SYMMETRY SLIPPAGE IN EQUIVARIANT ML: CREATION, COMPUTATION, AND CORRECTIONS

Anonymous authors
Paper under double-blind review

ABSTRACT

Symmetry-based inductive biases are prevalent throughout machine learning for science and beyond. Often, the symmetries realized by an architecture differ from those prescribed. This mismatch can occur in two ways: when prescribed symmetries are not realized (*under-equivariance*) and when *non*-prescribed symmetries *are* realized; we call this *symmetry slippage*. Although famous examples of symmetry slippage abound, a systematic study of its origin, estimation, and corrections has yet to exist. This work aims to initiate such a study. We formulate symmetry slippage in terms of subarchitectural decompositions, and use this framing to identify three common mechanisms by which contemporary techniques can unintentionally create it. These creation mechanisms recover the familiar examples, theoretically ground popular heuristics, and identify symmetry slippage across nine active research areas. While we find that guaranteeing an absence of symmetry slippage is hard in general, we prove conditions under which our decomposition framing is exact. We use this result to exactly compute, for the first time, the realized symmetries of a classical family subsuming e.g. convolutional architectures. Finally, we study *grippage encodings* (GEs), simple symmetry-correcting features that restore expressivity. We find that GEs improve empirical performance across our nine case studies.

1 INTRODUCTION

The ubiquity of symmetry in science motivates machine learning architectures $\mathscr{A}$ that leverage obedience to a prescribed transformation law A as inductive bias (Bronstein et al., 2021). Precisely, $\mathscr{A}$ is designed so that the transformation law it realizes, denoted $\mathrm{Sym}\,\mathscr{A}$, satisfies $\mathrm{A} = \mathrm{Sym}\,\mathscr{A}$. Equality requires two conditions: that every prescribed symmetry is realized ($\mathrm{A} \subset \mathrm{Sym}\,\mathscr{A}$), and that every realized symmetry was prescribed ($\mathrm{A} \supset \mathrm{Sym}\,\mathscr{A}$). The first condition, $\mathrm{A} \subset \mathrm{Sym}\,\mathscr{A}$, is known as *equivariance* of $\mathscr{A}$ with respect to A; its enforcement bedrocks the eponymous *equivariant machine learning* literature and Geometric Deep Learning (Bronstein et al., 2021) more broadly. The other condition, $\mathrm{A} \supset \mathrm{Sym}\,\mathscr{A}$, is subtler but relevant, with famous failure cases observed in fields ranging from molecular learning (Dumitrescu et al., 2025; Weaver & Lupo Pasini, 2025) to the theory of graph neural networks (Xu et al., 2019; Morris et al., 2019). Despite its prominence, the principles of creating, computing, and correcting symmetry slippage have yet to be systematically investigated in contemporary geometric ML.

The present work aims to initiate such a study. We develop a calculus for studying end-to-end symmetry slippage from that of local, modular architectural components, and from this derive three general mechanisms, called *group-closure*, *over-quotienting*, and *extension-mistyping*, by which design choices in equivariant ML can produce symmetry slippage. The first mechanism, group-closure, describes how the class of functions used to build $\mathscr{A}$ can impose a specific symmetry slippage which may be called *over-equivariance*. Group-closure examples include "automorphism group-closure", which includes (but is not limited to) a classical finite group setting, and "Zariski group-closure", which occurs e.g. when polynomial features incidentally extend equivariance from orientation-sensitive Lie subgroups such as $\mathrm{SO}^+(1,d)$, $\mathrm{GL}^+(d)$, or $(\mathbb{R}^+)^k$. The second mechanism, over-quotienting, describes architecture operations which can identify inputs more coarsely than prescribed. The symmetry slippage induced by over-quotienting is more general than the over-equivariance induced by group-closure; for instance, over-quotienting in message-passing graph neural networks converts S_n-invariance into *Weisfeiler-Lehman* (WL)-invariance (Weisfeiler & Lehman, 1968; Xu et al., 2019; Morris et al., 2019). The third mechanism, extension-mistyping,

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.

induces a still more general symmetry slippage. It occurs when a transformation law on a subgroup admits multiple extensions to a larger group, and certain architectural design choices incidentally choose the 'wrong' one.

To demonstrate the applicability of these creation mechanisms, we present nine case studies (three per mechanism) of current research areas in which symmetry slippage occurs but has been un- or incompletely-recognized. In some of these case studies, we locate existing heuristics that partially and implicitly address symmetry slippage in certain domains but not others. Because these heuristics are often motivated by domain-specific intuition, we hope that our creation mechanisms' theoretical, domain-agnostic grounding of them will boost their transferability.

The ease with which symmetry slippage can be unintentionally created leads naturally to inquiry into how its absence can be proven. While we show that exactly computing $\mathrm{Sym}\,\mathscr{A}$ is GI-hard in general, even when promised $\mathrm{A} \leq \mathrm{Sym}\,\mathscr{A}$, we prove conditions under which $\mathrm{Sym}\,\mathscr{A}$ can be exactly computed by composing the transformation laws of $\mathscr{A}$'s subarchitectures. To exhibit this structure theorem, we apply it to compute $\mathrm{Sym}\,\mathscr{A}$ exactly for the classical setting of sparse parameter-sharing equivariant MLPs (e.g. convolutional architectures (Cohen & Welling, 2016)), by composing per-layer transformation law computations. Although this classical setting is quite well-studied (Ravanbakhsh et al., 2017; Maron et al., 2018; 2019), this constitutes to our knowledge its first exact computation of $\mathrm{Sym}\,\mathscr{A}$.

To correct symmetry slippage and restore expressivity, we introduce *grippage encodings (GEs)*, a simple type of positional encoding (Vaswani et al., 2017; Dwivedi et al., 2023; Rampášek et al., 2022) tailored to our three creation mechanisms. For each creation mechanism, we derive a recipe for building GEs that repair symmetry slippage. GEs may be incorporated from the outset into a (re)design, especially when methods to avoid the symmetry slippage are bespoke or otherwise nonobvious. They may also be used to tweak the transformation law of an existing architecture when retraining would be expensive or the prescribed law varies with task, akin to how symmetry-breaking techniques have been recently used to modulate foundation model transformation laws (Goel et al., 2026). Across each of our nine case studies, we empirically demonstrate that GEs improve performance.

To summarize, our contributions are:

- (Creation) We develop a calculus for studying symmetry slippage via architectural decompositions. Our framing identifies three symmetry slippage creation mechanisms: *group-closure*, *over-quotienting*, and *extension-mistyping*. These mechanisms expose previously unrecognized or incompletely-characterized symmetry slippage across nine active areas of research.
- (Computation) While we show that exactly computing $\mathrm{Sym}\,\mathscr{A}$ is hard in general, we prove conditions under which $\mathrm{Sym}\,\mathscr{A}$ can exactly be computed from subarchitectures. We apply this to compute $\mathrm{Sym}\,\mathscr{A}$ for a well-studied classical setting subsuming e.g. convolutional architectures.
- (Corrections) We introduce *grippage encodings*: lightweight symmetry-correcting features that restore expressivity. We empirically verify that grippage encodings improve performance on our nine case studies.

2 Background

Transformation Laws. Let X, Y be input and output spaces and Θ a space of parameters. An *architecture* $\mathscr{A} : X \rightsquigarrow Y$ is a map $\Theta \to Y^X$. The *hypothesis class* of $\mathscr{A}$ is the image of this map, denoted $\mathcal{F}_{\mathscr{A}}$. Fundamentally, our approach to understanding the symmetries of $\mathscr{A}$ (or, more precisely, $\mathcal{F}_{\mathscr{A}}$) is based on (i) understanding the symmetries of its subarchitectures, and (ii) understanding in general how "local" subarchitectural symmetries relate to "global", end-to-end symmetries. For example, to understand the symmetries of an $\mathscr{A} = \mathrm{MLP}$, we seek to understand the extent to which the symmetries of its individual layers (subarchitectures) can be "composed" into symmetries of $\mathscr{A}$ itself. An MLP can be viewed as a computational graph of neurons $v \in V$ and weights; the layers L of an MLP correspond to node cuts $C \subset V$ in this graph, splitting the MLP into upstream and downstream computational subgraphs. To precisely reason about architectural pieces of more general $\mathscr{A}$, we take this viewpoint as foundational, identifying $\mathscr{A}$ with its computation graph $\mathscr{A} = \big(V, E, (Z_v)_{v\in V}, (\phi_v)_{v\in V}\big)$, and studying cut-induced subgraphs of the latter. Here, (V, E) is a DAG having input nodes $I \subset V$ and output nodes $O \subset V$, the Z_v are vector

spaces, and $\phi_v = \{\phi_v^\theta : Z_{\{u\to v\}} \to Z_v\}_{\theta\in\Theta}$ is the parameterized family of *node maps* at v, with $Z_B := \prod_{v\in B} Z_v$ for $B \subset V$ and $\{u \to v\} := \{u \in V : (u, v) \in E\}$).

We make precise the notion of "symmetries of $\mathscr{A}$" using *transformation laws*. Fix $J \leq \mathrm{Bij}(X)$, $K \leq \mathrm{Bij}(Y)$, and a group action $\Gamma \overset{\alpha}{\curvearrowright} X$ with $\alpha(\Gamma) \leq J$. A *transformation law* from X to Y is a subgroup $\mathrm{L} \leq J \times K$. A collection $\mathcal{F}$ of functions $f : X \to Y$ *obeys* L if $f\sigma = \tau f$ for every $(\sigma, \tau) \in \mathrm{L}$, $f \in \mathcal{F}$. Given group actions $H \overset{\alpha_X}{\curvearrowright} X$, $H \overset{\alpha_Y}{\curvearrowright} Y$ with $\alpha_X(H) \leq \alpha(\Gamma)$ and $\alpha_Y(H) \leq K$, an H-equivariant architecture $\mathscr{A} : X \rightsquigarrow Y$ is designed to obey a prescribed transformation law $\mathrm{A} \leq J \times K$. For instance, a standard prescription is the law $\mathrm{A} := \{\big(\alpha_X(h), \alpha_Y(h)\big) : h \in H\}$, which we shall denote by $\underline{H}^{\alpha_Y}_{\alpha_X}$. The *realized transformation law of* $\mathscr{A}$ is the maximal transformation law of its hypothesis class $\mathcal{F}_{\mathscr{A}}$,

$$\mathrm{Sym}\,\mathscr{A} := \{(\sigma, \tau) \in J \times K : f\sigma = \tau f \text{ for all } f \in \mathcal{F}_{\mathscr{A}}\}.$$

While strictly speaking $\mathrm{Sym}\,\mathscr{A}$ depends on J, K, we suppress them from the notation because they are usually self-evident (e.g. if Y is a vector space, it is often appropriate to take $K = \mathrm{GL}(Y)$).

Symmetry Slippage. Equivariant learning designs architectures $\mathscr{A} : X \rightsquigarrow Y$ obeying a prescribed transformation law $\mathrm{A} \leq J \times K$. The goal is $\mathrm{A} = \mathrm{Sym}\,\mathscr{A}$. One way equality fails is $\mathrm{Sym}\,\mathscr{A} \not\supset \mathrm{A}$, which can be called *under-equivariance*. Under-equivariance is extensively studied, insofar as the most fundamental motivation for equivariant learning is to avoid it. The other way equality fails is $\mathrm{Sym}\,\mathscr{A} \not\subset \mathrm{A}$, when $\mathscr{A}$ realizes transformations that were not prescribed. We call this *symmetry slippage*, and call elements of $\mathrm{Sym}\,\mathscr{A} - \mathrm{A}$ *slipped transformations*. Symmetry slippage contains *over-equivariance* as the special case when $\mathrm{A} := \underline{H}^{\alpha_Y}_{\alpha|_H}$ is prescribed for some $H \leq \Gamma$, but a group $\Gamma \geq G \gneq H$ exists and α_Y extends to an action of G on Y such that $\underline{H}^{\alpha_Y}_{\alpha|_H} \subsetneq \underline{G}^{\alpha_Y}_{\alpha|_G} \leq \mathrm{Sym}\,\mathscr{A}$. The notion is altogether more general than over-equivariance, however; a detailed contextualization may be found in Appendix A.

Additional Notation. Y^X denotes functions from X to Y. We let $\mathrm{Bij}(X) \subset X^X$ denote the bijections of a set X, and denote group actions $\alpha : G \to \mathrm{Bij}(X)$ by $\alpha : G \curvearrowright X$. Given two G-actions $\alpha_X : G \curvearrowright X$, $\alpha_Y : G \curvearrowright Y$, we let $\mathrm{Map}_G(X, Y)$ denote the maps f from X to Y that are G-equivariant with respect to α_X and α_Y, and, for linear actions, let $\mathrm{Hom}_G(X, Y)$ denote the linear such maps. If α_X, α_Y are clear from context we use the unadorned term 'f is equivariant'; if other actions on X, Y are also being discussed we write 'f is (α_X, α_Y)-equivariant' to clarify. Given a family of group actions $\alpha = \{\alpha_v : G \curvearrowright Z_v\}_{v\in V}$, we write α_B for the product action $g \mapsto \big(\alpha_v(g)\big)_{v\in B}$ on Z_B. The composition of transformation laws $\mathrm{L}_1 \subset J \times K'$, $\mathrm{L}_2 \subset K' \times K$ is $\mathrm{L}_2 \circ \mathrm{L}_1 := \{(\sigma, \tau) : \exists\delta \in K' \text{ s.t. } (\sigma, \delta) \in \mathrm{L}_1 \text{ and } (\delta, \tau) \in \mathrm{L}_2\} \subset J \times K$. The composition of function classes $\mathcal{F} \subset Y^X$, $\mathcal{G} \subset Z^Y$ is $\mathcal{G} \circ \mathcal{F} = \{g \circ f : f \in \mathcal{F}, g \in \mathcal{G}\} \subset Z^X$.

3 Creating Symmetry Slippage

3.1 Decomposing Transformation Laws

We now make precise the local-to-global viewpoint described in Section 2. Say a pair (C_1, C_2) of node cuts of V is *good* if every path from an input $i \in I$ to an output $o \in O$ meets C_1 before C_2. A good pair (C_1, C_2) uniquely determines a computation graph $\mathscr{A}_{(C_1,C_2)} : Z_{C_1} \rightsquigarrow Z_{C_2}$, which we call the *subarchitecture* of $\mathscr{A}$ corresponding to (C_1, C_2). Every good pair (C_1, C_2) induces, for each $\theta \in \Theta$, a factorization $f^\theta : X \longrightarrow Z_{C_1} \longrightarrow Z_{C_2} \longrightarrow Y$, where a^θ, Φ^θ, and e^θ are realized by $\mathscr{A}_{(I,C_1)}$, $\mathscr{A}_{(C_1,C_2)}$, and $\mathscr{A}_{(C_2,O)}$, respectively.

In particular, given a node cut $C \subset V$, the pairs (I, C) and (C, O) are always good, whence C splits $\mathscr{A}$ into subarchitectures $\mathscr{A} : X \overset{\mathscr{A}_{(I,C)}}{\rightsquigarrow} Z_C \overset{\mathscr{A}_{(C,O)}}{\rightsquigarrow} Y$. At the level of hypothesis classes, however, this decomposition is only *lax*: in general, $\mathcal{F}_{\mathscr{A}} \subset \mathcal{F}_{\mathscr{A}_{(C,O)}} \circ \mathcal{F}_{\mathscr{A}_{(I,C)}}$ but equality need not hold.

Proposition 1. *Any subarchitecture $\mathscr{A}_{(C_1,C_2)}$ gives rise to a (lax) decomposition*

$$\mathrm{Sym}\,\mathscr{A}_{(C_2,O)} \circ \mathrm{Sym}\,\mathscr{A}_{(C_1,C_2)} \circ \mathrm{Sym}\,\mathscr{A}_{(I,C_1)} \leq \mathrm{Sym}\,\mathscr{A}. \tag{1}$$

Equality need not hold in general.

Proposition 1 implies that to identify symmetry slippage in an architecture $\mathscr{A}$, it is sufficient to find a subarchitecture $\mathscr{A}_{(C_1,C_2)}$ whose realized law contains an unprescribed transformation preserved

by upstream and downstream computation. At the same time, it implies $\mathrm{Sym}\,\mathscr{A}$ cannot in general be computed from subarchitectures. Our creation mechanisms in Section 3 exploit the first fact. Our computational results in Section 4 investigate conditions under which equality in (1) does hold.

3.2 Local Symmetry Slippage

With the framing of Section 3.1 in mind, we describe three mechanisms by which symmetry slippage may be locally created.

3.2.1 Group-Closure

Group-closure gives local conditions under which over-equivariance (Section 2; App. A) occurs.

Definition 2. *Let $\mathscr{A}_{(C_1,C_2)}$ be a subarchitecture the node maps $\phi_v^\theta : Z_{\{u\to v\}} \to Z_v$ of which belong to a class $\mathcal{C}$ of functions. Let G be a group that acts on the spaces Z_v, and let $H \le G$ act by restriction. We say that $\mathcal{C}$* closes *H to G on $\mathscr{A}_{(C_1,C_2)}$ if for every node v of $\mathscr{A}_{(C_1,C_2)}$, H-equivariance implies G-equivariance for all $\phi \in \mathcal{C} \cap Z_v^{Z_{\{u\to v\}}}$.*

Proposition 3. *Let $H \lneq G \le \Gamma$, let $\alpha_X := \alpha|_G$, and let $G \overset{\alpha_Y}{\curvearrowright} Y$ with $\alpha_Y(G) \le K$, and suppose $\mathrm{A} = \underline{H}^{\alpha_Y}_{\alpha_X} \lneq \underline{G}^{\alpha_Y}_{\alpha_X}$. Let $\mathscr{A}_{(C_1,C_2)}$ be a subarchitecture on which $\mathcal{C}$ closes H to G. Suppose the node maps ϕ_v^θ of $\mathscr{A}_{(C_1,C_2)}$ are H-equivariant, and suppose that a^θ and e^θ are G-equivariant for all θ. Then $\underline{H}^{\alpha_Y}_{\alpha_X} \lneq \underline{G}^{\alpha_Y}_{\alpha_X} \le \mathrm{Sym}\,\mathscr{A}$, and group-closure therefore causes symmetry slippage.*

Example 1 (Automorphism closure). A common strategy is to identify elements of H with automorphisms of some object Σ, then constrain $\mathscr{A}$ to a class $\mathcal{C}$ of $\mathrm{Aut}(\Sigma)$-equivariant maps. This strategy only guarantees $H \le \mathrm{Aut}(\Sigma)$, not equality. Consequently, $\mathcal{C}$ closes H to $\mathrm{Aut}(\Sigma)$, with symmetry slippage occurring when $H \lneq \mathrm{Aut}(\Sigma)$.

Subexample 1. The classical setting of building $\mathscr{A}$ from morphisms of finite-group permutation representations, as is done e.g. in G-CNNs (Cohen & Welling, 2016), GCNs (Bruna et al., 2013; Henaff et al., 2015; Kipf & Welling, 2016), and DeepSets (Ravanbakhsh et al., 2016; Zaheer et al., 2017), is contained in Example 1. Indeed, let I and O be the sets of input and output neurons of a single linear MLP layer $\mathscr{A}_{(I,O)}$. Let $H \curvearrowright I$, $H \curvearrowright O$ be finite actions, inducing permutation representations $H \to J := S_I \le \mathrm{GL}(\mathbb{R}^I)$, $H \to K := S_O \le \mathrm{GL}(\mathbb{R}^O)$. One can check that, in this setting, $\mathrm{Hom}_H(\mathbb{R}^I, \mathbb{R}^O) = \mathcal{C}$, where $\mathcal{C}$ is the space of matrices constant on index orbits. Thus H-equivariance can be implemented by orbitwise parameter-sharing. The orbit construction, however, only yields $\underline{H}^{\alpha_Y}_{\alpha_X} \le \mathrm{Aut}(\Sigma)$ in general, with $\mathcal{C}$ closing H to $\mathrm{Aut}(\Sigma)$, where Σ is a coloring of the edges $E = O \times I$ of $\mathscr{A}_{(I,O)}$ by H-orbit. This perspective was studied by Ravanbakhsh et al. (2017), who introduced the term *uniquely equivariant* for what is in our present notation equality $\underline{H}^{\alpha_Y}_{\alpha_X} = \mathrm{Aut}(\Sigma)$, i.e. the absence of symmetry slippage by over-equivariance.

If the actions on I and O are in fact free, there is a (noncanonical) bijection $\Psi : (O \times I)/H \to H \times O/H \times I/H$ identifying H-orbits in $O \times I$ with triples of the form (a, Ho, Hi) for $a \in H$, $o \in O$, $i \in I$. This makes it possible to sparsify the above orbitwise parameter-sharing construction, by fixing $A \subset H$ and retaining only those H-orbits $H(o,i)$ for which $\Psi_1\big(H(o,i)\big) \in A$ (Ravanbakhsh et al., 2017). For instance, if $H = C_n$ acts by circular translation on $I = O = C_n$, then O/H, I/H are singletons (since the action is transitive) and so H-orbits in $O \times I$ are in bijection with relative translations $a \in C_n$ between input and output indices. Retaining only $a \in A$ then recovers circular convolution with a kernel having support A. An MLP which implements equivariance by sharing parameters according to A may be called an A-MLP; in Section 4, we will exactly compute $\mathrm{Sym}(A\text{- MLP})$.

Example 1 is earnestly more general than the classical finite setting above. For instance, our Case Study 3 (Section 5.2.1) applies Example 1 with $\Sigma = \mathfrak{g}$ a Lie algebra rather than a colored graph.

Example 2 (Zariski closure). Suppose G is the group of real points of a real algebraic group $\mathbf{G}$ and, for every relevant $\phi \in \mathcal{C}$ and z, the map $g \mapsto \phi(gz) - g\phi(z)$ is regular. If H is Zariski-dense in G, $\overline{H}^{\text{Zariski}} = G$, then H-equivariance of ϕ implies G-equivariance. Hence $\mathcal{C}$ closes H to G.

An exposition of the algebro-geometric concepts relevant to Example 2 may be found in Cox et al. (2007). The important takeaway is that Zariski-density $\overline{H}^{\text{Zariski}} = G$ is not pathological; for instance, it occurs when $\mathbf{G}$ is Zariski-connected and $H = G_0$ is the identity component of G. Instances include the identity components $\mathrm{SO}_0(p,q) \subset \mathrm{SO}(p,q)$ for $p, q \ge 1$, $\mathrm{GL}^+(d,\mathbb{R}) \subset \mathrm{GL}(d,\mathbb{R})$, $\mathrm{Aff}^+(d,\mathbb{R}) \subset \mathrm{Aff}(d,\mathbb{R})$, and $(\mathbb{R}^+)^d \subset (\mathbb{R}^\times)^d$. We will be most interested in the cases that $\mathcal{C} = \mathbb{R}[X]$ or $\mathcal{C} = K(X)$ consists of polynomial or rational functions.

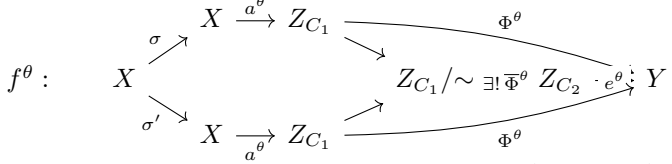
In linear settings, we are able to upgrade from $\mathcal{C} = \mathbb{R}[X]$ to the analytic class $C^\omega(X)$.

Proposition 4. *If the spaces Z_v for v in $\mathscr{A}_{(C_1,C_2)}$ are finite-dimensional G-representations and polynomials $\mathbb{R}[X]$ close H to G on $\mathscr{A}_{(C_1,C_2)}$, then $\mathcal{C} := C^\omega(X)$ closes H to G on $\mathscr{A}_{(C_1,C_2)}$.*

3.2.2 Over-Quotienting

While group-closure investigates the setting where the function class $\mathcal{C}$ from which $\mathscr{A}$ is built induces over-equivariance and hence symmetry slippage, *over-quotienting* investigates the setting where an operation in $\mathscr{A}$ itself can induce symmetry slippage.

Definition 5. *We say that $\mathscr{A}$* over-quotients A *if there exist a subarchitecture $\mathscr{A}_{(C_1,C_2)} : Z_{C_1} \rightsquigarrow Z_{C_2}$, an equivalence relation $\sim$ on Z_{C_1}, a pair $(\sigma, \tau) \in \mathrm{A}$, and $\sigma' \in J$, such that $(\sigma'\sigma^{-1}, I_Y) \notin \mathrm{A}$ and for all $\theta \in \Theta$ there exists a unique $\overline{\Phi}^\theta$ making each f^θ factor as a commutative diagram*



In practice, the over-quotienting condition is verified by checking $a^\theta\sigma(x) \sim a^\theta\sigma'(x)$ for all $x \in X$ and $\theta \in \Theta$, then checking that Φ^θ is constant on $\sim$-classes and so preserves this identification. Finally, one checks $(\sigma'\sigma^{-1}, I_Y) \notin \mathrm{A}$, which verifies this identification was not prescribed by A.

Proposition 6. *If $\mathscr{A}$ over-quotients A, then $f\sigma'\sigma^{-1} = f$ for every $f \in \mathcal{F}_{\mathscr{A}}$. Therefore, $(\sigma'\sigma^{-1}, I_Y) \in \mathrm{Sym}\,\mathscr{A} - \mathrm{A}$ and $\mathscr{A}$ exhibits symmetry slippage.*

If $\Gamma \curvearrowright X$ is a group action and $\mathrm{A} = \underline{H}^I_\alpha$ for a restriction of the Γ-action to $H \lneq \Gamma$, and the relation $\sim = \sim_G$ on Z_{C_1} is the orbit relation of the Γ-action restricted to $G \gneq H$, then factorization through $Z_{C_1}/\!\sim_G$ implies $\mathscr{A}$ is G-equivariant so long as a^θ is. Hence $\underline{H} \subsetneq \underline{G} \leq \mathrm{Sym}\,\mathscr{A}$ and the symmetry slippage is over-equivariance, as for group-closure. Unlike group-closure, not all over-quotienting instances are over-equivariance. For example, over-quotienting in vanilla MPNNs (Gilmer et al., 2017) converts S_n-invariance into constancy on WL-classes (Xu et al., 2019; Morris et al., 2019), even though these additional identifications are not induced by any $S_n \lneq G \leq \Gamma$ (see Appendix A). Our case studies (Section 5.2) reinforce that over-quotienting beyond over-equivariance also occurs into recent published work.

3.2.3 Extension-Mistyping

In both group-closure and over-quotienting, symmetry slippage $\mathrm{Sym}\,\mathscr{A} \not\subset \mathrm{A}$ generally occurs because more transformations are realized than prescribed, $\mathrm{A} \subsetneq \mathrm{Sym}\,\mathscr{A}$. We next introduce a creation mechanism, *extension-mistyping*, for which A and $\mathrm{Sym}\,\mathscr{A}$ may be earnestly incomparable. While group-closure concerns unintended-but-canonical extensions of group actions, extension-mistyping concerns intended-but-noncanonical extensions. While over-quotienting concerns incidental identification of features, extension-mistyping concerns incidental identification of "transformation types", in a sense the following definition makes precise. Let $H \lneq G$ and $\mathscr{A}_{(C_1,C_2)}$ be a subarchitecture with H-action $\boldsymbol{\rho} = (\rho_v : H \curvearrowright Z_v)$ on its spaces. Suppose there are $(Z_v)_{v \in T \subset V}$ that come equipped with G-actions $\eta_v : G \curvearrowright Z_v$ satisfying $\eta_v|_H = \rho_v$. Let $\widetilde{\rho}^{\mathrm{p}}$ be an extension of ρ.

Definition 7. *Say $\mathscr{A}$* extension-mistypes $\widetilde{\rho}^{\mathrm{p}}$ on $\mathscr{A}_{(C_1,C_2)}$ *if there is an extension $\widetilde{\rho}^{\mathrm{r}}$ of ρ such that $\widetilde{\rho}^{\mathrm{r}}_v = \eta_v$ for* all *$v \in T$, $\eta_v \neq \widetilde{\rho}^{\mathrm{p}}_v$ for* some *$v \in T$, and node maps ϕ^θ_v of $\mathscr{A}_{(C_1,C_2)}$ are G-equivariant.*

Proposition 8. *Let $H \lneq G$, and suppose $\mathscr{A}$ extension-mistypes $\widetilde{\rho}^{\mathrm{p}}$ on $\mathscr{A}_{(C_1,C_2)}$ via $\widetilde{\rho}^{\mathrm{r}}$, with actions $G \curvearrowright X, G \curvearrowright Y$ satisfying $\alpha^{\mathrm{r}}_X(G) \leq J$ and $\alpha^{\mathrm{r}}_Y(G) \leq K$, such that a^θ is $(\alpha^{r}_X, \widetilde{\rho}^{r}_{C_1})$-equivariant and e^θ is $(\widetilde{\rho}^{r}_{C_2}, \alpha^{r}_Y)$-equivariant. Then $\underline{G}^{\alpha^{r}_Y}_{\alpha^{r}_X} \leq \mathrm{Sym}\,\mathscr{A}$. Therefore, if $\underline{G}^{\alpha^{r}_Y}_{\alpha^{r}_X} \not\subset \mathrm{A}$, then $\mathrm{Sym}\,\mathscr{A} \not\subset \mathrm{A}$ and $\mathscr{A}$ exhibits symmetry slippage.*

Proposition 9. *Let $H \triangleleft G$ have index 2, and fix $s \in G - H$. Suppose there exists $\boldsymbol{S} = (S_v \in \mathrm{Aut}\, Z_v)$ with certain properties. If $\widetilde{\rho}^{\mathrm{p}}_v(s) \neq S_v$ for some $v \in T$, then $\mathscr{A}$ extension-mistypes $\widetilde{\boldsymbol{\rho}}^{\mathrm{p}}$ on $\mathscr{A}_{(C_1,C_2)}$.*

4 Computing Symmetry Slippage

Identifying symmetry slippage as in Section 3 does not require exactly computing $\mathrm{Sym}\,\mathscr{A}$, just the verification that $\mathrm{Sym}\,\mathscr{A} \not\subset \mathrm{A}$. Nevertheless, exact computation of $\mathrm{Sym}\,\mathscr{A}$ remains of interest in principle, for it is the only way to guarantee absence of symmetry slippage.

Theorem 10. *For every finite group H, computing additional generators B satisfying $\mathrm{Sym}\,\mathscr{A} = \langle \mathrm{A}, B \rangle$ is GI-hard, even when $\mathrm{A} = \underline{H}^{\alpha_Y}_{\alpha_X} \leq \mathrm{Sym}\,\mathscr{A}$ is promised.*

The first step to proving Theorem 10 returns to the setting of Subexample 1 and show that the inclusion described therein is exact. The following lightly strengthens Theorem 2.1 in Ravanbakhsh et al. (2017).

Proposition 11. *With the setting of Subexample 1,* $\mathrm{Sym}(\mathscr{A}) = \mathrm{Aut}(\Sigma)$ *for one linear layer. The result maintains for a nonlinear layer* $x \mapsto \mathrm{a}(Wx)$ *if* $(r \in \mathbb{R}) \mapsto (t \mapsto \mathrm{a}(tr))$ *is injective.*

Hardness makes us expect that deducing whether $\mathrm{A} = \mathrm{Sym}\,\mathscr{A}$ can in general necessitate architecture-specific knowledge. Since general architectures are usually built from modular pieces, it is natural to ask when $\mathrm{Sym}\,\mathscr{A}$ can be exactly computed from subarchitectures $\mathrm{Sym}\,\mathscr{A}_{(C_i,C_j)}$, that is, when a converse to Proposition 1 holds. The following result offers such conditions. Let $\mathscr{A} = \mathscr{A}_{(C_{L-1},O)} \circ \cdots \circ \mathscr{A}_{(I,C_1)}$ be a decomposition of $\mathscr{A}$ for which $\mathcal{F}_{\mathscr{A}} = \mathcal{F}_{\mathscr{A}(C_{L-1},O)} \circ \cdots \circ \mathcal{F}_{\mathscr{A}_{(I,C_1)}}$. Assume $\boldsymbol{\beta} = (\beta_v : G \curvearrowright Z_v)$ is an action on the spaces of $\mathscr{A}$ such that $\mathrm{Sym}\,\mathscr{A} \subset \beta_I(G) \times \beta_O(G)$. For $\mathcal{F} \subset Y^X$, write $\mathcal{F}^\top : X \to Y^{\mathcal{F}}$ for the map $x \mapsto (f(x))_{f\in\mathcal{F}}$.

Theorem 12. *For each* ℓ *suppose (i) there exists* $\mathcal{F}'_\ell \subset \mathcal{F}_{\mathscr{A}_{(C_{\ell-1},C_\ell)}} \cap \mathrm{Map}_G(Z_{C_{\ell-1}}, Z_{C_\ell})$ *such that* $\mathcal{F}'^{\top}_\ell$ *is injective, and (ii) for* $\ell > 1$, $fg \equiv gf$ *for all* $g \in G$ *whenever* $fgp = gfp$ *for all* $f \in \mathcal{F}_{\mathscr{A}_{(C_{\ell-1},C_\ell)}}, p \in \mathcal{F}'_{\ell-1} \circ \cdots \circ \mathcal{F}'_1$. *Then* $\mathrm{Sym}\,\mathscr{A} = \mathrm{Sym}\,\mathscr{A}_{(C_{L-1},O)} \circ \cdots \circ \mathrm{Sym}\,\mathscr{A}_{(I,C_1)}$.

Conditions (i) and (ii) are complementary, roughly corresponding to appropriate notions of injectivity and surjectivity. If the $fg - gf$ were linear, the first condition is equivalent to $\bigcap_{f\in\mathcal{F}'_\ell} \ker f = (0)$ and the second is $\mathrm{span}(\bigcup_{p\in\mathcal{F}'_{\ell-1}\circ\cdots\circ\mathcal{F}'_1} \mathrm{im}\, p) = Z_{C_{\ell-1}}$. Of course, linearity basically never holds, and the stated hypotheses are the generalizations of these natural notions to the nonlinear setting.

To exemplify Theorem 12, we apply it to compute $\mathrm{Sym}(A\text{-MLP})$ (Subexample 1) in closed-form. Before applying Theorem 12, we first compute $\mathrm{Sym}\,\mathscr{A}$ in closed form for $\mathscr{A}$ a single equivariant layer.

Proposition 13. *Let* $\mathscr{A} : \mathbb{R}^I \to \mathbb{R}^O$ *be a linear* A-MLP *layer as in Subexample 1,* $A \neq \emptyset$. *Put* $D_A := \langle AA^{-1}\rangle$, $k := [H : D_A]$. *Then* $\mathrm{Sym}\,\mathscr{A} \cong D_A \wr S_k$. *The conclusion maintains for a nonlinear layer* $x \mapsto \mathrm{a}(Wx)$ *provided* $(r \in \mathbb{R}) \mapsto (t \mapsto \mathrm{a}(tr))$ *is injective.*

Here, $\wr$ denotes the *wreath product* $D_A \wr S_k := D_A^k \rtimes S_k$ where S_k acts on D_A^k by permuting its factors. In the dense special case $A = H$, $D_A = H$ and $k = 1$ and so $\mathrm{Sym}\,\mathscr{A} \cong H$. The injectivity condition on a in Propositions 11 and 13 is not stringent; it is satisfied e.g. by ReLU, sigmoid, $\tanh$, leaky ReLU, ELU, etc.

Proposition 14. *Let* $\mathscr{A} = \mathscr{A}_{(C_{L-1},O)} \circ \cdots \circ \mathscr{A}_{(I,C_1)}$ *be an* A-MLP *as in Subexample 1 with activation* a, *such that* $(r \in \mathbb{R}) \mapsto (t \mapsto \mathrm{a}(tr))$ *is injective. Put* $D_A := \langle AA^{-1}\rangle$, $k := [H : D_A]$. *Then*

$$\mathrm{Sym}\,\mathscr{A} \cong D_A \wr S_k.$$

It is of future interest to derive further closed-form characterizations of $\mathrm{Sym}\,\mathscr{A}$ in other settings.

5 Correcting Symmetry Slippage

5.1 Grippage Encodings

We next discuss a simple architectural modification that restores expressivity loss incurred by symmetry slippage. Let us say that a modification $\mathscr{A}'$ of $\mathscr{A}$ *grips* a slipped transformation $s = (\sigma_s, \tau_s) \in \mathrm{Sym}\,\mathscr{A} - \mathrm{A}$ if $s \notin \mathrm{Sym}\,\mathscr{A}'$, i.e., there exists $f' \in \mathcal{F}_{\mathscr{A}'}$ such that $f'\sigma_s \neq \tau_s f'$. In this section, we study a lightweight intervention, *grippage encodings (GEs)*, for gripping slipped transformations by appending particular features. Throughout, we assume there is a "usual way" to append feature maps to $\mathscr{A}$. To be precise, we assume that for any cut C, there is a rule assigning any $c : X \to Z$ to an architecture $\mathscr{A} \oplus_C c$ whose functions factor as $X \xrightarrow{(b,c)} Z_C \times Z \xrightarrow{e'} Y$ and for which $\mathscr{A} \oplus_C 0 = \mathscr{A}$. This rule is usually self-evident. For example, if C is a layer of $\mathscr{A} = \mathrm{MLP}$, the rule replaces $z \mapsto \phi(Wz)$ with $\big((b(x), c(x))\big) \mapsto \phi(Wb(x) + Uc(x))$ for U trainable.

Definition 15. *A* grippage encoding for $S \subset \mathrm{Sym}\,\mathscr{A} - \mathrm{A}$ at C *is a map* $c : X \to Z$ *for which* $\mathscr{A} \oplus_C c$ *grips every* $s \in S$.

While the general principle of GEs is simple, their implementation can be subtle depending on the creation mechanism for S. The following result streamlines proving GEs in our case studies.

GEs are of course not the only method by which slipped transformations can be gripped. For certain over-quotienting and extension-mistyping examples in particular, there are bespoke repairs which

are just as simple. GEs remain attractive, however, in that their analysis is amenable to general principles. We codify these principles in the proposition below, which greatly streamlines proving GEs in our case studies.

Proposition 16 (Informal). *$c : X \to Z$ is a GE for S if any of the following hold:*

1. *Group-closure created S and $c\alpha_X(g) \neq gc$ for every $s = (\alpha_X(g), \alpha_Y(g)) \in S$.*[1]
2. *Over-quotienting created S and $c\sigma' \neq c\sigma$ for every $s = (\sigma'\sigma^{-1}, I) \in S$.*[2]
3. *Extension-mistyping created S and there exist $\kappa^p, \kappa^r : G \curvearrowright Z$, compatible with A and $\mathscr{A}$ respectively, such that $\operatorname{im} c \not\subset \operatorname{Eq}\big(\kappa^p(g), \kappa^r(g)\big)$ for every $s_g \in S$. If in fact $[G : H] = 2$, $\kappa^r|_H = \kappa^p|_H$, and $\kappa^p(s) = -\kappa^r(s)$ for some $s \in G - H$, it suffices that $c \neq 0$.*

Remark 1.

- GEs can furthermore be used to naturally modify an architecture's realized law with minimal retraining when the aforementioned bespoke methods often cannot; Case Study 5 includes an example of this.
- Abstractly, the purpose of GEs is to modify one obeyed transformation law L_1 into another transformation law L_2. For us, the ambient interpretation is that $\mathrm{L}_1 = \operatorname{Sym}\mathscr{A}$ is a realized transformation law, while $\mathrm{L}_2 = \mathrm{A}$ is a prescribed one. However, in principle L_1 and L_2 can be any pair of laws, and GEs can be used to modulate between them. This may be useful e.g. for adapting the equivariance properties of pretrained models to different tasks or targets Goel et al. (2026). Case Study 5 includes an example of this principle.

5.2 Experiments

We offer 9 case studies in which our creation mechanisms identify, and GEs provably repair, symmetry slippage within an active area. We overview each case study here, deferring details to Appendix C. Table 1 summarizes the effects of GEs across the case studies.

5.2.1 Case Studies for Group-Closure

Case Study 1: Time-Orientation in Lorentz-Equivariant ML. In areas such as experimental high-energy particle physics (HEP), architectures seek equivariance to the (proper orthochronous) Lorentz group $\mathrm{SO}^+(1, d)$ of spacetime-orientation-preserving maps that respect the Minkowski metric on $\mathbb{R}^{1,d}$ (Bogatskiy et al., 2020; Gong et al., 2022; Spinner et al., 2024; Bogatskiy et al., 2024; Batatia et al., 2023). Existing techniques for $\mathrm{SO}^+(1, d)$-equivariant learning use exclusively polynomial features; since $\mathrm{SO}^+(1, d)$ is Zariski-dense in $\mathrm{SO}(1, d)$, this can be shown to entail symmetry slippage under Example 2 and Proposition 3. Conceptually, these architectures $\mathscr{A}$ discard time-orientation information. While this information is not relevant on standard particle physics tasks, it is relevant on any task involving directed relations such as displacements in $\mathbb{R}^{1,d}$. To demonstrate this, we compare four Lorentz-equivariant architectures with and without GEs on a neural PDE task, namely, learning the retarded solution to the wave equation $\Box u = s$ in 2+1 dimensions (Lu et al., 2021; Li et al., 2021; Kovachki et al., 2023).

Case Study 2: Signs in Units-Equivariant ML. Physical laws should not depend on arbitrary choices of measurement units, motivating units-equivariant learning. A common construction, based on the Buckingham-Π theorem (Buckingham, 1914), first maps measured quantities to dimensionless Laurent-monomial features and then applies an unconstrained architecture (Villar et al., 2023). The physical group consists of positive changes of base units, $H = (\mathbb{R}^+)^k$, but H is Zariski-dense in $G = (\mathbb{R}^\times)^k$; consequently, such algebraic dimensionless features are also invariant to nonphysical sign reversals of the base units, giving symmetry slippage under Example 2 and Proposition 3. To demonstrate this, we compare units-equivariant architectures with and without GEs on a forecasting task with the Mauna Loa CO_2 dataset (Keeling et al., 1976; Rasmussen & Williams, 2006), using previous concentration changes to predict future ones.

We further observe that Zariski-density of $(\mathbb{R}^+)^k$ in $(\mathbb{R}^\times)^k$ appears also in weight-space learning for ReLU networks, causing polynomial path representations (Meng et al., 2019; Gonon et al., 2025) to inherit unprescribed sign-invariance by Example 2. This gives a group-closure interpretation of existing heuristics; in particular, the practice of explicitly retaining sign information acts as a GE for the entailed over-equivariance (Meng et al., 2019; Yi et al., 2019; Gonon et al., 2025).

[1] Recall that transformations slipped by group-closure take the form $s = \big(\alpha_X(g), \alpha_Y(g)\big)$ for $g \in G - H$,

[2] Recall transformations slipped by over-quotienting take the form $s = (\sigma'\sigma^{-1}, I_Y)$ for some $\sigma, \sigma' \in J$.

Case Study 3: Inner Automorphisms in Lie Algebra Learning. Lie-algebra-valued data arise in geometric learning problems such as dynamical-system modeling, registration, and homography-based vision (Lin et al., 2024). It is often desirable that architectures learning on the Lie algebra $(\mathfrak{k}, [_, _])$ of a Lie group K be equivariant to the *Adjoint representation* $\mathrm{Ad} : K \to \mathrm{Aut}\,\mathfrak{k}$; conceptually, this encodes a form of coordinate-independence. This turns out equivalent to being $H := \mathrm{Int}(\mathfrak{k})$-equivariant, where $\mathrm{Int}(\mathfrak{k})$ is the *inner automorphism group* of $\mathfrak{k}$. An attractive approach to implementing Ad-equivariance is offered by Lie Neurons (Lin et al., 2024; Kim et al., 2026). However, Lie Neuron architectures turn out to be built from operations that are not only H-equivariant, but G-equivariant, where $G = \mathrm{Aut}(\mathfrak{k})$ is the full automorphism group of $\mathfrak{k}$, yielding symmetry slippage under Example 1 and Proposition 3. We compare the architectures with and without GEs on a standard task in data-driven constitutive modeling (Dammaß et al., 2025; Holthusen et al., 2026).

5.2.2 Case Studies for Over-Quotienting

Case Study 4: Roots in Glycan Learning. Glycans are carbohydrate molecules made of multiple simple sugars linked together by chemical bonds. Many glycans have a *reducing end*, a distinguished sugar that contains a free anomeric carbon. These glycans are commonly abstracted as trees whose nodes are sugars and whose edges are bonds, who are rooted at the reducing end, and whose nodes are colored by sugar type and edges by bond type. Recent glycan-learning architectures such as GlycanML (Xu et al., 2025a) and GlycanAA (Xu et al., 2025b) use graph-based representations of these molecules that over-quotient away the reducing end position, causing symmetry slippage per Proposition 6 (although, unlike the group-closure case studies above, it is not over-equivariance). We compare GlycanML and GlycanAA with and without GEs on the domain prediction benchmark found in both papers.

Case Study 5: Normalization in ECG ML. Amplitude normalization, which may be viewed as (at least) $G := \mathbb{R}^+$-invariance, is a common practice in ECG ML (Ma et al., 2021; Weimann & Conrad, 2024; Zeng et al., 2026; Hu et al., 2026; Li et al., 2025). While deliberate and often desirable, normalization over-quotients certain targets that prescribe a trivial transformation law $\mathrm{A} = \underline{H}^I_\alpha = \underline{(e)}^I_\alpha = \{(I_X, I_Y)\}$, incurring symmetry slippage via Proposition 6. To demonstrate this, we evaluate the models (Weimann & Conrad, 2024; Li et al., 2025) with and without GEs on the well-known PTB-XL benchmark (Wagner et al., 2020; Strodthoff et al., 2021) on which they evaluate. Our first experiment demonstrates the slippage on a multilabel classification task restricted to two known amplitude-sensitive labels (Table 1). Our second experiment appends GEs to the papers' released pretrained architectures for the full multilabel task that includes targets for which normalization is desirable and for which it is not; GEs modestly but replicably improve performance here (see Appendix C). In this way, this case study draws upon our secondary GE perspective (Remark 1), that appending GEs to a pretrained foundation model allows the model's transformation law to be adjusted based on the task.

Case Study 6: Polarity in Time-Frequency Signal Processing. The Choi-Williams distribution is a classical quadratic time-frequency representation widely applied to nonstationary signal analysis (Choi & Williams, 1989). Being quadratic, $\mathrm{CWD}(x)$ is invariant under polarity reversal, over-quotienting the prescribed $\mathrm{Aff}^+(1, \mathbb{R})$-invariance of an established framework for CWD-based ECG analysis (Kayikcioglu et al., 2020) to $\mathrm{Aff}(1, \mathbb{R})$-invariance. This becomes problematic e.g. when the task is sensitive to ST elevation and ST depression, which are clinically distinct phenomena (Wagner et al., 2009). This is the case for the task in Kayikcioglu et al. (2020). We evaluate the architecture with and without GEs on the European ST-T database (Taddei et al., 1992) task of distinguishing ST-elevation and ST-depression episodes.

5.2.3 Case Studies for Extension-Mistyping

Case Study 7: (Pseudo)scalar mistyping in neural operators. $H := C_4$ acts on functions $\omega : \mathbb{R}^2 \to \mathbb{R}$ as $g\omega(x) := \omega(g^{-1}x)$. This action extends in two natural ways to $G := D_4$. One is via $g\omega(x) := \omega(g^{-1}x)$, in which case ω is a *scalar field* such as pressure; another is via $g\omega(x) := \det(g)\omega(g^{-1}x)$, in which case ω is a *pseudoscalar field*, such as vorticity (curl). Popular G-equivariant neural operator approaches such as Helwig et al. (2023), however, mistype the extension so that e.g. vorticity is treated as an ordinary scalar field, creating symmetry slippage under Proposition 8. To demonstrate, we compare the the architecture of Helwig et al. (2023) with and without GEs on two standard Navier-Stokes benchmarks, including the benchmark (NS-SYM) evaluated in Helwig et al. (2023).

Table 1: Effect of GEs across 9 case studies. For each metric M, we define $\Delta M := M_{\mathrm{GE}} - M_{\mathrm{baseline}}$. The two +GE rows give the range of effects across the evaluated architectures.

Case Study	1	2	3	4	5	6	7	8	9
Δ Metric	NMSE ↓	Acc. (pp) ↑	$\mathcal{L}$ ↓	Macro-F1 ↑	AUROC ↑	AUROC ↑	Rel. MSE ↓	RMSE ↓	RMSE ↓
+GE (min)	-0.571	$+23.6$	-0.268	$+0.036$	$+0.015$	$+0.117$	-0.662	-0.145	-0.041
+GE (max)	-0.458	$+23.6$	-0.242	$+0.046$	$+0.066$	$+0.127$	-0.662	-0.145	-0.039

Case Study 8: Tetrahedral Chirality in Molecular Graph Learning. A common source of molecular chirality is a tetrahedral atom bonded to four distinct ligands, whose spatial arrangement can occur in two opposite configurations conventionally labeled R and S. Because the ordering of the four ligands is arbitrary, S_4 acts by permuting them: even permutations A_4 preserve handedness, whereas odd permutations reverse it, so a stereochemical encoding defined relative to this ordering should transform according to the sign representation (Pattanaik et al., 2020). However, QGeoGNN, a popular GNN approach for chromatographic retention-time prediction (Xu et al., 2023), extends the representation on A_4 to the trivial representation on S_4, creating symmetry slippage by extension mistyping per Proposition 8. We compare QGeoGNN with and without GEs on its all-column retention-time benchmark (Xu et al., 2023).

Case Study 9: Pseudovector Mistyping in Magnetic Equivariant ML. Magnetic moments and spins are pseudovectors, transforming under the action of $G := \mathrm{O}(3)$ as $\widetilde{\rho}^{\mathrm{p}}(Q)w := \det(Q)Qw$. The distinction does not matter when only $\mathrm{SO}(3) =: H \triangleleft G$ is meaningfully acting, but becomes relevant to equivariant learning for magnetic materials, where spin and magnetic-moment pseudovectors are often represented alongside ordinary vectors such as atomic positions and displacements (Yuan et al., 2024; Yu et al., 2024; Cho et al., 2026). However, such architectures often incidentally extend the $\mathrm{SO}(3)$-action to the $\mathrm{O}(3)$-action for vectors when the $\mathrm{O}(3)$-action for pseudovectors should be prescribed, causing symmetry slippage via extension-mistyping per Proposition 9. To demonstrate, we compare two architectures (Yuan et al., 2024; Cho et al., 2026) with and without GEs on a noncollinear anomalous-Hall prediction task (Kipp et al., 2024).

6 Discussion

This paper studied *symmetry slippage* in contemporary equivariant ML, when an architecture $\mathscr{A}$ realizes symmetries that were not prescribed. We identified 3 creation mechanisms, *group-closure*, *over-quotienting*, and *extension-mistyping*, and showed how they manifest across 9 active areas. We showed that exactly computing $\mathrm{Sym}\,\mathscr{A}$ is hard in general, but also that, under certain conditions, $\mathrm{Sym}\,\mathscr{A}$ can be computed exactly from subarchitectures. We applied this result to exactly compute $\mathrm{Sym}(A\text{-MLP})$ for the first time, where the class A-MLP subsumes e.g. G-CNNs (Ravanbakhsh et al., 2017; Cohen & Welling, 2016).

To correct symmetry slippage, we introduced grippage encodings, and demonstrated them as a repair across our 9 case studies. A limitation of GEs is that they do not come with expressivity guarantees beyond promising that a targeted subset $S \subset \mathrm{Sym}\,\mathscr{A} - \mathrm{A}$ has been removed, $S \cap \mathrm{Sym}\,\mathscr{A}' = \emptyset$. In particular, they do not guarantee $\mathrm{Sym}\,\mathscr{A}' = \mathrm{A}$, perhaps unsurprisingly in light of Theorem 10. Furthermore, while the all-parameter nature of symmetry slippage makes existence a strong statement about an architecture, the GE condition is inversely weak, requiring only that *some* functions $f^\theta \in \mathcal{F}_{\mathscr{A}'}$ disobey S. Thus not all GEs are equivalent, and their efficacy can depend on how C and $e^{\theta'}$ are chosen. Nevertheless, we find GEs remain natural, non-bespoke, lightweight, and effective in restoring performance lost by symmetry slippage in many cases. This includes cases where they modify the transformation law of a pretrained model for which retraining is costly or otherwise undesirable, as when symmetry requirements are task-dependent (Goel et al., 2026).

Scientific machine learning is entering a differentiation phase (Polikarov, 1995) in which increasingly specalized architectures, data modalities, and scientific tools are being assembled into increasingly modular workflows (Smidt, 2026). In this era, we anticipate symmetry slippage to become both more salient and more subtle. While we expect that general principles for proving symmetry slippage's absence are rarer than general principles proving its creation, it is possible to envisage a broad program in exactly computing $\mathrm{Sym}\,\mathscr{A}$ for diverse architecture classes $\mathscr{A}$, allowing for provably exact equivariance to be deployed in complicated scientific workflows.

AI USE STATEMENT

We used generative AI to generate synthetic data sets, implement methods, and clean and reformat datasets. We used generative AI to assist in proofwriting by way of proofreading and polishing. We largely did not use generative AI to provide critical ingredients for proving mathematical claims; an exception is Theorem 10, for which our initial approach was flawed and generative AI provided some useful, albeit noisy, insights that ultimately helped repair it. We additionally used generative AI to assist in literature search. All LLM-generated code was checked by the authors, and no paper (LLM-retrieved or not) was referenced without the authors' understanding of its content. We did not use generative AI to help develop theoretical models or conceptual frameworks, formulate mathematical claims, to assist with translation, to propose or refine hypotheses, to design or provide feedback on research methodology or experiments, support qualitative and thematic data analysis, or interpret results. We take responsibility for the final content of this work.

REPRODUCIBILITY STATEMENT

The proofs of all theoretical results may be found in Appendix B. Full case study discussions may be found in Appendix C. Experimental details may be found in Appendix D.

REFERENCES

Rami Alazrai, Hisham Alwanni, Yara Baslan, Nasim Alnuman, and Mohammad I. Daoud. EEG-based brain-computer interface for decoding motor imagery tasks within the same hand using Choi–Williams time-frequency distribution. *Sensors*, 17(9):1937, August 2017. doi: 10.3390/s17091937.

François Auger, Patrick Flandrin, Paulo Gonçalvès, and Olivier Lemoine. *Time-Frequency Toolbox for Use with MATLAB*. CNRS and Rice University, 1998.

Joseph Bakarji, Jared Callaham, Steven L. Brunton, and J. Nathan Kutz. Dimensionally consistent learning with buckingham pi. *Nature Computational Science*, 2(12):834–844, 2022. doi: 10.1038/s43588-022-00355-5.

Ilyes Batatia, David Peter Kovacs, Gregor N. C. Simm, Christoph Ortner, and Gábor Csányi. MACE: Higher order equivariant message passing neural networks for fast and accurate force fields. In *NeurIPS*, 2022.

Ilyes Batatia, Mario Geiger, Jose Munoz, Tess Smidt, Lior Silberman, and Christoph Ortner. A general framework for equivariant neural networks on reductive lie groups. In *Advances in Neural Information Processing Systems*, volume 36, pp. 55260–55284, 2023. doi: 10.52202/075280-2412.

Ben Blum-Smith and Soledad Villar. Machine learning and invariant theory. *Notices of the American Mathematical Society*, 70(8):1205–1213, 2023. doi: 10.1090/noti2760.

Alexander Bogatskiy, Brandon Anderson, Jan T. Offermann, Marwah Roussi, David W. Miller, and Risi Kondor. Lorentz group equivariant neural network for particle physics. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 992–1002. PMLR, 2020.

Alexander Bogatskiy, Timothy Hoffman, David W. Miller, Jan T. Offermann, and Xiaoyang Liu. Explainable equivariant neural networks for particle physics: PELICAN. *Journal of High Energy Physics*, 2024(3):113, 2024. doi: 10.1007/JHEP03(2024)113.

Michael M. Bronstein, Joan Bruna, Taco Cohen, and Petar Veličković. Geometric deep learning: Grids, groups, graphs, geodesics, and gauges. *arXiv:2104.13478*, 2021.

Joan Bruna, Wojciech Zaremba, Arthur Szlam, and Yann LeCun. Spectral networks and locally connected networks on graphs. *arXiv preprint arXiv:1312.6203*, 2013.

Edgar Buckingham. On physically similar systems; illustrations of the use of dimensional equations. *Physical Review*, 4(4):345–376, 1914. doi: 10.1103/PhysRev.4.345.

R. S. Cahn, Christopher Ingold, and V. Prelog. Specification of molecular chirality. *Angewandte Chemie International Edition in English*, 5(4):385–415, 1966. doi: 10.1002/anie.196603851.

Shuhao Cao, Francesco Brarda, Ruipeng Li, and Yuanzhe Xi. Spectral-refiner: Accurate fine-tuning of spatiotemporal fourier neural operator for turbulent flows. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=MKP1g8wU0P.

Jiacheng Cen, Anyi Li, Ning Lin, Yuxiang Ren, Zihe Wang, and Wenbing Huang. Are high-degree representations really unnecessary in equivariant graph neural networks? *Advances in Neural Information Processing Systems*, 37:26238–26266, 2024.

Youngwoo Cho, Seunghoon Yi, Wooil Yang, Sungmo Kang, Young-Woo Son, Jaegul Choo, Joonseok Lee, Soo Kyung Kim, and Hongkee Yoon. Robust and interpretable adaptation of equivariant materials foundation models via sparsity-promoting fine-tuning. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=moBqB1CUym.

H-I Choi and William J Williams. Improved time-frequency representation of multicomponent signals using exponential kernels. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 37(6):862–871, 1989.

Taco Cohen and Max Welling. Group equivariant convolutional networks. In Maria Florina Balcan and Kilian Q. Weinberger (eds.), *Proceedings of The 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pp. 2990–2999, New York, New York, USA, 20–22 Jun 2016. PMLR. URL https://proceedings.mlr.press/v48/cohenc16.html.

Paul G. Constantine, Zachary del Rosario, and Gianluca Iaccarino. Data-driven dimensional analysis: Algorithms for unique and relevant dimensionless groups, 2017.

David A Cox, John B Little, Donal O'shea, and John Little. *Ideals, varieties, and algorithms: an introduction to computational algebraic geometry and commutative algebra*, volume 10. Springer, 2007.

Franz Dammaß, Karl A. Kalina, and Markus Kästner. When invariants matter: The role of I_1 and I_2 in neural network models of incompressible hyperelasticity. *Mechanics of Materials*, 210:105443, 2025. doi: 10.1016/j.mechmat.2025.105443.

Congyue Deng, Or Litany, Yueqi Duan, Adrien Poulenard, Andrea Tagliasacchi, and Leonidas J. Guibas. Vector neurons: A general framework for SO(3)-equivariant networks. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 12200–12209, 2021.

Alexandru Dumitrescu, Dani Korpela, Markus Heinonen, Yogesh Verma, Valerii Iakovlev, Vikas Garg, and Harri Lähdesmäki. E(3)-equivariant models cannot learn chirality: Field-based molecular generation. In *The Thirteenth International Conference on Learning Representations*, 2025.

Vijay Prakash Dwivedi, Chaitanya K Joshi, Anh Tuan Luu, Thomas Laurent, Yoshua Bengio, and Xavier Bresson. Benchmarking graph neural networks. *Journal of Machine Learning Research*, 24(43):1–48, 2023.

Nikolaos Evangelou, Noah J. Wichrowski, George A. Kevrekidis, Felix Dietrich, Mahdi Kooshkbaghi, Sarah McFann, and Ioannis G. Kevrekidis. On the parameter combinations that matter and on those that do not, 2021.

Marc Finzi, Max Welling, and Andrew Gordon Wilson. A practical method for constructing equivariant multilayer perceptrons for arbitrary matrix groups. In *International conference on machine learning*, pp. 3318–3328. PMLR, 2021.

Federico Frisone and Andrea Misiti. Buckingham theorem application to machine learning algorithms: Methodology and practical examples. Master's thesis, Politecnico di Milano, 2019.

Mario Geiger and Tess Smidt. e3nn: Euclidean neural networks. *arXiv preprint arXiv:2207.09453*, 2022. doi: 10.48550/arXiv.2207.09453. URL https://arxiv.org/abs/2207.09453.

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.

Abhinav Goel, Derek Lim, Hannah Lawrence, Stefanie Jegelka, and Ningyuan Huang. Any-subgroup equivariant networks via symmetry breaking. *arXiv preprint arXiv:2603.19486*, 2026.

Shiqi Gong, Qi Meng, Jue Zhang, Huilin Qu, Congqiao Li, Sitian Qian, Weitao Du, Zhi-Ming Ma, and Tie-Yan Liu. An efficient lorentz equivariant graph neural network for jet tagging. *Journal of High Energy Physics*, 2022(7):30, 2022. doi: 10.1007/JHEP07(2022)030.

Antoine Gonon, Nicolas Brisebarre, Elisa Riccietti, and Rémi Gribonval. A rescaling-invariant lipschitz bound based on path-metrics for modern ReLU network parameterizations. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 20047–20074. PMLR, 2025.

William L. Hamilton, Rex Ying, and Jure Leskovec. Inductive representation learning on large graphs, 2018. URL https://arxiv.org/abs/1706.02216.

Jiaqi Han, Wenbing Huang, Hengbo Ma, Jiachen Li, Joshua B. Tenenbaum, and Chuang Gan. Learning physical dynamics with subequivariant graph neural networks. In *Advances in Neural Information Processing Systems*, volume 35, 2022.

Jacob Helwig, Xuan Zhang, Cong Fu, Jerry Kurtin, Stephan Wojtowytsch, and Shuiwang Ji. Group equivariant fourier neural operators for partial differential equations. *arXiv preprint arXiv:2306.05697*, 2023.

Mikael Henaff, Joan Bruna, and Yann LeCun. Deep convolutional networks on graph-structured data. *arXiv preprint arXiv:1506.05163*, 2015.

Nigel Hitchin. Lie groups, 2015. C3.5 Lie Groups lecture notes, University of Oxford.

Hagen Holthusen, Moritz Flaschel, Denisa Martonová, and Ellen Kuhl. Benchmarking data-driven material models on the classic Treloar dataset. *arXiv preprint arXiv:2608.14063*, 2026.

Haoyang Hu, Zekai Yu, Feiwei Qin, and Fei Yang. From one to twelve: Feasibility and clinical utility of deep learning-derived 12-lead ecgs for remote cardiac monitoring. *Frontiers in Cardiovascular Medicine*, 13: 1856637, 2026. doi: 10.3389/fcvm.2026.1856637.

Sékou-Oumar Kaba and Siamak Ravanbakhsh. Symmetry breaking and equivariant neural networks. In *NeurIPS 2023 Workshop on Symmetry and Geometry in Neural Representations*, 2024. URL https://openreview.net/forum?id=d55JaRL9wh.

İlknur Kayıkçıoğlu, Güzin Ulutaş, and Temel Kayıkçıoğlu. Time-frequency analysis based detection of ECG ST segment change using large feature set. In *2017 25th Signal Processing and Communications Applications Conference (SIU)*, pp. 1–4, Antalya, Turkey, May 2017. doi: 10.1109/SIU.2017.7960408.

Ilknur Kayikcioglu, Fulya Akdeniz, Cemal Köse, and Temel Kayikcioglu. Time-frequency approach to ecg classification of myocardial infarction. *Computers & Electrical Engineering*, 84:106621, 2020.

Charles D. Keeling and Timothy P. Whorf. Atmospheric co2 concentrations derived from flask air samples at sites in the sio network. In *Trends: A Compendium of Data on Global Change*. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tennessee, USA, 2004.

Charles D. Keeling, Robert B. Bacastow, Arnold E. Bainbridge, Carl A. Ekdahl, Peter R. Guenther, Lee S. Waterman, and John F. S. Chin. Atmospheric carbon dioxide variations at mauna loa observatory, hawaii. *Tellus*, 28(6):538–551, 1976. doi: 10.1111/j.2153-3490.1976.tb00701.x.

Chankyo Kim, Sicheng Zhao, Minghan Zhu, Tzu-Yuan Lin, and Maani Ghaffari. Equivariant neural networks for general linear symmetries on lie algebras. In *Proceedings of the 43rd International Conference on Machine Learning*, 2026. URL https://reductive-lie-neuron.github.io/. Reductive Lie Neurons (ReLNs).

Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.

Jonathan Kipp, Fabian R. Lux, Thorben Pürling, Abigail Morrison, Stefan Blügel, Daniele Pinna, and Yuriy Mokrousov. Machine learning inspired models for hall effects in non-collinear magnets. *Machine Learning: Science and Technology*, 5(2):025060, 2024. doi: 10.1088/2632-2153/ad51ca.

Nikola Kovachki, Zongyi Li, Burigede Liu, Kamyar Azizzadenesheli, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Neural operator: Learning maps between function spaces with applications to PDEs. *Journal of Machine Learning Research*, 24(89):1–97, 2023. URL https://jmlr.org/papers/v24/21-1524.html.

Greg Landrum. RDKit: Open-source cheminformatics. https://www.rdkit.org, 2026. Software.

Hannah Lawrence, Vasco Portilheiro, Yan Zhang, and Sékou-Oumar Kaba. Improving equivariant networks with probabilistic symmetry breaking. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=1T5Mcc8n4y.

Sun Ung Lee, David Robb, and Colin Besant. The directional Choi–Williams distribution for the analysis of rotor-vibration signals. *Mechanical Systems and Signal Processing*, 15(4):789–811, 2001. doi: 10.1006/mssp.2000.1359.

Jun Li, Aaron D. Aguirre, Valdery Moura Junior, Jiarui Jin, Che Liu, Lanhai Zhong, Chenxi Sun, Gari D. Clifford, M. Brandon Westover, and Shenda Hong. An electrocardiogram foundation model built on over 10 million recordings. *NEJM AI*, 2(7):AIoa2401033, 2025. doi: 10.1056/AIoa2401033.

Zongyi Li, Nikola B. Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew M. Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=c8P9NQVtmnO.

Ning Lin, Jiacheng Cen, Anyi Li, Wenbing Huang, and Hao Sun. Reducing symmetry increase in equivariant neural networks. In *International Conference on Learning Representations*, volume 2026, pp. 44119–44165, 2026.

Tzu-Yuan Lin, Minghan Zhu, and Maani Ghaffari. Lie Neurons: Adjoint-equivariant neural networks for semisimple Lie algebras. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 30529–30545. PMLR, 2024. URL https://proceedings.mlr.press/v235/lin24aa.html.

Lennart Linden, Dominik K. Klein, Karl A. Kalina, Jörg Brummund, Oliver Weeger, and Markus Kästner. Neural networks meet hyperelasticity: A guide to enforcing physics. *Journal of the Mechanics and Physics of Solids*, 179:105363, 2023. doi: 10.1016/j.jmps.2023.105363.

Kevin Linka, Markus Hillgärtner, Kian P. Abdolazizi, Roland C. Aydin, Mikhail Itskov, and Christian J. Cyron. Constitutive artificial neural networks: A fast and general approach to predictive data-driven constitutive modeling by deep learning. *Journal of Computational Physics*, 429:110010, 2021. doi: 10.1016/j.jcp.2020.110010.

Lu Lu, Pengzhan Jin, Guofei Pang, Zhongqiang Zhang, and George Em Karniadakis. Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators. *Nature Machine Intelligence*, 3(3):218–229, 2021. doi: 10.1038/s42256-021-00302-5.

Hao Ma, Chao Chen, Qing Zhu, Haitao Yuan, Liming Chen, and Minglei Shu. An ecg signal classification method based on dilated causal convolution. *Computational and Mathematical Methods in Medicine*, 2021: 6627939, 2021. doi: 10.1155/2021/6627939.

Haggai Maron. *Deep and Convex Shape Analysis*. PhD thesis, Weizmann Institute of Science, August 2019.

Haggai Maron, Heli Ben-Hamu, Nadav Shamir, and Yaron Lipman. Invariant and equivariant graph networks. *arXiv preprint arXiv:1812.09902*, 2018.

Haggai Maron, Ethan Fetaya, Nimrod Segol, and Yaron Lipman. On the universality of invariant networks. In *International conference on machine learning*, pp. 4363–4371. PMLR, 2019.

James C. McWilliams. The emergence of isolated coherent vortices in turbulent flow. *Journal of Fluid Mechanics*, 146:21–43, 1984. doi: 10.1017/S0022112084001750.

Qi Meng, Shuxin Zheng, Huishuai Zhang, Wei Chen, Qiwei Ye, Zhi-Ming Ma, Nenghai Yu, and Tie-Yan Liu. G-SGD: Optimizing ReLU neural networks in its positively scale-invariant space. In *International Conference on Learning Representations*, 2019.

Christopher Morris, Martin Ritzert, Matthias Fey, William L Hamilton, Jan Eric Lenssen, Gaurav Rattan, and Martin Grohe. Weisfeiler and leman go neural: Higher-order graph neural networks. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pp. 4602–4609, 2019.

Behnam Neyshabur, Ruslan Salakhutdinov, and Nathan Srebro. Path-sgd: Path-normalized optimization in deep neural networks. In *Advances in Neural Information Processing Systems*, volume 28, pp. 2422–2430, 2015.

Lagnajit Pattanaik, Octavian-Eugen Ganea, Ian Coley, Klavs F Jensen, William H Green, and Connor W Coley. Message passing networks for molecules with tetrahedral chirality. *arXiv preprint arXiv:2012.00094*, 2020.

Asarja Polikarov. Concerning the integration of sciences: Kinds and stages. *Journal for general philosophy of science*, 26(2):297–312, 1995.

Ladislav Rampášek, Michael Galkin, Vijay Prakash Dwivedi, Anh Tuan Luu, Guy Wolf, and Dominique Beaini. Recipe for a general, powerful, scalable graph transformer. *Advances in neural information processing systems*, 35:14501–14515, 2022.

Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning*. Adaptive Computation and Machine Learning. MIT Press, Cambridge, MA, 2006. ISBN 978-0-262-18253-9.

Siamak Ravanbakhsh, Jeff Schneider, and Barnabas Poczos. Deep learning with sets and point clouds. *arXiv preprint arXiv:1611.04500*, 2016.

Siamak Ravanbakhsh, Jeff Schneider, and Barnabás Póczos. Equivariance through parameter-sharing. In Doina Precup and Yee Whye Teh (eds.), *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pp. 2892–2901. PMLR, 06–11 Aug 2017. URL https://proceedings.mlr.press/v70/ravanbakhsh17a.html.

Stephan Rudolph. On the context of dimensional analysis in artificial intelligence. In *Proceedings of the International Workshop on Similarity Methods*, pp. 147–161, Stuttgart, Germany, 1998.

Haitz Sáez de Ocáriz Borde, Anastasis Kratsios, Marc T Law, Xiaowen Dong, and Michael Bronstein. Neural spacetimes for dag representation learning. In *International Conference on Learning Representations*, volume 2025, pp. 56302–56341, 2025.

Franco Scarselli, Marco Gori, Ah Chung Tsoi, Markus Hagenbuchner, and Gabriele Monfardini. The graph neural network model. *IEEE Transactions on Neural Networks*, 20(1):61–80, 2009. doi: 10.1109/TNN.2008.2005605.

Michael Schlichtkrull, Thomas N. Kipf, Peter Bloem, Rianne van den Berg, Ivan Titov, and Max Welling. Modeling relational data with graph convolutional networks. In *The Semantic Web – 15th International Conference, ESWC 2018*, volume 10843 of *Lecture Notes in Computer Science*, pp. 593–607. Springer, 2018. doi: 10.1007/978-3-319-93417-4_38.

Ervin Sejdić, Igor Djurović, and Jin Jiang. Time–frequency feature representation using energy concentration: An overview of recent advances. *Digital Signal Processing*, 19(1):153–183, January 2009. doi: 10.1016/j.dsp.2007.12.004.

Ákos Seress. *Permutation Group Algorithms*, volume 152 of *Cambridge Tracts in Mathematics*. Cambridge University Press, 2003. doi: 10.1017/CBO9780511546549.

Tess Smidt. Physics is different: Context, culture & craft in effective aifor physics. 2026.

Tess E. Smidt, Mario Geiger, and Benjamin Kurt Miller. Finding symmetry breaking order parameters with euclidean neural networks. *Physical Review Research*, 3(1):L012002, 2021. doi: 10.1103/PhysRevResearch.3.L012002. URL https://doi.org/10.1103/PhysRevResearch.3.L012002.

Jonas Spinner, Victor Bresó, Pim de Haan, Tilman Plehn, Jesse Thaler, and Johann Brehmer. Lorentz-equivariant geometric algebra transformers for high-energy physics. In *Advances in Neural Information Processing Systems*, volume 37, pp. 22178–22205, 2024. doi: 10.52202/079017-0699.

Nils Strodthoff, Patrick Wagner, Tobias Schaeffter, and Wojciech Samek. Deep learning for ecg analysis: Benchmarks and insights from ptb-xl. *IEEE Journal of Biomedical and Health Informatics*, 25(5):1519–1528, 2021. doi: 10.1109/JBHI.2020.3022989.

Alessandro Taddei, Giovanni Distante, M Emdin, P Pisani, GB Moody, C Zeelenberg, and C Marchesi. The european st-t database: standard for evaluating systems for the analysis of st-t changes in ambulatory electrocardiography. *European heart journal*, 13(9):1164–1172, 1992.

L. R. G. Treloar. Stress-strain data for vulcanised rubber under various types of deformation. *Transactions of the Faraday Society*, 40:59–70, 1944. doi: 10.1039/TF9444000059.

Shikhar Vashishth, Soumya Sanyal, Vikram Nitin, and Partha Talukdar. Composition-based multi-relational graph convolutional networks. In *International Conference on Learning Representations*, 2020. URL https://openreview.net/forum?id=BylA_C4tPr.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

Soledad Villar, David W. Hogg, Kate Storey-Fisher, Weichi Yao, and Ben Blum-Smith. Scalars are universal: Equivariant machine learning, structured like classical physics. In *Advances in Neural Information Processing Systems*, volume 34, pp. 28848–28863, 2021.

Soledad Villar, Weichi Yao, David W Hogg, Ben Blum-Smith, and Bianca Dumitrascu. Dimensionless machine learning: Imposing exact units equivariance. *Journal of Machine Learning Research*, 24(109):1–32, 2023.

Galen S Wagner, Peter Macfarlane, Hein Wellens, Mark Josephson, Anton Gorgels, David M Mirvis, Olle Pahlm, Borys Surawicz, Paul Kligfield, Rory Childers, et al. Aha/accf/hrs recommendations for the standardization and interpretation of the electrocardiogram: part vi: acute ischemia/infarction: a scientific statement from the american heart association electrocardiography and arrhythmias committee, council on clinical cardiology; the american college of cardiology foundation; and the heart rhythm society: endorsed by the international society for computerized electrocardiology. *Circulation*, 119(10):e262–e270, 2009.

Patrick Wagner, Nils Strodthoff, Ralf-Dieter Bousseljot, Dieter Kreiseler, Fatima I. Lunze, Wojciech Samek, and Tobias Schaeffter. Ptb-xl, a large publicly available electrocardiography dataset. *Scientific Data*, 7(1):154, 2020. doi: 10.1038/s41597-020-0495-6.

Dian Wang, Xupeng Zhu, Jung Yeon Park, Mingxi Jia, Guanang Su, Robert Platt, and Robin Walters. A general theory of correct, incorrect, and extrinsic equivariance. *Advances in Neural Information Processing Systems*, 36:40006–40029, 2023.

Rylie Weaver and Massamiliano Lupo Pasini. Chi-geometry: A library for benchmarking chirality prediction of gnns. *arXiv preprint arXiv:2508.09097*, 2025. doi: 10.48550/arXiv.2508.09097.

Kuba Weimann and Tim O. F. Conrad. Federated learning with deep neural networks: A privacy-preserving approach to enhanced ecg classification. *IEEE Journal of Biomedical and Health Informatics*, 28(11):6931–6943, 2024. doi: 10.1109/JBHI.2024.3427787.

Boris Weisfeiler and Andrei A. Lehman. A reduction of a graph to a canonical form and an algebra arising during this reduction. *Nauchno-Tekhnicheskaya Informatsiya*, (9):12–16, 1968.

YuQing Xie and Tess Smidt. Equivariant symmetry breaking sets. *Transactions on Machine Learning Research*, 2024. URL https://openreview.net/forum?id=tHKH4DNSR5.

Hao Xu, Jinglong Lin, Dongxiao Zhang, and Fanyang Mo. Retention time prediction for chromatographic enantioseparation by quantile geometry-enhanced graph neural network. *Nature Communications*, 14(1): 3095, 2023.

Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How Powerful are Graph Neural Networks? In *International Conference on Learning Representations*, 2019. URL https://openreview.net/forum?id=ryGs6iA5Km.

Minghao Xu, Yunteng Geng, Yihang Zhang, Ling Yang, Jian Tang, and Wentao Zhang. Glycanml: A multi-task and multi-structure benchmark for glycan machine learning. In *International Conference on Learning Representations*, 2025a. URL https://openreview.net/forum?id=owEQ0FTfVj.

Minghao Xu, Jiaze Song, Keming Wu, Xiangxin Zhou, Bin Cui, and Wentao Zhang. Modeling all-atom glycan structures via hierarchical message passing and multi-scale pre-training. In Aarti Singh, Maryam Fazel, Daniel Hsu, Simon Lacoste-Julien, Felix Berkenkamp, Tegan Maharaj, Kiri Wagstaff, and Jerry Zhu (eds.), *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 69675–69688. PMLR, 13–19 Jul 2025b. URL https://proceedings.mlr.press/v267/xu25z.html.

Mingyang Yi, Qi Meng, Wei Chen, Zhi-Ming Ma, and Tie-Yan Liu. Positively scale-invariant flatness of ReLU neural networks. *arXiv preprint arXiv:1903.02237*, 2019.

Hongyu Yu, Boyu Liu, Yang Zhong, Liangliang Hong, Junyi Ji, Changsong Xu, Xingao Gong, and Hongjun Xiang. Physics-informed time-reversal equivariant neural network potential for magnetic materials. *Physical Review B*, 110(10):104427, 2024. doi: 10.1103/PhysRevB.110.104427.

Zilong Yuan, Zhiming Xu, He Li, Xinle Cheng, Honggeng Tao, Zechen Tang, Zhiyuan Zhou, Wenhui Duan, and Yong Xu. Equivariant neural network force fields for magnetic materials. *Quantum Frontiers*, 3(1):8, 2024. doi: 10.1007/s44214-024-00055-3.

Manzil Zaheer, Satwik Kottur, Siamak Ravanbakhsh, Barnabas Poczos, Russ R Salakhutdinov, and Alexander Smola. Deep sets. *Advances in neural information processing systems*, 30, 2017.

LiuPing Zeng, JingMei Pan, YangJie Lu, and Xian Pan. Stabilizing extreme few-shot ecg classification via self-supervised contrastive pretraining. *Annals of Noninvasive Electrocardiology*, 31(3):e70188, 2026. doi: 10.1111/anec.70188.

Ming Zhang, Lutao Liu, and Ming Diao. LPI radar waveform recognition based on time-frequency distribution. *Sensors*, 16(10):1682, October 2016. doi: 10.3390/s16101682.

Zinan Zheng, Yang Liu, Jia Li, Jianhua Yao, and Yu Rong. Relaxing continuous constraints of equivariant graph neural networks for broad physical dynamics learning. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 4548–4558. Association for Computing Machinery, 2024. doi: 10.1145/3637528.3671957.

CONTENTS

1 Introduction **1**

2 Background **2**

3 Creating Symmetry Slippage **3**
3.1 Decomposing Transformation Laws . . . 3
3.2 Local Symmetry Slippage . . . 4
3.2.1 Group-Closure . . . 4
3.2.2 Over-Quotienting . . . 5
3.2.3 Extension-Mistyping . . . 5

4 Computing Symmetry Slippage **5**

5 Correcting Symmetry Slippage **6**
5.1 Grippage Encodings . . . 6
5.2 Experiments . . . 7
5.2.1 Case Studies for Group-Closure . . . 7
5.2.2 Case Studies for Over-Quotienting . . . 8
5.2.3 Case Studies for Extension-Mistyping . . . 8

6 Discussion **9**

A Related Work **17**

B Elaborations Upon and Proofs of Results **18**
B.1 Proofs from Section 3.2 . . . 18
B.2 Elaboration and Proofs from Section 4 . . . 20
B.3 Elaboration and Proofs from Section 5.1 . . . 25

C Case Studies **27**
C.1 Group-Closure Case Studies . . . 27
C.1.1 Time-Orientation in Lorentz-Equivariant ML. . . . 27
C.1.2 Signs in Units-Equivariant ML . . . 28
C.1.3 Inner Automorphisms in Lie Algebra Learning . . . 29
C.2 Over-Quotienting Case Studies . . . 31
C.2.1 Roots in Glycan Learning . . . 31
C.2.2 Normalization in ECG ML . . . 32
C.2.3 Polarity in Time-Frequency Signal Processing . . . 33
C.3 Extension-Mistyping Case Studies . . . 35
C.3.1 (Pseudo)scalar Mistyping in Neural Operators . . . 35
C.3.2 Tetrahedral Chirality in Molecular Graph Learning . . . 36
C.3.3 Pseudovector Mistyping in Magnetic Equivariant ML . . . 37

D Experimental Details **39**

A Related Work

Over-Equivariance The most-studied example of symmetry slippage occurs when $\mathrm{A} := \underline{H}^{\alpha_Y}_{\alpha_X}$ is prescribed for some $H \leq \Gamma$, but a group $\Gamma \geq G \gneq H$ exists such that $\underline{H}^{\alpha_Y}_{\alpha_X} \subsetneq \underline{G}^{\alpha_Y}_{\alpha_X} \leq \mathrm{Sym}\,\mathscr{A}$; forms of this principle have have been called e.g. *excessive equivariance* (Han et al., 2022; Zheng et al., 2024) and *nonunique equivariance* Ravanbakhsh et al. (2017); Maron (2019). We adopt the term *over-equivariance*. Of our three creation mechanisms, group-closure always entails symmetry slippage by over-equivariance, over-quotienting sometimes does, and extension-mistyping never does. A familiar story of over-equivariance involves chirality in molecular learning: (Dumitrescu et al., 2025; Weaver & Lupo Pasini, 2025). The majority of such works follow a contour in which over-equivariance is identified by domain- and architecture-specific means, and a repair is tailored accordingly. In contrast, the present work considers the complementary question of how domain- and architecture-agnostic design choices can create over-equivariance (and more general symmetry slippage) to begin with, together with characterizing computational considerations and establishing general repair principles. The clearest precedent for such a program is the lineage (Ravanbakhsh et al., 2017; Maron et al., 2019), who systematically examined over-equivariance in the classical setting where $\mathscr{A}$ is built from morphisms of finite-group permutation representations. This paper may be viewed as a generalization of these works to the setting of contemporary equivariant ML. In particular, we identify the setting of Ravanbakhsh et al. (2017) as a doubly-special case of group-closure in Section 3.2, and in Section 4 we strengthen their computational results. Other precedents include Finzi et al. (2021), who characterize necessary (but, as we shall see, not sufficient) conditions for arbitrarily matrix-group equivariance, and Wang et al. (2023), who analyze the learning-theoretic costs of over-equivariance.

When symmetry slippage is caused by over-equivariance, our GEs may be viewed as symmetry-breaking features (Kaba & Ravanbakhsh, 2024; Xie & Smidt, 2024; Goel et al., 2026; Lawrence et al., 2025; Smidt et al., 2021) . The two nevertheless concern different aspects of symmetry and neither subsumes the other. On the one hand, symmetry slippage is an "architecture-level property", in the sense that belonging to $\mathrm{Sym}\,\mathscr{A}$ entails transformation-law obedience of every f^θ, $\theta \in \Theta$. In contrast, much symmetry breaking literature instead studies the refined question of how symmetry behaves pointwise under a fixed f^θ. This refined notion also underlies recent *symmetry increase* studies (Cen et al., 2024; Lin et al., 2026), which explore how equivariance forces individual inputs to map to outputs with strictly larger stabilizers, and how this increase can be reduced. On the other hand, not all symmetry slippage is over-equivariance, and accordingly not all GEs may be viewed as symmetry-breaking. To see an example of a well-known instance of symmetry slippage that is not over-equivariance, we turn to graph representation learning Scarselli et al. (2009); Gilmer et al. (2017); Hamilton et al. (2018).

Symmetry Slippage Beyond Over-Equivariance There is a natural action α_X of $\Gamma := S_n$ on the space $X = X_n$ of n-node graphs by node relabeling, and the most basic class of graph neural networks $\mathscr{A} : X \rightsquigarrow Y$ are designed to be invariant under α_X, i.e. invariant under graph isomorphism. In our notation, the transformation law $\mathrm{A} := \underline{H}^{I_Y}_{\alpha_X}$, where $H := \Gamma = S_n$, is prescribed. Equivalently, A prescribes that every realization $f^\theta : X \to Y$ is constant on the classes of the orbit relation $\sim_{S_n}$, and hence factors uniquely through X/S_n. It is known that message-passing graph neural networks $\mathscr{A}$ (Gilmer et al., 2017) are no more expressive than a coarser equivalence relation on X given by the Weisfeiler-Lehman Isomorphism Test (Weisfeiler & Lehman, 1968); we write $G \sim_{\mathrm{WL}} G'$ when WL fails to distinguish G and G'. Isomorphic graphs are WL-indistinguishable, so $\sim_{S_n} \subset \sim_{\mathrm{WL}}$, inducing a map $q : X/S_n \to X/\sim_{\mathrm{WL}}$ that is surjective but not generally injective. If $G \sim_{\mathrm{WL}} G'$, then $f^\theta(G) = f^\theta(G')$ for every realization $f^\theta \in \mathcal{F}_{\mathscr{A}}$. Thus every realization factors uniquely through $X/\sim_{\mathrm{WL}}$, and hence as $X \to X/S_n \xrightarrow{q} X/\sim_{\mathrm{WL}} \to Y$.

We claim this is over-quotienting in the sense of Definition 5, but not over-equivariance. Take $C_1 = I$, $C_2 = O$, and take $\sim := \sim_{\mathrm{WL}}$. Choose graphs $G_0, G_1 \in X$ with $G_0 \sim_{\mathrm{WL}} G_1$ but $G_0 \not\sim_{S_n} G_1$. Consider $(\sigma, \tau) := (I_X, I_Y) \in \mathrm{A}$. With $J = \mathrm{Bij}(X)$, define $\sigma' \in J$ to swap G_0 and G_1 and fix every other graph. Then $G \sim_{\mathrm{WL}} \sigma'(G)$ for every $G \in X$. Since every f^θ is constant on WL-classes, $f^\theta \sigma' = f^\theta$, and hence $(\sigma', I_Y) \in \mathrm{Sym}\,\mathscr{A}$. On the other hand, $(\sigma', I_Y) \notin \mathrm{A}$. Indeed, every input component of an element of A is a vertex relabeling $\alpha_X(\pi)$, which can send a graph only to an isomorphic graph, whereas σ' sends G_0 to the nonisomorphic graph G_1. Thus $\sigma' \notin \alpha_X(S_n)$, and therefore $(\sigma', I_Y) \in \mathrm{Sym}\,\mathscr{A} - \mathrm{A}$. Therefore, vanilla message-passing GNNs exhibit symmetry slippage by over-quotienting in light of Proposition 6. This slippage is not over-

equivariance, however. Indeed, here one has $H = \Gamma = S_n$, while over-equivariance requires group extension $H \lneq G \leq \Gamma$.

Our case studies provide further examples where symmetry slippage beyond over-equivariance occurs.

B Elaborations Upon and Proofs of Results

In this appendix section, we provide the precise statements and proofs of results referenced in the main text. The majority of results from Appendices B.1 and B.3 are straightforward once the relevant definitions are in place. The results from Section B.2 are considerably less straightforward. For a family $\mathcal{F} \subset Y^X$, recall that we write $\mathcal{F}^\top : X \to Y^{\mathcal{F}}$ for the map $x \mapsto (f(x))_{f\in\mathcal{F}}$; for an architecture $\mathscr{A}$, write $\mathscr{A}^\top := \mathcal{F}_{\mathscr{A}}^\top$.

B.1 Proofs from Section 3.2

Proposition 1. *Any subarchitecture $\mathscr{A}_{(C_1,C_2)}$ gives rise to a (lax) decomposition*

$$\operatorname{Sym}\mathscr{A}_{(C_2,O)} \circ \operatorname{Sym}\mathscr{A}_{(C_1,C_2)} \circ \operatorname{Sym}\mathscr{A}_{(I,C_1)} \leq \operatorname{Sym}\mathscr{A}. \tag{1}$$

Equality need not hold in general.

Proof. Let $(\sigma,\tau) \in \operatorname{Sym}\mathscr{A}_{(C_2,O)} \circ \operatorname{Sym}\mathscr{A}_{(C_1,C_2)} \circ \operatorname{Sym}\mathscr{A}_{(I,C_1)}$, so that there exist δ_1, δ_2 such that $(\sigma,\delta_1) \in \operatorname{Sym}\mathscr{A}_{(I,C_1)}$, $(\delta_1,\delta_2) \in \operatorname{Sym}\mathscr{A}_{(C_1,C_2)}$, and $(\delta_2,\tau) \in \operatorname{Sym}\mathscr{A}_{(C_2,O)}$. As described in Section 3.1, any $\theta \in \Theta$ gives rise to $f^\theta = e^\theta\Phi^\theta a^\theta$, where a^θ, Φ^θ, and e^θ are the maps realized by $\mathscr{A}_{(I,C_1)}$, $\mathscr{A}_{(C_1,C_2)}$, and $\mathscr{A}_{(C_2,O)}$, respectively. Then $f^\theta\sigma = e^\theta\Phi^\theta a^\theta\sigma = e^\theta\Phi^\theta\delta_1 a^\theta = e^\theta\delta_2\Phi^\theta a^\theta = \tau e^\theta\Phi^\theta a^\theta = \tau f^\theta$. Hence $(\sigma,\tau) \in \operatorname{Sym}\mathscr{A}$, hence $\operatorname{Sym}\mathscr{A}_{(C_2,O)} \circ \operatorname{Sym}\mathscr{A}_{(C_1,C_2)} \circ \operatorname{Sym}\mathscr{A}_{(I,C_1)} \leq \operatorname{Sym}\mathscr{A}$.

To see that equality need not hold, consider the following counterexample. Let $V = \{i, c_1, c_2, o\}$ with $I = \{i\}$, $C_1 = \{c_1\}$, $C_2 = \{c_2\}$, $O = \{o\}$, and $E = \{(i,c_1),(c_1,c_2),(c_2,o)\}$. Let $Z_v = \mathbb{R}^3$ for every $v \in V$, let $\Gamma = S_3$, and let $\boldsymbol{\alpha} = (\alpha_v : \Gamma \curvearrowright Z_v)_{v\in V}$ each be the standard representation of S_3 on $\mathbb{R}^3$. Take $J = \alpha_I(\Gamma)$ and $K = \alpha_O(\Gamma)$. Let $\Theta = S_3$ and define the realizations of the three subarchitectures by $a^\pi = P_\pi$, $\Phi^\pi = \mathrm{id}_{\mathbb{R}^3}$, and $e^\pi = P_\pi^{-1}$. Then f^π realized by $\mathscr{A}$ satisfies $f^\pi = e^\pi\Phi^\pi a^\pi = \mathrm{id}_{\mathbb{R}^3}$, and hence $\operatorname{Sym}\mathscr{A} = \{(P_\sigma, P_\sigma) : \sigma \in S_3\}$. On the other hand, if $(P_\sigma, P_\tau) \in \operatorname{Sym}\mathscr{A}_{(I,C_1)}$, then $P_\pi P_\sigma = P_\tau P_\pi$ for every $\pi \in S_3$. Taking $\pi = e$ gives $\tau = \sigma$, after which σ must commute with every element of S_3. But $Z(S_3) = \{e\}$, s0 $\operatorname{Sym}\mathscr{A}_{(I,C_1)} = \{(\mathrm{id},\mathrm{id})\}$. The same argument gives $\operatorname{Sym}\mathscr{A}_{(C_2,O)} = \{(\mathrm{id},\mathrm{id})\}$, while $\operatorname{Sym}\mathscr{A}_{(C_1,C_2)} = \{(P_\sigma,P_\sigma) : \sigma \in S_3\}$. Thus $\operatorname{Sym}\mathscr{A}_{(C_2,O)} \circ \operatorname{Sym}\mathscr{A}_{(C_1,C_2)} \circ \operatorname{Sym}\mathscr{A}_{(I,C_1)} = \{(\mathrm{id},\mathrm{id})\} \lneq \operatorname{Sym}\mathscr{A}$. □

Proposition 3. *Let $H \lneq G \leq \Gamma$, let $\alpha_X := \alpha|_G$, and let $G \overset{\alpha_Y}{\curvearrowright} Y$ with $\alpha_Y(G) \leq K$, and suppose $\mathrm{A} = \underline{H}_{\alpha_X}^{\alpha_Y} \lneq \underline{G}_{\alpha_X}^{\alpha_Y}$. Let $\mathscr{A}_{(C_1,C_2)}$ be a subarchitecture on which $\mathcal{C}$ closes H to G. Suppose the node maps ϕ_v^θ of $\mathscr{A}_{(C_1,C_2)}$ are H-equivariant, and suppose that a^θ and e^θ are G-equivariant for all θ. Then $\underline{H}_{\alpha_X}^{\alpha_Y} \lneq \underline{G}_{\alpha_X}^{\alpha_Y} \leq \operatorname{Sym}\mathscr{A}$, and group-closure therefore causes symmetry slippage.*

Proof. Since $\mathcal{C}$ closes H to G on $\mathscr{A}_{(C_1,C_2)}$ and every ϕ_v^θ is H-equivariant, every ϕ_v^θ is G-equivariant. Hence the $\Phi^\theta : Z_{C_1} \to Z_{C_2}$ realized by $\mathscr{A}_{(C_1,C_2)}$ is G-equivariant. For every $g \in G$,

$$\begin{aligned} f^\theta\alpha_X(g) &= e^\theta\Phi^\theta a^\theta\alpha_X(g) && (2)\\ &= e^\theta\Phi^\theta\alpha_{C_1}(g)a^\theta && (3)\\ &= e^\theta\alpha_{C_2}(g)\Phi^\theta a^\theta && (4)\\ &= \alpha_Y(g)e^\theta\Phi^\theta a^\theta && (5)\\ &= \alpha_Y(g)f^\theta && (6) \end{aligned}$$

whence every $f^\theta \in \mathcal{F}_{\mathscr{A}}$ is G-equivariant, so $\underline{G}_{\alpha_X}^{\alpha_Y} \leq \operatorname{Sym}\mathscr{A}$. Since by assumption $\mathrm{A} = \underline{H}_{\alpha_X}^{\alpha_Y} \lneq \underline{G}_{\alpha_X}^{\alpha_Y}$, we get $\mathrm{A} \lneq \underline{G}_{\alpha_X}^{\alpha_Y} \leq \operatorname{Sym}\mathscr{A}$, and therefore $\mathscr{A}$ exhibits symmetry slippage. □

Proposition 4. *If the spaces Z_v for v in $\mathscr{A}_{(C_1,C_2)}$ are finite-dimensional G-representations and polynomials $\mathbb{R}[X]$ close H to G on $\mathscr{A}_{(C_1,C_2)}$, then $\mathcal{C} := C^\omega(X)$ closes H to G on $\mathscr{A}_{(C_1,C_2)}$.*

Proof. Fix a node v of $\mathscr{A}_{(C_1,C_2)}$ and an H-equivariant $\phi \in C^\omega(X)$. Since the spaces are G-representations, $g0 = 0$ for all $g \in G$. Taylor-expanding ϕ at 0 gives on an open neighborhod of 0, $\phi(x) = \sum_{m=0}^{\infty} P_m(x)$, where $P_m(x) := \frac{1}{m!} D^m\phi(0)[x, \ldots, x]$ is a homogeneous polynomial map of degree m. Using H-equivariance and expanding both sides at 0 then gives $\sum_{m=0}^{\infty} P_m(hx) = \sum_{m=0}^{\infty} hP_m(x)$. Matching terms, $P_m(hx) = hP_m(x)$ for every m, from which we see each P_m is H-equivariant. Because polynomials close H to G on $\mathscr{A}_{(C_1,C_2)}$, each P_m is therefore G-equivariant.

Now fix $g \in G$. For x in a small enough neighborhood U of 0 that the Taylor expansions of both $\phi(x)$ and $\phi(gx)$ converge, $\phi(gx) = \sum_{m=0}^{\infty} P_m(gx) = \sum_{m=0}^{\infty} gP_m(x) = g\sum_{m=0}^{\infty} P_m(x) = g\phi(x)$. Thus the real-analytic map $x \mapsto \phi(gx) - g\phi(x)$ vanishes on U. But by analytic continuation it must then vanish everywhere. Hence ϕ is G-equivariant. Since v and ϕ were arbitrary, $C^\omega(X)$ closes H to G on $\mathscr{A}_{(C_1,C_2)}$. □

Proposition 6. *If $\mathscr{A}$ over-quotients* A*, then $f\sigma'\sigma^{-1} = f$ for every $f \in \mathcal{F}_{\mathscr{A}}$. Therefore, $(\sigma'\sigma^{-1}, I_Y) \in \operatorname{Sym}\mathscr{A} - \mathrm{A}$ and $\mathscr{A}$ exhibits symmetry slippage.*

Proof. Suppose $\mathscr{A}$ over-quotients A. Then for every $\theta \in \Theta$ and $x \in X$, $a^\theta\sigma(x)$ and $a^\theta\sigma'(x)$ have the same image in $Z_{C_1}/\!\sim$, and since Φ^θ factors through $Z_{C_1}/\!\sim$, $\Phi^\theta a^\theta\sigma(x) = \Phi^\theta a^\theta\sigma'(x)$. Applying e^θ gives $f^\theta\sigma = f^\theta\sigma'$ for every θ. Applying σ^{-1} to both sides, $f^\theta = f^\theta\sigma'\sigma^{-1}$. Since $\sigma, \sigma' \in J$, we have $\sigma'\sigma^{-1} \in J$, and $I_Y \in K$. Hence $(\sigma'\sigma^{-1}, I_Y) \in \operatorname{Sym}\mathscr{A}$. But Definition 5 asserts $(\sigma'\sigma^{-1}, I_Y) \notin \mathrm{A}$. Therefore $(\sigma'\sigma^{-1}, I_Y) \in \operatorname{Sym}\mathscr{A} - \mathrm{A}$, so $\operatorname{Sym}\mathscr{A} \not\subset \mathrm{A}$ and $\mathscr{A}$ exhibits symmetry slippage. □

Proposition 8. *Let $H \lneq G$, and suppose $\mathscr{A}$ extension-mistypes $\widetilde{\rho}^{\mathrm{p}}$ on $\mathscr{A}_{(C_1,C_2)}$ via $\widetilde{\rho}^{\mathrm{r}}$, with actions $G \curvearrowright_{\alpha^{\mathrm{r}}_X} X, G \curvearrowright_{\alpha^{\mathrm{r}}_Y} Y$ satisfying $\alpha^{\mathrm{r}}_X(G) \le J$ and $\alpha^{\mathrm{r}}_Y(G) \le K$, such that a^θ is $(\alpha^{\mathrm{r}}_X, \widetilde{\rho}^{\mathrm{r}}_{C_1})$-equivariant and e^θ is $(\widetilde{\rho}^{\mathrm{r}}_{C_2}, \alpha^{\mathrm{r}}_Y)$-equivariant. Then $\underline{G}^{\alpha^{\mathrm{r}}_Y}_{\alpha^{\mathrm{r}}_X} \le \operatorname{Sym}\mathscr{A}$. Therefore, if $\underline{G}^{\alpha^{\mathrm{r}}_Y}_{\alpha^{\mathrm{r}}_X} \not\subset \mathrm{A}$, then $\operatorname{Sym}\mathscr{A} \not\subset \mathrm{A}$ and $\mathscr{A}$ exhibits symmetry slippage.*

Proof. Let $H \lneq G$, and suppose $\mathscr{A}$ extension-mistypes $\widetilde{\rho}^{\mathrm{p}}$ on $\mathscr{A}_{(C_1,C_2)}$ as witnessed by some extension $\widetilde{\rho}^{\mathrm{r}}$ of ρ. Suppose there exist actions $\alpha^{\mathrm{r}}_X : G \curvearrowright X$, $\alpha^{\mathrm{r}}_Y : G \curvearrowright Y$ such that a^θ is G-equivariant with respect to α^{r}_X and $\widetilde{\rho}^{\mathrm{r}}_{C_1}$, and e^θ is G-equivariant with respect to $\widetilde{\rho}^{\mathrm{r}}_{C_2}$ and α^{r}_Y. By extension-mistyping, every node map ϕ^θ_v of $\mathscr{A}_{(C_1,C_2)}$ is G-equivariant with respect to $\widetilde{\rho}^{\mathrm{r}}$, and hence the realized map $\Phi^\theta : Z_{C_1} \to Z_{C_2}$ is G-equivariant with respect to $\widetilde{\rho}^{\mathrm{r}}_{C_1}$ and $\widetilde{\rho}^{\mathrm{r}}_{C_2}$. Thus, for every $g \in G$ and $\theta \in \Theta$,

$$
\begin{aligned}
f^\theta\alpha^{\mathrm{r}}_X(g) &= e^\theta\Phi^\theta a^\theta\alpha^{\mathrm{r}}_X(g) && (7)\\
&= e^\theta\Phi^\theta\widetilde{\rho}^{\mathrm{r}}_{C_1}(g)a^\theta && (8)\\
&= e^\theta\widetilde{\rho}^{\mathrm{r}}_{C_2}(g)\Phi^\theta a^\theta && (9)\\
&= \alpha^{\mathrm{r}}_Y(g)e^\theta\Phi^\theta a^\theta && (10)\\
&= \alpha^{\mathrm{r}}_Y(g)f^\theta && (11)
\end{aligned}
$$

Hence every f^θ is G-equivariant with respect to α^{r}_X and α^{r}_Y. Since $\alpha^{\mathrm{r}}_X(G) \le J$ and $\alpha^{\mathrm{r}}_Y(G) \le K$, it follows that $\underline{G}\alpha^{\mathrm{r}}_X{}^{\alpha^{\mathrm{r}}_Y} \le \operatorname{Sym}\mathscr{A}$. So if $\underline{G}\alpha^{\mathrm{r}}_X{}^{\alpha^{\mathrm{r}}_Y} \not\subset \mathrm{A}$, then some element of $\underline{G}^{\alpha^{\mathrm{r}}_Y}_{\alpha^{\mathrm{r}}_X}$ lives in $\operatorname{Sym}\mathscr{A} - \mathrm{A}$, and so $\operatorname{Sym}\mathscr{A} \not\subset \mathrm{A}$ and $\mathscr{A}$ exhibits symmetry slippage. □

Proposition 9. *Suppose $H \triangleleft G$ with $[G : H] = 2$; choose $s \in G - H$. Suppose the node maps ϕ^θ_v are H-equivariant with respect to $\boldsymbol{\rho}$, and suppose there exist linear maps $\boldsymbol{S} = \big(S_v \in \mathrm{GL}(Z_v)\big)_{v\in V}$ satisfying (i) $S_v = \eta_v(s)$ for all $v \in T$, $i_{S_v} \circ \rho_v = \rho_v \circ i_s$[3], (ii) $S_v^2 = \rho_v(s^2)$, and (iii) $\phi^\theta_v S_{\{u\to v\}} = S_v\phi^\theta_v$ for all v, θ. If $\widetilde{\rho}^{\mathrm{p}}_v(s) \neq S_v$ for some $v \in T$, then there exists an extension $\widetilde{\boldsymbol{\rho}}^{\mathrm{r}}$ of $\boldsymbol{\rho}$ to G such that $\mathscr{A}$ extension-mistypes $\widetilde{\boldsymbol{\rho}}^{\mathrm{p}}$ on $\mathscr{A}_{(C_1,C_2)}$.*

Proof. Since $[G : H] = 2$, $G = H \sqcup sH$, so every $g \in G$ equals either h or sh for some $h \in H$. For each v, define $\widetilde{\rho}^{\mathrm{r}}_v : G \to \operatorname{Aut} Z_v$ by $\widetilde{\rho}^{\mathrm{r}}_v(h) := \rho_v(h)$ and $\widetilde{\rho}^{\mathrm{r}}_v(sh) := S_v\rho_v(h)$. We claim $\widetilde{\rho}^{\mathrm{r}}_v$ witnesses that $\mathscr{A}$ extension-mistypes $\widetilde{\rho}^{\mathrm{p}}$. This requires showing (Step 1) that $\widetilde{\rho}^{\mathrm{r}}_v : G \to \operatorname{Aut} Z_v$

[3]Here $i_g(x)$ denotes conjugation of x by g.

defines action extending ρ_v, (Step 2) that $\widetilde{\rho}^{\mathrm{r}}_v = \eta_v$ forall $v \in T$, (Step 3) that $\eta_v \neq \widetilde{\rho}^{\mathrm{p}}_v$ for some $v \in T$, (Step 4) the ϕ^θ_v are equivariant with respect to $\widetilde{\rho}^{\mathrm{r}}_v$.

Step 1. We first claim $\widetilde{\rho}^{\mathrm{r}}_v : G \to \mathrm{Aut}\, Z_v$ defines action extending ρ_v. This requires checking (a) that $\widetilde{\rho}^{\mathrm{r}}_v|_H = \rho_v$ and (b) that $\widetilde{\rho}^{\mathrm{r}}_v$ is a homomorphism. (a) holds by construction; for (b), it suffices to check that $\widetilde{\rho}^{\mathrm{r}}_v$ respects products of the form h_1h_2, sh_1h_2, h_1sh_2, sh_1sh_2. The first is immediate from (a). The second is immediate, too, by the definition of $\widetilde{\rho}^{\mathrm{r}}_v$ and the fact that ρ_v is homomorphism. For the case h_1sh_2, note that since $H \triangleleft G$ it holds $s^{-1}h_1s \in H$, and (i) gives $\rho_v(h_1)S_v = S_v\rho_v(s^{-1}h_1s)$. Therefore

$$\begin{aligned}\widetilde{\rho}^{\mathrm{r}}_v(h_1)\widetilde{\rho}^{\mathrm{r}}_v(sh_2) &= \rho_v(h_1)S_v\rho_v(h_2)\\ &= S_v\rho_v(s^{-1}h_1s)\rho_v(h_2)\\ &= S_v\rho_v(s^{-1}h_1sh_2)\\ &= \widetilde{\rho}^{\mathrm{r}}_v(h_1sh_2),\end{aligned}$$

where the last equality we used $h_1sh_2 = s(s^{-1}h_1sh_2)$. Lastly, for the case sh_1sh_2, condition (i) gives $S_v\rho_v(h_1)S_v^{-1} = \rho_v(sh_1s^{-1})$, while condition (ii) gives $S_v^2 = \rho_v(s^2)$. Hence

$$\begin{aligned}\widetilde{\rho}^{\mathrm{r}}_v(sh_1)\widetilde{\rho}^{\mathrm{r}}_v(sh_2) &= S_v\rho_v(h_1)S_v\rho_v(h_2)\\ &= \rho_v(sh_1s^{-1})S_v^2\rho_v(h_2)\\ &= \rho_v(sh_1s^{-1})\rho_v(s^2)\rho_v(h_2)\\ &= \rho_v(sh_1sh_2)\\ &= \widetilde{\rho}^{\mathrm{r}}_v(sh_1sh_2).\end{aligned}$$

Here $sh_1sh_2 \in H$: normality gives $sh_1s^{-1} \in H$, while $[G : H] = 2$ implies $s^2 \in H$. Thus $\widetilde{\rho}^{\mathrm{r}}_v$ indeed defines a G-action extending ρ_v.

Step 2. We next claim that $\widetilde{\rho}^{\mathrm{r}}_v = \eta_v$ for every $v \in T$. Indeed, both actions restrict to ρ_v on H, while condition (i) gives $\widetilde{\rho}^{\mathrm{r}}_v(s) = S_v = \eta_v(s)$. Since every element of G is either h or sh for some $h \in H$, for $sh \in sH$ we have

$$\widetilde{\rho}^{\mathrm{r}}_v(sh) = S_v\rho_v(h) = \eta_v(s)\eta_v(h) = \eta_v(sh).$$

Thus $\widetilde{\rho}^{\mathrm{r}}_v = \eta_v$ on all of G for every $v \in T$.

Step 3. This is clear since by hypothesis $\widetilde{\rho}^{\mathrm{p}}_v(s) \neq S_v = \widetilde{\rho}^{\mathrm{r}}_v(s) = \eta_v(s)$ for some $v \in T$, so $\eta_v \neq \widetilde{\rho}^{\mathrm{p}}_v$ for that v.

Step 4. It remains to check that every node map is G-equivariant with respect to $\widetilde{\boldsymbol{\rho}}^{\mathrm{r}}$. By assumption, the node maps are already H-equivariant with respect to $\boldsymbol{\rho}$, and $\widetilde{\boldsymbol{\rho}}^{\mathrm{r}}|_H = \boldsymbol{\rho}$. Thus it remains only to check elements of sH. For $sh \in sH$, condition (iii) and H-equivariance give

$$\begin{aligned}\phi^\theta_v\widetilde{\rho}^{\mathrm{r}}_{\{u\to v\}}(sh) &= \phi^\theta_v S_{\{u\to v\}}\rho_{\{u\to v\}}(h)\\ &= S_v\phi^\theta_v\rho_{\{u\to v\}}(h)\\ &= S_v\rho_v(h)\phi^\theta_v\\ &= \widetilde{\rho}^{\mathrm{r}}_v(sh)\phi^\theta_v.\end{aligned}$$

Since $G = H \sqcup sH$, every node map is therefore G-equivariant with respect to $\widetilde{\boldsymbol{\rho}}^{\mathrm{r}}$. We have shown that $\widetilde{\boldsymbol{\rho}}^{\mathrm{r}}$ is an extension of $\boldsymbol{\rho}$, that $\widetilde{\rho}^{\mathrm{r}}_v = \eta_v$ for every $v \in T$, that $\eta_v \neq \widetilde{\rho}^{\mathrm{p}}_v$ for some $v \in T$, and that every node map is G-equivariant with respect to $\widetilde{\boldsymbol{\rho}}^{\mathrm{r}}$.

The result follows. □

B.2 Elaboration and Proofs from Section 4

This section more precisely states, then proves, the statements claimed in Section 4.

Theorem 17. *For every finite group H, computing additional generators B satisfying $\mathrm{Sym}\,\mathscr{A} = \langle \mathrm{A}, B\rangle$ is GI-hard, even when $\mathrm{A} = \underline{H}^{\alpha_Y}_{\alpha_X} \leq \mathrm{Sym}\,\mathscr{A}$ is promised.*

To prove Theorem 17 we first prove the following, which will also be useful for Proposition 20 below. Let I, O be finite index sets. Recall from Subexample 1 that the architecture $\mathscr{A}$ corresponding to a parameter-sharing linear MLP layer $X := \mathbb{R}^I \to \mathbb{R}^O =: Y$ may be encoded as a coloring c of the bipartite graph $(O, I, O \times I)$, and conversely that any colored bipartite graph

$\Sigma = (O, I, O\times I, c)$ determines a parameter-sharing linear layer $W : \mathbb{R}^I \to \mathbb{R}^O$ as $W = \sum_\gamma w_\gamma A_\gamma$, where $(A_\gamma)_{oi} = \mathbf{1}_{\{c(o,i)=\gamma\}}$.

Proposition 18. *Let $\mathscr{A}$ denote the architecture corresponding to a linear* MLP *layer $\mathbb{R}^I \to \mathbb{R}^O$, sharing parameters according to colored graph Σ. Then $\operatorname{Aut}\Sigma = \operatorname{Sym}\mathscr{A} \leq S_I \times S_O$. The result continues to hold for a nonlinear layer $x \mapsto \mathrm{a}(Wx)$ provided $(r \in \mathbb{R}) \mapsto (t \mapsto \mathrm{a}(tr))$ is injective.*

Proof. We first consider the linear case. For each $\gamma \in \operatorname{im} c$, let A_γ denote its indicator matrix and define $W_{\boldsymbol{w}} := \sum_\gamma w_\gamma A_\gamma$ where w_γ are trainable weights. Hence

$$(P, Q) \in \operatorname{Sym}\mathscr{A} \iff W_{\boldsymbol{w}} P = Q W_{\boldsymbol{w}} \text{ for all } \boldsymbol{w} \tag{12}$$
$$\iff A_\gamma P = Q A_\gamma \text{ for all } \gamma \tag{13}$$
$$\iff (P, Q) \in \operatorname{Aut}(\Sigma), \tag{14}$$

whence $\operatorname{Sym}\mathscr{A} = \operatorname{Aut}(\Sigma)$.

Now consider the nonlinear case $x \mapsto \mathrm{a}(W_{\boldsymbol{w}} x)$. The inclusion $\operatorname{Aut}(\Sigma) \subset \operatorname{Sym}\mathscr{A}$ follows from the linear case, since coordinate permutations commute with a. For the reverse inclusion, if $(P, Q) \in \operatorname{Sym}\mathscr{A}$, then for every $\boldsymbol{w}, x$, $\mathrm{a}(W_{\boldsymbol{w}} P x) = Q\mathrm{a}(W_{\boldsymbol{w}} x) = \mathrm{a}(Q W_{\boldsymbol{w}} x)$. Applying the same identity to $t\boldsymbol{w}$ gives $\mathrm{a}(tW_{\boldsymbol{w}} P x) = \mathrm{a}(tQW_{\boldsymbol{w}} x)$ for all $t \in \mathbb{R}$. Since $r \mapsto (t \mapsto \mathrm{a}(tr))$ is injective, this implies coordinatewise that $W_{\boldsymbol{w}} P x = Q W_{\boldsymbol{w}} x$. Thus $W_{\boldsymbol{w}} P = Q W_{\boldsymbol{w}}$ for every $\boldsymbol{w}$, and the linear case gives $(P, Q) \in \operatorname{Aut}(\Sigma)$. Thus $\operatorname{Sym}\mathscr{A} = \operatorname{Aut}(\Sigma)$, as claimed. □

Proof. Let H a finite group be given. Given any finite simple graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, take $I := \mathcal{V} \times H =: O$. H acts on O and I in the evident way, $h \cdot (o, h_o) = (o, hh_o)$, $h \cdot (i, h_i) = (i, hh_i)$, inducing permutation representations $\rho_X : H \to S_I \leq \mathrm{GL}(\mathbb{R}^I)$, $\rho_Y : H \to S_O \leq \mathrm{GL}(\mathbb{R}^O)$. Take the family of actions $\alpha_X := \rho_X$, $\alpha_Y := \rho_Y$. We construct in polynomial time an architecture $\mathscr{A}$ for which $\underline{H}^{\alpha_Y}_{\alpha_X} =: \mathrm{A} \leq \operatorname{Sym}\mathscr{A}$, and show this reduces Graph Isomorphism to the problem of computing additional generators B satisfying $\langle \mathrm{A}, B\rangle = \operatorname{Sym}\mathscr{A}$.

So let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a finite simple graph. Put $I := \mathcal{V} \times H =: O$. Define $\kappa : \mathcal{V} \times \mathcal{V} \to \{\text{“self”}, \text{“edge”}, \text{“nonedge”}\}$ by

$$\kappa(v_o, v_i) = \begin{cases} \text{“self”} & o = i \\ \text{“edge”} & \{o, i\} \in \mathcal{E} \\ \text{“nonedge”} & \text{else.} \end{cases}$$

Define

$$c : O \times I \to \{\text{“self”}, \text{“edge”}, \text{“nonedge”}\} \times H \tag{15}$$
$$\big((o, h_o), (i, h_i)\big) \mapsto \underbrace{\big(\kappa(o, i), h_o^{-1} h_i\big)}_{=:\left(c_1(o,h_0), c_2(i,h_i)\right)} \tag{16}$$

and put $\Sigma := (O, I, O \times I, c)$. Let $\mathscr{A}$ be the corresponding linear parameter-sharing MLP layer (cf. Subexample 1).

For all $h, h_o, h_i \in H$ and $o, i \in \mathcal{V}$, one has

$$c\big((o, hh_o), (i, hh_i)\big) = \big(\kappa(o, i), (hh_o)^{-1}(hh_i)\big) \tag{17}$$
$$= \big(\kappa(o, i), h_o^{-1} h_i\big) \tag{18}$$
$$= c\big((o, h_o), (i, h_i)\big) \tag{19}$$

from which it follows that $\underline{H}^{\alpha_Y}_{\alpha_X} = \mathrm{A} \leq \operatorname{Sym}\mathscr{A}$. By Proposition 18, then, $\operatorname{Sym}\mathscr{A} = \operatorname{Aut}\Sigma$. We claim that $\operatorname{Aut}\Sigma \cong \operatorname{Aut}(\mathcal{G}) \times H$, and thus $\operatorname{Sym}\mathscr{A} \cong \operatorname{Aut}(\mathcal{G}) \times H$.

To this end let $(P, Q) \in \operatorname{Aut}\Sigma$, $P = (P_1 : \mathcal{V} \times H \to \mathcal{V}, P_2 : \mathcal{V} \times H \to H)$, $Q = (Q_1 : \mathcal{V} \times H \to \mathcal{V}, Q_2 : \mathcal{V} \times H \to H)$. Given $v \in \mathcal{V}$, observe that $c_1\big((v, \cdot), (v, _)\big) \equiv$ “self”. So since (P, Q) preserves colors, $c_1\big(Q(v, \cdot), P(v, _)\big) =$ “self”, and hence $Q_1(v, \cdot) \equiv P_1(v, _) =: \pi(v)$ for

some $\pi \in S_{\mathcal{V}}$. Write $p_v(b) := P_2(v, b)$ and $q_v(a) := Q_2(v, a)$, so that $P(v, b) = (\pi(v), p_v(b))$, $Q(v, a) = (\pi(v), q_v(a))$. Let $g_v = p_v(e)$. We would like to define a prospective group isomorphism

$$\Phi : \operatorname{Aut}\Sigma \to \operatorname{Aut}(\mathcal{G}) \times H \tag{20}$$

$$(P, Q) \mapsto (\pi, g_v). \tag{21}$$

Checking this is well-defined requires (1) checking π is an automorphism of $\mathcal{G}$ and (2) checking independence of $v \in V$. We then must check (3) Φ is a group homomorphism and (4) Φ is invertible. We treat these in turn.

Step 1. To see that $\pi \in \operatorname{Aut}\mathcal{G}$, observe that since (P, Q) preserves colors, $c_1\big(Q(o, a), P(i, b)\big) = c_1\big((o, a), (i, b)\big)$ for all $o, i \in \mathcal{V}$ and $a, b \in H$. By the definition of π, the left-hand side is $\kappa(\pi(o), \pi(i))$, while the right-hand side is $\kappa(o, i)$. Hence $\kappa(\pi(o), \pi(i)) = \kappa(o, i)$ for all $o, i \in \mathcal{V}$. In particular, $\{o, i\} \in \mathcal{E}$ if and only if $\{\pi(o), \pi(i)\} \in \mathcal{E}$. Since $\pi \in S_{\mathcal{V}}$, it follows that $\pi \in \operatorname{Aut}\mathcal{G}$.

Step 2. To check that Φ is independent of $v \in \mathcal{V}$, first use preservation of second color component on $((v, a), (v, b))$. This gives $q_v(a)^{-1}p_v(b) = a^{-1}b$ for all $a, b \in H$. Taking $a = b$ gives $p_v(a) = q_v(a)$, while taking $a = e$ gives $p_v(b) = q_v(e)b$. Since $g_v := p_v(e) = q_v(e)$, we therefore have $p_v(b) = q_v(b) = g_v b$ for all $b \in H$. Thus $P(v, a) = Q(v, a) = (\pi(v), g_v a)$. Now let $u, v \in \mathcal{V}$. Preservation of the second color component on $((u, a), (v, b))$ gives $(g_u a)^{-1}(g_v b) = a^{-1}b$. Taking $a = b = e$ yields $g_u = g_v$. Hence g_v is independent of v.

Step 3. To check that Φ is a group homomorphism, let $\Phi(P_j, Q_j) = (\pi_j, g_j)$ for $j = 1, 2$. By Step 2, $P_j(v, a) = Q_j(v, a) = (\pi_j(v), g_j a)$. Hence $P_1P_2(v, a) = P_1(\pi_2(v), g_2 a) = (\pi_1\pi_2(v), g_1 g_2 a)$, and likewise $Q_1Q_2(v, a) = (\pi_1\pi_2(v), g_1 g_2 a)$. Therefore $\Phi\big((P_1, Q_1)(P_2, Q_2)\big) = (\pi_1\pi_2, g_1 g_2) = \Phi(P_1, Q_1)\Phi(P_2, Q_2)$.

Step 4. We claim the map $\Psi : \operatorname{Aut}(\mathcal{G}) \times H \to \operatorname{Aut}(\Sigma)$ given by $\Psi(\pi, g) := (\Lambda_{\pi,g}, \Lambda_{\pi,g})$, where $\Lambda_{\pi,g}(v, a) := (\pi(v), ga)$, is an inverse to Φ. First, Ψ is well-defined: for any $(\pi, g) \in \operatorname{Aut}(\mathcal{G}) \times H$, we have

$$c\big(\Lambda_{\pi,g}(o, h_o), \Lambda_{\pi,g}(i, h_i)\big)\big(\kappa(\pi(o), \pi(i)), (gh_o)^{-1}(gh_i)\big) \tag{22}$$

$$= \big(\kappa(o, i), h_o^{-1}h_i\big) \tag{23}$$

$$= c\big((o, h_o), (i, h_i)\big), \tag{24}$$

where the second equality uses $\pi \in \operatorname{Aut}(\mathcal{G})$. Thus $(\Lambda_{\pi,g}, \Lambda_{\pi,g}) \in \operatorname{Aut}(\Sigma)$. By construction, $\Phi(\Psi(\pi, g)) = (\pi, g)$ for every $(\pi, g) \in \operatorname{Aut}(\mathcal{G}) \times H$. Conversely, if $(P, Q) \in \operatorname{Aut}\Sigma$ and $\Phi(P, Q) = (\pi, g)$, then Step 2 gives $P(v, a) = Q(v, a) = (\pi(v), ga) = \Lambda_{\pi,g}(v, a)$ for all $v \in \mathcal{V}$ and $a \in H$, so $\Psi(\Phi(P, Q)) = (P, Q)$. Hence $\Psi = \Phi^{-1}$, and therefore $\operatorname{Aut}\Sigma \cong \operatorname{Aut}(\mathcal{G}) \times H$.

We therefore see $\operatorname{Sym}\mathscr{A} \cong \operatorname{Aut}(\mathcal{G}) \times H$. Now let $\mathcal{G}_1 = (\mathcal{V}_1, \mathcal{E}_1)$ and $\mathcal{G}_2 = (\mathcal{V}_2, \mathcal{E}_2)$ be connected finite simple graphs, and put $\mathcal{K} := \mathcal{G}_1 \sqcup \mathcal{G}_2$. Apply the preceding construction to $\mathcal{K}$, producing an architecture $\mathscr{A}_{\mathcal{K}}$, and suppose additional generators $B = \{(P_\ell, Q_\ell)\}_\ell$ satisfying $\langle \mathrm{A}, B\rangle = \operatorname{Sym}\mathscr{A}_{\mathcal{K}}$ are computed. Together with the known generators of A, the permutations P_ℓ generate the full input action of $\operatorname{Sym}\mathscr{A}_{\mathcal{K}}$.

Fix $v \in \mathcal{V}_1$, and let $e \in H$ denote the identity. Let us compute the orbit of (v, e) under this generated input action. Under the isomorphism $\Phi : \operatorname{Sym}\mathscr{A}_{\mathcal{K}} \to \operatorname{Aut}(\mathcal{K}) \times H$, every input permutation has the form $P(w, a) = (\pi(w), ga)$ for some $(\pi, g) \in \operatorname{Aut}(\mathcal{K}) \times H$. Hence the orbit of (v, e) intersects $\mathcal{V}_2 \times H$ if and only if there exists $\pi \in \operatorname{Aut}(\mathcal{K})$ with $\pi(v) \in \mathcal{V}_2$. Since $\mathcal{G}_1$ and $\mathcal{G}_2$ are the connected components of $\mathcal{K}$, this occurs if and only if $\mathcal{G}_1 \cong \mathcal{G}_2$.

The orbit of a point under a permutation group given by generators can be computed in polynomial time (Seress, 2003). The construction above is also polynomial-time. Therefore Graph Isomorphism reduces in polynomial time to computing additional generators B satisfying $\langle \mathrm{A}, B\rangle = \operatorname{Sym}\mathscr{A}$, so this problem is GI-hard. □

Hardness makes us expect that deducing whether $\mathrm{A} = \operatorname{Sym}\mathscr{A}$ can necessitate the use of architecture-specific knowledge. Since general architectures are usually built from modular pieces, it is natural to ask when $\operatorname{Sym}\mathscr{A}$ can be exactly computed from subarchitectures $\operatorname{Sym}\mathscr{A}_{(C_i, C_j)}$, that is, when a converse to Proposition 1 holds. The following theorem offers such conditions. Let $\mathscr{A} = \mathscr{A}_{(C_{\ell-1}, O)} \circ \cdots \circ \mathscr{A}_{(I, C_1)}$ be a decomposition of $\mathscr{A}$ for which $\mathcal{F}_{\mathscr{A}} = \mathcal{F}_{\mathscr{A}(C_{\ell-1}, C_O)} \circ \cdots \circ \mathcal{F}_{\mathscr{A}_{(I, C_1)}}$.

Assume $\boldsymbol{\beta} = (\beta_v : G \curvearrowright Z_v)$ is an action on the spaces of $\mathscr{A}$ such that $\mathrm{Sym}\,\mathscr{A} \subset \beta_I(G) \times \beta_O(G)$. For the following result, Sym on the space Z_C is taken with respect to $J = \beta_C(G)$.

Theorem 19. *For each ℓ suppose (i) there exists $\mathcal{F}'_\ell \subset \mathcal{F}_{\mathscr{A}_{(C_{\ell-1},C_\ell)}} \cap \mathrm{Map}_G(Z_{C_{\ell-1}}, Z_{C_\ell})$ such that $\mathcal{F}'^{\top}_\ell$ is injective, and (ii) for $\ell > 1$, $fg \equiv gf$ for all $g \in G$ whenever $fgp = gfp$ for all $f \in \mathcal{F}_{\mathscr{A}_{(C_{\ell-1},C_\ell)}}, p \in \mathcal{F}'_{\ell-1} \circ \cdots \circ \mathcal{F}'_1$. Then $\mathrm{Sym}\,\mathscr{A} = \mathrm{Sym}\,\mathscr{A}_{(C_{L-1},O)} \circ \cdots \circ \mathrm{Sym}\,\mathscr{A}_{(I,C_1)}$.*

If the $fg - gf$ were linear, the first condition is equivalent to $\bigcap_{f\in\mathcal{F}'_\ell} \ker f = (0)$ and the second is $\mathrm{span}(\bigcup_{p\in\mathcal{F}'_{\ell-1}\circ\cdots\circ\mathcal{F}'_1} \mathrm{im}\, p) = Z_{C_{\ell-1}}$. Of course, linearity basically never holds, and the stated hypotheses are the generalizations of these natural notions to the nonlinear setting.

Proof. We note that conditions (i) and (ii) are complementary, with (i) being a monic condition and (ii) being an epic condition.

The inclusion $\supset$ is nearly immediate and is Proposition 1. For $\subset$, let $(\sigma, \tau) \in \mathrm{Sym}\,\mathscr{A}$. Explicitly,

$$\mathrm{Sym}\,\mathscr{A}_{(C_{L-1},O)} \circ \cdots \circ \mathrm{Sym}\,\mathscr{A}_{(I,C_1)} = \left\{ (\sigma,\tau) \in J \times K : \begin{array}{l} \text{there exist } \delta_0 = \sigma, \delta_1, \ldots, \delta_{L-1}, \delta_L = \tau \\ \text{such that } (\delta_{\ell-1}, \delta_\ell) \in \mathrm{Sym}\,\mathscr{A}_{(C_{\ell-1},C_\ell)} \text{ for all } \ell \end{array} \right\}.$$

So the goal is to find $g_0, \ldots, g_L \in G$ such that $\beta_I(g_0) = \sigma$, $\beta_O(g_L) = \tau$, and, for every ℓ, $f_\ell \beta_{C_{\ell-1}}(g_{\ell-1}) = \beta_{C_\ell}(g_\ell) f_\ell$ for all $f_\ell \in \mathcal{F}_{\mathscr{A}_{(C_{\ell-1},C_\ell)}}$.

We first claim that WLOG $g_L = g_0 =: g$. Indeed, if $r = f'_L \circ \cdots \circ f'_1$, $f'_\ell \in \mathcal{F}'_\ell$, then by $(\sigma, \tau) \in \mathrm{Sym}\,\mathscr{A}$ one has $\beta_O(g_L) r = r \beta_I(g_0)$ and by G-equivariance of $f'_\ell \in \mathcal{F}'_\ell$ one further has $\beta_O(g_L) r = r \beta_I(g_L)$. Injectivity (monicness) of the $\mathcal{F}'^{\top}_\ell$ then lets us "cancel r" from $r\beta_I(g_0) = r\beta_I(g_L)$ to conclude $\beta_I(g_0) = \beta_I(g_L)$. From here we split into cases based on ℓ.

Case $\ell = 1$. Suppose $\ell = 1$. Let $f_1 \in \mathcal{F}_{\mathscr{A}_{(I,C_1)}}$ and $q = f'_L \circ \cdots \circ f'_2$, where $f'_\ell \in \mathcal{F}'_\ell$. Since $qf_1 \in \mathcal{F}_{\mathscr{A}}$, we have $qf_1\beta_I(g) = \beta_O(g) q f_1 = q\beta_{C_1}(g) f_1$, where the second equality follows from G-equivariance of q. By successive injectivity of the $\mathcal{F}'^{\top}_\ell$, we may cancel q to obtain $f_1\beta_I(g) = \beta_{C_1}(g) f_1$ for all $f_1 \in \mathcal{F}_{\mathscr{A}_{(I,C_1)}}$. Hence $(\beta_I(g), \beta_{C_1}(g)) \in \mathrm{Sym}\,\mathscr{A}_{(I,C_1)}$.

Case $\ell > 1$. Now let $\ell > 1$. Fix $f_\ell \in \mathcal{F}_{\mathscr{A}_{(C_{\ell-1},C_\ell)}}$ and $p = f'_{\ell-1} \circ \cdots \circ f'_1$, where $f'_j \in \mathcal{F}'_j$, and let $q = f'_L \circ \cdots \circ f'_{\ell+1}$, where $f'_j \in \mathcal{F}'_j$. Since p and q are G-equivariant and $qf_\ell p \in \mathcal{F}_{\mathscr{A}}$, we have $qf_\ell\beta_{C_{\ell-1}}(g)p = qf_\ell p \beta_I(g) = \beta_O(g) q f_\ell p = q\beta_{C_\ell}(g) f_\ell p$. By successive injectivity of the $\mathcal{F}'^{\top}_j$ for $j > \ell$, we may cancel q to obtain $f_\ell\beta_{C_{\ell-1}}(g)p = \beta_{C_\ell}(g) f_\ell p$ for all $p \in \mathcal{F}'_{\ell-1} \circ \cdots \circ \mathcal{F}'_1$. For $\ell = L$, the same equality follows directly, since there is no q. By condition (ii), $f_\ell\beta_{C_{\ell-1}}(g) = \beta_{C_\ell}(g) f_\ell$ for all $f_\ell \in \mathcal{F}_{\mathscr{A}_{(C_{\ell-1},C_\ell)}}$. Hence $(\beta_{C_{\ell-1}}(g), \beta_{C_\ell}(g)) \in \mathrm{Sym}\,\mathscr{A}_{(C_{\ell-1},C_\ell)}$ for every $\ell > 1$.

Taking $\delta_\ell := \beta_{C_\ell}(g)$, we have $\delta_0 = \sigma$, $\delta_L = \tau$, and $(\delta_{\ell-1}, \delta_\ell) \in \mathrm{Sym}\,\mathscr{A}_{(C_{\ell-1},C_\ell)}$ for every ℓ. Hence (σ, τ) belongs to the relational composition, proving $\subset$. □

To exemplify Theorem 19, we apply it in the classical finite setting described in Subexample 1. In particular, we use it to compute $\mathrm{Sym}(A\text{-MLP})$ in closed-form, where A-MLP is an MLP each of whose layers is equivariant to a free action by a finite group H, with parameter-sharing determined by $A \subset H$ (cf. Subexample 1). Before applying Theorem 19, we first compute $\mathrm{Sym}\,\mathscr{A}$ in closed form for $\mathscr{A}$ a single equivariant layer. This is achieved by the following result.

Proposition 20. *Let $\mathscr{A} : \mathbb{R}^I \to \mathbb{R}^O$ be a linear A-MLP layer as in Section 3.2, $A \neq \emptyset$. Put $D_A := \langle AA^{-1} \rangle$, $k := [H : D_A]$. Then*

$$\mathrm{Sym}\,\mathscr{A} \cong D_A \wr S_k.$$

The conclusion continues to hold for a nonlinear layer $x \mapsto \mathrm{a}(Wx)$ provided $(r \in \mathbb{R}) \mapsto (t \mapsto \mathrm{a}(tr))$ is injective.

Here, the notation $\wr$ denotes the *wreath product* $D_A \wr S_k := D_A^k \rtimes S_k$ where S_k acts on D_A^k by permuting its factors. In the dense special case $A = H$, $D_A = H$ and $k = 1$ and so $\mathrm{Sym}\,\mathscr{A} \cong H$.

Remark 2. Proposition 20 may be viewed as generalization and strengthening of Theorem 3.3 in (Ravanbakhsh et al., 2017), which amounts to proving one direction, $\mathrm{A} \leq \mathrm{Aut}(\Sigma_A)$, in the special case where A is a symmetric generating set. (The authors claim their proof shows equality, but their

argument of the reverse direction appears to be flawed.) In our case, this direction already follows abstractly via Proposition 18; the content the proof below lies in the reverse inclusion.

Proof. Assume first the linear case. By Proposition 18, $\mathrm{Sym}\,\mathscr{A} = \mathrm{Aut}(\Sigma_A)$; it remains to compute the latter. We first claim $\mathrm{Aut}(\Sigma_A)$ can be identified with the pairs (s,t) of permutations of H that satisfy $sga = tga$ for all $g \in H$, $a \in A$. This object we can then identify with $D_A \wr S_k$.

To this end choose representatives n_p of the input H-orbits and m_q of the output H-orbits. Since the H-actions are free, every input and output index can be written uniquely as gn_p or gm_q. Recall that the sharing classes are $\Delta_{p,q,a} = \{(gm_q, gan_p) : g \in H\}$. Let $(\sigma_I, \sigma_O) \in \mathrm{Aut}(\Sigma_A)$. Since every triple (p,q,a) has its own color, σ_I and σ_O must preserve each input and output H-orbit individually, implying there are bijections $s_p, t_q : H \to H$ such that $\sigma_I(gn_p) = s_p(g)n_p$ and $\sigma_O(gm_q) = t_q(g)m_q$. Preservation of $\Delta_{p,q,a}$ means that $s_p(ga) = t_q(g)a$ for all $g \in H$, $a \in A$, p, and q. Fix $a_0 \in A$. Then $s_p(ga_0) = t_q(g)a_0$ for all p, q, so the left-hand side is independent of p and the right-hand side is independent of q. Since $g \mapsto ga_0$ is a bijection of H, all s_p are in fact the same bijection s, and all t_q are in fact the same bijection t. Thus $s(ga) = t(g)a$ for all $g \in H$ and $a \in A$.

Now let $a, b \in A$. We have $s(ga) = t(g)a$, and and similarly gab^{-1} gives $s(ga) = t(gab^{-1})b$. Therefore $t(gab^{-1}) = t(g)ab^{-1}$. Since $D_A = \langle AA^{-1}\rangle$, it follows that $t(gd) = t(g)d$ for all $g \in H$ and $d \in D_A$. Conversely, suppose $t : H \to H$ is a bijection satisfying $t(gd) = t(g)d$ for all $d \in D_A$. Fix $a_0 \in A$ and define $s(x) := t(xa_0^{-1})a_0$. Then for every $a \in A$, $s(ga) = t(gaa_0^{-1})a_0 = t(g)aa_0^{-1}a_0 = t(g)a$, since $aa_0^{-1} \in D_A$. Hence (s,t) preserves every sharing class. Thus $\mathrm{Aut}(\Sigma_A)$ is exactly the group of bijections $t : H \to H$ satisfying $t(gd) = t(g)d$ for all $d \in D_A$.

The D_A-orbits under right multiplication are the $k = [H : D_A]$ right cosets of D_A. Thus t induces an arbitrary permutation $\pi \in S_k$ of these cosets. Choose representatives $r_1, \ldots, r_k$. For each i, there is a unique $d_i \in D_A$ such that $t(r_i) = r_{\pi(i)}d_i$. Then D_A-equivariance forces $t(r_i d) = r_{\pi(i)}d_i d$ for all $d \in D_A$. Hence t is determined by an arbitrary permutation $\pi \in S_k$ together with independent choices $d_1, \ldots, d_k \in D_A$. Therefore $\mathrm{Aut}(\Sigma_A) \cong D_A^k \rtimes S_k = D_A \wr S_k$.

For the nonlinear layer, Proposition 18 still gives $\mathrm{Sym}\,\mathscr{A} = \mathrm{Aut}(\Sigma_A)$ under the injectivity condition on a, so the same computation applies. □

We note the injectivity condition on a in Propositions 18 and 20 is not stringent; it is satisfied e.g. by ReLU, sigmoid, $\tanh$, leaky ReLU, ELU, etc.

Proposition 21. *Let $\mathscr{A} = \mathscr{A}_{(C_{L-1},O)} \circ \cdots \circ \mathscr{A}_{(I,C_1)}$ be an A-*MLP *as in Section 3.2 with activation a, such that $(r \in \mathbb{R}) \mapsto (t \mapsto \mathrm{a}(tr))$ is injective. Put $D_A := \langle AA^{-1}\rangle$, $k := [H : D_A]$. Then*

$$\mathrm{Sym}\,\mathscr{A} \cong D_A \wr S_k.$$

Proof. Write $Z_{C_\ell} \cong (\mathbb{R}^H)^{m_\ell}$ for the space of the ℓth layer of $\mathscr{A}$. Let $G := S_H$ act on H in the obvious way, inducing a permutation representation $\gamma : G \to \mathrm{GL}(\mathbb{R}^H)$ and in turn product actions $(\gamma_\ell : G \curvearrowright Z_{C_\ell})$. Fix $a_* \in A$ and let $P_{a_*} : \mathbb{R}^H \to \mathbb{R}^H$ be the permutation matrix induced by right-multiplication by a_*. Define representations $\rho_\ell : G \to \mathrm{GL}(\mathbb{R}^H)$ recursively by $\rho_0 := \gamma$ and $\rho_\ell(g) := P_{a_*}\rho_{\ell-1}(g)P_{a_*}^{-1}$, inducing in turn product actions $\boldsymbol{\beta} = (\beta_\ell : G \curvearrowright Z_{C_\ell})_{\ell=0}^{L}$.

We claim $\mathscr{A}$ and $\boldsymbol{\beta}$ satisfies the hypotheses of Proposition 19; combined with Proposition 20 this proves the result. Showing the hypotheses amounts to checking (a) $\mathrm{Sym}\,\mathscr{A} \le \beta_I(G) \times \beta_O(G)$ and that conditions (i) and (ii) hold. The checks are straightforward, though writing them out explicitly takes some bookkeeping.

Verification of (a). Let $(\sigma, \tau) \in \mathrm{Sym}\,\mathscr{A}$. For $p \in [m_0]$ and $q \in [m_L]$, choose indices $p = p_0, p_1, \ldots, p_L = q$ and consider the realizations $F_{p,q,\boldsymbol{t}} = f_{L,t_L} \circ \cdots \circ f_{1,t_1}$, where $f_{\ell,t_\ell}(x) = \mathrm{a}(t_\ell(E_{p_\ell p_{\ell-1}} \otimes P_{a_*})x)$. Each $F_{p,q,\boldsymbol{t}}$ depends only on the pth input factor, and among the output factors only the qth is nonconstant. Since $F_{p,q,\boldsymbol{t}}\sigma = \tau F_{p,q,\boldsymbol{t}}$ for every $p, q, \boldsymbol{t}$, varying one input factor at a time shows that σ preserves each input factor, and varying the qth output shows that τ preserves each output factor. Hence $\sigma = \mathrm{diag}(\pi_1, \ldots, \pi_{m_0})$ and $\tau = \mathrm{diag}(\eta_1, \ldots, \eta_{m_L})$ for some $\pi_p, \eta_q \in S_H$. On the qth output factor, $F_{p,q,\boldsymbol{t}}$ has the form $\psi_{\boldsymbol{t}}^{\times H} R\,\mathrm{pr}_p$, where $R := P_{a_*}^L$ and $\psi_{\boldsymbol{t}} : \mathbb{R} \to \mathbb{R}$ is the scalar map obtained by composing the scaled activations. Restricting $F_{p,q,\boldsymbol{t}}\sigma = \tau F_{p,q,\boldsymbol{t}}$ to this factor gives $\psi_{\boldsymbol{t}}^{\times H} R\pi_p = \eta_q \psi_{\boldsymbol{t}}^{\times H} R = \psi_{\boldsymbol{t}}^{\times H}\eta_q R$. As $\boldsymbol{t}$ varies, the family

$(\psi_{\boldsymbol{t}})_{\boldsymbol{t}}$ separates points of $\mathbb{R}$: if $r \neq s$, at each successive layer the injectivity of $u \mapsto (t \mapsto \mathrm{a}(tu))$ allows a choice of t_ℓ for which the two resulting values remain distinct. We may therefore cancel the family $(\psi_{\boldsymbol{t}}^{\times H})_{\boldsymbol{t}}$ to obtain $R\pi_p = \eta_q R$ for every p, q. Thus all π_p are equal to some $g \in S_H$, and all η_q are equal to RgR^{-1}. By construction of $\boldsymbol{\beta}$, $\sigma = \beta_I(g)$ and $\tau = \beta_O(g)$, proving $\mathrm{Sym}\,\mathscr{A} \leq \beta_I(G) \times \beta_O(G)$.

Verification of condition (i). For each ℓ, let $\mathcal{F}'_\ell = \{x \mapsto \mathrm{a}(t(E_{qp} \otimes P_{a_*})x) : t \in \mathbb{R},\ p \in [m_{\ell-1}],\ q \in [m_\ell]\}$. Since $\rho_\ell(g)P_{a_*} = P_{a_*}\rho_{\ell-1}(g)$, we have $\beta_\ell(g)(E_{qp} \otimes P_{a_*}) = (E_{qp} \otimes P_{a_*})\beta_{\ell-1}(g)$, and therefore every element of $\mathcal{F}'_\ell$ is G-equivariant. Now let $x \neq y \in Z_{C_{\ell-1}}$. Then $x_{p,h} \neq y_{p,h}$ for some $p \in [m_{\ell-1}]$ and $h \in H$. For any $q \in [m_\ell]$, the block $E_{qp} \otimes P_{a_*}$ sends these two values to the same output coordinate. Writing them as $r \neq s$, injectivity of $r \mapsto (t \mapsto \mathrm{a}(tr))$ gives $t \in \mathbb{R}$ such that $\mathrm{a}(tr) \neq \mathrm{a}(ts)$. Hence some $f \in \mathcal{F}'_\ell$ satisfies $f(x) \neq f(y)$, so $(\mathcal{F}'_\ell)^\top$ is injective.

Verification of condition (ii). Let $\ell > 1$ and suppose $fgp = gfp$ for every $f \in \mathcal{F}_{\mathscr{A}_{(C_{\ell-1}, C_\ell)}}$ and $p \in \mathcal{F}'_{\ell-1} \circ \cdots \circ \mathcal{F}'_1$. Write $f(x) = \mathrm{a}(Bx)$. Since tB is again a weight matrix for every $t \in \mathbb{R}$, the hypothesis gives $\mathrm{a}(tB\beta_{\ell-1}(g)p(x)) = \mathrm{a}(t\beta_\ell(g)Bp(x))$ for every t, p, x. By injectivity of $r \mapsto (t \mapsto \mathrm{a}(tr))$, $(B\beta_{\ell-1}(g) - \beta_\ell(g)B)p(x) = 0$ for every p and x. Moreover, the differences $p(x) - p(y)$ span $Z_{C_{\ell-1}}$. Indeed, fix a standard basis vector $e_{u,h}$. Choose indices $u_0, \ldots, u_{\ell-1} = u$ and $h_0, \ldots, h_{\ell-1} = h$ so that $E_{u_j u_{j-1}} \otimes P_{a_*}$ sends the coordinate (u_{j-1}, h_{j-1}) to (u_j, h_j). Choose $x, y \in Z_I$ differing only in coordinate (u_0, h_0). Inductively, if the two states after layer $j-1$ differ only in coordinate (u_{j-1}, h_{j-1}), with distinct values $r \neq s$, choose t_j such that $\mathrm{a}(t_j r) \neq \mathrm{a}(t_j s)$. Then after layer j they differ only in coordinate (u_j, h_j). Thus, for some $p \in \mathcal{F}'_{\ell-1} \circ \cdots \circ \mathcal{F}'_1$, we have $p(x) - p(y) = ce_{u,h}$ with $c \neq 0$. Since $e_{u,h}$ was arbitrary, these differences span $Z_{C_{\ell-1}}$. The resulting difference a nonzero multiple of $e_{u,h}$. Since $B\beta_{\ell-1}(g) - \beta_\ell(g)B$ is linear and vanishes on every $p(x)$, it vanishes on these differences and hence on all of $Z_{C_{\ell-1}}$. Therefore $B\beta_{\ell-1}(g) = \beta_\ell(g)B$, and since a commutes with coordinate permutations, $f\beta_{\ell-1}(g) = \beta_\ell(g)f$. Thus condition (ii) holds. □

B.3 Elaboration and Proofs from Section 5.1

We briefly elaborate upon the exposition given in Section 5.1, parsing apart the individual pieces of Proposition 16 from the main text to give their precise assumptions and provide some commentary. We taken together, the results below prove Proposition 16.

GEs for Group-Closure. Recall that group-closure creates symmetry slippage as over-equivariance $\underline{H}^{\alpha_Y}_{\alpha_X} \leq \underline{G}^{\alpha_Y}_{\alpha_X} \leq \mathrm{Sym}\,\mathscr{A}$, so that slipped transformations $s \in \mathrm{Sym}\,\mathscr{A} - \mathrm{A}$ take the form $s = \big(\alpha_X(g), \alpha_Y(g)\big)$ for $g \in G - H$. Roughly, the next result states that $c : X \to Z$ is a GE for s provided $s \notin \mathrm{Sym}\,c$.

Proposition 22. *Let $S \subset \mathrm{Sym}\,\mathscr{A} - \mathrm{A}$ be a collection of transformations slipped due to group-closure. Let $c : X \to Z$ and G act on Z. Write $\mathscr{A}' := \mathscr{A} \oplus_{C_2} c$. If $\mathscr{A}'^{\top}_{(C_2,O)}$ is injective and G-equivariant and c is such that $c\alpha_X(g) \neq gc$ for every $s_g = (\alpha_X(g), \alpha_Y(g)) \in S$, then c is a GE for S.*

Proof. Fix $s_g = \big(\alpha_X(g), \alpha_Y(g)\big) \in S$. Since $c\alpha_X(g) \neq gc$, there exists $x \in X$ such that $c(\alpha_X(g)x) \neq gc(x)$. Since b is G-equivariant, $b(\alpha_X(g)x) = gb(x)$, and hence

$$\big(b(\alpha_X(g)x), c(\alpha_X(g)x)\big) \neq g\big(b(x), c(x)\big).$$

By injectivity of $\mathscr{A}'^{\top}_{(C_2,O)}$, there therefore exists $e' \in \mathcal{F}_{\mathscr{A}'_{(C_2,O)}}$ such that

$$e'\big(b(\alpha_X(g)x), c(\alpha_X(g)x)\big) \neq e'\big(g(b(x), c(x))\big).$$

Since $\mathscr{A}'^{\top}_{(C_2,O)}$ is G-equivariant, the right-hand side equals $\alpha_Y(g)e'\big(b(x), c(x)\big)$. Thus, $f' := e' \circ (b, c) \in \mathcal{F}_{\mathscr{A}'}$,

$$f'\alpha_X(g)(x) \neq \alpha_Y(g)f'(x).$$

Hence $s_g \notin \mathrm{Sym}\,\mathscr{A}'$. Since $s_g \in S$ was arbitrary, $\mathscr{A}'$ grips every $s \in S$, and therefore c is a GE for S. □

GEs for Over-Quotienting. Recall that slipped transformations $s \in \mathrm{Sym}\,\mathscr{A} - \mathrm{A}$ created by over-quotienting take the form $s = (\sigma'\sigma^{-1}, I_Y)$ for some $\sigma, \sigma' \in J$. Roughly, the following result states that $c : X \to Z$ is a GE for s provided $c\sigma' \neq c\sigma$.

Proposition 23. *Let $S \subset \mathrm{Sym}\,\mathscr{A} - \mathrm{A}$ be a collection of transformations slipped due to over-quotienting, let $c : X \to Z$, and write $\mathscr{A}' := \mathscr{A} \oplus_{C_2} c$. If $\mathscr{A}'^{\top}_{(C_2,O)}$ is injective[4] and c is such that $c\sigma' \neq c\sigma$ for all $s = (\sigma'\sigma^{-1}, I_Y) \in S$, then c is a GE for S.*

Proof. Fix $s = (\sigma'\sigma^{-1}, I_Y) \in S$. Since $c\sigma' \neq c\sigma$, there exists $x \in X$ such that $c\sigma'(x) \neq c\sigma(x)$. By over-quotienting, $a\sigma(x) \sim a\sigma'(x)$, and since Φ is constant on $\sim$-classes, $b\sigma(x) = \Phi a\sigma(x) = \Phi a\sigma'(x) = b\sigma'(x)$. Hence $\big(b\sigma'(x), c\sigma'(x)\big) \neq \big(b\sigma(x), c\sigma(x)\big)$. By injectivity of $\mathscr{A}'^{\top}_{(C_2,O)}$, there exists $e' \in \mathcal{F}_{\mathscr{A}'_{(C_2,O)}}$ separating these two representations. Thus, for $f' := e' \circ (b, c) \in \mathcal{F}_{\mathscr{A}'}$, $f'\sigma'(x) \neq f'\sigma(x)$ so $f'\sigma' \neq f'\sigma$. Applying σ^{-1}, we get $f'\sigma'\sigma^{-1} \neq f'$. Therefore $(\sigma'\sigma^{-1}, I_Y) \notin \mathrm{Sym}\,\mathscr{A}'$. Since $s \in S$ was arbitrary, $\mathscr{A}'$ grips every $s \in S$, and hence c is a GE for S. □

GEs for Extension-Mistyping. Let $S \subset \mathrm{Sym}\,\mathscr{A} - \mathrm{A}$ be transformations slipped due to extension-mistyping; recall these take the form $s_g = \big(\alpha^{\mathrm{r}}_X(g), \alpha^{\mathrm{r}}_Y(g)\big) \in \mathrm{Sym}\,\mathscr{A} - \mathrm{A}$ for $g \in G$. Consider as before a modification $\mathscr{A}' = \mathscr{A} \oplus_C c$ of $\mathscr{A}$ for a cut C and some $c : X \to Z$, with realized functions $f' : X \xrightarrow{(b,c)} Z_C \times Z \xrightarrow{e'} Y$. Roughly, the following says $c : X \to Z$ is a GE if there exist $\kappa^{\mathrm{p}}, \kappa^{\mathrm{r}} : G \curvearrowright Z$, "compatible" with A and $\mathscr{A}$ respectively, such that $c(x) \notin \mathrm{Eq}(\kappa^{\mathrm{p}}(g), \kappa^{\mathrm{r}}(g))$ for some x.

Proposition 24. *Assume Z admits an action $\kappa^p : G \curvearrowright Z$ with respect to which c is equivariant. Assume b is equivariant with respect to α^r_X and ρ^r_C, and that Z admits an action $\kappa^r : G \curvearrowright Z$ such that $\mathscr{A}'^{\top}_{(C,O)}$ is equivariant with respect to (ρ^r_C, κ^r) and α^r_Y. Finally, suppose $\mathscr{A}'^{\top}_{(C,O)}$ is injective[5]. Then if $\mathrm{im}\, c \not\subset \mathrm{Eq}\,\big(\kappa^p(g), \kappa^r(g)\big)$[6] for all $s_g \in S$, c is a GE for S.*

Proof. Fix $s_g = \big(\alpha^{\mathrm{r}}_X(g), \alpha^{\mathrm{r}}_Y(g)\big) \in S$. Since $\mathrm{im}\, c \not\subset \mathrm{Eq}\big(\kappa^{\mathrm{p}}(g), \kappa^{\mathrm{r}}(g)\big)$, there exists $x \in X$ such that $\kappa^{\mathrm{p}}(g)c(x) \neq \kappa^{\mathrm{r}}(g)c(x)$. By equivariance of b and c,

$$\big(b\alpha^{\mathrm{r}}_X(g)(x), c\alpha^{\mathrm{r}}_X(g)(x)\big) = \big(\rho^{\mathrm{r}}_C(g)b(x), \kappa^{\mathrm{p}}(g)c(x)\big) \neq \big(\rho^{\mathrm{r}}_C(g)b(x), \kappa^{\mathrm{r}}(g)c(x)\big).$$

By injectivity of $\mathscr{A}'^{\top}_{(C,O)}$, there exists $e' \in \mathcal{F}_{\mathscr{A}'_{(C,O)}}$ separating these. Since $\mathscr{A}'^{\top}_{(C,O)}$ is equivariant, $e'\big(\rho^{\mathrm{r}}_C(g)b(x), \kappa^{\mathrm{r}}(g)c(x)\big) = \alpha^{\mathrm{r}}_Y(g)e'\big(b(x), c(x)\big)$. Thus, for $f' := e' \circ (b, c) \in \mathcal{F}_{\mathscr{A}'}$, $f'\alpha^{\mathrm{r}}_X(g)(x) \neq \alpha^{\mathrm{r}}_Y(g)f'(x)$, and hence $s_g \notin \mathrm{Sym}\,\mathscr{A}'$. Since $s_g \in S$ was arbitrary, $\mathscr{A}'$ grips every $s \in S$, and therefore c is a GE for S. □

The equalizer condition dissolves in an important special case.

Corollary 25. *Under the hypothesis of Proposition 24, suppose additionally that $[G : H] = 2$, that κ^r, κ^p are linear[7] with $\kappa^r|_H = \kappa^p|_H$, and that $\kappa^p(s) = -\kappa^r(s)$ for some $s \in G - H$. Then any nonzero $c : X \to Z$ is a GE.*

Proof. Fix $s_g \in S$. Since $\kappa^{\mathrm{r}}|_H = \kappa^{\mathrm{p}}|_H$, necessarily $g \in G - H$. As $[G : H] = 2$, we may write $g = sh$ for some $h \in H$. Then from $\kappa^{\mathrm{p}}|_H = \kappa^{\mathrm{r}}|_H$,

$$\kappa^{\mathrm{p}}(g) = \kappa^{\mathrm{p}}(s)\kappa^{\mathrm{p}}(h) = -\kappa^{\mathrm{r}}(s)\kappa^{\mathrm{r}}(h) = -\kappa^{\mathrm{r}}(g).$$

Since the actions are linear and $\kappa^{\mathrm{r}}(g)$ is an isomorphism, $\mathrm{Eq}\big(\kappa^{\mathrm{p}}(g), \kappa^{\mathrm{r}}(g)\big) = \{0\}$. Thus if c is nonzero, in particular $\mathrm{im}\, c \not\subset \{0\}$, and therefore $\mathrm{im}\, c \not\subset \mathrm{Eq}\big(\kappa^{\mathrm{p}}(g), \kappa^{\mathrm{r}}(g)\big)$. Since this holds for every $s_g \in S$, Proposition 24 implies that c is a GE for S. □

[4] More generally, it suffices to have injectivity on the pairs of representations witnessing $c\sigma' \neq c\sigma$.

[5] More generally, it suffices to have injectivity on the pairs of representations witnessing $\kappa^{\mathrm{p}}(g)c \neq \kappa^{\mathrm{r}}(g)c$.

[6] Recall that the *equalizer* of functions $f, g : X \to Y$ is $\mathrm{Eq}(f, g) = \{x \in X : f(x) = g(x)\} \subset X$.

[7] Strictly speaking, we should also assume Z is over a field of characteristic not 2.

C CASE STUDIES

C.1 GROUP-CLOSURE CASE STUDIES

C.1.1 TIME-ORIENTATION IN LORENTZ-EQUIVARIANT ML.

Many physical tasks are equivariant to subgroups of the *Lorentz group* $\mathrm{O}(1,d)$. The Lorentz group $\mathrm{O}(1,d)$ is a generalization of $\mathrm{O}(d)$, the group of linear automorphisms of Euclidean space, to linear automorphisms of *Minkowski spacetime* $V = (\mathbb{R}^{1,d}, \langle _, _ \rangle)$. Here, $\mathbb{R}^{1,d} := \underbrace{\mathbb{R}}_{\text{time } t} \times \underbrace{\mathbb{R}^d}_{\text{space } \boldsymbol{x}}$ and $\langle \overbrace{(t,\boldsymbol{x})}^{=:\boldsymbol{v}}, \overbrace{(s,\boldsymbol{y})}^{=:\boldsymbol{w}} \rangle = ts - \boldsymbol{x}\cdot\boldsymbol{y} = v^0 w^0 - \sum_{i=1}^d v^i w^i$. Like $\mathrm{O}(d)$, $\mathrm{O}(1,d)$ has a preferred subgroup of $\det$-1 transformations, called the *proper Lorentz group* and denoted $G := \mathrm{SO}(1,d)$. Unlike $\mathrm{O}(d)$, $\mathrm{O}(1,d)$ bifurcates into further components based on whether time-orientation is preserved; the subgroup of transformations preserving both space- and time-orientation is called the *proper orthochronous Lorentz group* and denoted $H := \mathrm{SO}^+(1,d)$. Explicitly,

$$H := \mathrm{SO}^+(1,d) = \{g \in \mathrm{SO}(1,d) : gC^+ = C^+\},$$

where $C^+ := \{v \in V : \langle v, v\rangle > 0 \text{ and } v^0 > 0\}$ is called the *future cone* of Minkowski space V; equivalently, $\mathrm{SO}^+(1,d)$ is the identity component of G. Further unlike $\mathrm{O}(d)$, $\mathrm{O}(1,d)$ is noncompact and so many "classical" equivariant learning techniques such as group-averaging don't easily apply. This motivates the implicit or explicit use of techniques from algebraic invariant theory which more readily extend (Blum-Smith & Villar, 2023; Villar et al., 2021). Because such techniques build exclusively polynomial features, and $\mathrm{SO}^+(1,d)$ is Zariski-dense in $\mathrm{SO}(1,d)$, this puts us in the setting of Proposition 3 by way of Example 2. This symmetry slippage occurs in each instrinsic Lorentz-equivariant architecture of which the authors are aware (Gong et al., 2022; Bogatskiy et al., 2024; 2020; Spinner et al., 2024).

Importantly, we do not suspect this symmetry slippage to injure expressivity within the published experiments for those architectures, as said experiments pertain to particle physics tasks whose inputs are all in $\overline{C^+}$. That said, any application involving displacements in V (e.g. convolution) or more general directed relations (e.g. spacetime embeddings of directed graphs (Sáez de Ocáriz Borde et al., 2025)) may present an expressivity barrier. To demonstrate this, we compare the four architectures with and without GEs on a standard neural PDE task. The goal of neural PDE (Li et al., 2021; Kovachki et al., 2023) is to learn the solution operator $S : s \mapsto u$ of a PDE $Lu = s$. When L, s, and the boundary conditions are nice enough, S is determined by convolution with a function h (called the *fundamental solution*), so that

$$u = S(s) = s * h.$$

For constant-coefficient hyperbolic PDEs in Minkowski space (e.g. the wave equation), there is a canonical h satisfying $\operatorname{supp} h \subset \overline{C^+}$ called the retarded fundamental solution. Our experiment focuses on learning the retarded solution to the wave equation $\Box u = s$ in $2+1$ dimensions. In this setting, h admits a closed form given by $h(\Delta) = \frac{1}{2\pi} \frac{\mathbf{1}_{\{\Delta^0 > 0\}} \mathbf{1}_{\{\langle \Delta, \Delta\rangle > 0\}}}{\sqrt{\langle \Delta, \Delta\rangle}}$. Following standard discretization-invariant neural-operator setups (Li et al., 2021; Kovachki et al., 2023), we sample irregular spacetime point clouds $P = (v_1, \ldots, v_n) \in V^n$ and scalar source fields $s = (s_1, \ldots, s_n)$, and approximate the solution operator per the discrete convolution $\sum_j w_j h^\theta(v_i - v_j) s_j$. With $X := V^n \times \mathbb{R}^n$, $Y = \mathbb{R}^n$ a published model $\mathscr{M}$ gives rise to an architecture $\mathscr{A}_{\mathscr{M}} : X \rightsquigarrow Y$ parameterizing functions

$$f^\theta : (P, s) \overset{b}{\mapsto} (\{\Delta_{ij}\}, s) \mapsto \left\{h^\theta_{\mathscr{M}}(\Delta_{ij})\right\} \mapsto \hat{u},$$

where $\Delta_{ij} = v_i - v_j$, $\hat{u}_i = \sum_{j \neq i} w_j f^\theta(\Delta_{ij}) s_j$, and $h^\theta_{\mathscr{M}}$ is a convolutional kernel parameterized by $\mathscr{M}$.

For $Z = \{-1, 0, 1\}^{n \times n}$, we define a GE $c : X \to Z$ given by $c\big((v_1, \ldots, v_n), s\big)_{ij} := \mathbf{1}_{\langle v_i - v_j, v_i - v_j\rangle > 0} \operatorname{sign}(v_i^0 - v_j^0)$. Since $g \in G$ reverse time-orientation iff $g \notin H$, we have $c(g(v_1, \ldots, v_n), s) = c(v_1, \ldots, v_n, s)$ if $g \in H$ and $c(g(v_1, \ldots, v_n)) = -c(v_1, \ldots, v_n, s)$ if $g \in G - H$, so c is indeed a GE. Note that c is clearly non-polynomial, as Example 2 would require. We append c at the cut C of $\mathscr{A}$ preceding the application of $h^\theta_{\mathscr{M}}$, so that $Z_C = V^{\{(i,j) : i \neq j\}} \times \mathbb{R}^n$ and $\mathscr{A}'_{\mathscr{M}} = \mathscr{A}_{\mathscr{M}} \oplus_C c$ is defined by

Architecture	ID NMSE — Raw	ID NMSE — +GE	OOD NMSE — Raw	OOD NMSE — +GE
LorentzNet (Gong et al., 2022)	0.6399 ± 0.0240	0.0688 ± 0.0253	0.5886 ± 0.0220	0.0576 ± 0.0217
PELICAN (Bogatskiy et al., 2024)	0.5154 ± 0.0013	0.0459 ± 0.0013	0.4784 ± 0.0011	0.0343 ± 0.0009
LGN (Bogatskiy et al., 2020)	0.5401 ± 0.0170	0.0335 ± 0.0031	1.1434 ± 0.8805	0.1271 ± 0.0879
L-GATr (Spinner et al., 2024)	0.5129 ± 0.0275	0.0545 ± 0.0118	0.4847 ± 0.0348	0.0542 ± 0.0129

Table 2

$$X \xrightarrow{(b,c)} Z_C \times Z \xrightarrow{e^{\theta'}} Y \tag{25}$$

$$x = (P, s) \xmapsto{(b,c)} \big((\Delta_{ij}, s)_{i\neq j}, (c_{ij})_{i\neq j}\big) \mapsto \{h^{\theta'}_{\mathscr{M}}(\Delta_{ij}; c_{ij})\}_{i\neq j} \to \hat{u}, \tag{26}$$

where the notation $h^{\theta'}_{\mathscr{M}}(\Delta_{ij}; c_{ij})$ indicates that c_{ij} has been appended as an input feature of $\mathscr{M}$.

On this task we evaluate four architectures: LorentzNet (Gong et al., 2022), PELICAN (Bogatskiy et al., 2024), LGN (Bogatskiy et al., 2020), and L-GATr (Spinner et al., 2024)[8]. In practice, we encode the displacement Δ_{ij} as the tuple $\left(\frac{1}{2}\Delta_{ij}, -\frac{1}{2}\Delta_{ij}\right)$ when inputting into these architectures, as they can become trivial when operating on a single object. For each architecture, we compare performance when an arbitrary scalar feature z has been appended $h^{\theta'}_{\mathscr{M}}(\Delta_{ij}; z)$ versus when the GE c has been appended. The results are shown in Table 2.

C.1.2 Signs in Units-Equivariant ML

Most numerical quantities in scientific learning problems have units on which the underlying laws of physics do not depend. Firsthand experience of science's "units-equivariance" is provided by the *dimensional analysis* practiced in introductory science classes. Dimensionless representations and related data-driven dimensional methods have been used for discovery of dimensional structure Constantine et al. (2017); Evangelou et al. (2021), dynamical-system identification Bakarji et al. (2022), and symbolic regression Villar et al. (2023), among other machine-learning settings Rudolph (1998); Frisone & Misiti (2019).

To describe this symmetry, suppose an input $x = (x_1, \ldots, x_d) \in X$ consists of d-many measured quantities x_i, expressed in k "base units" according to associated vectors $u_i \in \mathbb{Z}^k$, where the jth entry of u_i is the exponent of the jth base unit in x_i. For instance, a classical dynamics setup might have $k = 3$ base units (m, s, kg); in this case, a velocity x_i would have $u_i = (1, -1, 0)$, while a force x_ℓ would have $u_\ell = (1, -2, 1)$. A change of base unit is specified by positive "conversion factors" $g = (g_1, \ldots, g_k) \in H := (\mathbb{R}^+)^k$ and acts on inputs by

$$gx := \left(g_1^{-u_{11}} \ldots g_k^{-u_{1k}} x_1, \ldots, g_1^{-u_{d1}} \ldots g_k^{-u_{dk}} x_d\right)$$

and on an output y with associated vector $v \in \mathbb{Z}^k$ by $gy = g_1^{-v_1} \ldots g_k^{-v_k} y$. Units consistency then amounts to equivariance with respect to this H-action. If y is dimensionless (e.g. in a classification task), then $v = 0$ and the goal is H-invariance.

In the setting of (Villar et al., 2023), a *Buckingham* Π*-group* is a Laurent monomial $\Pi(x) = \prod_{i=1}^d x_i^{a_i}$ that is dimensionless, i.e., that has $Ua = \sum_i a_i u_i = 0$. The *Buckingham*-Π *Theorem* (Buckingham, 1914) roughly states that if a physical relationship $F(x_1, \ldots, x_d) = 0$ holds independently of the choice of units then there exists a Ψ and Buckingham Π-groups $\Pi_1(x), \ldots, \Pi_{d-r}(x)$, $r = \operatorname{rank} U$, such that $F(x) = 0$ iff $\Psi(\Pi_1(x), \ldots, \Pi_{d-r}(x)) = 0$. Leveraging this principle, (Villar et al., 2023) perform unconstrained learning after applying the dimensionless feature map $x \mapsto \big(\Pi_1(x), \ldots, \Pi_{d-r}(x)\big)$ for Π_a obtained by solving the (integer) linear system $Ua = 0$.

Let $\mathcal{C} = K(X)$, and write $G := (\mathbb{R}^\times)^k$. G acts on X, Y by evident extension of the H-actions. Conceptually, the G-action permits nonphysical sign reversals of base units whereas H consists only

[8] L-GATr (Spinner et al., 2024) also uses fixed reference multivectors such as $e_0 = (1, 0, 0, 0)$, for supplying time-orientation externally. Unlike a GE such as our c, however, a fixed e_0 need not preserve the full prescribed $\mathrm{SO}^+(1, d)$ symmetry.

Architecture	Accuracy
Baseline	$58.44 \pm 1.30\%$
Baseline +GE	$82.04 \pm 0.56\%$
GE only	$58.53 \pm 0.00\%$
Baseline + Shuffled GE	$58.25 \pm 1.09\%$

Table 3

of legitimate, positive changes of base units. Now, $\overline{H}^{\text{Zariski}} = G$, and the Π-features used by (Villar et al., 2023) are Laurent monomials, hence regular functions on $(\mathbb{R}^\times)^d$. Therefore, Proposition 3 gives, by way of Example 2, symmetry slippage for any architecture built solely from H-equivariant features. While this symmetry slippage is innocuous in many settings, such as when the target only depends on magnitude, it can pose a problem in any application wherein the sign of a measured quantity matters to the target. For $Z = \{-1, 0, 1\}^d \subset \mathbb{R}^d$ and G acting trivially on Z, we define a GE $c : X \to Z$ by $c(x)_i = \operatorname{sign}(x_i)$. Then c is H-equivariant but not G-equivariant. A general architecture $\mathscr{A}$ of the form in (Villar et al., 2023) realizes functions factoring as

$$X \xrightarrow{b=\phi_\Pi} Z_C = \mathbb{R}^{d-r} \xrightarrow{e^\theta} Y.$$

Appending c at C, $\mathscr{A}' = \mathscr{A} \oplus_C c$ is then

$$X \xrightarrow{(b,c)} Z_C \times Z \xrightarrow{e^{\theta'}} Y \tag{27}$$

$$x \xmapsto{(b,c)} \big((\Pi_1(x), \ldots, \Pi_{d-r}(x)), c(x)\big) \xrightarrow{e^{\theta'}} \hat{y}. \tag{28}$$

For concreteness, we assume $\mathscr{A}_{(C,O)}$ is an MLP, as in the experiments of (Villar et al., 2023) and our own, in which case $e^{\theta'}$ is defined as in the MLP example in Section 5.1 and in which case we know $\mathscr{A}'^{\top}_{(C,O)}$ is injective. We conclude from Proposition 22 that c is indeed a GE for $\mathscr{A}$.

To demonstrate the repair, we compare architectures with and without GEs on a natural time-series forecasting task from the Mauna Loa record of atmospheric CO_2 concentration (Keeling et al., 1976). At each time t, the input $h_t = (h_{t,0}, \ldots, h_{t,11}) \in \mathbb{R}^{12} = X$ consists of the preceding twelve weekly changes in CO_2 concentration, and the target $y_t = \mathbf{1}_{q_{t+4} > q_t} \in Y = \{0, 1\}$ records whether the concentration q four weeks later exceeds its current value. All input coordinates have the same unit $k = 1$, so $H = \mathbb{R}^+$ acts by $h_t \mapsto a h_t$ and trivially on Y. The task is H-invariant, since a positive change of concentration units preserves whether future concentration is higher. It is not $G = \mathbb{R}^\times$-invariant, since $h_t \mapsto -h_t$ reverses increases and decreases and so need not preserve the target.

We have $U = (1, \ldots, 1) \in \mathbb{Z}^{1\times 12}$, with $\ker U$ having integer basis $a_j = e_j - e_0$. This gives $\Pi_j(h) = \frac{h_j}{h_0}$, $j = 1, \ldots, 11$. Taking e^θ to be parameterized by a width-32 ReLU MLP with two hidden layers, we append GEs as above. We also evaluate against two controls given given by the GE alone and augmentation by a split-wise shuffled GE. Results are shown in Table 3.

Remark 3. The same group-closure occurs in weight-space learning for ReLU networks (Neyshabur et al., 2015; Meng et al., 2019). For a network with k independently rescalable hidden neurons, positive homogeneity of ReLU entails the prescribed law $\mathrm{A} = \underline{H}^I_\alpha$ for $H := (\mathbb{R}_+)^k$, but, as above, the Zariski-closure of H is $G := (\mathbb{R}^\times)^k$. Since path and basis-path coordinates are polynomial in the network weights (Meng et al., 2019; Gonon et al., 2025), H-invariance of these coordinates extends automatically to G-invariance, causing symmetry slippage by Example 2. This gives a group-closure interpretation of existing sign-based heuristics. For instance, Meng et al. (2019) supplement basis-path values with signs of free skeleton weights. Letting $c : X \to Z$ denote the retained sign information and letting G act trivially on Z, one has $c\alpha(h) = c$ for every $h \in H$, whereas $c\alpha(g) \neq c$ for each signed rescaling $g \in G - H$ detected by c. Thus, Proposition 16 implies that c is a GE for $S = G - H$.

C.1.3 Inner Automorphisms in Lie Algebra Learning

Lie-algebra-valued representations arise naturally in geometric learning problems such as dynamical-system modeling, registration, and homography-based vision (**?**). It is desirable for an

architecture $\mathscr{A}$ for learning on Lie algebras to be coordinate-independent in an appropriate sense. More precisely, the Lie algebra $(\mathfrak{k}, [_, _])$ of a Lie group K carries a linear action of K by conjugation, known as the *Adjoint representation*. $\mathrm{Ad} : K \to \mathrm{Aut}\,\mathfrak{k}$. Specifically, $g \xrightarrow{\mathrm{Ad}} dC_g|_e$, where $dC_g|_e$ is the differential of $C_g : h \mapsto ghg^{-1}$ at e. The Lie algebra $\mathfrak{g}$ also acts on itself via the *adjoint representation* ("small-a") $\mathrm{ad} : \mathfrak{k} \to \mathfrak{gl}(\mathfrak{k})$, $x \mapsto [x, _]$; for K connected, $H := \mathrm{Ad}(K) = \langle \exp(\mathrm{ad}_x) : x \in \mathfrak{k} \rangle =: \mathrm{Int}(\mathfrak{k})$ is the *inner automorphism group* of $\mathfrak{k}$. With these definitions in place, the desired notion of equivariance in Lie algebra learning is Ad-equivariance, or equivalently Int-equivariance.

To this end, the Lie neurons program (Lin et al., 2024; Kim et al., 2026) generalizes the well-known vector neurons approach (Deng et al., 2021), which may be viewed as developing $\mathrm{SO}(3)$-equivariant learning for $(\mathbb{R}^3, \times) \cong \mathfrak{so}(3)$, to more general matrix Lie algebras $\Sigma = (\mathfrak{k}, [_, _])$. Lie neuron-based architectures $\mathscr{A}$ are built from the function class $\mathcal{C}$ determined by linear channel mixing operations, $[_, _]$-operations, and $B(_, _)$-operations, where B denotes the reductive invariant bilinear form on $\mathfrak{k}$ (e.g. the Killing form κ when $\mathfrak{k}$ is semisimple). Such operations are equivariant not only to inner automorphisms $H = \mathrm{Int}(\Sigma)$, but to the full automorphism group $G = \mathrm{Aut}(\Sigma)$, leaving us in the setting of Example 1 with $\mathcal{C}$ closing H to G, and thus producing symmetry slippage whenever $\mathrm{Int}(\Sigma) \lneq \mathrm{Aut}(\Sigma)$.

While this symmetry slippage does not pose an expressivity problem if the data and task refer to a single H-orbit, it can become problematic on tasks whose targets transform nontrivially under G/H (called *outer automorphisms* of Σ). For concreteness, we focus on the case $\mathfrak{k} = \mathfrak{sl}_3$. In this case $G/H \cong \mathbb{Z}_2$, represented by the outer automorphism $\tau(x) = -x^\top$. For $x = (x_1, \ldots, x_d) \in \mathfrak{sl}_3^d =: X$, let $Z = \mathbb{R}^d$ with G acting trivially, and define $c : X \to Z$ by $c(x)_i = \mathrm{tr}(x_i^3) = 3 \det x_i$. Then c is H-equivariant, since trace is conjugation-invariant, hence inner automorphism-invariant, but not G-equivariant, since $c(\tau x) = -c(x)$. A Lie-Neuron architecture $\mathscr{A}$ realizes functions factoring as

$$X \xrightarrow{b^\theta} Z_C = \mathbb{R}^m \xrightarrow{e^\theta} Y$$

where b^θ is given by $X \xrightarrow{b_1} \mathfrak{sl}_3^m \xrightarrow{b_2} \mathbb{R}^m$ for b_1 the Lie Neuron equivariant backbone and $b_2(w)_j = \kappa(w_j, w_j)$ and e^θ is parameterized by a $\mathscr{A}_{(C,O)} = \mathrm{MLP}$. Appending c at C, $\mathscr{A}' = \mathscr{A} \oplus_C c$ gives

$$X \xrightarrow{(b^\theta, c)} Z_C \times Z \xrightarrow{e^{\theta'}} Y \tag{29}$$

$$x \mapsto \big(b^\theta(x), c(x)\big) \mapsto \hat{y}, \tag{30}$$

where $e^{\theta'}$ defined per the $\mathscr{A}_{(C,O)} = \mathrm{MLP}$ example in Section 5.1. Since $\mathscr{A}'_{(C,O)} = \mathrm{MLP}$, we know $\mathscr{A}'^{\top}_{(C,O)}$ is injective. We conclude from Proposition 22 that c is indeed a GE for $\mathscr{A}$.

To demonstrate the repair, we compare Lie-Neuron architectures with and without GEs on a natural continuum mechanics task: learning the strain-energy function $W = W(F)$ of a hyperelastic material. We use the classical experiments of Treloar (Treloar, 1944), which measure the response of vulcanized rubber under several homogeneous deformation modes, including uniaxial tension (UT) and equi-biaxial-tension (BT). These data have become a standard benchmark for data-driven constitutive modeling (Dammaß et al., 2025; Linka et al., 2021; Holthusen et al., 2026; Linden et al., 2023). For an incompressible deformation F, the Cauchy-Green tensor $F^\top F$ has $\det F^\top F = 1$, meaning $F^\top F \in \mathrm{SL}(3)$, and so the true strain $\varepsilon = \frac{1}{2} log(F^\top F) \in \mathfrak{sl}_3$. For an isotropic material, W is invariant under orthogonal conjugation, so that architectures equivariant to the restriction of $\mathrm{Ad} : \mathrm{SL}(3) \to \mathrm{Aut}\,\mathfrak{sl}_3$ to $\mathrm{SO}(3)$, such as Lie Neurons (LN) (Lin et al., 2024) or Reductive Lie Neureons (ReLN) (Kim et al., 2026), are appropriate for the task.

Following (Dammaß et al., 2025), we fit W to Treloar's measured UT and BT responses. The dataset contains 24 UT and 16 BT stress measurements. For each deformation F, we use the true strain $\varepsilon(F) = \frac{1}{2} \log(F^\top F)$ as input, and the network predicts $W(\varepsilon)$ by minimizing $\mathcal{L} = \mathrm{MSE}_{\mathrm{UT}} + 3\,\mathrm{MSE}_{\mathrm{BT}}$. We compare LN and ReLN with and without the GE $c(\varepsilon) = \mathrm{tr}(\varepsilon^3)$ on the task. As controls, we also add an identical model with an added feature, and one with the even feature $c(\varepsilon)^2$. Results are shown in Table 4.

Architecture	Arm	$\mathcal{L}_{\mathrm{UT+3BT}}$	UT R^2	BT R^2
LN (Lin et al., 2024)	Baseline	0.2817 ± 0.0844	0.9259 ± 0.0198	0.9874 ± 0.0087
LN	matched-zero	0.2776 ± 0.0453	0.9258 ± 0.0115	0.9900 ± 0.0028
LN	$+c^2$	0.2408 ± 0.0214	0.9352 ± 0.0055	0.9922 ± 0.0013
LN	$+c$ **(GE)**	0.0093 ± 0.0071	0.9978 ± 0.0015	0.9992 ± 0.0010
ReLN (Kim et al., 2026)	Baseline	0.2609 ± 0.0456	0.9301 ± 0.0116	0.9910 ± 0.0029
ReLN	matched-zero	0.2524 ± 0.0254	0.9323 ± 0.0065	0.9915 ± 0.0015
ReLN	$+c^2$	0.2443 ± 0.0134	0.9342 ± 0.0034	0.9922 ± 0.0009
ReLN	$+c$ **(GE)**	0.0100 ± 0.0033	0.9975 ± 0.0009	0.9992 ± 0.0004

Table 4

C.2 Over-Quotienting Case Studies

C.2.1 Roots in Glycan Learning

Glycans are carbohydrate molecules produced by organisms across the three domains of life (Eukarya, Bacteria, Archaea). A glycan is made of multiple simple sugars (monosaccharides) linked together by chemical bonds (called glycosidic linkages). Many glycans have a distinguished sugar, called the *reducing end*, that contains a free anomeric carbon. These glycans are commonly abstracted as trees whose nodes are sugars and whose edges are bonds, who are rooted at the reducing end, and whose nodes are colored by sugar type and edges by bond type. Define X_n to be the collection of rooted Glycan trees,

$$X_n := \{(T, r) : T \text{ is a Glycan tree on } [n], r \in [n]\}.$$

The group $\Gamma := H := S_n$ acts on X_n by simultaneously relabeling nodes and roots, $\alpha_X(\pi)(T, r) = (\pi T, \pi r)$. We may take $J := \mathrm{Bij}(X_n)$. Graph neural networks (GNNs) (Scarselli et al., 2009; Gilmer et al., 2017; Schlichtkrull et al., 2018; Vashishth et al., 2020) have recently been used extensively for glycan learning. Popular examples include GlycanML (Xu et al., 2025a), which employs a GNN backbone to aid in taxononimization tasks (e.g. Eukarya vs. Bacteria vs. Archaea), and GlycanAA (Xu et al., 2025b), which refines such approaches with atom-level structure and hierarchical message passing. The prescribed transformation law of graph-based approaches to glycan learning is $\mathrm{A} = \{\big(\alpha_X(\pi), I_Y\big) : \pi \in S_n\}$. In both aforementioned architectures $\mathscr{A}$, however, the GNN backbone operates on unrooted trees, in doing so over-quotienting the reducing end. To see this, take $Z_{C_1} = X_n$ and $Z_{C_2} = \{T : (T, r) \in X_n\}$, whence $\mathscr{A}$ factors as

$$X_n \xrightarrow{b=I_{X_n}} Z_{C_1} \xrightarrow{\Phi} Z_{C_2} \xrightarrow{d^\theta} Y$$

for $\Phi : (T, r) \mapsto T$ and d^θ parameterized by the GNN backbone. Define an equivalence relation $\sim$ on Z_{C_1} by $(T_1, r_1) \sim (T_2, r_2) \iff T_1 = T_2$ as colored graphs, and define $\sigma = I_X$, $\tau = I_Y$, and $\sigma' = \beta_\kappa$ for some $\kappa \neq e$, where $\beta_\kappa \in J$ is given by $\beta_\kappa(T, r) = (T, \kappa r)$. We see that $b\sigma(x) \sim b\sigma'(x)$ for all $x \in X$. Furthermore, Φ is by construction constant on $\sim$-classes and $(\sigma'\sigma^{-1}, I_Y) \notin \mathrm{A}$, and so the over-quotienting conditions are satisfied and symmetry slippage occurs per Proposition 6.

On many biological tasks, the reducing-end location carries relevant information. For an especially simple instance, in the dataset used by (Xu et al., 2025a) for their domain-taxonomization tasks (e.g. predicting whether a glycan was produced by Eukarya, Bacteria, Archaea, or Virus), $x_B = [\mathrm{Glc}(\beta 1-2)]\mathrm{Glc}(\beta 1-3)\mathrm{Glc}$ occurs in records labeled Bacteria while $x_E = \mathrm{Glc}(\beta 1-2)[\mathrm{Glc}(\beta 1-3)]\mathrm{Glc}$ occurs in records labeled Eukarya, despite the fact that over-quotienting implies every f^θ parameterized by $\mathscr{A}$ satifies $f^\theta(x_E) = f^\theta(x_B)$.

Let $Z = \mathbb{R}^{[n]}$ and define $c : X_n \to Z$ by $c(T, r)_v := \mathbf{1}_{\{v=r\}}$, so that $c(T, r)$ is the indicator vector for the tree's root node. Given a slipped transformation $s_\kappa = \big(\beta_\kappa, I_Y\big)$, $\kappa \neq e$, take $\sigma = I_X$ and $\sigma' = \beta_\kappa$.. Then $c\sigma'(T, r) = e_{\kappa r}$ and $e_r = c\sigma(T, r)$. Since the action by S_n is faithful, it follows that $c\sigma' \neq c\sigma$. Putting $C := C_2$, $\mathscr{A}$ realizes functions factoring as

$$X_n \xrightarrow{b} Z_C \xrightarrow{e^\theta} Y$$

where $b = \Phi \circ a$ and e^θ is the downstream GNN. (For GlycanAA, e^θ additionally contains a deterministic all-atom expansion before its hierarchical message-passing network.) Appending c at C,

Architecture	Macro-F1	+ GE Macro-F1
GlycanML, RGCN (Xu et al., 2025a)	0.6001 ± 0.0250	0.6358 ± 0.0193
GlycanML, CompGCN (Xu et al., 2025a)	0.6250 ± 0.0242	0.6710 ± 0.0212
GlycanAA (Xu et al., 2025b)	0.6409 ± 0.0420	0.6871 ± 0.0067

Table 5

$\mathscr{A}' = \mathscr{A} \oplus_C c$ gives

$$X_n \xrightarrow{(b,c)} Z_C \times Z \xrightarrow{e^{\theta'}} Y \tag{31}$$

$$(T, r) \mapsto (T, e_r) \mapsto \hat{y}, \tag{32}$$

where $e^{\theta'}$ is implemented by supplying c_v as an additional node feature to the GNN. To apply Proposition 23 and thus verify c is a GE, it remains to check that $\mathscr{A}'^{\top}_{(C,O)}$ is injective on the pairs of representations witnessing $c\sigma' \neq c\sigma$, that is, that $\mathscr{A}'^{\top}_{(C,O)}$ separates a pair of representations witnessing $c\sigma' \neq c\sigma$. Indeed, let $\sigma' = \beta_\kappa$ with $\kappa \neq e$,with $\kappa \neq e$, and choose $r \in [n]$ such that $\kappa r \neq r$. Choose T so that $\deg_T(r) \neq \deg_T(\kappa r)$. Then the representations at Z_C are $z = (T, e_r)$ and $z' = (T, e_{\kappa r})$. a usual sum-aggregation message-passing layer can propagate the root indicator to its neighbors and, after pooling, realize

$$\psi(T, e_r) = \sum_{u \in [n]} \sum_{v \in N_T(u)} (e_r)_v = \deg_T(r).$$

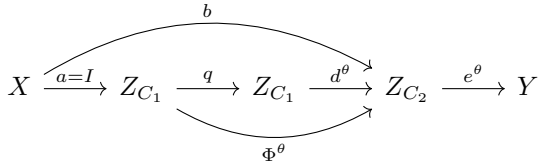
Hence $\psi(z) = \deg_T(r) \neq \deg_T(\kappa r) = \psi(z')$, so $\mathscr{A}'^{\top}_{(C,O)}$ separates this pair.

To demonstrate the repair, we compare GlycanML and GlycanAA with and without GEs on the proposed domain-level taxonomy prediction task benchmarked in both papers. Given a glycan, the task is to predict the domain (Eukarya, Bacteria, Archaea, or Virus) of the organism with which it is associated. We compare the released RGCN and CompGCN of GlycanML and the richer GlycanAA architecture, each with and without GEs, on their released 13209 domain-taxonomy records using 5-fold cross-validation with records having the same rootless graph placed in the same group. The results are shown in Table 5.

Note that this case study is over-quotienting, but not over-equivariance.

C.2.2 Normalization in ECG ML

Amplitude normalization is a common practice in ML for ECG analysis (Ma et al., 2021; Weimann & Conrad, 2024; Zeng et al., 2026; Hu et al., 2026; Li et al., 2025). Let $\mathbb{R}^{m \times T}$ denote the space of m-lead ECG signals. In the broadest sense, by a *normalization rule* we mean any function $q : X \subset \mathbb{R}^{m \times T} \to X$ satisfying $q(\lambda x) = q(x)$ for all $x \in X$, $\lambda > 0$. Equivalently, with $\Gamma := \mathbb{R}^\times \overset{\alpha}{\curvearrowright} X$ acting by scaling $\alpha_\lambda(x) = \lambda x$, a normalization rule q is an $G := \mathbb{R}^+$-invariant function on X. While many normalization rules exist, each imposes at least this symmetry. Let $H = (e) \leq G \leq \Gamma$. For certain tasks, a trivial transformation law $\mathrm{A} = \underline{H}^I_\alpha = \underline{(e)}^I_\alpha = \{(I_X, I_Y)\}$ is prescribed, and a normalizing architecture over-quotients it, incurring symmetry slippage. To see this, consider an architecture realizing functions $d^\theta : X \to Y$ and let $\mathscr{A}$ be the architecture obtained by imposing the normalization rule q, so that $\mathscr{A}$ realizes functions factoring as



where $Z_{C_1} = X$, Z_{C_2} is the original architecture's pooled feature space before its affine readout e^θ. Let $\sim_G$ be the orbit relation of $G \curvearrowright Z_{C_1}$. For $\sigma = I_X$ and $\sigma' = \alpha_\lambda$, $\lambda \neq 1$, we clearly have $a\sigma(x) \sim_G a\sigma'(x)$ for all $x \in X$. Furthermore, Φ is by construction constant on $\sim_G$-classes and

Architecture	Metric	Matched control	$+\log\rho$ GE
Weimann & Conrad (2024)	AUROC	0.8646 ± 0.0122	0.8792 ± 0.0159
	AUPRC	0.6839 ± 0.0283	0.7142 ± 0.0301
ECGFounder (Li et al., 2025)	AUROC	0.8832 ± 0.0039	0.9493 ± 0.0040
	AUPRC	0.6869 ± 0.0206	0.8310 ± 0.0275

Table 6

Architecture / task	Metric	Matched control	$+\log\rho$ GE
Weimann & Conrad (2024) native multilabel	AUROC	0.944762 ± 0.001881	0.945808 ± 0.002091
	AUPRC	0.608922 ± 0.005248	0.608099 ± 0.006468
ECGFounder PTB-XL	AUROC	0.947557 ± 0.008134	0.948464 ± 0.008039
	AUPRC	0.644723 ± 0.020212	0.646601 ± 0.020199

Table 7

$(\sigma'\sigma^{-1}, I_Y) \notin \mathrm{A}$, and so the over-quotienting conditions are satisfied and symmetry slippage occurs per Proposition 6, with $\{(\alpha_\lambda, I_Y) : \lambda > 0, \lambda \neq 1\} \subset \mathrm{Sym}\,\mathscr{A} - \mathrm{A}$. We examine two common ECG normalization rules. One rule, found e.g. in the recent ECG foundation model ECGFounder (Li et al., 2025), applies the z-score

$$q_{\text{glob}}(x) = \frac{x - \texttt{mean}(x)\mathbf{1}}{\texttt{stdev}(x)}.$$

A more refined rule, found e.g. in (Weimann & Conrad, 2024), standardizes leadwise thus:

$$q_{\text{lead}}(x)_{j,:} = \frac{x_{j,:} - \texttt{mean}(x_{j,:})}{\texttt{stdev}(x_{j,:})}.$$

While normalization is of course deliberate and often desirable, this over-quotienting can also erase task-relevant information. Two particularly direct examples may be found in the well-known PTB-XL benchmark (Strodthoff et al., 2021), which include e.g. the targets LVOLT (low QRS voltage) and VCLVH (voltage criteria for left-ventricular hypertrophy) that depend on amplitude.

Let $Z := \mathbb{R}$ and define $c : X \subset \mathbb{R}^{m\times T} \to Z$ by $c(x) = \log \frac{\|x\|_2}{\sqrt{mT}}$. Given a slipped transformation $s_\lambda = (\alpha_\lambda, I_Y)$, $\lambda > 0$, $\lambda \neq 1$, take $\sigma = I_X$ and $\sigma' = \alpha_\lambda$. Then $c\sigma'(x) = \log \frac{\|\lambda x\|_2}{\sqrt{mT}} = \log\lambda + c(x) \neq c\sigma(x)$. Putting $C := C_2$, $\mathscr{A}$ realizes functions factoring as $X \xrightarrow{b} Z_C \xrightarrow{e^\theta} Y$; appending c at C, $\mathscr{A}' = \mathscr{A} \oplus_C c$ gives

$$X \xrightarrow{(b,c)} Z_C \times Z \xrightarrow{e^{\theta'}} Y \tag{33}$$

$$x \mapsto \big(b(x), \log \frac{\|x\|_2}{\sqrt{mT}}\big) \mapsto \hat{y}, \tag{34}$$

where $e^{\theta'}$ is defined per the $\mathscr{A}_{(C,O)} = \mathrm{MLP}$ example in Section 5.1.

To demonstrate the repair, we compare the two architectures (Li et al., 2025) and (Weimann & Conrad, 2024) with and without GEs on the PTB-XL benchmark (Wagner et al., 2020; Strodthoff et al., 2021). We first take the PTB-XL multilabel classification task and restrict to the two labels LVOLT and VCLVH which depend explicitly on amplitude. We then compare the architectures in their published settings. For (Li et al., 2025), we follow its released PTB-XL setting, retaining the pretrained normalized architecture while training parameter-matched readouts e^θ with and without GEs. For Weimann and Conrad, we follow their released PhysioNet Challenge 2021 diagnostic setting (Weimann & Conrad, 2024). Results are shown in Tables 6 and 7.

C.2.3 Polarity in Time-Frequency Signal Processing

The Choi-Williams distribution (CWD) is a classical quadratic time-frequency representation widely applied to nonstationary signal analysis, including radar recognition (Zhang et al., 2016), machinery fault diagnosis (Lee et al., 2001), EEG Alazrai et al. (2017), and ECG (Kayıkçıoğlu et al.,

2017; Kayikcioglu et al., 2020) classification (Choi & Williams, 1989; Sejdić et al., 2009). Being quadratic, the CWD of a signal is invariant under sign-reversal, $\mathrm{CWD}(-x) = \mathrm{CWD}(x)$. Although such a property is often harmless, it is problematic for any task in which signs carry information. For example, a typical heartbeat is reflected in an ECG signal by a *P wave*, followed by a *QRS complex* (including the 'spike'), followed by a *T wave*. The constant stretch between the end of the QRS complex and start of the T wave is the *ST segment*. The time $R \subset \mathbb{R}$ between successive spikes is called an *R-R interval*. When the ST segment is not on the baseline, its sign relative to the baseline has physiological meaning, being used to help deduce what anomaly has occurred and where (Wagner et al., 2009). For simplicity, we focus on this CWD-in-ECG learning setting, following in particular the papers (Kayıkçıoğlu et al., 2017) and (Kayikcioglu et al., 2020).

Let X consist of pairs (r, i), where r is a single-lead ($m = 1$ in the previous ECG setting) ECG signal and i indexes an R-R interval $[R_i, R_{i+1}) \subset \mathbb{R}$. Let $\Gamma := \mathrm{Aff}(1, \mathbb{R}) \cong \mathbb{R}^\times \ltimes \mathbb{R}$ act on X as $\alpha_{\lambda,\delta}(r, i) = (\lambda r + \delta \mathbf{1}, i)$. Consider the subgroup $H := \mathrm{Aff}^+(1, \mathbb{R}) \cong \mathbb{R}^+ \ltimes \mathbb{R}$ of Γ. In the ECG setting, H models common gain and DC offset changes; Y is desired to be invariant under such nuisance transformations, giving the prescribed law

$$\mathrm{A} := \underline{H}^I_\alpha = \{(\alpha_{\lambda,\delta}, I_Y) : \lambda > 0, \delta \in \mathbb{R}\}.$$

Attaining $\mathrm{A} \leq \mathrm{Sym}\,\mathscr{A}$ can be done by standardizing each R-R interval as preprocessing. Indeed, for an interval $x_i(r) := r|_{[R_i, R_{i+1})}$, define $a(r, i) := \frac{x_i(r) - \mathtt{mean}\left(x_i(r)\right)\mathbf{1}}{\mathtt{stdev}\left(x_i(r)\right)}$, and observe that $a\big(\alpha_\lambda, \delta(r, i)\big) = a(r, i)$ for all $(r, i) \in X$. Put $Z_{C_1} := a(X)$. Following this preprocessing, the architecture $\mathscr{A}$ proposed in (Kayikcioglu et al., 2020) applies CWD feature extractor $\Phi : Z_{C_1} \to Z_{C_2} =: \mathbb{R}^{2\ell}$ given by $\Phi(u) = \big(E\big(\mathrm{CWD}(u)\big), E\big(\mathrm{CWD}(Du)\big)\big)$ where Du is the first-difference of u and E extracts ℓ-many pieces of summary information (e.g. min and max) from its input. Importantly, $\Phi(-u) = \Phi(u)$, invariance inherited from CWD. These features are fed into a downstream classification model parameterizing functions $e^\theta : Z_{C_2} \to Y$. Thus $\mathscr{A}$ realizes functions that factor as

$$X \xrightarrow{a} Z_{C_1} \xrightarrow{\Phi} Z_{C_2} \xrightarrow{e^\theta} Y.$$

Define an equivalence relation $\sim$ on Z_{C_1} by $u \sim v \iff v \in \{\pm u\}$. Note Φ is constant on $\sim$-classes because CWD is. Take $(\sigma, \tau) = (I_X, I_Y) \in \mathrm{A}$ and $\sigma' = \alpha_{-1,0}$. Then for every $(r, i) \in X$, $a\sigma(r, i) = a(r, i)$ and $a\sigma'(r, i) = a(-r, i) = -a(r, i)$, so that $a\sigma(r, i) \sim a\sigma'(r, i)$. Lastly, $(\sigma'\sigma^{-1}, I_Y) = (\alpha_{-1,0}, I_Y) \notin \mathrm{A}$ since $-1 \notin \mathbb{R}^+$. Hence the conditions of Definitoin 5 are satisfied and $\mathscr{A}$ over-quotients A, with Proposition 6 giving symmetry slippage as $(\alpha_{-1,0}, I_Y) \in \mathrm{Sym}\,\mathscr{A} - \mathrm{A}$.

We would like to define GEs repairing this symmetry slippage. Let $Z = \mathbb{R}$. Given $x = (r, i)$, define

$$s_i(r) := \frac{r(t_i^{\mathrm{ST}}) - \mathtt{mean}(r|_{B_i})}{\mathtt{stdev}\big(x_i(r)\big)}$$

where t_i^{ST} is the time at which the ST segment corresponding to heartbeat i's height was measured, and B_i is a short stretch of signal immediately before that heartbeat used to estimate the baseline. The idea is that $s_i(r)$ measures the ST height relative to the baseline, in units of the intervals standard deviation. We then define c as the comparison of this quantity to a reference value computed from beats near the beginning of r, $s_r^{\mathrm{ref}} := \mathtt{median}\{s_j(r) : j \text{ is a beat in the first 30s}\}$,

$$c(r, i) := s_i(r) - s_r^{\mathrm{ref}}.$$

Thus $c(r, i)$ is positive when the ST segment of beat i lies above the recording's reference level and negative when it lies below it. Given the slipped transformation $s = (\alpha_{-1,0}, I_Y)$, take $\sigma = I_X$ and $\sigma' = \alpha_{-1,0}$. Then, whenever $c(r, i) \neq 0$,

$$c\sigma'(r, i) = c(-r, i) = -c(r, i) \neq c(r, i) = c\sigma(r, i).$$

Putting $C := C_2$, $\mathscr{A}$ realizes functions factoring as

$$X \xrightarrow{b} Z_C \xrightarrow{e^\theta} Y.$$

where $b = \Phi \circ a$. Appending c at C, $\mathscr{A}' = \mathscr{A} \oplus_C c$ gives

$$X \xrightarrow{(b,c)} Z_C \times Z \xrightarrow{e^{\theta'}} Y, \tag{35}$$

$$(r, i) \mapsto \big(\Phi(a(r, i)), c(r, i)\big) \mapsto \hat{y}, \tag{36}$$

Architecture	AUROC
Baseline Logistic Regression (Kayikcioglu et al., 2020)	0.8505
Logistic Regression + GE	0.9671
Baseline 5-NN (Kayikcioglu et al., 2020)	0.8157
5-NN + GE	0.9427

Table 8

where $e^{\theta'}$ is just the classifier of $\mathscr{A}$ with c appended as a feature.

To demonstrate the repair, we evaluate the CWD representation on the natural task of distinguishing ST elevation from ST depression in the European ST-T Database (EDB) (Taddei et al., 1992). We compare the architecture $\mathscr{A}$ proposed in (Kayikcioglu et al., 2020) with and without GEs, with the non-GE version receiving instead the additional sign-invariant feature $|c|$. The cohort contains 68 subjects and 288 expert-annotated ST episodes. Following (Kayikcioglu et al., 2020), we train the same logistic and 5NN classifiers on both representations under leave-one-subject-out evaluation, taking 25 R-R intervals from each episode and averaging their predicted probabilities to obtain an episode-level prediction. Results are shown in Table 8.

C.3 Extension-Mistyping Case Studies

C.3.1 (Pseudo)scalar Mistyping in Neural Operators

The action on $\mathbb{R}^2$ of the dihedral group $G := D_4$ induces two natural actions on functions $\omega : \mathbb{R}^2 \to \mathbb{R}$. The first action is $g\omega(x) := \omega(g^{-1}x)$, in which case ω is called a *scalar (field)*. The second action is $g\omega(x) := \det(g)\omega(g^{-1}x)$, in which case ω is called a *pseudoscalar (field)*. The two actions, which we denote $\widetilde{\alpha}^{\mathrm{r}}, \widetilde{\alpha}^{\mathrm{p}} : G \curvearrowright \mathbb{R}^{\mathbb{R}^2}$, restrict to the same action $\alpha : C_4 =: H \curvearrowright \mathbb{R}^{\mathbb{R}^2}$, and so the distinction is not relevant when the only actions of interest are by rotations. It becomes relevant under reflections, however. For instance, two-dimensional vorticity (i.e., curl) is a pseudoscalar.

Helwig et al. (2023) introduce the *Group Equivariant Fourier Neural Operator (G-FNO)*, extending the Fourier neural operator of (Li et al., 2021) to impose equivariance to discrete rotations and reflections. They introduce two architectures, G-FNO-p4 and G-FNO-p4m. The first is $H = C_4$-equivariant, while the second extends the action to $G = D_4$. The architectures have the form

$$X \xrightarrow{P_G} Z_0 \xrightarrow{L_1} \cdots \xrightarrow{L_L} Z_L \xrightarrow{Q_G} Y$$

where the L_ℓ are group Fourier layers, $Z_\ell \cong (\mathbb{R}^{n_\ell} \otimes \mathbb{R}^K)^{\mathbb{R}^2}$ carries the representation λ_ℓ on functions $\mathbb{R}^2 \to \mathbb{R}^{n_\ell} \otimes \mathbb{R}^K$ induced by the tensor product of the trivial representation $K \curvearrowright \mathbb{R}^{n_\ell}$ and the regular representation on $\mathbb{R}^K$. Explicitly, $(\rho_\ell(k)z)(x) = (I_{n_\ell} \otimes \mathrm{Reg}_K(k))z(k^{-1}x)$. G-FNO-p4m, however, extends the action from α to $\widetilde{\alpha}^{\mathrm{r}}$ on vorticity rather than $\widetilde{\alpha}^{\mathrm{p}}$, thereby typing vorticity as an ordinary scalar rather than a pseudoscalar.

Let us verify this is extension-mistyping. Let $\mathscr{A}$ denote G-FNO-p4m, let $\lambda_\ell : G \curvearrowright Z_\ell$ denote the D_4-action on each Z_ℓ described above, and let $\boldsymbol{\rho} = (\rho_v : H \curvearrowright Z_v)_{v \in V}$ denote the H-actions obtained by restricting the G-actions realized by the architecture; in particular, $\rho_\ell = \lambda_\ell|_H$ on the Z_ℓ for $\ell = 1, \ldots, L$ and $\rho_I = \rho_O = \alpha$ on X and Y. Let $\widetilde{\boldsymbol{\rho}}^{\mathrm{p}}$ be the prescribed extension of $\boldsymbol{\rho}$ to $G = D_4$: on X and Y it is the pseudoscalar action $\widetilde{\alpha}^{\mathrm{p}}$, while on $Z_0, \ldots, Z_L$ it is λ_ℓ. Choose a reflection $s \in G - H$, and define $S_I = \widetilde{\alpha}^{\mathrm{r}}(s)$, $S_\ell = \lambda_\ell(s)$, and $S_O = \widetilde{\alpha}^{\mathrm{r}}(s)$. Since these operators arise from D_4-actions, they satisfy conditions (i) and (ii) of Proposition 9, and since each map in G-FNO-p4m is equivariant for these actions, condition (iii) is also satisfied. On the vorticity carriers, however, $\widetilde{\rho}^{\mathrm{p}}_I(s) = -\widetilde{\alpha}^{\mathrm{r}}(s) = -S_I$ and $\widetilde{\rho}^{\mathrm{p}}_O(s) = -\widetilde{\alpha}^{\mathrm{r}}(s) = -S_O$, so $\widetilde{\rho}^{\mathrm{p}}_v(s) \neq S_v$ for $v = I, O$. Since $H \triangleleft G$ and $[G : H] = 2$, Proposition 9 therefore gives an extension $\widetilde{\boldsymbol{\rho}}^{\mathrm{r}}$ of $\boldsymbol{\rho}$ for which G-FNO-p4m extension-mistypes $\widetilde{\boldsymbol{\rho}}^{\mathrm{p}}$.

To see this extension-mistyping creates symmetry slippage, take $\mathrm{A} := \underline{D_4}^{\widetilde{\alpha}^{\mathrm{p}}_Y}_{\widetilde{\alpha}^{\mathrm{p}}_X}$. In G-FNO-p4m, however, P_G is equivariant with respect to $\widetilde{\alpha}^{\mathrm{r}}_X$ and λ_0, each L_ℓ is equivariant with respect to $\lambda_{\ell-1}$ and λ_ℓ, and Q_G is equivariant with respect to λ_L and the $\widetilde{\alpha}^{\mathrm{r}}_Y$. Proposition 8 therefore gives $\underline{D_4}^{\widetilde{\alpha}^{\mathrm{r}}_Y}_{\widetilde{\alpha}^{\mathrm{r}}_X} \leq \mathrm{Sym}\,\mathscr{A}$. For every reflection $g \in D_4 - C_4$, one has $\widetilde{\alpha}^{\mathrm{p}}_X(g) = -\widetilde{\alpha}^{\mathrm{r}}_X(g)$ and $\widetilde{\alpha}^{\mathrm{p}}_Y(g) = -\widetilde{\alpha}^{\mathrm{r}}_Y(g)$, so the realized transformation $s_g := \big(\widetilde{\alpha}^{\mathrm{r}}_X(g), \widetilde{\alpha}^{\mathrm{r}}_Y(g)\big)$ is not contained in the prescribed law A.

Architecture	Rel. MSE (%)
G-FNO-p4m Baseline (Helwig et al., 2023)	2.363 ± 0.240
G-FNO-p4m + GE	1.701 ± 0.173
G-FNO-p4m Retyped	1.302 ± 0.130

Table 9

Thus $S := \{s_g : g \in D_4 - C_4\} \subset \operatorname{Sym} \mathscr{A} - \mathrm{A}$, and G-FNO-p4m exhibits symmetry slippage by extension-mistyping.

To repair, put $Z := (\mathbb{R}^G)^{\mathbb{R}^2}$ and define $c : X \to Z$ by $c(x)_g = \det(g)x$ for $g \in G$ (applied channelwise). Let $\kappa^{\mathrm{r}} : G \curvearrowright Z$ be the restriction of λ_0 to one channel of Z_0, and put $\kappa^{\mathrm{p}}(k) := \det(k)\kappa^{\mathrm{r}}(k)$. Then $c(\widetilde{\alpha}^{\mathrm{r}}_X(k)x) = \kappa^{\mathrm{p}}(k)c(x)$. Since $\det(h) = 1$ for $h \in H = C_4$, we have $\kappa^{\mathrm{p}}|_H = \kappa^{\mathrm{r}}|_H$, while for every reflection $s \in G - H$, $\kappa^{\mathrm{p}}(s) = -\kappa^{\mathrm{r}}(s)$. Putting $C := Z_0$, $\mathscr{A}$ realizes functions factoring as $X \xrightarrow{b} Z_C \xrightarrow{e^\theta} Y$, where $b = P_G$ and $e^\theta = Q_G \circ L_L \circ \cdots \circ L_1$. Appending c at C, $\mathscr{A}' = \mathscr{A} \oplus_C c$, gives

$$X \xrightarrow{(b,c)} Z_C \times Z \xrightarrow{e^{\theta'}} Y, \tag{37}$$

$$x \longmapsto \big(P_G x, c(x)\big) \longmapsto \hat{y}, \tag{38}$$

where $e^{\theta'}$ is obtained in the usual way (Section 5.1).

Since $c \not\equiv 0$, to apply Corollary 25 and thus verify that c is a GE it only remains check that $\mathscr{A}'^{\top}_{(C,O)}$ is injective on pairs of representations witnessing $\kappa^{\mathrm{p}}(s)c(x) \neq \kappa^{\mathrm{r}}(s)c(x)$. One example to see this is as follows. Choose any frequency ξ whose stabilizer under G is trivial, for example $\xi = (1, 2)$, and define ψ such that the only nonzero Fourier coefficient of ψ is ξ. Since the stabilizer of ξ is trivial, $\psi_- := \sum_{g \in G} \det(g) g\psi \neq 0$. A G-Fourier convolution followed by Q_G gives an e satisfying

$$e(0, \kappa^{\mathrm{p}}(s)c(x)) = -e(0, \kappa^{\mathrm{r}}(s)c(x)) \neq e(0, \kappa^{\mathrm{r}}(s)c(x)), \tag{39}$$

so the condition indeed holds.

To demonstrate the repair, we compare G-FNO-p4m with and without the GE on two vorticity-forecasting tasks. The first is the NS-SYM benchmark of (Helwig et al., 2023), governed by $\partial_t \omega + u \cdot \nabla \omega = \nu \Delta \omega + f$ with $\nu = 10^{-4}$ and $f = 0.1(\cos 4\pi x_1 + \cos 4\pi x_2)$; we refer the reader to (Helwig et al., 2023) for details. Results are shown in Table 9 We also experiment on an unforced decaying-turbulence benchmark (Cao et al., 2025). We use random peaked-spectrum initial conditions following the classical decaying-turbulence setup of McWilliams (1984) We generate the trajectories using a pseudospectral vorticity solver on the periodic domain $[0, 2\pi]^2$, following the modern `torch-cfd` implementation (Cao et al., 2025). We generate 768 examples at 64×64 resolution with peak wavenumber 4 and evolve them with viscosity $\nu = 10^{-3}$, time step $\Delta t = 10^{-3}$, and $T = 0.5$. The task is to approximate the finite-time solution operator $\omega(_, 0) \mapsto \omega(_, T)$. Given an initial vorticity field, the model predicts the full vorticity field after evolution to $T = 0.5$. For each sampled initial condition ω_0, we also include the sign-reversed initial condition $-\omega_0$ and evolve the two initial conditions separately. The idea is that the data contain pairs ω_0, $-\omega_0$, which we see causes symmetry slippage to become especially apparent empirically. Performance is measured by NMSE on held-out final-time vorticity fields. Results are shown in Table 10.

C.3.2 Tetrahedral Chirality in Molecular Graph Learning

Chiral molecules can occur in two mirror-image forms, called *enantiomers*, that have the same atoms/bonds but different handedness. A common source of chirality is a tetrahedral atom bonded to four distinct ligands, whose spatial arrangement can occur in two opposite configurations. While chemists conventionally describe these configurations using the Cahn-Ingold-Prelog (CIP) priority rules (Cahn et al., 1966), which assign the four ligands fixed priorities and label the resulting configuration R or S, the ligand ordering is in general arbitrary and $G := S_4$ acts by permuting it. Even

Architecture	NMSE
G-FNO-p4m Baseline (Helwig et al., 2023)	0.294442 ± 0.005787
G-FNO-p4m + GE	0.008005 ± 0.000336
G-FNO-p4m Retyped	0.005520 ± 0.000207

Table 10

permutations $H := A_4$ preserve handedness and odd permutations reverse it. Any representation of tetrahedral stereochemistry defined relative to an ordering of the four ligands should therefore be equivariant to the sign representation $\pi \mapsto \mathrm{sign}(\pi) \in \mathbb{R}$ of S_4 (Pattanaik et al., 2020).

For a concrete example, consider the architecture $\mathscr{A}$ given by QGeoGNN (Xu et al., 2023), a graph neural network for chromatographic retention-time prediction. QGeoGNN represents molecular chirality using an atomic *chiral tag* appended to ordinary atom features and applies a graph isomorphism network (GIN) (Xu et al., 2019). Encoding the two tetrahedral tag values at a stereocenter v by $\tau_v \in \{\pm 1\}$, the prescribed action is $\widetilde{\rho}^{\mathrm{p}}_v(\pi)\tau_v = \mathrm{sign}(\pi)\tau_v$: after permuting the four ligand positions by $\pi \in S_4$, the tag should flip iff π is odd in order to represent the same stereochemical configuration. QGeoGNN instead treats the tag as it does other node features, giving the realized action $\widetilde{\rho}^{\mathrm{r}}_v(\pi)\tau_v = \tau_v$. Thus $\widetilde{\rho}^{\mathrm{p}}_v|_{A_4} = \widetilde{\rho}^{\mathrm{r}}_v|_{A_4}$, while $\widetilde{\rho}^{\mathrm{p}}_v(s) = -\widetilde{\rho}^{\mathrm{r}}_v(s)$ for every $s \in S_4 - A_4$.

For every node map ϕ^θ_v in the GIN, $\phi^\theta_v \widetilde{\rho}^{\mathrm{r}}_{\{u\to v\}}(\pi) = \widetilde{\rho}^{\mathrm{r}}_v(\pi)\phi^\theta_v$ for $\pi \in S_4$, while $\widetilde{\rho}^{\mathrm{p}}_v(s) \neq \widetilde{\rho}^{\mathrm{r}}_v(s)$ on the chiral-tag carrier for $s \in S_4 - A_4$. Hence QGeoGNN extension-mistypes $\widetilde{\boldsymbol{\rho}}^{\mathrm{p}}$. Write $\widetilde{\rho}^{\mathrm{p}}_I, \widetilde{\rho}^{\mathrm{r}}_I : S_4 \curvearrowright X$ for the corresponding actions on molecular representations: both permute the four ligand positions by π, while $\widetilde{\rho}^{\mathrm{p}}_I$ sends $\tau_i \mapsto \mathrm{sign}(\pi)\tau_i$ and $\widetilde{\rho}^{\mathrm{r}}_I$ sends $\tau_i \mapsto \tau_i$. Taking $\mathrm{A} := \underline{S_4}^1_{\widetilde{\rho}^{\mathrm{p}}_I}$, every $f^\theta \in \mathcal{F}_{\mathscr{A}}$ satisfies $f^\theta \widetilde{\rho}^{\mathrm{r}}_I(\pi) = f^\theta$, so Proposition 8 gives $\underline{S_4}^1_{\widetilde{\rho}^{\mathrm{r}}_I} \leq \mathrm{Sym}\,\mathscr{A}$. Since $\widetilde{\rho}^{\mathrm{r}}_I(s) \neq \widetilde{\rho}^{\mathrm{p}}_I(s)$ for every $s \in S_4 - A_4$, $(\widetilde{\rho}^{\mathrm{r}}_I(s), 1) \notin \mathrm{A}$, and QGeoGNN exhibits symmetry slippage.

This symmetry slippage can pose a problem when enantiomers have different targets, as can occur in chromatographic enantioseparation. To repair, let $\chi_i(x) \in \{R, S, \varnothing\}$ denote the CIP configuration at atom i. Put $Z = (\mathbb{R}^4)^V$ and define $c : X \to Z$ by $c_i(x) = \big(\mathbf{1}_{\{\chi_i(x)=R\}}, \mathbf{1}_{\{\chi_i(x)=S\}}, \mathbf{1}_{\{\tau_i(x)=+1\}}, \mathbf{1}_{\{\tau_i(x)=-1\}}\big)$. Let P_{RS} exchange the first two coordinates of $\mathbb{R}^4$ and fix the latter two, and define $\kappa^{\mathrm{r}}, \kappa^{\mathrm{p}} : S_4 \curvearrowright \mathbb{R}^4$ by $\kappa^{\mathrm{r}}(\pi) = I_4$ and $\kappa^{\mathrm{p}}(\pi) = I_4$ for $\pi \in A_4$, while $\kappa^{\mathrm{p}}(\pi) = P_{RS}$ for $\pi \in S_4 - A_4$. It is readily checked that $\mathrm{im}\, c \not\subset \mathrm{Eq}(\kappa^{\mathrm{p}}(s), \kappa^{\mathrm{r}}(s))$.[9]

Let $a_i(x) \in \mathbb{R}^d$ denote the ordinary atom features at atom i immediately before the first GIN layer, so that the input to the GIN is $b(x) = ((a_i(x), \tau_i(x)))_{i\in V} \in Z_C$. Thus $\mathscr{A}$ factors as $X \xrightarrow{b} Z_C \xrightarrow{e^\theta} Y$. Appending c at this cut gives $\mathscr{A}' = \mathscr{A} \oplus_C c$ with $X \xrightarrow{(b,c)} Z_C \times Z \xrightarrow{e^{\theta'}} Y$ and $(b, c)(x) = ((a_i(x), \tau_i(x), c_i(x)))_{i\in V}$. Since $\mathrm{im}\, c \not\subset \mathrm{Eq}(\kappa^{\mathrm{p}}(s), \kappa^{\mathrm{r}}(s))$ for every $s \in S_4 - A_4$, together with the injectivity condition of Proposition 24, which we verified for this type of MPNN in Section C.2 this gives that c is a GE.

We demostate the repair on QGeoGNN's all-column retention-time benchmark (Xu et al., 2023) with a fixed $90/5/5$ IID split containing $23{,}020$ training, $1{,}278$ validation, and $1{,}278$ test molecules. We evaluate the model with and without GEs, comparing as well to a 'retyped' model that just replaces τ_i at assigned tetrahedral stereocenters by $\chi_i \in \{R, S\}$. Results are in Table 11.

C.3.3 Pseudovector Mistyping in Magnetic Equivariant ML

Magnetic moments and spins are pseudovectors w, acted on by $G := \mathrm{O}(3)$ as $\widetilde{\rho}^{\mathrm{p}}(Q)w := \det(Q)Qw$. This is in contrast to (ordinary) vectors, which are of course acted on as $\eta(Q)w = Qw$ with no determinant factor. The distinction is not relevant when only $\mathrm{SO}(3) := H \triangleleft G$ is acting as $\rho = \widetilde{\rho}^{p}|_H = \eta|_H$, but it is relevant to equivariant learning for magnetic materials, where spin and magnetic-moment pseudovectors are often represented alongside ordinary vectors such as atomic positions and displacements (Yuan et al., 2024; Yu et al., 2024; Cho et al., 2026). In the atomistic

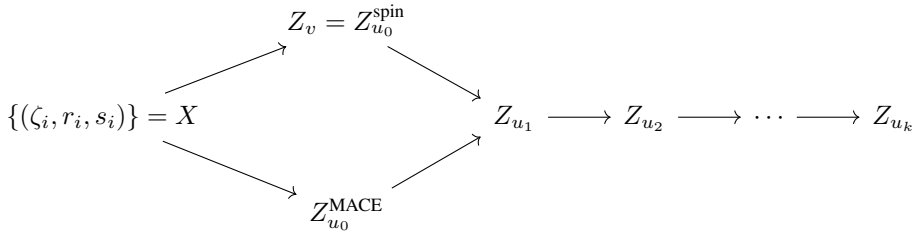
[9] Specifically: $\tau_i(\widetilde{\rho}^{\mathrm{r}}_I(\pi)x) = \tau_i(x)$, while $\chi_i(\widetilde{\rho}^{\mathrm{r}}_I(\pi)x) = \chi_i(x)$ for $\pi \in A_4$ and $\chi_i(\widetilde{\rho}^{\mathrm{r}}_I(\pi)x) = \overline{\chi_i(x)}$ for $\pi \in S_4 - A_4$, where $\overline{R} = S$ and $\overline{S} = R$. Hence $c(\widetilde{\rho}^{\mathrm{r}}_I(\pi)x) = \kappa^{\mathrm{p}}(\pi)c(x)$. Moreover, $\kappa^{\mathrm{p}}|_{A_4} = \kappa^{\mathrm{r}}|_{A_4}$, while for every $s \in S_4 - A_4$, $\mathrm{Eq}(\kappa^{\mathrm{p}}(s), \kappa^{\mathrm{r}}(s)) = \{(a, a, b, d) : a, b, d \in \mathbb{R}\}$. Since at an assigned stereocenter the first two coordinates of $c_i(x)$ are either $(1, 0)$ or $(0, 1)$, so $\mathrm{im}\, c \not\subset \mathrm{Eq}(\kappa^{\mathrm{p}}(s), \kappa^{\mathrm{r}}(s))$.

Architecture	RMSE
Baseline (Xu et al., 2023)	5.3713 ± 0.0971
Retyped	5.1707 ± 0.1258
+ GE	5.2268 ± 0.0454

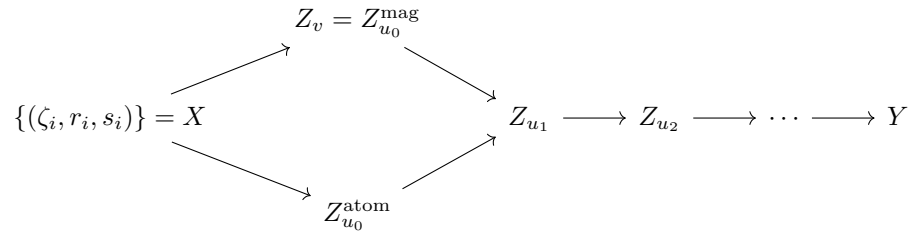
Table 11

equivariant learning literature (Geiger & Smidt, 2022), $\widetilde{\rho}^{\mathrm{p}}$ is denoted $1e$ and η is denoted $1o$ respectively. More generally, the notation $V_{\ell,p} = \ell p$ denotes the G-irrep obtained by extending the unique real degree-ℓ H-irrep (Hitchin, 2015) to G with $p \in \{e, o\}$ recording the action of $G - H$.

Suppose there is a node v in $\mathscr{A}$ whose representation space Z_v consists of spin or magnetic moment features and comes equipped with the standard action $\eta_v(Q)z = Qz$. With $s := -I \in G - H$, $S_v := -I_{Z_v}$, one has $\widetilde{\rho}^{\mathrm{p}}_v(s) \neq S_v$. Therefore, since $[G : H] = 2$, extension-mistyping occurs if S_v can be extended to a family $\boldsymbol{S} = (S_u)_{u \in V}$ of linear maps satisfying hypotheses (i)-(iii) of Proposition 9. For a straightforward example, let $\mathscr{A}$ denote the spin-aware augmentation of MACE (Batatia et al., 2022) proposed in (Cho et al., 2026). The authors propose a framework for augmenting equivariant materials foundation models, and instantiate the framework by augmenting MACE with additional inputs given by spin directions $s_i \in Z_v \cong \mathbb{R}^3$ where Z_v is equipped with the action $\eta_v = 1o$. With $X = \{(\zeta_i, r_i, s_i, \mathcal{G}_i)\}$ denoting tuples of atomic number, position, and spin direction and atomistic graphs $\mathcal{G}_i = (\mathcal{V}_i, \mathcal{E}_i)$ for which they are node features, the architecture $\mathscr{A}$ then has computation graph



where $Z_{u_i} = \bigoplus_{\ell,p} m_{,\ell_i,p_i} V_{\ell,p}$ for some multiplicities $m_{u_i,\ell,p}$ (Batatia et al., 2022). Let $S_{\ell,p} : V_{\ell,p} \to V_{\ell,p}$ be defined as I if $p = e$ and $-I$ if $p = o$. Define $\boldsymbol{S} = S_v \cup (S_{u_i})_{i=1}^k$ via $S_{u_i} = \bigoplus_{\ell_i,p_i} S_{\ell_i,p_i}$. With $\mathscr{A}$ parameterizing functions factoring as $X \xrightarrow{a} Z_{C_1} \xrightarrow{\Phi^\theta} Z_{C_2} \xrightarrow{e^\theta} Y$ where $Z_{C_1} = Z_v \oplus Z^{\mathrm{MACE}}_{u_0}$ and $Z_{C_2} = Z_{u_2}$, one has $S_v \in \boldsymbol{S}$ and it is readily checked that $\boldsymbol{S}$ satisfies the conditions (i)-(iii) of Proposition 9; we conclude that extension-mistyping occurs on $\mathscr{A}_{(C_1,C_2)}$. For another example, let $\mathscr{A}$ denote MagNet, the $E(3)$-equivariant force-field architecture for magnetic materials proposed in (Yuan et al., 2024). MagNet first processes atomic structure through a sequence of atomic interaction blocks, and then combines the resulting atomic features with additional inputs given by magnetic-moment directions $s_i \in Z_v \cong \mathbb{R}^3$, where Z_v is again equipped with the action $\eta_v = 1o$. The resulting features are processed through two magnetic interaction blocks. The architecture $\mathscr{A}$ therefore has computation graph of the form



where $Z^{\mathrm{atom}}_{u_0}$ is the output of the atomic interaction stage and Z_{u_1}, Z_{u_2} are the outputs of the two magnetic interaction blocks, where $Z_{u_i} = \bigoplus_{\ell,p} m_{u_i,\ell,p} V_{\ell,p}$ for some multiplicities $m_{u_i,\ell,p}$. For the

Architecture	Baseline RMSE	GE RMSE	GE R^2
(Cho et al., 2026)	0.071233 ± 0.000208	0.030384 ± 0.002293	0.8151 ± 0.0284
MagNet (Yuan et al., 2024)	0.070866 ± 0.000002	0.032227 ± 0.001077	0.7930 ± 0.0140

Table 12

purposes of our experiments below, we augment MagNet with an equivariant readout that first forms edge features $v_{ij} \in V_{1,o}$ and then combines each v_{ij} with the $r_{ij} \in V_{1,o}$ through a $1o \otimes 1o \to 1e$ tensor product, and regard this as the architecture $\mathscr{A}$. Let $S_{\ell,p} : V_{\ell,p} \to V_{\ell,p}$ be defined as before, and define $\boldsymbol{S}$ likewise as before. With $\mathscr{A}$ parameterizing functions factoring as $X \xrightarrow{a} Z_{C_1} \xrightarrow{\Phi^\theta} Z_{C_2} \xrightarrow{e^\theta} Y$, where $Z_{C_1} = Z_v \oplus Z_{u_0}^{\mathrm{atom}}$ and $Z_{C_2} = Z_{u_2}$, one has $S_v \in \boldsymbol{S}$, and it is readily checked that $\boldsymbol{S}$ satisfies conditions (i)-(iii) of Proposition 9; we conclude that extension-mistyping occurs on $\mathscr{A}_{(C_1,C_2)}$.Moreover, since a and e^θ are both equivariant in both cases above, Proposition 8 applies to give that this extension-mistyping creates symmetry slippage.

While this symmetry slippage does not pose an expressivity problem for tasks invariant under spin reversal (e.g. zero-field scalar energy prediction), it can be problematic otherwise. We define GEs for each aforementioned architecture in turn. Let $Z = \bigoplus_{i \neq j} \mathbb{R}^3$, and write $r_{ij} := r_i - r_j$. For the architecture $\mathscr{A}$ of (Cho et al., 2026), define $c : X \to Z$ by

$$c(x)_{ij} = \begin{cases} (r_{ij} \times s_i) \times r_{ij} & (i,j) \in \mathcal{E}_x \\ 0 & (i,j) \notin \mathcal{E}_x. \end{cases}$$

Note $c\big(\alpha_X^{\mathrm{r}}(Q)x\big)_{ij} = Qc(x)_{ij}$. Now define actions $\kappa^{\mathrm{p}}, \kappa^{\mathrm{r}} : G \curvearrowright Z$ as $\kappa^{\mathrm{p}}(Q) := \bigoplus_{i \neq j} \eta(Q)$ and $\kappa^{\mathrm{r}}(Q) := \bigoplus_{i \neq j} \widetilde{\rho}^{\mathrm{p}}(Q)$. Certainly $c\alpha_X^{\mathrm{r}}(Q) = \kappa^{\mathrm{p}}(Q) \circ c$. Putting $C := C_2$, $\mathscr{A}$ realizes functions as $X \xrightarrow{b} Z_C \xrightarrow{e^\theta} Y$; appending c at C, $\mathscr{A}' = \mathscr{A} \oplus_C c$ gives

$$X \xrightarrow{(b,c)} Z_C \times Z \xrightarrow{e^{\theta'}} Y,$$

where $e^{\theta'}$ is defined as with the $\mathscr{A}_{(C,O)} = \mathrm{MLP}$ example in Section 5.1. For the MagNet architecture $\mathscr{A}$ of (Yuan et al., 2024), instead define $c : X \to Z$ by

$$c(x)_{ij} = \begin{cases} r_{ij} \times s_i & (i,j) \in \mathcal{E}_x, \\ 0 & (i,j) \notin \mathcal{E}_x. \end{cases}$$

Then $c\big(\alpha_X^{\mathrm{r}}(Q)x\big)_{ij} = \det(Q)Q\, c(x)_{ij}$. Define $\kappa^{\mathrm{p}}(Q) := \bigoplus_{i \neq j} \widetilde{\rho}^{\mathrm{p}}(Q)$ and $\kappa^{\mathrm{r}}(Q) := \bigoplus_{i \neq j} \eta(Q)$. Hence $c\alpha_X^{\mathrm{r}}(Q) = \kappa^{\mathrm{p}}(Q)c$. Putting C immediately before our appended readout, $\mathscr{A}$ realizes functions as $X \xrightarrow{b} Z_C \xrightarrow{e^\theta} Y$; appending c at C, $\mathscr{A}' = \mathscr{A} \oplus_C c$ gives $X \xrightarrow{(b,c)} Z_C \times Z \xrightarrow{e^{\theta'}} Y$, with $e^{\theta'}$ defined as above.

In both constructions, $\kappa^{\mathrm{p}}|_H = \kappa^{\mathrm{r}}|_H$ since $\eta|_H = \widetilde{\rho}^{\mathrm{p}}|_H$. Moreover, for $s = -I \in G - H$, $\kappa^{\mathrm{p}}(s) = -\kappa^{\mathrm{r}}(s)$. Indeed, for the first architecture one has $\kappa^{\mathrm{p}}(s) = -I_Z$ and $\kappa^{\mathrm{r}}(s) = I_Z$, while for the second these are reversed. Since in each case $c \not\equiv 0$, and $\mathscr{A}_{(C,O)}$ satisfies the injectivity condition of Proposition 24, Corollary 25 gives that both c's are GEs for their respective architectures.

To demonstrate the repair, we compare the two architectures architectures with and without GEs on the noncollinear anomalous-Hall prediction task of (Kipp et al., 2024). Each input is a magnetic honeycomb configuration specified by two sublattice spin directions s_A, s_B, and the target is the symmetrized noncollinear Hall response $\sigma^{\mathrm{nc}}(s_A, s_B) = \frac{1}{2}\big(\sigma(s_A, s_B) + \sigma(s_B, s_A)\big)$. For each architecture, we compare the GE against a parameter-matched control appended at the same cut and processed by the same downstream architecture: for (Cho et al., 2026) we compare $c_{\mathrm{GE},ij} = (r_{ij} \times s_i) \times r_{ij}$ against $c_{\mathrm{ctl},ij} = r_{ij} \times s_i$, while for MagNet (Yuan et al., 2024) we compare $c_{\mathrm{GE},ij} = r_{ij} \times s_i$ against $c_{\mathrm{ctl},ij} = (r_{ij} \times s_i) \times r_{ij}$. Results are shown in Table 12.

D Experimental Details

We provide specific details to aid in reproducing the experiments described in Sections 5.2 and C. All training was done on an NVIDIA A100 GPU. In each case study, we report mean and standard

deviation over three seeds of the evaluation metric(s) used by the papers proposing the architectures we analyze, unless stated otherwise.

Experimental Details for Case Study 1 We use $64/16/24$ base geometries for training/validation/ID test, respectively, together with a time-orientation-reversed version of every geometry, and an independent 24-base-geometry OOD test set. Each contains $n = 64$ points sampled independently with $t \sim \mathrm{Unif}[-2, 2]$ and $x, y \sim \mathrm{Unif}[-1, 1]$, and four source fields, each generated as a sum of 4-7 anisotropic Gaussian bumps with amplitudes $a \sim \mathcal{N}(0, 1)$, centers $t_c \sim \mathrm{Unif}[-1.5, 1.5]$, $x_c, y_c \sim \mathrm{Unif}[-0.8, 0.8]$, widths $\sigma_t \sim \mathrm{Unif}[0.22, 0.55]$, $\sigma_x, \sigma_y \sim \mathrm{Unif}[0.18, 0.45]$, and normalized to unit standard deviation. We use uniform quadrature weights $w_j = 16/64 = 0.25$ and $q_{\min} = 0.005$. ID geometries are transformed by a boost with direction $\theta \sim \mathrm{Unif}[0, 2\pi)$ and rapidity $\eta \sim \mathrm{Unif}[-2, 2]$; OOD uses the same random directions with an equiprobable rapidity sign and $|\eta| \sim \mathrm{Unif}[4, 4.5]$. We use the released LorentzNet, PELICAN, LGN, and L-GATr models, modified for the task as described in Appendix C The control uses an identically zero scalar slot wherever the GE is supplied in the repaired arm, giving equal parameter counts.

Experimental Details for Case Study 2 We use the weekly Mauna Loa CO_2 series from `statsmodels.datasets.co2` (March 1958-December 2001 Keeling & Whorf (2004)). Missing weekly levels are linearly interpolated before forming first differences and the 12-week input windows. Because the literal Buckingham-Π features $\Pi_j(h) = h_j/h_0$ are undefined when $h_0 = 0$, we discard such windows, leaving 2100 examples. The resulting chronological train/validation/test split contains $1244/434/422$ examples. We repeat the experiment with a 16-sample purge after each split boundary, so that no weekly observation is shared across train, validation, and test windows; the purged test set contains 406 examples. All arms use full-batch Adam with learning rate 3×10^{-3}, weight decay 10^{-4}, at most 2000 epochs.

Experimental Details for Case Study 3 Following the Treloar protocol of Holthusen et al. (Holthusen et al., 2026), we construct incompressible deformation gradients from the reported stretches and fit jointly to the 24 uniaxial-tension (UT) and 16 biaxial-tension (BT) stress measurements, reserving pure shear for extrapolation. We parameterize the input by the true strain $\varepsilon = \frac{1}{2}\log(F^\top F)$, predict a strain energy $W_\theta(\varepsilon)$, and obtain the corresponding nominal stresses by automatic differentiation through the deformation parametrization. We optimize $\mathrm{MSE}_{\mathrm{UT}} + 3\,\mathrm{MSE}_{\mathrm{BT}}$.

Experimental Details for Case Study 4 For all three architectures, we use the same five outer stratified group folds over the 13,209 records, grouping records by exact rootless attributed graph. In each outer fold, the held-out group is used only for testing, while a second grouped stratified split of the remaining records supplies validation data for checkpoint selection. We otherwise retain the released domain-classification configurations and architecture-specific hierarchy; the common training configuration uses three 128-dimensional message-passing layers, dual readout, batch size 256, 50 epochs, Adam with learning rate 5×10^{-4} and weight decay 10^{-3}, and unweighted multiclass cross-entropy. RGCN and CompGCN checkpoints are selected by validation domain accuracy, while GlycanAA checkpoints are selected by validation Macro-F1. For each seed, predictions from the five held-out folds are concatenated to give one out-of-fold prediction over all 13,209 records; Table 5 reports Macro-F1 on the pooled OOF predictions, with mean and standard deviation taken over the three seeds rather than over folds.

Experimental Details for Case Study 5 For the targeted LVOLT/VCLVH experiments, we use the standard PTB-XL split (Wagner et al., 2020; Strodthoff et al., 2021) of folds 1-8 train, fold 9 validation, fold 10 test. We construct balanced target-aware cohorts by pairing each positive ECG with one fold-matched negative control, excluding target-positive patients from the control pool and preferentially matching sex, device, site, and age. Using all 182 LVOLT and 200 VCLVH positives gives 758 unique ECGs, we split 606/78/74. We report macro AUROC/AUPRC over the two labels. For ECGFounder (Li et al., 2025) and Weimann & Conrad (2024), we freeze respectively 1024-dimensional exact-global-z Net1D and 512-dimensional exact-z ResNet-34 representations from three validation-selected backbones trained on this cohort. For the broader native settings, ECGFounder uses the released pretrained model with ten-fold patient-disjoint OOF evaluation on 21,799 usable PTB-XL records: each fold is test, the next cyclic fold validation, and the remaining eight folds train the matched affine readout. Metrics are pooled OOF macro AUROC/AUPRC. For these entries, the central value is the pooled OOF metric and $\pm$ is patient-cluster-bootstrap standard deviation. For Weimann-Conrad, we train the released central ResNet-34 on the PTB client using the

released splits, yielding 16,667/1,198/4,488 train/validation/test records, for 100 epochs under the documented central-training configuration. We then freeze the 512-dimensional pooled representation and fit matched affine readouts with zero auxiliary input or standardized $\log \rho$; regularization is selected by validation macro AUROC. Final metrics are mask-aware macro averages over scored diagnoses, with $\pm$ denoting standard deviation across three independently trained seeds.

Experimental Details for Case Study 6 We compute the Choi-Williams distribution using the `tfrcw` implementation in the Time-Frequency Toolbox (TFTB) (Auger et al., 1998), with the number of frequency bins equal to the R-R interval length, kernel parameter $\sigma = 1$, and odd Hamming time- and frequency-smoothing windows of lengths $\lfloor N/10 \rfloor$ and $\lfloor N/4 \rfloor$. Each R-R interval is standardized and converted to its analytic signal; after removing 3% of CWD time columns from each boundary, we divide the remainder into five contiguous regions and record the maximum and minimum coefficient in each, repeating the procedure for the first-differenced waveform to obtain 20 CWD features. Although we exactly recovered the MIT-BIH healthy/arrhythmia portion of the dataset used by Kayikcioglu et al. (2020), the released annotations do not determine their ST-elevation/depression sample-selection procedure. We therefore instead devised the discussed experiment using expert-annotated ST episodes from the European ST-T Database (Taddei et al., 1992). For the GE, the ST sample is taken 120, 112, 104, or 100 ms after R_i for heart rates < 100, 100-110, 110-120, or ≥ 120 bpm, respectively; the local baseline is the flattest 40-ms window in $[R_i - 200\,\mathrm{ms}, R_i - 40\,\mathrm{ms}]$, and the recording reference s_r^{ref} is the median ST score over clean N-N intervals ending in the first 30 s. The final cohort contains 68 subjects and 288 episodes (207 depression, 81 elevation); we retain 25 adjacent clean N-N intervals wholly within each episode and nearest its annotated apex, requiring R-R duration in $[0.25, 2.0]$ s. Evaluation is leave-one-subject-out, holding out all recordings of each subject together and averaging the 25 interval probabilities for each episode. Features are standardized using training-fold statistics only. The repaired arm appends $c(r, i) = s_i(r) - s_r^{\mathrm{ref}}$, while the matched control appends $|c(r, i)|$. We use L_2-regularized logistic regression ($C = 1$) and 5-NN with Euclidean distance.

Experimental Details for Case Study 7 For the first experiment, we follow the released G-FNO protocol of Helwig et al. (Helwig et al., 2023). The baseline is the original scalar-typed G-FNO-$p4m$; the retyped arm changes the input/output vorticity representations to the pseudoscalar action, while the GE arm retains the scalar pathway and appends the GE at the first regular-carrier cut. For the decaying-turbulence experiment, we generate initial conditions following Cao et al. (Cao et al., 2025) and their `torch-cfd` implementation using the McWilliams spectrum (McWilliams, 1984): Gaussian white noise is spectrally filtered with density proportional to $[k(1 + (k/k_{\mathrm{peak}})^4)]^{-1}$, the resulting streamfunction is normalized to unit kinetic energy, and vorticity is obtained spectrally. We generate 768 base initial conditions with $k_{\mathrm{peak}} = 4$ on a 64^2 periodic grid and evolve each together with its sign reversal, giving 1,536 trajectories, with $\nu = 10^{-3}$, $\Delta t = 10^{-3}$, no forcing or drag, and $T = 0.5$. Base initial conditions are split $512/128/128$ into train/validation/test sets pre-pairing, so that each $\omega_0, -\omega_0$ pair remains within one split.

Experimental Details for Case Study 8 For QGeoGNN, we use the authors' $90/5/5$ all-column split and retain the released protocol. CIP assignments are computed in RDKit (Landrum, 2026) by explicitly assigning stereochemistry and reading the atom-level `_CIPCode`; assigned tetrahedral centers are mapped to R or S, while unassigned or unsupported stereochemistry is left unassigned. For the GE arm, $c_i = (\mathbf{1}_R, \mathbf{1}_S, \mathbf{1}_{\tau_i=+1}, \mathbf{1}_{\tau_i=-1})$ is zero for atoms without assigned tetrahedral stereochemistry and is injected immediately after the atom-feature encoder and before the first GIN layer through a zero-initialized bias-free $4 \to 128$ linear map, adding 512 trainable parameters. The Retyped arm changes only the stored chiral-tag feature to the corresponding absolute CIP R/S type.

Experimental Details for Case Study 9 We follow the anomalous-Hall setup of Kipp et al. (Kipp et al., 2024), using the ξ-even target at $E_F = -1.5\,\mathrm{eV}$, $\Gamma = 0.025\,\mathrm{eV}$, $t = 1$, $\alpha_R = 0.4\,\mathrm{eV}$, and $\lambda_{\mathrm{ex}} = 1.4\,\mathrm{eV}$. Following their random-spin setup, we sample s_A, s_B independently and uniformly on S^2 and apply the 12 C_{6v} spatial operations, including the identity, together with time reversal alone. We independently generate 2,048 base configurations and, unlike the original protocol, split them $1{,}536/256/256$ into train/validation/test sets before augmentation so that symmetry-related copies cannot cross splits, yielding $19{,}968/3{,}328/3{,}328$ augmented examples. Table 12 reports performance on the 3,328 augmented held-out examples; we additionally evaluate on the 256 untouched held-out base configurations as a robustness check. For Cho et al. (Cho et al., 2026), we reconstruct their published magnetic architecture, the parameter-matched control and GE each contain 3,560

trainable parameters. For MagNet (Yuan et al., 2024), we follow the published architectural specification; its matched control and GE each contain $58{,}378$ parameters.