



ORIGINAL ARTICLE

Open Access



# Equivariant neural network force fields for magnetic materials

Zilong Yuan<sup>1†</sup>, Zhiming Xu<sup>1†</sup>, He Li<sup>1,2†</sup>, Xinle Cheng<sup>1</sup>, Honggeng Tao<sup>1</sup>, Zechen Tang<sup>1</sup>, Zhiyuan Zhou<sup>1,3</sup>, Wenhui Duan<sup>1,2,4\*</sup> and Yong Xu<sup>1,4,5\*</sup>

## Abstract

Neural network force fields have significantly advanced ab initio atomistic simulations across diverse fields. However, their application in the realm of magnetic materials is still in its early stage due to challenges posed by the subtle magnetic energy landscape and the difficulty of obtaining training data. Here we introduce a data-efficient neural network architecture to represent density functional theory total energy, atomic forces, and magnetic forces as functions of atomic and magnetic structures. Our approach incorporates the principle of equivariance under the three-dimensional Euclidean group into the neural network model. Through systematic experiments on various systems, including monolayer magnets, curved nanotube magnets, and moiré-twisted bilayer magnets of CrI<sub>3</sub>, we showcase the method's high efficiency and accuracy, as well as exceptional generalization ability. The work creates opportunities for exploring magnetic phenomena in large-scale materials systems.

**Keywords:** Neural Networks, Force Fields, Magnetic Materials, Density Functional Theory, Deep Learning

## 1 Introduction

Ab initio calculations employing density functional theory (DFT) have become an indispensable tool in material discovery, but their practical applications are limited to small-size systems due to the high computational cost. Deep learning has been proposed as a viable solution to address the trade-off between efficiency and accuracy. Over the past decade, deep learning ab initio methods have revolutionized electronic and atomistic modeling [1–13]. For electronic modeling, a series of deep neural networks are developed to learn the relationship between DFT Hamiltonians and materials structures [8, 9, 13]. By satisfying the principle of equivariance, the neural-network approach has demonstrated exceptional accuracy in example studies of various non-magnetic systems [8–10, 13]. Remarkably, an extended deep-learning method has been developed

to learn the mapping from atomic structures and magnetic structures to DFT Hamiltonians, which preserves a generalized principle of equivariance for magnetic materials [11]. For atomistic modeling, neural network force fields (NNFFs) have been devised for non-magnetic materials and widely applied in molecular dynamics and Monte Carlo simulations [14–25]. The corresponding research on magnetic materials is of equal importance; however, it remains largely unexplored.

The development of NNFFs for magnetic materials faces a few challenges. Firstly, magnetic NNFFs double the degrees of freedoms ( $6N$ ) of conventional NNFFs ( $3N$ ), which requires substantial amount of extra data for machine learning. At the same time, the training data of magnetic materials are significantly more costly than conventional DFT datasets due to the additional computational workload for constraining magnetic configurations, exacerbating the situation of data scarcity. Yang et al. [26] recently proposed a neural network method based on an existing descriptor based model, but it suffers from data inefficiency problem for only including invariant features, which has been explicitly illustrated in

\*Correspondence: [duanw@tsinghua.edu.cn](mailto:duanw@tsinghua.edu.cn); [yongxu@mail.tsinghua.edu.cn](mailto:yongxu@mail.tsinghua.edu.cn)

<sup>†</sup>State Key Laboratory of Low Dimensional Quantum Physics and Department of Physics, Tsinghua University, Beijing, 100084, China

Full list of author information is available at the end of the article <sup>†</sup>Equal contributors

previous works [22, 24]. Alternatively, incorporating a prior knowledge of symmetry into neural network design can alleviate this problem. Equivariant neural networks (ENNs) [27–29] are aimed to satisfy the equivariance requirements by ensuring that the input, output and all internal features transform equivalently under the symmetry group. Therefore, ENNs can be extended to scenarios with limited data without necessitating data augmentation, rendering them more viable for magnetic material energy modeling task.

Secondly, the derivatives of total energy with respect to varying orientations of magnetic moments (called magnetic forces in this work) serve a role in analogy to atomic forces in conventional NNFFs, which are indispensable to the atomistic modeling of magnetic materials. Yu et al. [30] recently proposed a time-reversal equivariant neural network to map the energy of magnetic materials, where the training data of magnetic forces were not used and explicitly learned. For magnetic NNFFs, the absence of magnetic forces will seriously increase the demand for additional data to fit the energy profile, which is ignored in previous studies [31–37]. Furthermore, those models could not comprehensively explore the magnetic effects involving higher-order derivatives to magnetic moments accurately, such as the low-energy elementary excitations [38].

To address the above challenges, we propose MagNet, an equivariant deep-learning framework to represent DFT total energy  $E(\{\mathcal{R}\}, \{\mathcal{M}\})$  and its derivative forces as functions of atomic structures  $\{\mathcal{R}\}$  and complex non-collinear local atomic magnetic moments  $\{\mathcal{M}\}$ . As a critical innovation, we design an ENN architecture naturally integrating both atomic and magnetic degrees of freedom and incorporate with direct mapping of magnetic forces, enabling efficient and accurate learning of magnetic materials. The method is systematically tested to show high reliability in calculating the magnon dispersion and good generalization capabilities by example studies of magnetic CrI<sub>3</sub> nanotubes. Finally we implement our method for studying spin dynamics of moiré-twisted bilayer CrI<sub>3</sub>. Benefiting from the high efficiency and accuracy, there could be further promising applications of MagNet in magnetic materials computation at large length/time scales.

## 2 Methods

Deep learning methods have enabled efficient material simulations with DFT accuracy. A significant generalization of the methods is required for studying magnetic materials. For nonmagnetic systems, the total energy  $E$  as a function of atomic structure  $\{\mathcal{R}\}$  is calculated by self-consistent field (SCF) iterations in DFT. The function  $E(\{\mathcal{R}\})$  is the learning target of NNFFs. In contrast, for magnetic systems, the total energy depends not only on atomic structure  $\{\mathcal{R}\}$  but also on magnetic structure  $\{\mathcal{M}\}$ . To compute the total energy for varying  $\{\mathcal{M}\}$ , one needs

to apply constrained DFT that utilizes the Lagrangian approach to constrain specific magnetic configuration and introduces constraining fields as an additional potential in the Kohn–Sham equation [39], which significantly increases the computational workload and is much more time-consuming.

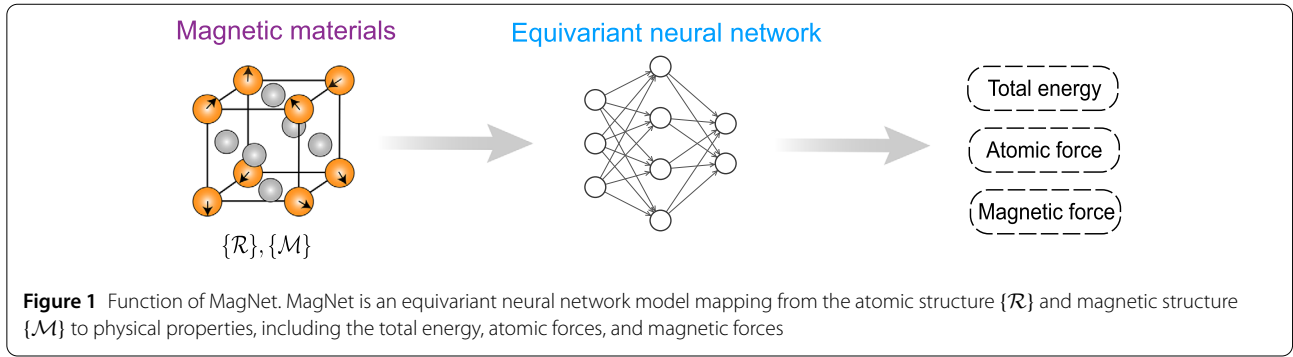
The function of MagNet is illustrated in Fig. 1. First, magnetic materials with different atomic and magnetic configurations are calculated by constrained DFT for preparing training datasets. Then the training datasets are fed into MagNet for predicting physical properties, including DFT total energy, atomic forces, and magnetic forces for atomic and magnetic structures unseen in the training datasets. By substituting the costly SCF calculation with neural networks, the method significantly lowers the computational overhead and enables the efficient and accurate mapping between properties and structures of magnetic materials. The critical point here is empowering neural networks by leveraging a priori knowledge, which will be discussed subsequently.

Noticeably, for most magnetic materials, varying  $\{\mathcal{R}\}$  will alter the strength of interatomic bonding energies in the total energy, whereas varying  $\{\mathcal{M}\}$  mainly modifies the relatively weak and localized magnetic exchange interactions, leading to minor changes in the total energy. Consequently, the effects on the total energy due to alterations in  $\{\mathcal{M}\}$  are expected to be weaker in magnitude and shorter in length scale. The subtle interactions require an appropriate design of neural networks, distinct from the description on changes induced by  $\{\mathcal{R}\}$ .

The equivariance is another essential point to consider in network design. For atomistic systems, the physical properties of materials are equivariant under the action of rotation, inversion, and translation — which comprise the three-dimensional Euclidean group  $E(3)$ . Scalar quantities like the total energy are invariant under these symmetry group operations, whereas vector quantities such as atomic forces and magnetic forces are equivariant and will change when the atomic geometry is transformed. Thus it is natural to incorporate the equivariance into the design of neural networks. Given information about one structure, the target property of all the symmetry-related structures can be obtained from neural networks via equivariant transformations, which enables a more efficient mapping in data limited cases.

Here we present an ENN architecture of MagNet. The equivariant building blocks of the neural network model are implemented following the scheme proposed by DeepH-E3 [9]. Formally, a function  $f$  relating the input vector space  $X$  and the output vector space  $Y$  is regarded equivariant, provided that for any input  $x \in X$ , output  $y \in Y$ , and any group element  $g$  within a transformation group  $G$ , the following condition is satisfied:

$$f(D_X(g)x) = D_Y(g)f(x), \quad (1)$$



where  $D_X(g)$  and  $D_Y(g)$  indicate transformation matrices in  $X$  and  $Y$ , parameterized by  $g$ . In MagNet, translation symmetry is guaranteed by operating on relative positions of atoms. For rotation, features  $v_m^l$  carry the irreducible representation of the  $SO(3)$  group of dimension  $2l + 1$ , where  $l$  represents the angular momentum quantum number, and  $m$  denotes the magnetic quantum number varying between  $-l$  and  $l$ . A key operation for interacting different  $l$  features is the tensor product, denoted as  $\otimes$ , which uses Clebsch–Gordan coefficients  $C_{l_1, m_1, l_2, m_2}^{l_3, m_3}$  to combine features  $x^{l_1}$  and  $y^{l_2}$  and produces output feature  $z^{l_3}$ :

$$\begin{aligned} z_{m_3}^{l_3} &= x_{m_1}^{l_1} \otimes y_{m_2}^{l_2} \\ &= \sum_{m_1, m_2} C_{l_1, m_1, l_2, m_2}^{l_3, m_3} x_{m_1}^{l_1} y_{m_2}^{l_2}. \end{aligned} \quad (2)$$

Since the features are equivariant under rotation, the physical quantities represented by the features will also change equivariantly under rotation. For spatial inversion, it is necessary to introduce an additional parity index into the features, which labels the spatial inversion either even ( $p = 1$ ) or odd ( $p = -1$ ). Parity equivariance is ensured by permitting contributions to an output feature with parity  $p_3$  from two features possessing parities  $p_1$  and  $p_2$  in the tensor product if the selection rule is satisfied:  $p_3 = p_1 p_2$ . In addition, for time-reversal symmetry, it could be treated essentially in the same way as parity and integrated into ENNs, as implemented in the previous work [11].

In the context of the building blocks illustrated in Fig. 2, the operation ‘E3Linear’ is formulated as:

$$\text{E3Linear}(v_{cm}^l) = \sum_{c'} W_{cc'}^l v_{c'm}^l + b_c^l, \quad (3)$$

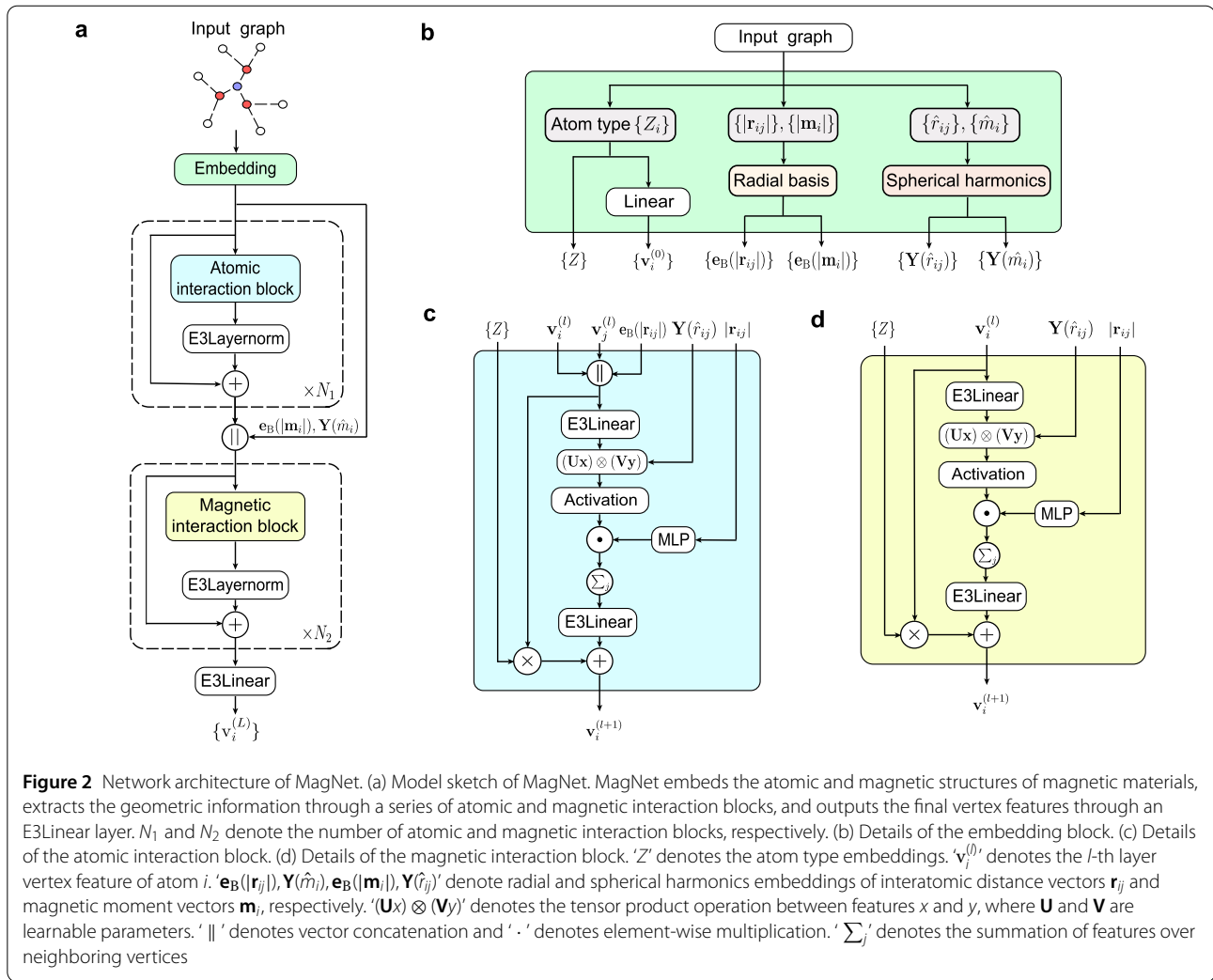
where  $c$  and  $c'$  denote the channel indices. The terms  $W_{cc'}^l$  and  $b_c^l$  are the learnable weights and biases, respectively. It is essential to note that the biases  $b_c^l$  are nonzero only for equivariant features  $v$  with  $l = 0$ , ensuring the preservation of equivariance requirements. ‘Activation’ introduces a non-linear activation function, or a scalar gate, on the

features depending on the index  $l$ . For features with  $l > 0$ , a linear scaling is employed. For features with  $l = 0$ , a non-linear SiLU function is employed. The normalization of the features while preserving equivariance is achieved by using the E3LayerNorm proposed in Ref. [9]:

$$\text{E3LayerNorm}(v_{cm}^l) = g_c^l \cdot \frac{v_{cm}^l - \mu_m^l}{\sigma_m^l + \epsilon} + h_c^l, \quad (4)$$

where  $\mu_m^l$  and  $\sigma_m^l$  are the mean and the standard deviation of features, respectively,  $g_c^l$  and  $h_c^l$  are learnable parameters, and  $\epsilon$  is a small constant introduced for enhancing numerical stability. The term  $h_c^l$  is subject to the same equivariance requirements as  $b_c^l$  in Eq. (3).

The network architecture of MagNet, as shown in Fig. 2, is built on an embedding block, followed by a series of atomic interaction blocks and magnetic interaction blocks, and output final vertex features after an E3Linear layer. For the embedding block, the purpose is to initialize equivariant features from the information of magnetic materials, including the interatomic distance vector  $\mathbf{r}_{ij}$ , the magnetic moment vector  $\mathbf{m}_i$ , and the atom species  $Z_i$ . For non-magnetic atom  $i$ , the magnetic moment  $\mathbf{m}_i$  is set to zero. The radial functions expand interatomic distances and magnetic moment length in the form of Gaussian basis [14]. The directions of  $\mathbf{r}_{ij}$  and  $\mathbf{m}_i$  are incorporated into the real spherical harmonics  $Y_m^l$  with indices  $l$  and  $m$ . Atomic interaction blocks encode interactions between neighboring atoms where different features are mixed and contracted through the tensor product. Gaussian functions and a polynomial envelope function [16] are implemented in multi-layer perceptrons (MLP) as the radial weights for coupled tensor production interactions. Then the vertex features are carried magnetic moment information to interact with other atomic features in a following series of magnetic interaction blocks. Since the influence of magnetic moment is relatively more localized, we set smaller number of layers for magnetic interaction blocks than for atomic interaction blocks. Final vertex features are obtained as the output of the E3Linear layer. The total energy is derived from the sum of final vertex features with a



rescaling as shown in Eq. (5). Atomic forces and magnetic forces are subsequently determined as the negative gradient of atomic positions and magnetic moments to the predicted total energy.

$$\hat{E} = \sum_i (\sigma_0 v_i + \mu_0), \quad (5)$$

$$\hat{F}_{i,\alpha} = -\frac{\partial \hat{E}}{\partial r_{i,\alpha}}, \quad (6)$$

$$\hat{F}_{\text{mag},i,\alpha} = -\frac{\partial \hat{E}}{\partial m_{i,\alpha}}, \quad (7)$$

where  $\sigma_0$  and  $N\mu_0$  are the standard deviation and the mean over the training set, respectively.  $N$  is the number of atoms,  $i$  is the atom index number and  $\alpha$  is the coordinate index. MagNet is trained using a loss function based on a weighted sum of total energy, atomic forces, and magnetic

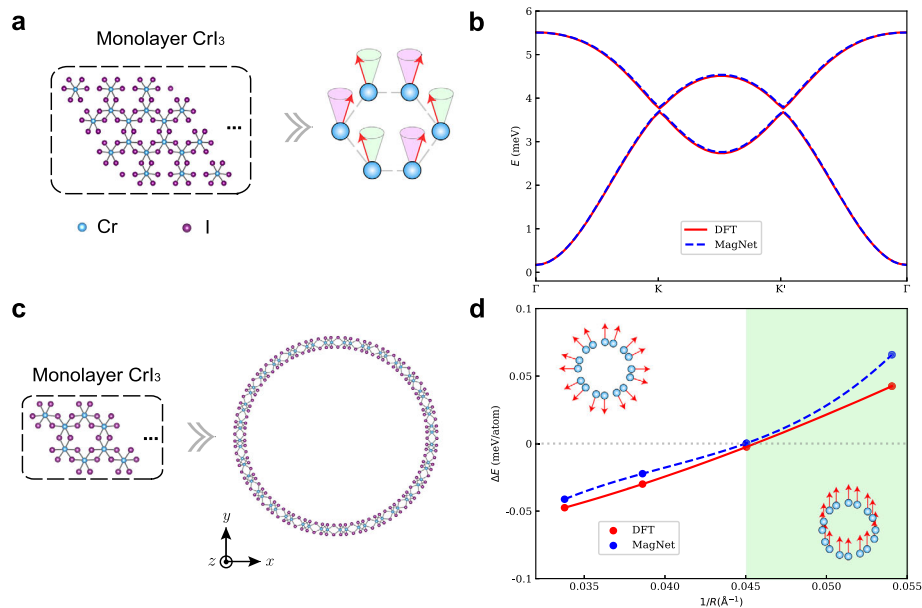
forces mean-squared error loss terms:

$$L = \lambda_E \|\hat{E} - E\|^2 + \frac{\lambda_F}{3N} \sum_{i=1}^N \sum_{\alpha=1}^3 \left\| -\frac{\partial \hat{E}}{\partial r_{i,\alpha}} - F_{i,\alpha} \right\|^2 + \frac{\lambda_{F_{\text{mag}}}}{3N_{\text{mag}}} \sum_{j=1}^{N_{\text{mag}}} \sum_{\beta=1}^3 \left\| -\frac{\partial \hat{E}}{\partial m_{j,\beta}} - F_{\text{mag},j,\beta} \right\|^2, \quad (8)$$

where  $\lambda_E$ ,  $\lambda_F$ , and  $\lambda_{F_{\text{mag}}}$  denote the weights of total energy, atomic forces, and magnetic forces, respectively.  $N$ ,  $N_{\text{mag}}$  are number of atoms and number of magnetic atoms, respectively.  $\alpha$ ,  $\beta$  are the coordinate indices.

### 3 Results and discussions

The capability of MagNet is tested by a series of example studies on the magnetic material  $\text{CrI}_3$ . Our results demonstrate that MagNet can well reproduce DFT results. Remarkably, once trained by DFT data on small structures



**Figure 3** Example applications of MagNet. (a) Schematic diagrams showing monolayer CrI<sub>3</sub> and its spin waves. MagNet is trained by DFT calculation results of monolayer CrI<sub>3</sub> and used to study spin waves. (b) Magnon dispersion of monolayer CrI<sub>3</sub> predicted by MagNet via neural networks automatic differentiation and further checked by DFT finite-difference calculations. (c) Generalization ability of MagNet. MagNet learns from DFT results of monolayer CrI<sub>3</sub> and generalizes to study CrI<sub>3</sub> nanotubes. (d) Energy difference ( $\Delta E$ ) between the two possible magnetic configurations displayed in the insets as a function of nanotube curvature  $1/R$

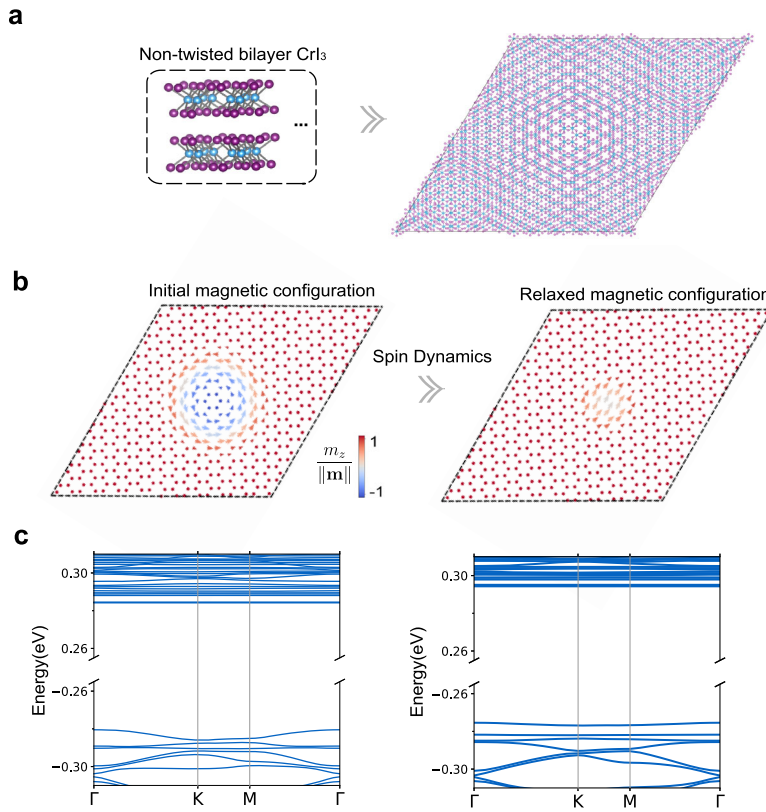
with random magnetic orientations, MagNet can accurately predict on new magnetic configurations unseen in the training datasets, especially the large-scale magnetic structures. To generate the dataset and benchmark results, we calculated DFT total energy, atomic forces, and magnetic forces for given magnetic configurations using constrained DFT as implemented in the DeltaSpin package [40], where the Kohn–Sham eigenstates and the constraining fields are updated alternately to obtain the target magnetic moments. Atomic and magnetic forces are obtained via the Hellmann–Feynman theorem [41].

Magnons as elementary excitations of spin waves in magnetic materials are regarded as prospective information carriers, which facilitates the realization of diverse spin-wave-based logic gates [42–44] for potential computing applications [45]. As an example study, we predicted the magnon dispersion of a magnetic material through MagNet using the neural-network automatic differentiation. We prepared DFT datasets by calculating supercells of monolayer CrI<sub>3</sub> with the equilibrium lattice structure and randomly perturbed magnetic moment orientations, up to 10° away from the ground state ferromagnetic configuration [Fig. 3(a)]. The neural network model of MagNet was trained by the DFT data and then used to predict the magnon dispersion. To verify the results of neural-network automatic differentiation, the finite difference method by DFT was performed to compute the derivative of magnetic

forces:  $f'(x) = [f(x + \Delta) - f(x - \Delta)]/2\Delta$ , where the step size  $\Delta$  refers to the change of magnetic-moment orientation and  $\Delta = 5^\circ$  was chosen. Details of deriving the magnon dispersion are described in Appendix C. The calculation results of DFT and MagNet are shown and compared in Fig. 3(b). MagNet achieves a mean-absolute error (MAE) of  $1.67 \times 10^{-2}$  meV/ $\mu_B$  for magnetic forces on the validation dataset, and the predicted magnon dispersion agrees well with the DFT reference, indicating the good reliability and high accuracy of MagNet.

Strain gradients can significantly affect the magnetism in curved structures [46, 47]. CrI<sub>3</sub> nanotubes have attracted considerable interest for the study of curved magnetism [48]. The first-principles calculations, however, are limited by the large-size structures and diverse magnetic configurations. Modeling  $E(\{\mathcal{R}\}, \{\mathcal{M}\})$  is a challenging task for neural networks when both  $\{\mathcal{R}\}$  and  $\{\mathcal{M}\}$  vary simultaneously. We prepared DFT datasets by calculating flat sheets of monolayer CrI<sub>3</sub> featuring randomly perturbed atomic and magnetic configurations, and applied the trained neural-network model of MagNet to investigate the CrI<sub>3</sub> nanotubes [Fig. 3(c)]. Furthermore, the energies of the two possible magnetic configurations are considered. One is a non-collinear magnet, with the magnetic moments aligned along the radial direction, and the other is a ferromagnet, as displayed in the insets of Fig. 3(d). The (10, 10), (12, 12), (14, 14), (16, 16) nanotubes





**Figure 4** Example applications of MagNet in studying moiré-twisted materials. (a) Generalization ability of MagNet. MagNet learns from DFT results of non-twisted bilayer CrI<sub>3</sub> and generalizes to study moiré-twisted bilayer CrI<sub>3</sub> with varying twist angles. (b) Initial magnetic configurations (adapted from Ref. [11]) and relaxed magnetic configurations of moiré-twisted bilayer CrI<sub>3</sub> with a twist angle of 63.48° (4,336 atoms per supercell) as predicted by MagNet. The magnetic moments of the top CrI<sub>3</sub> layer are represented by colored arrows, with the in-plane components denoted by the arrow length and the out-of-plane components denoted by the color. On the magnetic moments of the bottom CrI<sub>3</sub> layer, the out-of-plane components are the same as the top layer, whereas the in-plane components are opposite. (c) Band structure of the moiré-twisted bilayer CrI<sub>3</sub> (displayed in (b)) predicted by the xDeepH method [11]

of CrI<sub>3</sub> are used to investigate the size effects. The MAE of total energy predicted by MagNet reaches as low as 0.129 meV/atom. The energy differences between the two magnetic configurations as a function of nanotube curvature are predicted by DFT and MagNet. As shown in Fig. 3(d), the crossover from ferromagnet to non-collinear magnet as increasing the nanotube radius are well captured by MagNet, as checked by the DFT benchmark data. The good generalization ability of MagNet is thus demonstrated.

Finally, we tried a more challenging study on twisted bilayer CrI<sub>3</sub>, which has been reported to exhibit abundant non-collinear magnetic textures both theoretically and experimentally [49–52]. The Landau–Lifshitz–Gilbert equation [53] was applied to update magnetic moment configurations according to the predicted magnetic forces:

$$\frac{dm_i}{dt} = \gamma m_i \times \frac{\partial E}{\partial m_i} + \gamma \alpha m_i \times \left( m_i \times \frac{\partial E}{\partial m_i} \right), \quad (9)$$

where  $\gamma$  is the electron gyromagnetic ratio and  $\alpha$  is a phenomenological damping parameter. More specifically, the new magnetic moment orientations could be efficiently updated with the dissipative term proposed in Ref. [54]:

$$\hat{m}_i' = \hat{m}_i + \lambda \hat{m}_i \times \left( \hat{m}_i \times \frac{\partial E}{\partial \hat{m}_i} \right), \quad (10)$$

where  $\lambda$  represents the step size, and the magnitude of the magnetic moment is normalized after each update step.

As shown in Fig. 4(a), the non-twisted bilayer CrI<sub>3</sub> datasets were used to train MagNet. Our simulations of twisted bilayer CrI<sub>3</sub> were carried out on a supercell comprising 4,326 atoms with a twist angle of  $\theta = 63.48^\circ$ , which was predicted to host non-collinear magnetic configurations [49, 50]. Using the skyrmion state [Fig. 4(b)] predicted by Ref. [49] as the initial magnetic configuration, we performed the spin dynamics simulation according to Eq. (10) with the magnetic forces predicted by MagNet.

Converged within a few hundred steps, the skyrmion state transits to a more stable magnetic configuration, in which the out-of-plane components are positive and the in-plane components are in opposite directions between the top and bottom layers [Fig. 4(b)]. Furthermore, we applied the extended deep-learning DFT Hamiltonian method (named xDeepH) [11] to predict the electronic structure of the relaxed magnetic configuration. As shown in Fig. 4(c), the valence bands near the Fermi level become flatter after performing the spin dynamics. The isolated flat bands could be useful for exploring the correlated electronic and magnetic physics. This work demonstrates that the magnetic and electronic structures of magnetic superstructures can be predicted by deep learning methods.

#### 4 Conclusions

In brief summary, we proposed a general neural-network framework of MagNet to represent DFT total energy, atomic and magnetic forces as functions of atomic and magnetic structures by deep neural networks. MagNet incorporates the  $E(3)$  group symmetry, which significantly reduces the training complexity and the amount of training data required. High accuracy and exceptional generalization ability of the method are demonstrated by investigating various kinds of magnets formed by  $\text{CrI}_3$ . This approach creates opportunities for exploring novel magnetism and spin dynamics in magnetic structures at large length/time scales.

#### Appendix A: Neural-network methods

The MagNet model is implemented with PyTorch, PyTorch Geometric and e3nn [55] libraries. The initial vertex features were set to be  $32 \times 0e$ . For the interaction blocks, equivariant vertex features are set to be  $32 \times 0e + 16 \times 1o$ , where  $32 \times 0e$  denotes 32 equivariant vectors carrying an  $l = 0$  representation with even parity, and  $16 \times 1o$  denotes 16 equivariant vectors carrying an  $l = 1$  representation with odd parity. For the magnetic interaction blocks, equivariant vertex features were set to be  $32 \times 0e + 16 \times 1o$ . The number of atomic interaction blocks  $N_1$  and magnetic interaction blocks  $N_2$  are set to be 4 and 2, respectively. Spherical harmonics has a maximal angular momentum of  $l = 2$ . The learning rate is selected initially to be 0.001 and decreases by a factor of 0.5 when the loss does not decrease after 20 epochs. The cutoff of the radius on three datasets are set as 6 Å. All the network results are reported on validation sets. For the study of the magnon dispersion of  $\text{CrI}_3$ , we use  $\lambda_E = 10$ ,  $\lambda_F = 10^5$  and  $\lambda_{F_{\text{mag}}} = 10^8$ , getting mean-absolute errors (MAE) of  $5.57 \times 10^{-7}$  eV/atom,  $1.26 \times 10^{-4}$  eV/Å, and  $1.67 \times 10^{-5}$  eV/ $\mu_B$  for DFT total energy, atomic forces and magnetic forces, respectively. For the study of  $\text{CrI}_3$  nanotubes, we use  $\lambda_E = 10$ ,  $\lambda_F = 10^5$  and  $\lambda_{F_{\text{mag}}} = 10^6$ , getting MAEs of  $1.29 \times 10^{-4}$  eV/atom,  $3.77 \times$

$10^{-3}$  eV/Å, and  $1.30 \times 10^{-3}$  eV/ $\mu_B$  for total energy, atomic forces and magnetic forces, respectively. For the study of bilayer  $\text{CrI}_3$ , we use  $\lambda_E = 10$ ,  $\lambda_F = 10^5$  and  $\lambda_{F_{\text{mag}}} = 10^6$ , getting MAEs of  $1.32 \times 10^{-3}$  eV/atom,  $9.35 \times 10^{-3}$  eV/Å, and  $3.99 \times 10^{-3}$  eV/ $\mu_B$  for DFT total energy, atomic forces and magnetic forces, respectively. Neural-network training is performed on an NVIDIA RTX 3090 GPU with a batch size of 1. For spin dynamics simulations, the step size  $\lambda$  is set as 20.

#### Appendix B: DFT calculation methods

DFT calculations are performed by the Vienna ab initio simulation package [56] using the Perdew–Berke–Ernzerhof-type exchange-correlation functional. Constrained DFT calculations as implemented by the Deltaspin method [40] are applied to study systems with specified magnetic configurations. In all calculations, we constrain the magnetic orientation as well as the magnitude of magnetic moments. The spin–orbit coupling is included in the DFT calculations. The DFT +  $U$  method with a Hubbard correction of  $U = 3.0$  eV is applied to describe the  $3d$  orbitals of Cr.

The dataset of  $3 \times 3$  supercells of monolayer  $\text{CrI}_3$  contains 200 configurations with random perturbations (up to  $10^\circ$ ) away from the ground state ferromagnetic configuration in the equilibrium atomic structure. The energy cut off of plane wave basis is set to be 400 eV. A Monkhorst–Pack  $k$ -mesh of  $3 \times 3 \times 1$  is used.

For the dataset of  $2 \times 2$  supercells of monolayer  $\text{CrI}_3$ , 80 different atomic structures are prepared by introducing random atomic displacements (up to 0.1 Å) to each atom about their equilibrium positions. For each atomic structure, 5 random magnetic configurations are generated by arbitrarily arranging the orientation of the constrained magnetic moment for each magnetic atom Cr, giving 400 structures in total. The energy cut off of plane wave basis is set to be 400 eV. A Monkhorst–Pack  $k$ -mesh of  $4 \times 4 \times 1$  is used. For  $\text{CrI}_3$  nanotubes, a Monkhorst–Pack  $k$ -mesh of  $1 \times 1 \times 5$  is used.

For the dataset of supercells of bilayer  $\text{CrI}_3$ , the second layer is shifted with respect to the first layer. The shift is sampled by an  $8 \times 8$  grid of the supercell, yielding 64 atomic configurations. In addition, random displacements (up to 0.1 Å) are introduced to each atom about their equilibrium positions. For each atomic configuration, 5 random magnetic configurations, ferromagnetic and antiferromagnetic configurations with a random perturbation (up to  $30^\circ$ ) away from the  $z$ -axis configurations are generated, giving 448 structures. The energy cut off of plane wave basis is set to be 300 eV. A Monkhorst–Pack  $k$ -mesh of  $2 \times 2 \times 1$  is used.

All the datasets are divided into training and validation sets randomly by a ratio of 8: 2.

### Appendix C: Magnon dispersion computation

The energy of a magnetic material near a minimum can be described by a quadratic form:

$$E = \frac{1}{2} \sum_{ab} \mathbf{S}_a^\top \mathbf{J}_{ab} \mathbf{S}_b, \quad (\text{C.1})$$

where  $\mathbf{S}_a$  and  $\mathbf{S}_b$  denote the spin vector in global coordinates,  $a$  and  $b$  denote the index of spin site respectively, and  $\mathbf{J}_{ab}$  denotes the exchange coupling matrix. We choose a local-coordinate system on each magnetic atom, with the  $z$  axis parallel to the ground-state magnetic orientation. The transformation between local and global coordinates is a three by three orthogonal matrix  $\mathbf{R}_a$ :

$$\mathbf{S}_a = \mathbf{R}_a \mathbf{S}'_a, \quad (\text{C.2})$$

where  $\mathbf{S}'_a$  is the spin vector in local coordinates.

When the deviation from the ground-state magnetization is small,  $\mathbf{S}'_a$  can be written as:

$$\mathbf{S}'_a = S_a (\hat{\mathbf{z}} + \mathbf{P}^\top \chi_a), \quad (\text{C.3})$$

where  $S_a$  is the length of the  $a$ -th spin,

$$\hat{\mathbf{z}} = (0, 0, 1)^\top, \quad \mathbf{P} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \quad (\text{C.4})$$

and  $\chi_a$  is the projection of spin on the local  $x$ - $y$  plane, normalized by the length of spin. We note that  $\chi_a$  is a two-component object.

From the Heisenberg equation of motion, we obtain:

$$\frac{d\chi_a}{dt} = i \sum_b \sigma_y \mathbf{D}_{ab} \chi_b, \quad (\text{C.5})$$

where  $\sigma_y$  is the Pauli- $y$  matrix, and the spin dynamical matrix  $\mathbf{D}_{ab}$  is defined as:

$$\mathbf{D}_{ab} = \frac{1}{\sqrt{S_a S_b}} \left( \frac{\partial^2 E}{\partial \chi_a \partial \chi_b} \right). \quad (\text{C.6})$$

Here, the second derivative of energy  $\frac{\partial^2 E}{\partial \chi_a \partial \chi_b}$  can be computed by either neural network automatic differentiation or DFT finite difference.

In a system with translation symmetry, the dynamical matrix only depends on the difference between  $\mathbf{r}_a$  and  $\mathbf{r}_b$ , therefore we can write:

$$\mathbf{D}_{ab} = \mathbf{D}(\mathbf{r}_a - \mathbf{r}_b) := \mathbf{D}(\mathbf{r}), \quad (\text{C.7})$$

where  $\mathbf{r}_a$  is the position of the  $a$ -th spin.

Finally, the magnon dispersion is obtained by diagonalizing the Fourier transformation of the spin dynamical matrix  $\tilde{\mathbf{D}}(\mathbf{k})$ :

$$\tilde{\mathbf{D}}(\mathbf{k}) = \sum_{\mathbf{r}} e^{-i\mathbf{k} \cdot \mathbf{r}} \mathbf{D}(\mathbf{r}). \quad (\text{C.8})$$

Specifically, for the DFT benchmark, let us consider the CrI<sub>3</sub> unit cell with 2 Cr atoms, marked as atom1 and atom2. For brevity, we label the displacement of Cr atom's magnetic moment as (atom1<sub>*x*</sub>, atom1<sub>*y*</sub>, atom2<sub>*x*</sub>, atom2<sub>*y*</sub>). We need to perform constrained DFT to calculate 8 specific configurations near the ground states, noted as (+5°, 0, 0, 0), (−5°, 0, 0, 0), (0, +5°, 0, 0), (0, −5°, 0, 0), (0, 0, +5°, 0), (0, 0, −5°, 0), (0, 0, 0, +5°), (0, 0, 0, −5°), and obtain the derivatives of magnetic forces from DFT by finite difference method:  $f'(x) = [f(x + \Delta) - f(x - \Delta)]/2\Delta$  in the global coordinate frame. Then we transform them from the global coordinate frame into the local coordinate frame and calculate the spin dynamical matrices. Finally, the magnon dispersion is obtained by performing the Fourier transformations and diagonalizing the spin dynamical matrices.

#### Author contributions

WD and YX proposed and supervised the project. Under the guidance of WD and YX, ZY and ZT prepared the dataset, XC, ZY and ZZ performed the theoretical derivation, ZY and HL performed the code implementation and data analysis. YX, ZY, ZX, and HT contributed to the interpretation of the results. YX, ZY, and ZX prepared the paper with input from the other co-authors. All authors have read and approved the manuscript.

#### Funding

This work was supported by the Basic Science Center Project of NSFC (grant no. 52388201), the Ministry of Science and Technology of China (grant no. 2023YFA1406400), the National Natural Science Foundation of China (grant no. 12334003), the National Science Fund for Distinguished Young Scholars (grant no. 12025405), the Beijing Advanced Innovation Center for Future Chip (ICFC), and the Beijing Advanced Innovation Center for Materials Genome Engineering. The calculations were done on Hefei advanced computing center.

#### Data availability

All the relevant data discussed in the present paper are available from the authors on request.

#### Declarations

##### Ethics approval and consent to participate

Not applicable.

##### Consent for publication

Not applicable.

##### Competing interests

YX is an editorial board member for Quantum Frontiers and was not involved in the editorial review, or the decision to publish, this article. All authors declare that there are no competing interests.

##### Author details

<sup>1</sup>State Key Laboratory of Low Dimensional Quantum Physics and Department of Physics, Tsinghua University, Beijing, 100084, China. <sup>2</sup>Institute for Advanced Study, Tsinghua University, Beijing, 100084, China. <sup>3</sup>School of Physics, Peking University, Beijing, 100871, China. <sup>4</sup>Frontier Science Center for Quantum Information, Beijing, 100084, China. <sup>5</sup>RIKEN Center for Emergent Matter Science (CEMS), Wako, Saitama 351-0198, Japan.



Received: 7 February 2024 Revised: 24 March 2024 Accepted: 7 April 2024  
Published online: 22 April 2024

## References

- Behler J, Parrinello M (2007) Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys Rev Lett* 98:146401
- Bartók AP, Payne MC, Kondor R, Csányi G (2010) Gaussian approximation potentials: the accuracy of quantum mechanics, without the electrons. *Phys Rev Lett* 104(13):136403
- Behler J (2011) Atom-centered symmetry functions for constructing high-dimensional neural network potentials. *J Chem Phys* 134(7):074106
- Bartók AP, Kondor R, Csányi G (2013) On representing chemical environments. *Phys Rev B* 87(18):184115
- Thompson AP, Swiler LP, Trott CR, Foiles SM, Tucker GJ (2015) Spectral neighbor analysis method for automated generation of quantum-accurate interatomic potentials. *J Comput Phys* 285:316–330
- Shapeev AV (2016) Moment tensor potentials: a class of systematically improvable interatomic potentials. *Multiscale Model Simul* 14(3):1153–1173
- Zhang L, Han J, Wang H, Car R, E W (2018) Deep potential molecular dynamics: a scalable model with the accuracy of quantum mechanics. *Phys Rev Lett* 120(14):143001
- Li H, Wang Z, Zou N, Ye M, Xu R, Gong X, Duan W, Xu Y (2022) Deep-learning density functional theory Hamiltonian for efficient ab initio electronic-structure calculation. *Nat Comput Sci* 2(6):367–377
- Gong X, Li H, Zou N, Xu R, Duan W, Xu Y (2023) General framework for E(3)-equivariant neural network representation of density functional theory Hamiltonian. *Nat Commun* 14(1):2848
- Tang Z, Li H, Lin P, Gong X, Jin G, He L, Jiang H, Ren X, Duan W, Xu Y (2023) Efficient hybrid density functional calculation by deep learning. [arXiv:2302.08221](https://arxiv.org/abs/2302.08221)
- Li H, Tang Z, Gong X, Zou N, Duan W, Xu Y (2023) Deep-learning electronic-structure calculation of magnetic superstructures. *Nat Comput Sci* 3(4):321–327
- Li H, Tang Z, Fu J, Dong W-H, Zou N, Gong X, Duan W, Xu Y (2024) Deep-learning density functional perturbation theory. *Phys Rev Lett* 132(9):096401
- Wang Y, Li H, Tang Z, Tao H, Wang Y, Yuan Z, Chen Z, Duan W, Xu Y (2024) DeePh-2: enhancing deep-learning electronic structure via an equivariant local-coordinate transformer. [arXiv:2401.17015](https://arxiv.org/abs/2401.17015)
- Schütt KT, Sauceda HE, Kindermans P-J, Tkatchenko A, Müller K-R (2018) SchNet – a deep learning architecture for molecules and materials. *J Chem Phys* 148(24):241722
- Unke OT, Meuwly M (2019) PhysNet: a neural network for predicting energies, forces, dipole moments, and partial charges. *J Chem Theory Comput* 15(6):3678–3693
- Gasteiger J, Groß J, Günnemann S (2020) Directional message passing for molecular graphs. In: International conference on learning representations
- Wang Z, Wang C, Zhao S, Du S, Xu Y, Gu B-L, Duan W (2021) Symmetry-adapted graph neural networks for constructing molecular dynamics force fields. *Sci China, Ser G, Phys Mech Astron* 64(11):117211
- Schütt K, Unke O, Gastegger M (2021) Equivariant message passing for the prediction of tensorial properties and molecular spectra. In: Proceedings of the 38th international conference on machine learning, vol 139, pp 9377–9388
- Wang Z, Wang C, Zhao S, Xu Y, Hao S, Hsieh CY, Gu B-L, Duan W (2022) Heterogeneous relational message passing networks for molecular dynamics simulations. *NPJ Comput Mater* 8(1):53
- Gasteiger J, Becker F, Günnemann S (2021) Gemnet: universal directional graph neural networks for molecules. In: Advances in neural information processing systems, vol 34, pp 6790–6802
- Brandstetter J, Hesselink R, Pol E, Bekkers EJ, Welling M (2022) Geometric and physical quantities improve E(3) equivariant message passing. In: International conference on learning representations
- Batzner S, Musaelian A, Sun L, Geiger M, Mailoa JP, Kornbluth M, Molinari N, Smidt TE, Kozinsky B (2022) E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nat Commun* 13(1):2453
- Batatia I, Kovacs DP, Simm G, Ortner C, Csányi G (2022) Mace: higher order equivariant message passing neural networks for fast and accurate force fields. In: Advances in neural information processing systems, vol 35, pp 11423–11436
- Musaelian A, Batzner S, Johansson A, Sun L, Owen CJ, Kornbluth M, Kozinsky B (2023) Learning local equivariant representations for large-scale atomistic dynamics. *Nat Commun* 14(1):579
- Wang Y, Wang T, Li S, He X, Li M, Wang Z, Zheng N, Shao B, Liu T-Y (2024) Enhancing geometric representations for molecules with equivariant vector-scalar interactive message passing. *Nat Commun* 15(1):313
- Yang T, Cai Z, Huang Z, Tang W, Shi R, Godfrey A, Liu H, Lin Y, Nan C-W, Ye M, Zhang L, Wang H, Xu B (2023) Deep learning illuminates spin and lattice interaction in magnetic materials. [arXiv:2304.09606](https://arxiv.org/abs/2304.09606)
- Cohen T, Welling M (2016) Group equivariant convolutional networks. In: Proceedings of the 33rd international conference on machine learning, pp 2990–2999
- Thomas N, Smidt T, Kearnes S, Yang L, Li L, Kohlhoff K, Riley P (2018) Tensor field networks: rotation-and translation-equivariant neural networks for 3D point clouds. [arXiv:1802.08219](https://arxiv.org/abs/1802.08219)
- Kondor R, Lin Z, Trivedi S (2018) Clebsch–Gordan Nets: a Fully Fourier Space Spherical Convolutional Neural Network. In: Advances in Neural Information Processing Systems, vol 31
- Yu H, Liu B, Zhong Y, Hong L, Ji J, Xu C, Gong X, Xiang H (2024) General time-reversal equivariant neural network potential for magnetic materials. [arXiv:2211.11403](https://arxiv.org/abs/2211.11403)
- Eckhoff M, Behler J (2021) High-dimensional neural network potentials for magnetic systems using spin-dependent atom-centered symmetry functions. *NPJ Comput Mater* 7(1):1–11
- Unke OT, Chmiela S, Gastegger M, Schütt KT, Sauceda HE, Müller K-R (2021) SpookyNet: learning force fields with electronic degrees of freedom and nonlocal effects. *Nat Commun* 12(1):7273
- Novikov I, Grabowski B, Körmann F, Shapeev A (2022) Magnetic moment tensor potentials for collinear spin-polarized materials reproduce different magnetic states of bcc Fe. *NPJ Comput Mater* 8(1)
- Kotlykhov AS, Gubaev K, Hodapp M, Tantarini C, Shapeev AV, Novikov IS (2023) Constrained DFT-based magnetic machine-learning potentials for magnetic alloys: a case study of Fe–Al. *Sci Rep* 13(1):19728
- Yu H, Xu C, Li X, Lou F, Bellaiche L, Hu Z, Gong X, Xiang H (2022) Complex spin Hamiltonian represented by an artificial neural network. *Phys Rev B* 105(17):174422
- Chapman JBJ, Ma P-W (2022) A machine-learned spin-lattice potential for dynamic simulations of defective magnetic iron. *Sci Rep* 12(1):22451
- Yu H, Zhong Y, Hong L, Xu C, Ren W, Gong X, Xiang H (2023) Spin-dependent graph neural network potential for magnetic materials. [arXiv:2203.02853](https://arxiv.org/abs/2203.02853)
- Costa AT, Santos DLR, Peres NM, Fernández-Rossier J (2020) Topological magnons in CrI<sub>3</sub> monolayers: an itinerant fermion description. *2D Mater* 7(4):045031
- Dederichs PH, Blügel S, Zeller R, Akai H (1984) Ground states of constrained systems: application to cerium impurities. *Phys Rev Lett* 53:2512–2515
- Cai Z, Wang K, Xu Y, Wei S-H, Xu B (2023) A self-adaptive first-principles approach for magnetic excited states. *Quantum Front* 2(1):21
- Feynman RP (1939) Forces in molecules. *Phys Rev* 56:340–343
- Kostylev MP, Serga AA, Schneider T, Leven B, Hillebrands B (2005) Spin-wave logical gates. *Appl Phys Lett* 87(15):153501
- Schneider T, Serga AA, Leven B, Hillebrands B, Stamps RL, Kostylev MP (2008) Realization of spin-wave logic gates. *Appl Phys Lett* 92(2):022505
- Ustinov AB, Lähderanta E, Inoue M, Kalinikos BA (2019) Nonlinear spin-wave logic gates. *IEEE Magn Lett* 10:1–4
- Mahmoud A, Ciubotaru F, Vanderveken F, Chumak AV, Hamdioui S, Adelman C, Cotoana S (2020) Introduction to spin wave computing. *J Appl Phys* 128(16):161101
- Volkov OM, Kákay A, Kronast F, Mönch I, Mawass M-A, Fassbender J, Makarov D (2019) Experimental observation of exchange-driven chiral effects in curvilinear magnetism. *Phys Rev Lett* 123:077201
- Hertel R (2013) Curvature-induced magnetochirality. *SPIN* 03(03):1340009
- Edström A, Amoroso D, Picozzi S, Barone P, Stengel M (2022) Curved magnetism in CrI<sub>3</sub>. *Phys Rev Lett* 128(17):177202
- Zheng F (2023) Magnetic skyrmion lattices in a novel 2D-twisted bilayer magnet. *Adv Funct Mater* 33(2):2206923
- Akram M, LaBollita H, Dey D, Kapeghian J, Erten O, Botana AS (2021) Moiré skyrmions and chiral magnetic phases in twisted CrX<sub>2</sub> (X = I, Br, and Cl) bilayers. *Nano Lett* 21(15):6633–6639
- Song T, Sun Q-C, Anderson E, Wang C, Qian J, Taniguchi T, Watanabe K, McGuire MA, Stöhr R, Xiao D, Cao T, Wrachtrup J, Xu X (2021) Direct visualization of magnetic domains and moiré magnetism in twisted 2D magnets. *Science* 374(6571):1140–1144

52. Xu Y, Ray A, Shao Y-T, Jiang S, Weber D, Goldberger JE, Watanabe K, Taniguchi T, Muller DA, Mak KF et al (2021) Emergence of a noncollinear magnetic state in twisted bilayer  $\text{CrI}_3$ . [arXiv:2103.09850](https://arxiv.org/abs/2103.09850)
53. Gilbert TL (2004) Classics in magnetics a phenomenological theory of damping in ferromagnetic materials. *IEEE Trans Magn* 40(6):3443–3449
54. Ivanov AV, Uzdin VM, Jónsson H (2021) Fast and robust algorithm for energy minimization of spin systems applied in an analysis of high temperature spin configurations in terms of skyrmion density. *Comput Phys Commun* 260:107749
55. Geiger M, Smidt T (2022) e3nn: euclidean neural networks. [arXiv:2207.09453](https://arxiv.org/abs/2207.09453)
56. Kresse G, Furthmüller J (1996) Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys Rev B* 54:11169–11186

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

---

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)