



PAPER • OPEN ACCESS

Constructing realistic effective spin Hamiltonians with machine learning approaches

To cite this article: Xue-Yang Li *et al* 2020 *New J. Phys.* **22** 053036

Recent citations

- [PASP: Property analysis and simulation package for materials](#)
Feng Lou *et al*

View the [article online](#) for updates and enhancements.



PAPER

Constructing realistic effective spin Hamiltonians with machine learning approaches

Xue-Yang Li^{1,2}, Feng Lou^{1,2}, Xin-Gao Gong^{1,2,3} and Hongjun Xiang^{1,2,3,4}¹ Key Laboratory of Computational Physical Sciences (Ministry of Education), State Key Laboratory of Surface Physics, Department of Physics, Fudan University, Shanghai 200433, People's Republic of China² Collaborative Innovation Center of Advanced Microstructures, Nanjing 210093, People's Republic of China³ Shanghai Qizhi Institution, Shanghai, People's Republic of China⁴ Author to whom any correspondence should be addressedE-mail: hxiang@fudan.edu.cn**Keywords:** effective Hamiltonian, interaction terms, magnetic systems, machine learning, variable selectionSupplementary material for this article is available [online](#)

OPEN ACCESS

RECEIVED

3 February 2020

REVISED

26 March 2020

ACCEPTED FOR PUBLICATION

2 April 2020

PUBLISHED

26 May 2020

Original content from
this work may be used
under the terms of the
[Creative Commons
Attribution 4.0 licence](#).

Any further distribution
of this work must
maintain attribution to
the author(s) and the
title of the work, journal
citation and DOI.



Abstract

The effective Hamiltonian method has recently received considerable attention due to its power to deal with finite-temperature problems and large-scale systems. In this work, we put forward a machine learning (ML) approach to generate realistic effective Hamiltonians. In order to find out the important interactions among many possible terms, we propose some new techniques. In particular, we suggest a new criterion to select models with less parameters using a penalty factor instead of the commonly-adopted additional penalty term, and we improve the efficiency of variable selection algorithms by estimating the importance of each possible parameter by its relative uncertainty and the error induced in the parameter reduction. We also employ a testing set and optionally a validation set to help prevent over-fitting problems. To verify the reliability and usefulness of our approach, we take two-dimensional MnO and three-dimensional TbMnO₃ as examples. In the case of TbMnO₃, our approach not only reproduces the known results that the Heisenberg, biquadratic, and ring exchange interactions are the major spin interactions, but also finds out that the next most important spin interactions are three-body fourth-order interactions. In both cases, we obtain effective spin Hamiltonians with high fitting accuracy. These tests suggest that our ML approach is powerful for identifying the effective spin Hamiltonians. Our ML approach is general so that it can be adopted to construct other effective Hamiltonians.

1. Introductions

First-principles calculations using density functional theory (DFT) [1, 2] now become an important tool to study the structural and electronic properties of solid state materials. However, the traditional DFT methods have difficulties in computing finite temperature properties and dealing with large scale systems due to the high computational cost. Besides, a straightforward DFT calculation usually gives a final total result instead of a microscopic mechanism. Therefore, the effective Hamiltonian method, which is not only much more efficient, but also provides clearer physical insight, is widely adopted to deal with thermodynamic and kinetic problems. In practice, the parameters in the effective Hamiltonian models can be obtained by fitting results from *ab initio* calculations. In this way, we can achieve both high accuracy and high efficiency in the same time. This methodology has been applied to investigate many kinds of physical systems [3]. Taking spin systems as an example, the quantum Monte Carlo simulations on the spin model Hamiltonians are adopted to understand quantum many-body effect [4, 5], and the spin molecular dynamics simulations were developed to simulate the spin–lattice–electron interplay in magnetic systems [6–8].

Different kinds of effective Hamiltonians were proposed to deal with different degrees of freedom [9–13]. In this work, we focus on the effective spin Hamiltonians of magnetic systems to illustrate the power of our method. When considering two-body second-order interactions, the full spin Hamiltonian describing a magnetic system with localized electrons can be expressed as $H_{\text{spin}} = \sum_{i<j} J_{ij} \vec{S}_i \cdot \vec{S}_j + \sum_{i<j} \vec{D}_{ij} \cdot (\vec{S}_i \times \vec{S}_j) + \sum_i A_i S_{iz}^2$, which contains symmetric Heisenberg exchange interactions, antisymmetric Dzyaloshinskii–Moriya (DM) interactions, and single-ion anisotropy (SIA). To identify these undetermined parameters, we can adopt the DFT-based four-state method [14–16], which estimates each interaction parameter using the energies of four specially designated spin states. It is one of the simplest and most efficient methods which has been widely applied to many magnetic systems [14–16]. For instance, it was adopted to predict the spin–lattice order of frustrated systems MgCr_2O_4 [15] and determine DM interacting parameters in the multiferroic skyrmion material Cu_2OSeO_3 [16]. An energy-mapping analysis based on noncollinear spin states [17–19] was also developed to estimate interaction parameters. By rotating the spin moments, a series of noncollinear spin states are generated, each accompanied with a different rotation angle. In this case, one can fit the corresponding interaction parameters with the energies of these spin states. Another method for computing Heisenberg interactions is based on magnetic-force linear response theory [20–24], which considers perturbations of the spin ground state.

The full spin Hamiltonians with two-body second-order interactions are popular to simulate localized magnetic systems. However, sometimes high-order interactions are important to describe the physical properties, especially in metallic and narrow gap magnets. For example, the two-body second-order spin model Hamiltonian fails to explain the non-Heisenberg behaviors in TbMnO_3 [17]. To treat this problem, Scaramucci *et al* included the biquadratic and four-spin ring couplings to the effective spin Hamiltonians. They apply both collinear and noncollinear energy-mapping analyses to identify interaction parameters [17]. In particular, the two-body fourth-order biquadratic interactions cannot be treated by collinear energy-mapping analysis, so noncollinear calculations are necessary in this case. Although Scaramucci *et al* succeed in explaining the non-Heisenberg behaviors in TbMnO_3 , it is not a general method to be applied to other non-Heisenberg systems. Besides, some other fourth-order or even higher order interactions may also be important so that an automatic method for determining the effective spin Hamiltonians is desirable. A reliable method should consider all the possible interaction terms and select out the important ones. In this case, applying truncations is always necessary to limit the number of undetermined parameters in effective Hamiltonians. However, even after truncations, usually there are still too many different possible interaction terms to be identified. It is difficult to extract the important interaction terms and eliminate negligible ones. A method to take all the possible terms into account is the spin-cluster expansion (SCE) approach [25, 26]. Singer *et al* adopted this method to study the spin interactions in bcc and fcc Fe [27]. After performing truncations by restricting the interacting distance and interaction orders, they obtained 154 (179) different interaction terms (including Heisenberg interactions and non-Heisenberg interactions) in the case of bcc (fcc) Fe. To fit the effective Hamiltonians, 3954 (835) different spin states are randomly generated. In their method, starting from the model with only a constant term, they try adding each possible interaction term to the temporary model and accept the one that leads to the best fitting performance into the model, thus the number of terms is increased one by one. It was shown that bcc and fcc Fe can both be well described by 20 interaction terms selected out of the 154 and 179 possible terms, respectively. This method of picking out important terms is straightforward but may be not very reliable, since it may wrongly add unnecessary terms to the model (see section 2.4). Therefore, a more reliable and efficient algorithm to extract the important interaction terms is essential, not only in the context of SCE, but also for many other effective Hamiltonian models.

In this work, we develop a machine learning (ML) approach accompanied with multiple linear regression analyses to extract the important interaction terms. In our approach, several machine learning techniques are adopted. In particular, we use a testing set and optionally a validation set in order to reduce the risk of over-fitting [28]. Furthermore, we propose a new penalty factor ‘ λ^p ’ to give preference to the models with less parameters, which is inspired by the L_0 regularization term λp commonly used in machine learning [29]. Compared to the previous method, we try not only adding each term to the model, but also deleting and substituting terms from the temporary model during searching for the important terms. Thus the searching is more thorough and more reliable. We also make use of the uncertainty analyses to roughly estimate the importance of each possible term, which improves the search efficiency significantly. Our methods not only reduce the computational cost, but also produce the effective Hamiltonian model in a more concise form and with a clearer physical picture. What’s more, our approach is generally applicable to other physical systems.

Our article is organized as follows. In section 2, we present the main procedure of our method. First, we propose the general expansion form of our effective spin Hamiltonian model (see section 2.1). Then we simplify the model by truncation and symmetry analyses (see section 2.2). Finally, we apply our ML algorithm to extracting the effective Hamiltonians by eliminating the negligible interaction terms. At the same time, we evaluate the interaction parameters with the multiple linear regression approach (see section 2.3). In section 2.4, we will benchmark our method by considering a simple model function. In section 3, we apply our methods to magnetic materials. We take 2D-MnO (section 3.1) and TbMnO₃ (section 3.2) as examples of 2D and 3D cases, respectively. In section 4, we give a brief summary and give an outlook for our method.

2. Methods

2.1. General form of effective spin Hamiltonians

As we know, effective Hamiltonians can usually be described as linear combinations of some certain basis functions. In this case, the general form of effective Hamiltonians can be written as:

$$H_{\text{eff}} = \sum_{j=1}^{p_{\text{max}}} C_j h_j, \quad (1)$$

where $\{h_j\}$ is the set of basis functions, $\{C_j\}$ is the set of undetermined coefficients of $\{h_j\}$, and p_{max} is the number of all the possible terms $\{C_j h_j\}$.

As for an effective spin Hamiltonian, denoted as H_{spin} , we can expand it as:

$$H_{\text{eff}} = H_{\text{spin}} = H_{00} + \sum_{N=1}^{\infty} \sum_{M=1}^N H_{NM}, \quad (2)$$

where H_{NM} indicates the contributions of M -body N th-order interactions. For example, $H_{00} \equiv E_0$ is the constant term (the energy contribution which is not related to magnetism); and the three-body fourth-order terms are

$$H_{43} = \sum'_{l_1 l_2 l_3} \sum_{i_1, i_2, i_3, i_4=x, y, z} K^{l_1 l_2 l_3; i_1 i_2 i_3 i_4} S_{l_1 i_1} S_{l_1 i_2} S_{l_2 i_3} S_{l_3 i_4}, \quad (3)$$

where \vec{S}_{l_i} represents the spin vector of the l_i th site, with coordinate components $S_{l_i x}$, $S_{l_i y}$, and $S_{l_i z}$. $\sum'_{l_1 l_2 l_3}$ means l_1 , l_2 and l_3 are not equal to each other.

2.2. Simplification by truncations and symmetry analyses

In general, it is necessary to perform truncations in expansion order and interaction distances, which can restrict $\{C_j h_j\}$ terms to a finite number p_{max} . Besides, we can perform symmetry analyses to simplify the Hamiltonians. For example, terms with equal coefficients ($C_j = C_j$) should be combined, while those terms with zero coefficients ($C_j = 0$) should be eliminated.

In the case of magnetic systems, the total energy should be invariant under time inversion ($\vec{S}_k \rightarrow -\vec{S}_k$). Therefore, H_{spin} only contains even order terms, since odd order terms are all zeros. When the spin-orbit coupling (SOC) is negligible, H_{spin} should be invariant under any global spin rotations, which leads to the conclusion that H_{spin} only contains inner product terms of spins. In consequence, the second order terms should only consist of Heisenberg interaction terms $\{K^{kl}(\vec{S}_k \cdot \vec{S}_l)\}$, while fourth order terms comprise $\{K^{klmn}(\vec{S}_k \cdot \vec{S}_l)(\vec{S}_m \cdot \vec{S}_n)\}$. Further, we assume that the norms of the vectors $\{\vec{S}_k\}$ are constant (in other words, we only consider the spin directions as degrees of freedom), so that $(\vec{S}_k \cdot \vec{S}_k)$ is also constant for any k . Thereby, terms like $(\vec{S}_k \cdot \vec{S}_k)$ can be combined into the constant term, while terms like $(\vec{S}_k \cdot \vec{S}_l)(\vec{S}_m \cdot \vec{S}_n)$ can be simplified as $(\vec{S}_k \cdot \vec{S}_l)$ with a constant coefficient. As a result, we should only consider the inner products $\{(\vec{S}_k \cdot \vec{S}_l)\}$ with $k \neq l$. For convenience, the directions of $\{\vec{S}_k\}$ are denoted by unit vectors $\{\vec{e}_k\}$. We can rewrite the terms $\{K^{kl}(\vec{S}_k \cdot \vec{S}_l)\}$ as $\{J^{kl}(\vec{e}_k \cdot \vec{e}_l)\}$ ($k \neq l$), and $\{K^{klmn}(\vec{S}_k \cdot \vec{S}_l)(\vec{S}_m \cdot \vec{S}_n)\}$ as $\{J^{klmn}(\vec{e}_k \cdot \vec{e}_l)(\vec{e}_m \cdot \vec{e}_n)\}$ ($k \neq l, m \neq n$). Considering the symmetry of the lattice, some equal coefficients can be combined. In simple cases, n th nearest pairs k, l share the same coefficient $J^{kl} \equiv J_n$ (or $K^{kl} \equiv K_n$), so that the second order terms can be reduced to $\sum_{n=1}^{n_{\text{max}}} J_n \left(\sum_{\text{nth nearest pairs } k, l} \vec{e}_k \cdot \vec{e}_l \right)$, where n_{max} is determined by the truncation distance.

In our work, we keep the terms with orders up to four. We note that we can go to higher orders if necessary. The simplification of fourth order terms is similar to that of second order terms. The terms

$\{J^{klmn}(\vec{e}_k \cdot \vec{e}_l)(\vec{e}_m \cdot \vec{e}_n)\}$ corresponding to equivalent two-body, three-body or four-body interactions are combined. Suppose that there are n'_{\max} different types of equivalent terms $\{J^{klmn}(\vec{e}_k \cdot \vec{e}_l)(\vec{e}_m \cdot \vec{e}_n)\}$ ($k \neq l, m \neq n$), with the coefficients $\{J^{klmn}\}$ in each type sharing a same value. We denote the coefficient of the n' th type as $J'_{n'}$. Then the effective spin Hamiltonian can be expressed as:

$$H_{\text{eff}} = H_{\text{spin}} = E_0 + \sum_{n=1}^{n_{\max}} J_n \left(\sum_{n\text{th nearest pairs } k,l} \vec{e}_k \cdot \vec{e}_l \right) + \sum_{n'=1}^{n'_{\max}} J'_{n'} \sum_{n'\text{th type}} (\vec{e}_k \cdot \vec{e}_l)(\vec{e}_m \cdot \vec{e}_n), \quad (4)$$

where the basis functions $\{h_j\}$ include $\{1\}$, $\{\sum_{n\text{th nearest pairs } k,l} \vec{e}_k \cdot \vec{e}_l\}$, and $\{\sum_{n'\text{th type}} (\vec{e}_k \cdot \vec{e}_l)(\vec{e}_m \cdot \vec{e}_n)\}$; the undetermined coefficients $\{C_j\}$ include $\{E_0\}$, $\{J_n\}$, and $\{J'_{n'}\}$.

2.3. Extracting the major terms and fitting the parameters

After simplification by truncations and symmetry analyses, the number of undetermined coefficients (p_{\max}) is reduced. If p_{\max} is small, a straightforward fitting (using the linear least squares method, or equivalently multiple linear regression analysis) can be convenient. However, in most cases, p_{\max} is still a large number (usually 100–100 000). As we know, the number of sets of data for fitting must exceed the number of undetermined parameters if the linear least squares method is adopted. To generate such a large quantity of data is usually too time-consuming, or sometimes impossible. Besides, even if we have enough data for fitting, the large number of undetermined parameters may lead to over-fitting. What's more, an effective Hamiltonian with too many parameters is neither convenient for practical use, nor can it deliver a clear physical picture.

As we discussed above, we prefer an accurate model with parameters as few as possible. Our basic idea is to select the important interaction terms among the p_{\max} possible terms and discard the negligible ones. Supposing that we select p terms out of the p_{\max} possible terms $\{C_j h_j\}$, we can renumber the subscripts (j 's) and use $\{b_j\}_{j=1}^p$ to denote the corresponding coefficients $\{C_j\}$. Then, the effective Hamiltonian would be simplified to be:

$$H_{\text{eff}} = \sum_{j=1}^p b_j h_j. \quad (5)$$

After that, it is feasible to fit the p (usually less than 100) coefficients $\{b_j\}$ using the linear least squares method. In consequence, the difficult problem comes down to how to select p important terms out of the p_{\max} possible terms (note that p is also unknown) (which is usually called 'variable selection' [29]). As each of the p_{\max} possible terms can be either selected or not, there are $2^{p_{\max}}$ possible combinations of the terms, which means that using enumerating method to find out the best model is impractical.

In the following subsections, we shall first introduce briefly the multiple linear regression analysis since it is used extensively in our approach. Then the criterion for selecting better models and the methods of selecting the important interaction terms will be discussed.

2.3.1. Multiple linear regression analysis

Suppose H_{eff} possesses the form of equation (5) and that we use n sets of data for fitting the undetermined parameters $\{b_j\}$ in H_{eff} . These data can be acquired by randomly generating the microstates in the configuration space of interest, calculating the $\{h_j^{(i)}\}$ ($i = 1, \dots, n; j = 1, \dots, p$) according to the effective Hamiltonian model, and computing the 'exact value' $H_{\text{eff}}^{(i)}$ ($i = 1, \dots, n$) with *ab initio* methods (or using experimental data). We let $\hat{H}_{\text{eff}}^{(i)}$ denote the 'estimated value' of $H_{\text{eff}}^{(i)}$ predicted with a given model. The error between $H_{\text{eff}}^{(i)}$ and $\hat{H}_{\text{eff}}^{(i)}$, named as 'residual', usually approximately obey a normal distribution with a mean value 0 and a variance σ^2 (which is a common assumption in fitting problems, mainly based on the central limit theorem) [30]. The estimated value of the variance σ^2 (denoted by $\hat{\sigma}^2$) and the definition of the sum of squared residuals (denoted by Q_e) are given by:

$$\hat{\sigma}^2 = \frac{Q_e}{n - p}, \quad (6)$$

$$Q_e \equiv \sum_{i=1}^n \left(H_{\text{eff}}^{(i)} - \hat{H}_{\text{eff}}^{(i)} \right)^2 = \sum_{i=1}^n \left(H_{\text{eff}}^{(i)} - \sum_{j=1}^p b_j h_j^{(i)} \right)^2. \quad (7)$$

The values of the undetermined parameters $\{b_j\}$ can be obtained by minimizing Q_e and the uncertainties of the parameters can also be estimated easily [see part A of SM (<http://stacks.iop.org/NJP/22/053036/mmedia>)] [30].

2.3.2. The criterion for selecting better models

Before discussing the way to construct a good model, we should determine a criterion to judge the performance of models. The fitting variance σ^2 can be a good candidate: the smaller σ^2 is, the better the model performs. There are different approaches of estimating the variance σ^2 of the residual. Besides estimating σ^2 by the data for fitting (i.e., the training set) using equations (6, 7), we can also make use of some other sets of data (i.e., the testing set or the validation set) to estimate σ^2 [28]. For example, if we use the testing set (which contains n' sets of data) to estimate σ^2 , the estimated value of σ^2 (denoted by $\widehat{\sigma'^2}$) can be calculated with the following equations (which are similar to equations (6, 7)):

$$\widehat{\sigma'^2} = \frac{Q'_e}{n'}, \quad (8)$$

$$Q'_e \equiv \sum_{i=1}^{n'} (H_{\text{eff}}^{(i)} - \widehat{H}_{\text{eff}}^{(i)})^2, \quad (9)$$

where the variables related to the testing set are denoted by primes. Similarly, σ^2 can also be estimated by the validation set, and the estimated value will be denoted by $\widehat{\sigma_v^2}$. The training set, the validation set and the testing set are randomly generated from the same configuration space, and independent from each other. The parameters of the model must be determined by the training set without resorting to the validation set and the testing set. The difference between the validation set and the testing set is that, while searching for the best models, we can resort to the validation set (which is optional) to get rid of the models with over-fitting problem; but the testing set cannot be accessed. The testing set is only used to check whether the over-fitting problem occurs in those best candidate models finally suggested by the search and to provide reference for our final selection. The usage of the testing set (and optionally a validation set) is an important technique in machine learning, which can reduce the risk of over-fitting efficaciously.

In the case of effective Hamiltonians, we not only demand fitting accuracy, but also prefer a simpler model, which contains smaller number of undetermined parameters (denoted by p). Because a simpler model usually delivers a clearer physical picture and is less likely to over-fit. To give priority to the models with less parameters, we suggest using a new penalty factor λ^p ($\lambda \geq 1$), thus our criterion is $(\sigma^2 \cdot \lambda^p)$. Accordingly, we would search for the model with the least estimated value of $(\sigma^2 \cdot \lambda^p)$. Our proposed penalty factor λ^p is similar to the usual additional penalty term λp (i.e., the so-called L_0 regularization term) widely adopted in machine learning methods [29]. However, our new penalty factor is preferred since it is relatively easier to choose an appropriate value for λ , in contrast to the usual case where the appropriate magnitude of λ is highly dependent on the magnitudes of the data. More discussions on our criterion and the choices of the values of λ are given in part B of SM.

2.3.3. Methods for determining the important terms

Now that we have determined our criterion for better models, it is possible for us to search for better and better models according to the criterion. In this work, we propose three ML methods for searching for the best models (and determining the important terms). ‘Method 1’ and ‘Method 2’ are our new algorithms, the former using only a training set and a testing set, while the latter using an additional validation set. ‘Method 3’ is similar to the method used by Singer *et al* [27]. The flowcharts of these methods are shown in figure 1.

In our method 1 and method 2, we use a series of decreasing values for λ ($\lambda \geq 1$) (see more details in part B of SM), and for each of these values we search for a best model without using the testing set (details of the search will be discussed in the following paragraphs). Then we can use the testing set to compute the $\widehat{\sigma'^2}$ of these best models (corresponding to different values of λ). If a model with a larger p has a larger $\widehat{\sigma'^2}$, we can tell that over-fitting has occurred, thus the corresponding model must be abandoned. Among these best models, we usually choose the one with the smallest $\widehat{\sigma'^2}$ as our final model; sometimes we may choose another model with larger $\widehat{\sigma'^2}$ but smaller p instead, according to the physical meanings of the terms and the performance of the models. Starting from this final model, we shall decrease p (the number of parameters) one by one until $p = 1$. When decreasing the number of parameters from p to $(p - 1)$, we try the p different choices of which parameter to delete and select the model with the least σ^2 (estimated by the training set or the testing set). Therefore, we know the remaining several parameters are probably the most important ones, so that we can give the rankings of the importance of these parameters (and the corresponding interaction terms).

Now we shall discuss the way to find out the best (or a local best) model according to our criterion $(\sigma^2 \cdot \lambda^p)$ with a certain λ value (in methods 1 and 2). During searching for the best models, the testing set is not available, so we can only estimate the σ^2 with the training set or the validation set. We suppose there are

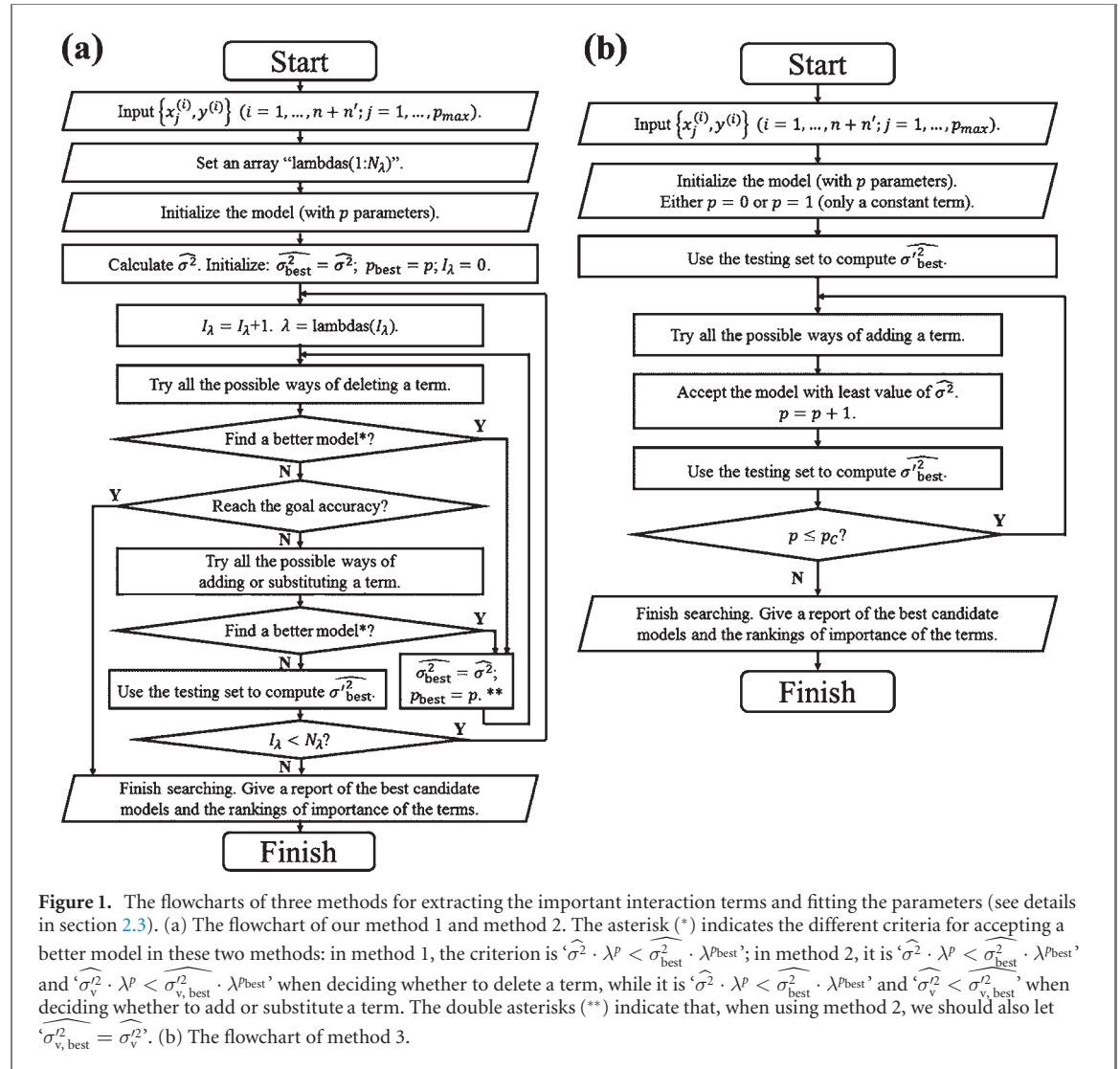


Figure 1. The flowcharts of three methods for extracting the important interaction terms and fitting the parameters (see details in section 2.3). (a) The flowchart of our method 1 and method 2. The asterisk (*) indicates the different criteria for accepting a better model in these two methods: in method 1, the criterion is ' $\hat{\sigma}^2 \cdot \lambda^p < \hat{\sigma}_{best}^2 \cdot \lambda^{p_{best}}$ '; in method 2, it is ' $\hat{\sigma}^2 \cdot \lambda^p < \hat{\sigma}_{best}^2 \cdot \lambda^{p_{best}}$ ' and ' $\hat{\sigma}_v^2 \cdot \lambda^p < \hat{\sigma}_{v,best}^2 \cdot \lambda^{p_{best}}$ ' when deciding whether to delete a term, while it is ' $\hat{\sigma}^2 \cdot \lambda^p < \hat{\sigma}_{best}^2 \cdot \lambda^{p_{best}}$ ' and ' $\hat{\sigma}_v^2 < \hat{\sigma}_{v,best}^2$ ' when deciding whether to add or substitute a term. The double asterisks (**) indicate that, when using method 2, we should also let ' $\hat{\sigma}_{v,best}^2 = \hat{\sigma}_v^2$ '. (b) The flowchart of method 3.

n' sets of data in the testing set, and another n sets of data can be accessed while searching for the best models. These n sets of data can either all be used as the training set ('Method 1'), or be divided into a training set and a validation set, with n_1 and n_2 sets of data separately ($n_1 + n_2 = n$) ('Method 2').

In our method 1, no validation set is used, thus the σ^2 has to be estimated by the training set (denoted as $\hat{\sigma}^2$) during the search. We usually start from the constant function fitting ($p = 1$): $H_{eff} = b_1 h_1$, with $h_1 \equiv 1$. Or we can start from the best model already found with another value of λ . Then we try deleting, adding or substituting a term to see whether we can find a better model (according to our criterion $\hat{\sigma}^2 \cdot \lambda^p$). If so, we accept the better model, and continue trying deleting, adding or substituting a term, until no better model can be found. During the search, we only change one term in a single step, which reduces the searching space significantly, thus enhancing the efficiency of the algorithm. If a more thorough search is needed, trying changing two terms in a single step can also be considered, but it would lead to an obviously lower efficiency. In this way, we search for a best model for each value of λ . If we want to stop the search as soon as the fitting accuracy is high enough, we can set a goal accuracy. When the estimated value of σ^2 is lower than the goal accuracy and no redundant term can be deleted from the temporarily best model, the search can be stopped immediately.

In our method 2, a validation set is used, thus the σ^2 can be estimated either by the training set (denoted as $\hat{\sigma}^2$) or by the validation set (denoted as $\hat{\sigma}_v^2$) during the search. Therefore, we actually have two criteria for better models: $\hat{\sigma}^2 \cdot \lambda^p$ and $\hat{\sigma}_v^2 \cdot \lambda^p$. We suppose that only if a model gives a smaller value of $\hat{\sigma}^2 \cdot \lambda^p$ and a smaller (or the same) value of $\hat{\sigma}_v^2 \cdot \lambda^p$ when compared with the previously best model, can we accept this model as the new best model. Since the requirements for accepting new models are stricter than method 1, this method tends to keep less terms and is more likely to avoid over-fitting. When deciding whether to delete a term, we let $\lambda' = \lambda$. But when deciding whether to add a term, we let $\lambda' = 1$, so that the

requirements for adding a new term to the model are not excessively strict. According to our experience, method 2 is usually less efficient than method 1, but it is better at avoiding over-fitting problems. As for the comprehensive performance of these two methods, either can behave better, depending on the specific cases.

In our methods 1 and 2, there are several techniques for significantly improving the efficiency of the search for best models. For example, by comparing the relative uncertainties of the parameters (see definition in part A of SM) in the temporary model, we can tell that the parameter with the largest relative uncertainty is most likely to be unimportant. Therefore, while trying substituting a term, by default we only consider replacing this most unimportant parameter by another one, which also reduces the searching space significantly. If a more thorough search is needed, more possibilities of substituting a term (such as replacing the second most unimportant parameter by another one) can also be considered, which would also lead to a much lower efficiency. While trying deleting a term, we also start from the term with the largest relative uncertainty. Another technique is that we arrange the terms appropriately so that the terms most likely to be important by prejudgment are earlier to be considered while trying adding a term. In our case, the sequencing of the parameters is based on the orders of the terms and the distances between interacting atoms. Terms with lower orders and nearer atomic distances are considered first. Besides, by analyzing the relative (or absolute) change of $\hat{\sigma}^2$ when adding or deleting a term, we give a rough estimation of the importance of this term. When we try adding or substituting a term, we always try adopting one or several temporarily unused terms with the highest rough estimation of importance in the first place. Some other techniques for speeding up the process are discussed in part C of SM.

The method of extracting the major terms used by Singer *et al* [27] is called ‘Method 3’ in our work. They do not have a testing set to avoid over-fitting. First, they also start with the constant function fitting ($p = 1$) (or we can start with the model $H_{\text{eff}} = 0$ where $p = 0$). Then they try every possible case of adding a term to the temporary model and choose the one with the least $\hat{\sigma}^2$ in the fitting (now $p = 2$). In this way, they continuously add terms one by one. The earlier-added terms are considered more important. By contrast, in our methods (methods 1 and 2), as mentioned earlier, the last several terms to be deleted from our final model are considered the most important ones. In their method, only adding terms are considered, while deleting terms and substituting terms are ignored. Therefore, if a term is incorrectly added, it will never be discarded. As an important improvement to this method, we use a testing set to avoid over-fitting. Despite of this improvement, we shall still call it ‘Method 3’. This method is simple and therefore easy to program, but it may be not very trustworthy since it may wrongly add some unimportant terms.

2.4. Verification of our algorithms

To verify the performance of our algorithms of extracting the important interaction terms, we test our algorithms with a known exact function. We suppose that there are 30 independent variables $a_i (i = 1, \dots, 30)$. Consider the Taylor expansion of these variables, with the orders of terms up to four. In this case, the basis functions $\{h_j\}$ in the model $H_{\text{eff}} = \sum_{j=1}^{p_{\text{max}}} C_j h_j$ include $\{1\}$, $\{a_i\}$, $\{a_i a_j\} (i \leq j)$, $\{a_i a_j a_k\} (i \leq j \leq k)$, and $\{a_i a_j a_k a_l\} (i \leq j \leq k \leq l)$. We find that the total number of terms p_{max} is $C_{34}^4 = 46\,376$.

We set 60 parameters to be nonzero, with 15 parameters belonging to terms with first, second, third, and fourth orders, respectively. The details of these parameters are listed in part D of SM. We generate the independent variables $\{a_k \in [-1, 1]\}$ according to uniform random distribution, and compute the corresponding $H_{\text{eff}} = \sum_{j=1}^{p_{\text{max}}} C_j h_j$. In this way, we generate 1500 sets of data in total, each set of data recording the values of $\{a_k\}$ and H_{eff} . The corresponding $\{h_j\}$ can be easily computed according to their definitions. Thus, we obtain 1500 sets of data $\left\{ \left\{ h_j^{(i)} \right\}, H_{\text{eff}}^{(i)} \right\}_{i=1}^{1500}$. Then we use three methods (see section 2.3) to find out nonzero parameters. We use 500 sets of data as testing set. Therefore, $n = 1000$ and $n' = 500$. In the case of method 2, $n_1 = n_2 = 500$.

In test A, the 60 nonzero parameters have values with different orders of magnitude, so that the difference in their importance is significant. The details of the values of these parameters are listed in part D of SM. All the three methods find the 60 nonzero parameters correctly. Method 1 and method 2 both find the 60 nonzero parameters during the second value of λ ($\lambda = 1.2$). The time costs of these three methods are 105.5 min, 47 min and 590 min, respectively. Our methods (method 1 and method 2) find out the correct model more efficiently. The reason why our methods have better efficiency is that while trying adding terms we do not search all the possibilities. Once we find a term that can give a smaller estimated value of $(\sigma^2 \cdot \lambda^p)$, we accept this term immediately and start the next round of search, which accelerates the search significantly.

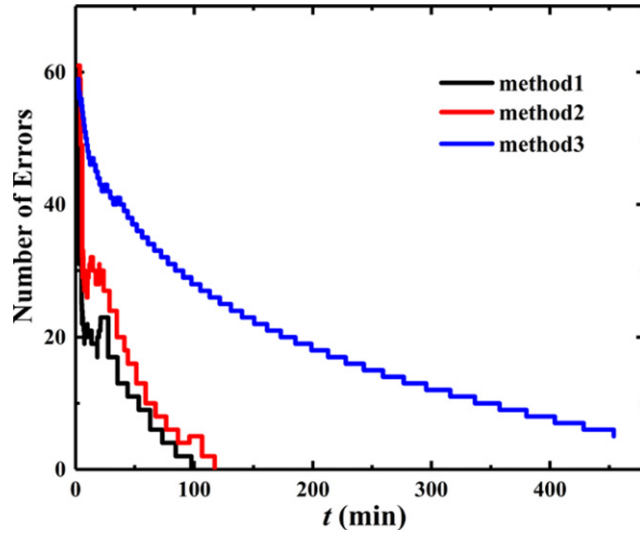


Figure 2. The plot of ‘Number of Errors’ (defined in section 2.4) versus time for three methods of extracting important terms (see section 2.3) in test B (see section 2.4). Method 1 and method 2 are our methods, and method 3 is similar to the method used by Singer *et al* [27].

In test B, we set the same 60 nonzero parameters to be nonzero, but their values are all set to be 1. Method 1 and method 2 both exactly find the 60 nonzero parameters during the sixth value of λ ($\lambda = 1.04$); but method 3 suggests a model with 65 nonzero parameters, 60 of which are correct, but the other five nonzero parameters are redundant. The time costs of these three methods are 98 min, 117 min, and 454 min, respectively. This time, our methods (method 1 and 2) obviously outperform method 3, not only in efficiency, but also in accuracy and reliability. Here we define ‘Number of Errors’ as the number of nonzero parameters not included in the model plus the number of parameters wrongly used in the model. The performances of these three methods in test B are illustrated by the plot of ‘Number of Errors’ versus time, as shown in figure 2.

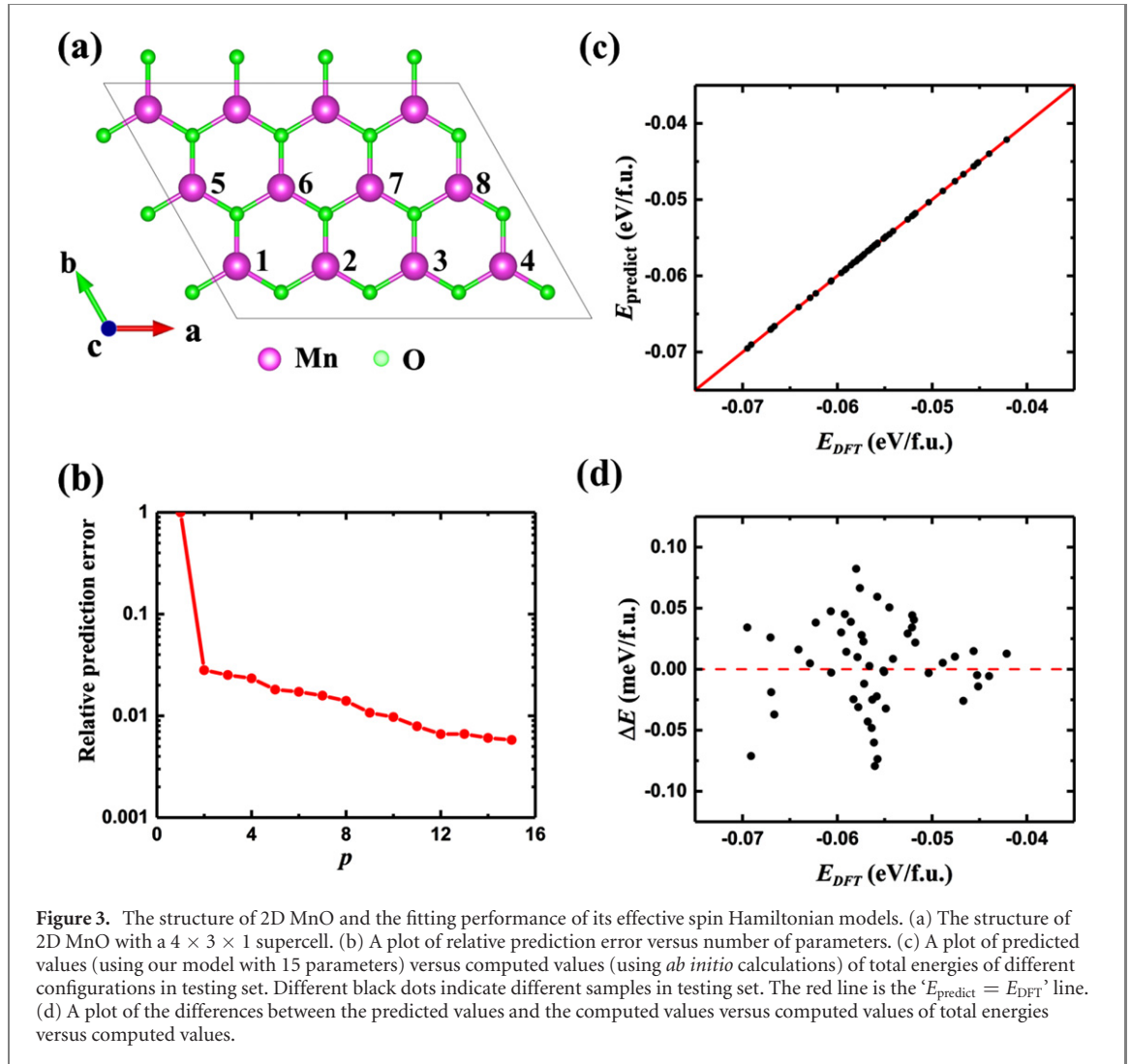
3. Applications

We now apply our approach to two prototypical magnetic systems, i.e., binary magnetic compound 2D MnO and spin-spiral multiferroic material TbMnO₃. The density functional theory (DFT) are performed to calculate the energies on the basis of the projector augmented wave (PAW) [31] method encoded in the Vienna *ab initio* simulation package (VASP) [32]. The exchange–correlation energy is described by the Perdew–Burke–Ernzerhof (PBE) functional of the generalized gradient approximation (GGA) [33]. The strong correlation effects of Mn d electrons are included by means of the DFT + U scheme. Spin–orbit couplings are not included while the directions of the magnetic moments are constrained. In our model, all the parameters have the same unit of eV, since we use dimensionless unit vectors instead of real spin vectors in the $\{x_j\}$. We use J_n to denote the coefficients of Heisenberg interactions between the n th nearest pairs [see equation (4)].

When we calculate the total energy of each configuration of spins, the total energy of FM state is used as a reference energy, and the total energy per primitive cell is used. When extracting the major terms, we try all the three methods [see section 2.3.3], and we find that our method 1 performs best in both cases. To describe the performance of the fitting, we use $\sqrt{\sigma'^2}$, the standard error of prediction. And we normalize it with $\sqrt{\sigma_0'^2}$, which is the value of $\sqrt{\sigma'^2}$ when we use the constant function fitting. We will use ‘relative prediction error’ to refer to the value of $\sqrt{\sigma'^2}/\sqrt{\sigma_0'^2}$.

3.1. 2D MnO

According to density functional calculations, ultrathin films of wurtzite MnO can automatically transform into a stable 2D graphitic structure [34]. The primitive cell of 2D MnO is identical to that of graphene, with Mn and O atoms substituting the two types of sites of C atoms, as shown in figure 3(a). Since this structure is simple and common, we use this predicted 2D binary magnetic compound as an example to test our approach. In our calculations, we adopt a $4 \times 3 \times 1$ supercell which consists of 12 chemical units of MnO. Typical computational parameters are a 400 eV plane-wave energy cutoff, a $2 \times 2 \times 1$ k -point sampling



mesh set, and a $0.01 \text{ eV } \text{\AA}^{-1}$ force tolerance for structural relaxation calculations ($a = 3.459 \text{ \AA}$). We set the on-site Coulomb repulsion and the effective on-site exchange interaction of Mn d states to be $U = 4.5 \text{ eV}$ and $J = 0.5 \text{ eV}$. A Γ -centered $2 \times 2 \times 1$ k -point mesh is applied. The truncation distances of second and fourth order terms are set to 20 Bohr and 18 Bohr (up to fifth nearest pairs and fourth nearest pairs) respectively. After symmetry analyses, the number of possible nonequivalent parameters (p_{max}) is 170 (including a constant term). The training set and the testing set only contain 100 and 50 sets of data, respectively.

By applying our method 1, we pick out the major interaction terms of MnO. We find that important spin interaction terms in MnO include the Heisenberg interactions between the first four nearest pairs, the two-body fourth-order interaction (biquadratic interaction) between the nearest pairs, three-body fourth-order interactions among atoms that constitute a regular triangle, a line, an isosceles obtuse triangle, or a 'y' shape, and four-body fourth-order interactions among atoms that constitute a rhombus. The ranking of importance of these interactions and the fitting result are shown in supplementary table S2 (see SM). The dominating spin interaction is antiferromagnetic (AFM) Heisenberg interaction between the nearest pairs, which is $J_1 = (1.9299 \pm 0.0029) \times 10^{-2} \text{ eV}$ given by our fitting, consistent with the result $J_1 = 1.8 \times 10^{-2} \text{ eV}$ estimated by the four-state method. The AFM Heisenberg interaction between the nearest pairs is consistent with Goodenough–Kanamori (GK) rule which predicts an AFM super-exchange interaction between two half-filled $\text{Mn}^{2+} (3d^5)$ ions [35, 36].

The relation between the relative prediction error (comparing with the results of our first principles calculations) and the number of parameters used in our effective spin Hamiltonian model for 2D MnO is shown in figure 3(b). When there is only one parameter in the model ($p = 1$), the model is the constant function fitting, so the relative prediction error is 1 according to our definition. When the AFM Heisenberg interaction between the nearest pairs is added to the model ($p = 2$), the relative prediction error is reduced to only 2.8%, proving the dominance of this Heisenberg interaction term. Using 15 parameters (including a

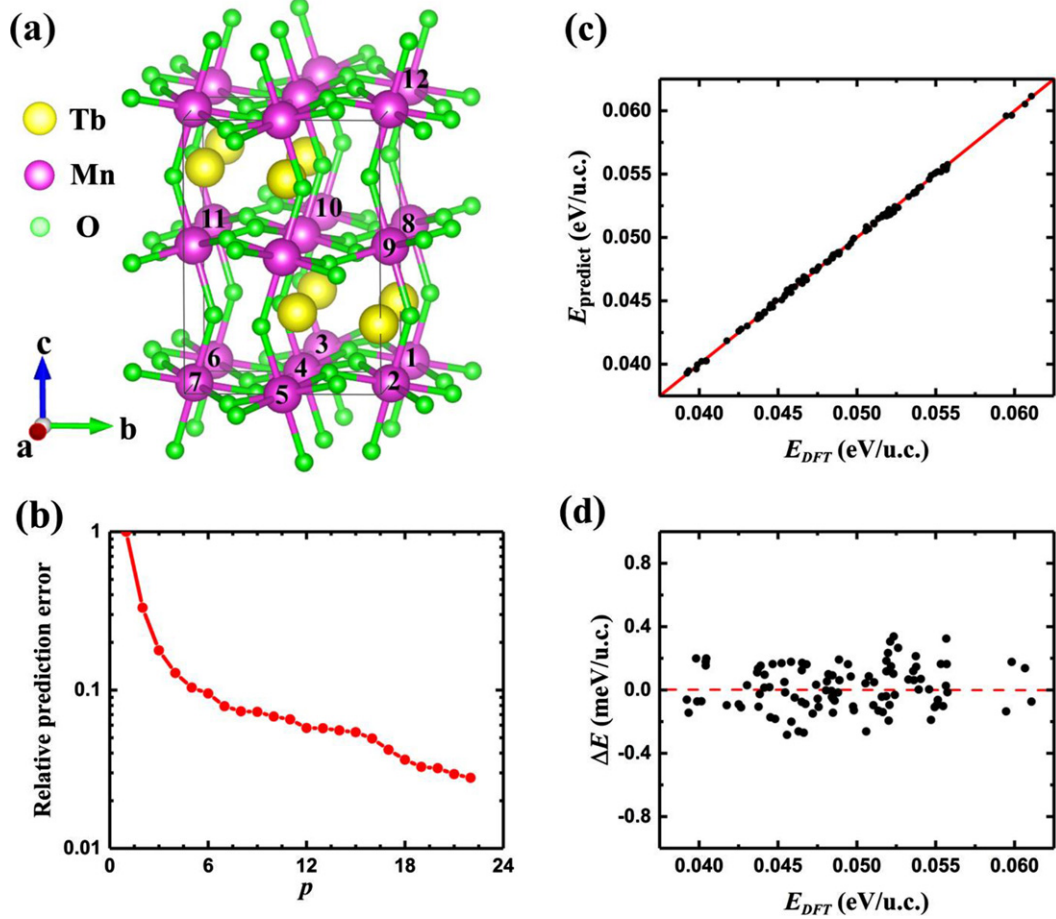


Figure 4. The structure of TbMnO₃ and the fitting performance of its effective spin Hamiltonian models. (a) The structure of the unit cell of TbMnO₃. (b) A plot of relative prediction error versus number of parameters. (c) A plot of predicted values (using our model with 22 parameters) versus computed values (using *ab initio* calculations) of total energies of different configurations in testing set. Different black dots indicate different samples in testing set. The red line is the ' $E_{\text{predict}} = E_{\text{DFT}}$ ' line. (d) A plot of the differences between the predicted values and the computed values of total energies versus computed values.

constant term) in our model, we get a relative prediction error of only 0.58%, which shows very high fitting accuracy. The predicting performance of our suggested model (with 15 parameters) is illustrated in figures 3(c) and (d), which also indicate that our suggested model has very high fitting accuracy with an approximately normally-distributed random error.

3.2. TbMnO₃

Now we apply our method to the prototypical multiferroic TbMnO₃, which displays a spin-spiral magnetic order [37–39]. It is of great significance to investigate the spin interactions in TbMnO₃. Fedorova *et al* find that biquadratic and ring exchange interactions are important in explaining the non-Heisenberg behaviors in orthorhombic perovskite manganites TbMnO₃ [17]. In our work, we want to verify the importance of these interactions, as well as search for other important interactions. The primitive cell of TbMnO₃ has *Pbnm* space symmetry with 20 atoms. The unit cell of TbMnO₃ is illustrated in figure 4(a). In our calculations, we adopt the $2 \times 2 \times 2$ supercell which consists of 32 chemical units of TbMnO₃. The typical calculation parameters are almost the same as those of 2D MnO, except that a Γ -centered $2 \times 2 \times 2$ k -point sampling mesh set is used in this case, and that on-site Coulomb repulsion of $U = 2$ eV and effective on-site exchange interaction of $J = 0$ eV are applied to the Mn d states [17]. The truncation distances of second and fourth order terms are both set to 15 Bohr (up to 11th nearest pairs). After symmetry analyses, the number of possible nonequivalent parameters (p_{max}) is 347 (including a constant term). The training set and the testing set contain 400 and 100 sets of data, respectively.

By applying our method 1, we not only reproduce the known result that the Heisenberg, biquadratic and ring exchange interactions are the important spin interactions in TbMnO₃ [17], but also find out that the next most important spin interactions are three-body fourth-order interactions among the atoms which constitute a line (in the diagonal direction in the xy plane, or along the z axis), or approximately form a right angle (in the xy plane). The ranking of importance of these interactions and our fitting results are

shown in supplementary table S3 (see SM). The most important interaction is the Heisenberg interaction between the second nearest pairs (which are the nearest pairs in the xy plane). The most important non-Heisenberg interaction is the biquadratic interaction between the second nearest pairs, which is ranked after the Heisenberg interactions between the second, tenth and fifth nearest pairs. The comparison between the estimation of the interactions given by our fitting and by Fedorova *et al* is also shown in supplementary table S2. Our results are consistent with theirs.

Using 22 parameters (including a constant term), a relative prediction error of only 2.8% is achieved. The relation between the relative prediction error and the number of parameters is shown in figure 4(b). The predicting performance of our model with 22 parameters is shown in figures 4(c) and (d).

4. Summary and discussions

In this work, we put forward a general approach to find out the effective Hamiltonians in magnetic systems. We provide a practical criterion to judge which models are better and three efficient ML methods for searching for the best model according to our criterion. By applying our approach, we successfully extract the important spin interactions in MnO and TbMnO₃. These two applications suggest that our ML approach is powerful for identifying the effective spin Hamiltonians. Besides, our approach not only reduces the computational cost, but also gives an effective Hamiltonian model with a more concise form and a clearer physical picture.

The performance of our approach depends mainly on the amount of the training data, the accuracy of *ab initio* calculations, and the rationality of the form of effective Hamiltonians. When enlarging the training set, we usually find more important terms with better fitting performance. When we improve the accuracy of *ab initio* calculations, our method usually gives predictions with higher accuracy. As for the rationality of the general form of effective Hamiltonians, we should adopt suitable initial Hamiltonians (usually with an expansion form) when treating different physical models. We note that our approach can be applied to many other physical systems such as alloy, ferroelectric or multiferroic systems. Our work suggests that artificial intelligence will play an important role in computational simulations and design of materials.

Acknowledgments

This work is supported by NSFC 11825403, the Special Funds for Major State Basic Research (Grant No. 2015CB921700), the Program for Professor of Special Appointment (Eastern Scholar), the Qing Nian Ba Jian Program, and the Fok Ying Tung Education Foundation.

References

- [1] Hohenberg P and Kohn W 1964 *Phys. Rev. B* **136** B864
- [2] Kohn W and Sham L J 1965 *Phys. Rev.* **140** 1133
- [3] Picozzi S and Ederer C 2009 *J. Phys.: Condens. Matter* **21** 303201
- [4] Foulkes W M C, Mitas L, Needs R J and Rajagopal G 2001 *Rev. Mod. Phys.* **73** 33
- [5] Sandvik A W and Kurkijarvi J 1991 *Phys. Rev. B* **43** 5950
- [6] Ma P W, Woo C H and Dudarev S L 2008 *Phys. Rev. B* **78** 024434
- [7] Ma P W, Dudarev S L and Woo C H 2012 *Phys. Rev. B* **85** 184301
- [8] Omelyan I P, Mryglod I M and Folk R 2001 *Phys. Rev. Lett.* **86** 898
- [9] Zhong W, Vanderbilt D and Rabe K M 1994 *Phys. Rev. Lett.* **73** 1861
- [10] Liu K, Lu J, Picozzi S, Bellaiche L and Xiang H 2018 *Phys. Rev. Lett.* **121** 027601
- [11] van de Walle A, Asta M and Ceder G 2002 *CALPHAD, Comput. Coupling Phase Diagr. Thermochem.* **26** 539
- [12] Xiang H, Kan E, Wei S-H, Whangbo M-H and Yang J 2009 *Nano Lett.* **9** 4025
- [13] Nelson L J, Ozolins V, Reese C S, Zhou F and Hart G L W 2013 *Phys. Rev. B* **88** 155105
- [14] Xiang H J, Lee C, Koo H J, Gong X G and Whangbo M H 2013 *Dalton Trans.* **42** 823
- [15] Xiang H J, Kan E J, Wei S H, Whangbo M H and Gong X G 2011 *Phys. Rev. B* **84** 224429
- [16] Yang J H, Li Z L, Lu X Z, Whangbo M H, Wei S H, Gong X G and Xiang H J 2012 *Phys. Rev. Lett.* **109** 107203
- [17] Fedorova N S, Ederer C, Spaldin N A and Scaramucci A 2015 *Phys. Rev. B* **91** 165122
- [18] Novak P, Chaplygin I, Seifert G, Gemming S and Laskowski R 2008 *Comput. Mater. Sci.* **44** 79
- [19] Zhu H-F, Cao H-Y, Xie Y, Hou Y-S, Chen S, Xiang H and Gong X-G 2016 *Phys. Rev. B* **93** 024511
- [20] Bruno P 2003 *Phys. Rev. Lett.* **90** 087205
- [21] Wan X G, Yin Q and Savrasov S Y 2006 *Phys. Rev. Lett.* **97** 266403
- [22] Pi S-T, Nanguneri R and Savrasov S 2014 *Phys. Rev. B* **90** 045148
- [23] Sotnikov O M and Mazurenko V V 2016 *Phys. Rev. B* **94** 195107
- [24] Yoon H, Kim T J, Sim J-H, Jang S W, Ozaki T and Han M J 2018 *Phys. Rev. B* **97** 125132
- [25] Drautz R and Fahnle M 2004 *Phys. Rev. B* **69** 104404
- [26] Singer R and Fahnle M 2006 *J. Math. Phys.* **47** 113503
- [27] Singer R, Dietermann F and Fahnle M 2011 *Phys. Rev. Lett.* **107** 017204

- [28] Mitchell T M 1997 *Machine Learning* (New York: McGraw-Hill)
- [29] Chen Y, Du P and Wang Y 2014 *WIREs Comput. Stat.* **6** 1
- [30] Weisberg S 2013 *Applied Linear Regression* 3rd edn (Hoboken, NJ: Wiley)
- [31] Kresse G and Joubert D 1999 *Phys. Rev. B* **59** 1758
- [32] Kresse G and Furthmüller J 1996 *Phys. Rev. B* **54** 11169
- [33] Perdew J P, Burke K and Ernzerhof M 1996 *Phys. Rev. Lett.* **77** 3865
- [34] Kan E, Li M, Hu S, Xiao C, Xiang H and Deng K 2013 *J. Phys. Chem. Lett.* **4** 1120
- [35] Goodenough J B 1955 *Phys. Rev.* **100** 564
- [36] Kanamori J 1959 *J. Phys. Chem. Solids* **10** 87
- [37] Kimura T, Goto T, Shintani H, Ishizaka K, Arima T and Tokura Y 2003 *Nature* **426** 55
- [38] Xiang H J, Wei S H, Whangbo M H and Da Silva J L F 2008 *Phys. Rev. Lett.* **101** 037209
- [39] Malashevich A and Vanderbilt D 2008 *Phys. Rev. Lett.* **101** 037210