = 10^{-3}.$$

Only the learning rate was explicitly supplied to the optimizer. With the PyTorch version used for these experiments, the remaining AdamW parameters were

$$(\beta_1, \beta_2) = (0.9, 0.999), \qquad \epsilon_{\text{Adam}} = 10^{-8}, \qquad \lambda_{\text{wd}} = 10^{-2},$$

with AMSGrad disabled. Here ϵ_{Adam} is the numerical-stability constant and λ_{wd} is the weight-decay coefficient.

The learning rate was controlled by a `ReduceLROnPlateau` scheduler monitoring the epoch-level validation loss:

$$\eta \leftarrow 0.5\,\eta$$

after a plateau with patience 20 epochs. The scheduler used minimization mode, an absolute improvement threshold of 10^{-12}, zero cooldown, a minimum learning rate of 0, and a numerical update threshold of $\epsilon_{\text{sched}} = 10^{-8}$. It was evaluated once per epoch.

Each toy-model run used one GPU, a batch size of 64 systems, four message-passing layers, eight data-loader workers, and an edge cutoff of 6 Å. The data were split into training, validation, and test subsets in the ratio

$$0.8 : 0.1 : 0.1.$$

The training loader was shuffled, whereas the validation and test loaders were not. The experiment seed was used for both the data split and Lightning's model and worker seeding. The toy-model experiments used three independent seeds (24, 42, 67) for both the data split and model initialization.

G.10 Real-system optimization and training conditions

Similar training scenario was applied to the real-system models, unless explicitly stated otherwise. The real-system model was trained for a maximum of 600 epochs. Runs that reached the cluster wall-time limit were resumed from the most recent checkpoint, which was saved every five epochs, until all 600 epochs had been completed. The checkpoint with the lowest validation loss was retained and used for the reported training, validation, and test metrics.

Each real-system run used four GPUs with distributed data-parallel training, a global batch size of 64 systems, and a per-GPU batch size of 16. Gradient accumulation was not used. Every model used four message-passing layers, eight data-loader workers, an edge cutoff of 6 Å, and at most 128 neighbors per atom.

The 15089 systems were split into 12071 training, 1509 validation, and 1509 test systems, corresponding to the ratio

$$0.8 : 0.1 : 0.1.$$

The training loader was shuffled, whereas the validation and test loaders were not. Three independently seeded data partitions initial model seed where used(seeds: 1, 24, 42). For the some of the models we observed unstable gradient behavior during the first 40 epochs of the training. For four force-head models, the global L_2 norm of the gradients over all trainable parameters was clipped at every optimizer step. The clipping thresholds were 1 for WP-Equi, 20 for WP-Inv, 30 for WP-O3, and 50 for WP-Reg. All combined-loss models and all remaining force-head models were trained without gradient clipping. The global gradient norm before and after clipping was logged every 100 optimization steps.

H TOY BENCHMARK RESULTS FOR ALL WALLPAPER GROUPS

Remaining observations on the averaged results. Among the WallpaperNet models in Table 1, WP-InvExt is distinctly the second best, with the remaining equivariant architectures performing very similarly to each other. Direct force-head learning is also the only target for which the full-cell e3nn models outperform all WallpaperNet models. Increasing the angular resolution of the full-cell e3nn baseline from $\ell_{\max} = 1$ to $\ell_{\max} = 2$ provides no systematic improvement in the averaged results, and the two models have essentially identical force accuracy. Rank-two features therefore do not appear to be necessary for achieving the accuracy level reached by the full-cell baselines on this benchmark.

The group-resolved results largely support the trends obtained from averaging over all 17 wallpaper groups, but they also reveal a pronounced dependence on the symmetry of the individual system. Figure 4 summarizes the energy and force errors across groups, and Tables 27–43 report the corresponding numerical results. Across the WallpaperNet architectures, WP-O3 is the most consistently accurate model. It gives the lowest energy MAE among the WallpaperNet models for 16 of the 17 groups and the best combined-training force metrics for all 17 groups. The only exception for energy is cmm (Table 37), where WP-InvExt gives a slightly lower error. WP-InvExt is otherwise remarkably consistent: it is typically the second-best WallpaperNet model for energy and for forces obtained from conservative training. In contrast, increasing the complexity of the regular-representation pathways from WP-Equi through WP-VecEqui, WP-Reg, and WP-RegTP does not lead to a systematic improvement. This indicates that explicit wallpaper-group equivariance alone is not sufficient to guarantee higher accuracy. Rather, the combination of wallpaper-group processing with $O(3)$ tensor features used by WP-O3 appears to provide the most effective representation of the geometric interactions in this benchmark.

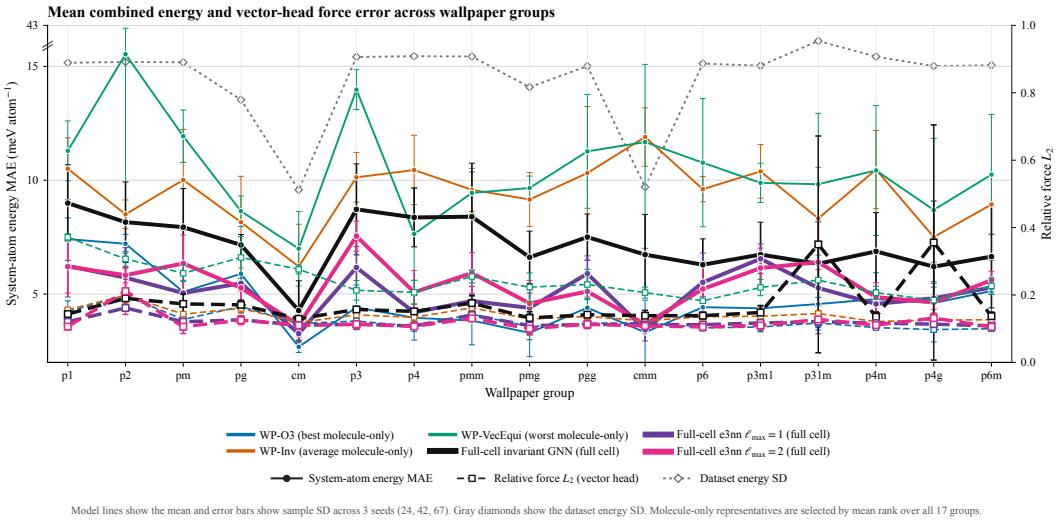


Figure 4: Group-resolved toy-benchmark performance. Solid curves show the mean system-atom energy MAE from combined training (left axis), and dashed curves show the mean relative force L_2 error from vector-head training (right axis). Error bars denote the sample standard deviation across three seeds. WP-O3, WP-Inv, and WP-VecEqui are respectively the best-, average-, and worst-ranked molecule-only representatives across all groups, while all three full-cell baselines are shown. Gray diamonds indicate the standard deviation of the mean residual energy per atom within each group.

For the lowest-symmetry groups, the advantage of exploiting wallpaper symmetry is less pronounced. For $p1$ (Table 27), the full-cell e3nn models provide the lowest errors: the $\ell_{\max} = 2$ model gives the lowest energy MAE and the best vector-head force results, while the $\ell_{\max} = 1$ model performs best for forces from combined training. WP-O3 is nevertheless the strongest WallpaperNet architecture for this group, substantially improving over the other WP models. A similar behavior is observed for $p2$ (Table 28), where the full-cell e3nn $\ell_{\max} = 1$ model gives the best result for every reported accuracy metric. The $p2$ results also show comparatively large seed-to-seed variation for some of the regular-representation models, particularly WP-VecEqui and WP-Reg, suggesting less robust optimization for these architectures in this group.

The mirror and glide groups pm (Table 29) and pg (Table 30) show two somewhat different regimes. For pm, WP-O3 gives the best combined-training force prediction, with a relative force error of 18.8% and $R^2 = 0.964$, while its energy error of 5.09 meV atom^{-1} is essentially comparable to the 5.05 meV atom^{-1} obtained by the full-cell e3nn $\ell_{\max} = 1$ model. For vector-head training, however, the full-cell e3nn models retain an advantage. In pg, the full-cell e3nn $\ell_{\max} = 2$ model performs best across the principal metrics, although WP-O3 remains the best WallpaperNet architecture for combined training. Interestingly, WP-Inv gives the best vector-head force result among the WallpaperNet models for pg, marginally outperforming WP-O3. Thus, in these low-order reflection/glide groups, exploiting the wallpaper representation does not consistently outperform a full-cell $O(3)$-equivariant treatment.

The picture begins to change for cm (Table 31) and for the groups containing higher-order rotations. For cm, WP-O3 already gives the lowest energy MAE and the best conservative-force prediction, whereas the vector-head results of the full-cell e3nn models remain slightly better. For $p3$ (Table 32), WP-O3 reduces the energy MAE to 4.42 meV atom^{-1} and the combined relative force error to 15.4%, outperforming all full-cell models under combined training. The full-cell e3nn models retain a small advantage for the vector head. The same pattern occurs for $p4$ (Table 33): WP-O3 gives the best energy and combined-force performance and reaches a vector-head relative force error of 10.7%, equal to the best full-cell value, although the full-cell $\ell_{\max} = 2$ model has a slightly smaller component MAE. For $p6$ (Table 38), the difference narrows further: WP-O3 gives the best energy, combined-force, and vector-head relative-force errors, while its vector-head component MAE of 229.2 meV Å^{-1} is essentially comparable to the 227.5 meV Å^{-1} of the full-cell $\ell_{\max} = 2$ model.

These groups therefore show that the advantage of the symmetry-reduced representation becomes increasingly apparent once nontrivial rotational symmetry is present.

A similar trend is found for the twofold groups pmm (Table 34), pmg (Table 35), pgg (Table 36), and cmm (Table 37). WP-O3 gives the best combined-training force performance in all four cases and the lowest energy error for pmm, pmg, and pgg. For cmm, WP-InvExt instead gives the lowest energy MAE, 2.98 meV atom^{-1}, compared with 3.32 meV atom^{-1} for WP-O3. This is the only wallpaper group for which another WallpaperNet architecture outperforms WP-O3 in energy. Under vector-head training, the full-cell e3nn models remain slightly more accurate for these four groups. The differences are particularly small for pmg and pgg: for example, in pmg, WP-O3 reaches a relative vector-force error of 10.6%, compared with 10.0% for the full-cell e3nn $\ell_{\max} = 2$ model. Thus, even where the full-cell baseline remains numerically best for the direct force head, the symmetry-reduced WP-O3 model is already operating in essentially the same accuracy regime.

The strongest results for WallpaperNet are obtained for the higher-symmetry groups $p3m1$ (Table 39), $p31m$ (Table 40), $p4m$ (Table 41), $p4g$ (Table 42), and $p6m$ (Table 43). WP-O3 gives the best force metrics under both training modes for all five groups. For $p3m1$ and $p31m$, it also gives the lowest energy error, meaning that WP-O3 outperforms all three full-cell baselines simultaneously for energy, conservative forces, and direct vector-force prediction. The same complete dominance is observed for $p4g$ and $p6m$. In $p4g$, for example, WP-O3 reaches a vector-head relative force error of 9.7%, a component MAE of 207.0 meV Å^{-1}, and $R^2 = 0.990$, while its energy MAE of 4.61 meV atom^{-1} is also marginally below that of the full-cell e3nn $\ell_{\max} = 2$ model. For $p6m$, WP-O3 similarly gives 10.0%, 212.4 meV Å^{-1}, and $R^2 = 0.990$ for the vector head. The only exception within this set is the $p4m$ energy prediction, where full-cell e3nn $\ell_{\max} = 1$ gives 4.57 meV atom^{-1} compared with 4.78 meV atom^{-1} for WP-O3. The reported seed variations substantially overlap, while WP-O3 remains the best model for every force metric.

The group-resolved results also refine the comparison between conservative forces and the explicitly learned vector force head. Direct vector-force training improves the force metrics in the overwhelming majority of model–group combinations, but the improvement is not literally universal. For WP-VecEqui and WP-RegTP on $p1$ (Table 27), vector-head training produces larger force errors than combined training, and WP-RegTP shows a small degradation in relative force error and R^2 for pmm (Table 34). More pronounced failures occur for the full-cell invariant GNN on $p31m$ (Table 40) and $p4g$ (Table 42), where the vector-head results have both substantially larger mean errors and extremely large seed-to-seed variations. Apart from these isolated cases, the vector head consistently reduces force errors, confirming that direct force supervision is substantially easier for these strongly nonequilibrium synthetic structures than recovering forces through the energy gradient.

Taken together, the individual-group results therefore strengthen rather than merely reproduce the averaged comparison. The benefit of wallpaper-aware processing is smallest for groups with little nontrivial point-group symmetry, where full-cell $O(3)$-equivariant models frequently remain more accurate. As the amount of rotational, reflection, and glide symmetry increases, WP-O3 becomes increasingly competitive and, for several of the most symmetric groups, surpasses the full-cell baselines even for direct vector-force prediction. At the same time, the strong and consistent performance of WP-InvExt shows that much of the gain over the basic invariant model can already be obtained by enriching scalar message passing with covariant geometric directions. The poorer and more variable results of the intermediate regular-representation architectures further show that increasing equivariant architectural complexity by itself does not guarantee better prediction. Instead, the results favor architectures in which the reduced wallpaper representation is supplemented by geometrically expressive features, most clearly realized by WP-O3.

In Tables 27–43, entries report the mean $\pm$ sample standard deviation across three seeds, C and V denote combined and vector-head training, and the energy MAE is taken from combined training. Bold marks the best mean for each metric and training mode. Best epochs are not ranked.

Table 27: Toy-model performance for wallpaper group p1 (1,000 structures, mean $\Delta E = -7.12$ eV atom^{-1}, energy SD 17.42 meV atom^{-1}).

Model	Energy MAE ↓ (meV atom^{-1})	Mode	Rel. force L_2 ↓ (%)	Force comp. MAE ↓ (meV Å^{-1})	Force comp. R^2 ↑	Best epoch
WallpaperNet models						
WP-Inv	10.50 ± 1.35	C	32.6 ± 2.3	751.1 ± 58.6	0.893 ± 0.015	739.7 ± 231.2
		V	15.6 ± 0.6	341.6 ± 9.3	0.975 ± 0.002	304.7 ± 47.0
WP-InvExt	9.28 ± 2.37	C	28.2 ± 5.6	635.0 ± 123.5	0.918 ± 0.033	511.7 ± 121.5
		V	16.7 ± 1.1	363.0 ± 33.3	0.972 ± 0.004	320.7 ± 94.2
WP-Equi	12.45 ± 6.14	C	34.9 ± 4.9	798.3 ± 136.3	0.877 ± 0.033	510.7 ± 345.5
		V	26.8 ± 0.9	616.5 ± 29.1	0.928 ± 0.005	233.3 ± 25.1
WP-VecEqui	11.29 ± 1.31	C	34.3 ± 1.2	790.4 ± 52.2	0.883 ± 0.008	722.3 ± 107.9
		V	37.1 ± 1.0	849.6 ± 24.6	0.862 ± 0.008	236.3 ± 30.7
WP-Reg	9.29 ± 3.34	C	30.9 ± 5.3	706.8 ± 124.2	0.903 ± 0.033	837.3 ± 125.8
		V	22.6 ± 3.1	492.9 ± 67.0	0.948 ± 0.014	306.0 ± 124.4
WP-RegTP	12.50 ± 3.87	C	31.0 ± 4.2	728.7 ± 112.3	0.903 ± 0.027	628.0 ± 37.3
		V	32.3 ± 2.1	764.2 ± 39.0	0.896 ± 0.014	166.7 ± 6.7
WP-O3	7.41 ± 0.93	C	22.3 ± 0.4	523.2 ± 4.5	0.950 ± 0.002	534.0 ± 107.6
		V	15.1 ± 3.2	338.5 ± 62.3	0.977 ± 0.009	309.7 ± 82.1
Full-cell baselines						
Full-cell invariant GNN	9.00 ± 1.68	C	29.7 ± 4.7	673.5 ± 115.1	0.910 ± 0.027	839.3 ± 250.6
		V	14.3 ± 0.3	310.4 ± 10.7	0.979 ± 0.001	804.7 ± 129.7
Full-cell e3nn ($\ell_{\max} = 1$)	6.23 ± 1.36	C	$\mathbf{19.5 \pm 2.3}$	$\mathbf{450.4 \pm 48.4}$	$\mathbf{0.961 \pm 0.009}$	840.3 ± 267.0
		V	12.0 ± 2.1	266.7 ± 48.5	0.985 ± 0.005	425.3 ± 111.3
Full-cell e3nn ($\ell_{\max} = 2$)	$\mathbf{6.21 \pm 1.16}$	C	20.8 ± 1.8	482.8 ± 40.6	0.956 ± 0.007	912.3 ± 133.7
		V	$\mathbf{10.6 \pm 0.9}$	$\mathbf{232.9 \pm 15.1}$	$\mathbf{0.989 \pm 0.002}$	662.3 ± 29.1

Table 28: Toy-model performance for wallpaper group p2 (1,000 structures, mean $\Delta E = -7.15$ eV atom^{-1}, energy SD 18.03 meV atom^{-1}).

Model	Energy MAE ↓ (meV atom^{-1})	Mode	Rel. force L_2 ↓ (%)	Force comp. MAE ↓ (meV Å^{-1})	Force comp. R^2 ↑	Best epoch
WallpaperNet models						
WP-Inv	8.50 ± 0.63	C	35.9 ± 4.3	741.0 ± 84.1	0.870 ± 0.030	558.7 ± 149.1
		V	19.9 ± 2.3	400.9 ± 50.4	0.960 ± 0.009	210.7 ± 38.5
WP-InvExt	11.61 ± 3.68	C	31.5 ± 3.2	672.2 ± 63.2	0.900 ± 0.020	343.0 ± 133.6
		V	19.1 ± 1.6	381.7 ± 35.3	0.963 ± 0.006	291.7 ± 37.8
WP-Equi	12.78 ± 4.05	C	39.8 ± 2.0	820.9 ± 42.7	0.841 ± 0.015	361.3 ± 39.6
		V	23.1 ± 3.2	472.2 ± 45.1	0.946 ± 0.015	256.3 ± 38.9
WP-VecEqui	23.37 ± 17.65	C	40.1 ± 2.0	846.5 ± 36.3	0.839 ± 0.016	431.3 ± 268.3
		V	30.7 ± 2.2	638.6 ± 39.6	0.905 ± 0.013	239.0 ± 37.2
WP-Reg	17.61 ± 10.76	C	39.4 ± 3.4	820.4 ± 79.1	0.844 ± 0.026	474.0 ± 347.5
		V	23.7 ± 1.9	493.2 ± 54.4	0.944 ± 0.009	236.0 ± 41.5
WP-RegTP	10.67 ± 0.50	C	37.4 ± 3.2	783.8 ± 54.2	0.860 ± 0.023	399.0 ± 25.1
		V	34.0 ± 4.6	744.5 ± 79.0	0.883 ± 0.032	171.0 ± 46.7
WP-O3	7.21 ± 0.41	C	25.0 ± 1.5	532.0 ± 36.4	0.937 ± 0.007	488.3 ± 199.4
		V	19.8 ± 1.5	388.2 ± 27.8	0.961 ± 0.006	165.7 ± 41.7
Full-cell baselines						
Full-cell invariant GNN	8.16 ± 1.77	C	30.6 ± 4.1	614.6 ± 81.3	0.906 ± 0.024	611.0 ± 195.5
		V	19.0 ± 2.0	360.0 ± 23.5	0.964 ± 0.007	566.0 ± 82.8
Full-cell e3nn ($\ell_{\max} = 1$)	$\mathbf{5.72 \pm 1.32}$	C	$\mathbf{23.7 \pm 1.4}$	$\mathbf{491.3 \pm 16.3}$	$\mathbf{0.944 \pm 0.007}$	815.3 ± 283.8
		V	$\mathbf{16.2 \pm 0.7}$	$\mathbf{310.2 \pm 18.8}$	$\mathbf{0.974 \pm 0.002}$	342.0 ± 67.5
Full-cell e3nn ($\ell_{\max} = 2$)	5.83 ± 0.34	C	24.3 ± 3.4	495.9 ± 76.6	0.940 ± 0.016	925.3 ± 87.9
		V	21.1 ± 6.9	426.0 ± 194.5	0.952 ± 0.032	410.3 ± 234.8

Table 29: Toy-model performance for wallpaper group pm (1,000 structures, mean $\Delta E = -7.11$ eV atom^{-1}, energy SD 17.79 meV atom^{-1}).

Model	Energy MAE ↓ (meV atom^{-1})	Mode	Rel. force L_2 ↓ (%)	Force comp. MAE ↓ (meV Å^{-1})	Force comp. R^2 ↑	Best epoch
WallpaperNet models						
WP-Inv	10.01 ± 2.22	C	30.7 ± 2.8	696.5 ± 62.6	0.905 ± 0.017	807.7 ± 160.8
		V	14.3 ± 2.6	306.6 ± 47.2	0.979 ± 0.007	416.0 ± 177.8
WP-InvExt	6.48 ± 1.66	C	23.6 ± 1.7	537.9 ± 39.4	0.944 ± 0.008	469.7 ± 115.5
		V	13.9 ± 2.0	298.7 ± 44.2	0.980 ± 0.005	413.7 ± 42.9
WP-Equi	11.87 ± 2.49	C	31.9 ± 0.9	723.2 ± 36.3	0.898 ± 0.006	383.0 ± 88.5
		V	21.5 ± 2.0	484.5 ± 44.6	0.954 ± 0.008	182.7 ± 45.9
WP-VecEqui	11.93 ± 1.15	C	33.4 ± 1.1	748.2 ± 11.6	0.889 ± 0.007	663.3 ± 306.6
		V	26.5 ± 2.5	579.3 ± 40.5	0.930 ± 0.013	305.3 ± 4.9
WP-Reg	11.64 ± 2.48	C	33.9 ± 3.3	727.9 ± 28.1	0.884 ± 0.023	610.7 ± 110.6
		V	18.3 ± 1.4	383.9 ± 31.2	0.966 ± 0.005	281.3 ± 36.1
WP-RegTP	12.23 ± 2.45	C	33.5 ± 1.6	777.7 ± 39.1	0.888 ± 0.011	718.3 ± 171.0
		V	23.4 ± 4.1	518.9 ± 93.3	0.944 ± 0.019	303.0 ± 94.9
WP-O3	5.09 ± 0.51	C	$\mathbf{18.8 \pm 2.0}$	$\mathbf{430.6 \pm 40.5}$	$\mathbf{0.964 \pm 0.007}$	730.7 ± 274.9
		V	12.7 ± 0.9	288.6 ± 18.4	0.984 ± 0.002	206.0 ± 102.2
Full-cell baselines						
Full-cell invariant GNN	7.94 ± 1.70	C	25.6 ± 3.1	588.3 ± 82.6	0.934 ± 0.016	951.7 ± 45.6
		V	17.4 ± 2.9	374.2 ± 83.4	0.969 ± 0.010	472.3 ± 222.4
Full-cell e3nn ($\ell_{\max} = 1$)	$\mathbf{5.05 \pm 1.24}$	C	20.3 ± 1.7	451.7 ± 49.0	0.959 ± 0.007	928.3 ± 116.4
		V	12.1 ± 2.4	258.0 ± 29.1	0.985 ± 0.006	519.0 ± 176.4
Full-cell e3nn ($\ell_{\max} = 2$)	6.35 ± 1.25	C	20.6 ± 2.9	471.5 ± 52.5	0.957 ± 0.012	899.3 ± 83.5
		V	$\mathbf{10.6 \pm 2.0}$	$\mathbf{226.2 \pm 29.0}$	$\mathbf{0.989 \pm 0.004}$	719.7 ± 214.2

Table 30: Toy-model performance for wallpaper group pg (1,000 structures, mean $\Delta E = -7.49$ eV atom^{-1}, energy SD 13.53 meV atom^{-1}).

Model	Energy MAE ↓ (meV atom^{-1})	Mode	Rel. force L_2 ↓ (%)	Force comp. MAE ↓ (meV Å^{-1})	Force comp. R^2 ↑	Best epoch
WallpaperNet models						
WP-Inv	8.16 ± 2.01	C	32.7 ± 1.3	739.6 ± 29.8	0.893 ± 0.008	573.3 ± 183.5
		V	16.0 ± 1.4	340.1 ± 16.9	0.974 ± 0.005	354.7 ± 41.9
WP-InvExt	6.68 ± 1.41	C	30.3 ± 2.2	669.1 ± 70.8	0.908 ± 0.013	496.3 ± 177.0
		V	16.3 ± 0.7	356.6 ± 9.2	0.973 ± 0.002	352.0 ± 92.0
WP-Equi	7.74 ± 0.98	C	33.7 ± 4.0	768.8 ± 65.3	0.885 ± 0.026	537.7 ± 42.2
		V	24.6 ± 2.8	540.9 ± 59.0	0.939 ± 0.014	296.3 ± 31.5
WP-VecEqui	8.64 ± 0.67	C	33.8 ± 3.6	769.9 ± 59.3	0.885 ± 0.024	872.0 ± 140.2
		V	31.1 ± 6.2	690.1 ± 95.4	0.901 ± 0.037	258.3 ± 27.2
WP-Reg	7.77 ± 0.21	C	32.9 ± 3.1	747.3 ± 50.3	0.891 ± 0.020	622.7 ± 172.6
		V	24.8 ± 1.8	520.6 ± 35.1	0.938 ± 0.009	259.7 ± 43.5
WP-RegTP	9.61 ± 1.10	C	33.5 ± 2.6	768.9 ± 40.8	0.887 ± 0.018	471.3 ± 164.9
		V	31.8 ± 5.2	727.8 ± 76.6	0.897 ± 0.034	240.0 ± 93.4
WP-O3	5.89 ± 1.24	C	23.5 ± 1.2	551.2 ± 41.3	0.945 ± 0.006	542.7 ± 47.2
		V	16.6 ± 1.7	360.0 ± 35.1	0.972 ± 0.006	262.0 ± 68.6
Full-cell baselines						
Full-cell invariant GNN	7.16 ± 0.46	C	29.5 ± 2.4	664.7 ± 27.2	0.913 ± 0.014	878.0 ± 202.7
		V	17.1 ± 1.4	365.4 ± 27.0	0.971 ± 0.005	655.3 ± 150.4
Full-cell e3nn ($\ell_{\max} = 1$)	5.47 ± 0.34	C	23.6 ± 1.5	540.9 ± 21.9	0.944 ± 0.007	993.0 ± 2.6
		V	12.7 ± 0.8	273.6 ± 24.0	0.984 ± 0.002	454.3 ± 65.3
Full-cell e3nn ($\ell_{\max} = 2$)	$\mathbf{5.27 \pm 0.57}$	C	$\mathbf{22.8 \pm 1.1}$	$\mathbf{517.9 \pm 25.7}$	$\mathbf{0.948 \pm 0.005}$	872.0 ± 218.2
		V	$\mathbf{12.5 \pm 1.3}$	$\mathbf{264.2 \pm 23.5}$	$\mathbf{0.984 \pm 0.003}$	672.7 ± 240.4

Table 31: Toy-model performance for wallpaper group cm (1,000 structures, mean $\Delta E = -7.01$ eV atom^{-1}, energy SD 9.58 meV atom^{-1}).

Model	Energy MAE ↓ (meV atom^{-1})	Mode	Rel. force L_2 ↓ (%)	Force comp. MAE ↓ (meV Å^{-1})	Force comp. R^2 ↑	Best epoch
WallpaperNet models						
WP-Inv	6.21 ± 1.86	C	30.7 ± 1.1	680.5 ± 34.4	0.906 ± 0.007	549.7 ± 34.9
		V	11.8 ± 1.1	247.0 ± 18.1	0.986 ± 0.003	495.7 ± 130.2
WP-InvExt	3.40 ± 0.55	C	24.6 ± 0.8	535.2 ± 10.7	0.940 ± 0.004	541.7 ± 143.2
		V	12.5 ± 1.0	263.7 ± 24.8	0.984 ± 0.002	579.7 ± 77.2
WP-Equi	8.29 ± 6.39	C	33.3 ± 0.6	731.2 ± 28.2	0.889 ± 0.004	488.3 ± 24.4
		V	20.5 ± 0.7	439.5 ± 4.9	0.958 ± 0.003	247.3 ± 41.5
WP-VecEqui	6.99 ± 1.64	C	36.3 ± 2.1	808.7 ± 83.3	0.868 ± 0.015	644.7 ± 79.5
		V	27.6 ± 1.3	575.1 ± 34.2	0.924 ± 0.007	260.0 ± 7.9
WP-Reg	7.61 ± 1.18	C	35.2 ± 0.6	773.0 ± 9.6	0.876 ± 0.005	533.3 ± 149.6
		V	17.4 ± 1.0	351.3 ± 26.3	0.970 ± 0.004	366.0 ± 87.0
WP-RegTP	6.78 ± 1.06	C	32.8 ± 2.3	737.8 ± 56.0	0.892 ± 0.015	536.3 ± 125.4
		V	27.2 ± 1.7	593.7 ± 3.0	0.926 ± 0.009	301.0 ± 79.5
WP-O3	$\mathbf{2.68 \pm 0.24}$	C	$\mathbf{18.8 \pm 1.0}$	$\mathbf{425.6 \pm 23.7}$	$\mathbf{0.964 \pm 0.004}$	694.3 ± 146.5
		V	11.5 ± 1.2	239.5 ± 24.0	0.987 ± 0.003	319.0 ± 32.7
Full-cell baselines						
Full-cell invariant GNN	4.28 ± 1.31	C	28.1 ± 5.7	614.9 ± 120.3	0.919 ± 0.030	934.0 ± 84.9
		V	12.9 ± 0.4	270.2 ± 11.9	0.983 ± 0.001	763.7 ± 102.7
Full-cell e3nn ($\ell_{\max} = 1$)	3.26 ± 0.40	C	22.0 ± 0.9	484.1 ± 20.0	0.952 ± 0.004	846.0 ± 141.3
		V	11.0 ± 0.7	228.1 ± 16.0	0.988 ± 0.002	466.7 ± 83.1
Full-cell e3nn ($\ell_{\max} = 2$)	3.64 ± 0.52	C	21.6 ± 1.7	488.2 ± 32.4	0.953 ± 0.007	961.0 ± 60.6
		V	$\mathbf{11.0 \pm 0.1}$	$\mathbf{226.7 \pm 5.1}$	$\mathbf{0.988 \pm 0.000}$	764.3 ± 200.2

Table 32: Toy-model performance for wallpaper group p3 (1,000 structures, mean $\Delta E = -6.78$ eV atom^{-1}, energy SD 21.43 meV atom^{-1}).

Model	Energy MAE ↓ (meV atom^{-1})	Mode	Rel. force L_2 ↓ (%)	Force comp. MAE ↓ (meV Å^{-1})	Force comp. R^2 ↑	Best epoch
WallpaperNet models						
WP-Inv	10.13 ± 1.09	C	30.8 ± 1.0	725.9 ± 22.5	0.905 ± 0.006	910.7 ± 131.8
		V	14.1 ± 0.7	315.9 ± 13.3	0.980 ± 0.002	358.7 ± 45.6
WP-InvExt	7.73 ± 1.22	C	25.6 ± 4.0	564.9 ± 67.8	0.933 ± 0.020	476.7 ± 336.2
		V	14.1 ± 1.5	311.9 ± 26.3	0.980 ± 0.004	468.3 ± 156.5
WP-Equi	11.05 ± 2.28	C	30.6 ± 0.7	715.3 ± 2.7	0.906 ± 0.004	878.7 ± 94.1
		V	16.1 ± 1.5	367.6 ± 25.9	0.974 ± 0.005	246.3 ± 24.8
WP-VecEqui	13.98 ± 0.88	C	35.6 ± 5.0	821.0 ± 108.6	0.871 ± 0.036	809.0 ± 222.9
		V	21.4 ± 2.9	455.3 ± 53.2	0.954 ± 0.013	316.7 ± 30.7
WP-Reg	10.02 ± 1.24	C	29.6 ± 1.4	673.8 ± 47.3	0.912 ± 0.008	833.3 ± 197.1
		V	19.7 ± 2.2	413.5 ± 26.4	0.961 ± 0.008	304.3 ± 34.4
WP-RegTP	10.46 ± 0.75	C	27.8 ± 1.3	665.0 ± 32.5	0.923 ± 0.007	836.0 ± 139.8
		V	19.9 ± 0.2	447.1 ± 3.5	0.960 ± 0.001	341.3 ± 14.2
WP-O3	$\mathbf{4.42 \pm 0.98}$	C	$\mathbf{15.4 \pm 1.1}$	$\mathbf{366.7 \pm 20.8}$	$\mathbf{0.976 \pm 0.004}$	744.7 ± 213.6
		V	12.3 ± 1.7	282.2 ± 28.0	0.985 ± 0.004	232.3 ± 67.7
Full-cell baselines						
Full-cell invariant GNN	8.72 ± 2.00	C	24.9 ± 2.9	570.0 ± 70.5	0.937 ± 0.014	634.7 ± 36.2
		V	15.7 ± 0.6	348.2 ± 16.3	0.975 ± 0.002	821.0 ± 181.7
Full-cell e3nn ($\ell_{\max} = 1$)	6.17 ± 0.92	C	19.2 ± 2.5	442.1 ± 58.2	0.963 ± 0.010	865.7 ± 212.8
		V	11.3 ± 1.5	$\mathbf{255.7 \pm 26.3}$	0.987 ± 0.003	459.3 ± 77.9
Full-cell e3nn ($\ell_{\max} = 2$)	7.54 ± 0.66	C	20.0 ± 1.0	460.1 ± 20.7	0.960 ± 0.004	958.3 ± 49.7
		V	$\mathbf{11.3 \pm 1.4}$	258.9 ± 26.5	$\mathbf{0.987 \pm 0.003}$	840.7 ± 167.0

Table 33: Toy-model performance for wallpaper group p4 (1,000 structures, mean $\Delta E = -6.77$ eV atom^{-1}, energy SD 21.95 meV atom^{-1}).

Model	Energy MAE ↓ (meV atom^{-1})	Mode	Rel. force L_2 ↓ (%)	Force comp. MAE ↓ (meV Å^{-1})	Force comp. R^2 ↑	Best epoch
WallpaperNet models						
WP-Inv	10.45 ± 1.52	C	33.7 ± 1.4	778.9 ± 34.3	0.886 ± 0.010	541.3 ± 110.9
		V	13.4 ± 1.1	280.8 ± 29.9	0.982 ± 0.003	419.3 ± 56.6
WP-InvExt	4.41 ± 0.41	C	21.3 ± 0.5	466.1 ± 22.7	0.955 ± 0.002	726.7 ± 263.4
		V	14.4 ± 1.8	310.9 ± 45.7	0.979 ± 0.005	375.3 ± 26.6
WP-Equi	12.66 ± 4.98	C	32.3 ± 3.0	718.7 ± 78.5	0.895 ± 0.019	547.7 ± 213.3
		V	15.2 ± 0.9	333.1 ± 20.0	0.977 ± 0.003	337.7 ± 29.1
WP-VecEqui	7.64 ± 2.03	C	30.3 ± 2.2	669.4 ± 61.1	0.908 ± 0.014	801.3 ± 71.8
		V	20.8 ± 3.5	433.0 ± 70.4	0.956 ± 0.015	351.0 ± 82.5
WP-Reg	9.89 ± 0.41	C	32.9 ± 2.0	743.1 ± 43.5	0.892 ± 0.013	687.7 ± 162.5
		V	18.2 ± 1.9	384.0 ± 39.4	0.967 ± 0.007	288.7 ± 77.9
WP-RegTP	10.54 ± 0.60	C	32.7 ± 1.0	753.5 ± 29.6	0.893 ± 0.006	546.3 ± 174.4
		V	25.8 ± 7.2	596.1 ± 170.7	0.930 ± 0.034	312.3 ± 15.2
WP-O3	$\mathbf{3.96 \pm 0.98}$	C	$\mathbf{16.9 \pm 0.6}$	$\mathbf{375.2 \pm 19.2}$	$\mathbf{0.971 \pm 0.002}$	830.0 ± 272.1
		V	10.7 ± 1.5	234.6 ± 26.2	0.988 ± 0.003	333.0 ± 121.6
Full-cell baselines						
Full-cell invariant GNN	8.36 ± 1.29	C	29.5 ± 4.7	661.9 ± 105.0	0.911 ± 0.028	763.3 ± 215.0
		V	15.1 ± 0.4	322.1 ± 13.1	0.977 ± 0.001	823.7 ± 103.4
Full-cell e3nn ($\ell_{\max} = 1$)	4.19 ± 0.37	C	18.0 ± 3.1	402.5 ± 68.1	0.967 ± 0.011	969.3 ± 24.0
		V	10.8 ± 0.4	232.9 ± 15.0	0.988 ± 0.001	499.7 ± 141.2
Full-cell e3nn ($\ell_{\max} = 2$)	5.08 ± 0.96	C	20.2 ± 2.3	447.2 ± 49.6	0.959 ± 0.009	977.0 ± 23.6
		V	$\mathbf{10.7 \pm 0.7}$	$\mathbf{231.8 \pm 8.4}$	$\mathbf{0.989 \pm 0.001}$	625.3 ± 58.3

Table 34: Toy-model performance for wallpaper group pmm (1,000 structures, mean $\Delta E = -6.77$ eV atom^{-1}, energy SD 21.78 meV atom^{-1}).

Model	Energy MAE ↓ (meV atom^{-1})	Mode	Rel. force L_2 ↓ (%)	Force comp. MAE ↓ (meV Å^{-1})	Force comp. R^2 ↑	Best epoch
WallpaperNet models						
WP-Inv	9.58 ± 0.94	C	30.7 ± 0.5	698.0 ± 23.1	0.906 ± 0.003	695.7 ± 266.6
		V	16.2 ± 0.7	332.0 ± 12.7	0.974 ± 0.002	593.3 ± 261.2
WP-InvExt	5.81 ± 1.04	C	24.5 ± 4.3	520.4 ± 84.3	0.939 ± 0.022	713.7 ± 267.5
		V	15.8 ± 0.4	320.8 ± 7.9	0.975 ± 0.001	463.3 ± 89.8
WP-Equi	15.18 ± 1.77	C	46.7 ± 19.6	1032.9 ± 384.1	0.756 ± 0.205	531.7 ± 220.9
		V	18.4 ± 0.8	382.2 ± 12.9	0.966 ± 0.003	306.0 ± 60.9
WP-VecEqui	9.44 ± 0.91	C	31.2 ± 1.4	705.3 ± 33.7	0.902 ± 0.009	771.0 ± 193.8
		V	25.5 ± 1.9	534.4 ± 25.1	0.935 ± 0.010	252.7 ± 39.4
WP-Reg	11.26 ± 0.78	C	31.4 ± 3.1	716.4 ± 67.7	0.901 ± 0.020	801.7 ± 89.6
		V	21.1 ± 2.4	430.7 ± 51.0	0.955 ± 0.010	333.0 ± 43.1
WP-RegTP	9.11 ± 0.50	C	30.2 ± 0.0	700.2 ± 11.7	0.909 ± 0.000	750.3 ± 136.2
		V	30.9 ± 2.8	690.7 ± 16.4	0.904 ± 0.017	237.3 ± 35.8
WP-O3	$\mathbf{3.84 \pm 1.06}$	C	$\mathbf{18.2 \pm 1.0}$	$\mathbf{401.9 \pm 13.0}$	$\mathbf{0.967 \pm 0.004}$	847.3 ± 113.2
		V	14.3 ± 1.1	304.7 ± 35.2	0.979 ± 0.003	261.7 ± 107.5
Full-cell baselines						
Full-cell invariant GNN	8.40 ± 2.35	C	29.1 ± 4.8	640.8 ± 120.8	0.914 ± 0.027	992.3 ± 5.5
		V	17.6 ± 2.3	345.2 ± 50.5	0.969 ± 0.008	862.0 ± 113.5
Full-cell e3nn ($\ell_{\max} = 1$)	4.69 ± 0.64	C	23.0 ± 4.9	468.5 ± 47.4	0.946 ± 0.023	836.7 ± 52.2
		V	14.0 ± 0.7	286.4 ± 20.9	0.980 ± 0.002	467.7 ± 122.8
Full-cell e3nn ($\ell_{\max} = 2$)	5.92 ± 0.89	C	22.0 ± 2.6	487.3 ± 48.7	0.951 ± 0.012	988.7 ± 16.2
		V	$\mathbf{13.1 \pm 0.7}$	$\mathbf{273.3 \pm 17.6}$	$\mathbf{0.983 \pm 0.002}$	661.3 ± 271.4

Table 35: Toy-model performance for wallpaper group pmg (1,000 structures, mean $\Delta E = -7.10$ eV atom^{-1}, energy SD 14.09 meV atom^{-1}).

Model	Energy MAE ↓ (meV atom^{-1})	Mode	Rel. force L_2 ↓ (%)	Force comp. MAE ↓ (meV Å^{-1})	Force comp. R^2 ↑	Best epoch
WallpaperNet models						
WP-Inv	9.15 ± 1.18	C	31.3 ± 2.1	717.3 ± 46.3	0.902 ± 0.013	553.7 ± 381.7
		V	12.9 ± 4.8	264.4 ± 83.4	0.982 ± 0.014	466.0 ± 187.3
WP-InvExt	4.43 ± 1.32	C	22.8 ± 3.1	495.8 ± 65.7	0.948 ± 0.015	571.7 ± 123.7
		V	12.8 ± 1.8	260.8 ± 25.2	0.983 ± 0.005	535.7 ± 196.1
WP-Equi	8.93 ± 1.21	C	33.8 ± 5.0	764.5 ± 128.4	0.884 ± 0.032	586.3 ± 186.6
		V	16.4 ± 3.4	340.9 ± 48.5	0.972 ± 0.012	398.0 ± 64.7
WP-VecEqui	9.66 ± 0.52	C	32.4 ± 2.6	747.2 ± 62.1	0.895 ± 0.017	500.3 ± 95.8
		V	22.3 ± 4.2	468.8 ± 84.5	0.949 ± 0.018	312.0 ± 29.8
WP-Reg	8.58 ± 1.50	C	30.6 ± 3.9	690.4 ± 98.4	0.905 ± 0.023	569.7 ± 49.1
		V	17.0 ± 2.1	342.3 ± 40.4	0.971 ± 0.007	375.0 ± 57.3
WP-RegTP	8.48 ± 0.34	C	29.7 ± 1.9	692.3 ± 42.5	0.911 ± 0.011	770.3 ± 270.7
		V	21.4 ± 1.6	471.0 ± 77.4	0.954 ± 0.007	579.0 ± 43.4
WP-O3	$\mathbf{3.30 \pm 1.04}$	C	$\mathbf{16.8 \pm 2.9}$	$\mathbf{373.1 \pm 49.8}$	$\mathbf{0.971 \pm 0.010}$	762.7 ± 115.7
		V	10.6 ± 3.9	215.2 ± 55.4	0.988 ± 0.009	298.0 ± 95.6
Full-cell baselines						
Full-cell invariant GNN	6.62 ± 1.15	C	27.1 ± 2.9	590.8 ± 49.2	0.926 ± 0.015	992.3 ± 2.9
		V	13.2 ± 1.8	266.1 ± 21.1	0.982 ± 0.005	766.0 ± 158.8
Full-cell e3nn ($\ell_{\max} = 1$)	4.39 ± 0.66	C	20.1 ± 2.8	445.9 ± 39.2	0.959 ± 0.012	905.3 ± 151.9
		V	10.9 ± 2.4	220.1 ± 40.3	0.988 ± 0.005	431.0 ± 40.4
Full-cell e3nn ($\ell_{\max} = 2$)	4.59 ± 1.01	C	19.7 ± 3.3	443.6 ± 55.2	0.960 ± 0.013	960.0 ± 57.5
		V	$\mathbf{10.0 \pm 1.7}$	$\mathbf{209.2 \pm 28.4}$	$\mathbf{0.990 \pm 0.003}$	766.0 ± 192.1

Table 36: Toy-model performance for wallpaper group pgg (1,000 structures, mean $\Delta E = -7.71$ eV atom^{-1}, energy SD 15.18 meV atom^{-1}).

Model	Energy MAE ↓ (meV atom^{-1})	Mode	Rel. force L_2 ↓ (%)	Force comp. MAE ↓ (meV Å^{-1})	Force comp. R^2 ↑	Best epoch
WallpaperNet models						
WP-Inv	10.31 ± 2.92	C	32.7 ± 5.6	754.7 ± 134.1	0.891 ± 0.035	423.3 ± 19.6
		V	13.6 ± 1.5	290.1 ± 33.9	0.981 ± 0.004	368.7 ± 31.0
WP-InvExt	6.62 ± 0.24	C	24.1 ± 1.2	529.6 ± 35.4	0.942 ± 0.006	595.0 ± 317.8
		V	13.2 ± 2.9	285.3 ± 59.0	0.982 ± 0.007	456.0 ± 46.5
WP-Equi	9.39 ± 1.96	C	33.8 ± 2.8	779.0 ± 64.8	0.885 ± 0.018	571.0 ± 76.0
		V	17.7 ± 2.6	385.1 ± 56.2	0.968 ± 0.009	271.7 ± 50.5
WP-VecEqui	11.27 ± 2.50	C	34.7 ± 3.4	804.0 ± 82.4	0.879 ± 0.023	536.7 ± 115.4
		V	23.1 ± 4.5	492.6 ± 82.4	0.945 ± 0.020	264.0 ± 86.5
WP-Reg	13.67 ± 6.96	C	34.3 ± 1.9	784.4 ± 53.2	0.882 ± 0.013	626.7 ± 415.6
		V	18.5 ± 1.4	378.8 ± 17.4	0.966 ± 0.005	298.7 ± 11.6
WP-RegTP	10.40 ± 2.25	C	32.2 ± 3.2	757.4 ± 85.5	0.895 ± 0.021	626.3 ± 306.3
		V	22.8 ± 1.0	525.7 ± 34.6	0.948 ± 0.005	349.0 ± 16.5
WP-O3	$\mathbf{4.40 \pm 0.39}$	C	$\mathbf{18.4 \pm 2.1}$	$\mathbf{418.8 \pm 49.2}$	$\mathbf{0.966 \pm 0.008}$	626.0 ± 21.7
		V	11.6 ± 1.0	247.5 ± 16.0	0.987 ± 0.002	316.7 ± 62.1
Full-cell baselines						
Full-cell invariant GNN	7.50 ± 1.01	C	26.9 ± 3.5	597.0 ± 59.4	0.927 ± 0.018	814.3 ± 180.5
		V	14.2 ± 0.6	302.7 ± 10.1	0.980 ± 0.002	796.7 ± 179.2
Full-cell e3nn ($\ell_{\max} = 1$)	5.90 ± 0.78	C	23.6 ± 1.9	520.2 ± 25.1	0.944 ± 0.009	842.0 ± 149.0
		V	$\mathbf{11.3 \pm 0.5}$	$\mathbf{240.6 \pm 17.5}$	$\mathbf{0.987 \pm 0.001}$	439.7 ± 60.9
Full-cell e3nn ($\ell_{\max} = 2$)	5.12 ± 0.25	C	21.7 ± 2.0	493.9 ± 53.8	0.953 ± 0.009	966.3 ± 47.1
		V	11.3 ± 1.3	249.0 ± 24.5	0.987 ± 0.003	680.7 ± 215.9

Table 37: Toy-model performance for wallpaper group cmm (1,000 structures, mean $\Delta E = -7.07$ eV atom^{-1}, energy SD 9.70 meV atom^{-1}).

Model	Energy MAE ↓ (meV atom^{-1})	Mode	Rel. force L_2 ↓ (%)	Force comp. MAE ↓ (meV Å^{-1})	Force comp. R^2 ↑	Best epoch
WallpaperNet models						
WP-Inv	11.90 ± 1.28	C	32.3 ± 1.1	723.7 ± 10.1	0.896 ± 0.007	797.7 ± 219.8
		V	12.3 ± 0.5	261.9 ± 11.6	0.985 ± 0.001	569.7 ± 191.3
WP-InvExt	$\mathbf{2.98 \pm 0.59}$	C	21.7 ± 2.5	480.8 ± 59.1	0.952 ± 0.011	428.0 ± 50.4
		V	12.1 ± 0.8	257.0 ± 25.3	0.985 ± 0.002	515.7 ± 151.3
WP-Equi	13.38 ± 3.62	C	33.1 ± 0.4	758.1 ± 13.5	0.890 ± 0.003	604.0 ± 126.2
		V	15.2 ± 1.6	334.3 ± 43.2	0.977 ± 0.005	327.0 ± 35.5
WP-VecEqui	11.68 ± 4.67	C	35.9 ± 1.4	799.9 ± 23.4	0.871 ± 0.010	695.0 ± 130.8
		V	20.7 ± 1.6	427.4 ± 25.8	0.957 ± 0.007	295.7 ± 23.0
WP-Reg	7.47 ± 1.70	C	33.4 ± 2.5	755.4 ± 51.5	0.888 ± 0.017	659.3 ± 319.3
		V	18.1 ± 5.0	377.6 ± 102.8	0.966 ± 0.020	388.7 ± 79.4
WP-RegTP	13.77 ± 1.36	C	33.3 ± 2.1	750.1 ± 31.3	0.889 ± 0.014	746.7 ± 322.8
		V	26.4 ± 1.8	601.3 ± 8.4	0.930 ± 0.009	347.7 ± 68.3
WP-O3	3.32 ± 1.42	C	$\mathbf{18.3 \pm 1.0}$	$\mathbf{409.3 \pm 26.0}$	$\mathbf{0.966 \pm 0.004}$	727.3 ± 171.6
		V	11.8 ± 0.3	255.0 ± 18.8	0.986 ± 0.001	266.0 ± 60.9
Full-cell baselines						
Full-cell invariant GNN	6.73 ± 1.77	C	32.7 ± 4.6	738.9 ± 102.7	0.892 ± 0.029	738.7 ± 192.3
		V	13.9 ± 1.6	298.3 ± 26.2	0.981 ± 0.004	755.0 ± 105.7
Full-cell e3nn ($\ell_{\max} = 1$)	3.43 ± 0.49	C	21.8 ± 2.0	491.9 ± 29.0	0.952 ± 0.008	866.0 ± 223.4
		V	11.2 ± 0.7	235.0 ± 23.4	0.987 ± 0.001	426.7 ± 48.2
Full-cell e3nn ($\ell_{\max} = 2$)	3.67 ± 0.36	C	21.8 ± 0.3	500.1 ± 1.8	0.953 ± 0.001	864.7 ± 178.4
		V	$\mathbf{10.8 \pm 0.8}$	$\mathbf{230.9 \pm 14.2}$	$\mathbf{0.988 \pm 0.002}$	687.3 ± 92.6

Table 38: Toy-model performance for wallpaper group p6 (1,000 structures, mean $\Delta E = -7.34$ eV atom^{-1}, energy SD 16.91 meV atom^{-1}).

Model	Energy MAE ↓ (meV atom^{-1})	Mode	Rel. force L_2 ↓ (%)	Force comp. MAE ↓ (meV Å^{-1})	Force comp. R^2 ↑	Best epoch
WallpaperNet models						
WP-Inv	9.61 ± 0.55	C	31.2 ± 0.9	709.5 ± 9.0	0.903 ± 0.006	883.3 ± 21.1
		V	13.5 ± 0.9	294.0 ± 4.7	0.982 ± 0.002	409.0 ± 91.0
WP-InvExt	7.31 ± 2.11	C	23.5 ± 1.0	525.2 ± 27.2	0.945 ± 0.005	554.3 ± 110.6
		V	13.7 ± 2.0	301.2 ± 24.6	0.981 ± 0.006	414.0 ± 116.0
WP-Equi	12.36 ± 1.58	C	33.8 ± 4.9	772.5 ± 123.7	0.884 ± 0.034	543.7 ± 253.1
		V	14.1 ± 2.7	306.5 ± 47.4	0.980 ± 0.008	336.7 ± 59.1
WP-VecEqui	10.77 ± 2.81	C	30.0 ± 4.9	668.4 ± 117.0	0.909 ± 0.028	827.3 ± 180.3
		V	18.3 ± 0.6	392.1 ± 26.1	0.967 ± 0.002	372.3 ± 18.2
WP-Reg	13.01 ± 0.55	C	34.3 ± 1.8	774.1 ± 23.4	0.882 ± 0.012	679.7 ± 192.7
		V	17.3 ± 2.2	358.4 ± 30.1	0.970 ± 0.007	356.0 ± 46.8
WP-RegTP	8.89 ± 2.33	C	27.0 ± 0.2	624.9 ± 22.7	0.927 ± 0.001	834.7 ± 144.7
		V	22.1 ± 2.8	507.3 ± 66.9	0.951 ± 0.013	326.3 ± 49.8
WP-O3	$\mathbf{4.42 \pm 0.48}$	C	$\mathbf{17.1 \pm 2.5}$	$\mathbf{385.6 \pm 36.4}$	$\mathbf{0.970 \pm 0.009}$	883.7 ± 102.6
		V	$\mathbf{10.2 \pm 0.9}$	229.2 ± 16.2	$\mathbf{0.989 \pm 0.002}$	298.3 ± 86.2
Full-cell baselines						
Full-cell invariant GNN	6.29 ± 1.14	C	22.8 ± 3.0	500.4 ± 61.6	0.947 ± 0.013	823.0 ± 195.4
		V	13.9 ± 0.7	304.1 ± 19.6	0.981 ± 0.002	741.0 ± 184.5
Full-cell e3nn ($\ell_{\max} = 1$)	5.51 ± 1.28	C	19.5 ± 1.7	433.5 ± 27.2	0.962 ± 0.007	786.0 ± 288.1
		V	11.2 ± 0.9	240.8 ± 9.8	0.987 ± 0.002	445.3 ± 82.1
Full-cell e3nn ($\ell_{\max} = 2$)	5.22 ± 1.20	C	18.0 ± 2.7	412.4 ± 63.8	0.967 ± 0.009	971.0 ± 47.6
		V	10.5 ± 0.7	$\mathbf{227.5 \pm 13.6}$	0.989 ± 0.001	708.0 ± 182.9

Table 39: Toy-model performance for wallpaper group $\mathrm{p3m1}$ (1,000 structures, mean $\Delta E = -7.34$ eV atom^{-1}, energy SD 15.43 meV atom^{-1}).

Model	Energy MAE ↓ (meV atom^{-1})	Mode	Rel. force L_2 ↓ (%)	Force comp. MAE ↓ (meV Å^{-1})	Force comp. R^2 ↑	Best epoch
WallpaperNet models						
WP-Inv	10.39 ± 1.18	C	29.8 ± 4.5	688.5 ± 112.8	0.910 ± 0.027	751.7 ± 211.9
		V	13.7 ± 2.8	286.9 ± 53.8	0.981 ± 0.008	449.0 ± 65.3
WP-InvExt	5.57 ± 0.56	C	22.8 ± 2.1	510.4 ± 49.5	0.948 ± 0.010	659.0 ± 315.1
		V	13.6 ± 2.3	292.9 ± 37.5	0.981 ± 0.007	519.0 ± 155.9
WP-Equi	10.71 ± 2.33	C	35.3 ± 4.5	830.0 ± 97.6	0.874 ± 0.032	560.0 ± 268.3
		V	14.9 ± 2.0	329.1 ± 28.5	0.978 ± 0.006	351.3 ± 33.9
WP-VecEqui	9.88 ± 0.87	C	30.7 ± 3.1	705.7 ± 68.8	0.905 ± 0.019	787.0 ± 111.5
		V	22.2 ± 2.4	478.9 ± 51.0	0.950 ± 0.011	319.0 ± 41.1
WP-Reg	12.40 ± 1.16	C	33.6 ± 2.4	773.7 ± 57.9	0.887 ± 0.016	812.7 ± 230.7
		V	16.6 ± 1.4	350.2 ± 26.4	0.972 ± 0.005	462.3 ± 127.9
WP-RegTP	9.97 ± 2.09	C	30.3 ± 4.6	717.4 ± 105.2	0.907 ± 0.028	515.7 ± 30.0
		V	22.0 ± 4.5	490.9 ± 99.1	0.950 ± 0.019	388.3 ± 150.4
WP-O3	$\mathbf{4.38 \pm 0.13}$	C	$\mathbf{17.9 \pm 2.5}$	$\mathbf{407.5 \pm 43.7}$	$\mathbf{0.967 \pm 0.009}$	778.3 ± 166.5
		V	$\mathbf{10.5 \pm 0.8}$	$\mathbf{227.4 \pm 12.1}$	$\mathbf{0.989 \pm 0.002}$	318.3 ± 41.0
Full-cell baselines						
Full-cell invariant GNN	6.74 ± 1.42	C	22.7 ± 4.4	515.5 ± 91.9	0.947 ± 0.021	900.3 ± 161.5
		V	14.8 ± 2.1	307.0 ± 36.1	0.978 ± 0.006	808.3 ± 96.7
Full-cell e3nn ($\ell_{\max} = 1$)	6.55 ± 0.65	C	21.1 ± 3.8	478.7 ± 64.9	0.954 ± 0.017	688.7 ± 106.6
		V	11.7 ± 2.5	254.0 ± 50.1	0.986 ± 0.006	512.3 ± 92.5
Full-cell e3nn ($\ell_{\max} = 2$)	6.14 ± 0.87	C	20.8 ± 2.5	465.1 ± 42.1	0.956 ± 0.010	768.7 ± 256.7
		V	10.9 ± 0.7	242.9 ± 15.1	0.988 ± 0.002	814.0 ± 154.0

Table 40: Toy-model performance for wallpaper group $\mathrm{p31m}$ (1,000 structures, mean $\Delta E = -6.22$ eV atom^{-1}, energy SD 32.38 meV atom^{-1}).

Model	Energy MAE ↓ (meV atom^{-1})	Mode	Rel. force L_2 ↓ (%)	Force comp. MAE ↓ (meV Å^{-1})	Force comp. R^2 ↑	Best epoch
WallpaperNet models						
WP-Inv	8.32 ± 2.25	C	27.9 ± 3.2	613.7 ± 78.3	0.921 ± 0.017	796.7 ± 175.0
		V	14.5 ± 1.0	305.2 ± 19.8	0.979 ± 0.003	337.7 ± 47.6
WP-InvExt	7.68 ± 1.36	C	27.4 ± 3.6	568.1 ± 48.5	0.924 ± 0.019	556.3 ± 193.3
		V	15.2 ± 1.7	309.9 ± 40.0	0.977 ± 0.005	407.0 ± 56.0
WP-Equi	9.47 ± 0.92	C	32.6 ± 9.0	697.6 ± 159.9	0.888 ± 0.061	507.7 ± 30.9
		V	16.1 ± 1.0	335.6 ± 23.7	0.974 ± 0.003	336.3 ± 39.3
WP-VecEqui	9.82 ± 3.11	C	29.5 ± 3.4	632.0 ± 60.6	0.912 ± 0.020	931.0 ± 73.5
		V	24.4 ± 5.1	480.3 ± 66.5	0.939 ± 0.025	262.0 ± 50.2
WP-Reg	9.31 ± 1.48	C	31.6 ± 2.7	677.5 ± 73.2	0.900 ± 0.017	526.7 ± 189.0
		V	18.7 ± 3.2	368.3 ± 48.4	0.964 ± 0.013	248.7 ± 42.1
WP-RegTP	7.98 ± 0.48	C	24.9 ± 1.5	549.0 ± 25.7	0.938 ± 0.008	735.0 ± 152.0
		V	20.2 ± 3.9	426.2 ± 80.0	0.958 ± 0.016	335.3 ± 108.6
WP-O3	$\mathbf{4.56 \pm 1.14}$	C	$\mathbf{17.2 \pm 1.8}$	$\mathbf{374.3 \pm 49.1}$	$\mathbf{0.970 \pm 0.006}$	655.0 ± 93.4
		V	$\mathbf{11.7 \pm 1.4}$	$\mathbf{236.0 \pm 25.8}$	$\mathbf{0.986 \pm 0.003}$	281.0 ± 24.0
Full-cell baselines						
Full-cell invariant GNN	6.33 ± 1.84	C	24.7 ± 2.4	512.6 ± 69.6	0.938 ± 0.012	954.0 ± 74.5
		V	35.0 ± 32.2	805.7 ± 830.8	0.808 ± 0.285	436.7 ± 294.1
Full-cell e3nn ($\ell_{\max} = 1$)	5.27 ± 2.01	C	18.3 ± 2.8	401.4 ± 66.7	0.966 ± 0.011	992.0 ± 9.6
		V	12.1 ± 1.3	249.5 ± 17.8	0.985 ± 0.003	458.3 ± 94.3
Full-cell e3nn ($\ell_{\max} = 2$)	6.40 ± 1.92	C	20.2 ± 1.9	446.5 ± 49.9	0.959 ± 0.008	968.3 ± 25.7
		V	12.8 ± 1.7	262.9 ± 34.2	0.984 ± 0.004	893.3 ± 104.5

Table 41: Toy-model performance for wallpaper group $\mathrm{p4m}$ (1,000 structures, mean $\Delta E = -6.77$ eV atom^{-1}, energy SD 21.66 meV atom^{-1}).

Model	Energy MAE ↓ (meV atom^{-1})	Mode	Rel. force L_2 ↓ (%)	Force comp. MAE ↓ (meV Å^{-1})	Force comp. R^2 ↑	Best epoch
WallpaperNet models						
WP-Inv	10.47 ± 1.72	C	30.1 ± 2.6	708.2 ± 55.7	0.909 ± 0.015	618.7 ± 169.2
		V	12.1 ± 1.6	272.2 ± 30.3	0.985 ± 0.004	415.7 ± 58.6
WP-InvExt	5.70 ± 1.63	C	21.4 ± 2.9	475.9 ± 58.7	0.953 ± 0.013	429.7 ± 44.6
		V	13.2 ± 1.6	291.0 ± 26.5	0.982 ± 0.004	686.0 ± 110.1
WP-Equi	10.88 ± 0.81	C	31.6 ± 2.7	716.5 ± 74.9	0.900 ± 0.017	723.7 ± 253.0
		V	13.8 ± 1.1	311.5 ± 10.5	0.981 ± 0.003	489.7 ± 159.8
WP-VecEqui	10.43 ± 2.85	C	28.3 ± 2.2	645.6 ± 68.9	0.919 ± 0.012	848.0 ± 157.5
		V	20.7 ± 1.9	440.3 ± 38.3	0.957 ± 0.008	464.7 ± 111.4
WP-Reg	12.73 ± 1.67	C	31.3 ± 0.4	725.4 ± 21.2	0.902 ± 0.003	529.3 ± 159.7
		V	15.8 ± 1.2	336.3 ± 36.4	0.975 ± 0.004	408.7 ± 117.9
WP-RegTP	11.41 ± 3.63	C	28.5 ± 1.9	662.8 ± 56.5	0.919 ± 0.011	556.0 ± 92.0
		V	21.0 ± 3.7	495.1 ± 80.9	0.955 ± 0.016	517.7 ± 15.4
WP-O3	4.78 ± 1.16	C	$\mathbf{16.2 \pm 0.9}$	$\mathbf{369.8 \pm 30.6}$	$\mathbf{0.974 \pm 0.003}$	861.3 ± 116.2
		V	$\mathbf{10.3 \pm 0.6}$	$\mathbf{230.3 \pm 11.3}$	$\mathbf{0.989 \pm 0.001}$	237.0 ± 34.4
Full-cell baselines						
Full-cell invariant GNN	6.87 ± 1.70	C	24.4 ± 2.5	553.5 ± 56.3	0.940 ± 0.013	967.3 ± 34.9
		V	13.6 ± 1.4	306.7 ± 32.0	0.981 ± 0.004	704.3 ± 131.3
Full-cell e3nn ($\ell_{\max} = 1$)	$\mathbf{4.57 \pm 0.55}$	C	18.2 ± 1.4	412.8 ± 26.4	0.967 ± 0.005	864.3 ± 137.3
		V	11.6 ± 1.0	252.9 ± 16.5	0.986 ± 0.002	455.7 ± 160.3
Full-cell e3nn ($\ell_{\max} = 2$)	4.87 ± 0.63	C	18.6 ± 2.2	430.5 ± 41.7	0.965 ± 0.008	938.0 ± 65.1
		V	11.1 ± 0.5	246.0 ± 5.2	0.988 ± 0.001	670.0 ± 149.0

Table 42: Toy-model performance for wallpaper group $\mathrm{p4g}$ (1,000 structures, mean $\Delta E = -6.92$ eV atom^{-1}, energy SD 15.21 meV atom^{-1}).

Model	Energy MAE ↓ (meV atom^{-1})	Mode	Rel. force L_2 ↓ (%)	Force comp. MAE ↓ (meV Å^{-1})	Force comp. R^2 ↑	Best epoch
WallpaperNet models						
WP-Inv	7.51 ± 1.11	C	33.5 ± 3.2	753.4 ± 66.7	0.887 ± 0.021	614.3 ± 218.9
		V	12.6 ± 0.4	266.6 ± 18.3	0.984 ± 0.001	471.3 ± 71.8
WP-InvExt	4.90 ± 0.25	C	24.5 ± 2.9	519.8 ± 33.5	0.939 ± 0.015	756.7 ± 136.0
		V	12.4 ± 0.8	260.1 ± 13.5	0.985 ± 0.002	433.3 ± 119.7
WP-Equi	11.13 ± 5.53	C	37.1 ± 3.2	804.2 ± 58.0	0.862 ± 0.024	844.3 ± 136.8
		V	14.2 ± 0.8	309.9 ± 14.8	0.980 ± 0.002	367.7 ± 70.7
WP-VecEqui	8.69 ± 3.14	C	32.0 ± 3.4	704.5 ± 79.0	0.897 ± 0.022	826.7 ± 144.8
		V	18.4 ± 3.9	394.7 ± 83.2	0.965 ± 0.015	346.7 ± 131.2
WP-Reg	7.81 ± 1.46	C	33.8 ± 1.6	755.4 ± 27.1	0.886 ± 0.010	690.3 ± 205.2
		V	16.6 ± 0.9	340.7 ± 21.3	0.972 ± 0.003	344.3 ± 46.9
WP-RegTP	8.51 ± 0.30	C	31.9 ± 1.9	721.5 ± 38.0	0.898 ± 0.012	477.7 ± 53.5
		V	31.3 ± 5.8	699.0 ± 106.5	0.900 ± 0.037	250.7 ± 104.2
WP-O3	$\mathbf{4.61 \pm 1.71}$	C	$\mathbf{17.7 \pm 1.0}$	$\mathbf{388.9 \pm 15.5}$	$\mathbf{0.969 \pm 0.003}$	712.3 ± 227.4
		V	$\mathbf{9.7 \pm 0.8}$	$\mathbf{207.0 \pm 14.4}$	$\mathbf{0.990 \pm 0.002}$	238.7 ± 32.9
Full-cell baselines						
Full-cell invariant GNN	6.20 ± 2.88	C	26.8 ± 6.4	586.7 ± 160.4	0.925 ± 0.037	820.7 ± 116.7
		V	35.6 ± 34.9	772.7 ± 773.9	0.792 ± 0.318	543.0 ± 414.1
Full-cell e3nn ($\ell_{\max} = 1$)	4.81 ± 0.66	C	20.4 ± 1.9	455.4 ± 30.9	0.958 ± 0.008	697.3 ± 133.2
		V	11.3 ± 1.5	235.4 ± 34.6	0.987 ± 0.003	548.0 ± 26.2
Full-cell e3nn ($\ell_{\max} = 2$)	4.63 ± 0.32	C	20.0 ± 0.3	453.9 ± 22.1	0.960 ± 0.001	811.3 ± 209.9
		V	13.1 ± 1.3	284.4 ± 29.7	0.983 ± 0.004	764.0 ± 208.9

Table 43: Toy-model performance for wallpaper group $\mathrm{p6m}$ (1,000 structures, mean $\Delta E = -7.34$ eV atom^{-1}, energy SD 15.73 meV atom^{-1}).

Model	Energy MAE ↓ (meV atom^{-1})	Mode	Rel. force L_2 ↓ (%)	Force comp. MAE ↓ (meV Å^{-1})	Force comp. R^2 ↑	Best epoch
WallpaperNet models						
WP-Inv	8.94 ± 1.30	C	29.4 ± 3.1	665.0 ± 73.4	0.913 ± 0.018	775.0 ± 171.8
		V	12.6 ± 0.6	270.7 ± 14.0	0.984 ± 0.002	349.7 ± 159.0
WP-InvExt	6.70 ± 0.42	C	22.7 ± 1.8	502.7 ± 24.4	0.948 ± 0.008	576.7 ± 86.4
		V	12.5 ± 0.5	265.3 ± 20.8	0.984 ± 0.001	526.7 ± 284.7
WP-Equi	11.53 ± 4.12	C	37.5 ± 15.2	848.9 ± 350.7	0.844 ± 0.127	533.3 ± 138.0
		V	12.4 ± 1.4	275.2 ± 33.5	0.985 ± 0.003	380.7 ± 75.1
WP-VecEqui	10.25 ± 2.64	C	31.3 ± 0.3	708.5 ± 24.6	0.902 ± 0.002	922.3 ± 112.9
		V	22.7 ± 2.4	480.5 ± 42.1	0.948 ± 0.011	419.0 ± 105.2
WP-Reg	9.02 ± 1.51	C	30.6 ± 2.5	694.0 ± 40.9	0.906 ± 0.016	939.3 ± 89.5
		V	19.6 ± 3.0	386.9 ± 36.6	0.961 ± 0.012	333.7 ± 21.2
WP-RegTP	9.26 ± 3.13	C	26.6 ± 4.6	614.4 ± 104.9	0.928 ± 0.025	765.7 ± 108.7
		V	23.5 ± 1.7	560.5 ± 53.7	0.944 ± 0.008	446.3 ± 187.9
WP-O3	$\mathbf{5.14 \pm 1.35}$	C	$\mathbf{17.2 \pm 2.3}$	$\mathbf{387.2 \pm 52.7}$	$\mathbf{0.970 \pm 0.008}$	839.3 ± 102.6
		V	$\mathbf{10.0 \pm 0.8}$	$\mathbf{212.4 \pm 6.6}$	$\mathbf{0.990 \pm 0.002}$	268.0 ± 32.8
Full-cell baselines						
Full-cell invariant GNN	6.65 ± 2.25	C	21.7 ± 2.9	479.0 ± 72.6	0.953 ± 0.012	821.7 ± 149.1
		V	13.7 ± 1.2	287.7 ± 24.7	0.981 ± 0.003	649.0 ± 196.4
Full-cell e3nn ($\ell_{\max} = 1$)	5.29 ± 0.31	C	17.5 ± 0.8	392.1 ± 26.8	0.969 ± 0.003	991.0 ± 4.4
		V	10.8 ± 0.2	237.2 ± 11.3	0.988 ± 0.001	437.0 ± 68.9
Full-cell e3nn ($\ell_{\max} = 2$)	5.61 ± 0.39	C	19.1 ± 1.5	428.8 ± 31.2	0.963 ± 0.006	942.7 ± 81.4
		V	10.5 ± 0.2	228.9 ± 9.0	0.989 ± 0.000	720.3 ± 285.1

I MOLECULE–GRAPHENE SYSTEM RESULTS

Table 44 reports the complete test-set results averaged over the three random seeds, including the zero-force score, with the C and V training modes in separate rows.

Table 44: Complete test-set results of the seven wallpaper-symmetry models and three full-cell baselines on graphene, including the zero-force score. Energy metrics are taken from the combined experiments, while force metrics are reported in separate C and V rows, where C denotes force-weighted conservative training and V denotes direct vector-head training. Values are means and sample standard deviations over three random seeds. Lower is better for the error metrics, and higher is better for the zero-force score and R^2.

Model	Energy MAE ↓ (meV atom^{-1})	Mode	Rel. force L_2 ↓ (%)	Zero-force score ↑ (%)	Force comp. MAE ↓ (meV Å^{-1})	Force comp. R^2 ↑
Wallpaper-symmetry models						
WP-Inv	$\mathbf{1.73 \pm 0.02}$	C	80.79 ± 1.35	34.72 ± 2.19	139.9 ± 0.6	0.35 ± 0.02
		V	41.35 ± 7.79	82.50 ± 6.64	75.0 ± 16.3	0.82 ± 0.07
WP-InvEq	4.75 ± 5.09	C	78.68 ± 1.15	38.08 ± 1.81	137.1 ± 2.0	0.38 ± 0.02
		V	25.67 ± 0.63	93.41 ± 0.33	43.5 ± 1.0	0.93 ± 0.00
WP-Equi	1.90 ± 0.17	C	80.59 ± 0.94	35.05 ± 1.51	140.7 ± 1.4	0.35 ± 0.02
		V	27.95 ± 0.80	92.18 ± 0.44	48.8 ± 0.9	0.92 ± 0.00
WP-VecInv	2.24 ± 0.06	C	62.49 ± 15.95	59.26 ± 21.39	109.5 ± 25.4	0.59 ± 0.21
		V	26.62 ± 0.93	92.91 ± 0.50	44.8 ± 0.8	0.93 ± 0.01
WP-Reg	2.09 ± 0.12	C	30.70 ± 1.90	90.55 ± 1.19	54.0 ± 2.7	0.91 ± 0.01
		V	27.15 ± 1.07	92.62 ± 0.58	45.7 ± 1.8	0.93 ± 0.01
WP-RegTP	1.86 ± 0.24	C	30.49 ± 1.33	90.69 ± 0.82	52.7 ± 2.3	0.91 ± 0.01
		V	26.21 ± 1.14	93.12 ± 0.60	44.6 ± 1.1	0.93 ± 0.01
WP-O3	1.80 ± 0.08	C	$\mathbf{19.78 \pm 0.70}$	$\mathbf{96.09 \pm 0.28}$	$\mathbf{32.9 \pm 0.9}$	$\mathbf{0.96 \pm 0.00}$
		V	$\mathbf{19.22 \pm 0.75}$	$\mathbf{96.30 \pm 0.29}$	$\mathbf{31.3 \pm 0.2}$	$\mathbf{0.96 \pm 0.00}$
Full-cell baselines						
Full cell invariant GNN	2.58 ± 0.23	C	67.51 ± 1.03	54.42 ± 1.39	119.9 ± 2.5	0.54 ± 0.01
		V	47.22 ± 6.46	77.42 ± 5.93	83.6 ± 12.2	0.77 ± 0.06
Full cell e3nn $\ell_{\max} = 1$	1.92 ± 0.07	C	25.89 ± 0.89	93.29 ± 0.46	44.7 ± 2.1	0.93 ± 0.00
		V	25.31 ± 3.50	93.51 ± 1.83	43.9 ± 7.2	0.94 ± 0.02
Full cell e3nn $\ell_{\max} = 2$	2.04 ± 0.22	C	21.43 ± 0.75	95.40 ± 0.32	36.2 ± 1.2	0.95 ± 0.00
		V	25.55 ± 5.82	93.24 ± 2.81	45.0 ± 11.1	0.93 ± 0.03

I.1 DISCUSSION OF THE REAL-SYSTEM RESULTS

The real-system experiments address three related questions, summarized in Table 2: how accurately the models predict the total energy, how accurately they predict forces directly, and whether comparable forces can be recovered conservatively as derivatives of a learned energy. The energy values, conservative-force metrics, and direct-force metrics therefore correspond to three different training objectives. The force-weighted conservative objective and the per-seed results are given in Appendix I.2.

For the first task, training is dominated by the total-energy objective, with forces included as an auxiliary target. Under this setting, all WallpaperNet models remain far below the commonly used chemical-accuracy threshold of $43\,\mathrm{meV\,atom^{-1}}$. WP-Inv gives the lowest energy MAE at $1.73\,\mathrm{meV\,atom^{-1}}$, followed by WP-O3 at $1.80\,\mathrm{meV\,atom^{-1}}$, WP-RegTP at $1.86\,\mathrm{meV\,atom^{-1}}$, and WP-Equi at $1.90\,\mathrm{meV\,atom^{-1}}$. These values are also below the full-cell e3nn baselines at 1.92 and $2.04\,\mathrm{meV\,atom^{-1}}$ for $\ell_{\max} = 1$ and 2, respectively. The exception is WP-InvEq, which reaches $4.75 \pm 5.09\,\mathrm{meV\,atom^{-1}}$ and is therefore considerably less stable across seeds. Overall, the energy results show that the additional equivariant complexity is not required for this scalar target: the simplest invariant WP-Inv model gives the lowest mean error.

The same energy-dominated objective does not, however, provide useful conservative forces. The full-system energy residual has a much larger numerical scale than the component-wise force loss, while force supervision is restricted to the non-frozen adsorbate atoms. With equal loss weights, optimization therefore primarily reduces the energy error and the resulting energy gradients tend toward small forces. This motivates treating force prediction as a separate problem. The toy experiments already showed that direct vector-head training generally gives substantially more accurate forces than differentiating the energy, so we first evaluate the same strategy on the real system.

For direct force prediction, WP-O3 gives the lowest error of all ten models, with a relative force error of 19.22%, a component MAE of $31.3\,\mathrm{meV\,\AA^{-1}}$, and $R^2 = 0.96$. The full-cell e3nn baselines reach 25.31% and 25.55% relative error for $\ell_{\max} = 1$ and 2, respectively. Among the other WallpaperNet models, WP-InvEq, WP-VecInv, WP-RegTP, and WP-Reg lie in a narrow range of 25.67–27.15%, whereas WP-Equi reaches 27.95%. WP-Inv performs substantially worse at 41.35%, similar to the full-cell invariant GNN at 47.22%. Thus, the model ordering for forces differs clearly from that for energies: WP-Inv gives the lowest energy error but one of the largest direct-force errors, whereas WP-O3 is competitive for energy and gives the lowest force error.

We next test whether accurate forces can also be obtained conservatively. To prevent the energy term from dominating the optimization, we rescale it using the empirical ratio between the force-MSE and energy-L1 scales and the mean adsorbate-to-full-system atom-count ratio, resulting in an energy weight of approximately 1.22×10^{-6} relative to a unit force weight. The latter atom-count factor is an additional weighting choice rather than a required normalization, since both losses are already mean-reduced.

The reweighted conservative results show a strong dependence on architecture. WP-O3 reaches 19.78% relative force error, $32.9\,\mathrm{meV\,\AA^{-1}}$ component MAE, and $R^2 = 0.96$, compared with 19.22%, $31.3\,\mathrm{meV\,\AA^{-1}}$, and $R^2 = 0.96$ for its direct head. Thus, for WP-O3, deriving the forces from the energy introduces only a small loss in accuracy once the objective is appropriately balanced. The full-cell e3nn models also perform well conservatively, reaching relative errors of 25.89% for $\ell_{\max} = 1$ and 21.43% for $\ell_{\max} = 2$.

The same recovery is not observed for all architectures. WP-Reg and WP-RegTP reach conservative relative errors of 30.70% and 30.49%, respectively, while WP-VecInv increases to $62.49 \pm 15.95\%$. WP-Inv, WP-InvEq, and WP-Equi remain around 79–81%, and the full-cell invariant GNN reaches 67.51%. In particular, an equivariant force output alone is insufficient: WP-InvEq improves substantially over WP-Inv for direct force regression (25.67% versus 41.35%) but gives almost the same poor conservative performance (78.68% versus 80.79%). The conservative results therefore distinguish architectures that can produce an equivariant vector output from those that can encode sufficient directional information in the scalar energy surface itself. Among the tested models, WP-O3 does this most effectively.

Finally, WallpaperNet obtains these results on a substantially smaller message-passing graph. The WallpaperNet representation contains on average 25 nodes and 512 edges per system, whereas the full-cell representation contains 386 nodes and 22,929 edges, corresponding to approximately $15\times$ more nodes and $45\times$ more edges. The $45\times$ edge reduction should not be interpreted as a $45\times$ wall-clock speedup, since actual cost also depends on representation size, tensor-product operations, hardware, batch size, and training objective. Peak GPU memory and the accuracy–memory trade-off are reported in Appendix I.3. We additionally benchmark all ten architectures under both conservative and direct-force training at approximately 90% GPU-memory utilization. These maximum-batch throughput results are reported in Appendix I.4.

Additional observations on the averaged results. Beyond the discussion above, the averaged results in Table 44 show the following. WP-O3 reaches the highest zero-force score in both training modes, 96.09% with conservative and 96.30% with direct-head training. In direct-force training, the full-cell invariant GNN is the weakest full-cell model, and WP-Inv reaches a component MAE of $75.0\,\mathrm{meV\,\AA^{-1}}$. The explicit equivariant output maps of WP-InvEq reduce this to $43.5\,\mathrm{meV\,\AA^{-1}}$, making WP-InvEq the second-best WallpaperNet model for direct-force component MAE despite being the least stable model for energy prediction. WP-VecInv, WP-Reg, and WP-RegTP reach a similar direct-force regime, with component MAEs between 44.6 and $45.7\,\mathrm{meV\,\AA^{-1}}$, so increasing

the complexity of the wallpaper-equivariant architecture does not by itself guarantee improved accuracy. Finally, the contrast between the energy and the conservative-force rankings indicates that accurate prediction of scalar energy values and accurate prediction of the local derivatives of the energy surface are related but not equivalent learning problems: an invariant representation can be sufficient for the former, whereas the latter benefits much more strongly from explicit directional equivariant features.

I.2 Per-seed results and force-weighted conservative objective

Tables 45, 46, and 47 report the test-set results for the three random seeds separately. Energy metrics are taken from the combined experiments, whereas force metrics are taken from the direct force-head experiments. The energy and force-head best epochs are reported as E/F, respectively. For the real-system conservative force experiment, we found that the unweighted joint objective, $\lambda_E = \lambda_F = 1$, was strongly imbalanced because the numerical scale of the full-system energy residual was much larger than that of the component-wise molecular force MSE. In practice, this caused optimization to be dominated by the energy term and the predicted energy surface to become nearly flat with respect to the adsorbate coordinates, resulting in forces that collapsed toward zero.

We therefore introduced a force-weighted conservative objective. The energy coefficient was determined exclusively from the cached training split as

$$\lambda_E = \frac{\left\langle F_{\mathrm{mol}}^2 \right\rangle}{\left\langle |E_{\mathrm{total}} - E_{\mathrm{atom-ref}} - E_{\mathrm{graph-ref}}| \right\rangle} \frac{\langle N_{\mathrm{mol}} \rangle}{\langle N_{\mathrm{full}} \rangle}, \qquad \lambda_F = 1. \tag{111}$$

The first ratio approximately matches the numerical scales of the energy-L1 and force-MSE terms. The additional atom-count factor accounts for the fact that the energy target corresponds to the complete system, whereas force supervision is restricted to the adsorbate atoms. Tables 48, 49, and 50 report the corresponding force-weighted conservative experiments. Across the three training splits, the resulting coefficients were 1.21915×10^{-6}, 1.22801×10^{-6}, and 1.22606×10^{-6}, with a mean of 1.22441×10^{-6}. The force coefficient was fixed at $\lambda_F = 1$.

The force-weighted conservative experiments show that the poor force performance of the original 1:1 combined objective was primarily a loss-balancing problem rather than a fundamental limitation of energy-derived force prediction. In contrast to the fixed-size toy systems, the molecule–graphene dataset contains full cells ranging from 158 to 699 atoms, with a mean of 386.825 atoms and a standard deviation of 51.593. The corresponding total system energies range from $-40{,}589.4$ to $-9{,}197.05\,\mathrm{eV}$. After rescaling the energy contribution, several architectures recover meaningful gradients while retaining forces that are conservative by construction.

WP-O3 benefits most clearly from the reweighted objective. Across the three force-weighted runs, it reaches relative force errors of 20.57%, 19.48%, and 19.28%, component MAEs of 33.7, 33.2, and $31.9\,\mathrm{meV\,\AA^{-1}}$, and $R^2 = 0.96$ in all three cases. These values are remarkably close to the direct vector-head results, for which WP-O3 gives relative force errors of 18.46%–19.97% across the three reported seeds and component MAEs of approximately $31\,\mathrm{meV\,\AA^{-1}}$. Thus, for WP-O3, most of the accuracy advantage previously associated with direct force regression can be recovered while enforcing the physically stronger constraint that forces derive from a scalar energy. Among the force-weighted conservative models, WP-O3 provides the clearest simultaneous satisfaction of both energy and force requirements. Its energy MAE remains below the commonly used $43\,\mathrm{meV\,atom^{-1}}$ chemical-accuracy threshold in both reported runs. WP-Reg and WP-RegTP also remain below the energy threshold in both runs, but their force errors are substantially larger, at approximately 30%. The full-cell e3nn $\ell_{\max} = 2$ model approaches WP-O3 in force accuracy, but crosses the chemical-accuracy threshold in one of the reported runs. Thus, WP-O3 gives the most consistent balance between chemically accurate energies and accurate conservative forces.

WP-VecInv is particularly unstable under the force-weighted conservative objective. The results therefore indicate that accurate direct equivariant force prediction does not automatically imply that the same architecture can learn an energy surface with equally accurate derivatives.

Among the full-cell baselines, the $O(3)$-equivariant models also respond much more successfully to force-weighted conservative training than the invariant GNN. In particular, the full-cell e3nn $\ell_{\max} = 2$ model reaches relative force errors of 21.21%, 20.81%, and 22.27%, approaching WP-O3 but remaining consistently less accurate in the reported runs. The $\ell_{\max} = 2$ model also performs

Table 45: Test-set performance for random seed 1. Lower is better for the error metrics, and higher is better for the zero-force score and R^2.

Model	Energy MAE ↓ (meV system^{-1})	Energy MAE ↓ (meV atom^{-1})	Rel. force L_2 ↓ (%)	Zero-force score ↑ (%)	Force comp. MAE ↓ (meV Å^{-1})	Force comp. R^2 ↑	Best epoch E / F
Wallpaper-symmetry models							
WP-Inv	672.20	1.75	34.58	88.04	61.8	0.88	352 / 33
WP-InvEq	4077.70	10.62	26.38	93.04	44.6	0.93	46 / 166
WP-Equi	761.76	2.02	28.66	91.78	49.8	0.92	546 / 273
WP-VecInv	817.05	2.17	26.23	93.12	44.1	0.93	429 / 174
WP-Reg	740.18	1.97	25.93	93.28	43.7	0.93	544 / 246
WP-RegTP	**647.53**	**1.73**	26.23	93.12	44.4	0.93	525 / 172
WP-O3	709.95	1.88	**19.23**	**96.30**	**31.4**	**0.96**	462 / 200
Full-cell baselines							
Full cell invariant GNN	957.12	2.56	52.46	72.48	93.2	0.72	590 / 47
Full cell e3nn $\ell_{\max} = 1$	727.08	1.92	29.25	91.44	52.1	0.91	591 / 87
Full cell e3nn $\ell_{\max} = 2$	771.98	2.02	29.89	91.07	52.7	0.91	584 / 57

Table 46: Test-set performance for random seed 24. Lower is better for the error metrics, and higher is better for the zero-force score and R^2.

Model	Energy MAE ↓ (meV system^{-1})	Energy MAE ↓ (meV atom^{-1})	Rel. force L_2 ↓ (%)	Zero-force score ↑ (%)	Force comp. MAE ↓ (meV Å^{-1})	Force comp. R^2 ↑	Best epoch E / F
Wallpaper-symmetry models							
WP-Inv	664.42	1.72	49.86	75.14	93.3	0.75	504 / 9
WP-InvEq	676.98	1.80	25.17	93.66	43.6	0.94	587 / 188
WP-Equi	753.69	1.98	27.09	92.66	47.9	0.93	587 / 289
WP-VecInv	864.21	2.28	25.94	93.27	44.6	0.93	352 / 266
WP-Reg	848.52	2.22	27.59	92.39	47.0	0.92	436 / 508
WP-RegTP	**653.28**	**1.71**	25.05	93.72	43.6	0.94	595 / 157
WP-O3	665.37	1.72	**18.46**	**96.59**	**31.1**	**0.97**	595 / 189
Full-cell baselines							
Full cell invariant GNN	871.53	2.36	49.20	75.79	87.7	0.76	598 / 56
Full cell e3nn $\ell_{\max} = 1$	703.26	1.85	22.55	94.91	39.0	0.95	591 / 228
Full cell e3nn $\ell_{\max} = 2$	869.10	2.27	18.93	96.41	32.4	0.96	594 / 208

noticeably better than the $\ell_{\max} = 1$ model for conservative forces, suggesting that the higher-order geometric features become more useful when accurate force gradients must be encoded through a scalar energy surface. This behavior differs from the simpler averaged toy benchmark, where increasing $\ell_{\max}$ did not provide a systematic advantage. On average, for force prediction, models whose internal message-passing representation remains predominantly invariant and scalar tend to perform worse than architectures that propagate explicit equivariant directional features. This trend is visible both for the WallpaperNet models and for the full-cell baselines.

The improved conservative forces come at a clear cost in energy accuracy. In the energy-oriented combined experiments, the strongest models achieve errors of approximately 1.7–$2.0\,\mathrm{meV\,atom^{-1}}$, whereas the force-weighted conservative WP-O3 runs give 21.78–$33.24\,\mathrm{meV\,atom^{-1}}$. The reweighted objective should therefore dose not benefit both, energy and forces, but shifts the optimization toward the local derivatives of the energy surface. The direct force head remains the simplest route to high force accuracy, whereas the force-weighted conservative formulation demonstrates that comparable force accuracy can be approached while retaining exact energy-force consistency.

Taken together, these experiments reveal a three-way trade-off between energy accuracy, force accuracy, and conservativity. The unweighted combined objective favors accurate total energies but fails to resolve the molecular forces. Direct vector regression gives accurate forces without constraining them to derive from an energy. The force-weighted conservative objective sacrifices some energy accuracy to recover accurate gradients. WP-O3 performs particularly well in this third regime, indicating that its coupled $O(3)$–wallpaper representation is effective not only for predicting equivariant vectors directly, but also for representing the directional structure required in the derivatives of a scalar potential-energy surface.

I.3 Memory footprint of molecule–graphene systems

The roughly $45\times$ smaller edge count of the WallpaperNet graphs (Section 4.2) primarily indicates the reduction in system size and memory footprint. We report the peak training VRAM consumption across three independently initialized runs. Figure 5 compares this memory footprint with test-set

Table 47: Test-set performance for random seed 42. Lower is better for the error metrics, and higher is better for the zero-force score and R^2.

Model	Energy MAE ↓ (meV system^{-1})	Energy MAE ↓ (meV atom^{-1})	Rel. force L_2 ↓ (%)	Zero-force score ↑ (%)	Force comp. MAE ↓ (meV Å^{-1})	Force comp. R^2 ↑	Best epoch E / F
Wallpaper-symmetry models							
WP-Inv	**638.28**	**1.70**	39.61	84.31	70.0	0.84	529 / 25
WP-InvEq	687.08	1.83	25.46	93.52	42.5	0.94	556 / 213
WP-Equi	639.31	1.71	28.11	92.10	48.9	0.92	584 / 287
WP-VecInv	852.27	2.27	27.68	92.34	45.6	0.92	400 / 180
WP-Reg	774.71	2.08	27.93	92.20	46.3	0.92	422 / 375
WP-RegTP	810.42	2.14	27.34	92.53	45.9	0.93	516 / 173
WP-O3	670.47	1.81	**19.97**	**96.01**	**31.3**	**0.96**	508 / 205
Full-cell baselines							
Full cell invariant GNN	1021.78	2.82	40.00	84.00	69.9	0.84	597 / 86
Full cell e3nn $\ell_{\max} = 1$	738.95	1.99	24.12	94.18	40.5	0.94	582 / 232
Full cell e3nn $\ell_{\max} = 2$	678.18	1.83	27.84	92.25	50.1	0.92	597 / 80

Table 48: Test-set performance for random seed 1 using the force-weighted energy/force objective. Lower is better for the error metrics, and higher is better for the zero-force score and R^2.

Model	Energy MAE ↓ (meV system^{-1})	Energy MAE ↓ (meV atom^{-1})	Rel. force L_2 ↓ (%)	Zero-force score ↑ (%)	Force comp. MAE ↓ (meV Å^{-1})	Force comp. R^2 ↑	Best epoch
Wallpaper-symmetry models							
WP-Inv	321,562.00	889.75	80.04	35.94	139.6	0.36	380
WP-InvEq	290,840.00	785.90	79.53	36.75	139.1	0.37	379
WP-Equi	290,464.00	773.53	79.51	36.79	139.1	0.37	398
WP-VecInv	3,694,400.00	9573.70	52.84	72.08	92.7	0.72	82
WP-Reg	**5250.41**	**13.79**	32.89	89.18	57.1	0.89	399
WP-RegTP	8162.55	21.39	29.96	91.02	51.0	0.91	394
WP-O3	11,930.60	30.96	**20.57**	**95.77**	**33.7**	**0.96**	347
Full-cell baselines							
Full cell invariant GNN	49,695.00	128.90	68.67	52.85	122.8	0.53	172
Full cell e3nn $\ell_{\max} = 1$	25,607.20	65.83	25.60	93.45	43.7	0.93	382
Full cell e3nn $\ell_{\max} = 2$	28,302.40	73.69	21.21	95.50	35.3	0.96	373

accuracy for energy prediction and direct force-head prediction. Peak VRAM is measured during training, while accuracy is evaluated on the held-out test set. Table 51 reports the peak-memory values used to construct the figure.

For energy training, the molecule-only WallpaperNet models require 0.72–$10.15\,\mathrm{GiB}$, whereas the full-cell invariant, e3nn $\ell_{\max} = 1$, and e3nn $\ell_{\max} = 2$ baselines require 22.42, 23.43, and $43.27\,\mathrm{GiB}$, respectively. For direct force-head training, the molecule-only models require 0.16–$1.57\,\mathrm{GiB}$, compared with 3.94, 4.82, and $10.79\,\mathrm{GiB}$ for the corresponding full-cell baselines.

WP-O3 provides the best force accuracy, reaching a mean force-component MAE of $31.3\,\mathrm{meV\,\AA^{-1}}$ with a mean peak memory consumption of $0.61\,\mathrm{GiB}$. For energy prediction, WP-Inv achieves the lowest mean system-atom MAE of $1.73\,\mathrm{meV\,atom^{-1}}$ while requiring only $0.72\,\mathrm{GiB}$. These results show that the molecule-only models can match or improve upon the accuracy of the full-cell baselines while using substantially less training memory.

I.4 One-batch maximum utilization speed tests

We benchmarked all ten architectures in both combined and direct force-head training modes on a single NVIDIA GH200 GPU, giving 20 runs in total. For each architecture and mode, the maximum batch size was estimated from the median optimizer-training peak measured at a batch size of 16 per GPU. We targeted 90% of the available $93\,\mathrm{GiB}$, or $83.7\,\mathrm{GiB}$, and used

$$B_{\max} = \left\lfloor \frac{16 \times 83.7\,\mathrm{GiB}}{M_{16}} \right\rfloor, \tag{112}$$

where M_{16} is the measured median peak memory at batch size 16. Each timed fit comprised 30 epochs and included validation, checkpointing, and offline Weights & Biases logging. Inference was performed from the best checkpoint after fitting, and the complete time covers the entire Python run.

For combined training, the molecule-only models complete in 1.7–$77.3\,\mathrm{min}$, with test inference requiring 0.9–$7.5\,\mathrm{s}$. The corresponding full-cell ranges are 46.0–$193.0\,\mathrm{min}$ and 5.3–$17.3\,\mathrm{s}$, respectively.

Table 49: Test-set performance for random seed 24 using the force-weighted energy/force objective. Lower is better for the error metrics, and higher is better for the zero-force score and R^2.

Model	Energy MAE ↓ (meV system^{-1})	Energy MAE ↓ (meV atom^{-1})	Rel. force L_2 ↓ (%)	Zero-force score ↑ (%)	Force comp. MAE ↓ (meV Å^{-1})	Force comp. R^2 ↑	Best epoch
Wallpaper-symmetry models							
WP-Inv	318,885.00	855.62	79.98	36.03	139.5	0.36	282
WP-InvEq	295,801.00	779.89	77.37	40.14	135.0	0.40	354
WP-Equi	308,116.00	788.64	81.09	34.25	141.8	0.34	164
WP-VecInv	3,677,060.00	9370.42	53.73	71.13	96.9	0.71	142
WP-Reg	12,556.60	33.10	29.65	91.21	53.0	0.91	89
WP-RegTP	**5998.09**	**15.28**	29.51	91.29	51.7	0.91	395
WP-O3	12,959.50	33.24	**19.48**	**96.21**	**33.2**	**0.96**	258
Full-cell baselines							
Full cell invariant GNN	60,007.70	148.33	66.72	55.49	119.1	0.55	126
Full cell e3nn $\ell_{\max} = 1$	15,967.20	40.84	26.89	92.77	47.2	0.93	296
Full cell e3nn $\ell_{\max} = 2$	17,415.70	46.21	20.81	95.67	35.6	0.96	390

Table 50: Test-set performance for random seed 67 using the force-weighted energy/force objective. Lower is better for the error metrics, and higher is better for the zero-force score and R^2.

Model	Energy MAE ↓ (meV system^{-1})	Energy MAE ↓ (meV atom^{-1})	Rel. force L_2 ↓ (%)	Zero-force score ↑ (%)	Force comp. MAE ↓ (meV Å^{-1})	Force comp. R^2 ↑	Best epoch
Wallpaper-symmetry models							
WP-Inv	319,690.00	847.65	82.34	32.20	140.7	0.32	381
WP-InvEq	357,664.00	892.79	79.14	37.36	137.3	0.37	301
WP-Equi	297,028.00	773.34	81.17	34.12	141.2	0.34	381
WP-VecInv	3,749,750.00	9635.12	80.90	34.56	138.7	0.35	37
WP-Reg	13,407.80	35.74	29.56	91.26	51.9	0.91	108
WP-RegTP	**6391.74**	**16.49**	32.01	89.75	55.3	0.90	392
WP-O3	8415.36	21.78	**19.28**	**96.28**	**31.9**	**0.96**	375
Full-cell baselines							
Full cell invariant GNN	12,966.90	33.49	67.14	54.93	118.0	0.55	196
Full cell e3nn $\ell_{\max} = 1$	24,553.90	63.70	25.18	93.66	43.4	0.94	384
Full cell e3nn $\ell_{\max} = 2$	14,300.30	36.89	22.27	95.04	37.6	0.95	367

Under direct force-head training, the molecule-only models complete in 1.2–$31.4\,\mathrm{min}$ and require 0.7–$2.4\,\mathrm{s}$ for test inference, compared with 21.0–$58.8\,\mathrm{min}$ and 3.6–$7.0\,\mathrm{s}$ for the full-cell baselines.

Figure 6 relates the timing measurements to the corresponding seed-42 test errors. The complete runtime includes training, validation, checkpointing, and inference, whereas the test-inference panels isolate prediction from the best checkpoint. Logarithmic time axes accommodate the large runtime range across architectures.

The timing results show that reducing the explicit support graph can lead to substantial computational savings, but that the magnitude of this advantage depends strongly on the operations used within the symmetry-reduced model. The simplest invariant WallpaperNet architectures occupy an especially low-cost regime. Under combined training, WP-Inv and WP-InvEq complete the full 30-epoch run in approximately $2\,\mathrm{min}$ and process more than 10^4 test structures per second, whereas the full-cell baselines require 46–$193\,\mathrm{min}$ and reach approximately 10^2–4×10^2 structures per second. This large difference reflects both the substantially smaller molecule-only graph and the comparatively inexpensive scalar message-passing operations.

The reduction in graph size does not, however, make every WallpaperNet architecture faster than every full-cell baseline. WP-Reg and WP-RegTP contain multiple regular-representation pathways and explicit tensor-product operations and are correspondingly much more expensive than the invariant WallpaperNet models. Under combined training, their complete runtimes of 64.6 and $77.3\,\mathrm{min}$, respectively, exceed that of the full-cell invariant GNN ($46.0\,\mathrm{min}$). This demonstrates that computational cost is controlled by both graph size and the operations performed on each edge: symmetry reduction decreases the number of interactions that must be processed, but sufficiently expensive equivariant operations can partially or completely offset this advantage.

WP-O3 provides a particularly favorable compromise between these two effects. Despite combining an $O(3)$-equivariant tensor pathway with wallpaper-group-equivariant processing, its combined-training run completes in $50.1\,\mathrm{min}$, compared with $73.2\,\mathrm{min}$ for the full-cell e3nn $\ell_{\max} = 1$ model and $193.0\,\mathrm{min}$ for $\ell_{\max} = 2$. This corresponds to approximately $1.5\times$ and $3.9\times$ shorter complete

Table 51: Peak training VRAM at batch size 16 per GPU for the three random seeds. Values are in GiB per GPU. The final column in each training block gives the mean and sample standard deviation across seeds.

	Combined training				Direct force-head training			
Model	Seed 1	Seed 24	Seed 42	Mean ± SD	Seed 1	Seed 24	Seed 42	Mean ± SD
WP-Inv	0.725	0.726	0.721	0.724 ± 0.002	0.161	0.162	0.161	0.1612 ± 0.0005
WP-InvEq	0.741	0.743	0.737	0.741 ± 0.003	0.160	0.160	0.160	0.1601 ± 0.0004
WP-Equi	3.479	3.483	3.461	3.474 ± 0.012	0.379	0.380	0.378	0.379 ± 0.001
WP-VecInv	1.643	1.646	1.634	1.641 ± 0.006	0.388	0.389	0.386	0.388 ± 0.001
WP-Reg	8.487	8.498	8.442	8.476 ± 0.030	0.887	0.888	0.882	0.886 ± 0.003
WP-RegTP	10.161	10.175	10.107	10.148 ± 0.036	1.568	1.570	1.560	1.566 ± 0.005
WP-O3	6.364	6.373	6.329	6.355 ± 0.023	0.608	0.608	0.604	0.607 ± 0.002
Full cell invariant GNN	22.531	22.543	22.184	22.419 ± 0.204	3.961	3.963	3.901	3.942 ± 0.036
Full cell e3nn $\ell_{\max} = 1$	23.547	23.560	23.184	23.430 ± 0.214	4.838	4.850	4.764	4.817 ± 0.047
Full cell e3nn $\ell_{\max} = 2$	43.487	43.510	42.816	43.271 ± 0.394	10.847	10.852	10.680	10.793 ± 0.098

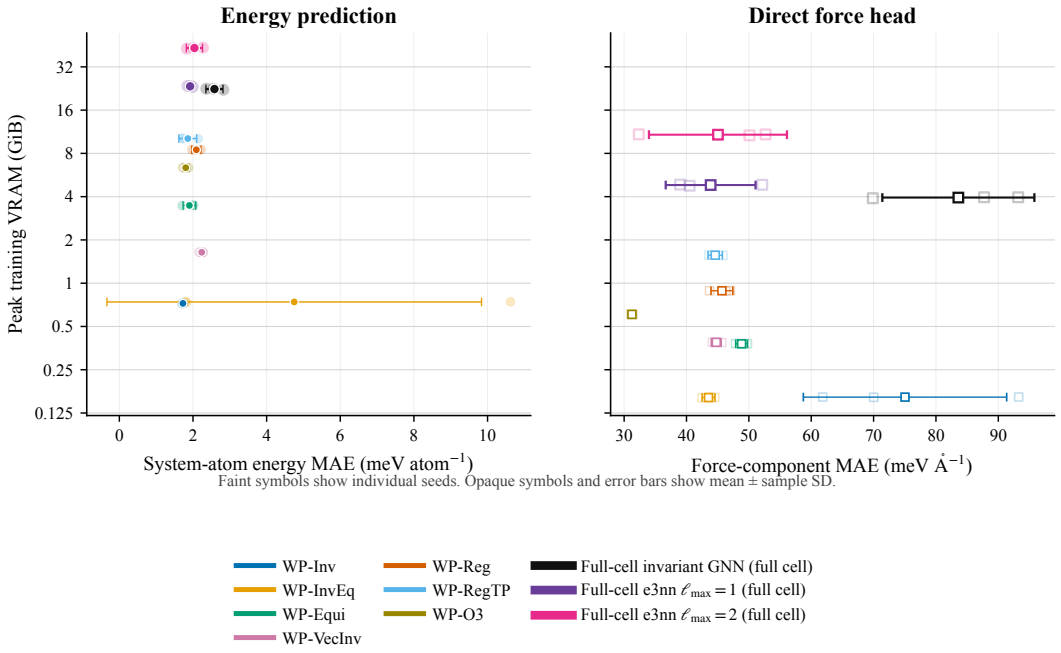


Figure 5: Peak training VRAM as a function of test-set accuracy for energy prediction (left) and direct force-head prediction (right). Energy accuracy is measured by the system-atom MAE, and force accuracy by the Cartesian force-component MAE. Faint markers show the three individual runs, opaque markers and error bars show the mean and sample standard deviation. Colors identify model architectures, and heavier full-cell markers and legend strokes distinguish the full-cell baselines. The VRAM axis is logarithmic.

runtimes, respectively. The same trend is visible during inference: WP-O3 evaluates approximately 335 test structures per second, compared with 245 and 97 structures per second for the two full-cell e3nn models. The full-cell invariant GNN remains slightly faster than WP-O3 in this mode, illustrating that the benefit should be interpreted relative to models of comparable geometric expressivity rather than as an unconditional speed advantage over every full-cell architecture.

The computational advantage becomes stronger for direct force-head training. Direct force prediction avoids differentiating forces through the coordinate gradient of the scalar energy, substantially reducing both memory requirements and computational cost. Consistent with this, every architecture supports a larger batch and a shorter runtime in the direct-head experiment than in combined training. For WP-O3, the estimated maximum batch size increases from 422 to 771, while the complete runtime decreases from 50.1 to 19.0 min. The corresponding reductions for the full-cell e3nn models

Table 52: One-GPU maximum-batch timing results for all ten architectures under combined and direct force-head training. Inference times are measured from the best checkpoint.

Model	Batch	Fit (min)	Infer train (s)	Infer val (s)	Infer test (s)	Infer total (s)	Complete (min)	Test structures s^{-1}	Test peak (GiB GPU^{-1})
Combined training									
WP-Inv	3280	1.4	2.0	0.8	0.9	3.7	1.9	10128.1	13.82
WP-InvEq	3196	1.4	2.0	0.7	0.9	3.5	1.7	10177.2	14.17
WP-Equi	887	8.2	6.9	1.3	1.4	9.7	8.6	1922.9	41.26
WP-VecInv	1547	3.9	3.7	1.0	1.1	5.7	4.3	4346.8	33.05
WP-Reg	338	63.3	48.8	6.4	6.6	61.8	64.6	257.3	40.69
WP-RegTP	275	75.9	56.5	7.4	7.5	71.4	77.3	220.8	38.97
WP-O3	422	49.0	37.5	5.1	5.2	47.8	50.1	335.1	38.08
Full cell invariant GNN	112	45.0	34.2	5.3	5.3	44.8	46.0	411.1	41.85
Full cell e3nn $\ell_{\max} = 1$	117	71.6	56.3	7.8	7.9	72.0	73.2	244.6	45.47
Full cell e3nn $\ell_{\max} = 2$	59	189.9	131.3	19.7	17.3	168.4	193.0	96.6	44.32
Direct force-head training									
WP-Inv	5326	0.9	2.0	0.7	0.9	3.5	1.2	22397.3	2.67
WP-InvEq	5256	0.9	2.0	0.6	0.7	3.3	1.2	21652.0	2.57
WP-Equi	1241	6.0	5.1	1.1	1.2	7.3	6.3	3232.7	5.11
WP-VecInv	3060	1.4	2.2	0.7	0.9	3.8	1.6	9223.1	7.41
WP-Reg	568	25.3	13.2	2.0	2.1	17.4	25.8	992.7	6.80
WP-RegTP	469	30.9	15.8	2.3	2.4	20.4	31.4	815.6	10.23
WP-O3	771	18.6	10.7	1.7	1.8	14.2	19.0	1322.2	6.16
Full cell invariant GNN	252	20.2	23.1	3.3	3.6	30.0	21.0	841.0	15.30
Full cell e3nn $\ell_{\max} = 1$	311	23.5	27.1	4.0	4.0	35.1	24.4	640.9	22.68
Full cell e3nn $\ell_{\max} = 2$	155	57.5	49.3	6.8	7.0	63.1	58.8	284.7	26.77

are also large, reflecting the additional cost of gradient-derived force supervision during combined training.

Most importantly, the direct-force timing and accuracy results show that the computational reduction does not come at the expense of predictive accuracy. WP-O3 simultaneously gives the lowest direct-force error in the real-system experiment and is faster than all three full-cell baselines. Its complete runtime is $19.0\,\mathrm{min}$, compared with 21.0, 24.4, and $58.8\,\mathrm{min}$ for the full-cell invariant, e3nn $\ell_{\max} = 1$, and e3nn $\ell_{\max} = 2$ models, respectively. At inference, WP-O3 reaches approximately 1322 test structures per second, corresponding to approximately $1.6\times$, $2.1\times$, and $4.6\times$ the throughput of these three baselines. At the same time, its force-component MAE is lower than that of every full-cell model. WP-O3 therefore occupies a favorable region of the accuracy–runtime trade-off rather than achieving its improved force accuracy through increased computational cost.

The comparison also illustrates why parameter count alone is an incomplete measure of computational complexity. All architectures were constructed with approximately matched trainable parameter counts, yet their runtimes, feasible batch sizes, and inference throughputs differ by more than an order of magnitude. Regular-representation tensor products, $O(3)$ tensor products, the number of explicit graph edges, and differentiation through conservative force predictions all contribute substantially to the actual computational cost. The timing experiment therefore complements the parameter-matched accuracy comparison by showing the practical computational consequences of the different symmetry representations.

Overall, these results support the central motivation of the symmetry-reduced construction: when the substrate symmetry is known, its effect can be represented without explicitly propagating messages through every substrate atom. The largest gains are obtained by the simpler invariant architectures, while more expressive equivariant models trade some of this efficiency for geometric capacity. WP-O3 provides the strongest balance in the present real-system experiment, retaining much of the computational advantage of the molecule-only graph while exceeding the accuracy of the substantially larger full-cell $O(3)$-equivariant models. It's important to note that the proposed WP-Nets are not the best performing designs, they are selected as illustrations of application of different parts of the proposed WallpaperNet framework. Additional, keeping the same amount of learnable parameters for all models is a choice that was made to simplify the comparison, but it is not a requirement of the proposed framework, therefore additional gains can be made.

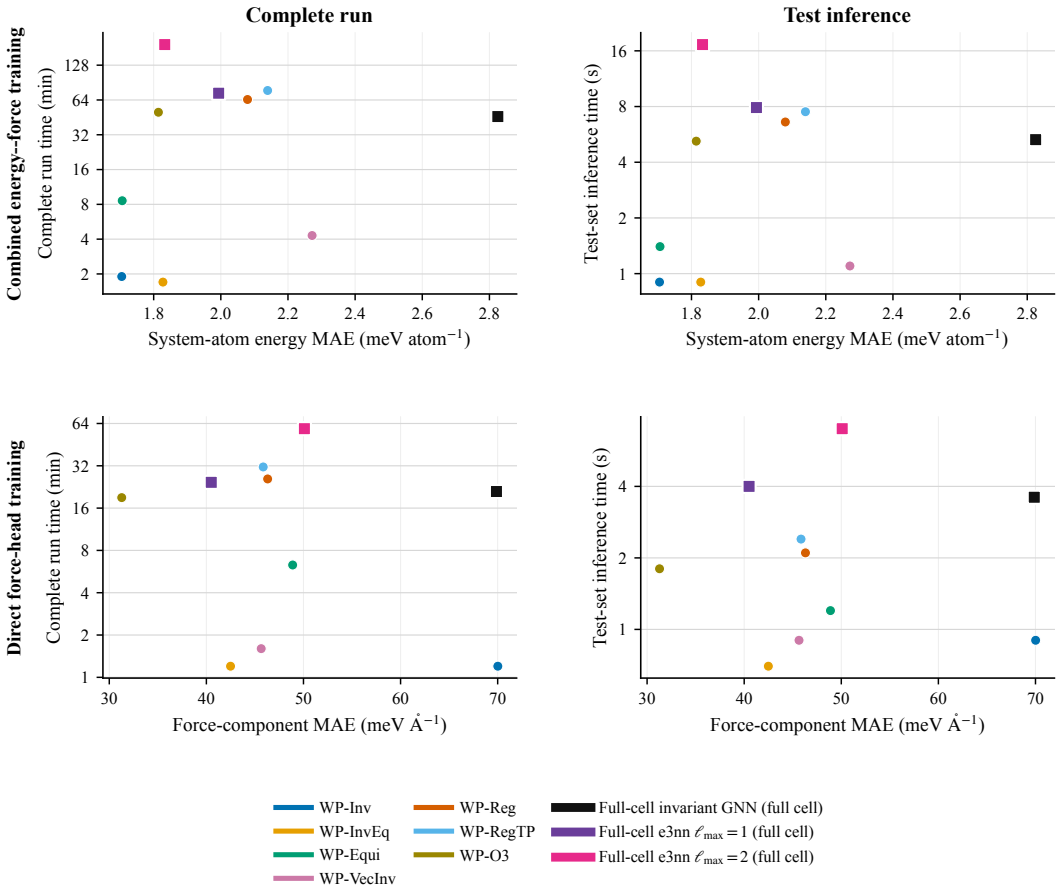


Figure 6: Accuracy–runtime trade-offs for the one-GPU maximum-batch experiments. The upper row uses the system-atom energy MAE from combined training, and the lower row uses the force-component MAE from direct force-head training. The left column reports complete Python-run time and the right column reports test-set inference time. Square markers denote full-cell baselines and circles denote molecule-only WallpaperNet models.