Spectral neighbor representation for vector fields: Machine learning potentials including spin

M. Domina, M. Cobelli, and S. Sanvito
School of Physics and CRANN Institute, Trinity College, Dublin, Ireland

(Received 23 February 2022; revised 17 May 2022; accepted 10 June 2022; published 30 June 2022)

We introduce a translational and rotational invariant local representation for vector fields, which can be employed in the construction of machine learning energy models of solids and molecules. This allows us to describe, on the same footing, the energy fluctuations due to the atomic motion, the longitudinal and transverse excitations of the vector field, and their mutual interplay. The formalism can then be applied to physical systems where the total energy is determined by a vector density, as in the case of magnetism. Our representation is constructed over the power spectrum of the combined angular momentum describing the local atomic positions and the vector field, and it can be used in conjunction with different machine learning schemes and data taken from accurate *ab initio* electronic structure theories. We demonstrate the descriptive power of our representation for a range of classical spin Hamiltonian and machine learning algorithms. In particular, we construct energy models based on both linear Ridge regression, as in conventional spectral neighbor analysis potentials, and the Gaussian approximation. These are both built to represent a Heisenberg-type Hamiltonian including a longitudinal energy term and spin-lattice coupling.

DOI: 10.1103/PhysRevB.105.214439

I. INTRODUCTION

The modeling of the structural, electronic, and magnetic properties of materials at finite temperature requires the exploration of complex energy surfaces, a task that is usually performed in the configuration space or through extensive time-dependent simulations. The gold standard is set by *ab initio* methods, where one solves directly an electronic problem, for instance through density functional theory. The accuracy of the method is then given by the accuracy of the underline electronic structure theory, which often can be determined by the level of approximation taken. Notably, also the computational overheads are set by the electronic structure theory, so that there is always a tradeoff between the accuracy of the prediction, the duration of the simulation, and the maximum size of the system to simulate. Real-time time-dependent simulations of purely electronic quantities, such as the dynamics of spins [1–3], are limited to a few atoms and a few hundreds of femtoseconds, while *ab initio* molecular dynamics (MD) simulations can reach well within the picosecond range and may involve several hundreds of atoms.

The general strategy for extending the range of dynamics simulations across both the time and length scale is to abandon completely the *ab initio* description and replace the solution of the electronic problem with some parametric functions, constructed to reproduce the *ab initio* potential energy surface, namely classical force fields [4]. In their most canonical form, one describes the interaction among the ions by introducing energy contributions that account for the various physical forces at play (covalent bond, dispersive forces, etc.). Once the force field is defined, only the atomic positions determine the total energy. A similar approach has been recently introduced for spin-dynamics. In this case, one associates with each atom a classical spin vector, $\mathbf{S}_i$, so that the total energy is defined over a continuous vector field with values at the atomic positions [5,6]. The total energy then takes the form of a classical Heisenberg model and may include both anisotropy and friction terms. Furthermore, the formalism can be extended to include spin-lattice coupling [7], longitudinal spin fluctuations [8], and possibly the effects of a spin current [9].

In general, force fields constructed in this way have two main drawbacks. On the one hand, their accuracy is significantly inferior to that of an *ab initio* electronic structure theory, although this varies depending on the class of compounds one wants to study. On the other hand, they tend to be specific to the particular type of bond they describe. In addition, spin-type force fields may not be able to describe entire excitations types. For instance, magnetic Stoner excitations are not part of the spectrum of a classical Heisenberg model.

Recently, a new class of force fields, named machine learning force fields (MLFFs), have been shown to solve both the accuracy and specificity issues. The general philosophy of MLFFs is quite different from that of their classical counterparts, since one does not pretend to construct an energy function of the atomic coordinates with terms bearing a physical interpretation, but instead tries to reproduce extremely accurately the *ab initio* potential energy surface. MLFFs comprise of two parts: an abstract representation of the atom density distribution [10], and a machine learning model that correlates such representation to the system's total energy. For the two parts to work together, the representation should be translational, atom-permutational, and rotational invariant. The first two conditions are usually met by local representations (also known as atomic-neighbor descriptions), where the

energy is expressed in terms of atomic contributions, while the last condition must be explicitly enforced in the descriptor's construction. Symmetry functions [11] and bispectra [12] are two examples of such rotational invariant local representations.

Importantly, the spin degrees of freedom are not explicitly included in the representation, which typically describes an atomic density only. For this reason, MLFFs are currently unable to describe the energy difference between inequivalent magnetic phases, for instance between a ferromagnetic and an antiferromagnetic ground state, unless the different phases are also associated with different structures. This limitation can be addressed by combining the MLFF with a classic spin Hamiltonian in order to create a model that can predict energies dependent on both the atomic positions and the spin order. This route was recently explored by Nikolov and co-workers [13], who equipped a spectral neighbor analysis potential (SNAP) with a classical Heisenberg Hamiltonian to compute thermodynamic properties (e.g., the Curie temperature) of α-Fe.

Another possible strategy is to make the MLFF aware of the spin configuration by including such information in the input features of the model. This is a nontrivial task, since such features need to retain the aforementioned symmetries in order for the model to perform well. A recent attempt along these lines consists in the introduction of a novel definition of symmetry functions carrying spin information [14]. Such reformulation was constructed for the spin-collinear case and for fixed spin magnitude (no longitudinal spin information is available). Similarly, the atomic-cluster expansion method was recently extended to vector fields [15] in a way that enables the description of noncollinear spin configurations. Notably, by design both methods require a large number of features for an accurate description of the magnetic environment, so that their training needs large datasets. Thus, a compact representation describing a vector field, which is hence able to compute atomic and magnetic excitations on the same footing, remains elusive. Our paper aims to fill that gap.

Here, we propose a local representation for vector fields, which can be used with either linear and nonlinear machine learning models. This is based on the power spectrum of the combined angular momentum describing the local atomic positions and the vector field. The representation is rotationally invariant and can be further generalized to tensorial densities. To test its descriptive power, such a representation is combined with either linear Ridge regression or the Gaussian approximation to construct MLFFs describing the potential energy surface of a Heisenberg model with longitudinal fluctuations and spin-lattice coupling. Our results show that extremely accurate energy predictions can be obtained with a rather moderate number of training data.

II. METHODS

A. Density for a vector field

The starting point of any atomic-neighbor description of solids and molecules consists in defining the local particle density associated with the ith atom,

$$\rho_i(\mathbf{r}) = \sum_{r_{ai}<r_{\rm cut}} w_a h_a(\boldsymbol{r}-\boldsymbol{r}_{ai}). \tag{1}$$

Here, $\boldsymbol{r}_{ai} = \boldsymbol{r}_a - \boldsymbol{r}_i$ is the distance between the atoms at the position $\boldsymbol{r}_a$ and $\boldsymbol{r}_i$, $r_{ai} = |\boldsymbol{r}_{ai}|$, so that the coordinates of the ith atom define the origin of the local reference frame. The sum in Eq. (1) runs over all the atoms inside a sphere of radius $r_{\rm cut}$ with the center at $\boldsymbol{r}_i$, while w_a are weights usually associated with the atomic species of the ath atom. In this expansion, h_a is a localization function, such as a Gaussian or a Dirac δ, centered at $\boldsymbol{r}_a$, whose specific shape, in general, can depend on the ath atom type. Atomic-neighbor descriptions are then constructed by defining rotationally invariant combinations of the coefficients of expansion of h_a over an appropriate local basis [10].

In the same spirit, we can now define a local *vector* density, $\boldsymbol{\rho}(\mathbf{r})$, through Eq. (1) by associating a vector $\boldsymbol{v}_a$ with each position $\boldsymbol{r}_a$, namely

$$\boldsymbol{\rho}(\mathbf{r}) = \sum_a w_a h_a(\boldsymbol{r}-\boldsymbol{r}_a)\boldsymbol{v}_a, \tag{2}$$

where for simplicity we have dropped the index i. In this formulation, $\boldsymbol{v}_a$ may, for instance, represent the local moment of the ions in a magnetic compound, so that $\boldsymbol{\rho}(\mathbf{r})$ describes the local magnetization field. In Eq. (2), the vector $\boldsymbol{v}_a$ is defined through its Cartesian components,

$$\boldsymbol{v}_a = v_{a,x}\hat{\boldsymbol{e}}_x + v_{a,y}\hat{\boldsymbol{e}}_y + v_{a,z}\hat{\boldsymbol{e}}_z = \sum_{i=x,y,z} v_{a,i}\hat{\boldsymbol{e}}_i, \tag{3}$$

with $\hat{\boldsymbol{e}}_i$ being the unit vector along $i = x, y, z$. However, it is convenient to replace the decomposition of Eq. (3) with one using the spherical versors [16],

$$\hat{\boldsymbol{e}}_{\pm1} = \mp\frac{1}{\sqrt{2}}(\hat{\boldsymbol{e}}_x \pm i\hat{\boldsymbol{e}}_y) \quad \text{and} \quad \hat{\boldsymbol{e}}_0 = \hat{\boldsymbol{e}}_z, \tag{4}$$

so $\boldsymbol{v}_a$ becomes

$$\boldsymbol{v}_a = \sum_{q=0,\pm1} v_{a,q}\hat{\boldsymbol{e}}_q \quad \text{with} \quad \begin{cases} v_{a,\pm1} = \mp\frac{1}{\sqrt{2}}(v_{a,x} \mp i v_{a,y}), \\ v_{a,0} = v_{a,z}. \end{cases} \tag{5}$$

This decomposition is a particular case of the more general one for a tensor of order λ in its irreducible spherical components. Therefore, the spherical components $v_{a,q}$ transform under rotations as the spherical harmonic Y_1^q [16,17].

To construct covariant descriptors for the local vector density of Eq. (2), one first needs to expand the spatial part, h_a, over an orthonormal radial basis. Here we use the product between a radial basis, $R_{nl}(r)$, and the three-dimensional spherical harmonics, $Y_l^m(\hat{\boldsymbol{r}})$, where as usual n, l, and m are, respectively, the principal, the angular momentum, and the third component of the angular momentum quantum numbers. The local vector density then becomes

$$\boldsymbol{\rho}(\boldsymbol{r}) \simeq \sum_{n=0}^{n_{\rm max}} \sum_{l=0}^{n} \sum_{m=-l}^{l} \sum_{q=0,\pm1} c_{nlmq} R_{nl}(r) Y_l^m(\hat{\boldsymbol{r}})\hat{\boldsymbol{e}}_q, \tag{6}$$

where the equality holds only for a complete basis but not for the one truncated at $n_{\rm max}$, and where the coefficients of

expansion are calculated as

$$c_{nlmq} = \sum_a w_a v_{a,q} \int dr\, d\hat{\boldsymbol{r}}\, r^2 R_{nl}(r) Y_l^{m*}(\hat{\boldsymbol{r}}) h_a(\boldsymbol{r} - \boldsymbol{r}_a). \quad (7)$$

In this work, we choose the radial-basis set introduced for the spherical-Bessel descriptors [18], namely $R_{nl}(r) = g_{n-l,l}(r)$. These are orthonormal on the sphere and smoothly vanish at the cutoff radius. Note that the choice of the radial basis set is not unique or crucial, and an alternative basis can be selected. The only practical criterion is that they should approximate completeness with a relative small number of basis functions, namely the convergence must be rapid. Note also that, in what follows, we will always assume orthonormality within the radial basis set, although the expressions derived could be easily generalized to the nonorthogonal case.

The use of the spherical components of a tensor has already been exploited in the construction of covariant kernels for vectorial and tensorial properties related to an atomic environment [19–21]. Here, we follow the same basic idea and formalism, which stems from the decomposition of a tensorial object into spherical components. The main difference is that we cannot just associate a tensorial object with our density, since the density itself is the vectorial field. As such, we will avail ourselves of the same concepts and methods that are typically used to describe the coupling of angular momenta and vector fields in atomic physics [16]. Another difference with existing literature is that our primary target is the construction of invariant quantities instead of covariant ones. A similar goal was already pursued in Ref. [15], where a vector field was described using the same strategy employed in dealing with the atomic positions. More explicitly, in [15] the magnetic vectors are encoded in Dirac's δ distributions, which are then expanded on a suitable basis set and coupled with the analogous expansion arising from the positions of the atoms. Here, however, we will preserve the vectorial nature of the field at each step in the derivation, a strategy that results in a simpler coupling scheme, as will be shown in detail below. With this in mind, we will first proceed with deriving an invariant power spectrum for the vector density of Eq. (2), and then we will implement a linear regression to fit such a power spectrum. This second step is similar to what is commonly done with the formalism of the spectral neighbor analysis potentials (SNAPs) [22].

For the remainder of the paper, we will develop the formalism by using a conventional Dirac notation, which allows one to appreciate better the structure of the representation. In fact, as shown in Ref. [23], the Dirac notation gives us a natural tool for dealing with local atomic densities. In this way, the expansion of the vector density defined in Eq. (6) can be written in a compact form as

$$|\boldsymbol{\rho}\rangle = \sum_{nlmq} c_{nlmq} |nlmq\rangle, \quad (8)$$

where

$$\langle \boldsymbol{r}|nlm\rangle = g_{n-l,l}(r) Y_l^m(\hat{\boldsymbol{r}}) \quad \text{and} \quad |q\rangle \equiv \hat{\boldsymbol{e}}_q. \quad (9)$$

We can then express all the relevant quantities over the the basis, $|nl1JM\rangle$, of the combined angular momenta, $\boldsymbol{L} + \boldsymbol{1} = \boldsymbol{J}$, by using the standard addition scheme [17],

$$|nl1JM\rangle = \sum_{m=-l}^{l} \sum_{q=-1}^{1} C_{lm1q}^{JM} |nlmq\rangle, \quad (10)$$

where C_{lm1q}^{JM} are the Clebsch-Gordan coefficients. As usual, J and M, are the quantum numbers for the total angular momentum and its projection, while "1" refers to the angular momentum of **1**, describing the vector nature of the field. Hence, we have

$$|l-1| \leqslant J \leqslant l+1 \quad \text{and} \quad -J \leqslant M \leqslant J. \quad (11)$$

The states $|nlJM\rangle$, when projected over the position representation, are the products of *vector* spherical harmonics and radial functions. By inverting Eq. (10),

$$|nlmq\rangle = \sum_{JM} C_{lm1q}^{JM} |nlJM\rangle, \quad (12)$$

we can write the vector density as

$$|\boldsymbol{\rho}\rangle = \sum_{nlJM} u_{nlJM} |nlJM\rangle \quad (13)$$

with

$$u_{nlJM} = \langle nlJM|\boldsymbol{\rho}\rangle = \sum_{mq} C_{lm1q}^{JM} \langle nlmq|\boldsymbol{\rho}\rangle = \sum_{mq} C_{lm1q}^{JM} c_{nlmq}. \quad (14)$$

The last equality follows from the fact that the Clebsch-Gordan coefficients are real and from the orthonormality of the $g_{n-l,l}$ functions. The form of $|\boldsymbol{\rho}\rangle$ given by Eq. (13) contains the expansion of the vector density over the combined angular-momenta basis. It should be noted that the presence of the Clebsch-Gordan coefficients imposes that the values of J and M must satisfy the conditions (11). The Clebsch-Gordan coefficients impose also that the nonzero terms in the double sum of (14) are such that $M = m + q$.

B. Invariant power spectrum for a vector field

In this section, we are going to introduce an invariant power spectrum for the density given in Eq. (13). Let us initially restrict our formulation to the case in which we ignore the atom at the origin of the local reference frame, i.e., we assume that the magnitude of the vector field is zero at the origin. One way to obtain the power spectrum, p_{nlJ}, is through the construction of the following inner product:

$$\langle \boldsymbol{\rho}|\boldsymbol{\rho}\rangle = \sum_{nlJ} p_{nlJ}, \quad (15)$$

which explicitly reads

$$p_{nlJ} = \sum_M |u_{nlJM}|^2, \quad (16)$$

where we have used the orthogonality of the $|nlJM\rangle$ basis. Since the vectors $|nlJM\rangle$ correspond to the coupled angular momenta, they transform under system rotation, $\hat{R}$, as the spherical harmonics Y_J^M, namely

$$\hat{R} Y_J^M = \sum_{M'} D_{M'M}^J(\hat{R}) Y_J^{M'}, \quad (17)$$

where $D_{MM'}^J(\hat{R})$ is the Wigner D-matrix associated with the rotation $\hat{R}$. It must be noted that, when one considers the

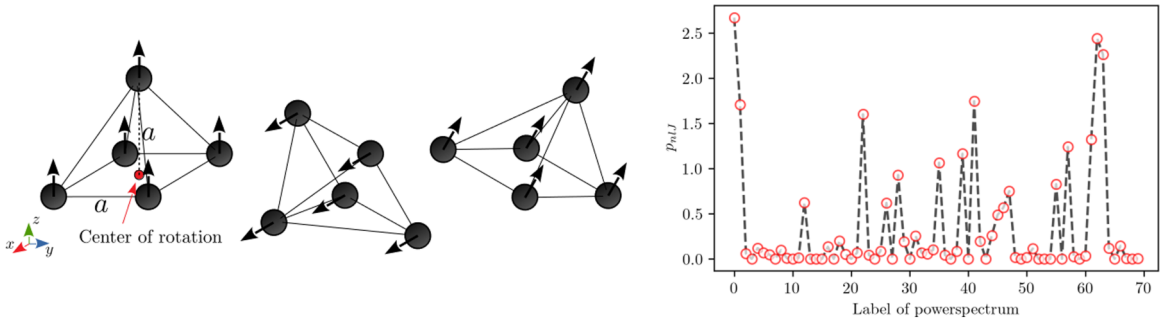

FIG. 1. The invariant power spectrum (right-hand side panel) for a pyramid-shaped set of atoms, where a vector field is associated with each atom (left-hand side image). The power spectrum is evaluated with respect to the shown center of rotation (red dot) and it is invariant against the simultaneous rotation of the system and the vectorial field around that center. We chose the value $n_{\max} = 6$ for this example. The number of independent power-spectrum components is given by $(n_{\max}+1)(3n_{\max}+2)/2 = 70$. The x-axis labels the power-spectrum components, while the y-axis shows their actual value.

original angular momentum basis, the rotation $\hat{R}$ appears as a simultaneous rotation of both the positions of the atoms and the vector field. By applying this rotation to the density in Eq. (13), we obtain

$$\begin{aligned}\hat{R}|\boldsymbol{\rho}\rangle &= \sum_{nlJM} u_{nlJM}\hat{R}|nlJM\rangle \\ &= \sum_{nlJM} u_{nlJM}\sum_{M'} D^J_{M'M}(\hat{R})|nlJM'\rangle \\ &= \sum_{nlJM'} u'_{nlJM'}|nlJM'\rangle, \end{aligned} \tag{18}$$

from which we can infer the transformation rule for the expansion coefficients,

$$\hat{R} : u_{nlJM} \to \sum_{M'} u_{nlJM'} D^J_{MM'}(\hat{R}). \tag{19}$$

Therefore, under rotation $\hat{R}$, the power spectrum, $p_{nlJ} = \sum_M u^*_{nlJM}u_{nlJM}$, transforms as

$$\begin{aligned}\hat{R} : p_{nlJ} &\to \sum_{M'M''} u^*_{nlJM'}u_{nlJM''}\underbrace{\sum_M \left(D^J_{MM'}\right)^* D^J_{MM''}}_{=\delta_{M'M''}} \\ &= \sum_{M'} u^*_{nlJM'}u_{nlJM'} = p_{nlJ}, \end{aligned} \tag{20}$$

where we have used the unitarity of the Wigner D-matrices [note that we have shortened the notation $D^J_{MM'}(\hat{R})$ into $D^J_{MM'}$]. This proves that the power spectrum obtained from Eq. (16) is rotationally invariant for simultaneous rotations of the atomic positions and the vector field. Such invariance is shown numerically in Fig. 1, where the power spectra computed for different rotations are shown to overlap perfectly. In the Appendixes, we will briefly discuss the generalized power spectrum connecting different radial channels. Furthermore, we will also extend our construction to the more general case of a tensorial density.

If we now take a localization function of the form

$$h_a(\boldsymbol{r}-\boldsymbol{r}_i) = \delta(\boldsymbol{r}-\boldsymbol{r}_a), \tag{21}$$

namely a Dirac-δ function centered on the ath atom, then the local vector density reads

$$\boldsymbol{\rho}(\boldsymbol{r}) = \sum_a w_a \delta(\boldsymbol{r}-\boldsymbol{r}_a)\boldsymbol{v}_a. \tag{22}$$

In this case, the expansion coefficients of Eq. (14) are readily evaluated by using Eq. (7) as

$$u_{nlJM} = \sum_a w_a g_{n-l,l}(r_a)\sum_{mq} C^{JM}_{lm1q} Y^{m*}_l(\hat{\boldsymbol{r}}_a)v_{a,q}. \tag{23}$$

In what follows, we will use this expression to explicitly evaluate the power spectrum.

If we now consider a vector field having a vector-bearing atom at the origin, it can be proven (see the Appendixes) that the power spectrum is not generally invariant under rotations. We can interpret this rotational-symmetry breaking by noting that a vector field at the origin introduces an inner preferential direction for the local reference frame. One pragmatic solution to recover the invariance is to always rotate the system so that the vector field at the origin points along the z-axis. After such alignment, we obtain a power spectrum, which is invariant under rotations around the z-axis. Another possible solution is to choose a suitable radial basis set so that all the noninvariant terms are automatically removed. As shown in the Appendixes, we proved that the spherical Bessel functions have this property. In the following, we will always consider power spectra with a central atom. As an example of its explicit evaluation, it is useful to obtain the complete expression for the p_{n0J} power spectra. The component u_{n0JM} is proportional to

$$u_{n0JM} \propto \delta_{J1}\sum_a w_a g_{n0}(r_a)v_{a,M}, \tag{24}$$

where we used the equalities $q = M$ and $J = 1$ enforced through the Clebsh-Gordan coefficients C^{JM}_{0010}, and we did not carry over the spherical harmonics values and the Clebsh-Gordan coefficients, which, in this case, are unessential constants. The power spectrum is then proportional to

$$\begin{aligned} p_{n0J} &\propto \delta_{J1}\sum_{ab} w_a w_b g_{n0}(r_a)g_{n0}(r_b)\sum_M v_{a,M}v^*_{b,M} \\ &\propto \delta_{J1}\sum_{ab} w_a w_b g_{n0}(r_a)g_{n0}(r_b)\boldsymbol{v}_a\cdot\boldsymbol{v}_b. \end{aligned} \tag{25}$$

It is interesting to note that if the vector field is made of local spins, the p_{n0J} power spectrum component will have a structure similar to that of a Heisenberg model, i.e., it depends on the inner product between the spins with distances-only-dependent coefficients.

Having derived a set of invariant power spectra also for the case of an atom at the origin, we can now introduce the models used to test our formalism and the machine learning scheme that implements the power spectrum.

C. Training a machine learning model

The machine learning model used here is a linear regression constructed over the power spectrum $\{p_{nlJ}\}$, following the same philosophy of SNAP [22]. Thus, let us assume to have a system of N atoms, each one of them bearing a local spin. Let us also define the power spectrum vector, $\boldsymbol{p}$, as the one-dimensional vector, whose entries are the p_{nlJ} components describing a specific local neighborhood. Specifically, the ith power spectrum vector $\boldsymbol{p}^{(i)}$ is the vector obtained by centering the local reference frame on the ith atom, and then by evaluating the ith power spectrum set $\{p^{(i)}_{nlJ}\}$ with respect to that frame. Thus, given our N-atoms system, we obtain N power spectrum vectors, $\boldsymbol{p}^{(i)}$.

Our main working hypothesis is that the energy, or any other quantity that we wish to represent, can be written as the sum of short-ranged contributions, $\varepsilon_i(\boldsymbol{q}^i)$, located on each atom [12], namely

$$E = \sum_i \varepsilon_i(\boldsymbol{q}^i), \tag{26}$$

where $\boldsymbol{q}^i$ is a vector describing the local environment of the ith atom. Then, following the same idea behind the SNAP [22], we further assume that the power spectrum vectors $\boldsymbol{p}^{(i)}$ form a suitable set of descriptors to represent such decomposition, so that the local energies can be expressed as a linear combination of power spectrum vectors,

$$E \simeq \boldsymbol{\theta} \cdot \sum_i \boldsymbol{p}^{(i)} = \sum_{nlJ} \theta_{nlJ} \sum_i p^{(i)}_{nlJ}. \tag{27}$$

Here $\{\theta_{nlJ}\}$ is an appropriate set of weights. The validity of these assumptions cannot be determined from the outset and must be tested on a case-by-case base. Within our formalism, the power spectrum vectors can be seen as the descriptors of a linear regression problem, where the target is the energy of the system.

In what follows, we will first calculate the energies of several atomic and spin configurations obtained by displacing the position of the atoms and the direction and magnitude of the magnetic vectors. Then, we will evaluate the power spectrum vectors for each atom and for each of the configurations considered. Finally, we will train a Ridge regression and optimize the weights vector, $\boldsymbol{\theta}$, to predict the total energy. This will allow us to investigate the descriptive power of our vectorial representation and of the full method proposed. In the next section, we describe the different models investigated.

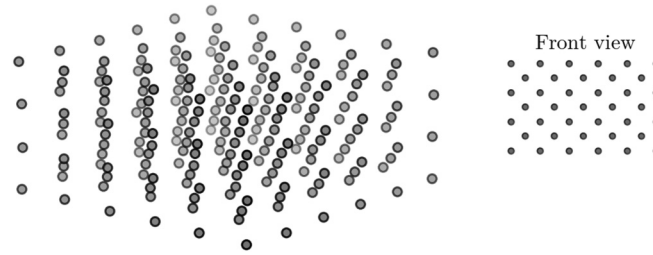

FIG. 2. The physical system investigated in the present work: a tetrahedral cluster of bcc iron. The system is made of seven stacked square-shaped layers of five or six atoms per side, as shown in the inset. The total number of atoms is 219.

D. The physical system investigated

In this work, we consider a rectangular cluster of sites arranged over a bcc lattice, containing 219 atoms of the same species. The cell is rectangular with a six-atom-wide square base and no periodic boundary conditions, as shown in Fig. 2. Each atom bears a local spin and can be displaced from the ideal high-symmetry bcc site. The training data, namely the atomic and spin configurations and their associated total energies, may come from a suitable total energy theory. This is usually some *ab initio* method such as spin-polarized density functional theory or a quantum-chemistry wave-function scheme. Since the generation of such a dataset is rather time consuming, and our objective here is simply that of introducing our vector field representation, we use instead a range of analytical energy models.

In particular, we assume that the total energy is determined by the Hamiltonian

$$H = H_{\mathrm{H}} + H_{\mathrm{L}}, \tag{28}$$

where

$$H_{\mathrm{H}} = -\frac{1}{2} \sum_{\langle i,j \rangle} J_{ij}(r_{ij}) \mathbf{S}_i \cdot \mathbf{S}_j,$$

$$H_{\mathrm{L}} = \sum_i \left(A S_i^2 + B S_i^4 + C S_i^6 \right). \tag{29}$$

Here, H_{H} describes a Heisenberg model, where the exchange parameter between the pair of atoms $\langle i, j \rangle$, bearing spin $\mathbf{S}_i$ and $\mathbf{S}_j$, depends on the atom distance r_{ij}. Note that the spin vectors are in units of $\hbar$, so that $S_i = M_i / g_e \mu_{\mathrm{B}}$, with M_i being the ith local magnetic moment and μ_{B} the Bohr magneton. In particular, in this work we choose the following functional form [24] for J_{ij}:

$$J_{ij}(r_{ij}) = J_n (1 - \Delta r_{ij}/r_n)^3 \quad \text{with} \quad \Delta r_{ij} = r_{ij} - r_n, \tag{30}$$

where the index n indicates that the atoms i and j form an nth-neighbor pair. The distance r_n is that between two nth-neighbor atoms in the undistorted bcc lattice (the nth-neighbor equilibrium distance). Similarly, the constants J_n are the Heisenberg coupling elements between two nth neighbors at equilibrium. It should be noted that H_{H} describes coupling between the position and the spin degrees of freedom by means of the coupling constants, $J_{ij}(r_{ij})$. The Hamiltonian is then completed by a Landau-like term, H_{L}, which describes the dependence of the energy on the longitudinal local

magnetization (the magnitude of local spins) [8], where A, B, and C are constants to be determined.

In this work, we set the various parameters to describe bcc iron [8]. Thus, the Heisenberg exchange interaction extends to second nearest neighbors with $J_1 = 22.52$ meV and $J_2 = 17.99$ meV, while the Landau parameters are chosen to be $A = -440.987$ meV, $B = 150.546$ meV, and $C = 50.769$ meV. We will now proceed to show how our descriptors are able to capture the potential energy surface of the Hamiltonian of Eq. (28).

III. NUMERICAL SIMULATIONS

The Hamiltonian given in Eq. (28) consists of two qualitatively different terms, H_{H} and H_{L}. The first describes transverse energy excitations and spin-lattice coupling, while the second accounts for longitudinal excitations. To investigate the descriptive power of the power spectrum and of our linear energy model over these two different types of excitation, we first consider only the Heisenberg term with fixed magnetic momenta lengths, $M_i = 2.2\mu_{\mathrm{B}}$, and later the full model.

A. Representing the Heisenberg model with spin-lattice interaction

The dataset has been built by displacing the atomic positions from the ideal equilibrium bcc structure and by choosing different orientations of the local magnetic moments. For the atomic positions, we have chosen three sets presenting a different maximum displacement of 5%, 10%, and 20% of the lattice constant, respectively. The sampling of the displacements is uniform in space. In contrast, we have used two different strategies to define the spin structure. In the first one, we align the majority of the spins along the z-axis, while the remaining magnetic moments point in a random direction. More specifically, out of the available 219 magnetic moments, we randomly choose always more than 200 spins (the actual number is between 200 and 219, and it also randomly selected), to be aligned along the z-axis, thus forming an almost ferromagnetic structure (this training set is called the "ferromagnetic" one). The second strategy, instead, consists in assigning to all the magnetic moments a random orientation ("random" training set). Considering the three different choices for the maximum atomic displacement, and the two for the spin alignment, we have thus built a total of six datasets, each consisting of 100 configurations.

To test the predictions made by our fitted spin potentials, we build three additional test sets for each of the six datasets explored. The first set consists of 219 configurations, with the nth configuration having n randomly chosen spins aligned along the z-axis, while the remaining ones are randomly oriented in space. In this case, the lattice is chosen to be pristine bcc (no atomic displacements) so as to test independently the vectorial character of the potential. In contrast, the second and the third test sets are designed to investigate also the atomic displacements. They consist of 50 configurations each, and the atomic displacement has the same maximum magnitude as that of the dataset used to train the model. The spin configuration of the second test set has 200 randomly chosen spins aligned along the z-axis while the remaining ones are randomly oriented in space. The third test set, instead, has all the spins randomly oriented in space. As such, these sets have been designed to test the predictions in a mostly ferromagnetic environment and in a paramagnetic one, respectively.

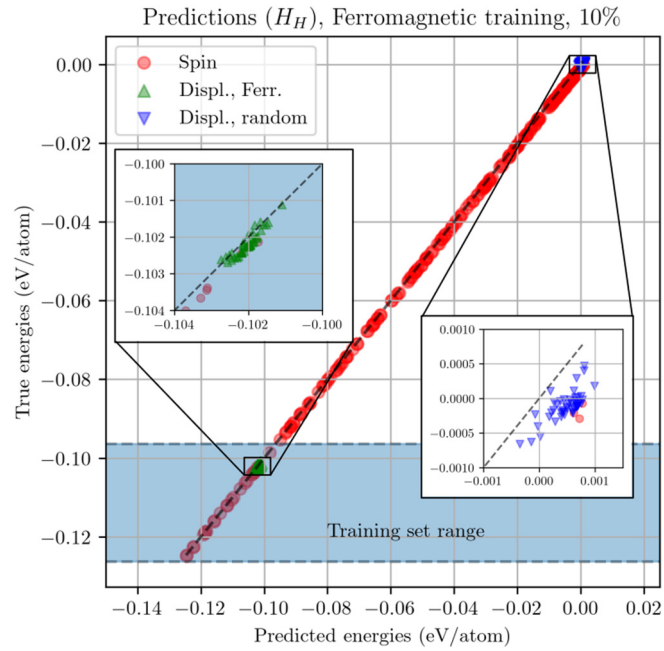

FIG. 3. Predicted against actual energies for a Ridge linear regression trained on the ferromagnetic dataset with 10% maximum atomic displacement. The actual energies are from the Heisenberg model with spin-lattice coupling. The results are tested against three different test sets. The red dots represent the configurations with undistorted bcc atomic positions and progressively z-aligned magnetic moments. The other dots represent different displacements of the atoms for near to ferromagnetic and paramagnetic configurations, respectively. The figure demonstrates the good agreement of the potential also for configurations, which are energetically far from that used in the training (blue region). Zoom-ins around different energy regions are displayed in the insets.

In Fig. 3 we show the results for the ferromagnetic-trained potential with a 10% maximum atomic displacement, for which we explicitly report the procedure and the results. The results on the other training sets are reported in the supplemental material [25]. The optimal potential parameters are found to be $n_{\mathrm{max}} = 4$, corresponding to 35 features only, and $r_{\mathrm{cut}} = 1.4$ (lattice units), while the regularization constant of the Ridge regression is $\alpha = 3.2 \times 10^3$. In the cross-validation procedure, we split the dataset in training and test sets five times with an 80:20 ratio with respect to the total dataset. We obtain an energy mean absolute error (MAE) of $(4.83 \pm 0.15) \times 10^{-5}$ eV/at on the training set, and of $(6.8 \pm 0.7) \times 10^{-5}$ eV/at on the test one, which roughly corresponds to an error smaller than 0.1%. When looking more specifically at the model predictions, the analysis on the first test set (red points in Fig. 3) returns us a MAE of 5.6×10^{-4} eV/at (∼1%). Notably we find that the MAE of configurations with an energy above −0.01 eV/at, namely those that are further away from the energy range of the training set, is 8.2×10^{-4} eV/at. This means that the prediction of the

model is still effective also in the portion of the configuration space far from that of the training. The MAEs on the second (green points in Fig. 3) and third test sets (blue points in the figure) are, respectively, 1.9×10^{-4} and 6.0×10^{-4} eV/at. This suggests that the model can extrapolate rather well across the configuration space.

The remaining trained models show us MAEs similar to that reported above, when the training is performed over the ferromagnetic datasets. In contrast, when training on the three training sets denoted as "random" spin configurations, we notice a significantly larger MAE for large atomic displacements. Explicitly, the MAE is 1.3×10^{-4} eV/at for the random training set with 5% maximum atomic displacement, but this already increases to 2.5×10^{-3} eV/at for 10% maximum displacement and reaches up to 0.01 eV/at for 20% maximum displacement. The failure for the largest maximum displacement can be attributed to the fact that the random dataset explores a much smaller portion of the energy landscape of the model. This is because the random configurations are all characterized by a small total magnetization and hence rather similar energies. For all the other cases, the good agreement obtained between our model and the true potential energy surface demonstrates that the Heisenberg model, including spin-lattice coupling, is accurately described by our potential, which is able to extrapolate the entire energy landscape. It is crucial to remark at this point that we did not introduce any prior knowledge of the functional dependence of the coupling constant on the pairwise distance between the atoms, i.e., the model and the descriptors are able to autonomously interpolate the spin-phonon coupling. Moreover, given the modest size of the dataset, we found it quite remarkable that when using a small dataset and a reduced number of features simultaneously, we are able to reach the accuracies reported here.

Next we will consider the full Hamiltonian, including H_{L}, describing both transverse and longitudinal spin excitations.

B. Heisenberg model with longitudinal excitations

The investigation of the model described by the complete Hamiltonian of Eq. (28) follows the same approach used for the analysis on the Heisenberg part. In this case, we build a dataset corresponding only to one maximum displacement of the atomic positions, namely 10% of the lattice constant. The spin configurations correspond to the "ferromagnetic" case described in the previous section. However, having to deal with longitudinal excitations as well, we also vary the magnetic moment's length. In particular, the magnetic moments aligned along the z-axis are chosen to be $2.25\mu_{\mathbf{B}}$, while the randomly oriented ones have a length randomly chosen in the range 1.8–$2.3\mu_{\mathbf{B}}$.

When testing the predictions, we build an additional test set, corresponding to the first one presented in the previous section, namely containing an increasing number of aligning spins. In this case, the length of the z-aligned magnetic moments is again fixed to $2.25\mu_{\mathbf{B}}$, while the randomly oriented ones have a length in the range 1.9–$2.3\mu_{\mathbf{B}}$. Also the cross-validation procedure is similar to the one employed before with a five-time split of the dataset into training and test sets,

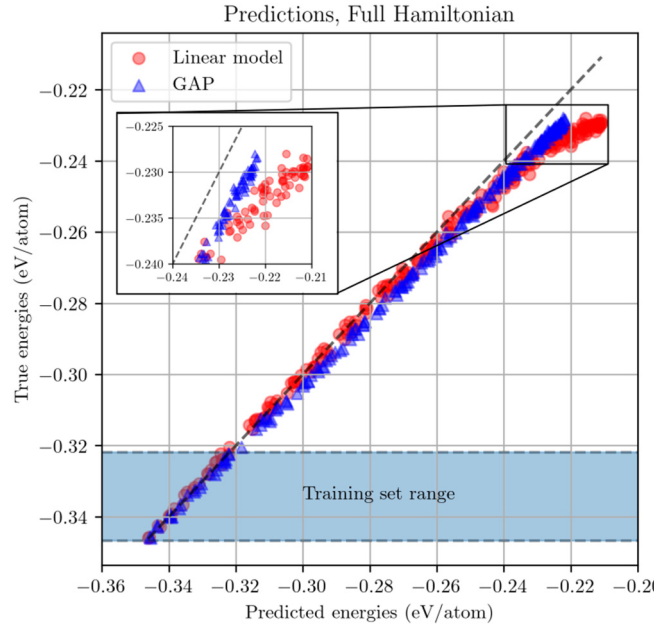

FIG. 4. Predicted against actual energies for both a Ridge linear regression (red circles) and a GAP (blue triangles) constructed to predict the potential energy surface of the full Hamiltonian of Eq. (28) (Heisenberg model with spin-lattice coupling and longitudinal spin fluctuations). Results are presented for the test set. While the accuracy of the two models is similar for energies in the training set region, this becomes significantly different for the high-energy data points. In particular, the linear model accurately predicts energies up to about −0.26 eV/at, but then significantly deviates from the parity line (dashed black line). In contrast, the GAP, while it appears to slightly underestimate the actual energies, outperforms the linear model at extrapolating away from the training set region.

with an 80:20 ratio. The parameters chosen are then $n_{\max} = 4$, $r_{\mathrm{cut}} = 1.4$ (lattice constant), $\alpha = 2 \times 10^5$.

The MAEs obtained in this case are $(4.9 \pm 0.1) \times 10^{-4}$ and $(6.0 \pm 0.5) \times 10^{-4}$ eV/at, respectively, for the training and test set. These values are about one order of magnitude larger than those obtained previously for the Heisenberg model with spin-lattice coupling. We can understand such accuracy loss by noticing that the descriptors are quadratic in the spin magnitude, as is evident from Eqs. (16), (23), and (25). Therefore, a linear machine learning model, such as that employed here, will not be able to capture the energy contributions to the fourth and sixth power in the magnetization, which defines the longitudinal part of the Hamiltonian, H_{L}. In fact, it may be surprising that the model still performs accurately even in this case. This is because we are exploring a region of the potential energy surface relatively close to the minimum, where the energy contributions in S^4 and S^6 remain modest.

To corroborate this hypothesis, we evaluate the model predictions on a test set containing progressively aligned spins, for which we obtain a MAE of 6.0×10^{-3} eV/at. These results are shown as red dots in Fig. 4, where it is clear that the Ridge regression performs poorly as we progressively explore energy regions away from the training range. Such behavior must be associated with the limit of the machine learning

linear model constructed over our description. In more detail, we find that the low-energy regions are still well described, with an MAE of 7.1×10^{-4} eV/at for energies less than -0.32 eV/at (in the training range). In contrast, the potential rapidly departs from the parity line at higher energies, where we compute an MAE of 1.0×10^{-2} eV/at for data above -0.26 eV/at.

We can improve on the error and go beyond the quadratic nature of our descriptors by combining the power-spectrum representation of the atomic and vector field with a nonlinear machine learning model. In particular, we consider here a Gaussian approximation potential (GAP) [26]. The GAP expresses the atomic energy of the ith atom as

$$\varepsilon_i = \sum_{t=0}^{N_{\text{train}}} \theta_t S(\boldsymbol{p}^i, \boldsymbol{p}^t) = \sum_t \theta_t e^{-\frac{1}{2\sigma}(\boldsymbol{p}^i - \boldsymbol{p}^t)^2}, \tag{31}$$

where the sum is extended over all the atoms in the training set, and where $\boldsymbol{p}^i$ and $\boldsymbol{p}^t$ are the power spectrum, respectively, of the ith atom and of the training set. The nonlinearity of the similarity kernel, S, allows us to describe energy contributions going beyond the quadratic order in the spin magnitude.

The GAP predictions obtained over the test set are shown in Fig. 4 as blue triangles. We notice that the MAE associated with the configurations having energy smaller than -0.32 eV/at remains very close to that obtained with Ridge regression. However, the total MAE decreases to 4.5×10^{-4} eV/at, and, most importantly, the MAE for energies larger than -0.26 eV/at is now reduced to 5.8×10^{-3} eV/at, i.e., it is halved. In fact, the figure clearly shows that the nonlinear GAP improves the ability of the model to extrapolate away from the training set range. Since the actual potential energy surface for a spin system, such as the one obtainable from density functional theory, is expected to include energy contributions going beyond a quadratic dependence on the magnetization, we conclude that the best use of our representation will be in conjunction with nonlinear machine learning models.

IV. CONCLUSION

In this work, we have introduced an invariant power spectrum representation for vectorial fields. After having presented an in-depth analysis of its rotational invariance and basic properties, we have designed a linear energy model, closely following the SNAP [22] approach. Such spin SNAP was then put to the test against a bcc iron model described by a Hamiltonian containing spin-lattice coupling and both transverse and longitudinal energy excitations. Spin-lattice coupling is introduced by mean of a Heisenberg model with exchange parameters depending on the interatomic distance, while the longitudinal spin excitations are described by a simple Landau term containing even powers of the magnetization.

We then trained a first potential, linear in the power spectrum, for the situation in which the longitudinal spin excitations are neglected. This was trained over a dataset obtained by displacing the atomic positions and the orientations of the atomic-magnetic moments, comprising a total of only 100 different configurations. Our results showed that the power spectrum is able to describe the entire energy surface by accurately extrapolating far beyond the energy range covered by the training set. This proves that a linear model using the power spectrum is sufficient to describe both the Heisenberg model and the spin-lattice coupling, already from a small dataset. Crucially, no prior information on the dependence of the exchange constants on the atomic position was used by the model.

We then repeated the exercise for the complete Hamiltonian, containing also the Landau term, by training over a dataset containing spins of different magnitude. Our results are highly accurate for configurations with energies within the range explored by our training set, but the model does not perform well in extrapolating. We have attributed this result to the inability of the power spectrum, combined with a linear machine learning model, to describe energy contributions scaling beyond a quadratic dependence on the spin magnitude. Such a shortcoming can be overcome by employing a nonlinear model. Thus, we have investigated a Gaussian approximation potential and shown that extrapolation over a much larger portion of the energy landscape is indeed possible.

All in all, our analysis has shown that a power spectrum representation of the magnetization field can be used, together with nonlinear machine learning models, as an efficient descriptor of spin potential energy surfaces. This can now be used in conjunction with training sets obtained from accurate electronic structure theory to predict finite-temperature properties of magnets.

ACKNOWLEDGMENTS

This work has been supported by the Irish Research Council Advanced Laureate Award (IRCLA/2019/127), and by the Irish Research Council postgraduate program (M.C.). We acknowledge the DJEI/DES/SFI/HEA Irish Centre for High-End Computing (ICHEC) and Trinity Centre for High Performance Computing (TCHPC) for the provision of computational resources.

APPENDIX A: ATOM-CENTERED POWER SPECTRUM

Let us prove that, if the vector field has an atom at the origin, the power spectrum of (16) will not be generally rotationally invariant. For simplicity, we can consider first the trivial case in which our system is made of just a single atom, e.g., there are no other atoms inside the cutoff radius. The component u_{n100} of the coefficients of Eq. (23) reads

$$\begin{aligned} u_{n100} &= w_0 R_{n1}(0) \sum_{mq} Y_1^{m*}(\mathbf{0}) C^{00}_{1m1q} v_q \\ &\propto R_{n1}(0) \sum_q C^{00}_{101q} v_q \\ &\propto R_{n1}(0) v_0, \end{aligned} \tag{A1}$$

where in the first step we have used the fact that the spherical harmonics along the z-axis vanish unless $m = 0$, and in the second step we have considered the equality $q = M - m$ implemented through the Clebsch-Gordan coefficients (for simplicity, we do not carry over an unessential constant). The

power spectrum component for this case then is simply

$$p_{n10} \propto |R_{n1}(0)v_0|^2. \tag{A2}$$

Equation (A2) establishes that the $n10$ component of the power spectrum is proportional to the magnitude of the z component of the vector field at the origin. Crucially, for a general rotation of the reference frame, the z component becomes mixed with the other components, so that it changes its value. We then deduce that the power spectrum is not rotational invariant if the vector field does not vanish at the origin of the local reference frame. It is worth stressing that this proof holds, since the origin is a fixed point for the rotation. If the center or rotation is not the origin, then also the argument of the spherical harmonics will rotate, making the $m \neq 0$ terms relevant, too. This proof can be generalized also to the case in which there are other atoms within the cutoff radius. As pointed out in the main text, however, we can recover a cylindrical symmetry by simply rotating the system so that the vector fields in the origin point in the same direction as the z-axis. Another possible solution stems from the presence of the radial-basis set in Eq. (A2), which, if carefully chosen, could remove the symmetry-breaking terms of the power spectrum. Spherical Bessel functions are suitable for this purpose, as we will now show. We first notice that for $l \neq 0$ the function vanishes at the origin, namely $g_{n-l,l}(0) = 0$. This is a consequence of the fact that the basis set is defined in terms of the spherical-Bessel functions $j_l(x)$, which also vanish at the origin for nonzero l. Thus, for $l \neq 0$ all the contributions arising from an atom at the origin are removed, ensuring the rotational invariance. We are then left to prove that the power spectrum is invariant also for $l = 0$. This is easily done by noticing that the spherical harmonic Y_0^0 is just a constant, with no angular dependence, completing the demonstration.

APPENDIX B: POWER SPECTRUM CONNECTING DIFFERENT RADIAL CHANNELS

Following the same approach used in Ref. [12], we can generalize the expression for the power spectrum so that different radial channels are coupled. A generalized expression for the power spectrum may read

$$p_{nn'lJ} = \sum_M u^*_{nlJM} u_{n'lJM}. \tag{B1}$$

The rotational invariance is still ensured by the fact that the transformation rules for the expansion coefficients u_{nlJM} involve only the Wigner D-matrix belonging to the angular momentum J [see Eq. (18)]. However, it is apparent that the number of components defining the descriptors are increased with this coupling choice. Note that we must ensure the same l for the two factors in the sum above, so that the resulting quantity is real: the only complex part is found in the angular-dependent terms since the radial functions are real.

APPENDIX C: GENERALIZATION TO A TENSORIAL DENSITY

In this Appendix, we generalize our formalism from a vectorial field to a tensorial one. In this case, the local density reads

$$\boldsymbol{\rho}(\mathbf{r}) = \sum_a w_a h_a(\mathbf{r} - \mathbf{r}_a) \boldsymbol{T}^{n_a}_a, \tag{C1}$$

where $\boldsymbol{T}^{n_a}_a$ is an n_a-rank tensor associated with the ath atom. Note that tensors associated with different atoms can have different ranks, as, for example, when dealing simultaneously with a scalar field and a vector field. We can then use the spherical decomposition of the tensors and write $\boldsymbol{T}^{n_a}_a$ in Dirac notation as

$$\left|\boldsymbol{T}^{n_a}_a\right\rangle = \sum_\lambda^{n_a} \sum_{\mu=-\lambda}^{\lambda} (T_a)^\mu_\lambda |\lambda\mu\rangle, \tag{C2}$$

where $(T_a)^\mu_\lambda$ is the spherical component of the tensor $\boldsymbol{T}^{n_a}_a$ relative to the spherical basis $|\lambda\mu\rangle$. This is analogous to $v_{a,q}$ and $|q\rangle$ for the case of a vectorial field. The components $(T_a)^\mu_\lambda$ transform as the spherical harmonic Y^μ_λ under rotation. We can then write the density as

$$|\boldsymbol{\rho}\rangle = \sum_{n=0}^{n_{\max}} \sum_{l=0}^{n} \sum_{m=-l}^{l} \sum_{\lambda=0}^{\Lambda} \sum_{\mu=-\lambda}^{\lambda} c_{nlm\lambda\mu} |nlm\lambda\mu\rangle, \tag{C3}$$

where the expansion has been truncated at $n_{\max}$, and with $\Lambda = \max_a(n_a)$ being the highest rank of the tensors in the tensorial field, additional zero coefficients can be introduced to have a homogeneous representation in the highest-tensorial rank. The expansion coefficients are then obtained by projection as

$$c_{nlm\lambda\mu} = \sum_a w_a (T_a)^\mu_\lambda \int dr\, d\hat{\mathbf{r}}\, r^2 R_{nl}(r) Y_l^{m*}(\hat{\mathbf{r}}) h_a(\boldsymbol{r} - \boldsymbol{r}_a), \tag{C4}$$

with $(T_a)^\mu_\lambda = 0$ if $\lambda > n_a$. Following the same procedure outlined previously for the case of a vectorial field, we can express the density over the coupled basis as

$$|\boldsymbol{\rho}\rangle = \sum_{nl\lambda JM} u_{nl\lambda JM} |nl\lambda JM\rangle, \tag{C5}$$

by means of the coupling scheme,

$$|nl\lambda JM\rangle = \sum_{m=l}^{l} \sum_{\mu=-\lambda}^{\lambda} C^{JM}_{lm\lambda\mu} |nlm\lambda\mu\rangle, \tag{C6}$$

with $|l - \lambda| \leqslant J \leqslant l + \lambda$. The coupled-basis coefficients are given in terms of the uncoupled ones as

$$u_{nl\lambda JM} = \sum_{m\mu} C^{JM}_{lm\lambda\mu} c_{nlm\lambda\mu}. \tag{C7}$$

Finally, the power spectrum is again given by squaring the coefficients

$$p_{nl\lambda J} = \sum_M |u_{nl\lambda JM}|^2, \tag{C8}$$

and it is invariant under simultaneous rotations of the frame of reference and the tensorial field. A further generalization to multichannel coupling can be obtained by using the same argument presented in Appendix B.

[1] P. Elliott, T. Müller, J. K. D. S. Sharma, and E. K. U. Gross, Ultrafast laser induced local magnetization dynamics in Heusler compounds, Sci. Rep. **6**, 38911 (2016).

[2] J. Simoni, M. Stamenova, and S. Sanvito, Ab initio dynamical exchange interactions in frustrated antiferromagnets, Phys. Rev. B **96**, 054411 (2017).

[3] J. Simoni, M. Stamenova, and S. Sanvito, Ultrafast demagnetizing fields from first principles, Phys. Rev. B **95**, 024412 (2017).

[4] D. Frenkel and B. Smit, *Understanding Molecular Simulation* (Academic, San Diego, 2002).

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

[6] O. Eriksson, A. Bergman, L. Bergqvist, and J. Hellsvik, *Atomistic Spin Dynamics: Foundations and Applications* (Oxford University Press, Oxford, 2017).

[7] P.-W. Ma and S. L. Dudarev, *Atomistic Spin-Lattice Dynamics*, in *Handbook of Materials Modeling*, edited by W. Andreoni and S. Yip (Springer, 2018).

[8] P.-W. Ma and S. L. Dudarev, Longitudinal magnetic fluctuations in Langevin spin dynamics, Phys. Rev. B **86**, 054416 (2012).

[9] M. O. A. Ellis, M. Stamenova, and S. Sanvito, Multiscale modeling of current-induced switching in magnetic tunnel junctions using ab initio spin-transfer torques, Phys. Rev. B **96**, 224410 (2017).

[10] F. Musil, A. Grisafi, A. P. Bartók, C. Ortner, G. Csányi, and M. Ceriotti, Physics-inspired structural representations for molecules and materials, Chem. Rev. **121**, 9759 (2021).

[11] J. Behler and M. Parrinello, Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces, Phys. Rev. Lett. **98**, 146401 (2007).

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

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

[14] 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).

[15] R. Drautz, Atomic cluster expansion of scalar, vectorial, and tensorial properties including magnetism and charge transfer, Phys. Rev. B **102**, 024104 (2020)

[16] M. Weissbluth, *Atoms and Molecules* (Academic, 1978).

[17] D. A. Varshalovich, A. N. Moskalev, and V. K. Khersonskii, *Quantum Theory of Angular Momentum* (World Scientific, Singapore, 1988).

[18] E. Kocer, J. K. Mason, and H. Erturk, Continuous and optimally complete description of chemical environments using spherical Bessel descriptors, AIP Adv. **10**, 015021 (2020).

[19] A. Grisafi, D. M. Wilkins, G. Csányi, and M. Ceriotti, Symmetry-Adapted Machine Learning for Tensorial Properties of Atomistic Systems, Phys. Rev. Lett. **120**, 036002 (2018).

[20] A. Glielmo, P. Sollich, and A. De Vita, Accurate interatomic force fields via machine learning with covariant kernels, Phys. Rev. B **95**, 214302 (2017).

[21] V. H. A. Nguyen and A. Lunghi, Predicting tensorial molecular properties with equivariant machine-learning models, Phys. Rev. B **105**, 165131 (2022).

[22] 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).

[23] M. J. Willatt, F. Musil, and M. Ceriotti, Atom-density representations for machine learning, J. Chem. Phys. **150**, 154110 (2019).

[24] P.-W. Ma, C. H. Woo, and S. L. Dudarev, Large-scale simulation of the spin-lattice dynamics in ferromagnetic iron, Phys. Rev. B **78**, 024434 (2008).

[25] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevB.105.214439. for training parameters and prediction plots regarding the investigated datasets.

[26] 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).