Spin-dependent graph neural network potential for magnetic materials

Hongyu Yu [1,2] Yang Zhong,[1,2] Liangliang Hong,[1,2] Changsong Xu,[1,2] Wei Ren,[3] Xingao Gong,[1,2] and Hongjun Xiang [1,2,*]

[1]*Key Laboratory of Computational Physical Sciences (Ministry of Education), Institute of Computational Physical Sciences,*
State Key Laboratory of Surface Physics, and Department of Physics, Fudan University, Shanghai 200433, China
[2]*Shanghai Qi Zhi Institute, Shanghai 200030, China*
[3]*Department of Physics, International Center of Quantum and Molecular Structures and Materials Genome Institute,*
Shanghai University, Shanghai 200444, China

The development of machine-learning interatomic potentials has immensely contributed to the accuracy of simulations of molecules and crystals. However, creating interatomic potentials for magnetic systems that account for both magnetic moments and structural degrees of freedom remains a challenge. This work introduces SpinGNN, a spin-dependent interatomic potential approach that employs the graph neural network (GNN) to describe magnetic systems. SpinGNN consists of two types of edge GNNs: Heisenberg edge GNN (HEGNN) and spin-distance edge GNN (SEGNN). HEGNN is tailored to capture Heisenberg-type spin-lattice interactions, while SEGNN accurately models multibody and high-order spin-lattice coupling. The effectiveness of SpinGNN is demonstrated by its exceptional precision in fitting a high-order spin Hamiltonian and two complex spin-lattice Hamiltonians with great precision. Furthermore, it successfully models the subtle spin-lattice coupling in BiFeO$_3$ and performs large-scale spin-lattice dynamics simulations, predicting its antiferromagnetic ground state, magnetic phase transition, and domain-wall energy landscape with high accuracy. Finally, we perform spin-lattice simulations over one million atoms across GPUs in parallel. Our study broadens the scope of graph neural network potentials to magnetic systems, serving as a foundation for carrying out large-scale spin-lattice dynamic simulations in first-principle accuracy on such systems.

I. INTRODUCTION

Machine-learning (ML) interatomic potentials have transformed the field of computational condensed-matter physics by enabling highly accurate simulations of materials [1,2]. These potentials are trained on a dataset of atomic configurations and their corresponding total energies or forces using a machine-learning model. This allows the model to learn the potential-energy surface (PES) of the system and simulate its behavior under different conditions such as temperature, pressure, or deformation. Compared to traditional methods like density-functional theory, ML interatomic potentials are much faster, making them a powerful tool for investigating large and complex materials systems over long timescales [3–8]. Kernel-based ML interatomic potentials employ material descriptors as inputs and learn to map them to the corresponding energy of the material. These methods include kernel ridge regression and Gaussian process regression. The descriptors, which are derived from the atomic positions and atomic numbers, remain invariant under uniform translations, rotations, and permutations of identical atoms in the material. The construction of the descriptors significantly affects the performance of the machine-learning interatomic potentials (MLIP) model. Various descriptors have been developed to encode material information such as Behler-Parrinello neural networks [9], Gaussian Approximation Potential (GAP) [10], Spectral Neighbor Analysis Potential (SNAP) [11], Moment Tensor Potentials [12], DeepMD [13], Atomic Cluster Expansion (ACE) [14], Smooth Overlap of Atomic Positions (SOAP) [15], and Atom-Centered Symmetry Functions (ACSFs) [16].

A. Graph neural network and message-passing neural network

Graph neural networks (GNNs) with message-passing deep-learning architecture have shown remarkable performance in condensed-matter physics. Message-passing neural network (MPNN) interatomic potentials induce an atomistic graph by connecting each atom (node) to neighboring atoms within a finite cutoff sphere. This structure representation is learned directly from the input structure, which makes it more natural than manual descriptor construction methods. The graph representations of GNN naturally include symmetry invariants such as translation, rotation, and permutation. Recently, molecular and crystal graph neural networks have been well developed as end-to-end learned natural descriptors for crystals [17–25]. The Line Graph Neural Network, a recently proposed extension of GNN, explicitly includes bond angles and updates edge features in the graph, exemplified by ALIGNN [26], and Dimenet++ [20]. Furthermore, equivariant neural networks have been developed to process noninvariant geometric inputs, such as displacement vectors, while maintaining symmetry [27]. By using only E(3)-equivariant operations, these models achieve internal features that are equivariant with respect to the 3D Euclidean group. Equivariant architectures provide an approach for

*hxiang@fudan.edu.cn

creating interatomic potential models such as NequIP [23] and MACE [28]. In particular, the NequIP model, along with other equivariant implementations [29–33], has demonstrated remarkable performance in accurately describing the structural and kinetic properties of complex materials, achieving unprecedented error rates across a broad range of systems and remarkable sample efficiency. The MPNN has shown outstanding performance in training potential, interpreting materials, predicting properties, and inverse design [17,18,34–37]. While MPNNs can learn many-body correlations and access nonlocal information beyond the local cutoff, their parallelism presents significant limitations for large-scale system simulations. The iterative propagation of information in MPNNs results in large receptive fields with many effective neighbors per atom, hindering parallel computation and restricting the accessible length scales for atom-centered MLIPs. A potential solution to these challenges is Allegro [24], a recently proposed interatomic potential architecture that uses strictly local equivariant deep neural networks for scalable and accurate modeling. As a result, molecular dynamics simulations can now predict the structural and kinetic properties of complex systems containing millions of atoms with near first-principles fidelity.

B. Magnetic interatomic potential

In magnetic materials, the PES depends not only on the positions of the atoms but also on the spin configurations of the magnetic atoms. At finite temperatures, atomic magnetic moments fluctuate, and a single magnetic state is insufficient to describe the system. Even at ambient temperatures, atomic spin flips occur frequently due to small energy differences between different spin configurations, typically in the range of a few meV/atom. Therefore, a thorough understanding of magnetic materials necessitates incorporating the effects of temperature on atomic spin configurations. The effective spin-lattice Hamiltonian is traditionally used to understand the spin-lattice coupling in multiferroic materials, such as $BiFeO_3$, based on the system's symmetry [38–42]. However, it is challenging to determine the terms and coefficients required to build the potential, and the effective Hamiltonian method is limited in its ability to describe systems with significant perturbations from the ideal structure.

MLIPs have advanced significantly in their ability to investigate complex systems and physical phenomena, but their failure to consider the diverse range of spin arrangements and magnetic interactions remains a limitation. Many descriptors only work with potentials that do not include atomic magnetic moments, which makes most MLIP models unable to describe the intricate spin-lattice coupling present in magnetic materials. Representing magnetic PES is even more challenging than nonmagnetic PES due to the dimensional crisis. For illustration, for a system with N particles, a nonmagnetic potential function has $3 \times N$ freedom degrees as input of $E = f(\{\vec{r_i}\})$, while for a magnetic potential function, it has $6 \times N$ freedom degrees, including atomic positions and magnetic moments as $E = f(\{\vec{r_i}\}, \{\vec{s_i}\})$, which makes it inaccessible to describe a magnetic system well with traditional atomic potentials. Also, one significant limitation is the cost involved in generating a dataset that is well sampled in $\{\vec{r_i}, \vec{s_i}\}$ configuration

spaces with high accuracy. The spin-polarized first-principle calculation is much more expensive than the ones without spin polarization. Consequently, there is a significant need for a general and robust magnetic neural network potential with high accuracy and efficiency to investigate and better understand magnetic materials.

Several machine-learning methods have been developed for magnetic systems. In our previous work, Yu *et al.* developed the spin Hamiltonian based on the proposed spin descriptors without considering the atomic displacement, so could not consider the spin-lattice coupling [43]. Magnetic moment tensor potentials (mMTPs) [44], developed by Ivan et al., incorporate the contributions of collinear magnetic moments on moment tensor potentials (MTPs) to add collinear magnetic degrees of freedom to MLIP. Although mMTPs have been demonstrated for body-centered cubic iron with calculations and simulations, it should be noted that they can only handle collinear spin and do not explicitly consider spin-inversion symmetry. Marco *et al.* developed the magnetic high-dimensional neural network potential (mHDNNP), which modifies the ACSFs to spin-dependent ACSFs [45]. However, like mMTPs, mHDNNP can only handle collinear magnetic moments, hindering the study of noncollinear magnetic states like spirals, skyrmions, and bimerons. Note that considering noncollinear magnetic states is indispensable for calculating magnetic transition temperature with Monte Carlo simulation and performing spin-lattice dynamics [46,47]. Meanwhile, Svetoslav *et al.* proposed magnetoelastic machine-learning interatomic potentials (ML-IAP) [48] with SNAP [11] potential and a spin Hamiltonian considering Heisenberg and biquadratic terms, mainly considering the spin-spiral state. While magnetoelastic ML-IAP can deal with a noncollinear magnetic state, the spin-lattice part is the traditional explicit term, and the complex interaction cannot be described well under this framework. Moreover, in a recent study, Behnam *et al.* [49] discovered that both SOAP and ACSFs, the most commonly used descriptors, fail to adequately handle four-body interactions, leading to a failure to learn the full energy landscape of materials. To date, the freedom of magnetic moments has not been included in GNN, resulting in underdeveloped MLIP for noncollinear magnetic degrees of freedom.

In this work, we introduce SpinGNN, a generalized spin-dependent graph neural network framework designed to incorporate noncollinear spin into graph neural networks. This framework extends molecular and crystal GNNs to encompass both structural and magnetic states. In our study, we find that Spin-Dimenet++ and Spin-Allegro show high accuracy and effectively capture fine spin-lattice interactions in some magnetic datasets. We further demonstrate that Spin-Allegro is not only efficient in parallelizing across devices for spin-lattice simulations that involve millions of atoms, but it can also predict magnetic ground states accurately. Finally, we show that Spin-Allegro can predict consistent magnetic phase transitions, related transition temperatures, and energy landscape of domain walls of $BiFeO_3$, a commonly studied multiferroic material.

This paper follows the subsequent structure: First, we review related work on kernel-based interatomic potential, graph neural network interatomic potential, and magnetic
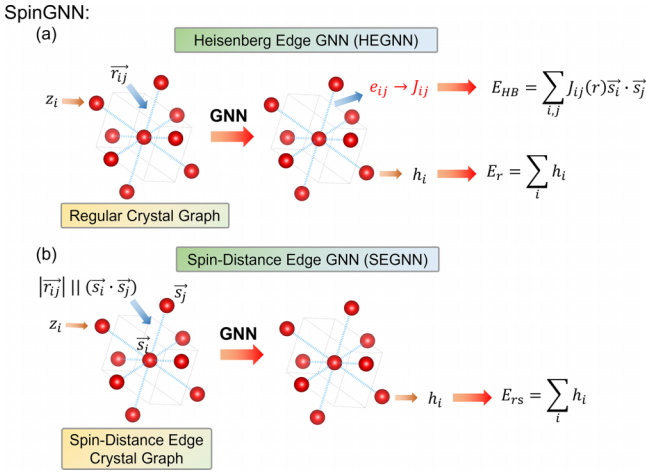
FIG. 1. The SpinGNN framework. Illustration of the SpinGNN including the Heisenberg edge GNN (HEGNN) and spin-distance edge GNN (SEGNN). (a) HEGNN utilizes the updated edge feature of GNN as the Heisenberg coefficients and builds a Heisenberg-based magnetic potential. (b) SEGNN utilizes the spin-distance edge crystal graph which initializes the edge with the dot product of the spin vector and bond length and builds a high-order general magnetic potential.

interatomic potential. Next, we present the core concepts of the SpinGNN framework, as well as the detailed design of Spin-Dimenet++ and Spin-Allegro model based on the popular Dimenet++ and Allegro. Then, we introduce the edge-vector based atomic virial method to enable parallel stress calculation for Allegro. Following that, we present a series of results from magnetic datasets. Finally, we demonstrate the results of spin-lattice dynamics simulation for the multiferroic $BiFeO_3$ on a large scale.

II. RESULTS

The following introduces our proposed method for learning magnetic potential-energy surfaces, utilizing a spin-dependent graph neural network architecture that incorporates noncollinear spin with a stress parallel calculation method for Allegro.

A. The SpinGNN framework

The SpinGNN framework, illustrated in Fig. 1, enhances the standard GNN potential by integrating the degrees of freedom related to atomic positions and collinear or non-collinear magnetic moments. The framework involves two separate neural network architectures, namely the Heisenberg edge graph neural network (HEGNN) and the spin-distance edge graph neural network (SEGNN). HEGNN uses updated edge features in GNN to map the variational Heisenberg coefficients between magnetic atoms. Conversely, SEGNN utilizes both the atomic distance and the dot product of the noncollinear spin between two atoms to initialize the edges in the network. By incorporating HEGNN and SEGNN in a GNN potential, we can create and train a magnetic potential model for describing the magnetic material's complicated PES. While HEGNN supports the discovery of the fundamental Heisenberg magnetic potential, SEGNN enables the modeling of complex high-order magnetic potentials. We can

use both models independently or collectively to develop the magnetic potential of an ensemble. HEGNN is proficient in representing uncomplicated Heisenberg 2-order magnetic interactions, whereas SEGNN can generally illustrate both 2-order and higher-order interactions through graph convolution in GNN, rendering it competent in accommodating most spin-lattice interactions. Since the Heisenberg interactions are mostly prevalent in magnetic potentials, HEGNN is adept at constructing the magnetic potential landscape for materials with simple magnetic interactions that can be expressed through the Heisenberg model. Conversely, SEGNN should be utilized for modeling the magnetic potential landscape for materials with high-order magnetic interactions, where the Heisenberg model falls short. Hence, the choice of the SpinGNN model for creating the magnetic potential landscape is determined by the type of magnetic interactions dominating the material of interest.

B. Concerning general magnetic potentials, HEGNN and SEGNN work in ensembles that are expressed as

$$E^{\text{total}} = E^{\text{HEGNN}} + E^{\text{SEGNN}}, \tag{1}$$

$$= \lambda_r \sum_i h_i^{\text{HEGNN}}(\vec{r}) + \lambda_{HB} \sum_{i,j} J_{ij}^{\text{HEGNN}}(r) \vec{s_i} \cdot \vec{s_j}, \tag{2}$$

$$E^{\text{SEGNN}}(r, \vec{s}) = \lambda_{rs} \sum_i h_i^{\text{SEGNN}}(r, \vec{s}), \tag{3}$$

where $h_i^{\text{HEGNN}}(r)$ is the HEGNN component of the local atomic energy of atom i and $J_{ij}^{\text{HEGNN}}(r)$ represents the Heisenberg coefficient of the bond between atoms i and j. Both terms are solely dependent on the neighboring atomic position environment and do not involve the spin configuration. Here, $h_i^{\text{SEGNN}}(r, \vec{s})$ represents the SEGNN component of the local atomic energy of atom i which relates to the position and spin vector of the surrounding atoms. The coefficients, λ_r, λ_{HB} and λ_{rs}, are used in the potential, as PES is usually dominated by position-related energy, then refined by the Heisenberg interaction and high-order spin-lattice interaction. The spin force, like force and stress, can be calculated using autodifferentiation, as per its definition: the negative gradient of the total energy with respect to the spin vector of atom i. It is denoted as $\vec{\omega_i} = -\partial E / \partial \vec{s_i}$ and is used for carrying spin-lattice dynamics. It can also be used for training if provided in the dataset.

The input vector $\vec{s}$ in HEGNN and SEGNN is a unit vector. The length of the magnetic moment can be predicted by a separate MLP in the output of HEGNN or SEGNN if it varies with the atomic structure and spin configuration. When the length of the magnetic moment remains the same across different atomic structures and spin configurations, such as in multiferroic [50,51], it can be excluded from the SpinGNN prediction and treated as a constant during spin-lattice dynamics.

1. HEGNN

We first introduce the HEGNN based on the Heisenberg model and GNN. The Heisenberg model is commonly used to describe magnetic interactions in materials, and it is ex-
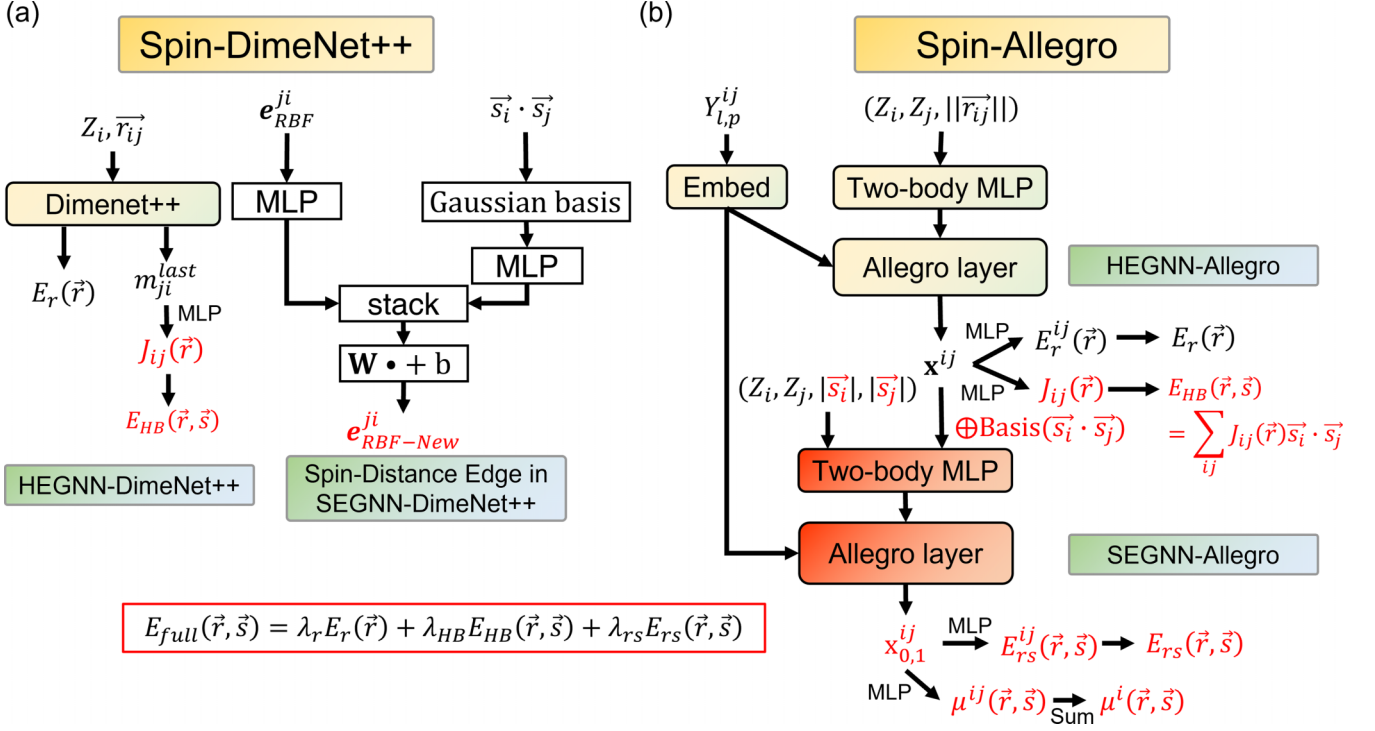
FIG. 2. The Spin-Dimenet++ and Spin-Allegro architecture. (a) Details of the HEGNN-DimeNet++ and spin-distance edge in SEGNN-DimeNet and (b) Details of the Spin-Allegro, which is stacked by HEGNN-Allegro and SEGNN-Allegro. Spin-related parts are highlighted in red.

pressed as $H_{HB} = \sum_{i>j} J_{ij}(r)\vec{s_i} \cdot \vec{s_j}$, where $J_{ij}(r)$ depends on the atomic environment of the bond between two magnetic atoms. However, defining the explicit Heisenberg coefficient $J_{ij}(r)$ is challenging as it is affected by the relative positions of the surrounding atoms, which are difficult to define explicitly. Typically, the Heisenberg coefficient is simplified as $J_{ij}(r_{ij})$ assuming it is only influenced by the atomic distance between two magnetic atoms, thus ignoring the many-body effect. GNNs such as Dimenet++ overcome this limitation by updating the edges of the graph and reflecting the surrounding atomic information of the magnetic bond. As a result, the updated edge features in the line graph can represent the learned Heisenberg coefficient $J_{ij}(r)$ including information about the neighbor atoms of the magnetic bond. The updated node features, on the other hand, represent the local atomic energy, similar to the usual GNN potential. Thus, the magnetic potential of HEGNN can be expressed as Eq. (2), where $E_i^{pos}(r)$ represents the traditional structure potential, derived from the learned node features, and $J_{ij}(r)$ represents the Heisenberg coefficient, derived from the learned edge feature.

2. SEGNN

Then, we introduce the SEGNN model, which utilizes a spin-distance edge crystal graph based on the molecular and crystal GNN with the spin-distance edge feature. Radial basis functions, such as Gaussian expansion [18] and Bessel functions with atomic distance as the only input [19], are typically used for edge initialization of the graph. We propose a spin-distance edge that incorporates bond information on not only the distance but also the spin product, $\vec{s_i} \cdot \vec{s_j}$. SEGNN has two types of edges: magnetic edges and non-

magnetic edges. Magnetic edges exist between two magnetic atoms with magnetic moments, whereas nonmagnetic edges occur when at least one of the two atoms lacks magnetic moments. To construct the spin edge, we use an expansion basis of the value of $\vec{s_i} \cdot \vec{s_j}$ for magnetic edges, and we use a zero array for nonmagnetic edges to distinguish them from the magnetic ones with $\vec{s_i} \cdot \vec{s_j} = 0$. Then, we stack the spin edge and distance edge to form the final spin-distance edge. The implementation of the spin-distance edge could be expressed as

$$e_{ij}^{\text{distance}} = f(\text{RBF}(|\vec{r_{ij}}|)), \tag{4}$$

$$e_{ij}^{\text{spin}} = f(\text{basis}(\vec{s_i} \cdot \vec{s_j})) \text{ for magnetic edges}, \tag{5}$$

$$e_{ij}^{\text{spin}} = f(\text{basis}(0.0) \times 0.0) \text{ for nonmagnetic edges}, \tag{6}$$

$$e_{ij}^{\text{spin-distance}} = f(e_{ij}^{\text{distance}} \oplus e_{ij}^{\text{spin}}), \tag{7}$$

where "RBF" refers to radial basis function, 'basis' refer to a basis function, and $\oplus$ denotes the concatenation. The addition of spin-distance edges provides a simple and effective way to include spin information in the original crystal graph. Therefore, any GNN model can be used to represent magnetic interactions using this modification in the edge-initialization process to construct the magnetic potential.

3. Spin-Dimenet++

Dimenet++ [19,20], a remarkable line graph neural network for materials, is applied within SpinGNN and used to construct the magnetic potential shown in Fig. 2(a). In HEGNN-DimeNet++, the last layer's edge message

features $\boldsymbol{m}_{ji}$ of DimeNet++ are employed to establish the edge Heisenberg coefficients using multilayer perceptrons (MLPs) to map high-dimensional features to a single-value scalar. Meanwhile, in SEGNN-DimeNet++, the spin-distance edge $e_{\text{RBF}-\text{new}}^{(ji)}$ replaces the distance representations $e_{\text{RBF}}^{(ji)}$, an amalgamation of the original $e_{\text{RBF}}^{(ji)}$ and the spin dot product of the two atoms on the bond $\langle \overrightarrow{s_i}, \overrightarrow{s_j} \rangle$ obtained employing some MLPs with details shown in Fig. 2(a).

4. Spin-Allegro

Allegro [24], a strictly local equivariant neural network interatomic potential architecture, is implemented within SpinGNN and utilized to create the magnetic potential and perform large-scale spin-lattice dynamics, shown in Fig. 2(b). HEGNN-Allegro and SEGNN-Allegro are stacked within the Spin-Allegro. HEGNN-Allegro constructs the local edge energy along with the edge Heisenberg coefficients, utilizing distinct MLPs, using updated edge features as input. For SEGNN-Allegro, the edge features from HEGNN-Allegro are combined with $\langle \overrightarrow{s_i}, \overrightarrow{s_j} \rangle$ expansion based on the Fourier basis, the one-hot embedding of atom species, and the spin information of the center and neighbor atoms to serve as input features for the next layer of Allegro:

$$\mathbf{x}_{\text{SEGNN}}^{ij,\, L=0} = \text{MLP}_{\text{two}-\text{body}}(1\text{hot}(Z_i) \oplus 1\text{hot}(Z_j) \oplus |\overrightarrow{s_i}| \oplus |\overrightarrow{s_j}|$$
$$\oplus\, \mathbf{x}_{\text{HEGNN}}^{ij,\, L=\text{last}} \oplus B(\overrightarrow{s_i} \cdot \overrightarrow{s_j})) \cdot u(\overrightarrow{r_{ij}}), \tag{8}$$

where $\oplus$ denotes concatenation, $1\text{hot}(\cdot)$ is a one-hot encoding of the center and neighbor atom species Z_i and Z_j, $|\overrightarrow{s_i}|$ is the normalization of the spin vector, $\mathbf{x}_{\text{HEGNN}}^{ij,\, L=\text{last}}$ is the edge invariant features from the last layer of the HEGNN-Allegro, $B(\overrightarrow{s_i} \cdot \overrightarrow{s_j})$ is the projection of the $\overrightarrow{s_i} \cdot \overrightarrow{s_j}$ onto a Fourier basis, and $u(\overrightarrow{r_{ij}})$ is the polynomial envelope function as proposed in Ref. [19]. While the edges that involve nonmagnetic atoms use an array of zero to replace the $\langle \overrightarrow{s_i}, \overrightarrow{s_j} \rangle$ expansion $B(\overrightarrow{s_i} \cdot \overrightarrow{s_j})$, the magnetic edges require both atoms of the edges to be magnetic. When the magnetic moment μ_i is included in the prediction, an additional MLP is applied on the updated invariant latent features from the last Allegro layer, shown in Fig. 2(b). The trained Spin-Allegro model can be utilized to perform spin-lattice calculations using the SPIN package in LAMMPS [46,52,53]. Furthermore, it is parallelizable across multiple workers and can efficiently perform large-scale spin-lattice dynamics, similar to the large-scale molecular dynamics achievable with Allegro [24].

C. Stress calculation for Allegro in parallel

In neural network potential, stress can be calculated by the derivatives of a zero strain on the cell over the energy with the help of autograd technology. However, stress calculation with the cell is limited during parallel calculation in large systems as the cell is not usually treated as an input. In LAMMPS [53], a widely used molecular dynamics program, ghost atoms instead of the cell and other periodic boundary information are involved during parallel calculation, which further limits stress calculation with the cell in parallelism. We propose an edge-vector based atomic virial method for the parallel stress

calculation of Allegro [24], which makes it possible to carry out *NPT* simulations in parallel.

We first calculate the virial Ξ of the system and stress is equal to $\frac{\Xi}{\Omega}$, where Ω is the volume of the cell. According to the virial definition, $\Xi = \sum_i \boldsymbol{R}_i \boldsymbol{F}_i$, with $\boldsymbol{R}_i$ and $\boldsymbol{F}_i$ as the position and force vector of i's atom, respectively. In Allegro, the energy can be expressed as $E = \sum_{i,j} E_{ij}$ where coefficients related to all the species are ignored for clarity. We define node energy as $E_i = \sum_j E_{ij}$ and note that $E_{ij} \neq E_{ji}$ for the edge energy in Allegro. As the atomic energy is strictly local in Allegro and the position information of atoms is collected as $\boldsymbol{R}_{ij}$ with the center atom i and neighbor atom j, the edge energy can be expressed as $E_{ij} = f_1(Z_i, \{Z_j, \boldsymbol{R}_{ij}, j \in N(i)\})$ and similarly, node energy as $E_i = f_2(Z_i, \{Z_j, \boldsymbol{R}_{ij}, j \in N(i)\})$. Here, $\boldsymbol{R}_{ij}$ is defined as $\boldsymbol{R}_{ij} = \boldsymbol{R}_j - \boldsymbol{R}_i$ and the system energy is $E = \sum_i E_i$. For atom i as the center atom, we can derive that $\frac{\partial E_i}{\partial \boldsymbol{R}_i} = -\sum_{j \neq i} \frac{\partial E_i}{\partial \boldsymbol{R}_{ij}}$ for atom i as the center atom and $\frac{\partial E_j}{\partial \boldsymbol{R}_i} = \frac{\partial E_j}{\partial \boldsymbol{R}_{ji}}$ for atom i as the neighbor atoms. As a result, the force is expressed as $\boldsymbol{F}_i = -\frac{\partial E}{\partial \boldsymbol{R}_i} = -\frac{\partial \sum_j E_j}{\partial \boldsymbol{R}_i} = -\frac{\partial E_i}{\partial \boldsymbol{R}_i} - \sum_{j \neq i} \frac{\partial E_j}{\partial \boldsymbol{R}_i} = \sum_{j \neq i} \frac{\partial E_i}{\partial \boldsymbol{R}_{ij}} - \sum_{j \neq i} \frac{\partial E_j}{\partial \boldsymbol{R}_{ji}}$. Hence, the virial can be expressed as

$$\Xi = \sum_i \boldsymbol{R}_i \boldsymbol{F}_i = \sum_i \boldsymbol{R}_i \left(\sum_{j \neq i} \frac{\partial E_i}{\partial \boldsymbol{R}_{ij}} - \sum_{j \neq i} \frac{\partial E_j}{\partial \boldsymbol{R}_{ji}} \right)$$
$$= \sum_i \boldsymbol{R}_i \sum_{j \neq i} \frac{\partial E_i}{\partial \boldsymbol{R}_{ij}} - \sum_i \boldsymbol{R}_i \sum_{j \neq i} \frac{\partial E_j}{\partial \boldsymbol{R}_{ji}}$$
$$= \sum_i \boldsymbol{R}_i \sum_{j \neq i} \frac{\partial E_i}{\partial \boldsymbol{R}_{ij}} - \sum_j \boldsymbol{R}_j \sum_{i \neq j} \frac{\partial E_i}{\partial \boldsymbol{R}_{ij}}$$
$$= -\sum_i \sum_{j \neq i} \boldsymbol{R}_{ij} \frac{\partial E_i}{\partial \boldsymbol{R}_{ij}}. \tag{9}$$

Considering the symmetry of the virial and the numerical error in the calculation, the final virial is calculated with $\Xi_i = -\sum_{j \neq i} (\boldsymbol{R}_{ij} \frac{\partial E_i}{\partial \boldsymbol{R}_{ij}} + \boldsymbol{R}_{ji} \frac{\partial E_j}{\partial \boldsymbol{R}_{ji}})/2$ as the atomic virial for each atom and $\Xi = \sum_i (\Xi_i + \Xi_i^T)/2$ as the system virial. Ξ_i can be calculated with only center and neighbor atoms' positions in parallel and $\frac{\partial E_i}{\partial \boldsymbol{R}_{ij}}$ can be easily calculated with the autograd technique. This method requires only the positions of the atoms as input without the cell, and the virial of the entire system can be calculated by the accumulation of atomic virial that can be correctly calculated in parallel. Additionally, the atomic virial provided in this method helps obtain the correct heat current for simulations of heat transport [54,55]. A careful numerical check is carried out, ensuring consistency with the results of the zero-strain cell stress method and LAMMPS parallel computation. With this edge-vector based atomic virial method, we can carry out *NPT* simulations in parallel and simulate heat transport with Allegro. See Refs. [56,57] for available implementation.

D. Fitting the artificially constructed spin model

SpinGNN's capability to precisely learn energies from different artificially constructed spin models is demonstrated with Spin-DimeNet++. As depicted in Fig. 2(a),
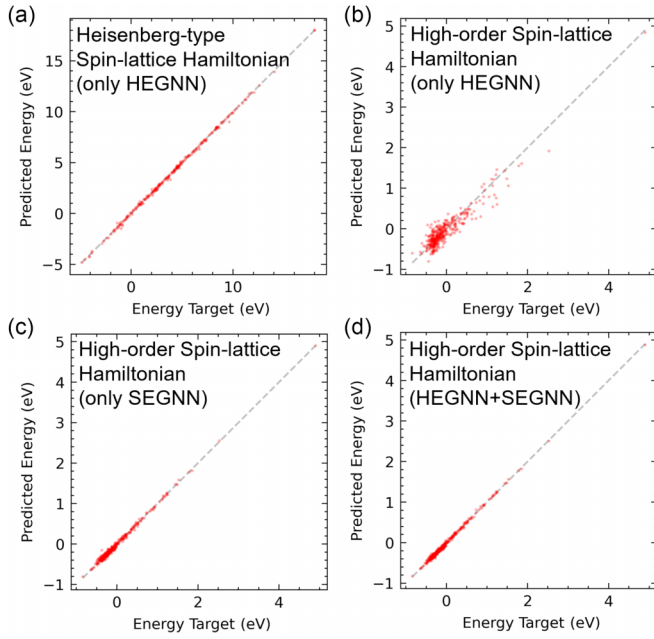
FIG. 3. The prediction results of Spin-DimeNet++ on the energy of test datasets generated from the artificially constructed spin model.

Spin-DimeNet++ precisely models the intricate spin-spin and spin-lattice interactions, achieving high accuracy.

1. Spin Hamiltonian of an artificial model

We first validate Spin-DimeNet++ by developing a spin Hamiltonian with a fixed structure that only accounts for spin degrees of freedom. We utilize an artificial model that includes Heisenberg interactions, fourth-order biquadratic interactions, and four-body, fourth-order interactions to describe the spin interactions between magnetic ions occupying the Al sites of the $CuAlO_2$-type [58] structure. The spin-Hamiltonian model consists of intralayer interactions within a triangular lattice and interlayer couplings:

$$H = \sum_{\langle ij \rangle_n}^{n=1,\ldots,5} J_n \mathbf{S}_i \cdot \mathbf{S}_j + \sum_{\langle ij \rangle_n}^{n=1,5,7} K_n (\mathbf{S}_i \cdot \mathbf{S}_j)^2$$
$$+ \sum_{\langle ijkl \rangle_1} L_1 (\mathbf{S}_i \cdot \mathbf{S}_j)(\mathbf{S}_k \cdot \mathbf{S}_l). \tag{10}$$

For more details regarding this model, please refer to Ref. [43]. Since high-order interactions must be considered, HEGNN and SEGNN are used to construct the GNN spin Hamiltonian, fitting the model well, as shown in Fig. 3(a), with a mean absolute error (MAE) defined as $\frac{\sum |\tilde{f}(x_i) - y_i|}{n}$)

of 1.024×10^{-5} meV per site. This performance is slightly better than that of the machine-learning spin Hamiltonian based on the spin descriptor, with an MAE of 1.03×10^{-5} meV per site [43]. Thus, Spin-DimeNet++ can effectively reconstruct the energy landscape of a complex spin Hamiltonian.

2. Two artificial spin-lattice models

We assess the capability of Spin-DimeNet++ to develop a magnetic potential that considers both structural and magnetic degrees of freedom. To fully validate Spin-DimeNet++, we used two complex spin-lattice model potentials that include different types of magnetic interactions. The first model potential is a Heisenberg type, while the second one includes higher-order magnetic interactions. These model potentials are meant to describe the spin-lattice couplings in the magnetic B sites of the ABO_3 perovskite structure and embody a total of 50 atoms.

The first model potential considered does not include high-order spin interactions. Instead, it consists of 108 terms expressed below, encompassing up to three-body and second-order interactions that are expressed as

$$H = \sum_{\langle ijk \rangle_r}^{r<7\text{Å}} J_{33}(r_i, r_j, r_k) u_{i/j/k} (\mathbf{S}_i \cdot \mathbf{S}_j) + J_{22}(r_i, r_j, r_k)(\mathbf{S}_i \cdot \mathbf{S}_j)$$
$$+ E_3(r_i, r_j, r_k) + \sum_{\langle ij \rangle_r}^{r<7\text{Å}} J_{23}(r_i, r_j) u_{i/j} (\mathbf{S}_i \cdot \mathbf{S}_j)$$
$$+ J_{22}(r_i, r_j)(\mathbf{S}_i \cdot \mathbf{S}_j) + E_2(r_i, r_j), \tag{11}$$

where $r < 7$ Å corresponds to considering up to third-nearest magnetic neighbor B site of the center B site, J_{ij} indicates the Heisenberg coefficient of the i body and jth order which is all set as 0.1 eV, while all other coefficients of terms only related to positions are set to 1 eV. The spin-lattice model potential used in this study is a complex Heisenberg model that does not consider high-order magnetic interactions. Since the first model only involves Heisenberg interactions, we have utilized only HEGNN-DimeNet++ for training the model. The results, as shown in Fig. 3(a), are impressive, with a mean absolute error of 5.04 meV per site. We have also attempted to improve the results by ensembling HEGNN-DimeNet++ and SEGNN-DimeNet++. However, this did not produce a better outcome, and the MAE was 5.57 meV per site due to the overfitting of SEGNN-DimeNet++. As a result, we conclude that HEGNN-DimeNet++ is efficient enough to learn the Heisenberg-type magnetic potential, indicating its ability to grasp the Heisenberg coefficients accurately.

The second model potential introduces fourth-order spin interactions, consisting of 274 terms, encompassing up to three-body and fourth-order interactions that are expressed as

$$H = \sum_{\langle ijk \rangle_r}^{r<5.71\,\text{Å}} J_{34}(r_i, r_j, r_k) u_{i/j/k} u_{i/j/k} (\mathbf{S}_i \cdot \mathbf{S}_j) + J_{33}(r_i, r_j, r_k) u_{i/j/k} (\mathbf{S}_i \cdot \mathbf{S}_j) J_{32}(r_i, r_j, r_k)(\mathbf{S}_i \cdot \mathbf{S}_j) + L(r_i, r_j, r_k)(\mathbf{S}_i \cdot \mathbf{S}_j)(\mathbf{S}_i \cdot \mathbf{S}_k)$$
$$+ E_3(r_i, r_j, r_k) + \sum_{\langle ij \rangle_r}^{r<5.71\,\text{Å}} J_{24}(r_i, r_j) u_{i/j} u_{i/j} (\mathbf{S}_i \cdot \mathbf{S}_j) + J_{23}(r_i, r_j) u_{i/j} (\mathbf{S}_i \cdot \mathbf{S}_j) + J_{22}(r_i, r_j)(\mathbf{S}_i \cdot \mathbf{S}_j) + E_2(r_i, r_j), \tag{12}$$

TABLE I. The energy difference per atom between different structural phases and magnetic states of $BiFeO_3$ are calculated using first-principle calculations and Spin-Allegro and expressed in units of meV per site.

Phase	Spin-Allegro	DFT
$R3c$-AFM	0.00	0.00
$R3c$-FM	44.17	44.35
$R3m$-AFM	35.07	32.10
$R3m$-FM	78.34	81.92
$Pnma$-AFM	10.24	11.53
$Pnma$-FM	56.99	53.38
$Pm3m$-AFM	200.78	201.27
$Pm3m$-FM	246.97	267.18

where $r < 5.71$ Å corresponds to considering up to second-nearest magnetic neighbor B site of the center B site and the coefficients are set the same as the first model that spin-related coefficients at 0.1 eV and all others at 1 eV. The model potential comprises both Heisenberg-type terms and fourth-order terms $L(r_i, r_j, r_k)(\mathbf{S}_i \cdot \mathbf{S}_j)(\mathbf{S}_i \cdot \mathbf{S}_k)$. The fourth-order terms cannot be learned by HEGNN-DimeNet++ alone. Initially, only HEGNN-DimeNet++ was used for fitting, resulting in a high MAE of 2.26 meV per atom in tests as shown in Fig. 3(b). Due to the high-order interactions in the second model, HEGNN-DimeNet++ alone is insufficient, and SEGNN-DimeNet++ is required. Then, only SEGNN-DimeNet++ was used for fitting, resulting in a relatively good MAE of 0.596 meV per atom as shown in Fig. 3(c). Subsequently, HEGNN and SEGNN were used jointly for fitting, resulting in a better MAE of 0.466 meV per atom as shown in Fig. 3(d). SEGNN-DimeNet++ effectively learns high-order spin-lattice interactions, while HEGNN-DimeNet++ excels in learning Heisenberg interactions, and the combination of the two models improves the overall accuracy.

E. Application of multiferroic BiFeO$_3$

We utilized SpinGNN to explore its capability to represent magnetic materials featuring large-scale spin-lattice dynamics. Specifically, we employed Spin-Allegro to investigate

$BiFeO_3$ (BFO), the room-temperature multiferroic materials [39,40,59–64]. BFO has garnered significant interest due to its exclusive combination of ferroelectric and antiferromagnetic properties. BFO possesses a perovskite crystal structure, and its magnetic ground state is antiferromagnetic. BFO displays a phase-transition temperature around 643 K [64], identified as antiferromagnetic Néel temperature, that results from the interplay of different magnetic interactions. Comprehending the mechanism behind the phase transitions in BFO is crucial to designing and optimizing materials for a wide range of applications, including but not limited to data storage, spintronics, and energy harvesting.

A BFO dataset comprising 4857 data points is generated through first-principles calculations. Spin-Allegro, comprising two layers of HEGNN-Allegro and one layer of SEGNN-Allegro, achieves high accuracy on the test dataset, with an MAE of 1.40 meV per site and 23.2 meV/Å for energy and force accuracy, respectively. In contrast, the original Allegro with the same configuration is trained with an MAE of 3.55 meV per site and 30.1 meV/Å for energy and force accuracy, respectively. Energy of $BiFeO_3$ with different structural phases and magnetic phases calculated by density-functional theory (DFT) and Spin-Allegro model are compared, shown in Table I.

To validate the potential accuracy, we utilized the potential along with the conjugate gradient (CG) algorithm [65] to explore the magnetic ground state of BFO. The investigation began with a $12 \times 12 \times 12$ supercell composed of 8640 atoms and 1728 Fe atoms with random spin configurations. The subsequent 1000 CG iterations produced a zero-spin temperature perfect G-type antiferromagnetic state, which agrees with the experimental data [59–61]. A small cell of BFO structure with a G-type antiferromagnetic state is shown in Fig. 4(a).

Then, we utilized spin-lattice semi-implicit B (SIB)-NPT simulation [51,66] over a range of temperatures to examine the magnetic and structural phase transition of BFO, starting from the same structure as earlier. We collected the antiferromagnetic vector from 50 to 1200K with two supercells including 8640 and 109 760 atoms, respectively, as shown in Fig. 4(b). Also, we collected the polarization vector from 50 to 1300 K with the supercells including 8640 atoms, shown
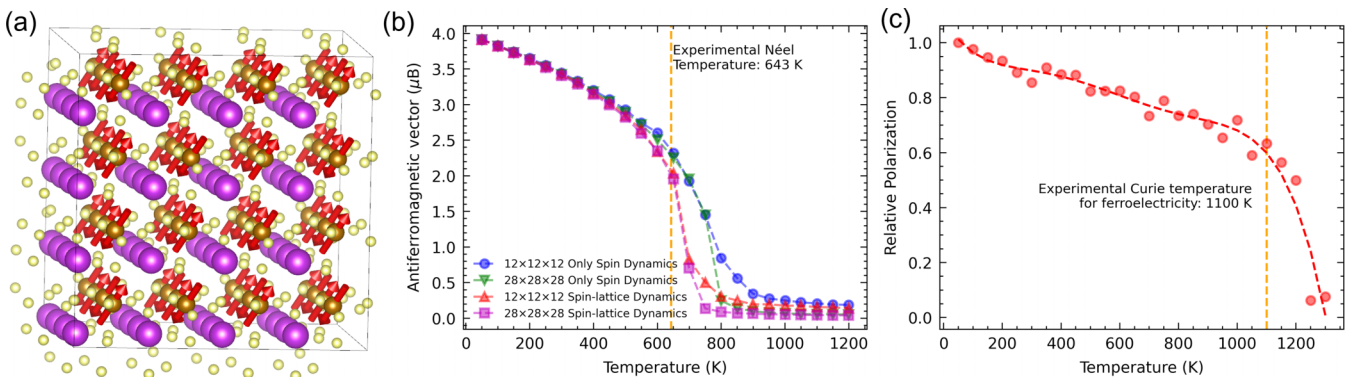

FIG. 4. (a) The structure of $BiFeO_3$ with ground G-type antiferromagnetic state. (b), (c) The antiferromagnetic vector and relative polarization in different temperatures of $BiFeO_3$ from simulations of spin-lattice dynamics and spin dynamics with Spin-Allegro. Obvious magnetic and ferroelectric transition is observed and transition temperatures can be estimated, which are close to the experimental values.

in Fig. 4(c). The polarization vector is estimated by the Born effective charge and atomic displacement. As the temperature increased, the antiferromagnetic vector and relative polarization decreased, indicating a magnetic transition from the antiferromagnetic to the paramagnetic state and a structural transition from the ferroelectric to the paraelectric state. The Néel and Curie temperature were estimated to be 650 and 1200 K, respectively, consistent with experiments [59–61]. We conducted a spin dynamics simulation without atomic motions and confirmed that spin-lattice coupling drives a lower and more accurate transition temperature shown in Fig. 4(b). The Spin-Allegro model provides a precise depiction of various atomic and spin phenomena at different temperatures. These results indicate that the model effectively characterizes the intricate spin-lattice interaction of BFO.

Furthermore, we investigated the domain wall (DW) of BFO, which plays a crucial role in its properties and applications. Different DW types in BFO have been observed through experimental studies, such as DW109, DW180, and DW71, with characteristic angles between the polarization of 109°, 180°, and 71°, respectively. These DWs can be controlled by external conditions such as magnetic fields, temperature, and pressure, and have gained attention for their potential applications [67–69]. However, the complex and large structure of the DWs in BFO presents challenges for effectively optimizing their spin orientation and configuration and understanding their various properties [70–74]. To address this, we optimized several DW structures using the CG and annealing algorithm [75] with Spin-Allegro and investigated their metastable structures and related energy. We analyzed the DW energies of DW109, DW180, and DW71 by calculating $(E_{\rm DW} - E_{\rm ferro})/(2/A)$, where A is the area of the supercell's topmost layer and $E_{\rm ferro}$ is the energy of pure ferroelectric structure. Initially, we created relatively small-cell DWs with 160, 80, and 80 atoms for DW109, DW180, and DW71, respectively. The DW energies of these DWs were in the order of 65.70, 166.4, and 182.0 mJ/m^2, respectively, which was in line with previous studies [71–74] and presented a relatively consistent value. The optimized structures were analyzed using DFT, and the forces on the atoms were minimal, demonstrating the efficacy of optimization with Spin-Allegro. Subsequently, we constructed larger DW structures with 1280, 640, and 640 atoms for DW109, DW180, and DW71, respectively, by increasing the size of the small DW structures' supercell to $2 \times 2 \times 2$. The DW energies of these structures were found to be 65.27, 170.5, and 184.6 mJ/m^2, respectively, which were relatively close to those of the small structures. All simulations ended with an antiferromagnetic state. Detailed energy values of the structures and workflow of optimization are provided in Supplemental Material [76].

To demonstrate the computational efficiency of the Spin-Allegro in parallel, several spin-lattice simulations were performed with different numbers of GPUs, shown in Table II. We used different sizes of bulk BiFeO$_3$ from a DFT-afforded size to over one million atoms with up to 96 GPUs. For comparison, it takes about 2000 s to carry out the spin-polarized calculation on 80-atom BiFeO$_3$ with 80 CPUs. Spin-lattice simulations for over one million atoms can be carried out with close to first-principle calculations. With a smaller

TABLE II. The computational efficiency of the spin-lattice simulations for different sizes of bulk BiFeO$_3$ performed with different numbers of GPUs. The time step of 0.5 fs is used.

Number of atoms	Number of GPUs	Speed in ps/day	Time steps (s)
80	1	1792	41.485
8 640	1	55	1.262
8 640	2	106	2.449
8 640	4	202	4.668
8 640	8	371	8.595
8 640	16	641	14.845
8 640	32	1024	23.714
109 760	8	34	0.794
109 760	16	67	1.553
109 760	24	99	2.282
233 280	16	32	0.745
1 080 000	48	21	0.479
1 080 000	64	28	0.642
1 080 000	96	41	0.958

Spin-Allegro model with fewer parameters, better efficiency could be achieved with less accuracy.

III. DISCUSSION

We introduced a deep-learning magnetic interatomic potential suitable for magnetic materials that accounts for the freedom degrees of atomic position and noncollinear magnetic moment. We achieved high accuracy in this endeavor by proposing a Spin Graph Neural Network. Our resulting Spin-DimeNet++ and Spin-Allegro models, based on the SpinGNN framework, exhibited good performance on the magnetic dataset, furnishing a dependable energy landscape for an artificial high-order spin Hamiltonian, two artificial complex spin-lattice models, as well as the multiferroic BiFeO$_3$. Spin-Allegro is equipped to scale to large system sizes of over one million atoms with ease thanks to the strict locality of its geometric representations in Allegro, along with parallelization in large-scale calculations. This renders it feasible to examine the magnetic, structural, and kinetic properties of complex magnetic systems that comprise millions of atoms with spin-lattice dynamics simulations at nearly first-principles fidelity. Compared to the spin-Hamiltonian method based on the spin descriptor and neural network without considering the atomic displacement in our previous work, SpinGNN serves as a more general framework including both freedom degrees of structural and magnetic states, which are important for most magnetic systems such as the multiferroics. Meanwhile, SpinGNN is based on the more expressive message-passing neural network and shows better performance. Given that the magnetic potential is important for simulations of magnetic materials and understanding magnetism, we expect that our SpinGNN framework and proposed parallelizable Spin-Allegro will find wide use in the study of magnetism.

Our framework facilitates the study of previously inaccessible magnetic materials systems with graph neural networks. It is worth noting that the SpinGNN framework can be

implemented with most graph neural network architectures, such as MACE [28], GEMNET [21], NequIP [23], and ALIGNN [26] to be implemented just as DimeNet++ [20] and Allegro [24]. Compared to previous works that built magnetic potential based on different descriptors with different implementations, SpinGNN serves as a general framework to transfer the common neural network potential into a magnetic potential. Furthermore, just as the GNN potential usually outperforms potential based on descriptors [77], the magnetic GNN potential with the SpinGNN framework is expected to be more powerful than the previously proposed magnetic potential based on descriptors, although a benchmark magnetic dataset is not yet available. With the continuous development of the neural network potential, particularly in recent years, we expect a better magnetic neural network potential to be built with a more powerful graph neural network potential and SpinGNN framework. Spin-Allegro is based on one of the most popular graph neural network potentials and has been proven to be not only powerful enough but also easy to parallelize for large-scale spin-lattice dynamics. Also, the time-reversal symmetry is naturally conserved in SpinGNN, where the feature of the spin-distance edge is about $\overrightarrow{s_i} \cdot \overrightarrow{s_j} = (-\overrightarrow{s_i}) \cdot (-\overrightarrow{s_j})$ and the Heisenberg-type spin-lattice formula $J_{ij}^{\mathrm{HEGNN}}(r)\overrightarrow{s_i} \cdot \overrightarrow{s_j} = J_{ij}^{\mathrm{HEGNN}}(r)(-\overrightarrow{s_i}) \cdot (-\overrightarrow{s_j})$.

The stress implementations for NequIP, Allegro, and LAMMPS interface are available. See Refs. [56,57].

APPENDIX

1. Software

We developed our codes utilizing the modified NequIP code available (see Ref. [56]) under git commit bb10cc6f91ca42fda452d7b22cc314d33c0e3d95, Allegro code available (see Ref. [78]) under git commit a7899a9c4c6be0620e11bef0bca8d06d9f7f32a8, as well as e3nn [27] with version 0.5.0, PYTORCH [79] with version 1.11.0 and PYTHON with 3.9.13. Spin-DimeNet++ is developed based on the implementation in PYTORCH_geometric [80]; see Ref. [81]. The LAMMPS was built based on the LAMMPS code [82] under git commit 9b989b186026c6fe9da354c79cc9b4e152ab03af with the modified pair_allegro code available [57], git commit 101d2c8db123c9549490fa7285ce5d85a616d46f. The VESTA [83] software was used to generate Figs. 1 and 4(a). MATPLOTLIB [84] was used for plotting results.

2. Spin-lattice dynamics simulations

We performed spin-lattice dynamics simulations using the SPIN package [46] in LAMMPS [53]. To ensure the accurate integration of the Landau-Lifshitz-Gilbert (LLG) [85] equation

and enforce the conservation of the magnetic moments' magnitude [51], we implemented the semi-implicit SIB method introduced by Mentink *et al.* [66] in LAMMPS and the SPIN package. The dynamics of spins and atomic motions were simulated using the molecular dynamics scheme (*NVT* or *NPT*) and the SIB method. Our simulations of BFO were carried out on the $12 \times 12 \times 12$ periodic supercell comprising 8640 atoms and the $28 \times 28 \times 28$ periodic supercell including 109 760 atoms with a time step of 0.5 fs in *NPT*-SIB. Temperature control was exerted by implementing the Nosé-Hoover thermostat with a temperature damping parameter of 100 time steps, and pressure control was maintained using the Parrinello-Rahman barostat with a temperature damping parameter of 1000 time steps, as implemented in LAMMPS [52]. To model the temperature-driven antiferromagnetic-paramagnetic, we simulated the dynamics of atomic motions and spin motions using *NPT* simulations, respectively, with a damping coefficient in the LLG equation of 0.01. The simulation commenced with BFO in the $R3c$ phase, in which the antiferromagnetic configurations were initialized randomly in 100 K with a Gaussian distribution of velocity. The system was heated from 100 to 1200 K in 30 ps and then cooled from 1200 to 50 K in 120 ps in steps of 50 K. At each given temperature, an equilibrium run of 5 ps was followed by a production run of 10 ps.

3. Reference training sets

We build the dataset of artificially constructed spin models using the effective Hamiltonian in the Property Analysis and Simulation Package (PASP) [41]. The first dataset of spin Hamiltonian is established on $CuMO^2$, where M is a magnetic ion, with a fixed $CuAlO_2$-type structure. We generate a dataset of 10 000 fully randomized spin configurations based on a $4 \times 4 \times 2$ supercell with 96 magnetic sites, i.e., 6 layers of 4×4 triangular lattices, and obtain their corresponding energies from Eq. (10). We refer the reader to our previous work [43] for more details. For the two spin-lattice effective Hamiltonian datasets, we construct atomic configurations based on the ideal structures of ABO_3 perovskite with 50 atoms in total, comprising magnetic B sites. We randomly sample spin configurations and calculate energy based on the constructed effective Hamiltonian from Eq. (11) and Eq. (12). In total, we generate 2416 and 3611 data with different distortions and spin configurations for the two spin-lattice Hamiltonians, respectively.

For the BFO dataset, we perform density-functional theory (DFT) calculations using VASP [86–88] with the projector augmented-wave (PAW) method [89] and the Perdew-Burke-Ernzerhof (PBE) functional [90]. We generate atomic configurations through several *NPT* molecular dynamics simulations starting from $4 \times 2 \times 2$ supercells with 80 atoms under different phases, including $R3c$, $R3m$, $Pnma$, and Pm-$3m$. The temperatures are sampled from 600 to 1500 K including 600, 800, 1200, and 1500 K. We also randomly sample spin configurations in the calculations without considering spin-orbit couplings [91]. We choose an energy cutoff for the plane-wave basis of 500 eV and a $2 \times 4 \times 4$ k-point grid. Total energy calculations are converged to 10^{-6} eV per supercell to ensure high-accuracy DFT data. We use a self-consistent value of the

effective Hubbard U [92] of 3.8 eV [38] for the localized $3d$ electrons of Fe ions according to the calculation result from the linear response approach [93].

4. Training details

Models were trained on an NVIDIA A100 GPU in single-GPU training with the mean-squared-error (MSE) loss function and the Adam optimizer [94] in PYTORCH [79] with default parameters of $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\varepsilon = 10^{-8}$. We lowered the learning rate with an on-plateau scheduler that responded to the validation loss. The learning stopped when the learning rate dropped below 10^{-6} or no advancement was made in the validation loss for multiple epochs. All models were trained with float32 precision. We utilized the SiLU activation function [95].

(a) Spin Hamiltonian and spin-lattice Hamiltonians. Spin-DimeNet++ was implemented with identical parameters across all three datasets. The model was trained on energy prediction as the target variable, with 4 layers, each consisting of 128 hidden channels. In addition, we employed 64 interaction triplet embeddings, 8 basis transformation embeddings, 256 output embedding channels, 7 spherical harmonics basis, and 5 for envelope exponent. Three linear layers were used for the output blocks, and both interaction blocks before and after the skip connection were composed of 1 and 2 residual layers, respectively. The batch sizes were 10, 6, and 8 for the spin Hamiltonian, Heisenberg type, and high-order spin-lattice Hamiltonian, respectively. We initialized the learning rate to be 0.0005 for all models, with a learning rate patience of 5, a decay factor of 0.8, and a stop patience of 20. We chose cutoff values of 8, 7.2, and 6 Å based on the Hamiltonian construction to ensure sufficient information was included. The Adam optimizer's weight decay was 0.01. Finally, we randomly partitioned the data into three sets: 80% for training, 10% for validation, and 10% for testing.

(b) $BiFeO_3$. We split our data into training (4200 samples), validation (357 samples), and testing (300 samples) sets. The target variables were energies, forces, and stresses. To define the loss function, we used per-atom mean-squared error [24], with weights of 1 assigned to the energy, force, and stress terms, and no weight decay in the Adam optimizer. For HBGNN-Allegro, we employed two layers, 16 features for both even and odd irreps, and an $\ell max = 2$. We used one layer for SEGNN-Allegro. The two-body latent MLP for both models consisted of three hidden layers with dimensions [64, 128, 128] using SiLU nonlinearities. The later latent MLPs consisted of one hidden layer of dimension 128 using SiLU nonlinearity. We implemented the embedding weight projection using a single matrix multiplication. For HEGNN-Allegro, the final edge energy and Heisenberg coefficient MLPs both had three hidden layers with dimensions [128, 128, 64], and SiLU nonlinearities, while for SEGNN-Allegro, the final edge energy MLP had one hidden layer of dimension 128 without nonlinearity. The weight initialization for all MLPs used a uniform distribution of unit variance. The coefficients λ_r, λ_{HB}, and λ_{rs} in Eqs. (1)–(3) and Fig. 2 were set to 1.0, 0.05, and 0.05, respectively. We set the radial cutoff at 6.0 Å and utilized an 8 nontrainable Bessel function basis for encoding with the polynomial envelope function using $p = 48$. The learning rate was set to 0.005, batch size to 1, with an on-plateau scheduler reducing the learning rate based on the validation loss of force every time it hit a plateau. The patience was set to 5, and the decay factor was set to 0.3. An exponential moving average with a weight of 0.99 was used to evaluate the validation set and the final model. We stopped the training when the learning rate dropped below 1×10^{-6}.

[1] G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto, and L. Zdeborová, Machine learning and the physical sciences, Rev. Mod. Phys. **91**, 045002 (2019).

[2] A. Agrawal and A. Choudhary, Perspective: Materials informatics and big data: Realization of the "fourth paradigm" of science in materials science, APL Mater. **4**, 053208 (2016).

[3] C. Chen and S. P. Ong, A universal graph deep learning interatomic potential for the periodic table, Nat. Comput. Sci. **2**, 718 (2022).

[4] X. Li, Z. Li, and J. Chen, Ab initio calculation of real solids via neural network ansatz, Nat. Commun. **13**, 1 (2022).

[5] V. L. Deringer, N. Bernstein, G. Csányi, C. Ben Mahmoud, M. Ceriotti, M. Wilson, D. A. Drabold, and S. R. Elliott, Origins of structural and electronic transitions in disordered silicon, Nature (London) **589**, 59 (2021).

[6] S. Chmiela, H. E. Sauceda, K.-R. Müller, and A. Tkatchenko, Towards exact molecular dynamics simulations with machine-learned force fields, Nat. Commun. **9**, 3887 (2018).

[7] T. W. Ko, J. A. Finkler, S. Goedecker, and J. Behler, A fourth-generation high-dimensional neural network potential with accurate electrostatics including non-local charge transfer, Nat. Commun. **12**, 398 (2021).

[8] P. Friederich, F. Häse, J. Proppe, and A. Aspuru-Guzik, Machine-learned potentials for next-generation matter simulations, Nat. Mater. **20**, 750 (2021).

[9] J. Behler and M. Parrinello, Generalized neural-network representation of high-dimensional potential-energy surfaces, Phys. Rev. Lett. **98**, 146401 (2007).

[10] A. P. Bartók, M. C. Payne, R. Kondor, and G. Csányi, Gaussian approximation potentials: The accuracy of quantum mechanics, without the electrons, Phys. Rev. Lett. **104**, 136403 (2010).

[11] A. P. Thompson, L. P. Swiler, C. R. Trott, S. M. Foiles, and G. J. Tucker, Spectral neighbor analysis method for automated generation of quantum-accurate interatomic potentials, J. Comput. Phys. **285**, 316 (2015).

[12] A. V. Shapeev, Moment tensor potentials: A class of systematically improvable interatomic potentials, Multiscale Model. Simul. **14**, 1153 (2016).

[13] L. Zhang, J. Han, H. Wang, R. Car, and E. W., Deep potential molecular dynamics: A scalable model with the accuracy of quantum mechanics, Phys. Rev. Lett. **120**, 143001 (2018).

[14] R. Drautz, Atomic cluster expansion for accurate and transferable interatomic potentials, Phys. Rev. B **99**, 014104 (2019).

[15] A. P. Bartók, R. Kondor, and G. Csányi, On representing chemical environments, Phys. Rev. B **87**, 184115 (2013).

[16] J. Behler, Atom-centered symmetry functions for constructing high-dimensional neural network potentials, J. Chem. Phys. **134**, 074106 (2011).

[17] K. T. Schütt, H. E. Sauceda, P.-J. Kindermans, A. Tkatchenko, and K.-R. Müller, SchNet–a deep learning architecture for molecules and materials, J. Chem. Phys. **148**, 241722 (2018).

[18] T. Xie and J. C. Grossman, Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties, Phys. Rev. Lett. **120**, 145301 (2018).

[19] J. Klicpera, J. Groß, and S. Günnemann, Directional message passing for molecular graphs, arXiv:2003.03123.

[20] J. Klicpera, S. Giri, J. T. Margraf, and S. Günnemann, Fast and uncertainty-aware directional message passing for non-equilibrium molecules, arXiv:2011.14115.

[21] J. Klicpera, F. Becker, and S. Günnemann, Gemnet: Universal directional graph neural networks for molecules, arXiv:2106.08903.

[22] J. Schmidt, L. Pettersson, C. Verdozzi, S. Botti, and M. A. L. Marques, Crystal graph attention networks for the prediction of stable materials, Sci. Adv. **7**, eabi7948 (2021).

[23] S. Batzner, A. Musaelian, L. Sun, M. Geiger, J. P. Mailoa, M. Kornbluth, N. Molinari, T. E. Smidt, and B. Kozinsky, E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials, Nat. Commun. **13**, 2453 (2022).

[24] A. Musaelian, S. Batzner, A. Johansson, L. Sun, C. J. Owen, M. Kornbluth, and B. Kozinsky, Learning local equivariant representations for large-scale atomistic dynamics, Nat. Commun. **14**, 579 (2023).

[25] M. D. Witman, A. Goyal, T. Ogitsu, A. H. McDaniel, and S. Lany, Defect graph neural networks for materials discovery in high-temperature clean-energy applications, Nat. Comput. Sci. **3**, 675 (2023).

[26] K. Choudhary and B. DeCost, Atomistic line graph neural network for improved materials property predictions, npj Comput. Mater. **7**, 185 (2021).

[27] M. Geiger and T. Smidt, E3nn Euclidean Neural Networks, arXiv:2207.09453.

[28] I. Batatia, D. P. Kovács, G. N. C. Simm, C. Ortner, and G. Csányi, MACE: Higher order equivariant message passing neural networks for fast and accurate force fields, arXiv:2206.07697.

[29] K. T. Schütt, O. T. Unke, and M. Gastegger, Equivariant message passing for the prediction of tensorial properties and molecular spectra, arXiv:2102.03150.

[30] Z. Qiao, A. S. Christensen, F. R. Manby, M. Welborn, A. Anandkumar, and T. F. Miller III, Unite: Unitary n-body tensor equivariant network with applications to quantum chemistry, arXiv:2105.14655.

[31] M. Haghighatlari *et al.*, NewtonNet: A Newtonian message passing network for deep learning of interatomic potentials and forces, Digit. Discov. **1**, 333 (2022).

[32] P. Thölke and G. De Fabritiis, TorchMD-NET: Equivariant transformers for neural network based molecular potentials, arXiv:2202.02541.

[33] J. Brandstetter, R. Hesselink, E. van der Pol, E. J. Bekkers, and M. Welling, Geometric and physical quantities improve E(3) equivariant message passing, arXiv:2110.02905.

[34] Q. Wang and L. Zhang, Inverse design of glass structure with deep graph neural networks, Nat. Commun. **12**, 5359 (2021).

[35] T. Xie, X. Fu, O.-E. Ganea, R. Barzilay, and T. Jaakkola, Crystal diffusion variational autoencoder for periodic material generation, arXiv:2110.06197.

[36] H. Li, Z. Wang, N. Zou, M. Ye, R. Xu, X. Gong, W. Duan, and Y. Xu, Deep-learning density functional theory Hamiltonian for efficient ab initio electronic-structure calculation, Nat. Comput. Sci. **2**, 367 (2022).

[37] D. Wang, Y. Wang, J. Chang, L. Zhang, H. Wang, and E. W., Efficient sampling of high-dimensional free energy landscapes using adaptive reinforced dynamics, Nat. Comput. Sci. **2**, 20 (2022).

[38] I. A. Kornev, S. Lisenkov, R. Haumont, B. Dkhil, and L. Bellaiche, Finite-temperature properties of multiferroic $BiFeO_3$, Phys. Rev. Lett. **99**, 227602 (2007).

[39] C. Xu, B. Xu, B. Dupé, and L. Bellaiche, Magnetic interactions in $BiFeO_3$: A first-principles study, Phys. Rev. B **99**, 104420 (2019).

[40] B. Xu, B. Dupé, C. Xu, H. Xiang, and L. Bellaiche, Revisiting spin cycloids in multiferroic $BiFeO_3$, Phys. Rev. B **98**, 184420 (2018).

[41] F. Lou, X. Y. Li, J. Y. Ji, H. Yu, J. S. Feng, X. G. Gong, and H. J. Xiang, PASP: Property analysis and simulation package for materials, J. Chem. Phys. **154**, 114103 (2021).

[42] X. Li, H. Yu, F. Lou, J. Feng, M.-H. Whangbo, and H. Xiang, Spin Hamiltonians in magnets: Theories and computations, Molecules **26**, 803 (2021).

[43] H. Yu, C. Xu, X. Li, F. Lou, L. Bellaiche, Z. Hu, X. Gong, and H. Xiang, Complex spin Hamiltonian represented by an artificial neural network, Phys. Rev. B **105**, 174422 (2022).

[44] I. Novikov, B. Grabowski, F. Körmann, and A. Shapeev, Magnetic moment tensor potentials for collinear spin-polarized materials reproduce different magnetic states of Bcc Fe, npj Comput. Mater. **8**, 13 (2022).

[45] M. Eckhoff and J. Behler, High-dimensional neural network potentials for magnetic systems using spin-dependent atom-centered symmetry functions, npj Comput. Mater. **7**, 170 (2021).

[46] J. Tranchida, S. J. Plimpton, P. Thibaudeau, and A. P. Thompson, Massively parallel symplectic algorithm for coupled magnetic spin dynamics and molecular dynamics, J. Comput. Phys. **372**, 406 (2018).

[47] K. Hukushima and K. Nemoto, Exchange Monte Carlo method and application to spin glass simulations, J. Phys. Soc. Jpn. **65**, 1604 (1996).

[48] S. Nikolov, M. A. Wood, A. Cangi, J.-B. Maillet, M.-C. Marinica, A. P. Thompson, M. P. Desjarlais, and J. Tranchida, Data-driven magneto-elastic predictions with scalable classical spin-lattice dynamics, npj Comput. Mater. **7**, 153 (2021).

[49] B. Parsaeifard and S. Goedecker, Manifolds of quasi-constant SOAP and ACSF fingerprints and the resulting failure to machine learn four-body interactions, J. Chem. Phys. **156**, 034302 (2022).

[50] J. B. Neaton, C. Ederer, U. V. Waghmare, N. A. Spaldin, and K. M. Rabe, First-principles study of spontaneous polarization in multiferroic $BiFeO_3$, Phys. Rev. B **71**, 014113 (2005).

[51] D. Wang, J. Weerasinghe, and L. Bellaiche, Atomistic molecular dynamic simulations of multiferroics, Phys. Rev. Lett. **109**, 067203 (2012).

[52] S. Plimpton, Fast parallel algorithms for short-range molecular dynamics, J. Comput. Phys. **117**, 1 (1995).

[53] A. P. Thompson *et al.*, LAMMPS - a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales, Comput. Phys. Commun. **271**, 108171 (2022).

[54] Z. Fan, L. F. C. Pereira, H.-Q. Wang, J.-C. Zheng, D. Donadio, and A. Harju, Force and heat current formulas for many-body potentials in molecular dynamics simulations with applications to thermal conductivity calculations, Phys. Rev. B **92**, 094301 (2015).

[55] A. J. Gabourie, Z. Fan, T. Ala-Nissila, and E. Pop, Spectral decomposition of thermal conductivity: Comparing velocity decomposition methods in homogeneous molecular dynamics simulations, Phys. Rev. B **103**, 205421 (2021).

[56] https://github.com/Hongyu-yu/nequip.

[57] https://github.com/Hongyu-yu/pair_allegro.

[58] T. Ishiguro, A. Kitazawa, N. Mizutani, and M. Kato, Single-crystal growth and crystal structure refinement of $CuAlO_2$, J. Solid State Chem. **40**, 170 (1981).

[59] C. Blaauw and F. van der Woude, Magnetic and structural properties of $BiFeO_3$, J. Phys. C: Solid State Phys. **6**, 1422 (1973).

[60] P. Fischer, M. Polomska, I. Sosnowska, and M. Szymanski, Temperature dependence of the crystal and magnetic structures of $BiFeO_3$, J. Phys. C: Solid State Phys. **13**, 1931 (1980).

[61] D. V. Karpinsky *et al.*,, Thermodynamic potential and phase diagram for multiferroic bismuth ferrite ($BiFeO_3$), npj Comput. Mater. **3**, 1 (2017).

[62] S. M. Selbach, T. Tybell, M.-A. Einarsrud, and T. Grande, The ferroic phase transitions of $BiFeO_3$, Adv. Mater. **20**, 3692 (2008).

[63] Y. Song, B. Xu, and C.-W. Nan, Lattice and spin dynamics in multiferroic $BiFeO_3$ and $RMnO_3$, Natl. Sci. Rev. **6**, 642 (2019).

[64] N. A. Spaldin, S.-W. Cheong, and R. Ramesh, Multiferroics: past, present, and future, Phys. Today **63**, 38 (2010).

[65] M. R. Hestenes and E. Stiefel, Methods of conjugate gradients for solving linear systems, J. Res. Natl. Bur. Stand. **49**, 409 (1952).

[66] J. H. Mentink, M. V. Tretyakov, A. Fasolino, M. I. Katsnelson, and T. Rasing, Stable and fast semi-implicit integration of the stochastic Landau–Lifshitz equation, J. Phys.: Condens. Matter **22**, 176001 (2010).

[67] Z. Chen, J. Liu, Y. Qi, D. Chen, S.-L. Hsu, A. R. Damodaran, X. He, A. T. N'Diaye, A. Rockett, and L. W. Martin, 180° Ferroelectric stripe nanodomains in $BiFeO_3$ thin films, Nano Lett. **15**, 6506 (2015).

[68] Y.-H. Chu *et al.*, Nanoscale domain control in multiferroic $BiFeO_3$ thin films, Adv. Mater. **18**, 2307 (2006).

[69] H. Huyan, L. Li, C. Addiego, W. Gao, and X. Pan, Structures and electronic properties of domain walls in $BiFeO_3$ thin films, Natl. Sci. Rev. **6**, 669 (2019).

[70] A. Lubk, S. Gemming, and N. A. Spaldin, First-principles study of ferroelectric domain walls in multiferroic bismuth ferrite, Phys. Rev. B **80**, 104110 (2009).

[71] O. Diéguez, P. Aguado-Puente, J. Junquera, and J. Íñiguez, Domain walls in a perovskite oxide with two primary structural order parameters: First-principles study of $BiFeO_3$, Phys. Rev. B **87**, 024102 (2013).

[72] W. Ren, Y. Yang, O. Diéguez, J. Íñiguez, N. Choudhury, and L. Bellaiche, Ferroelectric domains in multiferroic $BiFeO_3$ films under epitaxial strains, Phys. Rev. Lett. **110**, 187601 (2013).

[73] Y. Wang, C. Nelson, A. Melville, B. Winchester, S. Shang, Z.-K. Liu, D. G. Schlom, X. Pan, and L.-Q. Chen, $BiFeO_3$ domain wall energies and structures: A combined experimental and density functional Theory + **U** Study, Phys. Rev. Lett. **110**, 267601 (2013).

[74] F. Xue, Y. Gu, L. Liang, Y. Wang, and L.-Q. Chen, Orientations of low-energy domain walls in perovskites with oxygen octahedral tilts, Phys. Rev. B **90**, 220101 (2014).

[75] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, Optimization by simulated annealing, Science **220**, 671 (1983).

[76] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevB.109.144426 for calculation and optimization of the domain-wall structure.

[77] M. Pinheiro, F. Ge, N. Ferré, P. O. Dral, and M. Barbatti, Choosing the right molecular machine learning potential, Chem. Sci. **12**, 14396 (2021).

[78] https://github.com/mir-group/allegro.

[79] A. Paszke *et al.*, Pytorch An imperative style, high-performance deep learning library, Adv. Neural Inf. Process. Syst. **32**, 8026 (2019).

[80] M. Fey and J. E. Lenssen, Fast Graph Representation Learning with PyTorch Geometric, arXiv:1903.02428.

[81] https://github.com/pyg-team/pytorch_geometric/blob/2.0.4/torch_geometric/nn/models/dimenet.py

[82] https://github.com/lammps/lammps.git.

[83] K. Momma and F. Izumi, VESTA: A Three-dimensional visualization system for electronic and structural analysis, J. Appl. Crystallogr. **41**, 653 (2008).

[84] J. D. Hunter, Matplotlib: A 2D graphics environment, Comput. Sci. Eng. **9**, 90 (2007).

[85] J. L. García-Palacios and F. J. Lázaro, Langevin-dynamics study of the dynamical properties of small magnetic particles, Phys. Rev. B **58**, 14937 (1998).

[86] G. Kresse and J. Hafner, Ab initio molecular dynamics for liquid metals, Phys. Rev. B **47**, 558 (1993).

[87] G. Kresse and J. Hafner, Ab initio molecular-dynamics simulation of the liquid-metal–amorphous-semiconductor transition in germanium, Phys. Rev. B **49**, 14251 (1994).

[88] G. Kresse and J. Furthmüller, Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set, Comput. Mater. Sci. **6**, 15 (1996).

[89] P. E. Blöchl, Projector augmented-wave method, Phys. Rev. B **50**, 17953 (1994).

[90] J. P. Perdew, K. Burke, and M. Ernzerhof, Generalized gradient approximation made simple, Phys. Rev. Lett. **77**, 3865 (1996).

[91] P.-W. Ma and S. L. Dudarev, Constrained density functional for noncollinear magnetism, Phys. Rev. B **91**, 054420 (2015).

[92] S. L. Dudarev, G. A. Botton, S. Y. Savrasov, C. J. Humphreys, and A. P. Sutton, Electron-energy-loss spectra and the structural stability of nickel oxide: An LSDA+U study, Phys. Rev. B **57**, 1505 (1998).

[93] M. Cococcioni and S. de Gironcoli, Linear response approach to the calculation of the effective interaction parameters in the LDA + U method, Phys. Rev. B **71**, 035105 (2005).

[94] D. P. Kingma and J. Ba, Adam: A method for stochastic optimization, arXiv:1412.6980.

[95] S. Elfwing, E. Uchibe, and K. Doya, Sigmoid-weighted linear units for neural network function approximation in reinforcement learning, Neural Netw. **107**, 3 (2018).