THROUGH THE LOOKING-GLASS: EFFICIENT PARITY-COMPLETE LEARNING VIA LOCAL $O(2)$ FRAMES

Anonymous authors
Paper under double-blind review

ABSTRACT

Making all symmetry-allowed tensor representations computationally accessible remains a central challenge in equivariant atomistic learning. Clebsch–Gordan tensor products (CGTPs) are computationally demanding, whereas efficient $SO(2)$-based alternatives lack a unified treatment of natural- and unnatural-parity representations. We first introduce a generalized Wigner-$6j$ convolution that exactly recouples interactions involving additional node representations, replacing the first edge-level tensor-product intermediate with reusable node features. This reduces edge computation and storage but leaves the $\mathcal{O}(L^5)$ angular scaling of the sparse CGTPs unchanged. To address this remaining bottleneck, we develop a local $O(2)$ framework whose convolutions scale as $\mathcal{O}(L^3)$. The framework explicitly accounts for both rotations and reflections. Its linear maps, tensor products, and gated nonlinearities, combined with frame transformations, ensure global $O(3)$ equivariance for both spatial parities. Incorporating time-reversal labels extends this construction to $O(3) \times \mathbb{Z}_2^{\mathcal{T}}$. We apply the framework to magnetic TACE (mTACE), using distinct interactions for systems with and without spin–orbit coupling. On collinear CrN and noncollinear Fe benchmarks, mTACE achieves substantial reductions in atomic- and magnetic-force errors. We also provide EquivariantX, a library for global and local $O(2)$-equivariant computation.

1 INTRODUCTION

Modern equivariant machine learning interatomic potentials (MLIPs) have achieved high accuracy and strong transferability across chemical and structural environments, enabling simulations far beyond the reach of direct density-functional theory (DFT) calculations (Batzner et al., 2022; Batatia et al., 2022; Xu et al., 2026). Extending these models to additional physical degrees of freedom, however, raises a central challenge: how to retain the required symmetry information while keeping computation affordable, especially, for supporting spatial and time-reversal parities. For a maximum angular degree L, standard dense Clebsch-Gordan tensor-product (CGTP) evaluation has $\mathcal{O}(L^6)$ angular complexity. Although exploiting coefficient sparsity reduces this cost to $\mathcal{O}(L^5)$ (Bharadwaj et al., 2025; NVIDIA, 2025), retaining both spatial parities doubles the number of irrep types relative to a natural-parity model, while tracking time-reversal parity further expands the representation space to $O(3) \times \mathbb{Z}_2^{\mathcal{T}}$. Cartesian tensor provides an alternative route, but storing a rank-k Cartesian tensor in full requires 3^k components, also leading to exponential growth in feature storage and expensive high-rank contractions (Wang et al., 2026).

Local-frame methods address this angular bottleneck by aligning features with edge directions and reducing $SO(3)$ convolutions to local $SO(2)$ operations (Passaro & Zitnick, 2023; Yu et al., 2026; Xu et al., 2026) . However, equivariance under rotations about a reference direction alone does not specify the action of reflections or time reversal. Spatial inversion reverses the edge direction and induces a local reflection after realignment. Features with identical local rotation laws can therefore require different reflection actions, which must be reconciled with their global spatial parities. In particular, the $m = 0$ channels used for scalar nonlinearities and gating are not necessarily reflection-even or time-even. Operations permitted by $SO(2)$ symmetry can consequently mix incompatible representations or violate symmetry through their nonlinearities. The difficulty lies in consistently resolving these constraints across representation maps, linear layers, tensor products, and gates, while

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.

ensuring that the resulting local computations guarantee global $O(3) \times \mathbb{Z}_2^{\mathcal{T}}$ equivariance and retain efficient angular scaling.

Magnetic materials provide a concrete setting in which these distinctions are essential. Their coupled atomic structure and magnetic order govern phenomena such as magnetoelasticity and spin-orbit-induced anisotropy (Wu et al., 2024; Weber et al., 2022; Ding et al., 2024; Wang et al., 2026), with applications in information storage, energy conversion, sensing, and spin-based electronics. Atomic displacements are time-even polar vectors, whereas magnetic moments are time-odd axial vectors. For example, $\boldsymbol{m}_i \cdot \boldsymbol{n}_{ij}$ is unchanged by rotations about the edge direction $\boldsymbol{n}_{ij}$, but changes sign under either spatial inversion or time reversal. Although it is a scalar under local $SO(2)$ rotations, it cannot be passed through an arbitrary activation or used as an invariant gate without violating these additional symmetries. Despite recent progress in magnetic ACE, time-reversal-equivariant networks, and magnetic message passing (Rinaldi et al., 2024; Yu et al., 2024; Ho et al., 2026; Wang et al., 2026), this area remains considerably less mature than conventional MLIPs.

To address these challenges, we first introduce a generalized Wigner-$6j$ convolution that exactly recouples interactions among node features, edge spherical harmonics, and additional node representations, enabling reuse of node-level intermediates across edges to reduce redundant computation. We establish a local $O(2)$ theory that preserves complete $O(3)$ interactions and extend it with time-reversal parity to realize $O(3) \times \mathbb{Z}_2^{\mathcal{T}}$ equivariance. We implement the full framework in EquivariantX, a unified library of representation maps, frame transformations, linear layers, tensor products, and gates that makes parity-complete local computation reusable across equivariant architectures. For demonstration, we adopt the TACE architecture, the related TECE shows that real-valued local weights can preserve global $O(3)$ equivariance for natural-parity features (Xu et al., 2026) and achieves state-of-the-art accuracy on benchmarks such as Matbench Discovery (Riebesell et al., 2025). Using EquivariantX, we develop magnetic TACE (mTACE) with separate formulations for systems with and without spin-orbit coupling (SOC). mTACE achieves substantially better performance in benchmarks of magnetic systems than the reported baselines.

2 Related Work

2.1 Magnetic Machine Learning Interatomic Potentials

Early magnetic MLIPs were typically limited to collinear magnetic moments, while subsequent magnetic ACE and equivariant neural networks incorporated noncollinear magnetic moments and higher-order interactions (Rinaldi et al., 2024; Yu et al., 2024). mMACE (Ho et al., 2026) represents moments with spherical features and couples them to a MACE many-body expansion. Its most expressive formulation applies one CGTP to form a geometric edge feature and a second CGTP to couple the source magnetic solid harmonics. The high cost of these successive CGTPs motivates a reduced symmetry constraint, with the practical model enforcing only $SO(3)$ equivariance. HotPP-spin (Wang et al., 2026) models exchange-dominated and SOC systems with a Cartesian tensor network. Coupled structural and magnetic features carry explicit spatial and time-reversal labels, which multiply under tensor couplings, while the energy readout admits only spatially even and time-even scalar channels. This construction enforces the required symmetries. Its compression of some tensor-product paths ties weights across certain irreducible components, reducing the number of independently parameterized couplings.

2.2 Time-Reversal e3nn

Time-reversal e3nn (Yu et al., 2024) augments each $O(3)$ irrep with a $\mathbb{Z}_2^{\mathcal{T}}$ parity and propagates this label through linear maps, tensor products, gates, and other operations. It thus provides a general framework for incorporating time-reversal symmetry into CGTP-based networks, while retaining the high computational cost of CGTPs.

2.3 Wigner-$6j$ Convolution

E2Former (Li et al., 2026) introduced Wigner-$6j$ recoupling as a new convolution for geometry-based energy, force, and stress prediction. It uses the binomial expansion of regular solid harmonics to rewrite $\mathcal{R}^{(\ell)}(\boldsymbol{r}_j - \boldsymbol{r}_i)$ in terms of $\mathcal{R}^{(u)}(\boldsymbol{r}_i)$ and $\mathcal{R}^{(\ell-u)}(\boldsymbol{r}_j)$, allowing angular computation to be

reused across nodes. However, due to its reliance on solid harmonics, this construction is generally inequivalent to standard CGTP networks unless specifically designed to recover such equivalence. It is specialized to the positional solid-harmonic expansion and does not describe how independently supplied invariant or equivariant node inputs can directly modulate the message.

2.4 $SO(2)$ Local Frames

eSCN (Passaro & Zitnick, 2023) showed that aligning representations with an edge reduces an $SO(3)$ convolution to $SO(2)$ operations. QHNetV2 (Yu et al., 2026) extended this framework with $SO(2)$ tensor products and gates for Hamiltonian prediction. TECE (Xu et al., 2026) showed that real-valued weights preserve $O(3)$ equivariance for natural-parity features. The more challenging question here is the complete local representation needed when polar tensors, pseudotensors, and both time parities coexist.

3 From Wigner-$6j$ Recoupling to Local $O(2)$ Frames

A direct CGTP construction first couples a source node feature to an edge spherical harmonic, then couples this edge intermediate to an additional source-node representation, such as magnetic solid harmonics. These successive edge-level tensor products incur substantial computation and intermediate storage. Our generalized Wigner-$6j$ convolution instead couples the two node inputs once per node and reuses the result across outgoing edges. With the recoupling coefficients and path normalizations included, it exactly reproduces the two-stage CGTP without storing its first edge intermediate. Fused backends, such as OpenEquivariance (Bharadwaj et al., 2025) and CuEquivariance (NVIDIA, 2025), can thus be used to avoid explicit materialization of edge-level intermediates. Appendix C gives the recoupling, weight construction, and backend requirements. Recoupling reduces edge-level intermediate storage but still involves a substantial number of coupling paths. Their remaining cost motivates a complementary approach: align features with a reference direction and perform the learnable update using local $O(2)$ operators. The following section establishes the rotation and reflection constraints that ensure global $O(3)$ equivariance, and then incorporates time-reversal parity.

4 Local $O(2)$ Framework

We label $O(3) \times \mathbb{Z}_2^{\mathcal{T}}$ irreps by ℓpt, where p and t denote spatial and time-reversal parities. Relative positions and magnetic moments transform as `1oe` and `1eo`, respectively, while the energy is an invariant `0ee` scalar. With SOC, spatial and spin rotations are coupled; without SOC, they are independent. Magnetic regular solid harmonics carry the same parities as spherical harmonics and remain smooth at zero magnetic moment. Appendix A and B summarize notation and symmetry rules. Here, we use an edge direction, given by the normalized relative position $\boldsymbol{n} = \boldsymbol{r}_{ij}/\|\boldsymbol{r}_{ij}\|$ (`1o`), to illustrate the construction of local $O(2)$ frames and $O(3)$-equivariant maps. The construction is analogous for axial reference directions (`1e`), but the stabilizer subgroup is $SO(2) \times \mathbb{Z}_2$.

4.1 Local frames and stabilizer

Choose a proper rotation $R_{\boldsymbol{n}} \in SO(3)$ that aligns $\boldsymbol{n}$ with $\boldsymbol{e}_y = (0,1,0)^{\mathsf{T}}$:

$$R_{\boldsymbol{n}}\boldsymbol{n} = \boldsymbol{e}_y. \tag{1}$$

The stabilizer of the aligned direction is

$$H = \{g \in O(3) : g\boldsymbol{e}_y = \boldsymbol{e}_y\} \simeq O(2) \simeq SO(2) \rtimes \mathbb{Z}_2. \tag{2}$$

Its elements act orthogonally on the perpendicular xz plane. Those with determinant $+1$ are rotations about the y axis; those with determinant -1 are reflections through planes containing the y axis. A global transformation $Q \in O(3)$ acts on both the features and $\boldsymbol{n}$. The corresponding transformation in the aligned frame is

$$g(Q,\boldsymbol{n}) = R_{Q\boldsymbol{n}} Q R_{\boldsymbol{n}}^{-1}, \qquad g(Q,\boldsymbol{n})\boldsymbol{e}_y = R_{Q\boldsymbol{n}} Q\boldsymbol{n} = \boldsymbol{e}_y, \qquad \det g(Q,\boldsymbol{n}) = \det Q. \tag{3}$$

By definition, $g(Q,\boldsymbol{n})R_{\boldsymbol{n}} = R_{Q\boldsymbol{n}}Q$: first applying the global transformation Q and then aligning is equivalent to first aligning and then applying the corresponding local transformation $g(Q,\boldsymbol{n})$.

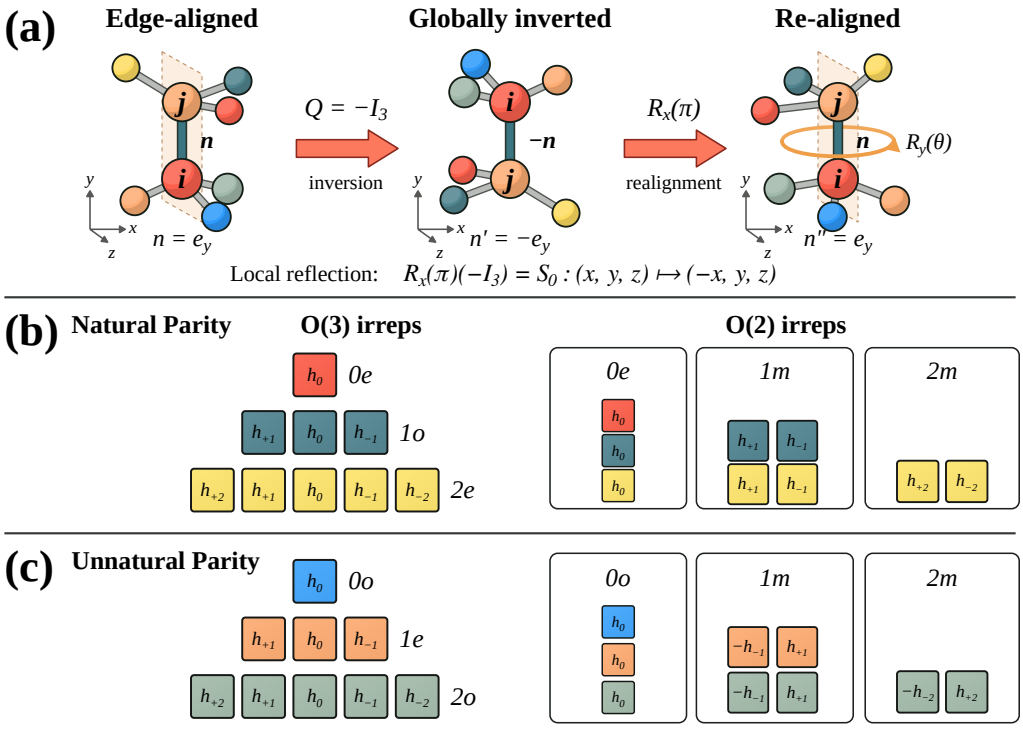


Figure 1: Global $O(3)$ to local $O(2)$ decomposition. (a) The edge direction $\boldsymbol{n}$ is rotated onto the y axis, leaving an $O(2)$ action in the plane perpendicular to the edge. (b) Restriction of natural-parity $O(3)$ irreps with $\ell \leq 2$ into the $m = 0, 1, 2$ local irreps. (c) The corresponding restriction of unnatural-parity irreps, where the $m = 0$ irreps are reflection-odd. For $m > 0$, we apply a change of basis.

The condition $g(Q, \boldsymbol{n})\boldsymbol{e}_y = \boldsymbol{e}_y$ shows that $g(Q, \boldsymbol{n}) \in H$: it leaves the aligned direction unchanged and acts as an $O(2)$ transformation on the perpendicular xz plane. Since both alignment rotations are proper, $\det g(Q, \boldsymbol{n}) = \det Q$. Thus proper and improper global transformations induce local rotations and reflections, respectively. Local reflection equivariance is therefore required for global $O(3)$ equivariance. The group H is generated by rotations $R_y(\theta)$ and one reflection S_0. Every reflection in H has the form $R_y(\theta)S_0$. We choose

$$S_0 = R_x(\pi)(-I_3) = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{4}$$

This reflects positions through the yz plane, or equivalently across the z axis in the xz plane. Although inversion $-I_3$ is not in H, it induces a local reflection after realignment. For example, the valid alignments $R_{\boldsymbol{e}_y} = I_3$ and $R_{-\boldsymbol{e}_y} = R_x(\pi)$ give $g(-I_3, \boldsymbol{e}_y) = S_0$.

4.2 Restriction and Local Basis

Rotations. Let $\rho_{\ell,p}$ denote the real $O(3)$ irrep of degree ℓ and spatial parity p. We use a real spherical harmonic basis with angular factors $\cos m\phi$ and $\sin m\phi$ for h_{+m} and h_{-m}, respectively, where ϕ is measured from z toward x. Rotations about y leave h_0 invariant and act on each two-dimensional subspace with $m = 1, \ldots, \ell$ as

$$\begin{pmatrix} h_{+m} \\ h_{-m} \end{pmatrix} \mapsto R_m(\theta) \begin{pmatrix} h_{+m} \\ h_{-m} \end{pmatrix}, \qquad R_m(\theta) = \begin{pmatrix} \cos m\theta & -\sin m\theta \\ \sin m\theta & \cos m\theta \end{pmatrix}. \tag{5}$$

These proper rotations are independent of p.

Reflection and change of basis. The rotation matrices above are independent of spatial parity, but the reflection action is not. In the basis defined above, $S_0 = (-I_3)R_x(\pi)$ gives

$$\rho_{\ell,p}(S_0) = pD^{(\ell)}(R_x(\pi)) = p(-1)^\ell \left(1 \oplus \bigoplus_{m=1}^{\ell} F\right), \qquad F = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{6}$$

The factor p comes from spatial inversion, while $(-1)^\ell$ follows from the Wigner-D matrix in this basis. At $m = 0$, the reflection character $q = p(-1)^\ell$ distinguishes the one-dimensional irreps $(0, +1) = 0e$ and $(0, -1) = 0o$. Their opposite signs cannot be related by a one-dimensional change of basis. For $m > 0$, natural- and unnatural-parity $O(3)$ irreps therefore give the same rotation matrix $R_m(\theta)$ but opposite reflection matrices F and $-F$. This difference is inherited from the parent $O(3)$ irreps, so their features cannot be mixed directly as channels of the same local irrep. We leave the natural-parity features unchanged and transform those with $p(-1)^\ell = -1$ by

$$\begin{gathered} J = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \qquad \begin{pmatrix} h'_{+m} \\ h'_{-m} \end{pmatrix} = J \begin{pmatrix} h_{+m} \\ h_{-m} \end{pmatrix} = \begin{pmatrix} -h_{-m} \\ h_{+m} \end{pmatrix}, \\ JR_m(\theta)J^{-1} = R_m(\theta), \qquad J(-F)J^{-1} = F. \end{gathered} \tag{7}$$

After this change of basis, features with the same $m > 0$ use identical rotation and reflection matrices and can be mixed across channels. We denote this two-dimensional real irrep by $(m, 0)$, where 0 indicates the absence of an independent reflection-parity label.

Restriction rule. Under $H \simeq O(2)$, an $O(3)$ irrep decomposes as

$$(\ell, p)_{O(3)}\downarrow_{O(2)} = \underbrace{(0, p(-1)^\ell)}_{\text{1D},\, m=0} \oplus \bigoplus_{m=1}^{\ell} \underbrace{(m, 0)}_{\text{2D}}. \tag{8}$$

Figure 1 illustrates the restriction for O(3) natural- and unnatural-parity irreps.

4.3 Global equivariance

Let ρ_{in} and ρ_{out} be direct sums of $O(3)$ irreps for the input and output features, respectively. Let B_{in} and B_{out} be the fixed orthogonal transformations into the local $O(2)$ basis: they apply J to every positive-order subspace with $p(-1)^\ell = -1$, and leave the remaining subspaces unchanged. In this basis, the restricted representations are

$$\sigma_{\text{in}}(g) = B_{\text{in}}\rho_{\text{in}}(g)B_{\text{in}}^{-1}, \qquad \sigma_{\text{out}}(g) = B_{\text{out}}\rho_{\text{out}}(g)B_{\text{out}}^{-1}, \qquad g \in H. \tag{9}$$

Let Φ be an $O(2)$-equivariant operator in this basis:

$$\Phi(\sigma_{\text{in}}(g)h) = \sigma_{\text{out}}(g)\Phi(h), \qquad g \in H. \tag{10}$$

The resulting learnable map

$$\mathcal{F}_{\boldsymbol{n}}(h) = \rho_{\text{out}}(R_{\boldsymbol{n}}^{-1})B_{\text{out}}^{-1}\Phi(B_{\text{in}}\rho_{\text{in}}(R_{\boldsymbol{n}})h) \tag{11}$$

is then globally equivariant:

$$\mathcal{F}_{Q\boldsymbol{n}}(\rho_{\text{in}}(Q)h) = \rho_{\text{out}}(Q)\mathcal{F}_{\boldsymbol{n}}(h), \qquad Q \in O(3). \tag{12}$$

Proof. With $g = g(Q, \boldsymbol{n}) \in H$ from Eq. (3), the frame transformations satisfy

$$\begin{aligned} B_{\text{in}}\rho_{\text{in}}(R_{Q\boldsymbol{n}})\rho_{\text{in}}(Q) &= B_{\text{in}}\rho_{\text{in}}(g)\rho_{\text{in}}(R_{\boldsymbol{n}}) = \sigma_{\text{in}}(g)B_{\text{in}}\rho_{\text{in}}(R_{\boldsymbol{n}}), \\ \rho_{\text{out}}(R_{Q\boldsymbol{n}}^{-1})B_{\text{out}}^{-1}\sigma_{\text{out}}(g) &= \rho_{\text{out}}(R_{Q\boldsymbol{n}}^{-1})\rho_{\text{out}}(g)B_{\text{out}}^{-1} = \rho_{\text{out}}(Q)\rho_{\text{out}}(R_{\boldsymbol{n}}^{-1})B_{\text{out}}^{-1}. \end{aligned} \tag{13}$$

Applying these identities together with local equivariance gives

$$\begin{aligned} \mathcal{F}_{Q\boldsymbol{n}}(\rho_{\text{in}}(Q)h) &= \rho_{\text{out}}(R_{Q\boldsymbol{n}}^{-1})B_{\text{out}}^{-1}\Phi(\sigma_{\text{in}}(g)B_{\text{in}}\rho_{\text{in}}(R_{\boldsymbol{n}})h) \\ &= \rho_{\text{out}}(R_{Q\boldsymbol{n}}^{-1})B_{\text{out}}^{-1}\sigma_{\text{out}}(g)\Phi(B_{\text{in}}\rho_{\text{in}}(R_{\boldsymbol{n}})h) \\ &= \rho_{\text{out}}(Q)\mathcal{F}_{\boldsymbol{n}}(h). \end{aligned} \tag{14}$$

Independence of the alignment choice. Any other proper alignment of the same direction has the form $R'_{\boldsymbol{n}} = R_y(\theta)R_{\boldsymbol{n}}$. Substituting it into Eq. (11) inserts $\sigma_{\text{in}}(R_y(\theta))$ before Φ and $\sigma_{\text{out}}(R_y(\theta))^{-1}$ after it. These transformations cancel by Eq. (10), giving $\mathcal{F}'_{\boldsymbol{n}}(h) = \mathcal{F}_{\boldsymbol{n}}(h)$. Thus the output is independent of the arbitrary rotation of the frame about the aligned direction, including direction-dependent choices of θ.

4.4 EquivariantX for $O(2)$-Equivariant Operators

EquivariantX adopts a flattened `ir_mul` layout and provides the frame transforms above together with native $O(2)$ linear maps, gates, and tensor products. These operators are equivariant under the rotation and reflection representations of their declared irreps and can also be used independently of a local frame. Figure 1 and Figure 2 summarize them; Appendix G gives an executable example. **O(2) Local Frame.** We establish the transformation between global $O(3)$ features and local $O(2)$ features in a form compatible with the `e3nn` (Geiger & Smidt, 2022). **O(2) Linear.** Linear mixes identical irreps using real-valued weights. A bias is allowed only for `0e`. **O(2) Gate.** For label-preserving activations, `0e` admits arbitrary nonlinearities, while `0o` requires odd ones. Positive-order irreps are multiplied by `0e` gates. Since local scalars may arise from $\ell > 0$ irreps, the nonlinearities are not restricted to $\ell = 0$ features. **O(2) Tensor Product.** Tensor products decompose according to Table 1. In particular, coupling a `0o` scalar s to a positive-order feature $h_m = (h_{+m}, h_{-m})^{\mathsf{T}}$ uses the map sJh_m, rather than direct scalar multiplication, to retain the reflection convention F.

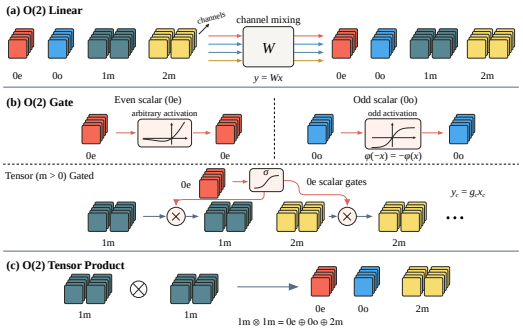


Figure 2: Core $O(2)$-equivariant operations in EquivariantX: (a) Linear transformations. (b) Gated nonlinearities. (c) Tensor products.

Table 1: Spatial tensor-product rules for real $O(2)$ irreps.

Input irreps	Output irreps
`0e`⊗`0e`	`0e`
`0e`⊗`0o`	`0o`
`0o`⊗`0o`	`0e`
`0e`⊗mm	mm
`0o`⊗mm	mm
m_1m⊗m_2m, $m_1 \neq m_2$	$\lvert m_1 - m_2\rvert$m⊕$(m_1 + m_2)$m
mm⊗mm	`0e`⊕`0o`⊕2mm

4.5 Time-Reversal Symmetry

A time-even reference direction leaves the frame unchanged under time reversal. The time label t is therefore preserved after restriction:

$$(\ell, p, t)_{O(3)\times\mathbb{Z}_2^{\mathcal{T}}}\Big\downarrow_{O(2)\times\mathbb{Z}_2^{\mathcal{T}}} = \big(0, p(-1)^\ell, t\big) \oplus \bigoplus_{m=1}^{\ell} (m, 0, t). \tag{15}$$

Requiring Φ to be $O(2) \times \mathbb{Z}_2^{\mathcal{T}}$-equivariant gives global $O(3) \times \mathbb{Z}_2^{\mathcal{T}}$ equivariance by the same proof. Linear maps preserve t, and tensor products multiply it. Label-preserving scalar activations must be odd if $q = -1$ or $t = -1$; invariant gates and biases carry $(m, q, t) = (0, +1, +1)$. Time-reversal e3nn (Yu et al., 2024) and EquivariantX supply the global and local operators, respectively.

5 Magnetic TACE

mTACE augments every spatial edge with magnetic information. Unlike positions, magnetic moments are unchanged by a translation. Therefore, we evaluate the angular basis on nodes and use regular solid harmonics to ensure smoothness at $\boldsymbol{m}_i = \mathbf{0}$. The magnetic cutoff $m_{\text{cut},Z}$ and basis normalization are described in Appendix E.

SOC and non-SOC Magnetic Local $O(2)$ Convolution. The SOC model is capable of capturing the full range of interactions. However, generating SOC training data is computationally expensive.

Therefore, our model is designed to provide a fully symmetry-preserving SOC formulation together with a compatible non-SOC counterpart. This unified framework makes it possible to pretrain foundation models on large-scale, relatively inexpensive non-SOC datasets and subsequently fine-tune them on smaller SOC datasets. Both variants use the magnetic node attributes $\mathcal{M}_i = \bigoplus_{\ell=0}^{L_{\mathrm{mag}}} \mathcal{R}^{(\ell)}(\boldsymbol{m}_i)$ and differ in the magnetic output degrees retained on each edge:

$$\mathcal{M}_{ij}^{\mathrm{SOC}} = \bigoplus_{\substack{0\le \ell_i,\ell_j \le L_{\mathrm{mag}} \\ |\ell_i-\ell_j| \le L \le \min(\ell_i+\ell_j, L_{\mathrm{mag}})}} \left[\mathcal{M}_i^{(\ell_i)} \otimes \mathcal{M}_j^{(\ell_j)}\right]_L, \quad \mathcal{M}_{ij}^{\text{non-SOC}} = \bigoplus_{\ell=0}^{L_{\mathrm{mag}}} \left[\mathcal{M}_i^{(\ell)} \otimes \mathcal{M}_j^{(\ell)}\right]_{L=0}. \tag{16}$$

Here ℓ_i, ℓ_j are node angular degrees and L is the magnetic edge output degree, all truncated at L_{mag}. Both products are unweighted, and coupling paths remain separate. The non-SOC architecture retains only $L = 0$, which forces $\ell_i = \ell_j$ and gives $(L_{\mathrm{mag}} + 1)$ copies of `0ee`. These channels preserve exchange correlations, including $\boldsymbol{m}_i \cdot \boldsymbol{m}_j$, while remaining invariant under spin rotations. The SOC architecture additionally retains non-scalar channels that can couple spin orientation to the lattice. Let $\widetilde{m}_i$ denote the transformed magnetic magnitude. In layer t, the magnetic node features $T(\widetilde{m}_i)$, where T denotes the Chebyshev polynomials of the first kind, are mapped by an element-dependent linear embedding and then passed through independent source and target MLPs on nodes. Their outputs are gathered onto each edge and multiplied element-wise ($\odot$) to generate the weights for subsequent linear layers that project all irreps to the same number of channels.

$$W_{ij}^{(t)} = \mathrm{MLP}_{\mathrm{tgt}}^{(t)}\left(W_{Z_i}^{(t)} T(\widetilde{m}_i)\right) \odot \mathrm{MLP}_{\mathrm{src}}^{(t)}\left(W_{Z_j}^{(t)} T(\widetilde{m}_j)\right), \quad \widetilde{\mathcal{M}}_{ij}^{(t)} = \mathrm{Linear}\left(\mathcal{M}_{ij}; W_{ij}^{(t)}\right). \tag{17}$$

For each edge (i, j), we gather the node features $h_i^{(t)}$ and $h_j^{(t)}$ and combine them with the magnetic edge features $\widetilde{\mathcal{M}}_{ij}^{(t)}$ from Eq. (17). Writing D_{ij} for the representation-specific Wigner-D transform into edge-aligned frames, including the fixed local basis change above, the local input is

$$h_{ij}^{(t)} = D_{ij} h_i^{(t)} \oplus D_{ij} h_j^{(t)} \oplus D_{ij} \widetilde{\mathcal{M}}_{ij}^{(t)}. \tag{18}$$

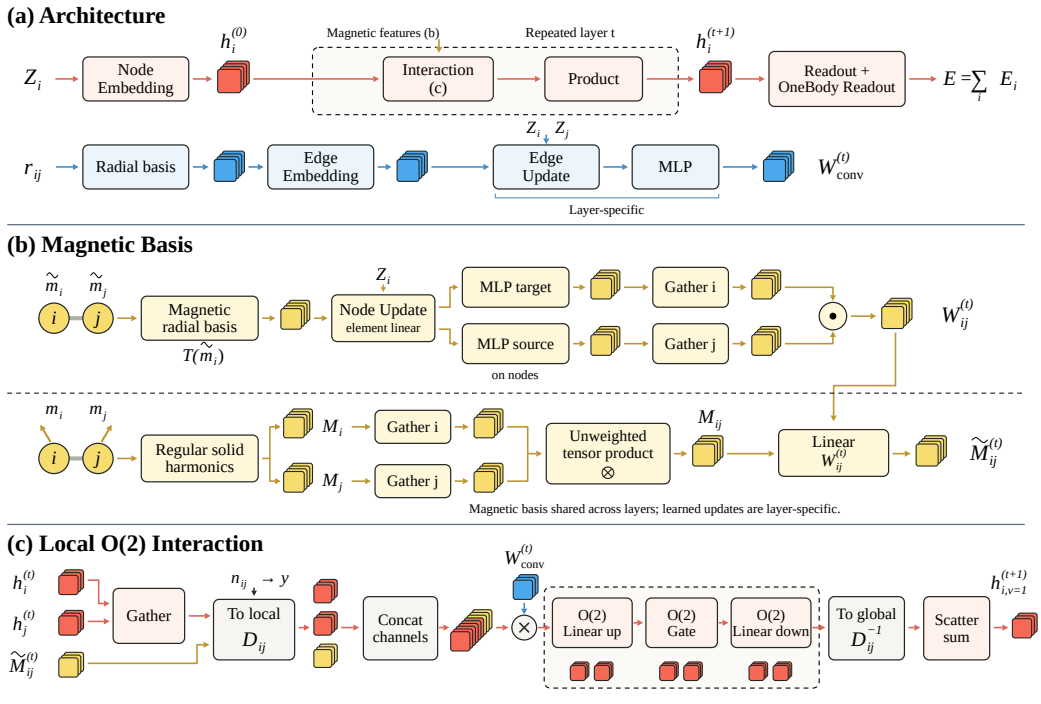


Figure 3: mTACE architecture. (a) Model overview with embedding, update, interaction, product, and readout blocks. (b) Construction of magnetic edge attributes and layer-specific radial weights. (c) A minimal magnetic local $O(2)$ convolution, where $\nu = 1$ denotes correlation order one before the product-basis expansion.

Table 2: RMSEs for the magnetic benchmarks. Energy, atomic-force, stress, and magnetic-force errors are denoted by E, F, S, and F^{mag} and reported in meV/atom, meV/Å, meV/Å^3, and meV/μ_{B}, respectively. A dash means that no result was reported. The best results are highlighted in **bold**.

Collinear CrN

Model	Type	E	F	S	F^{mag}
mMTP	non-SOC	1.7	108	2.50	64
mMACE	SOC[a]	1.21	24	1.56	19
mTACE	non-SOC	0.23	8.34	**0.39**	4.15
mTACE	SOC[a]	**0.18**	**6.85**	0.40	**3.34**

Noncollinear Fe-DeepSpin

Model	Type	E	F	S	F^{mag}
DeePSPIN-DZP	SOC	8.60	111	–	31.2
HotPP-spin	non-SOC	1.21	13.55	–	6.94
mTACE	non-SOC	0.36	5.19	0.64	**1.05**
mTACE	SOC	**0.27**	**5.10**	**0.56**	1.06

[a]For CrN, SOC models use e3nn without explicit time-reversal labels; independent global spin rotations and time reversal are imposed by data augmentation.

The $O(2)$ convolution used in all benchmark models is

$$h_{ij}^{(t)} \xrightarrow{\boldsymbol{W}_{\mathrm{conv},ij}^{(t)}} O(2)\ \mathrm{Linear}_{\mathrm{up}} \longrightarrow O(2)\ \mathrm{Gate} \longrightarrow O(2)\ \mathrm{Linear}_{\mathrm{down}} \longrightarrow \mathrm{RRA}, \tag{19}$$

where $\boldsymbol{W}_{\mathrm{conv},ij}^{(t)}$ are generated from spatial radial features and the source–target element embedding. All benchmark models use Radial Rotary Complex Attention (RRA) (Xu et al., 2026) after the final Linear to incorporate radial information. The message is rotated back to the global frame and scattered to the target node. The subsequent TACE product basis performs the explicit many-body expansion, and the energy readout retains only the invariant `0ee` scalar. See Fig. 3 for the complete architecture of $O(2)$-mTACE. The Wigner-6j interaction is described separately in Appendix C and is not used in these benchmarks.

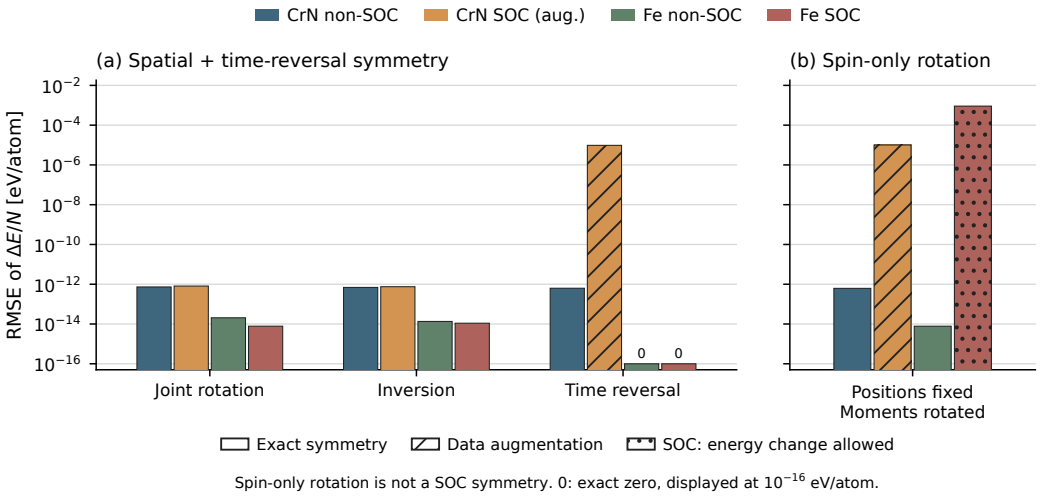


Figure 4: Energy symmetry tests in float64. (a) Joint spatial rotations, inversion, and time reversal. (b) Independent spin rotations, which need not preserve SOC energies. Colors identify models; diagonal hatching marks augmented symmetries and dotted hatching marks the SOC diagnostic. Exact zeros are displayed at a positive plotting floor on the logarithmic scale.

6 Magnetic Benchmarks

We evaluate mTACE on two public magnetic datasets: collinear CrN (Kotykhov et al., 2025) and noncollinear Fe-DeepSpin (Zheng et al., 2026). CrN and the Fe_{16} subset of Fe-DeepSpin are each split into training, validation, and test sets at a ratio of $0.8 : 0.1 : 0.1$. For the CrN benchmark, we use e3nn without explicit time-reversal labels (Geiger & Smidt, 2022). When training the SOC architecture, we augment the data with spin rotations and time reversal. The Fe-DeepSpin labels

were computed with SOC. However, the source study reports that SOC effects in elemental Fe are smaller than its intrinsic model errors and retained SOC primarily to validate the noncollinear spin-constrained DFT workflow. Following HotPP-spin (Wang et al., 2026), we therefore train both SOC and non-SOC architectures on this dataset. The non-SOC model treats it as an exchange-dominated benchmark: it preserves relative-spin correlations while neglecting the unresolved SOC-induced coupling between spin orientation and the lattice. Table 2 compares mTACE with published mMTP, HotPP-spin, mMACE, and DeePSPIN-DZP results (Kotykhov et al., 2025; Wang et al., 2026; Ho et al., 2026; Zheng et al., 2026). Across its variants, mTACE achieves the lowest RMSE for each evaluated property in this comparison. The reference results retain their original evaluation protocols; Appendix H details the splits and our training settings. The local $O(2)$ construction retains the theoretical angular scaling of its $SO(2)$ counterpart (Passaro & Zitnick, 2023), as analyzed in Appendix F.

Symmetry tests Alongside the prediction errors in Table 2, we evaluate energy invariance for the four models on their respective test sets in float64 precision, reporting the RMSE of the energy change per atom in Figure 4. Our local $O(2)$ construction guarantees equivariance in exact arithmetic; symmetries enforced by construction exhibit only small finite-precision residuals in our numerical tests. In contrast, the CrN SOC model, which learns time reversal and independent spin rotations through data augmentation, retains substantially larger deviations under both transformations. Thus, data augmentation does not guarantee exact physical symmetry, motivating explicit representation and coupling constraints. Independent spin rotations test an additional non-SOC symmetry: the energy change of the Fe SOC model under these rotations is allowed by SOC and is not an equivariance violation.

7 Conclusion

We introduce two complementary approaches to efficient parity-complete equivariant computation. Generalized Wigner-$6j$ convolution exactly recouples two-stage, three-input equivariant tensor products, replacing the first edge-level intermediate with reusable node features. Our parity-complete local $O(2)$ framework resolves reflection and time-reversal actions through compatible representations and operators, guaranteeing global $O(3) \times \mathbb{Z}_2^{\mathcal{T}}$. At fixed channel multiplicities, its local linear-and-gating convolution retains the $\mathcal{O}(L^3)$ angular scaling of its $SO(2)$ counterpart. EquivariantX supplies a closed set of reusable linear, tensor product, gating, and frame operations across architectures, while mTACE built on EquivariantX demonstrates accurate magnetic learning with distinct SOC and non-SOC formulations. Together, these developments make complete spatial-parity and time-reversal constraints compatible with efficient local computations. Dedicated $O(2)$ operators, including optimized fused kernels, could further improve efficiency, while applications involving external fields and additional physical degrees of freedom offer broader uses for the framework.

AI Use Statement

Large language models (LLMs) were used as assistive tools during this project. They assisted with code development and debugging, refinement of experimental designs, editing and proofreading of the manuscript, and verification of the implementation and mathematical derivations for potential errors. The research questions, ideas, research direction, experimental decisions, interpretation of the results, and conclusions were independently developed and determined by the authors. All LLM-generated content and suggestions were critically reviewed and verified by the authors, who take full responsibility for the final content of this work.

References

Ilyes Batatia, David P Kovacs, Gregor Simm, Christoph Ortner, and Gábor Csányi. Mace: Higher order equivariant message passing neural networks for fast and accurate force fields. *Advances in neural information processing systems*, 35:11423–11436, 2022.

Simon Batzner, Albert Musaelian, Lixin Sun, Mario Geiger, Jonathan P Mailoa, Mordechai Kornbluth, Nicola Molinari, Tess E Smidt, and Boris Kozinsky. E (3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nature communications*, 13(1):2453, 2022.

Vivek Bharadwaj, Austin Glover, Aydın Buluç, and James Demmel. An efficient sparse kernel generator for o (3)-equivariant deep networks. In *2025 Proceedings of the Conference on Applied and Computational Discrete Algorithms (ACDA)*, pp. 32–46. SIAM, 2025.

Chi Ding, Qing Lu, Zhaopeng Guo, Tianheng Huang, Junjie Wang, Yu Han, Dingyu Xing, and Jian Sun. Quasi-2d spin-peierls transition through interstitial anionic electrons in k (nh3) 2. *Science Bulletin*, 69(8):1027–1036, 2024.

Mario Geiger and Tess Smidt. e3nn: Euclidean neural networks. *arXiv preprint arXiv:2207.09453*, 2022.

Cheuk Hin Ho, Cas van der Oord, James P Darby, Theo Keane, Raz L Benson, Cristian Rebolledo Espinoza, Rutvij Kulkarni, Elina Spinu, Michail Papanikolaou, Richard Tomsett, et al. Equivariant many-body message passing interatomic potentials for magnetic materials. *arXiv preprint arXiv:2604.08143*, 2026.

Alexey S Kotykhov, Max Hodapp, Christian Tantardini, Konstantin Kravtsov, Ivan Kruglov, Alexander V Shapeev, and Ivan S Novikov. Actively trained magnetic moment tensor potentials for mechanical, dynamical, and thermal properties of paramagnetic crn. *Physical Review B*, 111(9): 094438, 2025.

Yunyang Li, Lin Huang, Zhihao Ding, Xinran Wei, Chu Wang, Han Yang, Zun Wang, Chang Liu, Yu Shi, Peiran Jin, et al. E2former: An efficient and equivariant transformer with linear-scaling tensor products. *Advances in Neural Information Processing Systems*, 38:23593–23630, 2026.

Aaron Meurer, Christopher P Smith, Mateusz Paprocki, Ondřej Čertík, Sergey B Kirpichev, Matthew Rocklin, AMiT Kumar, Sergiu Ivanov, Jason K Moore, Sartaj Singh, et al. Sympy: symbolic computing in python. *PeerJ Computer Science*, 3:e103, 2017.

NVIDIA. cuEquivariance. Software, 2025. URL https://github.com/NVIDIA/cuEquivariance.

Saro Passaro and C Lawrence Zitnick. Reducing so (3) convolutions to so (2) for efficient equivariant gnns. In *International conference on machine learning*, pp. 27420–27438. PMLR, 2023.

Janosh Riebesell, Rhys EA Goodall, Philipp Benner, Yuan Chiang, Bowen Deng, Gerbrand Ceder, Mark Asta, Alpha A Lee, Anubhav Jain, and Kristin A Persson. A framework to evaluate machine learning crystal stability predictions. *Nature Machine Intelligence*, 7(6):836–847, 2025.

Matteo Rinaldi, Matous Mrovec, Anton Bochkarev, Yury Lysogorskiy, and Ralf Drautz. Non-collinear magnetic atomic cluster expansion for iron. *npj Computational Materials*, 10(1):12, 2024.

Junjie Wang, Yijie Zhu, Zhongwei Zhang, Zhiyue Guo, Xudong Zhu, Lixin He, Chi Ding, and Jian Sun. Cartesian tensor equivariant machine-learning force field for spin-dependent atomistic simulations. *arXiv preprint arXiv:2608.30338*, 2026.

Mads C Weber, Mael Guennou, Donald M Evans, Constance Toulouse, Arkadiy Simonov, Yevheniia Kholina, Xiaoxuan Ma, Wei Ren, Shixun Cao, Michael A Carpenter, et al. Emerging spin–phonon coupling through cross-talk of two magnetic sublattices. *Nature communications*, 13(1):443, 2022.

Na Wu, Shengjie Zhang, Daqiang Chen, Yaxian Wang, and Sheng Meng. Three-stage ultrafast demagnetization dynamics in a monolayer ferromagnet. *Nature Communications*, 15(1):2804, 2024.

Zemin Xu, Wenbo Xie, and P Hu. Edge cluster expansion with radial rotary attention for interatomic potentials. *arXiv preprint arXiv:2607.10664*, 2026.

Haiyang Yu, Yuchao Lin, Xuan Zhang, Xiaofeng Qian, and Shuiwang Ji. Efficient prediction of SO(3)-equivariant hamiltonian matrices via SO(2) local frames. In *Forty-third International Conference on Machine Learning*, 2026. URL https://openreview.net/forum?id=f9i6VLHT5v.

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.

Daye Zheng, Xingliang Peng, Yike Huang, Yinan Wang, Duo Zhang, Zhengtao Huang, Zefeng Cai, Linfeng Zhang, Mohan Chen, Ben Xu, et al. Integrating deep-learning-based magnetic model and non-collinear spin-constrained method: methodology, implementation and application. *npj Computational Materials*, 12(1):52, 2026.

A NOTATION

$O(d),\ SO(d)$	Orthogonal and special orthogonal groups, $d \in \{2, 3\}$.
$\mathbb{Z}_2^{\mathcal{T}},\ \mathcal{T},\ \tau$	Time-reversal group, operation, and sign.
$G_{\mathrm{SOC}},\ G_{\text{non-SOC}}$	SOC and non-SOC symmetry groups.
$Q,\ Q_r,\ Q_s$	Joint, spatial, and spin transformations.
$H,\ g(Q, \boldsymbol{n})$	Stabilizer of $\boldsymbol{e}_y$ and the induced stabilizer element.
$\ell,\ p,\ t$	Degree, inversion parity, and time parity.
$p_v,\ t_v$	Parities of the input vector.
$m,\ q$	$O(2)$ order and reflection character.
$\rho_{\ell,p},\ \rho_{\mathrm{in}},\ \rho_{\mathrm{out}}$	Irrep and input/output representations.
$R_{\boldsymbol{n}},\ R_x,\ R_y,\ \boldsymbol{e}_y,\ I_3$	Alignment and axis rotations, y-axis unit vector, and identity.
$R_m(\theta),\ \theta,\ \phi$	Order-m rotation, rotation angle, and azimuth.
$S_0,\ F$	Reference reflection and its two-dimensional representation.
J	Change-of-basis matrix.
$B_{\mathrm{in}},\ B_{\mathrm{out}}$	Input/output basis transformations.
$\sigma_{\mathrm{in}},\ \sigma_{\mathrm{out}}$	Restricted representations in the $O(2)$ basis.
$D^{(\ell)},\ D_{ij}$	Degree-ℓ Wigner-D matrix and edge-alignment representation.
$\Phi,\ \mathcal{F}_{\boldsymbol{n}}$	H-equivariant map and induced $O(3)$-equivariant map.
$i,\ j,\ Z_i,\ \mathcal{N}(i)$	Target, source, species, and neighborhood.
$\boldsymbol{r}_i,\ \boldsymbol{r}_{ij},\ \boldsymbol{n}$	Position, relative position, and unit reference direction.
$\boldsymbol{m}_i,\ \boldsymbol{v}$	Magnetic moment and generic vector.
$h,\ h_0,\ h_{\pm m},\ h_m,\ s$	$O(3)$ or $O(2)$ features, depending on the context.
$h_i^{(t)},\ h_{ij}^{(t)}$	Node features and edge-aligned input at layer t.
ν	Product-basis correlation order.
$\mathcal{Y}^{(\ell)},\ \mathcal{R}^{(\ell)},\ Y^{(\ell)}$	Spherical, regular solid, and edge harmonics.
$\mathcal{M}_i,\ \mathcal{M}_{ij}$	Magnetic node and edge attributes.
$\ell_i,\ \ell_j,\ L,\ L_{\mathrm{mag}}$	Input, output, and cutoff degrees; L also denotes the complexity cutoff.
$\widetilde{\mathcal{M}}_{ij}^{(t)}$	Weighted magnetic edge representation.
$W_{Z_i}^{(t)},\ W_{ij}^{(t)}$	Element embedding and magnetic Linear weights.
$\boldsymbol{W}_{\text{conv},ij}^{(t)},\ \odot$	Convolution weights and Hadamard product.
$E,\ \boldsymbol{F}_i,\ \boldsymbol{F}_i^{\mathrm{mag}}$	Energy, atomic force, and magnetic force.
$F,\ F^{\mathrm{mag}},\ S$	Atomic-force, magnetic-force, and stress RMSEs.
$A_{\mathcal{T}},\ A_S,\ A_P$	Time-reversal, spin-rotation, and inversion augmentations.
$L_{\max}$	Node-feature degree cutoff for the product-basis output.
$\ell_{\max}$	Degree cutoff for edge features and interaction intermediates.
$m_{\max}$	Local $O(2)$ order cutoff.
$k,\ u$	Cartesian rank and solid-harmonic split degree.
$f,\ f_i,\ m_{\mathrm{cut},Z,f},\ \mathcal{D}_{Z,f}$	Fidelity indices, magnetic cutoff, and training-set indices.
$u_i^2,\ \widetilde{m}_i$	Scaled squared moment and transformed magnitude.
$T_n,\ T,\ N_{\mathrm{mag}}$	Chebyshev polynomial, radial basis, and basis size.
$E_{1\mathrm{b}},\ W_{Z,f,n}$	One-body energy and element/fidelity coefficients.
$\pi,\ \ell_1,\ \ell_2,\ \ell_3$	Coupling path and input degrees.
$a_j^{(\ell_3)},\ \ell_{12},\ \ell_{23}$	Additional representation and intermediate degrees.
$[\cdot]_L$	Projection onto degree L.
$\Phi_{ij,L}^{(\ell_{12})},\ g_{j,\pi}^{(\ell_{23})},\ A_{i,\pi}^{(L)}$	Edge contraction, recoupled intermediate, and path aggregate.
$\Gamma,\ \widehat{\Gamma}$	Wigner-6j coefficient and normalized coefficient.
$s^{\mathrm{dir}},\ s^{\mathrm{rec}},\ \bar{s}^{\mathrm{dir}},\ \bar{s}^{\mathrm{rec}}$	Path-normalization factors.
$W_{ij,\pi}^{\mathrm{edge}},\ W_{ij,\pi}^{\mathrm{extra}}$	Reference tensor-product weights.

B SYMMETRY FOUNDATIONS

For the real-valued tensor features considered here, an irrep of $O(3) \times \mathbb{Z}_2^{\mathcal{T}}$ is labeled by an angular degree ℓ, a spatial parity p, and a time-reversal parity t. We write this as ℓpt, representing the signs $p, t \in \{-1, +1\}$ by $\circ$ and $\mathtt{e}$, respectively. The first suffix always refers to spatial inversion and the second to time reversal. The full set of $O(3) \times \mathbb{Z}_2^{\mathcal{T}}$ irreps is therefore

$$\bigoplus_{\ell=0}^{\infty} (\ell\mathtt{ee} \oplus \ell\mathtt{eo} \oplus \ell\mathtt{oe} \oplus \ell\mathtt{oo}) \,. \tag{20}$$

Tensor-product paths obey the parity-selection rules $p_{\text{out}} = p_1 p_2$ and $t_{\text{out}} = t_1 t_2$, together with the angular-momentum rule $|\ell_1 - \ell_2| \le \ell_{\text{out}} \le \ell_1 + \ell_2$.

Polar and axial vectors. Let $Q \in O(3)$ and let $\mathcal{T}$ denote time reversal. A relative position is a time-even polar vector, whereas a noncollinear magnetic moment is a time-odd axial vector:

$$\begin{aligned} Q: &\quad \boldsymbol{r}_{ij} \mapsto Q\boldsymbol{r}_{ij}, \quad \boldsymbol{m}_i \mapsto \det(Q) Q \boldsymbol{m}_i, \\ \mathcal{T}: &\quad \boldsymbol{r}_{ij} \mapsto \boldsymbol{r}_{ij}, \quad \boldsymbol{m}_i \mapsto -\boldsymbol{m}_i. \end{aligned} \qquad \boldsymbol{r}_{ij} \sim \mathtt{1oe}, \quad \boldsymbol{m}_i \sim \mathtt{1eo}. \tag{21}$$

After suppressing the time label, $\boldsymbol{r}_{ij}$ and $\boldsymbol{m}_i$ reduce to the familiar $O(3)$ irreps $\mathtt{1o}$ and $\mathtt{1e}$, respectively. In particular, spatial inversion reverses $\boldsymbol{r}_{ij}$ but leaves $\boldsymbol{m}_i$ unchanged, while time reversal has exactly the opposite action.

Spherical harmonics. Consider a nonzero vector $\boldsymbol{v} \sim 1 p_v t_v$, identifying $\mathtt{e}$ with $+1$ and $\circ$ with -1. The degree-ℓ spherical harmonic is the highest-degree irreducible component of $\widehat{\boldsymbol{v}}^{\otimes \ell}$ and hence transforms as $\mathcal{Y}^{(\ell)}(\widehat{\boldsymbol{v}}) \sim \ell\, p_v^{\ell} t_v^{\ell}$. For relative positions and magnetic moments, respectively, this gives

$$\mathcal{Y}(\widehat{\boldsymbol{r}}_{ij}) \sim \mathtt{0ee} \oplus \mathtt{1oe} \oplus \mathtt{2ee} \oplus \mathtt{3oe} \oplus \cdots, \quad \mathcal{R}(\boldsymbol{m}_i) \sim \mathtt{0ee} \oplus \mathtt{1eo} \oplus \mathtt{2ee} \oplus \mathtt{3eo} \oplus \cdots. \tag{22}$$

All degrees constructed from an axial vector are spatially even, while their time parity alternates with ℓ. The corresponding regular solid harmonics $\mathcal{R}^{(\ell)}(\boldsymbol{v}) = \|\boldsymbol{v}\|^{\ell} \mathcal{Y}^{(\ell)}(\widehat{\boldsymbol{v}})$ carry the same spatial and temporal parities, while providing a smooth extension at $\boldsymbol{v} = \mathbf{0}$.

With and without spin–orbit coupling. With spin–orbit coupling (SOC), the same rotation acts on the structure and magnetic moments. In the absence of an external time-reversal-breaking field, the energy transforms as the invariant scalar $\mathtt{0ee}$. Equivalently, for every $Q \in O(3)$ and $\tau \in \{+1, -1\} \simeq \mathbb{Z}_2^{\mathcal{T}}$,

$$E(\{Q\boldsymbol{r}_i, \tau \det(Q) Q \boldsymbol{m}_i, Z_i\}) = E(\{\boldsymbol{r}_i, \boldsymbol{m}_i, Z_i\}), \qquad G_{\text{SOC}} = O(3) \times \mathbb{Z}_2^{\mathcal{T}}. \tag{23}$$

Here, $\tau = -1$ denotes time reversal. Because the same Q acts on both the spatial and spin degrees of freedom, this symmetry allows the energy to depend on the orientation of the magnetic moments relative to the lattice. Without SOC, spatial and spin rotations are independent. For $Q_r \in O(3)$ and $Q_s \in SO(3)$, the symmetry is instead

$$\left.\begin{aligned} E(\{Q_r\boldsymbol{r}_i, Q_s\boldsymbol{m}_i, Z_i\}) &= E(\{\boldsymbol{r}_i, \boldsymbol{m}_i, Z_i\}) \\ E(\{\boldsymbol{r}_i, -\boldsymbol{m}_i, Z_i\}) &= E(\{\boldsymbol{r}_i, \boldsymbol{m}_i, Z_i\}) \end{aligned}\right\}, \qquad \begin{aligned} G_{\text{non-SOC}} &= O(3)_{\text{space}} \times SO(3)_{\text{spin}} \times \mathbb{Z}_2^{\mathcal{T}} \\ &\simeq O(3)_{\text{space}} \times O(3)_{\text{spin}}. \end{aligned} \tag{24}$$

C GENERALIZED WIGNER-6j CONVOLUTION

C.1 DEFINITION

For notational brevity, parity and channel indices are omitted throughout this section. The *generalized Wigner-6j convolution* couples three factors: a source node feature $h_j^{(\ell_2)}$, an edge spherical harmonic $Y^{(\ell_1)}(\widehat{\boldsymbol{r}}_{ij})$, and an independently supplied node representation $a_j^{(\ell_3)}$. The additional representation may contain regular solid harmonics of an applied electric-field direction or an atomic magnetic moment. More generally, it may be any invariant or equivariant node input. An invariant input may enter the convolution weights, whereas a nonzero-degree input must enter the angular contraction

if its orientation is to affect the message. In the direct coupling order, the node feature and edge spherical harmonic first produce degree ℓ_{12}, after which the result is coupled to the additional node representation:

$$\Phi_{ij,L}^{(\ell_{12})} = \left[\left[h_j^{(\ell_2)} \otimes Y^{(\ell_1)}(\widehat{\boldsymbol{r}}_{ij}) \right]_{\ell_{12}} \otimes a_j^{(\ell_3)} \right]_L . \tag{25}$$

Here $[\,\cdot\,]_L$ denotes projection onto the degree-L irreducible component. Each possible coupling path is labeled by the following tuple:

$$\pi = (\ell_1, \ell_2, \ell_{12}, \ell_3, L). \tag{26}$$

For repeated irreps, π also identifies the input index, which we omit here for simplicity. The implementation uses channel-preserving `uvu` tensor products, with unit multiplicity for the edge harmonic and additional node representation. This order is costly because the first product is evaluated after source gather. Its intermediate lives on edges, and the second product further expands edge computation. To make the problem more concrete, we can interpret $a_j^{(\ell_3)}$ here as the regular solid harmonics of the noncollinear magnetic moment. Such a convolution scheme is in fact an essential ingredient in the convolution design of mMACE (Ho et al., 2026). The generalized Wigner-6j convolution rearranges the coupling order so that terms involving only the source atom j are evaluated before source gather. This removes the first edge-level intermediate. Avoiding storage of the final edge message additionally requires a fused tensor-product and aggregation backend, such as OpenEquivariance (Bharadwaj et al., 2025) and CuEquivariance (NVIDIA, 2025).

C.2 Exact Wigner-6j recoupling

We first construct the source-node intermediate

$$g_{j,\pi}^{(\ell_{23})} = \left[h_j^{(\ell_2)} \otimes a_j^{(\ell_3)} \right]_{\ell_{23}} , \tag{27}$$

and then couple it to the edge harmonic:

$$\Phi_{ij,L}^{(\ell_{12})} = \sum_{\ell_{23}} \Gamma^L_{\ell_1\ell_2\ell_{12}\ell_3\ell_{23}} \left[g_{j,\pi}^{(\ell_{23})} \otimes Y^{(\ell_1)}(\widehat{\boldsymbol{r}}_{ij}) \right]_L . \tag{28}$$

For this operand order and the component-normalized CG convention used by `e3nn` (Geiger & Smidt, 2022), the recoupling coefficient is

$$\Gamma^L_{\ell_1\ell_2\ell_{12}\ell_3\ell_{23}} = (-1)^{\ell_1+\ell_3+\ell_{12}+\ell_{23}} \sqrt{(2\ell_{12}+1)(2\ell_{23}+1)} \begin{Bmatrix} \ell_1 & \ell_2 & \ell_{12} \\ \ell_3 & L & \ell_{23} \end{Bmatrix} . \tag{29}$$

Here the braces denote the Wigner-6j symbol, evaluated using `SymPy` (Meurer et al., 2017). For each original path π, Eqs. (25) and (28) are exactly equivalent when all allowed ℓ_{23} with nonzero coefficients are retained. Matching the implemented tensor products also requires accounting for path normalization. Our reference uses `irrep_normalization=component` and `path_normalization=element`. Element normalization depends on the contributing paths and their multiplicities, so its factors can differ between the two coupling trees. Let s_π^{dir} and $s_{\pi,\ell_{23}}^{\mathrm{rec}}$ denote the products of the two generated instruction-normalization factors along the corresponding paths. These are fixed factors, not learned weights. Removing the component-normalization factors gives the element scales

$$\bar{s}_\pi^{\mathrm{dir}} = \frac{s_\pi^{\mathrm{dir}}}{\sqrt{(2\ell_{12}+1)(2L+1)}}, \qquad \bar{s}_{\pi,\ell_{23}}^{\mathrm{rec}} = \frac{s_{\pi,\ell_{23}}^{\mathrm{rec}}}{\sqrt{(2\ell_{23}+1)(2L+1)}}. \tag{30}$$

The coefficient used by the recoupled implementation is consequently

$$\widehat{\Gamma}^L_{\ell_1\ell_2\ell_{12}\ell_3\ell_{23}} = \Gamma^L_{\ell_1\ell_2\ell_{12}\ell_3\ell_{23}} \frac{\bar{s}_\pi^{\mathrm{dir}}}{\bar{s}_{\pi,\ell_{23}}^{\mathrm{rec}}}. \tag{31}$$

Let $W_{ij,\pi}^{\mathrm{edge}}$ and $W_{ij,\pi}^{\mathrm{extra}}$ be the channel-wise scalar weights for the first and second reference tensor products, respectively. In the magnetic interaction, both are generated on edges from the concatenation of the spatial edge features and the source magnetic radial basis $T(\widetilde{m}_j)$. Since scalar weights commute with the angular contractions, their product is applied in the final node-edge tensor product. The

node-node tensor product uses fixed unit weights. Using the implementation-normalized tensor products for g and the final contraction, the aggregated output of each original path is

$$A_{i,\pi}^{(L)} = \sum_{j\in\mathcal{N}(i)} \sum_{\ell_{23}} W_{ij,\pi}^{\text{edge}} W_{ij,\pi}^{\text{extra}} \widehat{\Gamma}^{L}_{\ell_1\ell_2\ell_{12}\ell_3\ell_{23}} \left[g_{j,\pi}^{(\ell_{23})} \otimes Y^{(\ell_1)}(\widehat{\boldsymbol{r}}_{ij}) \right]_L . \tag{32}$$

Distinct original paths retain separate weights and output channels; a subsequent equivariant linear map mixes these outputs.

C.3 Intermediate reuse and execution

The implementation evaluates each distinct combination of source-feature copy, additional-attribute copy, and intermediate irrep once per node. Original paths that share this combination reuse the same $g_{j,\pi}^{(\ell_{23})}$. The intermediates are stored in irrep-sorted order, making equal irrep types contiguous without summing independent channels or requiring a runtime permutation. Recoupling preserves every original path and includes every allowed intermediate contribution with a nonzero coefficient. It removes one edge-level tensor product stage and reuses node intermediates, but does not eliminate the coupling-path growth of global CGTPs. This motivates the local $O(2)$ formulation developed in the main text.

D Data Augmentation

With SOC, spatial and spin transformations are locked: a single $Q \in O(3)_{\text{space+spin}}$ acts on positions as $\boldsymbol{r}_i \mapsto Q\boldsymbol{r}_i$ and on axial magnetic moments as $\boldsymbol{m}_i \mapsto \det(Q)Q\boldsymbol{m}_i$. Without SOC, the two sectors transform independently. An arbitrary $Q_r \in O(3)_{\text{space}}$ acts on the positions, while an independent $Q_s \in SO(3)_{\text{spin}}$ acts on all magnetic moments. Their actions are $\boldsymbol{r}_i \mapsto Q_r\boldsymbol{r}_i$ and $\boldsymbol{m}_i \mapsto Q_s\boldsymbol{m}_i$. The SOC architecture admits the broader function space: it may depend on the orientation of the magnetic moments relative to the lattice. This is required for SOC data but is an unwanted degree of freedom for a non-SOC target. Thus, when the more general SOC architecture is applied to non-SOC data, data augmentation is used to encourage the appropriate non-SOC symmetry when it is not enforced by the architecture. Let $A_{\mathcal{T}}$ denote time reversal, A_P spatial inversion, and A_S a spin rotation. Their actions on positions, magnetic moments, atomic forces, and magnetic forces are

$$A_{\mathcal{T}}: \quad (\boldsymbol{r}_i, \boldsymbol{m}_i, \boldsymbol{F}_i, \boldsymbol{F}_i^{\text{mag}}) \mapsto (\boldsymbol{r}_i, -\boldsymbol{m}_i, \boldsymbol{F}_i, -\boldsymbol{F}_i^{\text{mag}}), \tag{33}$$

$$A_P: \quad (\boldsymbol{r}_i, \boldsymbol{m}_i, \boldsymbol{F}_i, \boldsymbol{F}_i^{\text{mag}}) \mapsto (-\boldsymbol{r}_i, \boldsymbol{m}_i, -\boldsymbol{F}_i, \boldsymbol{F}_i^{\text{mag}}), \tag{34}$$

$$A_S(R): \quad (\boldsymbol{r}_i, \boldsymbol{m}_i, \boldsymbol{F}_i, \boldsymbol{F}_i^{\text{mag}}) \mapsto (\boldsymbol{r}_i, R\boldsymbol{m}_i, \boldsymbol{F}_i, R\boldsymbol{F}_i^{\text{mag}}), \qquad R \in SO(3). \tag{35}$$

For non-SOC data without an external time-reversal-breaking field, the energy is unchanged under all three transformations. For SOC data, the preferred choice is the SOC architecture with complete spatial parity and explicit time-reversal parity; its group is $O(3)_{\text{space+spin}} \times \mathbb{Z}_2^{\mathcal{T}}$ and no augmentation is needed. For non-SOC data, the preferred choice is the explicit non-SOC architecture, whose group is $O(3)_{\text{space}} \times SO(3)_{\text{spin}} \times \mathbb{Z}_2^{\mathcal{T}}$ and likewise requires no augmentation. An SOC architecture should be used for non-SOC data only when its broader function space is deliberately retained and constrained through augmentation.

E Magnetic Cutoff and Basis Normalization

Magnetic cutoff. Let $m_{\text{cut},Z,f} > 0$ denote the magnetic cutoff for element Z and fidelity f. For each fidelity, it may be supplied as a positive scalar shared by all elements or as an element-dependent mapping; supplied values are used directly, without further rescaling. Otherwise, the cutoff is estimated independently from that fidelity's training statistics (Ho et al., 2026):

$$m_{\text{cut},Z,f} = 1.2 \max_{i\in\mathcal{D}_{Z,f}} \|\boldsymbol{m}_i\| + 0.1\,\mu_{\text{B}}, \tag{36}$$

where $\mathcal{D}_{Z,f}$ indexes the training-set moments of element Z at fidelity f, and magnetic moments and cutoffs are expressed in Bohr magnetons μ_{B}. A zero or missing element entry in an available statistics mapping gives a cutoff of $0.1\,\mu_{\text{B}}$. Cutoffs are stored separately for each element and fidelity and remain fixed during training.

Transformed magnitude features. The magnetic cutoff controls only the radial magnitude map, not the angular basis. Let f_i denote the fidelity of atom i's configuration. The implementation forms the squared scaled magnitude $u_i^2 = \sum_{\alpha=x,y,z}(m_{i,\alpha}/m_{\mathrm{cut},Z_i,f_i})^2$. The transformed scalar magnitude $\widetilde{m}_i$ and magnetic node features $T(\widetilde{m}_i)$ are

$$\widetilde{m}_i = 1 - 2\min(u_i^2, 1) \in [-1, 1], \qquad T(\widetilde{m}_i) = \big(T_1(\widetilde{m}_i), \ldots, T_{N_{\mathrm{mag}}}(\widetilde{m}_i)\big), \tag{37}$$

where T_n is the first-kind Chebyshev polynomial, evaluated by the recurrence $T_0(x) = 1$, $T_1(x) = x$, and $T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x)$ for $n \geq 1$. The parameter N_{mag} counts the returned nonconstant basis functions: $T_1, \ldots, T_{N_{\mathrm{mag}}}$. The radial constant T_0 is excluded from the interactions.

Smoothness and saturation. Saturation at $\|\boldsymbol{m}_i\| = m_{\mathrm{cut},Z_i,f_i}$ is intentional: it keeps the Chebyshev input within $[-1, 1]$ beyond the chosen magnetic range. Inside the open domain $\|\boldsymbol{m}_i\| < m_{\mathrm{cut},Z_i,f_i}$, the magnitude map is a polynomial in the magnetic components, the Chebyshev basis is polynomial, and the regular solid harmonics are Cartesian polynomials. With smooth network activations, the energy is therefore smooth with respect to the magnetic moments, including at zero. Magnetic-structure optimization confined to this domain encounters no discontinuity introduced by these basis functions. The hard saturation is continuous but generally not differentiable at the cutoff; crossing it can produce a jump in magnetic forces. The released magnetic radial implementations of mMACE (Ho et al., 2026) and HotPP-spin (Wang et al., 2026) likewise hard-clamp the moment magnitude at an upper cutoff. Such possible force discontinuities when crossing the cutoff are therefore an expected consequence of hard saturation, rather than a limitation specific to the local $O(2)$ formulation. The cutoff should therefore be chosen above the moment range explored in optimization.

Angular basis. The angular representation is evaluated on the original magnetic vector $\boldsymbol{m}_i$, without applying the magnetic cutoff:

$$\mathcal{M}_i^{(\ell)} = \mathcal{R}^{(\ell)}(\boldsymbol{m}_i) = \|\boldsymbol{m}_i\|^\ell \mathcal{Y}^{(\ell)}(\widehat{\boldsymbol{m}}_i), \qquad 0 \leq \ell \leq L_{\mathrm{mag}}. \tag{38}$$

The implementation uses integral normalization, multiplying the component-normalized solid harmonics by $1/\sqrt{4\pi}$. It evaluates the Cartesian polynomials directly, without division by the moment magnitude or element- or fidelity-dependent angular rescaling. Each $\ell > 0$ irrep is therefore homogeneous of degree ℓ and vanishes smoothly at zero, while the angular scalar $\mathcal{M}_i^{(0)} = 1/\sqrt{4\pi}$ is retained. The $\ell > 0$ angular features are unbounded and scale as $\|\boldsymbol{m}_i\|^\ell$.

One-body magnetic readout. To make the physical interpretation of the energy decomposition explicit, one may additionally impose the readout constraints proposed in mMACE (Ho et al., 2026); we do not discuss this extension further here. When enabled, the magnitude-only one-body readout is an element-dependent linear map with one output head per fidelity, no hidden layers, and no separate bias. Selecting the head for each configuration gives

$$E_{1\mathrm{b}} = \sum_i \left(W_{Z_i,f_i,0} + \sum_{n=1}^{N_{\mathrm{mag}}} W_{Z_i,f_i,n} T_n(\widetilde{m}_i) \right). \tag{39}$$

The constant input $T_0 = 1$ is appended only at readout time and supplies the independent coefficient $W_{Z_i,f_i,0}$. The element- and fidelity-dependent atomic-energy baseline is added separately.

F Computational Complexity

EquivariantX supports arbitrary direct sums of `e3nn` irreps with entry-dependent multiplicities through degree-wise Wigner-D rotations. Entries with equal multiplicity and distinct degrees are grouped into shared rotation applications to reduce kernel-launch overhead. At a fixed degree, the proper-rotation matrices are independent of spatial and time-reversal parities and can be reused across these sectors. The fixed basis change can also be absorbed into the corresponding frame matrices, together with its inverse in the local-to-global transformation, rather than applied separately. The additional global CGTP coupling paths introduced by complete $O(3)$ and time-reversal labels further motivate the local-frame approach, which retains order-wise operations and cubic theoretical angular complexity instead of explicitly evaluating all global angular couplings.

For maximum angular degree L and fixed channel multiplicity, a dense CGTP has angular complexity $\mathcal{O}(L^6)$; exploiting the sparsity of the Clebsch–Gordan coefficients reduces it to $\mathcal{O}(L^5)$. In contrast, the $SO(2)$ formulation of eSCN (Passaro & Zitnick, 2023) has $\mathcal{O}(L^3)$ angular complexity. The analysis for our local $O(2)$ frame proceeds analogously.

With all local orders retained, the degree-ℓ Wigner-D matrix acts on $2\ell+1$ angular features, giving quadratic cost in $2\ell+1$ per rotation. At local order m, each degree $\ell = m, \ldots, L$ contributes one copy of the irrep, so a Linear mixes channel widths proportional to $L - m + 1$. Summing over degrees and orders gives Wigner-D rotations: $\mathcal{O}\left(\sum_{\ell=0}^{L}(2\ell+1)^2\right) = \mathcal{O}(L^3)$, and Local Linear: $\mathcal{O}\left(\sum_{m=0}^{L}(L-m+1)^2\right) = \mathcal{O}(L^3)$. The one- and two-dimensional real irrep dimensions contribute only constant factors. Element-wise activations and gate multiplication cost $\mathcal{O}(L^2)$; generating the scalar gates from local invariant features costs at most $\mathcal{O}(L^3)$. For unnatural-parity representations, the additional fixed orthogonal basis change exchanges and changes signs of the two real features at each positive order. Its cost is $\mathcal{O}(\sum_{\ell=0}^{L}\ell) = \mathcal{O}(L^2)$, leaving the total theoretical angular complexity $\mathcal{O}(L^3)$. Complete spatial parity and time-reversal labels change block multiplicities without changing this exponent.

G EquivariantX

G.1 Local O(2) Convolution

Listing 1 constructs a minimal edge-aligned local $O(2)$ convolution and checks its global $O(3)$ equivariance under rotations, inversion, and their composition. The global input/output irreps and test transformations are created with `e3nn.o3`; `eqx.o2` provides the Wigner-D matrices, local frames, and `linear_up` → `o2_gate` → `linear_down` pipeline. The example includes both spatial parities at every degree up to $\ell = 2$, with 64 channels per global irrep, so it exercises the basis changes for unnatural-parity features and their mixing with natural-parity features. Edge-gathered node features and messages have shape `(num_edges, 18, 64)` and are flattened to `(num_edges, 1152)` in `ir_mul` order. The restricted representation has 192 copies each of `0e` and `0o`, 256 copies of `1m`, and 128 copies of `2m`. The gate applies SiLU to `0e`, tanh to `0o`, and sigmoid-activated `0e` gates to positive-order irreps. Rotations act on the angular axis, and all computations use float64.

Listing 1: Executable global-$O(3)$/local-$O(2)$ equivariance check.

```python
import torch
from e3nn import o3
from eqx import o2

torch.set_default_dtype(torch.float64)
torch.manual_seed(0)
num_edges = 16
num_channels = 64
lmax = 2
angular_irreps = o3.Irreps("0e + 0o + 1o + 1e + 2e + 2o")
irreps_in = o3.Irreps([(num_channels, ir) for _, ir in angular_irreps])
irreps_out = irreps_in

local_frame_in = o2.LocalFrame(irreps_in, mmax=lmax)
local_frame_out = o2.LocalFrame(irreps_out, mmax=lmax, reverse=True)
wignerd = o2.WignerD(mmax=lmax, lmax=lmax)

o2_gate = o2.Gate(
    f"{3 * num_channels}x0e + {3 * num_channels}x0o",
    [torch.nn.SiLU(), torch.nn.Tanh()],
    f"{6 * num_channels}x0e", [torch.nn.Sigmoid()],
    f"{4 * num_channels}x1m + {2 * num_channels}x2m",
)
linear_up = o2.Linear(local_frame_in.irreps_out, o2_gate.irreps_in)
linear_down = o2.Linear(o2_gate.irreps_out, local_frame_out.irreps_out)
```

```


@torch.no_grad()
def o2_interaction(node_feats, edge_vector):
    wigner, wigner_inv = wignerd(edge_vector)
    node_features = local_frame_in.to_local(node_feats.flatten(-2),
    wigner)
    projected = linear_up(node_features)
    hidden = o2_gate(projected)
    message = linear_down(hidden)
    return local_frame_out.to_global(message, wigner_inv).view_as(
    node_feats)


edge_vector = torch.randn(num_edges, 3)
node_feats = torch.randn(num_edges, angular_irreps.dim, num_channels)
rotation = o3.rand_matrix()

message = o2_interaction(node_feats, edge_vector)
for name, Q in (
    ("rotation", rotation),
    ("inversion", -torch.eye(3)),
    ("rotation + inversion", -rotation),
):
    representation = angular_irreps.D_from_matrix(Q)
    transformed_message = o2_interaction(
        representation @ node_feats,
        edge_vector @ Q.T,
    )
    expected_message = representation @ message
    torch.testing.assert_close(
        transformed_message, expected_message,
        atol=1.0e-10, rtol=1.0e-10,
    )
    error = (transformed_message - expected_message).abs().max().item()
    print(f"{name}: max error = {error:.3e}")
```

G.2 O(2) Tensor product

Listing 2 uses `o2.TensorProduct` directly, without frame rotations, to evaluate $\texttt{1m} \otimes \texttt{1m} = \texttt{0e} \oplus \texttt{0o} \oplus \texttt{2m}$ with 64 channels per irrep. Each instruction specifies `(i_in1, i_in2, i_out, mode, train)`: the indices select irrep entries, `uuu` couples matching channels, and `True` enables a learnable weight for each channel on that path. By default, these weights are stored internally and shared across the batch; the Clebsch–Gordan coefficients are fixed. Using `uvw` instead enables dense mixing of input and output channels. Each irrep entry uses flattened `ir_mul` storage, giving input shapes `(16, 128)` and output shape `(16, 256)`.

Listing 2: Channel-wise $O(2)$ tensor product with learnable path weights.

```
import torch
from eqx import o2

tp = o2.TensorProduct(
    irreps_in1="64x1m",
    irreps_in2="64x1m",
    irreps_out="64x0e + 64x0o + 64x2m",
    instructions=[
        (0, 0, 0, "uuu", True),  # 1m x 1m -> 0e
        (0, 0, 1, "uuu", True),  # 1m x 1m -> 0o
        (0, 0, 2, "uuu", True),  # 1m x 1m -> 2m
    ],
)
h1 = torch.randn(16, 2, 64).flatten(-2)
```

```
h2 = torch.randn(16, 2, 64).flatten(-2)
h_out = tp(h1, h2)
assert h_out.shape == (16, 256)
print(h_out.shape)
```

H TRAINING HYPERPARAMETERS

Table 3 summarizes the $O(2)$-mTACE configurations used for CrN and Fe-DeepSpin. We use training, validation, and test sets with an $80\%/10\%/10\%$ split; the validation set is reserved for model selection and the test set for final evaluation. For Fe-DeepSpin, this split applies only to Fe_{16}. HotPP-spin (Wang et al., 2026) uses 30,806 Fe_{16} configurations for training and 3,422 for testing, approximately a $90\%/10\%$ train/test split. The mMACE CrN results (Ho et al., 2026) are evaluated on the original benchmark test set (Kotykhov et al., 2025); its paper does not specify a numerical train/validation/test split for this benchmark. We regard our strict three-way split as a conservative and fair literature comparison. This does not establish identical test samples across studies, and the reported baseline errors retain their original evaluation protocols. All our models were trained in float32 precision on a single NVIDIA GeForce RTX 4090 GPU.

Table 3: Hyperparameters used for our training runs on the two magnetic benchmarks.

Hyperparameter	CrN	Fe-DeepSpin
Training split	80%	80%
Validation split	10%	10%
Test split	10%	10%
Optimizer	AdamW	AdamW
Initial learning rate	1×10^{-3}	1×10^{-3}
Learning-rate scheduler	ReduceLROnPlateau	ReduceLROnPlateau
Learning-rate reduction factor	0.5	0.5
Learning-rate patience	25	10
Early-stopping patience	50	20
Total batch size	4	8
Weight decay	1×10^{-8}	1×10^{-8}
Loss properties	$E/F/S/F^{\mathrm{mag}}$	$E/F/S/F^{\mathrm{mag}}$
Loss coefficients	1/5/1/10	1/5/1/10
Loss types	MSE/MSE/MSE/MSE	MSE/MSE/MSE/MSE
Gradient clipping norm threshold	10	10
Model EMA decay	0.999	0.999
Spatial cutoff radius (Å)	5	5
Number of layers	6	4
Interaction channels	64	48
Radial Rotary Complex Attention	Enabled	Enabled
$L_{\max}$	2	2
$\ell_{\max}$	2	2
$m_{\max}$	2	2
L_{mag}	2	2
Correlation order	2	2