Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Equivariant Many-body Message Passing Interatomic Potentials for Magnetic Materials

Cheuk Hin Ho
University of British Columbia
Cas van der Ord
Symmetric Group
James Darby
Symmetric Group
Theo Keane
MatNex
Raz Benson
MatNex
Cristian Rebolledo Espinoza
MatNex
Rutvij Kulkarni
MatNex
Elina Spinu
MatNex
Michail Papanikolaou
MatNex
Robert M. Forrest
MatNex
Jonathan Bean
MatNex
G´abor Cs´anyi
Max Planck Institute for Polymer Research
Christoph Ortner
christoph.ortner@ubc.ca

University of British Columbia

Article

Keywords:

Posted Date: November 20th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11670201/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: Competing interest reported. GC, CO, CvdO, JD are partners of Symmetric Group LLP that licenses force fields commercially. GC also has equity interest in Angstrom AI.

Equivariant Many-body Message Passing Interatomic Potentials for Magnetic Materials

Cheuk Hin Ho[1,2], Cas van der Oord[2], James P. Darby[2], Theo Keane[3], Raz L. Benson[3], Cristian Rebolledo Espinoza[3], Rutvij Kulkarni[3], Elina Spinu[3], Michail Papanikolaou[3], Richard Tomsett[3], Robert M. Forrest[3], Jonathan J. Bean[3,*], Gábor Csányi[2,4,5,*], Christoph Ortner[1,2,*]

[1] *Department of Mathematics, University of British Columbia*

[2] *Symmetric Group, 71-75 Shelton Street, Covent Garden, London, WC2H 9JQ, United Kingdom*

[3] *MatNex, Salisbury House, Station Road, Cambridge, CB1 2LA, United Kingdom*

[4] *Engineering Laboratory, University of Cambridge, Trumpington St, Cambridge, United Kingdom*

[5] *Max Planck Institute for Polymer Research, Ackermannweg 10, Mainz, 55128, Germany*

(Dated: November 9, 2026)

Magnetism governs key properties of materials used in energy, data storage, and spintronic technologies, yet its complex coupling to lattice and electronic degrees of freedom challenges conventional first-principles approaches. We introduce an equivariant message-passing graph neural network that embeds atomic magnetic moments as explicit degrees of freedom, enabling the learning of magnetic interactions beyond collinear approximations. The model learns physically consistent and transferable representations of magnetic behaviour and can incorporate spin–orbit coupling, achieving near density-functional-theory accuracy with strong data efficiency across diverse magnetic systems by fine-tuning from a pre-trained model. Applications to structural transformations, finite-temperature magnetic phenomena, and materials screening for strongly spin-orbit coupled materials demonstrate transferable magnetic behaviour, establishing a practical foundation for data-driven, high-throughput discovery of complex magnetic materials.

I. INTRODUCTION

Magnetism plays a fundamental role in determining the properties of a wide range of materials central to modern technology, including structural alloys, permanent magnets, and functional electronic and spintronic systems. These applications span information storage [31, 55], turbine generators [27, 49], electronic devices [15], and robotic systems [10, 57]. They increasingly extend to the emerging study of two-dimensional materials, where magnetic effects enable qualitatively new physical behavior [1]. Across these systems, magnetic degrees of freedom are not a perturbative correction but give rise to qualitatively distinct physical behavior.

Accurately describing such behavior requires a consistent treatment of both atomic structure and electronic degrees of freedom, including spin. In particular, the coupling between atomic positions and magnetic moments gives rise to a complex, high-dimensional energy landscape with multiple competing configurations. This makes predictive modeling of magnetic materials inherently challenging, especially when exploring different magnetic orderings or thermodynamic conditions.

Density functional theory (DFT) has been the primary computational tool for capturing magnetic effects, but its high computational cost limits simulations to small systems and short time scales, restricting the study of finite-temperature and mesoscale phenomena.

Recent advances in machine-learned interatomic potentials (MLIPs) have extended atomistic simulations to substantially larger length and time scales while retaining near-DFT accuracy. Both descriptor-based [21, 46, 53] and neural-network-based [61] approaches have been developed for magnetic systems. Domina et al. [20] extended spectral neighbor descriptors to vector fields, enabling machine-learning potentials that include spin. Machine-learned spin–lattice potentials for Fe have also been constructed by combining machine-learned interatomic potentials with parametrized classical spin Hamiltonians [11, 44]. Universal potentials such as CHGNet [19] use magnetic moments as auxiliary prediction targets that inform the latent features, but do not treat them as input degrees of freedom. Behler and co-workers [22] introduced spin-dependent atom-centered symmetry functions within high-dimensional neural-network potentials, while Xu et al. [58] developed a spin-informed graph neural network; both formulations are restricted to collinear magnetism. Yu et al. [61] introduced SpinGNN, based on Heisenberg-edge and spin-distance-edge graph neural networks, in which magnetic interactions are encoded through scalar combinations such as spin dot products. Its extension, SpinGNN++ [60], incorporates time-reversal-equivariant tensor features and explicit spin–lattice interaction terms, providing a more general treatment of non-collinear magnetism. Rinaldi et al. [53] developed a systematic non-collinear magnetic atomic cluster expansion, but considered systems without spin–orbit coupling and relied on constrained-DFT [9, 18] reference calculations to sample magnetic configurations. DeepSPIN [59, 63] instead introduced a pseudo-atom representation in which structural and magnetic degrees of freedom are encoded as coupled features. More recently, MagNet [62] introduced an $E(3)$-equivariant message-passing architecture with distinct

* christoph.ortner@ubc.ca; gc121@cam.ac.uk; jonathan@materialsnexus.com

atomic and magnetic interaction blocks, embedding both the magnitude and orientation of local magnetic moments within tensor-product features.

In this work, we present an equivariant message-passing graph neural network architecture (mMACE) that introduces atomic magnetic moments as explicit equivariant degrees of freedom. Our implementation builds on the MACE architecture [3, 6], but the modifications we introduce are generic and applicable to other frameworks. mMACE differs from these approaches in three respects. (i) *Symmetry:* mMACE treats atomic positions and magnetic moments within a single jointly $O(3)$-equivariant representation that accommodates SOC, with time-reversal symmetries enforced exactly through a lightweight symmetrization (Section II B). Unlike approaches that incorporate explicitly parameterized spin-lattice interactions, such as SpinGNN++ [58, 61], mMACE learns magnetic interactions through general equivariant many-body features without prescribing specific interaction forms. (ii) *Magnetic features:* each edge couples bond-direction spherical harmonics with solid harmonics of the neighbouring moment, so moment orientation and magnitude enter on the same footing as the geometry, rather than through pseudo-atoms (DeepSPIN [59]), explicit spin-Hamiltonian terms (SpinGNN++), or separate magnetic interaction blocks (MagNet [62]). (iii) *Body order:* symmetric tensor products of these joint features, as in magnetic ACE [21], include mixed spin–lattice correlations systematically up to order ν per layer, whereas MagNet stacks two-body tensor-product layers. Finally, mMACE provides one of the first large-scale pre-trained MLIPs explicitly parameterized by atomic magnetic moments, trained on MATPES and MP-ALOE, and demonstrates their use as starting points for data-efficient fine-tuning on specific magnetic systems. The magnetic moments transform under the symmetry group appropriate to the physics—$O(3) \times \mathbb{Z}_2^{\mathcal{T}}$ when spin–orbit coupling is present (or, $O_{\text{space}}(3) \times O_{\text{spin}}(3)$ when it is not) yielding a systematically improvable description of non-collinear magnetism and spin–orbit effects that includes collinear magnetism as a special case, all at modest additional computational cost.

We demonstrate the accuracy and broad applicability of mMACE across a range of magnetic phenomena: On collinear benchmarks (FeAl, CrN), mMACE reduces force and stress errors by factors of three to five relative to existing magnetic machine-learned potentials (c.f. Section III A). When trained on large foundational datasets (MATPES, MP-ALOE), the model systematically improves on magnetically active configurations with comparable accuracy on non-magnetic configurations, and serves as an effective starting point for targeted fine-tuning with minimal additional data (c.f. Section III B): a small number of configurations suffices to recover the FeNi Bain-path energetics across magnetic states (c.f. Section III C), and the frustrated non-collinear ground state of Mn_3Pt is reliably found from random initial spin orientations that are outside the training dataset (c.f. Section III D). Beyond energies and forces, the framework enables prediction of derived magnetic properties: spin–orbit coupling is incorporated to resolve magnetocrystalline anisotropy at the sub-meV scale across chemically diverse structures (c.f. Section III E), Heisenberg exchange constants are extracted in quantitative agreement with first-principles references, and mMACE-driven spin Monte Carlo simulations, which include many-body magnetic interactions explicitly, yield Curie temperatures in closer agreement with experiment than pairwise classical Heisenberg models parameterized from the same model (c.f. Section III F).

II. METHODS

A. Incorporating magnetism into EMPNNs

Equivariant message passing neural networks (EMPNNs) have been widely used to design MLIPs as surrogate models for DFT [6–8, 25, 28, 54]. We outline a framework for EMPNN parameterization of potential energy surfaces that include atomic magnetic moments $\mathbf{m}_i$ as atomic features. They can be thought of as a coarse-grained projection of the spin density onto atomic sites, either via a specified partitioning scheme or latent variables that are implicitly defined during model training.

An atomic structure is described by a graph with nodes indexed by i representing atoms and edges (i, j) specifying an atomic neighbourhood relation $j \in \mathcal{N}(i)$. In each layer t of an EMPNN, the *state* of a node (atom) i is described by a tuple

$$\sigma_i^{(t)} := (\mathbf{r}_i, \mathbf{m}_i, z_i, h_i^{(t)}), \tag{1}$$

where $\mathbf{r}_i \in \mathbb{R}^3$ denotes the atomic position, $\mathbf{m}_i \in \mathbb{R}^3$ the (generally non-collinear) atomic magnetic moment, $z_i \in \mathbb{Z}$ the atomic number (chemical element) and $h_i^{(t)}$ the hidden features at layer t, which can be scalar, vector or higher order. The atomic neighbourhood, denoted by $\mathcal{N}(i)$, is the set of all atoms j within a cutoff radius r_{cut} from atom i. At each message passing step, a message $\mathcal{M}_i^{(t)}$ is constructed by aggregating information from neighbouring atoms

$$\mathcal{M}_i^{(t)} := \bigoplus_{j \in \mathcal{N}(i)} M_t(\sigma_i^{(t)}, \sigma_j^{(t)}) \tag{2}$$

where M_t is a learnable message passing function and $\bigoplus_{j\in\mathcal{N}(i)}$ is a learnable permutation invariant pooling operation. At each update step, the message $\mathcal{M}_i^{(t)}$ is then transformed into a new feature vector for the next layer through a learnable update function U_t

$$h_i^{(t+1)} := U_t(\sigma_i^{(t)}, \mathcal{M}_i^{(t)}). \tag{3}$$

After $t_{\max}$ layers of message passing and updates (in our setting, $t_{\max} = 2$ is often sufficient), learnable readout functions $\mathcal{R}_t$ map the node states to interaction site energies (atomic energies)

$$E_i^{\text{inter}} := \sum_{t=1}^{t_{\max}} \mathcal{R}_t(h_i^{(t)}). \tag{4}$$

Finally, the total potential energy is obtained by summing the contributions from all sites,

$$E := \sum_i E_i^{\text{inter}} + \sum_i E_{0,z_i}(|\mathbf{m}_i|) \tag{5}$$

and one-body contributions E_0 that depend only on the atomic magnetic moments. Forces are obtained by taking the negative gradients with respect to atomic positions,

$$F_i := -\nabla_{\mathbf{r}_i} E\left(\{\mathbf{r}_i, \mathbf{m}_i, z_i)\}_i\right). \tag{6}$$

The corresponding magnetic forces on the atomic magnetic moment of atom i are given by

$$F_i^{\text{mag}} := -\nabla_{\mathbf{m}_i} E\left(\{\mathbf{r}_i, \mathbf{m}_i, z_i)\}_i\right). \tag{7}$$

These gradients enable deterministic, gradient-based optimization of the magnetic degrees of freedom to obtain the equilibrated (self-consistent) total energy. For periodic systems, stresses (or virials) on the unit cell can be derived in an analogous fashion to typical MLIPs.

Evaluating the energy in a configuration $\{\mathbf{r}_i, \mathbf{m}_i, z_i\}_i$ with non-zero magnetic forces $\nabla_{\mathbf{m}_i} E \neq 0$ effectively corresponds to a constrained DFT calculation, in which the magnetic moments are held fixed at non-equilibrium orientations or magnitudes. In practice, effective forces acting on atoms may be defined in different ways depending on the intended application. For example, in equilibrium geometry optimization, one typically relaxes the magnetic degrees of freedom via self-consistent cycles (at fixed atomic positions) and evaluates nuclear forces on the relaxed surface. In contrast, for spin-lattice dynamics, one uses the instantaneous forces evaluated at the evolving spin configuration. We note that local magnetic moments are not uniquely defined: they depend on the scheme used to partition the magnetization density among atoms, which differs between the electronic-structure codes used in this work (VASP for the pretraining datasets; GPAW and Quantum ESPRESSO for the fine-tuning datasets). The moments seen during pretraining and fine-tuning are therefore not defined identically. However, every individual training run uses data computed with a single code and partitioning scheme, so the moments are defined consistently within each training set. A systematic study of the effect of the partitioning scheme is not part of this work.

B. Equivariance and spin-orbit coupling

$O(3)$-equivariant MPNNs utilize architectures for $M_t, U_t, \mathcal{R}_t$ such that the hidden features $h_i^{(t)}$ and outputs E_i transform under rotation in a precise, consistent way. In most MPNN based MLIPs, a rotation Q acts only on the positions $\mathbf{r}_i \mapsto Q\mathbf{r}_i$. Introducing magnetic degrees of freedom means that rotations can now also act on the atomic magnetic moments $\mathbf{m}_i$. Therefore we need to determine the symmetry relations that the outputs E_i should satisfy under various operations involving these new degrees of freedom. The symmetry analysis below follows that of the magnetic atomic cluster expansion [21, 53], where positions and magnetic moments are treated on an equal footing; we restate it here to fix notation and to motivate the architectural choices of mMACE.

When spin-orbit coupling is present, since magnetic moments transform as axial vectors, the symmetry under a transformation $Q \in O(3)$ is

$$E\big(\{Q\mathbf{r}_i, \det(Q)\, Q\mathbf{m}_i, z_i\}_i\big) = E\big(\{\mathbf{r}_i, \mathbf{m}_i, z_i\}_i\big), \tag{8}$$

as well as time reversal symmetry

$$E\big(\{\mathbf{r}_i, -\mathbf{m}_i, z_i\}_i\big) = E\big(\{\mathbf{r}_i, \mathbf{m}_i, z_i\}_i\big). \tag{9}$$

This is equivalent to having the following symmetries, for $S \in SO(3)$,

$$\begin{aligned} E\big(\{S\mathbf{r}_i, S\mathbf{m}_i, z_i\}_i\big) &= E\big(\{\mathbf{r}_i, \mathbf{m}_i, z_i\}_i\big), \\ E\big(\{\mathbf{r}_i, -\mathbf{m}_i, z_i\}_i\big) &= E\big(\{\mathbf{r}_i, \mathbf{m}_i, z_i\}_i\big), \\ E\big(\{-\mathbf{r}_i, \mathbf{m}_i, z_i\}_i\big) &= E\big(\{\mathbf{r}_i, \mathbf{m}_i, z_i\}_i\big). \end{aligned} \tag{10}$$

In the absence of spin-orbit coupling, the spatial and spin reference frames are independent. The energy therefore possesses the larger product-group symmetry

$$E\big(\{Q\mathbf{r}_i, S\mathbf{m}_i, z_i\}_i\big) = E\big(\{\mathbf{r}_i, \mathbf{m}_i, z_i\}_i\big), \tag{11}$$

where $Q \in O_{\text{space}}(3)$ and $S \in SO_{\text{spin}}(3)$. In the non-relativistic limit, the relevant symmetry can be written as

$$O(3)_{\text{space}} \times SO(3)_{\text{spin}} \times \mathbb{Z}_2^{\mathcal{T}} \;\cong\; O(3)_{\text{space}} \times O(3)_{\text{spin}}.$$

Physically, the additional symmetry means that the orientation of the spin frame can be changed independently of the atomic geometry. However, as we explain in Section II C, imposing this symmetry explicitly leads to much larger intermediate tensors in the model, which is computationally unfavourable.

This decision leads to different group symmetries and hence different choices in the design of the EMPNN model architecture. An architecture *with* SOC can be obtained as a relatively straightforward extension of existing $O(3)$-equivariant MPNN architectures, whereas an architecture *without* SOC requires a more careful extension to account for the product group $O(3) \times O(3)$. In the next section we explain the architecture for the SOC case and include the extension to a non-SOC architecture in Appendix F.

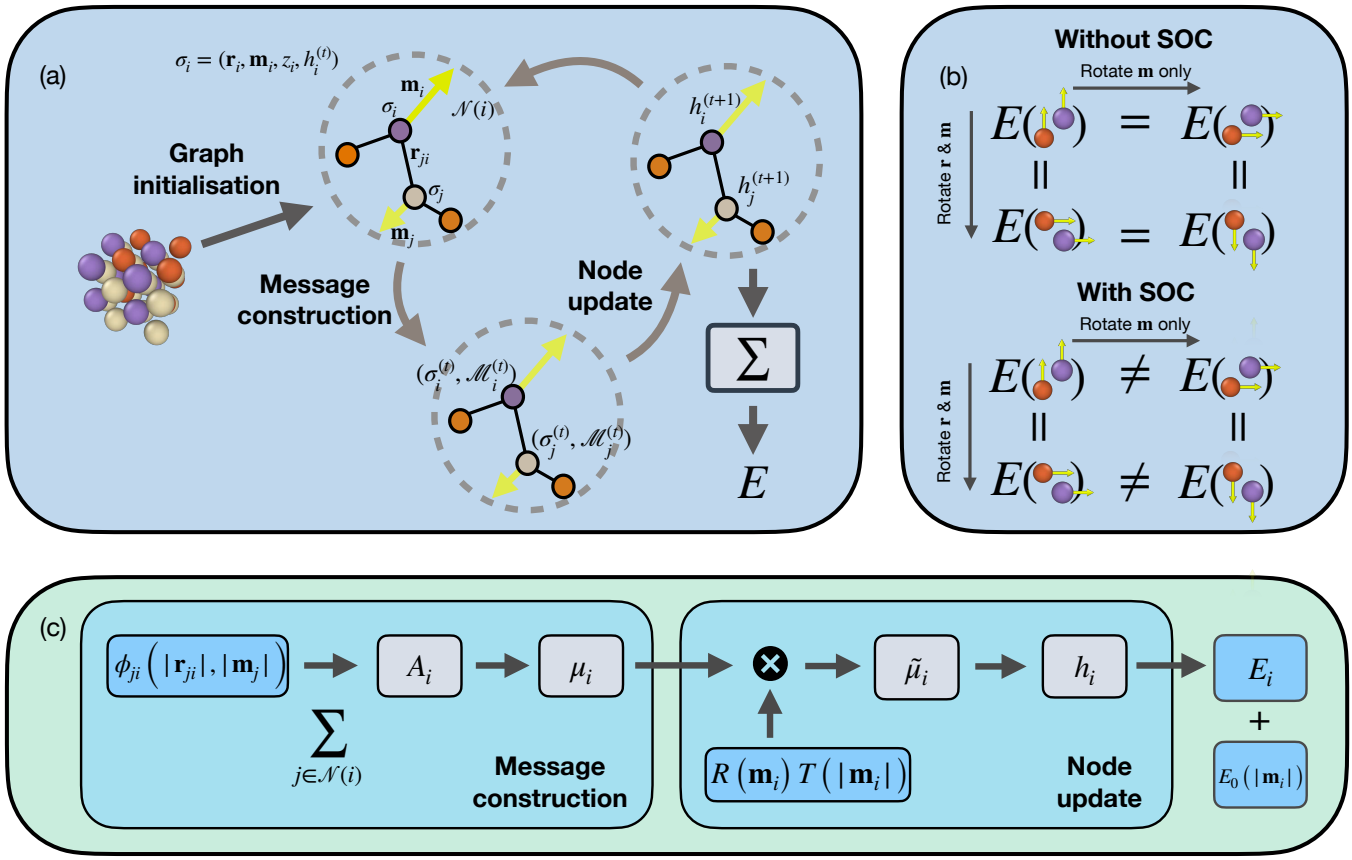


FIG. 1: Schematic overview of the magnetic message-passing architecture in mMACE. (a) Graph construction and message-passing process, where each node represents an atom with associated position, species, and magnetic moment. Messages exchanged along atomic edges encode magnetic moment-dependent interactions and are aggregated to produce local atomic energies (b) Inclusion of spin–orbit coupling (SOC) breaks rotational degeneracy, enabling the model to capture magnetocrystalline anisotropy. (c) Neural network architecture of mMACE, consisting of coupled positional and magnetic message-construction blocks and node-update operations. The network outputs atomic energies, forces, magnetic forces, and stress tensors while maintaining full rotational equivariance.

C. Model architecture

Building on the foregoing symmetry considerations, we introduce the spin-coupled magnetic MACE (mMACE) architecture, which is equivariant under joint rotations of positions and magnetic moments, *i.e.*, satisfying the first two relations in (10). This choice maintains the flexibility to capture SOC effects while keeping the architecture a straightforward extension of the standard MACE framework. The model is naturally invariant under joint spatial–spin rotations and reflections. i.e.

$$E\big(\{-\mathbf{r}_i, -\mathbf{m}_i, z_i\}_i\big) = E\big(\{\mathbf{r}_i, \mathbf{m}_i, z_i\}_i\big). \qquad (12)$$

We do not encode the magnetic moment directly as an axial vector, as this increases the dimension of the irreducible representation space from a single parity sector for each angular momentum, e.g. $0e \oplus 1o \oplus 2e \oplus \cdots$, to both parity sectors, $0e \oplus 0o \oplus 1e \oplus 1o \oplus 2e \oplus 2o \oplus \cdots$, thereby approximately doubling the feature dimension and the size of the intermediate tensors. Instead, we enforce (12) through the architecture and one can impose time-reversal symmetry $\mathbb{Z}_2^{\mathcal{T}}$ explicitly to recover the full SOC symmetry (Appendix A). This keeps the intermediate tensors small, and the resulting model still satisfies the full SOC symmetry. When training on data without SOC, the more stringent symmetry (11) can be effectively achieved to within high precision through data augmentation; further details are provided in Appendix B 4. An alternative architecture that explicitly enforces the stronger $O(3) \times O(3)$ symmetry (11) is presented in Appendix F.

Let $\sigma_i^{\text{init}} := (\mathbf{r}_i, \mathbf{m}_i, z_i) \in \mathbb{R}^3 \times \mathbb{R}^3 \times \mathbb{Z}$ denote the initial state of an atom. At each message passing iteration $t = 1, 2, \ldots$, node features from the previous layer are denoted by $h^{(t)}_{i,kl_2m_2}$. For the first layer ($t = 1$) we only allow invariant features (hence only the values $l_2 = m_2 = 0$ are required), which are initialized as a learnable embedding

of the chemical elements,

$$h^{(1)}_{i,k00} := W_{kz_i}. \tag{13}$$

This defines the node state $\sigma^{(1)}_i := (\mathbf{r}_i, \mathbf{m}_i, z_i, h^{(1)}_i)$ for message passing, as in (1).

For each atom i, information along its edges (i, j) is embedded using two learnable invariant bases $P_{kl_1l_2l_3}$ and $K_{kl_1l_2l_3}$

$$J_{ji} := J(|\mathbf{r}_{ji}|, z_i, z_j), \tag{14}$$
$$T_j := T(|\mathbf{m}_j|, z_j), \tag{15}$$
$$P^{(t)}_{ji,kl_1l_2l_3} := \mathrm{MLP}^{(t),\mathrm{P}}_{kl_1l_2l_3}\left(J_{ji} \oplus T_j\right), \tag{16}$$
$$K^{(t)}_{ji,kl_1l_2l_3} := \mathrm{MLP}^{(t),\mathrm{K}}_{kl_1l_2l_3}\left(J_{ji} \oplus T_j\right), \tag{17}$$

where $\oplus$ denotes a concatenation, and J and T are (non-trainable) Bessel and Chebyshev polynomial functions with pre-defined degree and element-dependent transformation; details are provided in Appendix B 2. We did not include $|\mathbf{m}_i|$ in the embedding of the edge; these will be included at a later stage in the architecture.

Both MLPs acts over the batch dimension (nodes or edges), and transform the concatenated input features to an output feature with dimension indexed by $kl_1l_2l_3$. The construction of P and K are analogous to the learnable radial basis R in the original MACE architecture [6], but now incorporate information about the spin magnitudes $|\mathbf{m}_j|$, which is an additional invariant input.

Next, the equivariant edge features with magnetic information can be constructed through the following sequence of operations

$$\phi^{(t),\mathrm{pos}}_{ji,kl_3m_3} := \sum_{l_1m_1l_2m_2} C^{l_3m_3}_{l_1m_1l_2m_2} \cdot P^{(t)}_{ji,kl_1l_2l_3} Y^{m_1}_{l_1}(\hat{\mathbf{r}}_{ji}) h^{(t)}_{j,kl_2m_2} \tag{18}$$

$$\phi^{(t),\mathrm{mag+pos}}_{ji,kl'_3m'_3} := \sum_{l'_1m'_1l'_2m'_2} C^{l'_3m'_3}_{l'_1m'_1l'_2m'_2} \cdot K^{(t)}_{ji,kl'_1l'_2l'_3} R^{m'_1}_{l'_1}(\mathbf{m}_j) \phi^{(t),\mathrm{pos}}_{ji,kl'_2m'_2}, \tag{19}$$

where $\hat{\mathbf{r}}_{ji} := \mathbf{r}_{ji}/|\mathbf{r}_{ji}|$, $Y^{m_1}_{l_1}$ are real spherical harmonics, and $R^{m'_1}_{l'_1}$ are real solid harmonics which guarantee smoothness in the limit $|\mathbf{m}_j| \to 0$. When $t = 1$, (18) simplifies to

$$\phi^{(1),\mathrm{pos}}_{ji,kl_3m_3} := P^{(t)}_{ji,k} Y^{m_3}_{l_3}(\hat{\mathbf{r}}_{ji}) h^{(1)}_{j,k00}, \tag{20}$$

reducing the computational cost of the first layer. The two contractions (18) and (19) can be merged into a single operation, requiring only one learnable radial basis, however, the resulting intermediate tensor would be significantly larger, leading to increased computational cost and memory usage. For this reason, we choose to break the contraction into two stages.

Next, the two-body basis on each node can be formed by pooling over the neighbourhood of atom i

$$A^{(t)}_{i,klm} := \sum_{j \in \mathcal{N}(i)} \phi^{(t),\mathrm{mag+pos}}_{ji,klm}. \tag{21}$$

Optionally, density normalization [4] can be applied to the two-body features to ensure internal normalization of the weights, which can result in better training dynamics and smooth extrapolation to high pressure.

The resulting joint positional–magnetic basis is closely related to the magnetic ACE basis of Drautz [21], in which magnetic moments are attached to atoms as vectorial degrees of freedom; mMACE augments it with learnable radial functions and embeds it in a message-passing architecture. Once the two-body basis $A^{(t)}_{i,klm}$ is formed, the many-body basis for each correlation order ν is

$$\mathbf{B}^{(t)}_{i,\eta_\nu kLM} := \sum_{\mathbf{lm}} C^{LM}_{\eta_\nu \mathbf{lm}} \prod_{\xi=1}^{\nu} \sum_{\tilde{k}} W^{(t)}_{k\tilde{k}l_\xi} A^{(t)}_{i,\tilde{k}l_\xi m_\xi}, \tag{22}$$

where $C^{LM}_{\eta_\nu \mathbf{lm}}$ are generalized Clebsch–Gordan coefficients [5]. Before we define the many-body message $\mathcal{M}^{(t)}$ of the current iteration t, we first specify intermediate features, which are linear in the many-body basis,

$$\mu^{(t)}_{i,kLM} := \sum_{\nu=1}^{\nu_{\max}} \sum_{\eta_\nu} W^{(t),\nu}_{z_i\eta_\nu kL} \mathbf{B}^{(t)}_{i,\eta_\nu kLM}. \tag{23}$$

We correlate the magnetic moment of the center atom i itself with its atomic environment, resulting in the many-body message

$$Q^{(t)}_{i,kl_0l_1L} := \mathrm{MLP}^{(t),\mathrm{Q}}_{kl_0l_1L}\left(T(|\mathbf{m}_i|, z_i)\right), \tag{24}$$

$$\mathcal{M}^{(t)}_{i,kLM} := \sum_{l_0m_0l_1m_1} C^{LM}_{l_0m_0l_1m_1} Q^{(t)}_{i,kl_0l_1L} R^{m_0}_{l_0}(\mathbf{m}_i) \mu^{(t)}_{i,kl_1m_1}. \tag{25}$$

Here $Q^{(t)}_{i,kl_0l_1L}$ is again a learnable radial function analogous to (17) but depends only on the element and magnitude of magnetic moment of atom i.

The hidden node feature update is completed via a residual connection

$$h^{(t+1)}_{i,kLM} := \sum_{\tilde{k}} W^{(t),\mathrm{self}}_{kL,\tilde{k}} \mathcal{M}^{(t)}_{i,\tilde{k}LM} + \sum_{\tilde{k}} W^{(t)}_{kL,\tilde{k}} \mu^{(t)}_{i,\tilde{k}LM} + \sum_{\tilde{k}} W^{(t)}_{kz_iL,\tilde{k}} h^{(t)}_{i,\tilde{k}LM}. \tag{26}$$

The atomic *interaction energies* are then obtained as a readout applied to the invariant node features,

$$E^{\mathrm{inter}}_i := \sum_{t=1}^{t_{\max}} \left(\mathcal{R}_t\left(h^{(t)}_{i,\tilde{k}LM}\right)\right). \tag{27}$$

The readout functions $\mathcal{R}_t$ are implemented as nonlinear MLPs in the final layer, while all preceding message-passing layers remain linear.

Finally, a one-body magnetic moment contribution is introduced to ensure the correct large volume limit (*i.e.* as $|\mathbf{r}_{ji}| \to \infty$ for all i, j). This term is pre-calculated as the energy of isolated atoms in different magnetic states or fitted using a simple linear model,

$$E_{0,z_i}(|\mathbf{m}_i|) := \sum_{n,z} W_{nz}\delta_{zz_i}\tilde{T}_n(|\mathbf{m}_i|), \tag{28}$$

where $(\tilde{T}_n)_n$ is a Chebyshev polynomial basis with an appropriate transformation applied to the input (See Appendix B 1 and Appendix B 2 for details).

The final atomic energies are given by the sum of interaction and one-body contributions,

$$E_i := E_i^{\text{inter}} + E_{0,z_i}(|\mathbf{m}_i|). \tag{29}$$

Further details on implementation can be found in Appendix B. The implementation of mMACE for reproducing the experiments is available at [30]. Details on parameter estimation ("training") and dataset generation can be found in Appendix G and H.

III. RESULTS

A. Training from Scratch: Benchmarks on Public Datasets

1. FeAl benchmark

We first benchmark the mMACE architecture on the collinear spin-polarized FeAl dataset of Kotykhov *et al.* [33], which was originally published for fitting a magnetic MTP (mMTP) model. Figure 2a compares energy, force and stress predictions of MACE and mMACE against the DFT reference. The standard MACE model exhibits regions where it predicts zero forces and stresses despite non-zero DFT values, reflecting its inability to capture magnetic moment contributions. In contrast, mMACE shows no such artefacts, indicating that the explicit inclusion of magnetic degrees of freedom enables it to represent spin-dependent energy surfaces more accurately.

These observations are quantified in Fig. 2a,c, where mMACE achieves the lowest RMSE across energies, forces, and stresses on the test set of [33], outperforming both mMTP and the standard MACE model. In particular, the force RMSE is reduced by nearly a factor of five and the stress RMSE by roughly a factor of three.

2. CrN benchmark

Figure 2b presents the analogous comparison for the collinear spin-polarized CrN dataset of Kotykhov *et al.* [34]. Here, the standard MACE model fails to resolve energy differences associated with different spin states, whereas mMACE successfully captures these distinctions and shows improved correlation with the DFT reference across energies, forces, and stresses.

The error metrics in Fig. 2b,d confirm that mMACE achieves the lowest errors across all quantities. Most notably, this dataset includes magnetic force labels, and the magnetic force RMSE is reduced by more than a factor of three relative to the mMTP baseline. A scatter plot of the magnetic force predictions is provided in Appendix E 1.

B. Pretrained Collinear Magnetic Models

To demonstrate the ability of mMACE to fit broad, general-purpose datasets, we train on two publicly available sources that include magnetic information from collinear spin-polarized DFT: MATPES [32], which provides both PBE and r^{2}SCAN reference calculations [50], and MP-ALOE [35], which contains r^{2}SCAN data. We fit mMACE models to each dataset independently and use the MATPES-PBE model as the starting point for the fine-tuning studies in Sections III C and III D.

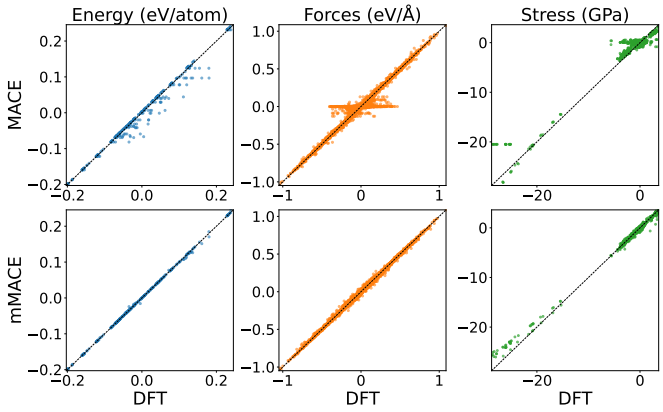
The results are summarized in Table I. On the MATPES-PBE dataset, mMACE consistently achieves lower mean absolute errors across energies, forces, and stresses compared to the baseline MACE model for configurations with non-zero magnetic moments. For zero-moment systems, where spin effects are minimal, mMACE performs comparably to MACE. The additional magnetic degrees of freedom therefore improve accuracy on magnetically active configurations while retaining comparable accuracy on non-magnetic ones.

The improvement is consistent across all three datasets, though more modest than the five- to tenfold error reductions observed on the FeAl and CrN benchmarks (Section III A). This is expected: MATPES and MP-ALOE are dominated by ferromagnetic configurations, so the magnetic diversity against which mMACE can differentiate itself is limited. Nevertheless, the pretrained mMACE model serves as a reliable baseline for fine-tuning on specific magnetic systems, as we demonstrate in Sections III C and III D.

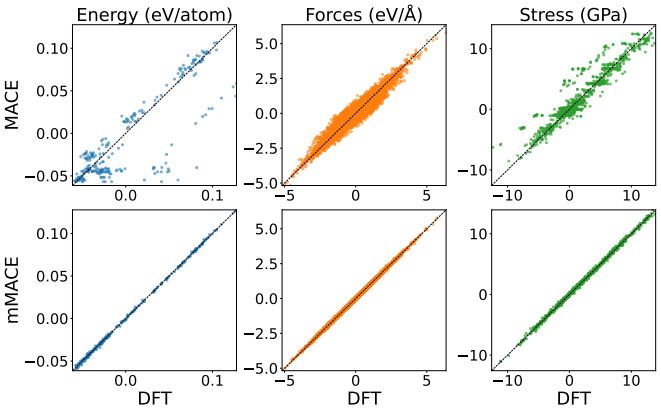
We note that single-point error metrics do not fully reflect a model's ability to reproduce self-consistent magnetic states under energy minimization. An analysis of relaxed magnetic moments and its implications for the stability and transferability of magnetic MLIPs is provided in Appendix E 2.

C. Fine-tuning study: FeNi Binary System

We next apply mMACE to the FeNi binary alloy, a prototypical transition-metal system of structural and magnetic relevance [23]. This section serves two purposes: (i) to assess whether mMACE can resolve magnetically sensitive phenomena, in particular the Bain transformation



(a) FeAl



(b) CrN

	mMTP	MACE	mMACE
Energy (meV/atom)	4.03	13.96	**0.97**
Force (meV/Å)	48	52	**6.86**
Stress (GPa)	0.79	0.92	**0.25**
Magnetic force (meV/μ_B)	–	–	–

(c) FeAl RMSE test errors

	mMTP	MACE	mMACE
Energy (meV/atom)	1.7	31.2	**1.21**
Forces (meV/Å)	108	219	**24**
Stress (GPa)	0.4	1.29	**0.25**
Magnetic force (meV/μ_B)	64	–	**19**

(d) CrN RMSE test errors

FIG. 2: Summary of benchmark results. (a-b) Scatter plots comparing mMACE and MACE to DFT reference on the FeAl and CrN datasets. (c-d) Corresponding RMSE values on the test sets [33, 34]. A qualitative plot on magnetic forces of the CrN dataset can be found in Appendix E 1. In contrast, the standard MACE model cannot predict magnetic forces at all, as this functionality is not included.

Dataset	**Model**	**Non-zero magmom**			**Zero magmom**		
		E	F	S	E	F	S
MATPES–PBE	MACE	44	123	0.908	**35**	**92**	**0.897**
	mMACE	**34**	**119**	**0.808**	39	96	0.947
MATPES–r2SCAN	MACE	47	154	1.040	41	117	1.117
	mMACE	**35**	**133**	**0.853**	**36**	**106**	**1.003**
MP–ALOE	MACE	57	96	1.362	53	92	1.430
	mMACE	**48**	**86**	**1.133**	**50**	**88**	**1.358**

TABLE I: Mean absolute errors on the test set of mMACE and MACE across three datasets. E, F, and S correspond to energy (meV/atom), forces (meV/Å), and stress (GPa).

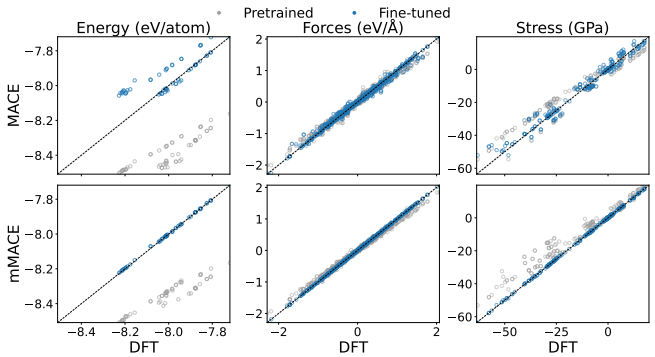
connecting the fcc and bcc phases; and (ii) to demonstrate a data-efficient fine-tuning strategy starting from the MATPES-PBE pre-trained model.

Before constructing a dedicated dataset, we assess the pre-trained model on the Bain path. Figure 3c shows that it correctly captures the energy scale separating non-magnetic and ferromagnetic states along the bcc–fcc transformation. However, it fails to resolve the relative ordering of the ferromagnetic and antiferromagnetic configurations. This is expected, as only 1,380 structures (0.35%) of the MATPES training dataset are antiferromagnetic, as classified by the `CollinearMagneticStructureAnalyzer` implemented in `pymatgen` [47], version 2025.10.7.

To address this, we generate targeted FeNi configurations using the protocol in Appendix H 2, progressively enriching structural coverage along the corresponding cell ratio of interest without explicitly sampling the Bain path (see the lower panels Fig. 3c). This results in a sequence of refined models with systematically improved accuracy.

The resulting Bain path after the second fine-tuning iteration (iter2 in Fig. 3c closely matches the DFT reference, although Bain-path configurations were not explicitly included in the training set for all magnetic orderings.

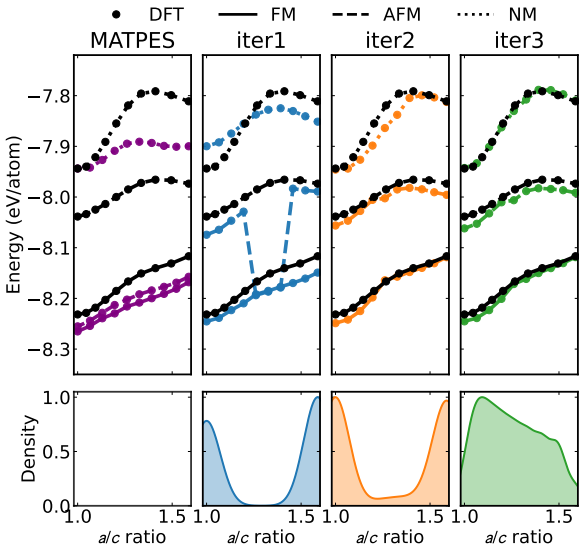
To further assess the mMACE model, we perform an additional fine-tuning stage (labeled as iter3) in which a small number of explicitly computed Bain-path configurations is incorporated into the training set. Following this fine-tuning step, the accuracy along all three trans-



(a)

Model	Energy (meV/atom)	Forces (meV/Å)	Stress (GPa)
MACE-MATPES-PBE	382.7	69.76	2.479
mMACE-MATPES-PBE	349.8	54.57	2.552
MACE-FT (iter3)	66.7	37.14	1.392
mMACE-FT (iter3)	**2.44**	**3.06**	**0.117**

(b)



(c)

FIG. 3: Finetuning of the FeNi database using mMACE. **(a, b) Left:** Energy, force and stress predictions of (m)MACE models against DFT reference with MAE values on the FeNi finetuning database. Errors after energy minimization (with DFT magnetic moment as initial guess) with mMACE-FT (iter3) are 34.6 meV/atom (energy), 18.2 meV/Å (forces), and 0.883 GPa (stress). **(c) Right:** mMACE Bain path transformation for FM, AFM, and NM states as the dataset is iteratively enriched (Section III C). Density plots indicates the distribution of additional data used in the face-centered tetragonal representation in each iteration, where $a/c = 1$ and $a/c = \sqrt{2} \approx 1.414$ corresponds to fcc and bcc respectively. Energies of mMACE-MATPES are shifted by matching the $a/c = 1.0$ NM state to remove the constant offset arising from differences between DFT implementations for comparison between different magnetic states.

formation paths improves, with the most pronounced gains observed for the NM transformation. This demonstrates that mMACE can fit a broad range of configurations, provided the training data adequately span the relevant structural and magnetic degrees of freedom. As summarized in Table 3b, the fine-tuned mMACE model significantly outperforms the baseline MACE-MATPES-PBE model across energy, force, and stress metrics. This improvement highlights the ability of mMACE to capture the underlying magnetic and structural physics with substantially higher fidelity.

Without introducing further training data, we additionally evaluate elastic constants. All equilibrium lattice parameters agree with DFT to within 0.01 Å across all phases and magnetic states, and all elastic constants reproduce the correct trends. Further details are provided in Appendix E 3.

Overall, these results demonstrate that magnetically sensitive phase stability and mechanical properties in FeNi can be recovered through targeted, physically motivated fine-tuning of a pre-trained magnetic model, without large-scale retraining.

D. Fine-tuning study: non-collinear Mn_3Pt

To evaluate mMACE on non-collinear magnetism, we consider the frustrated antiferromagnet Mn_3Pt, which features a triangular spin texture on a kagome-like Mn sublattice [64]. Strong geometric frustration and competing exchange interactions stabilize a non-collinear ground state, making Mn_3Pt a demanding test case: resolving the small energy differences between competing magnetic configurations requires an accurate treatment of spin-dependent interactions beyond collinearity.

The data generation procedure detailed in Appendix H 4 yields 147 training and 100 validation configurations. We use this dataset to fine-tune the MATPES-PBE pre-trained mMACE model, with the goal of robustly recovering the kagome-like ground state as the energy minimum.

In the absence of SOC, magnetic configurations are symmetry equivalent under global spin rotations. To define an error metric that is invariant under this symmetry, we consider the relative Gramian error for a matrix A containing stacked magnetic moment vectors

$$\varepsilon_{\mathrm{rel}}(A, A_{\mathrm{ref}}) = \frac{\|G(A) - G(A_{\mathrm{ref}})\|_F}{\|G(A_{\mathrm{ref}})\|_F}. \tag{30}$$

where $G(A) := A^T A$ is the Gramian matrix and $\|\cdot\|_F$ is the Frobenius norm. As shown in Fig. 4a, both the Gramian error and the angle alignment error decrease rapidly during the energy minimization and converge to near-zero values, confirming that the optimized spin configuration matches the kagome-like ground state up to a global rotation. Figure 4b illustrates the corresponding evolution of the Mn magnetic moment orientations, which undergo substantial rotations from their random initial values before stabilizing into the characteristic frustrated spin arrangement.

To demonstrate that mMACE can navigate the highly non-convex magnetic energy landscape, we perform 100 relaxations from random initial spin orientations, all outside the training and validation distributions. Figure 4c shows that the energy residue decreases systematically with decreasing magnetic misalignment, and that all runs converge to the kagome-like ground state. The MACE reference point (red triangle), relaxed from the DFT ground state, lies at zero misalignment and negligible energy residual, confirming that magnetic relaxation restores a well-converged self-consistent solution. Details on the relaxation procedure are provided in Appendix E 4.

E. Case study: magnetocrystalline anisotropy

Magnetocrystalline anisotropy (MCA) describes the directional preference of spin alignment along specific crystallographic axes. It originates from the coupling between the electronic spin and the crystal lattice *via* spin–orbit coupling (SOC) [36, 42].

For a fixed atomic configuration, the magnetic anisotropy energy E_A is defined as the total energy relative to a reference magnetization orientation (θ_0, ϕ_0),

$$E_A(\theta, \phi) := E(\theta, \phi) - E(\theta_0, \phi_0), \tag{31}$$

where (θ, ϕ) denote the polar and azimuthal angles of the magnetization vector. The orientation that minimizes the energy defines the *easy axis* of the system.

Typically, the energy scale associated with SOC-induced anisotropy is on the order of sub-meV per atom (see Fig. 6a), which is comparable to or smaller than the intrinsic error of modern machine-learned interatomic potentials in single-point energy predictions. Consequently, the anisotropy should be regarded as a small perturbative correction to the total energy of a configuration and is fitted separately. Throughout this work, all SOC reference data are obtained from non-self-consistent evaluations within the magnetic force theorem (Appendix C): for each magnetization orientation, the SOC energy is the band-energy difference computed from the eigenstates of a single collinear self-consistent calculation. mMACE is trained on these differences, and therefore learns the magnetic-force-theorem approximation to the anisotropy energy rather than a self-consistent SOC total energy.

In this work, we are primarily interested in identifying configurations that exhibit a strong anisotropic response to magnetization reorientation. To quantify this, we define the anisotropy strength of a configuration $\{\sigma_i\}_i := \{(\mathbf{r}_i, \mathbf{m}_i, z_i)\}_i$ as the maximal variation of the total energy over all magnetization directions,

$$\Delta E_A(\{\sigma_i\}_i) := \max_{\theta_1,\phi_1,\theta_2,\phi_2} \Big[E(\theta_1, \phi_1; \{\sigma_i\}_i) - E(\theta_2, \phi_2; \{\sigma_i\}_i)\Big]. \tag{32}$$

This quantity characterizes the intrinsic anisotropy scale associated with a given atomic and magnetic configuration. Details of the dataset construction and computational protocol are provided in Appendix H 3.

1. FeMn

We consider a ferrimagnetic configuration of $L1_0$ FeMn within a two-atom unit cell. SOC is included non-self-consistently in all calculations. The magnetic anisotropy energy is obtained by rotating the spin axis while keeping the atomic positions fixed. For this uniaxial system, the magnetic anisotropy energy E_A can be well approximated by a fourth-order angular expansion [13]

$$\frac{E_A}{V} \approx K_1 \sin^2\theta + K_2 \sin^4\theta. \tag{33}$$

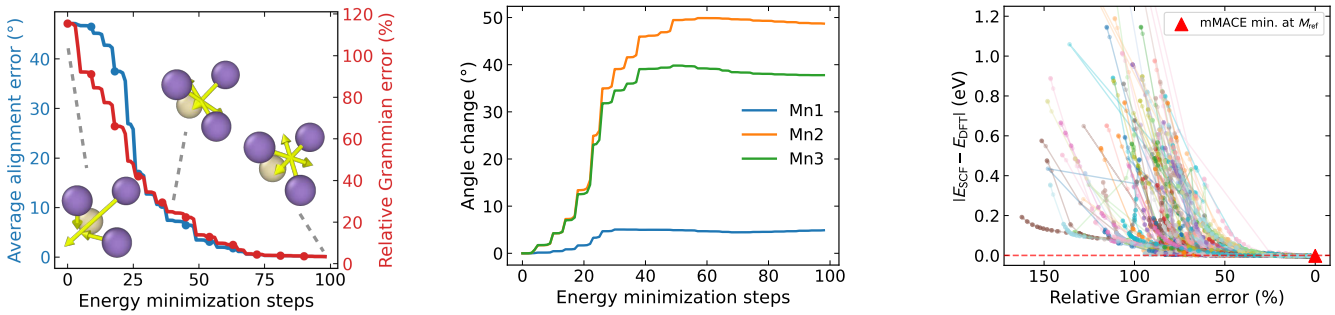
where V denotes the system volume and the z-axis coincides with the principal symmetry axis. The sign of K_1 determines whether this axis corresponds to an *easy* or *hard* magnetization direction, and typically $|K_1| \gg |K_2|$. Details on dataset construction can be found in Appendix H 3 a.

Figure 5 compares the angular dependence obtained from DFT and the mMACE model. The fitted values of K_1 are 3.52 MJ m^{-3} (mMACE) and 3.56 MJ m^{-3} (DFT), demonstrating excellent agreement. The fitted K_2 term is several orders of magnitude smaller than K_1 and cannot be reliably determined within numerical noise.

2. Fitting a random structure search (RSS) database

To assess the generalization capability of mMACE in predicting E_A across diverse chemical environments, we construct a database of random structures spanning a broad range of metallic compositions. The dataset includes transition and post-transition metals with varying strengths of spin–orbit coupling—namely Al, Mn, Fe, Co, Ni, Pt, and Y—and additionally incorporates up to one light element selected from B, C, or Si. A total of 242 *parent structures* were created, and for each parent structure 2000 orientations were sampled over the unit sphere, resulting in 484k configurations.

To account for the wide dynamic range of spin–orbit coupling energies, the results are subdivided by E_{SOC}



(a) (b) (c)

FIG. 4: (a) Energy minimization of magnetic moments on the Mn_3Pt lattice starting from a random orientation, ultimately relaxing to the kagome-like spin structure. (b) Rotation angle relative to a reference orientation as a function of minimization steps. Each energy minimization step corresponds to a single energy evaluation. (c) Energy error with respect to ground state decreases monotonically as the magnetic moments become better aligned. Large misalignments produce substantial energy residue. The red triangle indicates the mMACE-relaxed magnetic configuration, which lies at zero misalignment and zero energy residue.

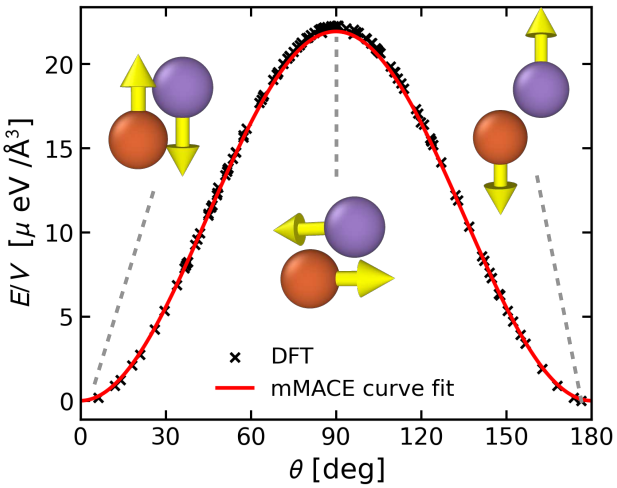


FIG. 5: Comparison between DFT data and mMACE energy density curve for FeMn.

range, following the distribution in Fig. 6a. Since ΔE_A spans several orders of magnitude, we report relative rather than absolute errors. We extract varying numbers of training configurations N_{train} per parent structure, using 190 configurations each for training and validation, with the remainder reserved for testing. The resulting learning curves are shown in Fig. 6c, reporting the relative mean absolute error. The model converges to a relative error of approximately 5%, although the relative error is larger for configurations with stronger spin–orbit coupling.

To assess the model's ability to identify configurations with strong SOC, we perform a 90–10 train-test split at the level of *parent configurations*, *i.e.* all SOC-rotated structures derived from the same parent are assigned exclusively to either the training or the test set.

Figure 6b compares the maximum SOC-induced energy ΔE_A per parent configuration on the test set, predicted by mMACE against the corresponding DFT reference values. Both axes are shown on logarithmic scales, and the dashed line indicates perfect agreement. The strong monotonic correlation between mMACE and DFT, quantified by a Spearman rank coefficient of $\rho = 0.894$, demonstrates that the model reliably identifies parent configurations associated with large SOC energy variations. While deviations in absolute magnitude remain, the relative ordering of SOC sensitivity across configurations is well preserved, which is the key requirement for SOC-aware configuration screening.

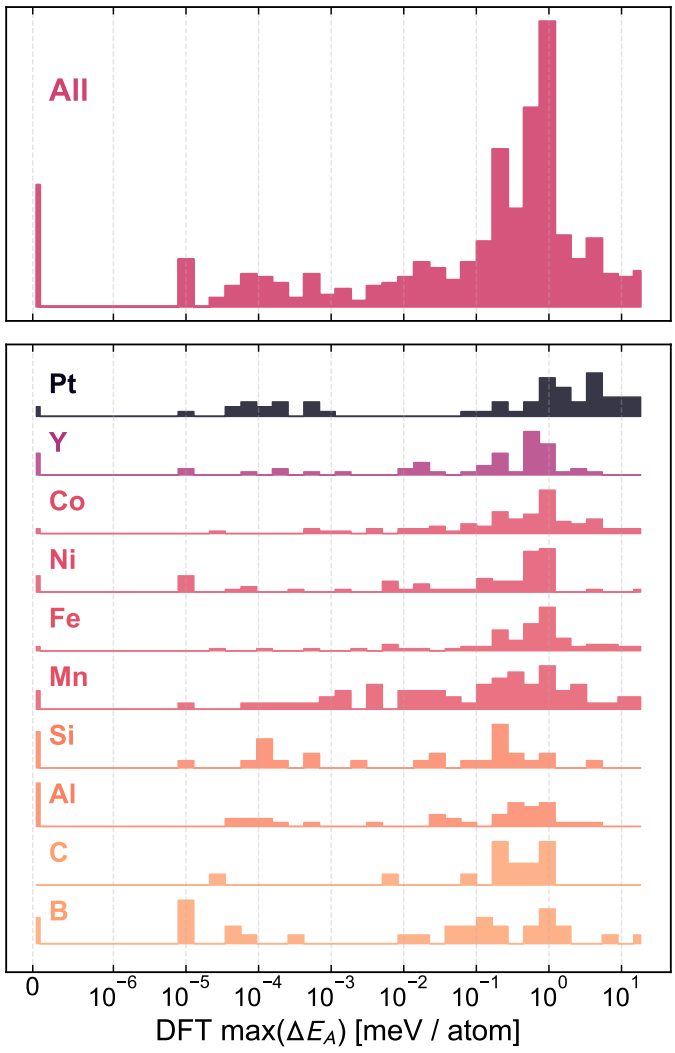
F. Case study: non-collinear Fe model

Rinaldi *et al.* [53] introduced a comprehensive dataset for pure Fe covering an extensive range of atomic configurations, including non-collinear magnetic data. The ACE model developed in that work was shown to reproduce a broad spectrum of physical properties with high accuracy, including Curie temperatures. We fit this Fe database with mMACE to benchmark its performance, and compare methods of Curie temperature estimation—in particular contrasting mMACE-driven simulations with a classical Heisenberg spin model—to highlight the benefits of magnetically informed ML potentials.

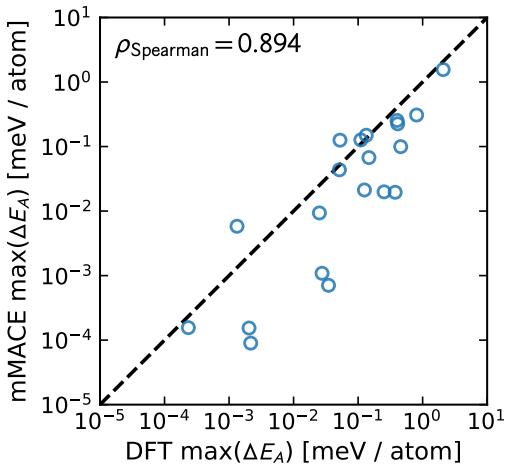
1. *Heisenberg exchange constant*

A widely used framework for describing magnetism in condensed matter is the classical Heisenberg model

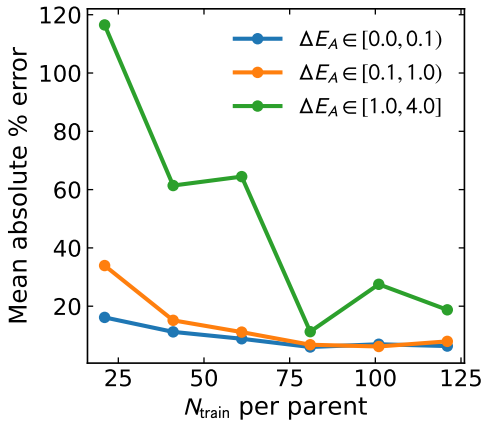
$$\mathcal{H} := \mathcal{H}_0 - \sum_{\langle ij \rangle} J_{ij} \hat{\mathbf{S}}_i \cdot \hat{\mathbf{S}}_j \tag{34}$$



(a)



(b)



(c)

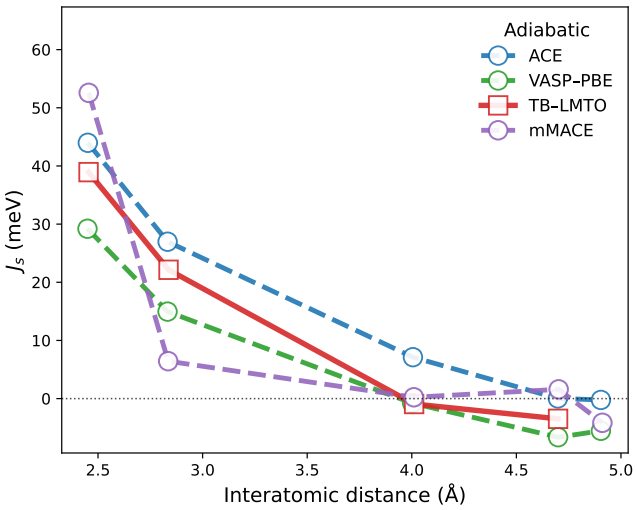
FIG. 6: Spin–orbit coupling (SOC) energy sensitivity across the RSS dataset. **Left:** Distribution of maximum SOC-induced magnetic anisotropy energies ΔE_A over all configurations containing a given element. Elements are ordered by median ΔE_A, illustrating the broad dynamic range of SOC sensitivity. **Right:** (Top) Maximum SOC energy variation per atom predicted by mMACE versus DFT on logarithmic scales on a test set; the dashed line denotes perfect agreement. (Bottom) Learning curve for SOC energy differences with percentage error regularized by 1 μ eV / atom, showing systematic error reduction with increasing training data.

where $\hat{\mathbf{S}}_i$ are normalized spin vectors and J_{ij} are exchange coefficients between pairs of atomic sites. It has been actively used in magnetic materials research and the development of electronic devices [41].

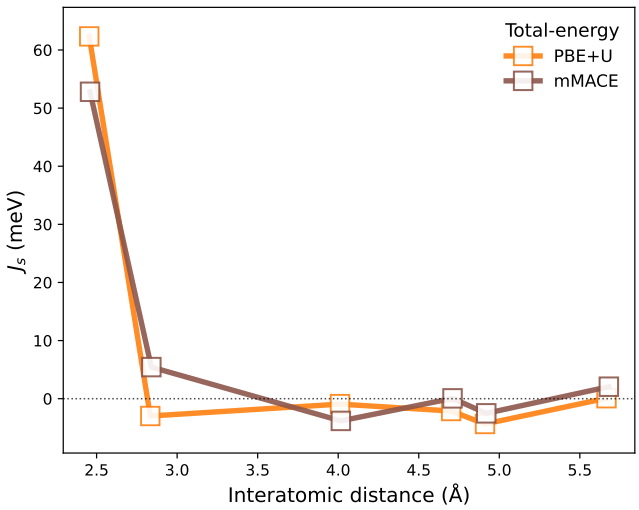
Two widely used approaches for estimating Heisenberg exchange coefficients are the total-energy–difference method [51] (see Appendix D 1 for details) and the real-space approach of Liechtenstein *et al.* [37–39]. These approaches rest on different approximations and can yield different estimates of the exchange constants. Figures 7a and 7b present the exchange coefficients grouped into symmetry-equivalent shells by interatomic distance. We find that mMACE successfully reproduces the qualitative trends of reference exchange constants reported in [51–53]. This agreement indicates that mMACE captures essential magnetic interactions, despite quantitative differences arising from the approximations used to extract the exchange parameters.

2. *Curie temperature*

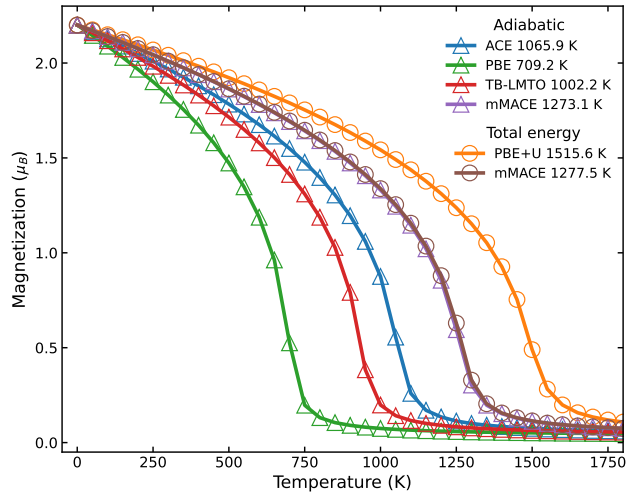
Iron has an experimental Curie temperature of approximately 1043 K [56]. To assess the predictive capability of different magnetic modeling strategies, we compare Curie temperatures obtained from (i) classical Heisenberg mod-



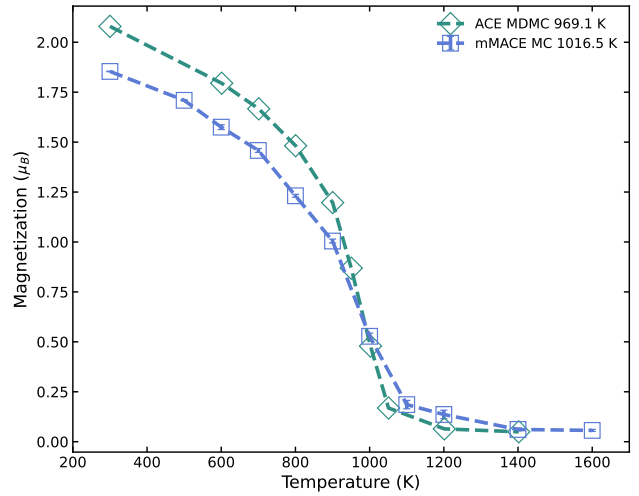
(a) Heisenberg exchange constants obtained using the real-space approach of Liechtenstein et al. [37–39].



(b) Heisenberg exchange constants obtained using the total-energy method in [51].



(c) VAMPIRE MC simulations with J_{ij} from the adiabatic and total-energy methods.



(d) MD–MC and MC simulations using different ACE and mMACE. ACE reference is obtained from [53].

FIG. 7: Heisenberg exchange interactions and resulting magnetic phase transitions in bcc Fe. (a,b) Variation of Heisenberg exchange constants J_s as a function of interatomic distance obtained using the adiabatic and total-energy methods, respectively. VASP–PBE reference data are taken from Ref. [52], and TB-LMTO data from Ref. [48]. (c,d) Temperature-dependent magnetisation curves showing the ferromagnetic–paramagnetic phase transition around the Curie temperature, computed using VAMPIRE Monte Carlo simulations and MD–MC simulations with machine-learned interatomic potentials.

els parameterised by exchange constants J_s (Figs. 7b and 7a) and simulated using VAMPIRE [24][45], and (ii) direct finite-temperature simulations with mMACE using Metropolis Monte Carlo (MC) sampling of the non-collinear spin degrees of freedom at fixed atomic positions (Appendix E 5), compared with published molecular dynamics–Monte Carlo (MDMC) results obtained with ACE [53].

The key difference between these approaches lies in the energy model. The VAMPIRE simulations use a classical Heisenberg Hamiltonian restricted to pairwise exchange within a finite number of neighbour shells, whereas mMACE represents the full energy of the non-collinear spin configuration, including many-body magnetic interactions explicitly. mMACE is therefore expected to describe finite-temperature magnetic disorder more accurately than the pairwise Heisenberg models.

All models exhibit a collapse of the net magnetisation

from approximately 2.2 μ_B/atom to zero near the Curie point (Figs. 7c and 7d). However, the Curie temperatures predicted using VAMPIRE depend sensitively on the reference method used to extract the exchange constants.

mMACE-driven Monte Carlo (MC) simulations on the non-collinear spin degrees of freedom yield $T_C \approx 1016$ K, close to the experimental value, whereas a Heisenberg model parameterized by exchange constants extracted from the same mMACE model overestimates T_C (Fig. 7c). The truncation of the mMACE energy to pairwise exchange therefore shifts T_C substantially, while the full model, which retains many-body magnetic interactions, gives a Curie temperature close to experiment. For reference, Fig. 7d also shows the temperature–magnetisation curve from published ACE MDMC simulations [53]. Since these additionally sample atomic positions, the ACE curve serves only as a reference rather than a direct comparison. We do not perform MDMC simulations here, as reliable estimation of T_C can be sensitive to the MD/MC step ratio, which was not reported in the original work, and sampling the positional degrees of freedom requires a more careful convergence study, which we defer to future work.

Overall, these results show that spin-only MC with mMACE can yield Curie temperatures close to experiment. Since it does not require equilibrating the lattice, such sampling is expected to converge faster in practice than MDMC, making it more suited to high-throughput screening of finite-temperature magnetic properties.

IV. DISCUSSION AND OUTLOOK

We have introduced mMACE, an equivariant message-passing architecture that extends MACE to incorporate atomic magnetic moments as explicit degrees of freedom. The architecture is equivariant under joint rotations of positions and moments, with spin–orbit coupling effects captured naturally; the stronger $O(3) \times O(3)$ symmetry appropriate for systems without SOC can be enforced through data augmentation or an alternative architecture presented in the appendix. On public benchmarks (FeAl, CrN) mMACE substantially reduces force and stress errors relative to both standard MACE and mMTP baselines. Pre-trained models on the MATPES and MP-ALOE foundational datasets yield systematic improvements on magnetically active configurations with comparable accuracy on non-magnetic configurations, and serve as effective starting points for targeted fine-tuning.

The fine-tuning studies on FeNi and Mn_3Pt illustrate two complementary strengths of the approach. For FeNi, a small number of targeted configurations suffices to recover the Bain-path energetics across ferromagnetic, antiferromagnetic, and non-magnetic states, together with accurate lattice parameters and elastic constants. For Mn_3Pt, mMACE reliably recovers the frustrated kagome-like non-collinear ground state from random initial spin orientations, demonstrating that the learned energy landscape is qualitatively correct even far from the training distribution. The magnetocrystalline anisotropy study shows that mMACE can resolve SOC-induced energy differences at the sub-meV scale across chemically diverse random structures, achieving relative errors of approximately 5% and preserving the ranking of anisotropy strengths. Finally, the Fe case study shows that mMACE-driven Monte Carlo simulations yield a Curie temperature close to experiment, whereas pairwise Heisenberg models parameterized from the same mMACE model overestimate it, highlighting the role of magnetic interactions beyond pairwise exchange.

Limitations to be addressed in future work include: First, the pre-trained models are fitted to datasets dominated by collinear, predominantly ferromagnetic configurations; extending the foundational training data to include antiferromagnetic, non-collinear, and frustrated magnetic orderings would broaden the applicability of pre-trained mMACE models. Second, while we have demonstrated that SOC-induced anisotropy can be fitted as a perturbative correction, integrating it self-consistently into the training procedure remains an open challenge. Third, extending the model architecture, along the lines of Appendix F, to effectively treat both non-SOC and SOC systems without the need for data augmentation would benefit both accuracy and generality.

Finally, coupling mMACE to spin-lattice dynamics frameworks would open the door to studying magnon-phonon interactions, demagnetisation, and other dynamical phenomena beyond the static relaxations and Monte Carlo sampling employed here.

ACKNOWLEDGMENTS

CH, JD, GC, CO and CvdO would like to acknowledge Materials Nexus for providing computational resources with regards to DFT data generation for the random structure search dataset for studying magnetocrystalline anisotropy.

AUTHOR CONTRIBUTIONS

CH developed the mathematical foundations and code implementation of mMACE and conducted all computational experiments. CH also led the database generation, manuscript writing, and created all figures/tables. CvdO and JD contributed to interpreting the results and guided the project on a daily basis, with CvdO taking a leading role and contributing sections to the manuscript. TK, RB, CE, RK and ES provided suggestions of generating datasets and generated datasets for studying the anisotropy constant. TK and RT validated the mMACE implementation and TK, RT, RB, MP and GC provided project guidance on a weekly basis. RF and JB suggested test cases and systems, and provided project over-

sight. GC and CO assisted in developing the mathematical foundations of the mMACE architecture. JB, GC and CO provided general project oversight.

COMPETING INTERESTS

GC, CO, CvdO, JD are partners of Symmetric Group LLP that licenses force fields commercially. GC also has equity interest in Ångström AI.

DATA AVAILABILITY

The data that support the findings of this study are publicly available. The MATPES dataset is available at [14]. The MP-ALOE dataset is available via Figshare at [17]. The non-collinear Fe dataset is available via Zenodo at [12]. Additional data are available in the code repositories listed below.

CODE AVAILABILITY

The implementation of Magnetic MACE is publicly available in [30], and scripts used to reproduce the results will be made publicly available in [29] upon publication.

Appendix A: Axial-vector symmetry and time-reversal symmetrization

Magnetic moments are axial rather than polar vectors. Consequently, under a spatial transformation $Q \in O(3)$, atomic positions and magnetic moments transform according to

$$\mathbf{r}_i \mapsto Q\mathbf{r}_i, \qquad \mathbf{m}_i \mapsto \det(Q)\, Q\mathbf{m}_i. \tag{A1}$$

For proper rotations, $Q \in SO(3)$, the two transformation laws coincide, whereas for improper transformations they differ by a sign. In the absence of an external magnetic field, the energy is invariant under global time reversal,

$$E(\mathbf{R}, \mathbf{M}) = E(\mathbf{R}, -\mathbf{M}). \tag{A2}$$

In the implementation used in the main text, the magnetic moment is encoded internally using the standard $l = 1$ polar-vector representation. The unsymmetrized model therefore satisfies

$$E_\theta(Q\mathbf{R}, Q\mathbf{M}) = E_\theta(\mathbf{R}, \mathbf{M}), \qquad Q \in O(3), \tag{A3}$$

where Q acts simultaneously on the positional and magnetic spherical harmonics. While Eq. (A3) does not by itself implement the physical axial-vector transformation of magnetic moments under improper rotations, the correct symmetry can be recovered exactly by defining the symmetrized energy

$$E_\theta^{\mathrm{sym}}(\mathbf{R}, \mathbf{M}) = \frac{1}{2}\left[E_\theta(\mathbf{R}, \mathbf{M}) + E_\theta(\mathbf{R}, -\mathbf{M})\right]. \tag{A4}$$

which can be evaluated as a batch for performance. Combining Eq. (A3) with Eq. (A4) gives the correct transformation law for an axial magnetic moment under arbitrary $Q \in O(3)$:

$$E_\theta^{\mathrm{sym}}\left(Q\mathbf{R}, \det(Q)\, Q\mathbf{M}\right) = E_\theta^{\mathrm{sym}}(\mathbf{R}, \mathbf{M}). \tag{A5}$$

To see this directly, for $\det(Q) = +1$, this follows immediately from the original joint $O(3)$ invariance. For $\det(Q) = -1$,

$$\begin{aligned} E_\theta^{\mathrm{sym}}(Q\mathbf{R}, -Q\mathbf{M}) &= \frac{1}{2}\left[E_\theta(Q\mathbf{R}, -Q\mathbf{M}) + E_\theta(Q\mathbf{R}, Q\mathbf{M})\right] \\ &= \frac{1}{2}\left[E_\theta(\mathbf{R}, -\mathbf{M}) + E_\theta(\mathbf{R}, \mathbf{M})\right] \\ &= E_\theta^{\mathrm{sym}}(\mathbf{R}, \mathbf{M}), \end{aligned} \tag{A6}$$

where the second equality uses Eq. (A3). Thus, although the internal magnetic representation uses the polar $l = 1$ convention, the symmetrized scalar energy transforms exactly as required for axial magnetic moments.

The corresponding transformation properties of the atomic and magnetic forces follow directly by differentiation of the invariant scalar energy. Therefore, the use of a polar-vector representation for the magnetic spherical harmonics should be regarded as an internal representation choice rather than a statement about the physical parity of the magnetic moment. After the projection in Eq. (A4), the final energy, atomic forces, and magnetic forces obey the physically appropriate axial-vector and time-reversal symmetries exactly.

In practice, Eq. (A4) can be applied as a lightweight wrapper around the original mMACE model and requires no change to the underlying message-passing architecture. Alternatively we found that with rotational data augmentation only, the models typically violate the time reversal symmetry on all datasets with MAE 6.2e-3 to 4.5e-1 meV/atom, which is much lower then the model error.

Appendix B: Implementation details

1. One-body magnetic contribution in MLIPs

The fitting of one-body magnetic contributions can suffer from overfitting or result in non-smooth curves when optimized directly with gradient-based methods such as Adam or LBFGS, which can compromise the physicality of the model. In this appendix, we explain how to obtain a physical $E(m)$ via regularization. To compare more precisely, all plots are shifted by the isolated atom energy at zero magnetic moment so that the one-body magnetic contributions of different elements can be compared. We apply fixed-spin moment (FSM) via occupation control in GPAW to obtain isolated atom reference energies for different magnetic states, enforcing the magnetic moment constraint by adjusting the orbital occupations within the SCF loop and fixing the difference between spin-up and spin-down electron numbers.

To obtain a smooth $E(m)$ one-body contribution, one can regularize the coefficients in the linear model $E(m)$ in one of the following ways. Suppose $E(m)$ is simply a Chebyshev polynomial expansion; one can solve the following matrix equation obtained from a regularized least-squares system

$$(\mathbf{X}^\top\mathbf{X} + \lambda\mathbf{\Gamma}^2)\mathbf{c} = \mathbf{X}^\top\mathbf{y}, \tag{B1}$$

where $\mathbf{X} \in \mathbb{R}^{N\times B}$ is the Chebyshev basis evaluated at N sampled magnetic moments with B basis functions, $\mathbf{\Gamma} \in \mathbb{R}^{B\times B}$ encodes a regularity prior [16], and $\mathbf{y} \in \mathbb{R}^N$ contains the DFT-computed one-body magnetic energies. The regularization matrix is diagonal,

$$\Gamma_{kk} = k^p + 1, \tag{B2}$$

where k is the polynomial degree and p controls the strength of the smoothness penalty, normally chosen as $p = 4$. For each λ, the coefficients $\mathbf{c} \in \mathbb{R}^B$ are computed via direct linear algebra, and the fitted curves are compared against DFT data to visualize the trade-off between accuracy and smoothness induced by regularization, as demonstrated in Figure 8.

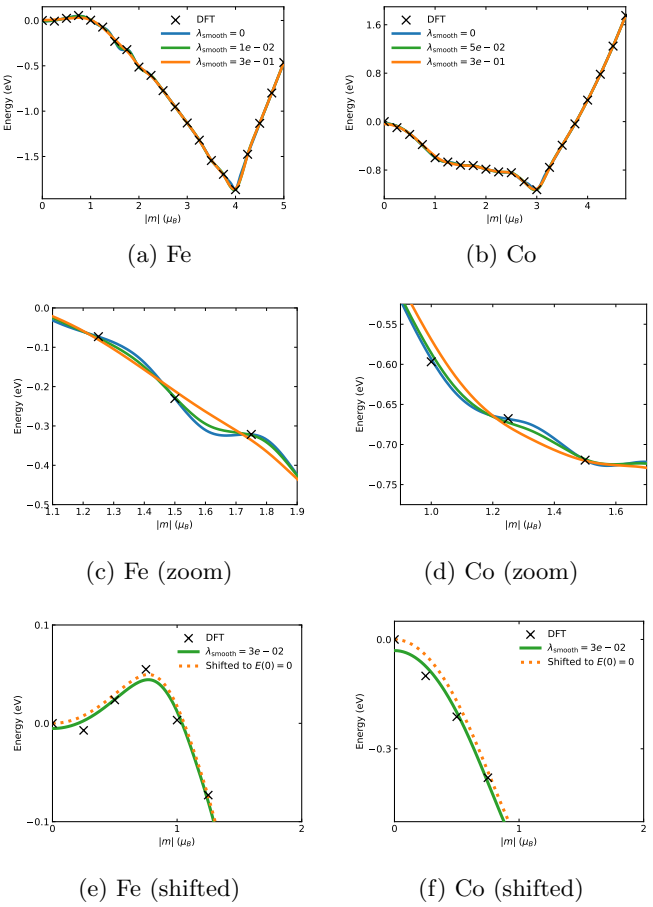
A more automated way of regularizing $E(m)$ is to augment the loss function, which we found works most robustly when combined with cubic spline interpolation.

More precisely, the one-body magnetic energy $E(m)$ is obtained by fitting a smooth interpolant to the discrete DFT reference points. In the present implementation, this is achieved using an even cubic spline representation $E(m) = f(m^2)$, which enforces invariance under spin inversion by construction. The fitting objective consists of a data fidelity term combined with a curvature penalty that suppresses large second derivatives of the spline with respect to its argument. Concretely, the optimized spline coefficients minimize

$$\mathcal{L}_{\text{iso}} = \frac{1}{N}\sum_{i=1}^{N}\big(E(m_i) - y_i\big)^2 + \lambda \sum_j \big(\Delta^2 f_j\big)^2, \qquad \text{(B3)}$$

where y_i denote the DFT reference energies evaluated at sampled magnetic moments m_i, and $\Delta^2 f_j$ represents the discrete second finite difference of the spline coefficients, providing a numerical approximation to the curvature of the energy with respect to m^2. The regularization parameter λ controls the relative weight of the smoothness penalty: small values allow the spline to closely interpolate the reference data, while larger values progressively suppress oscillations and enforce a more globally smooth energy landscape. By similarly inspecting the fitted curves obtained across a range of λ, one can identify a regime in which the resulting $E(m)$ is both smooth and faithful to the underlying DFT data, providing a robust one-body magnetic reference suitable for subsequent use in the machine-learned potential. The effect of different regularization strengths is demonstrated with two example elements Fe and Co in Figure 8.

Finally, to match the non-magnetic contribution exactly, the origin of the resulting fit can optionally be shifted to the isolated atom energy at zero magnetic moment, which yields regularized and smoother curves with the correct limit as $|m| \to 0$, as shown in Figures 8e and 8f. The techniques discussed above provide sufficient flexibility for designing the one-body magnetic contribution of MLIPs.



(a) Fe (b) Co

(c) Fe (zoom) (d) Co (zoom)

(e) Fe (shifted) (f) Co (shifted)

FIG. 8: Regularization of one-body magnetic energy curves using curvature-controlled fitting. Panels (a) and (b) show the fitted one-body magnetic energy $E(m)$ for Fe and Co, respectively, obtained from DFT reference data (markers) using different smoothness parameters λ. Panels (c) and (d) present zoomed-in views of selected regions of panels (a) and (b), highlighting how increasing regularization progressively suppresses short-wavelength curvature while preserving the overall trend of the DFT reference. (e) and (f) demonstrate how the resulting $E(m)$ can be shifted to 0 so that the model reduces to a non-magnetic many-body expansion exactly.

2. Radial basis

In this section, we give additional details for the construction of the radial bases involved in the above discussion in (16), (17) and (29). The construction of the Bessel function J is identical to regular, non-magnetic MACE; we refer the reader to [6] for details.

For the magnetic radial basis, let $m_{z_i} \in \mathbb{R}$ be the magnitude of the magnetic moment of an atom with species z_i. T maps a single input tuple (m, z_i) to a vector of length equal to the pre-defined maximum degree of the Chebyshev polynomial of the first kind, excluding the constant, via

$$T(m, z_i) := T(x_{z_i}(m)) \qquad \text{(B4)}$$

where x_{z_i} is a species dependent transformation

$$x_{z_i}(\cdot) : m \to 1 - 2\left(\frac{m}{M_{\max,z_i}}\right)^2 \qquad \text{(B5)}$$

so that it is mapped from $[0, M_{\max,z_i}] \to [-1, 1]$. The constant $M_{\max,z_i}$ is pre-computed depending on the range of magnetic moment magnitudes in the data. When unspecified in the paper, we choose

$$M_{\max,z_i} = aM^{(data)}_{\max,z_i} + b \qquad \text{(B6)}$$

where $a > 1$, $b > 0$ and $M^{(data)}_{\max,z_i}$ is the upper bound of

the magnetic moments in the dataset. In the above experiments, we chose $a = 1.2$ and $b = 0.1$, and found that errors and simple in-distribution properties are insensitive to these parameters.

3. Parameter Estimation

The term $E_{0,z_i}(|\mathbf{m}_i|)$ is a one-dimensional function that depends only on the magnitude of the magnetic moment of atom i. Notably, estimating E_{0,z_i} directly via gradient-based optimization can lead to severe overfitting due to the low intrinsic dimensionality of this term. To mitigate this, we fit the corresponding linear coefficients separately before fitting the remaining part of the model, using regularized approaches, as detailed in Appendix B 1.

Apart from the two-stage estimation for the one-body magnetic contribution, the remainder of the training procedure follows standard MACE hyperparameters and optimization strategies as described in [4]. Further details on dataset generation and hyperparameters for parameter estimation can be found in Appendix G.

4. Data Augmentation

Given that the model is capable of capturing spin–orbit coupling, fitting to datasets without spin–orbit effects (or collinear DFT data) requires the use of data augmentation. We find that data augmentation over only rotational data augmentation ensures the preservation of corresponding SOC or non-SOC symmetry up to one or two orders of magnitude smaller than the corresponding prediction error, with only a few rare outliers of comparable magnitude (See Figure 9 as an example evaluated on the FeNi dataset) We also verified that explicitly symmetrising the model under time reversal leads to negligible changes in all reported prediction errors. Therefore, for computationally efficiency the results are obtained without explicit time-reversal symmetrisation throughout. At the same time, the level of symmetry error violation is also in line with [40], where data augmentation is similarly applied for training the corresponding MLIPs.

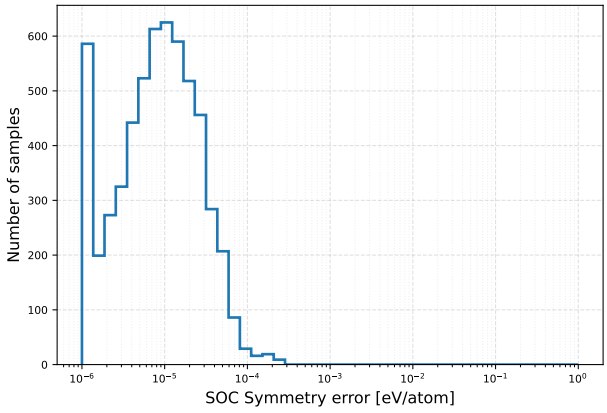


FIG. 9: Distribution of the per-atom energy deviation induced by random spatial rotations in mMACE models on the FeNi dataset after fine-tuning. The histogram is shown on a logarithmic scale and quantifies the symmetry error between unrotated and rotated configurations. The vast majority of samples exhibit deviations below 0.1 meV/atom, with only a small number of rare outliers, demonstrating that data augmentation preserves rotational symmetry to high numerical accuracy.

Appendix C: Magnetic Anisotropy Constant

The most rigorous approach to computing magnetic anisotropy energy (MAE) within DFT involves performing *non-collinear self-consistent-field* (SCF) calculations with SOC and constrained magnetic moments for each magnetization direction, using fully relativistic pseudopotentials or all-electron methods. Such calculations are computationally demanding and difficult to converge.

For most materials, however, magnetocrystalline anisotropy (MCA) is effectively a collinear phenomenon: atomic spins remain parallel or antiparallel, and only the orientation of the *net* magnetization changes. Therefore, the SOC contribution to the total energy can often be approximated from *non self-consistent* calculations based on a collinear reference state, where the SOC-induced energy correction is evaluated second-variationally from the scalar relativistic eigenstates.

A common and computationally efficient approach for evaluating such energy differences is the *magnetic force theorem* (MFT). Within the MFT approximation, the change in total energy due to SOC or a rotation of the magnetization direction can be estimated from the difference in non-self-consistent band energies obtained using the eigenstates of a collinear calculation [43]:

$$\mathrm{MAE} \approx \sum_i f_i(\theta_1,\phi_1)\,\epsilon_i(\theta_1,\phi_1) - \sum_i f_i(\theta_0,\phi_0)\,\epsilon_i(\theta_0,\phi_0), \tag{C1}$$

where $\epsilon_i(\theta,\phi)$ and $f_i(\theta,\phi)$ denote the eigenvalues and occupation numbers of the Kohn–Sham states for magnetization direction (θ,ϕ), respectively. This method exploits the variational property of the total energy with respect to the density, allowing one to bypass the expensive SCF procedure for each spin orientation.

Within the mMACE framework, the same orientation-dependent band-energy difference is learned directly from reference SOC DFT data from MFT. Typically, the scale of SOC effect is sub-meV per atom (see Fig. 6a), which is comparable to or smaller than the typical errors of modern MLIPs in single-point energy predictions. Nevertheless, within the standard workflow based on the magnetic force theorem, the task of approximating E_A reduces to estimating the SOC-induced difference in band energy

between a magnetization orientation (θ, ϕ) and a reference orientation (θ_0, ϕ_0).

$$\begin{aligned} E_A(\theta, \phi; \{\sigma_i\}_i) &\approx \delta E_{\text{band,soc}}(\theta, \phi; \{\sigma_i\}_i) \\ &= E_{\text{band,soc}}(\theta, \phi; \{\sigma_i\}_i) \\ &- E_{\text{band,soc}}(\theta_0, \phi_0; \{\sigma_i\}_i). \end{aligned} \tag{C2}$$

where $\{\sigma_i\}_i := \{(\mathbf{r}_i, \mathbf{m}_i, z_i)\}_i$ denotes a single configuration.

This allows the model to approximate the SOC-induced anisotropy landscape in a fully differentiable and computationally efficient manner, without explicitly invoking relativistic electronic-structure calculations.

Appendix D: Heisenberg exchange constants

1. Total energy difference method for Heisenberg exchange constant

In what follows we describe how Heisenberg exchange constants can be obtained from a trained mMACE model. We consider the Heisenberg Hamiltonian of the system

$$\mathcal{H} := \mathcal{H}_0 - \sum_{\langle ij \rangle} J'_{ij} \hat{\mathbf{S}}_i \cdot \hat{\mathbf{S}}_j \tag{D1}$$

where $\mathcal{H}_0$ denotes the non-magnetic part of the energy and J'_{ij} is the exchange interaction between spins $\hat{\mathbf{S}}_i$ and $\hat{\mathbf{S}}_j$. The summation goes over all spin pairs once. By considering symmetry and equivalent crystal sites (with equivalence classes)

$$\mathcal{H}_{\text{Heis}} = -\sum_k J_k c_k \tag{D2}$$

$$c_k = \sum_{\langle ij \rangle \in X_k} \hat{\mathbf{S}}_i \cdot \hat{\mathbf{S}}_j \tag{D3}$$

In practice, a finite cutoff $d_{\max}$ is chosen, which results in a finite number of equivalence classes N_k and all other exchange interaction energies are considered zero. The Heisenberg Hamiltonian then becomes

$$\mathcal{H}_{\text{Heis}} = -\sum_{k<N_k} J_k c_k \tag{D4}$$

For each magnetic configuration (indexed by l), one can construct an expression for the energy E^l with J_k as unknown

$$E^{(l)} = E_0 - \sum_{k<N_k} J_k c_k^{(l)}, \tag{D5}$$

where E_0 is the non-magnetic part of the total energy. This leads to a linear least squares system from which we can estimate J_k, taking the following form

$$\mathbf{E} = E_0 \mathbf{1} - \mathbf{C}\mathbf{J}, \tag{D6}$$

where $\mathbf{E} = [E^{(1)}, E^{(2)}, \ldots, E^{(L)}]^{\mathrm{T}}$ contains the total energies from L magnetic configurations and $\mathbf{C}_{lk} = c_k^{(l)}$ are the corresponding spin–spin correlation coefficients obtained from the dot products $\hat{\mathbf{S}}_i \cdot \hat{\mathbf{S}}_j$ summed over each neighbor shell k.

In practice, we include E_0 directly as an additional parameter to be fitted, rather than eliminating it by mean subtraction. This is achieved by appending a column of ones to the design matrix, resulting in the augmented linear system

$$\mathbf{E} = \mathbf{C}'\mathbf{x}, \qquad \mathbf{C}' = \begin{bmatrix} -c_1^{(1)} & \cdots & -c_{N_k}^{(1)} & 1 \\ \vdots & \ddots & \vdots & \vdots \\ -c_1^{(L)} & \cdots & -c_{N_k}^{(L)} & 1 \end{bmatrix}, \tag{D7}$$

where $\mathbf{x} = (J_1 \cdots J_{N_k} E_0)$. The parameters are then determined by solving the above least–squares problem, which yields both the exchange constants $\{J_k\}_{k<N_k}$ and the reference energy E_0 simultaneously.

2. Adiabatic approximation for Heisenberg exchange constant

In the real-space approach of Liechtenstein et al. [37–39], the adiabatic magnon energy of a spin spiral with wavevector $\mathbf{q}$ is given by [53]

$$E(\mathbf{q}) = \frac{1}{2} \sum_j J_{0j} \left[1 - \cos\big(\mathbf{q} \cdot \mathbf{r}_{0j}\big)\right], \tag{D8}$$

where J_{0j} are the real–space Heisenberg exchange interactions between a reference atom and neighbours at displacement $\mathbf{r}_{0j}$. The extra $\frac{1}{2}$ factor arises as convention of Eq. (34) counted each pair once.

To extract individual J_{ij} values, we use the small–angle rotation protocol. Starting from a collinear ferromagnetic state, the magnetic moments of sites i and j are rotated by $\pm\theta/2$ in opposite directions, while all other spins remain aligned. Denoting the resulting total–energy variations by

$$\delta E_i(\theta), \qquad \delta E_j(\theta), \qquad \delta E_{ij}(\theta),$$

the pair interaction energy is defined as

$$\Delta_{ij}(\theta) = \delta E_{ij}(\theta) - \big[\delta E_i(\theta) + \delta E_j(\theta)\big]. \tag{D9}$$

For sufficiently small rotation angles, the Heisenberg model predicts the quadratic relation

$$\Delta_{ij}(\theta) = J_{ij}\Big[(1-\cos\theta) - 2\big(1-\cos\tfrac{\theta}{2}\big)\Big] = \tfrac{1}{4} J_{ij}\, \theta^2 + \mathcal{O}(\theta^4). \tag{D10}$$

Hence, J_{ij} can be obtained via a linear fit of $\Delta_{ij}(\theta)$ versus $\theta^2/4$. In Refs. [48, 53] the corresponding relations are written in the convention $\mathcal{H} = -\sum_{i \neq j} J_{ij} \hat{\mathbf{S}}_i \cdot \hat{\mathbf{S}}_j$, in which each pair is counted twice; exchange constants in

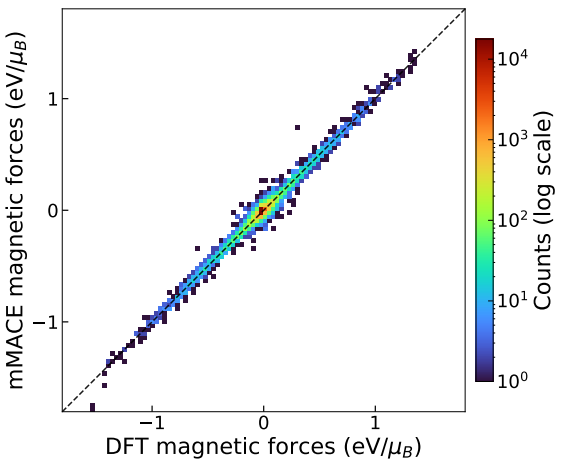


FIG. 10: 2D histogram comparing mMACE to DFT reference magnetic forces on the CrN dataset.

that convention are half of those used here, and reference values taken from these works are rescaled accordingly. This provides a computationally efficient route to estimating exchange interactions without explicitly constructing long–wavelength spirals.

It is important to emphasize that the effective Heisenberg exchange constants obtained from these calculations are not necessarily identical to those derived from the total energy difference method. Each approach serves only as an estimation of the exchange interaction through a different physical pathway, and the resulting values may therefore differ systematically (see Figs. 7a and 7b).

Appendix E: Experiment supplements

1. CrN public benchmark

Figure 10 shows a strong concentration of points along the diagonal, indicating excellent agreement between mMACE-predicted and DFT magnetic forces over the full force range. The narrow spread around the identity line, even in the low-force regime, demonstrates that mMACE accurately captures both the magnitude and sign of magnetic force responses in CrN.

2. Energy minimization on pre-trained datasets

Single-point evaluation metrics do not fully capture the behavior of magnetic degrees of freedom under minimization of the magnetic moments. The ground-state energy defined for an atomic configuration should be minimized with respect to the underlying internal variable $\{\mathbf{m}_i\}_i$ representing local magnetic moment. Different initial local magnetic moments can lead to qualitatively different relaxed spin configurations, especially in systems with competing magnetic states.

To quantify this effect, we analyze the consistency between energy-minmized and reference collinear magnetic moments across the MATPES-PBE, MATPES-r2SCAN, and MP-ALOE datasets. A magnetic moment minimization (i.e. relaxation) is classified as *fail* if the resulting minimization gives *NaN*, the energy goes above 100 eV / atom or forces go above 30 eV/Å. Around 2.02% (MATPES-PBE), 1.72% (MATPES-r2SCAN), and 1.29% (MP-ALOE) of the energy minimizations fail on the testing set. This is expected because we are only sampling a relatively narrow range of magnetic configurations from a small dataset; we hypothesize that including constrained DFT and/or applied fields would greatly increase the magnetic diversity.

Dataset	Energy	Force	Stress
MATPES-PBE	55.6	130.8	0.9969
MATPES-r2SCAN	61.8	154.4	1.1515
MP-ALOE	70.8	100.8	1.4239

TABLE II: Error metrics on testing set across the MATPES-PBE, MATPES-r2SCAN, and MP-ALOE datasets, reported as mean absolute errors (MAE) (units: energy in meV/atom, forces in meV/Å, stress in GPa).

For completeness, the corresponding accuracy on the subsets where the minimizations were successful are summarized in Table II. Figure 11 shows parity plots for energy, forces, stress, and magnetic moments across the three datasets. While energies and stresses exhibit strong agreement with DFT, the magnetic moments display a broader distribution, particularly in the at around $m_{\mathrm{DFT}} \approx 0$.

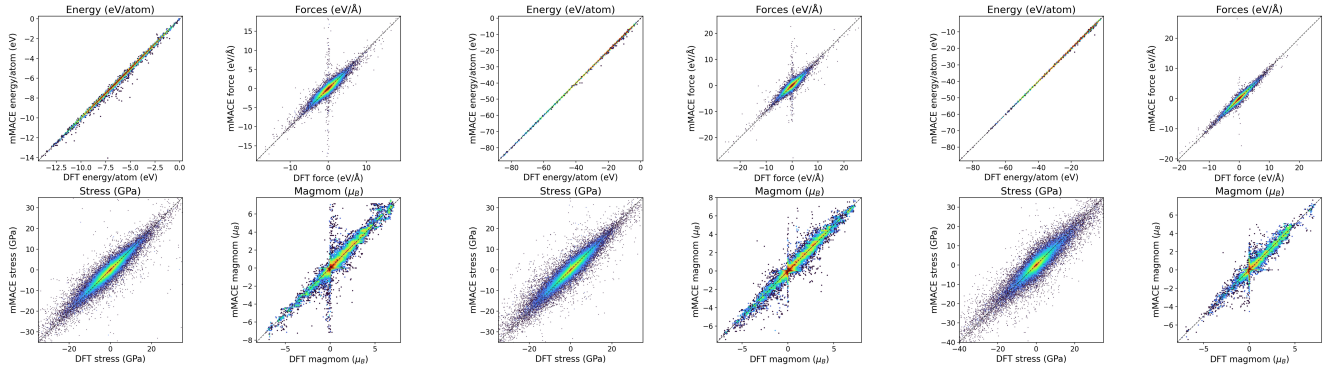
These discrepancies are not merely due to inaccurate model, but reflect an intrinsic ambiguity in the mapping between atomic structure and equilibrium magnetic configuration within spin DFT, leading to inconsistent dataset. More precisely, the mapping

$$\mathbf{R} \mapsto \{\mathbf{m}_i^*\} \tag{E1}$$

where $\mathbf{m}^*$ indicates a set of local magnetic moment from self-consistent magnetization density, is not single-valued.

We also note that local magnetic moments are not uniquely defined quantities, but depend on the partitioning scheme used to decompose the electronic density. Different schemes (e.g., Wigner-Seitz, Hirshfeld, or Bader-type approaches) can assign different values to atomic moments even when derived from the same charge density. This dependence is particularly relevant in systems with delocalized magnetization or strong hybridization, where the spatial distribution of spin density is diffuse.

a. Representative failure case. We illustrate this behavior using a representative 3-atom TbSbPt bulk configuration from the MATPES-PBE test set (MATPES



(a) MATPES-PBE (b) MATPES-r2SCAN (c) MP-ALOE

FIG. 11: Parity plots for energy, forces, stress, and magnetic moments across the three datasets. Energies and forces show strong agreement with DFT, while magnetic moments exhibit increased dispersion, particularly near $m \approx 0$, reflecting the presence of competing magnetic states.

id: `mp-16313`). The reference DFT solution corresponds to a nearly non-magnetic state with total moment $m \approx 0.34\,\mu_B$, whereas mMACE predicts a strongly spin-polarized configuration with a local moment of approximately $5.67\,\mu_B$ on the Tb site, indicating a qualitative change in magnetic state.

To determine which solution is energetically favorable, due to DFT software availability, we perform self-consistent DFT calculations in GPAW (rather than VASP used in MATPES) using two different initializations: (i) the reference DFT magnetic moments and (ii) the mMACE-predicted magnetic configuration. The resulting total energies are $E_{\mathrm{DFT}}^{\mathrm{ref}} = -16.55$ eV and $E_{\mathrm{DFT}}^{\mathrm{mMACE}} = -21.84$ eV, showing that the spin-polarized state obtained from the mMACE initialization is significantly lower in energy.

Method	Tb	Sb	Pt
DFT (VASP reference)	0.339	0.001	0.001
mMACE (predicted)	5.67	0.0004	0.363
GPAW (init from DFT)	-0.002	-0.00017	-0.0012
GPAW (init from mMACE)	5.82	-0.0098	0.0013

TABLE III: Converged local magnetic moments (in μ_B) for TbSbPt under different initializations. The VASP reference (and hence the mMACE prediction, which is trained on VASP moments) and the GPAW calculations use different schemes to partition the magnetization density among atoms, so moments from different codes should be compared only qualitatively.

Table III shows that the two initializations converge to distinct self-consistent solutions: a non-magnetic state when initialized from the DFT reference, and a strongly spin-polarized state when initialized from the mMACE prediction. Because local moments from VASP and GPAW are obtained with different partitioning schemes (see above), we do not interpret the precise values across codes; the conclusion relies only on the qualitative distinction between a nearly non-magnetic state and a strongly polarized Tb moment of ~ 5.7–$5.8\,\mu_B$, which is far larger than any partition-dependent differences, and on the total energies computed within GPAW. The qualitative agreement between the mMACE-predicted moments and the DFT solution obtained from the same initialization further indicates that mMACE identifies a stable magnetic branch of the electronic structure, rather than producing an unphysical configuration.

b. Internal inconsistency, role of similar structures and implications for MLIPs. To understand why mMACE converges to a strongly spin-polarized state for the TbSbPt configuration, we examine whether similar magnetic environments are present in the MATPES training data. We find that the dataset contains structurally similar configurations with magnetic states closely matching the mMACE prediction.

In particular, structurally similar bulk configurations that is derived from the same entry in Materials Project as the above structure exhibits a spin states (with mean absolute error up to $0.2\mu B$.) that is very close to mMACE-relaxed value. The existence of these configuration in the training set, together with similar bulk configuration such as TbNiSb (MATPES id: `mp-3716`) that also exhibits a similar local magnetic moment, contributes to the fact that mMACE converges to the lower energy spin states with lower DFT energy of -21.57 eV in the TbSbPt structure as given by GPAW in Table III above.

More broadly, this highlights a fundamental limitation when fitting non-magnetic MLIPs to datasets such as MATPES: they implicitly mix multiple self-consistent magnetic solutions of the Born-Oppenheimer surface without distinguishing between them. As a result, similar atomic environments can correspond to qualitatively different magnetic states and substantially different energies, violating the assumption of a single-valued mapping $E(\mathbf{R})$.

From a modeling perspective, this introduces intrinsic ambiguity into the learning problem. Apparent errors in predicted magnetic moments may instead reflect selection of an alternative valid magnetic branch rather than model failure. In this case, the mMACE prediction corresponds to a state that is both self-consistently stable under DFT and lower in energy than the provided reference.

These observations suggest that magnetic datasets should be interpreted as sampling a multi-branch energy landscape unless only the global minimum branch across different configurations is included, and that evaluation of magnetic MLIPs should go beyond single-point comparisons to incorporate relaxation-based validation or branch-aware metrics.

3. FeNi lattice parameters

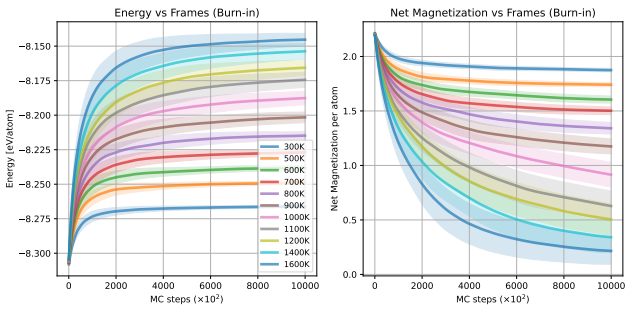
As shown in Table IV, equilibrium lattice parameters agree with DFT within 0.01 Å across all phases and magnetic states. Elastic constants are reproduced with correct trends and typical deviations of 10%, with larger errors confined to non-magnetic phases due to limited sampling.

4. Mn_3Pt relaxations

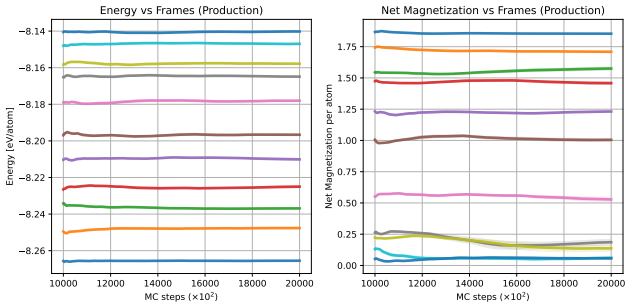
To perform the relaxation experiment, we first set up a fixed atomic configuration from a reference geometry optimization of the standard cubic Mn_3Pt structure. We then apply random 3D rotations to the reference magnetic moments with angles sampled uniformly from $[0, \pi]$, followed by uniform random scaling from $[0, 1.5]$ on each Mn atom. Exactly 100 such magnetic orientations are generated, on which we perform relaxation of the magnetic moments with the LBFGS optimizer until the maximum gradient reaches 10^{-4}. The magnetic moment of the Pt atom is kept fixed as it is known to be highly non-magnetic in this system, with a local magnetic moment on the order of $10^{-4}\,\mu_B$.

5. MC simulation details

The mMACE MC simulations sample only the magnetic degrees of freedom; atomic positions are kept fixed at the relaxed bcc geometry. Each MC step proposes a single-site move: an atom i is chosen at random and its moment $\mathbf{m}_i$ is rotated by an angle drawn uniformly from $[-\theta_{\max}, \theta_{\max}]$, with $\theta_{\max} = 1$ rad, about a random axis perpendicular to $\mathbf{m}_i$. This move changes only the orientation of the moment, so moment magnitudes are held fixed (no longitudinal fluctuations). The trial energy is evaluated directly with mMACE, without any relaxation of the magnetic moments, and the move is accepted with the Metropolis probability



(a) Burn-in stage to equilibrate the system.



(b) Production stage for statistical averaging.

FIG. 12: Convergence behavior of MC simulations.

$\min\{1, \exp(-\Delta E/k_B T)\}$. Figure 12 demonstrates convergence of the MC simulation in production and burn-in stages as mentioned in Section III F 2.

Curie temperature calculations were performed using a bcc Fe supercell containing 3456 atoms. The Curie temperature T_C was extracted by fitting $M(T) \propto (T_C - T)^\beta$ where both T_C and the critical exponent β were treated as fitting parameters.

The Curie temperature was estimated by fitting $M(T) \propto (T_C - T)^\beta$ where both T_C and β were treated as fitting parameters. The VAMPIRE simulations use the default settings [2, 24]. The VAMPIRE simulations were performed using J_{ij} for up to 6 and 5 nearest neighbours for the adiabatic and total-energy methods, respectively, to match the number of Heisenberg exchange constants included in existing references. We found that the results tend to be insensitive to exchange interactions beyond the 4th neighbour.

Appendix F: A Non-SOC mMACE Variant

We briefly summarize a variant of mMACE that implements the more restrictive non-SOC symmetry (11). The main modification is that the tensors' (l, m) channels are replaced by (l, m, l', m') channels to independently encode rotations acting on positions and magnetic moments. With notation as in Section II, for the first layer $(t = 1)$ we initialize the features as a one-hot embedding

Structure	**Mag.**	a (Å)	C_{11}	C_{12}	C_{44}	ϵ_{M_z} (μ_B/atom)
fcc	FM	3.57 (3.57)	1.666 (1.507)	1.080 (0.885)	0.812 (0.787)	0.007
fcc	AFM	3.57 (3.56)	1.637 (1.332)	1.071 (0.961)	0.815 (0.644)	0.141
fcc	NM	3.50 (3.50)	2.230 (2.228)	1.280 (1.259)	0.964 (1.181)	0.000
bcc	FM	3.59 (3.59)	1.992 (2.026)	0.273 (0.243)	0.865 (0.789)	0.101
bcc	AFM	3.59 (3.58)	1.992 (1.628)	0.272 (0.189)	0.868 (0.776)	0.013
bcc	NM	3.51 (3.51)	1.833 (2.176)	0.408 (0.421)	0.892 (1.121)	0.000

TABLE IV: Equilibrium lattice constants a, magnetization error ϵ_{M_z}, and elastic constants C_{11}, C_{12}, and C_{44} (in eV/Å^3) for bcc and fcc FeNi under different magnetic states. Elastic constants are reported as mMACE predictions with reference (DFT) values shown in parentheses.

of the chemical species

$$h^{(1)}_{j,k0000} := \sum_z W_{kz_j}\delta_{zz_j}. \tag{F1}$$

This defines the node state $\tilde{\sigma}_i := (\mathbf{r}_i, \mathbf{m}_i, z_i, h_i^{(t)})$ for message passing, as in (1).

With identical initialization of $P^{(t)}_{ji,kl_1l_2l_3}$ and $K^{(t)}_{ji,kl_1l_2l_3}$ as in (16) and (17) respectively, the two-body features with magnetic information can be constructed through the following sequence of operations:

$$\begin{aligned}\phi^{(t),\mathrm{pos}}_{ji,kl_3m_3} :=& \sum_{l_1m_1l_2m_2} C^{l_3m_3}_{l_1m_1l_2m_2} \\ &\cdot P^{(t)}_{ji,kl_1l_2l_3} Y^{m_1}_{l_1}(\hat{\mathbf{r}}_{ji}) h^{(t)}_{j,kl_2m_200}\end{aligned} \tag{F2}$$

$$\begin{aligned}\phi^{(t),\mathrm{mag}}_{ji,kl_3'm_3'} :=& \sum_{l_1'm_1'l_2'm_2'} C^{l_3'm_3'}_{l_1'm_1'l_2'm_2'} \\ &\cdot K^{(t)}_{ji,kl_1'l_2'l_3'} R^{m_1'}_{l_1'}(\mathbf{m}_j) h^{(t)}_{j,k00l_2'm_2'}\end{aligned} \tag{F3}$$

$$\phi^{(t),\mathrm{mag+pos}}_{ji,klml'm'} := \phi^{(t),\mathrm{pos}}_{ji,klm}\phi^{(t),\mathrm{mag}}_{ji,kl'm'} \tag{F4}$$

$$A^{(t)}_{i,klml'm'} := \sum_{j\in\mathcal{N}(i)} \phi^{(t),\mathrm{mag+pos}}_{ji,klml'm'}. \tag{F5}$$

Here, $\hat{\mathbf{r}}_{ji} := \mathbf{r}_{ji}/|\mathbf{r}_{ji}|$, $Y^{m_1}_{l_1}$ are real spherical harmonics, and $R^{m_1'}_{l_1'}$ are solid harmonics which guarantee smoothness in the limit $|\mathbf{m}| \to 0$.

As in the main text, this basis follows the non-collinear magnetic ACE construction of Drautz [21]. Once the two-body basis $A^{(t)}_{i,klml'm'}$ is formed, the many-body basis for each correlation order is constructed as

$$\begin{aligned}\mathbf{B}^{(t)}_{i,\eta_\nu\eta'_\nu kLML'M'} :=& \sum_{\mathbf{l}'\mathbf{m}'} C^{L'M'}_{\eta'_\nu\mathbf{l}'\mathbf{m}'} \sum_{\mathbf{lm}} C^{LM}_{\eta_\nu\mathbf{lm}} \\ &\prod_{\xi=1}^{\nu}\sum_{\tilde{k}} W^{(t)}_{k\tilde{k}l_\xi l'_\xi} A^{(t)}_{i,\tilde{k}l_\xi m_\xi l'_\xi m'_\xi},\end{aligned} \tag{F6}$$

where, $C^{LM}_{\eta_\nu\mathbf{lm}}$ are generalized Clebsch–Gordan coefficients [5]. Note that in this case, no coupling is involved between $\mathbf{l}$ and $\mathbf{l}'$, hence the CG contractions can be performed independently and parallelized. The many-body message is obtained by summing over contributions from each correlation order ν

$$\mu^{(t)}_{i,kLML'M'} := \sum_\nu \sum_{\eta_\nu\eta'_\nu} W^{(t),\nu}_{z_i\eta_\nu\eta'_\nu kLL'} \mathbf{B}^{(t)}_{i,k\eta_\nu\eta'_\nu LML'M'}. \tag{F7}$$

To include the contribution from the magnetic moment of atom i itself and correlate it with its atomic environment, an additional contraction is performed,

$$\begin{aligned}&\tilde{\mu}^{(t)}_{i,klmL'M'} \\ &:= \sum_{l_0m_0l'm'} C^{L'M'}_{l_0m_0l'm'} T^{(t)}_{i,kl_0l'L'} R^{m_0}_{l_0}(\mathbf{m}_i)\mu^{(t)}_{i,klml'm'}.\end{aligned} \tag{F8}$$

Here $T^{(t)}_{i,kl_0l'L}$ is again a learnable radial function analogous to (17) but depends only on the species and magnetic moment magnitude of atom i.

Across layers, the node feature update is completed via a residual connection, following standard practice in modern neural network architectures

$$\begin{aligned}\tilde{h}^{(t+1)}_{i,kLML'M'} :=& \sum_{\tilde{k}} W^{(t),\mathrm{self}}_{kLL',\tilde{k}} \tilde{\mu}^{(t)}_{i,\tilde{k}LML'M'} \\ &+ \sum_{\tilde{k}} W^{(t)}_{kLL',\tilde{k}} \mu^{(t)}_{i,\tilde{k}LML'M'} \\ &+ \sum_{\tilde{k}} W^{(t)}_{kz_iLL',\tilde{k}} h^{(t)}_{i,\tilde{k}LML'M'}.\end{aligned} \tag{F9}$$

The total energy is then obtained as a sum of readouts (either linear or non-linear MLPs) applied to the invariant node features. Additionally, a one-body magnetic moment contribution is introduced to ensure the correct large volume limit (*i.e.* as $|\mathbf{r}_{ji}| \to \infty$ for all i, j). This term is pre-trained or fitted as a simple linear expansion of Chebyshev polynomials as discussed in the main model architecture as described in Section II:

$$\begin{aligned}E_i :=& \sum_t \left(\mathcal{R}_t\left(h^{(t)}_{i,\tilde{k}LML'M'}\right)\right) + \sum_{nz} W_{nz}\delta_{zz_i} T_n(x(|\mathbf{m}_i|)) \\ =:& E^{\mathrm{inter}}_i + E_{0,z_i}(|\mathbf{m}_i|).\end{aligned} \tag{F10}$$

This completes a single forward pass of a non-SOC model.

Appendix G: Hyperparameters for parameter estimation

Unless otherwise stated, all models are optimized with Adam optimizer with learning rate 0.005. AMSgrad and EMA are applied with default parameters in the MACE package. A gradient clip of 100 is applied. Models are trained using a composite loss function that balances energies, forces, and stresses, with relative weights $w_E = 1$, $w_F = 10$, and $w_\sigma = 5$, respectively. For datasets that include magnetic-force labels $-\nabla_{\mathbf{m}_i} E$ (here, the CrN dataset), an additional magnetic-force term with weight $w_{\mathrm{mag}} = 1$ is included in the loss. Data augmentation is applied whenever the underlying DFT of the input data does not include SOC effects. A list of major hyperparameters for the model architecture is summarized in Table V.

1. FeAl public dataset

We employ a single layer model with scalar hidden irreducible representation `64×0e`. Body order correlations are included up to order $\nu = 3$, with a real-space interaction cutoff of $r_{\max} = 5.0$ Å. Angular dependencies of position and magnetic moments are resolved using spherical harmonics up to $\ell^{\mathrm{pos}}_{\max} = 3$ and $\ell^{\mathrm{mag}}_{\max} = 1$ respectively. The radial dependence of position and magnetic degree of freedom is expanded in a 10 and 8 Bessel basis and Chebyshev basis respectively. The magnetic moments are restricted to the range $M_{\mathrm{max,Al}} = 0.15$ and $M_{\mathrm{max,Fe}} = 4.0$ respectively. We also enforce one-body contributions to be simply constant for fair benchmarking with mMTP results without additional data.

2. CrN public dataset

We employ a single layer model with scalar hidden irreducible representation `64×0e`. Body order correlations are included up to order $\nu = 3$, with a real-space interaction cutoff of $r_{\max} = 5.0$ Å. Angular dependencies of position and magnetic moments are resolved using spherical harmonics up to $\ell^{\mathrm{pos}}_{\max} = 3$ and $\ell^{\mathrm{mag}}_{\max} = 1$ respectively. The radial dependence of position and magnetic degree of freedom is expanded in a 10 and 8 Bessel basis and Chebyshev basis respectively. The magnetic moments are restricted to the range $M_{\mathrm{max,Al}} = 0.15$ and $M_{\mathrm{max,Fe}} = 4.0$ respectively. We also enforce one-body contributions to be simply constant for fair benchmarking with mMTP results without additional data.

3. Pretrained models on foundational datasets

We employ a two layer model with scalar hidden irreducible representation `128×0e + 128 ×1o`. Body order correlations are included up to order $\nu = 3$, with a real-space interaction cutoff of $r_{\max} = 6.0$ Å. Angular dependencies of position and magnetic moments are resolved using spherical harmonics up to $\ell^{\mathrm{pos}}_{\max} = 3$ and $\ell^{\mathrm{mag}}_{\max} = 1$ respectively. The radial dependence of position and magnetic degree of freedom is expanded in a 10 and 8 Bessel basis and Chebyshev basis respectively. The magnetic moments are restricted to the range $M_{\mathrm{max,Al}} = 0.15$ and $M_{\mathrm{max,Fe}} = 4.0$ respectively. We also enforce one-body contributions to be simply constant for fair benchmarking without additional data to existing MACE architecture.

4. FeNi dataset

Identical model hyperparameters to the pre-trained model MATPES-PBE are used as the model is fine-tuned from the MATPES-PBE mMACE model. A lower learning rate of 0.0005 and a gradient clip of 10 instead of 100 are applied.

5. Magnetocrystalline anisotropy datasets

For fitting both FeMn and the general database constructed from random structure search, we employ a two layer model with scalar hidden irreducible representation `64×0e + 64×1o`. Body order correlations are included up to order $\nu = 3$, with a real-space interaction cutoff of $r_{\max} = 6.0$ Å. Angular dependencies of position and magnetic moments are resolved using spherical harmonics up to $\ell^{\mathrm{pos}}_{\max} = 3$ and $\ell^{\mathrm{mag}}_{\max} = 3$ respectively. The radial dependence of position and magnetic degree of freedom is expanded in a 10 and 8 Bessel basis and Chebyshev basis respectively.

The SOC configuration identification task uses the same model architecture as the MATPES pre-trained model without the constant energy shift, but is trained from scratch (i.e. randomly initialised rather than fine-tuned from the MATPES model). The value of ΔE_A are computed from evaluating the trained models on the 2000 sampled (θ, ϕ) which we generated the reference values from for consistency.

During training, the validation error is observed to increase monotonically after the first epoch, indicating rapid overfitting due to the limited number of parent configurations. As a result, we use the model trained for a single epoch for all SOC identification results. Even after this single epoch of training, the model already exhibits strong rank correlation with DFT reference SOC energy spreads, indicating that SOC-sensitive configurations can be identified from very limited SOC data.

Because SOC-induced energy differences are typically below 1 meV/atom, all SOC energy labels were multiplied by a factor of 10^3 (i.e. expressed in meV rather than eV) during training to improve the numerical conditioning of the loss. Model predictions are rescaled by 10^{-3}

Hyperparameter	Description
Hidden irreps	Feature dimension in the latent representation
Number of layers	Depth of the message-passing architecture
Body order (ν)	Maximum order of many-body correlations included
Cutoff radius ($r_{\max}$)	Spatial range of atomic interactions
Angular order (position, $\ell_{\max}^{\text{pos}}$)	Maximum angular resolution for positional features
Angular order (magnetic, $\ell_{\max}^{\text{mag}}$)	Maximum angular resolution for magnetic features
Radial basis size (position)	Number of radial basis functions for positions
Radial basis size (magnetic)	Number of radial basis functions for magnetic moments
Magnetic cutoff ($M_{\max,Z}$)	Maximum allowed magnetic moment for atoms of species type Z

TABLE V: Summary of key hyperparameters in the mMACE architecture.

before evaluation, so that all reported energies, errors, and anisotropy values are given in the original units.

6. Mn_3Pt dataset

Identical model hyperparameters to the pre-trained model MATPES-PBE are used as the model is fine-tuned from the MATPES-PBE mMACE model. A lower learning rate of 0.0005 and a gradient clip of 10 instead of 100 are applied.

7. Fe dataset

We employ a single layer model with scalar hidden irreducible representation `64×0e`. Body order correlations are included up to order $\nu = 3$, with a real-space interaction cutoff of $r_{\max} = 4.5$ Å. Angular dependencies of position and magnetic moments are resolved using spherical harmonics up to $\ell_{\max}^{\text{pos}} = 3$ and $\ell_{\max}^{\text{mag}} = 3$ respectively. The radial dependence of position and magnetic degree of freedom is embedded in a 10 and 8 Bessel basis and Chebyshev basis respectively before feeding into MLPs. The magnetic moments are restricted to $M_{\max,\text{Fe}} = 4.0$.

Appendix H: Data generation

1. Pretrained models on foundational datasets

A filtering process was performed to filter configurations that contain extreme DFT ground truth forces, as we found it improves the stability of the model in each corresponding reference dataset in [32] and [35] respectively. Configurations with forces larger than 20 eV/Å are removed from the dataset. For MATPES-PBE and MATPES-r2SCAN, training, validation and testing set from the official split with the above filtering. For MP-ALOE, only the configuration from the final ionic step of each configuration are included as those are the only configurations given with magnetic moment.

2. FeNi dataset

First, we generate a broad family of unit-cell–distorted structures by applying random perturbations to an ordered bcc FeNi alloy unit cell, with Fe and Ni arranged in an Fe–Ni–Fe–Ni sequence. The fcc base structures and a structure with cell ratio $a/c = 1.6$ are used to generate 5 variants by applying an isotropic volume scaling with a random scale factor uniformly sampled from $[0.6, 1.4]$, followed by adding Gaussian random displacements with $\sigma = 0.05\mathring{A}$ to all atomic positions. We then assign part of the data an initial guess of magnetic moments $[2.2, 0.6, 2.2, 0.6]$ μ_{B}, and the corresponding AFM magnetic arrangement, corresponding to the equilibrium values for bcc Fe and Ni, respectively. We also apply spin-flip operations, as well as scaling the magnetic moment of each atom with uniform random scaling sampled from $[0, 1]$. This stage provides diverse and reasonably off-equilibrium configurations.

Second, we perform direct DFT relaxations of both base structures and sample intermediate configurations along each relaxation trajectory. These data capture the energetics and magnetic responses along the physically relevant transformation pathway between the two phases and their respective minima.

This completes the first iteration of dataset generation. We use the resulting dataset to perform a the first round of fine-tuning. The resulting Bain path is shown by the blue line in Figure 3c (labeled as iter1). The poor performance, in particular the model's confusion between the FM and AFM states, indicates that more extensive sampling is required.

In the second iteration of data generation we enrich the dataset with off-equilibrium yet physically sensible configurations. To that end, we generate additional structures by applying Monte Carlo–like perturbations to both atomic positions and lattice degrees of freedom. Starting from a reference structure, random Gaussian displacements are applied to the atomic positions, together with perturbations to the simulation cell. Configurations with trial Monte Carlo moves (with admissible moves sampled from random position and cell rattling) are evaluated us-

ing the fine-tuned mMACE model, and acceptance is determined via a Metropolis criterion based on the Boltzmann probability at 2000 K to sufficiently generate a wide range of cell and positions samples. This ensures dense coverage of the a/c ratio and cell volumes relevant to the Bain path, as shown in Figure 3c.

To further improve the quality of the Bain path, we additionally sample the AFM Bain path generated by the previous model. These configurations are subsequently perturbed in atomic positions and magnetic moments, and evaluated using single-point DFT calculations. This final stage enriches the dataset with configurations that are energetically accessible but rarely sampled by direct relaxations.

A second fine-tuning procedure is performed using the resulting dataset, which contains approximately 244 configurations. The resulting Bain path is shown in Figure 3c (labeled as iter2), and already closely matches the DFT reference. This provides strong evidence that mMACE is sufficiently expressive to capture the physics governing this transformation.

Testing set are generated from randomly drawing samples from the configurations generated in iteration 1. For all the calculations in this example, spin-polarized density functional theory calculations were performed using GPAW with a plane-wave basis set and the PBE exchange–correlation functional. A kinetic energy cutoff of 600 eV was used for the plane-wave expansion. Brillouin-zone integrations were carried out using Monkhorst–Pack grids generated from a reciprocal-space density of approximately 0.03 Å^{-1}. All calculations were performed within the collinear spin approximation without spin–orbit coupling. Atomic positions were relaxed until the residual forces were below 1×10^{-4} eV/Å, while the simulation cell was kept fixed. Electronic self-consistency was achieved using GPAW's default convergence settings.

3. Magnetic anisotropy constant

When generating data in this section, for each dataset all angles are defined with a fixed coordinate frame, i.e., ϕ and θ are always pointing to x and z respectively as it has to be consistent over different parent configurations for the fitting to make sense.

In the experiment that follows, for each parent configuration, we uniformly sample spin orientations over the unit sphere and evaluate the corresponding band energy differences (C2) for each orientation. These energy differences are then used as the training targets, and the fitting procedure is carried out in the same manner as for energy fitting in MLIPs with mMACE.

a. FeMn dataset

To construct the training dataset, we start from a fully relaxed collinear DFT structure. From this fixed atomic configuration, we uniformly sample 10,000 magnetization orientations over the unit sphere and compute the corresponding SOC band energies. Calculations are run with spin-polarized DFT in GPAW with the PBE functional with identical settings as FeNi. Magnetocrystalline anisotropy energies were computed using `spinorbit.soc_eigenstates` by evaluating SOC-induced band energy differences for fixed magnetization directions. We emphasize that this dataset is intentionally over-sampled and is designed primarily to assess the expressiveness and fidelity of the proposed framework, rather than to represent a minimal or production-level training protocol.

b. Random structure search (RSS) dataset

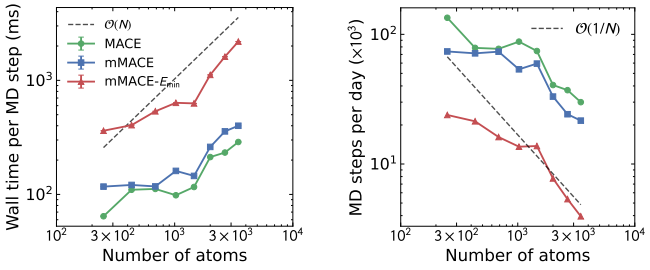
We generated a large database of random ternary structures within the chemical space spanned by Al, Mn, Fe, Co, Ni, Y, Pt, optionally combined with one light element from B, C, Si, using random structure search (RSS) with unit cells containing up to 10 atoms. Structure generation was performed using `buildcell` driven by MACE-MPA-0, and the resulting configurations were screened using MACE-MatPES-PBE-0 to evaluate energies, forces, and pressures.

From this pool, 1000 structures were sampled with a bias to match the elemental distribution of the parent dataset. Magnetic orderings were generated for each structure using the `MagneticStructureEnumerator` implemented in `pymatgen` [47], version 2025.10.7. These structures were relaxed using DFT (Quantum ESPRESSO [26]), comprising non-magnetic, ferromagnetic, ferrimagnetic, and antiferromagnetic states.

Magnetic anisotropy calculations were subsequently performed using GPAW [43]. For nominally non-magnetic systems, the self-consistent calculations were initialized with small random magnetic moments to enable SOC-induced anisotropy. In total, around 242 collinear calculations were used in the subsequent analysis and 2000 are sampled on the sphere for each parent configuration.

4. Mn_3Pt dataset

Training structures were generated from the four-atom primitive unit cell of Mn_3Pt, comprising three Mn atoms forming the frustrated kagome sublattice and one Pt atom occupying the face-centered position. We first performed a reference density-functional theory (DFT) calculation by fully relaxing the structure initialized according to Ref. [64], using Quantum ESPRESSO [26] with the PBE exchange–correlation functional and non-relativistic



(a) (b)

FIG. 13: **Performance scaling of magnetic interatomic potentials on bcc Fe.** (**a**) Wall-clock time per molecular dynamics step (energy and forces) as a function of system size. (**b**) Corresponding molecular dynamics throughput, reported as the number of MD steps per day. Both panels compare the baseline MACE model and its magnetic extension (mMACE), demonstrating near-linear scaling with system size and a moderate overhead associated with the magnetic degrees of freedom.

pseudopotentials. Single-point density functional theory (DFT) calculations were performed for a total of 147 configurations using Quantum ESPRESSO [26], with a plane-wave energy cutoff of 816 eV and a k-point density of 0.29 Å^{-1}.

Appendix I: Performance

In this section, we perform a performance analysis with the current implementation of mMACE against the original MACE, using the MATPES pre-trained model in Section III B.

To evaluate the difference in performance between original and mMACE we compare the time per MD steps in Fig. 13. Usual MACE, mMACE and mMACE with energy minimized at every MD step until the gradient of energy with respect to the magnetic moment converges up to 10^{-3} are compared. In the case where energy is minimized, the energy between subsequent minimization updates is less than 10^{-3} meV / atom. We found that mMACE takes longer to converge in the first energy minimization (for example, on a 250 atoms supercell, it requires 18 energy minimization steps), but requires fewer iteration later (4-10 energy minimization steps), analogous to re-using electron densities in DFT.

[1] Jaime Garrido Aldea, Dorye L Esteras, Stephan Roche, and Jose H Garcia. Challenges and strategies for first-principles simulations of two-dimensional magnetic phenomena. *Nanoscale*, 17(43):24955–24989, 2025.

[2] J. D. Alzate-Cardona, D. Sabogal-Suárez, R. F. L. Evans, and E. Restrepo-Parra. Optimal phase space sampling for Monte Carlo simulations of Heisenberg spin systems. *Journal of Physics: Condensed Matter*, 31(9):095802, 2019.

[3] Ilyes Batatia, Simon Batzner, Dávid Péter Kovács, Albert Musaelian, Gregor N. C. Simm, Ralf Drautz, Christoph Ortner, Boris Kozinsky, and Gábor Csányi. The design space of E(3)-equivariant atom-centred interatomic potentials. *Nature Machine Intelligence*, 7:56–67, 2025.

[4] Ilyes Batatia, Philipp Benner, Yuan Chiang, Alin M Elena, Dávid P Kovács, Janosh Riebesell, Xavier R Advincula, Mark Asta, Matthew Avaylon, William J Baldwin, et al. A foundation model for atomistic materials chemistry. *The Journal of chemical physics*, 163(18), 2025.

[5] Ilyes Batatia, Mario Geiger, Jose Munoz, Tess Smidt, Lior Silberman, and Christoph Ortner. A general framework for equivariant neural networks on reductive lie groups. *Advances in Neural Information Processing Systems*, 36:55260–55284, 2023.

[6] 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.

[7] 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.

[8] Anton Bochkarev, Yury Lysogorskiy, and Ralf Drautz. Graph atomic cluster expansion for semilocal interactions beyond equivariant message passing. *Physical Review X*, 14(2):021036, 2024.

[9] Zefeng Cai, Ke Wang, Yong Xu, Su-Huai Wei, and Ben Xu. A self-adaptive first-principles approach for magnetic excited states. *Quantum Frontiers*, 2:21, 2023.

[10] Peter Campbell. Permanent magnet materials and their application. 1994.

[11] Jacob B. J. Chapman and Pui-Wai Ma. A machine-learned spin-lattice potential for dynamic simulations of defective magnetic iron. *Scientific Reports*, 12:22451, 2022.

[12] Non collinear Fe dataset authors. Non-collinear fe dataset, 2025.

[13] Jack B Collings, Ricardo Rama-Eiroa, Rubén M Otxoa, Richard FL Evans, and Roy W Chantrell. Generalized form of the magnetic anisotropy field in micromagnetic and atomistic spin models. *Physical Review B*, 107(6):064413, 2023.

[14] MATPES Consortium. Matpes dataset. `https://matpes.ai/`, 2024.

[15] Jun Cui, Matthew Kramer, Lin Zhou, Fei Liu, Alexander Gabay, George Hadjipanayis, Balamurugan Balasubramanian, and David Sellmyer. Current progress and future challenges in rare-earth-free permanent magnets. *Acta Materialia*, 158:118–137, 2018.

[16] James Darby, Joe David Morrow, Albert Bartók, Volker L. Deringer, Gábor Csányi, and Christoph Ortner. Regularity priors for the linear atomic cluster expansion. *Machine Learning: Science and Technology*, 2026.

[17] MP-ALOE dataset authors. Mp-aloe dataset, 2025.

[18] P. H. Dederichs, S. Blügel, R. Zeller, and H. Akai. Ground states of constrained systems: Application to cerium impurities. *Physical Review Letters*, 53(26):2512–2515, 1984.

[19] Bowen Deng, Peichen Zhong, KyuJung Jun, Janosh Riebesell, Kevin Han, Christopher J. Bartel, and Gerbrand Ceder. CHGNet as a pretrained universal neural network potential for charge-informed atomistic modelling. *Nature Machine Intelligence*, 5(9):1031–1041, 2023.

[20] M. Domina, M. Cobelli, and S. Sanvito. Spectral neighbor representation for vector fields: Machine learning potentials including spin. *Physical Review B*, 105(21):214439, 2022.

[21] Ralf Drautz. Atomic cluster expansion of scalar, vectorial, and tensorial properties including magnetism and charge transfer. *Physical Review B*, 102(2):024104, 2020.

[22] Marco Eckhoff and Jörg Behler. High-dimensional neural network potentials for magnetic systems using spin-dependent atom-centered symmetry functions. *npj Computational Materials*, 7(1):170, 2021.

[23] Alexander Edström, Jonathan Chico, Adam Jakobsson, Anders Bergman, and Jan Rusz. Electronic structure and magnetic properties of l 1 0 binary alloys. *Physical Review B*, 90(1):014402, 2014.

[24] R F L Evans, W J Fan, P Chureemart, T A Ostler, M O A Ellis, and R W Chantrell. Atomistic spin model simulations of magnetic nanomaterials. *Journal of Physics: Condensed Matter*, 26(10):103202, feb 2014.

[25] Johannes Gasteiger, Florian Becker, and Stephan Günnemann. Gemnet: Universal directional graph neural networks for molecules. *Advances in Neural Information Processing Systems*, 34:6790–6802, 2021.

[26] Paolo Giannozzi, Stefano Baroni, Nicola Bonini, Matteo Calandra, Roberto Car, Carlo Cavazzoni, Davide Ceresoli, Guido L Chiarotti, Matteo Cococcioni, Ismaila Dabo, et al. Quantum ESPRESSO: a modular and open-source software project for quantumsimulations of materials. *Journal of physics: Condensed matter*, 21(39):395502, 2009.

[27] Oliver Gutfleisch, Matthew A. Willard, Ekkes Brück, Christina H. Chen, S. G. Sankar, and J. Ping Liu. Magnetic materials and devices for the 21st century: Stronger, lighter, and more energy efficient. *Advanced Materials*, 23(7):821–842, 2011.

[28] Mojtaba Haghighatlari, Jie Li, Xingyi Guan, Oufan Zhang, Akshaya Das, Christopher J Stein, Farnaz Heidar-Zadeh, Meili Liu, Martin Head-Gordon, Luke Bertels, et al. NewtonNet: a newtonian message passing network for deep learning of interatomic potentials and forces. *Digital Discovery*, 1(3):333–343, 2022.

[29] Cheuk Hin Ho. magmace-exaple. `https://github.com/CheukHinHoJerry/magmace-examples`, 2026.

[30] Cheuk Hin Ho. Magnetic mace branch. `https://github.com/CheukHinHoJerry/mace/tree/2fc7d9b305f012c5b1a59c5b92f2d132888b7d75`, 2026. Commit 2fc7d9b305f012c5b1a59c5b92f2d132888b7d75.

[31] Igor Žutić, Jaroslav Fabian, and S. Das Sarma. Spintronics: Fundamentals and applications. *Rev. Mod. Phys.*, 76:323–410, Apr 2004.

[32] Aaron D Kaplan, Runze Liu, Ji Qi, Tsz Wai Ko, Bowen Deng, Janosh Riebesell, Gerbrand Ceder, Kristin A Persson, and Shyue Ping Ong. A foundational potential energy surface dataset for materials. *arXiv preprint arXiv:2503.04070*, 2025.

[33] Alexey S Kotykhov, Konstantin Gubaev, Max Hodapp, Christian Tantardini, Alexander V Shapeev, and Ivan S Novikov. Constrained DFT-based magnetic machine-learning potentials for magnetic alloys: a case study of Fe–Al. *Scientific Reports*, 13(1):19728, 2023.

[34] 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.

[35] Matthew C Kuner, Aaron D Kaplan, Kristin A Persson, Mark Asta, and Daryl C Chrzan. MP-ALOE: an r2SCAN dataset for universal machine learning interatomic potentials. *npj Computational Materials*, 11(1):352, 2025.

[36] Ludovic Le Laurent, Cyrille Barreteau, and Troel Markussen. Magnetocrystalline anisotropy of Fe, Co, and Ni slabs from density functional theory and tight-binding models. *Physical Review B*, 100(17):174426, 2019.

[37] A Il Liechtenstein, MI Katsnelson, VP Antropov, and VA Gubanov. Local spin density functional approach to the theory of exchange interactions in ferromagnetic metals and alloys. *Journal of Magnetism and Magnetic Materials*, 67(1):65–74, 1987.

[38] AI Liechtenstein, MI Katsnelson, VP Antropov, and VA Gubanov. LSDF-approach to the theory of exchange interactions in magnetic metals. *Journal of magnetism and magnetic materials*, 54:965–966, 1986.

[39] AI Liechtenstein, MI Katsnelson, and VA Gubanov. Exchange interactions and spin-wave stiffness in ferromagnetic metals. *Journal of Physics F: Metal Physics*, 14(7):L125, 1984.

[40] Arslan Mazitov, Filippo Bigi, Matthias Kellner, Paolo Pegolo, Davide Tisi, Guillaume Fraux, Sergey Pozdnyakov, Philip Loche, and Michele Ceriotti. PET-MAD as a lightweight universal interatomic potential for advanced materials modeling. *Nature Communications*, 16(1):10653, 2025.

[41] Mengjuan Mi, Han Xiao, Lixuan Yu, Yingxu Zhang, Yuanshuo Wang, Qiang Cao, and Yilin Wang. Two-dimensional magnetic materials for spintronic devices. *Materials Today Nano*, 24:100408, 2023.

[42] Yoshio Miura and Jun Okabayashi. Understanding magnetocrystalline anisotropy based on orbital and quadrupole moments. *Journal of Physics: Condensed Matter*, 34(47):473001, 2022.

[43] Jens Jørgen Mortensen, Ask Hjorth Larsen, Mikael Kuisma, Aleksei V Ivanov, Alireza Taghizadeh, Andrew Peterson, Anubhab Haldar, Asmus Ougaard Dohn, Christian Schäfer, Elvar Örn Jónsson, et al. GPAW: An open python package for electronic structure calculations. *The Journal of Chemical Physics*, 160(9), 2024.

[44] Svetoslav Nikolov, Mitchell A. Wood, Attila Cangi, Jean-Bernard Maillet, Mihai-Cosmin Marinica, Aidan P. Thompson, Michael P. Desjarlais, and Julien Tranchida. Data-driven magneto-elastic predictions with scalable classical spin-lattice dynamics. *npj Computational Materials*, 7:153, 2021.

[45] `https://vampire.york.ac.uk/`.

[46] Ivan Novikov, Blazej Grabowski, Fritz Körmann, and Alexander Shapeev. Magnetic moment tensor potentials for collinear spin-polarized materials reproduce different magnetic states of bcc Fe. *npj Computational Materials*, 8(1):13, 2022.

[47] Shyue Ping Ong, William Davidson Richards, Anubhav Jain, Geoffroy Hautier, Michael Kocher, Shreyas Cholia, Dan Gunter, Vincent L. Chevrier, Kristin A. Persson, and Gerbrand Ceder. Python materials genomics (pymatgen): A robust, open-source python library for materials analysis. *Computational Materials Science*, 68:314–319, 2013.

[48] M. Pajda, J. Kudrnovský, I. Turek, V. Drchal, and P. Bruno. Ab initio calculations of exchange interactions, spin-wave stiffness constants, and Curie temperatures of Fe, Co, and Ni. *Physical Review B*, 64(17):174402, 2001.

[49] Claudiu C Pavel, Christian Thiel, Stefanie Degreif, Darina Blagoeva, Matthias Buchert, Doris Schüler, and Evangelos Tzimas. Role of substitution in mitigating the supply pressure of rare earths in electric road transport applications. *Sustainable materials and technologies*, 12:62–72, 2017.

[50] John P Perdew, Kieron Burke, and Matthias Ernzerhof. Generalized gradient approximation made simple. *Physical review letters*, 77(18):3865, 1996.

[51] Jan Priessnitz and Dominik Legut. OstravaJ: a tool for calculating magnetic exchange interactions via DFT. *arXiv preprint arXiv:2501.08251*, 2025.

[52] Matteo Rinaldi. *Modelling magnetism from the electronic structure to continuum for iron and its alloys*. PhD thesis, Dissertation, Bochum, Ruhr-Universität Bochum, 2022, 2023.

[53] 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.

[54] Vıctor Garcia Satorras, Emiel Hoogeboom, and Max Welling. E (n) equivariant graph neural networks. In *International conference on machine learning*, pages 9323–9332. PMLR, 2021.

[55] Viktor Sverdlov. Editorial for the topic on magnetic materials and devices, 2025.

[56] AL Udovskii and DA Vasil'ev. Application of physical-empirical models to calculate a fragment of the phase diagram and the physical properties of bcc Fe–Cr alloys: I. formulation of a model and the estimation and approximation of experimental data. *Russian Metallurgy (Metally)*, 2015(3):237–243, 2015.

[57] Tiantian Xu, Jiangfan Yu, Xiaohui Yan, Hongsoo Choi, and Li Zhang. Magnetic actuation based motion control for microrobots: An overview. *Micromachines*, 6(9):1346–1364, 2015.

[58] Wenbin Xu, Rohan Yuri Sanspeur, Adeesh Kolluru, Bowen Deng, Peter Harrington, Steven Farrell, Karsten Reuter, and John R Kitchin. Spin-informed universal graph neural networks for simulating magnetic ordering. *Proceedings of the National Academy of Sciences*, 122(27):e2422973122, 2025.

[59] Teng Yang, Zefeng Cai, Zhengtao Huang, Wenlong Tang, Ruosong Shi, Andy Godfrey, Hanxing Liu, Yuanhua Lin, Ce-Wen Nan, Meng Ye, et al. Deep learning illuminates spin and lattice interaction in magnetic materials. *Physical Review B*, 110(6):064427, 2024.

[60] Hongyu Yu, Boyu Liu, Yang Zhong, Liangliang Hong, Junyi Ji, Changsong Xu, Xingao Gong, and Hongjun Xiang. General time-reversal equivariant neural network potential for magnetic materials. *Physical Review B*, 110:104427, 2024.

[61] Hongyu Yu, Yang Zhong, Liangliang Hong, Changsong Xu, Wei Ren, Xingao Gong, and Hongjun Xiang. Spin-dependent graph neural network potential for magnetic materials. *Physical Review B*, 109(14):144426, 2024.

[62] 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*, 3:8, 2024.

[63] Daye Zheng, Xingliang Peng, Yike Huang, Yinan Wang, Duo Zhang, Zhengtao Huang, Linfeng Zhang, Mohan Chen, Ben Xu, and Weiqing Zhou. Integrating deep-learning-based magnetic model and non-collinear spin-constrained method: Methodology, implementation and application. *arXiv preprint arXiv:2501.14382*, 2025.

[64] Belén E Zuniga-Cespedes, Kaustuv Manna, Hilary ML Noad, Po-Ya Yang, Michael Nicklas, Claudia Felser, Andrew P Mackenzie, and Clifford W Hicks. Observation of an anomalous hall effect in single-crystal Mn3Pt. *New Journal of Physics*, 25(2):023029, 2023.