PHYSICAL REVIEW MATERIALS **3**, 053807 (2019)

Bayesian optimization of chemical composition: A comprehensive framework and its application to $R\mathrm{Fe}_{12}$-type magnet compounds

Taro Fukazawa,[1,2,*] Yosuke Harashima,[1,2] Zhufeng Hou,[3] and Takashi Miyake[1,2,3]
[1]*CD-FMat, National Institute of Advanced Industrial Science and Technology, Tsukuba, Ibaraki 305-8568, Japan*
[2]*ESICMM, National Institute for Materials Science, Tsukuba, Ibaraki 305-0047, Japan*
[3]*CMI^2, National Institute for Materials Science, Tsukuba, Ibaraki 305-0047, Japan*

(Received 22 March 2019; published 28 May 2019)

We propose a framework for optimization of the chemical composition of multinary compounds with the aid of machine learning. The scheme is based on first-principles calculation using the Korringa-Kohn-Rostoker method and the coherent potential approximation (KKR-CPA). We introduce a method for integrating datasets to reduce systematic errors in a dataset, where the data are corrected using a smaller and more accurate dataset. We apply this method to values of the formation energy calculated by KKR-CPA for nonstoichiometric systems to improve them using a small dataset for stoichiometric systems obtained by the projector-augmented-wave method. We apply our framework to optimization of $R\mathrm{Fe}_{12}$-type magnet compounds $(R_{1-\alpha}Z_\alpha)(\mathrm{Fe}_{1-\beta}\mathrm{Co}_\beta)_{12-\gamma}\mathrm{Ti}_\gamma$, and benchmark the efficiency in determination of the optimal choice of elements (R and Z) and ratio (α, β, and γ) with respect to magnetization, Curie temperature, and formation energy. We find that the optimization efficiency depends on descriptors significantly. The variables β, γ, and the number of electrons from the R and Z elements per cell are important in improving the efficiency. When the descriptor is appropriately chosen, the Bayesian optimization becomes much more efficient than random sampling.

DOI: 10.1103/PhysRevMaterials.3.053807

I. INTRODUCTION

Machine learning is attracting much attention these days, and its application to data obtained by first-principles calculation is a promising way to accelerate the exploration of novel materials. The basic idea is as follows: (i) introduce a numerical representation x for the materials, which is called a descriptor, (ii) calculate a property y for materials from the search space of the descriptor by first-principles calculation, and (iii) infer a relation $y = f(x)$ between x and y from the thus obtained data by modeling f. Many efforts have been made to develop models and descriptors that work in materials discovery [1–9]. These models can be used to identify promising candidates by predicting the property $f(x')$ for unknown materials x'. It is also possible to perform the modeling and the sampling alternately to obtain the optimal x as quickly as possible, which is called optimization.

Bayesian optimization (BO) is a powerful technique to find the maximum (or the minimum) of an unknown function along this idea. It is based on Bayesian modeling using a dataset collected in the previous sampling-modeling iterations, and it does not require an explicit form of the function $y = f(x)$. This method is efficient because it takes account of the uncertainty of a model in addition to the mean value. Figure 1 illustrates a typical situation in which BO is efficient. The dashed line is the true model. Suppose we have four sampled points that are denoted by black circles. By Bayesian modeling, we obtain the mean value (solid line) and the uncertainty (gray region). In this situation, the mean value does not give good prediction for the highest-score point. However, by considering the information of the uncertainty, one can find a significant probability that the true model has the maximum between the two rightmost data points.

BO has been recently applied in various problems in materials science [10–12]. It has also a potential for application to optimization of a chemical composition [13], but there were no reports on quantitative estimation of efficiency in such a problem, avoiding possible overestimation by mere luck to the best of our knowledge. In such applications, we need to properly choose a search space, a descriptor for the candidate systems, and a score to describe the performance that are suitable for the problem. Otherwise, the efficiency of the scheme is deteriorated. A descriptor—a form to which the input data are encoded—is especially crucial [2,5–8,14].

Accuracy of the first-principles calculation is also of great importance in the computer-aided materials search. However, conventional methods are often insufficient to achieve enough accuracy, while sophisticated schemes are too time-consuming for the purpose. For example, the magnetic transition temperature is overestimated in the mean-field approximation. Systematic errors also come in from numerical factors, such as a limited number of basis functions.

It is a promising idea to improve the data by using a smaller dataset from more accurate calculations or experiments [15,16]. This idea is also seen in the notion of transfer learning, which uses referential datasets that are different from the target dataset, and transfers the knowledge from the reference to the target [17,18]. However, there is no method that works for any purposes, and we need to devise a method that is suitable for each of the problems on the basis of knowledge about the origin of the error.

---

[*] taro.fukazawa@aist.go.jp

2475-9953/2019/3(5)/053807(15)

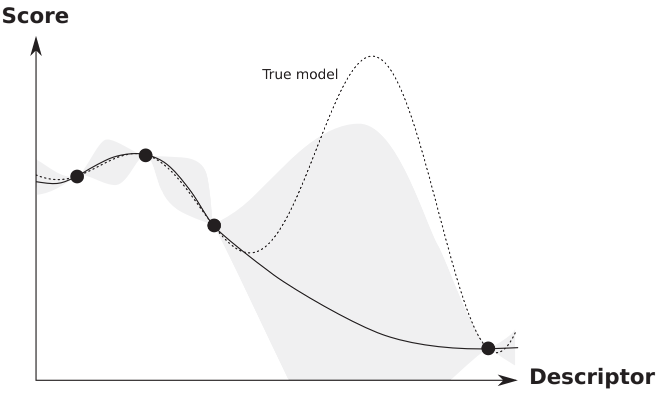

FIG. 1. Schematic illustration of a typical situation in which Bayesian optimization works efficiently. The dashed line represents the true model, the black circles denote sampling points, the solid line shows a model obtained from the sampling points, and the gray region shows a confidence interval of the model.

In this paper, we propose a practical framework for optimizing nonstoichiometric composition of multinary compounds based on Bayesian optimization and first-principles calculation. We perform a benchmark of our scheme and discuss its efficiency in the optimization using a dataset obtained by first-principles calculation with the Korringa-Kohn-Rostoker method and the coherent potential approximation (KKR-CPA) for systems with nonstoichiometric compositions. We investigate the performance of descriptors, and we discuss how we can choose an efficient one for problems. To set up a pragmatic problem, we deal with an issue from the cutting edge of the materials study on hard magnets. We also present a method for correcting systematic errors in the formation energy by using a smaller but more accurate dataset. Our idea is to construct a model of errors on the basis of our understanding of it.

This paper is organized as follows: In Sec. II, we describe our problem setup for the benchmark, providing the background. In the first part of Sec. III, we present a brief summary of the whole framework. We then provide details of Bayesian optimization and the first-principles calculation in the subsequent subsections. Section III C is devoted to the method for integrating datasets that we use to improve the formation energy. We present the results of the benchmark and the data integration in Sec. IV. Finally, we conclude the paper with a summary in Sec. V.

II. PROBLEM SETUP AND ITS BACKGROUND

$R\mathrm{Fe}_{12}$-type compounds having ThMn_{12} structure have been considered as a possible main phase of a strong hard-magnet because they are expected to have high magnetization due to its high Fe content, and to have high magnetocrystalline anisotropy if the R element is properly chosen [19–37]. The magnetic properties of $\mathrm{NdFe}_{12}\mathrm{N}$ were evaluated theoretically a few years ago [38], and its high magnetization and anisotropy field were confirmed by a successful synthesis of $\mathrm{NdFe}_{12}\mathrm{N}_x$ film [39,40].

Unfortunately, $\mathrm{NdFe}_{12}\mathrm{N}$ and its mother compound NdFe_{12} are thermally unstable. They cannot be synthesized as a bulk without substituting another element for a part of the Fe elements. Titanium is typical of such a stabilizing element [41,42]. Introduction of Ti, however, reduces the magnetization significantly. Co also has a potential to work as a stabilizing element according to a prediction by first-principles calculation [43]. Compared to Ti, Co is favorable in terms of magnetization. In fact, a recent experiment on $\mathrm{Sm(Fe}_{0.8}\mathrm{Co}_{0.2})_{12}$ film showed that it has superior saturation magnetization and anisotropy field to $\mathrm{Nd}_2\mathrm{Fe}_{14}\mathrm{B}$, the main phase of the current strongest magnet [44]. Chemical substitution at the R site also affects structural stability. Zirconium has attracted attention as a stabilizing element at the rare-earth site [45–48]. Recent first-principles calculation predicted that Dy also works as a stabilizer [49]. Therefore, optimization of the chemical composition of $R\mathrm{Fe}_{12}$-type compounds in terms of stability and magnetic properties is an important issue for the development of next-generation permanent magnets.

Bearing these in mind, we set $R\mathrm{Fe}_{12}$-type magnet compounds as target systems. In particular, we optimize the chemical formula of $(R_{1-\alpha}Z_\alpha)(\mathrm{Fe}_{1-\beta}\mathrm{Co}_\beta)_{12-\gamma}\mathrm{Ti}_\gamma$ ($R =$ Y, Nd, Sm; $Z =$ Zr, Dy) so that it maximizes magnetization, the Curie temperature, or minimizes the formation energy from the unary systems in the benchmark. Therefore, the problem is a combination of optimization with respect to the compositions (α, β, and γ) and optimization with respect to the choice of elements for R and Z. We discuss the efficiency of the optimization by comparing a number of iterations required in Bayesian optimization with that in random sampling. We also study how the efficiency is affected by the choice of descriptor.

III. METHODS

Figure 2 shows the workflow in our optimization framework. At the beginning of the scheme, the user prepares a list of candidate compounds. The candidates are expressed in the form of a descriptor. In this study, we prepare 11 types of descriptors for the $(R_{1-\alpha}Z_\alpha)(\mathrm{Fe}_{1-\beta}\mathrm{Co}_\beta)_{12-\gamma}\mathrm{Ti}_\gamma$ ($R =$ Y, Nd, Sm; $Z =$ Zr, Dy) systems, which we discuss in Sec. III A.

Then, the candidate list is passed to the optimizer. The role of the optimizer is to pick one system from the candidate list so that a system with a high score is quickly found in the whole scheme. Because it does not have enough data to perform Bayesian optimization at the beginning, it randomly chooses a system from the list. It receives feedback from a scorer later in the scheme, and records it. When the record reaches a certain size, the sampling method is switched to Bayesian optimization. To cover the role of the optimizer, we use a Python module called "COMmon Bayesian Optimization library" (COMBO) [10,50]. We present the parameters used in our benchmark in Sec. III A.

In the next stage, a quantum simulator calculates physical properties for the system chosen, which is the most time-demanding process in the scheme. The details in our simulation are described in Sec. III B.

Then, the scorer integrates the calculated properties to a score. It also improves the estimated values by using the referential data before generating the score. In our application, we use the value of magnetization, Curie temperature, or the

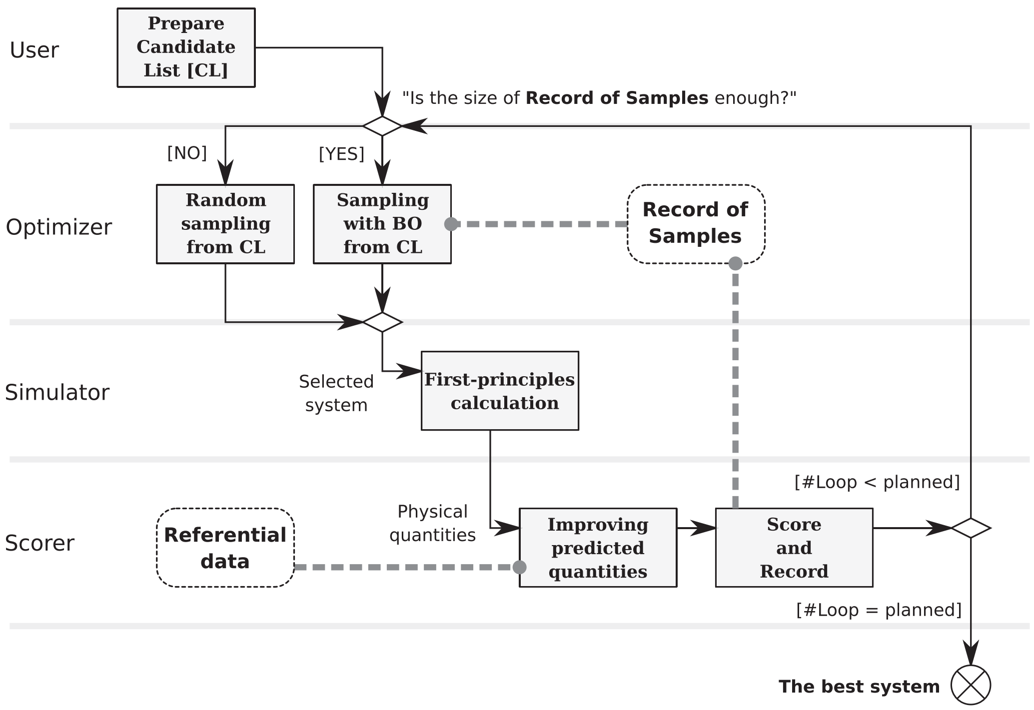

FIG. 2. The workflow of the proposed scheme for materials search. The solid squares denote activities; the dashed squares with round corners represent datasets. BO stands for Bayesian optimization, and CL stands for candidate list.

formation energy as a score. As for the formation energy, we improve it by the method for integrating datasets presented in Sec. III C. The preparation of the referential data is described in Sec. III B. The score is fed back to the optimizer to increase the size of the data used in Bayesian optimization.

The iteration loop is repeated until the number of iterations reaches a criterion. Otherwise, the workflow goes back to the optimizer. After the loop has ended, the candidate with the best score found in the iterations is output.

A. Bayesian optimization

As mentioned above, we use COMBO [10,50] to cover the role of the optimizer in Fig. 2. We use Thompson sampling as a heuristic to the exploration-exploitation problem in optimization. The dimension of the random feature maps, which determines the degree of approximation for the Gaussian kernel, is set to 5000. The first 10 samples are chosen randomly without using Bayesian optimization. The number of iterations is set to 100, including the first 10 iterations with the random sampling.

The candidate list consists of $(R_{1-\alpha}Z_{\alpha})(\mathrm{Fe}_{1-\beta}\mathrm{Co}_{\beta})_{12-\gamma}\mathrm{Ti}_{\gamma}$ systems for all combinations of $R = \mathrm{Y, Nd, Sm}$; $Z = \mathrm{Zr, Dy}$; $\alpha = 0, 0.1, \ldots, 1$; $\beta = 0, 0.1, \ldots, 1$; $\gamma = 0, 0.5, \ldots, 2$. There are duplicates on the list (e.g., $\mathrm{YZr_0Fe_{12}}$ and $\mathrm{YDy_0Fe_{12}}$), and the number of unique items is 3245 out of the $3 \times 2 \times 11 \times 11 \times 5 = 3630$ systems.

We use 11 different sets of descriptors listed in Table I. The descriptors consist of the number of electrons per cell (N), the number of electrons from the R element per cell (N_R), the number of electrons from the Z element per cell (N_Z), $N_R + N_Z$ ($\equiv N_{2\mathrm{a}}$), the number of electrons from the transition elements—namely Fe,Co,Ti—per cell (N_{T}), the atomic number of the R element (Z_R), the atomic number of the Z element (Z_Z), an index for the R element ($n_R = 0, 1, 2$ corresponding to $R = \mathrm{Y, Nd, Sm}$), an index for the Z element ($n_Z = 0, 1$ corresponding to $Z = \mathrm{Zr, Dy}$), the Z content per cell (α), the Co/(Fe+Co) ratio (β), the Ti content (γ), and the values of α_1, α_2, α_3, and α_4 when the chemical formula is expressed in the form of $(\mathrm{Y}_{1-\alpha_1-\alpha_2-\alpha_3-\alpha_4}\mathrm{Nd}_{\alpha_1}\mathrm{Sm}_{\alpha_2}\mathrm{Zr}_{\alpha_3}\mathrm{Dy}_{\alpha_4})(\mathrm{Fe}_{1-\beta}\mathrm{Co}_{\beta})_{12-\gamma}\mathrm{Ti}_{\gamma}$.

B. First-principles calculation

In the "Simulator" block in Fig. 2, we perform first-principles calculation based on density functional theory with the local density approximation [51,52]. We use the open-core approximation [53–55] to the f-electrons in Nd, Sm, and Dy and apply the self-interaction correction [56].

We assume the $\mathrm{ThMn_{12}}$ structure (Fig. 3) for the $(R_{1-\alpha}Z_{\alpha})(\mathrm{Fe}_{1-\beta}\mathrm{Co}_{\beta})_{12-\gamma}\mathrm{Ti}_{\gamma}$ systems. The lattice parameters are determined by linear interpolation from those for $R\mathrm{Fe}_{12}$, $R\mathrm{Fe}_{11}\mathrm{Ti}$, $Z\mathrm{Fe}_{12}$, $Z\mathrm{Fe}_{11}\mathrm{Ti}$, and $R\mathrm{Co}_{12}$. These values for the stoichiometric systems were calculated with the

TABLE I. 11 Descriptors used in the Bayesian optimization. See the text for a description of the variables N, N_R, N_Z, $N_{2\mathrm{a}}$, N_{T}, Z_R, Z_Z, n_R, n_Z, α, β, γ, α_1, α_2, α_3, and α_4.

no. 1	N		
no. 2	$N_{2\mathrm{a}}, N_{\mathrm{T}}$	no. 7	$N_{2\mathrm{a}}, \beta, \gamma$
no. 3	N_R, N_Z, N_{T}	no. 8	N_R, N_Z, β, γ
no. 4	$Z_R, Z_Z, \alpha, N_{\mathrm{T}}$	no. 9	$Z_R, Z_Z, \alpha, \beta, \gamma$
no. 5	$n_R, n_Z, \alpha, N_{\mathrm{T}}$	no. 10	$n_R, n_Z, \alpha, \beta, \gamma$
no. 6	$\alpha_1, \alpha_2, \alpha_3, \alpha_4, N_{\mathrm{T}}$	no. 11	$\alpha_1, \alpha_2, \alpha_3, \alpha_4, \beta, \gamma$

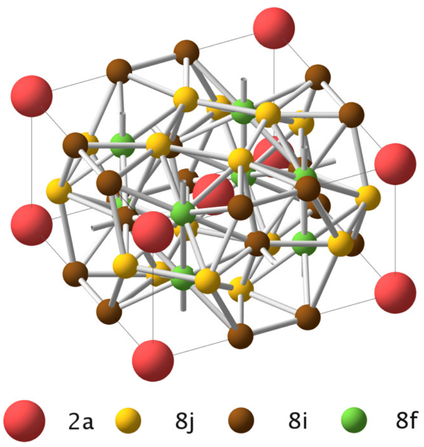

FIG. 3. The 2a, 8j, 8i, and 8f Wyckoff positions in the ThMn_{12} structure. Some bonds are shown as guides to the eye to illustrate the three-dimensional structure.

projector-augmented-wave (PAW) method [57,58] using a software package QMAS [59]. We use the Perdue-Burke-Ernzerhof (PBE) exchange-correlation functional [60] of the generalized gradient approximation (GGA) to obtain adequate structures. The values for $Z\mathrm{Fe}_{11}\mathrm{Ti}$ and $R\mathrm{Co}_{12}$ are presented in Appendix A. Those for $R\mathrm{Fe}_{12}$, $R\mathrm{Fe}_{11}\mathrm{Ti}$, and $Z\mathrm{Fe}_{12}$ are in Refs. [49,61,62].

In the treatment with the CPA [63–65], we assume quenched randomness for random occupation of the elements: the element R and Z are assumed to occupy the 2a site (see Fig. 3 for the Wyckoff positions). Titanium is assumed to occupy the 8i site. Iron (cobalt) is assumed to occupy the 8f, 8i, and 8j sites with a common probability of $1-\beta$ (β) to these sites.

We calculate the magnetization, the Curie temperature, and the formation energy from the unary systems. We use the raw value of magnetization from KKR-CPA. To estimate the Curie temperature, we calculate intersite magnetic couplings by using Liechtenstein's method [66], and we convert them to the Curie temperature using the mean-field approximation [61]. Although this procedure overestimates the Curie temperature, we can expect from previous results that it can capture material trends because theoretical values within the mean-field approximation had a significant correlation with experimental Curie temperatures [67].

The best value among the candidates is 1.76 T (DyFe_{12}) for magnetization, 1310 K [$\mathrm{Sm}(\mathrm{Fe}_{0.2}\mathrm{Co}_{0.8})_{12}$] for Curie temperature, and -2.85 eV ($\mathrm{SmCo}_{10}\mathrm{Ti}_2$) for the formation energy. It should be noted, however, that the values on the list cannot be directly used as information for experimental synthesis because the data do not include information of phase competition (especially with $\mathrm{Th}_2\mathrm{Zn}_{17}$-type and $\mathrm{Th}_2\mathrm{Ni}_{17}$-type phases), magnetic anisotropy, and contribution to the magnetization from the f-electrons. We cover this subject in Appendix D, and we provide lists of some of the best systems with the physical properties there.

As for the formation energy, KKR needs too many computational resources to obtain an accurate energy difference between systems when they have far different structures from each other. We use the method that we describe in the following subsection to correct the energy difference calculated by KKR-CPA with referential data of total energy obtained by PAW.

C. A method for integration datasets

Let us consider the formation energy from the unary systems defined as follows:

$$\begin{aligned}\Delta E \equiv\ & E[(R_{1-\alpha}Z_\alpha)(\mathrm{Fe}_{1-\beta}\mathrm{Co}_\beta)_{12-\gamma}\mathrm{Ti}_\gamma] \\ & - E[(\text{the unary systems})],\end{aligned} \tag{1}$$

where "(the unary systems)" is defined as

$$\begin{aligned}&(\text{the unary systems}) \\ &= (1-\alpha)R + \alpha Z \\ &\quad + (1-\beta)(12-\gamma)\mathrm{Fe} + \beta(12-\gamma)\mathrm{Co} + \gamma\mathrm{Ti},\end{aligned} \tag{2}$$

and $E[\cdot]$ denotes the total energy of the system in the square brackets. Because the structures of $(R_{1-\alpha}Z_\alpha)(\mathrm{Fe}_{1-\beta}\mathrm{Co}_\beta)_{12-\gamma}\mathrm{Ti}_\gamma$ and each of the unary systems are much different from one another, it is not efficient to directly calculate this formation energy with the KKR method, although it can deal with nonstoichiometric systems with CPA. Our idea is to calculate the formation energy of stoichiometric systems more accurately by another method, and use calculated energies as reference data.

We construct a stochastic model for the total energy of $(R_{1-\alpha}Z_\alpha)(\mathrm{Fe}_{1-\beta}\mathrm{Co}_\beta)_{12-\gamma}\mathrm{Ti}_\gamma$ based on the expectation that the smaller the structural difference is between the two systems, the more accurate is the energy difference given by KKR-CPA. To quantify the difference between the systems, we consider a descriptor with which the difference between the systems ($\vec{x}$ and $\vec{y}$) is well-described by the distance ($|\vec{x}-\vec{y}|$). Let us denote the reference systems in the form of the descriptor by $\vec{x}_1^R, \vec{x}_2^R, \dots, \vec{x}_M^R$, where M is the number of the reference systems. The descriptor here does not have to be identical to the descriptor used in the Bayesian optimization. In the demonstration, we use a set of $(\alpha', \beta', \gamma') \equiv (\alpha, \beta(12-\gamma)/12, \gamma/12)$ with which the search space can be expressed as $(R_{1-\alpha'}Z_{\alpha'})(\mathrm{Fe}_{1-\beta'-\gamma'}\mathrm{Co}_{\beta'}\mathrm{Ti}_{\gamma'})_{12}$ irrespective of the choice of the descriptor in the optimization.

For each of the reference points $\vec{x}_i^R$, we construct a stochastic model $\tilde{E}_i[\vec{y}]$ for the total energy of a system $\vec{y}$ (see also the graphs outside the box in Fig. 4):

$$\tilde{E}_i[\vec{y}] - E\big[\vec{x}_i^R\big] = E'[\vec{y}] - E'\big[\vec{x}_i^R\big] + \varepsilon_i. \tag{3}$$

The two E's on the right-hand side (to which primes are attached) are the total energy calculated with KKR-CPA, whereas $E[\vec{x}_i^R]$ on the left-hand side is evaluated by a more accurate method, for which we use PAW in the present work. ε_i is a random variable whose distribution is $\mathcal{N}(0, S_i^2)$, i.e., the normal distribution whose mean is zero and variance is S_i^2, where $S_i^2 \equiv \sigma^2|\vec{y}-\vec{x}_i^R|$ and σ^2 is a parameter we will estimate later. This model describes the expectation that the deviation of the energy difference (the first two terms on the right-hand side) from the true difference (the left-hand side) tends to be large when the difference, $|\vec{y}-\vec{x}_i^R|$, is large. The graphs outside the box in Fig. 4 depict how $\tilde{E}_i$ behave: there are three models corresponding to the three reference systems

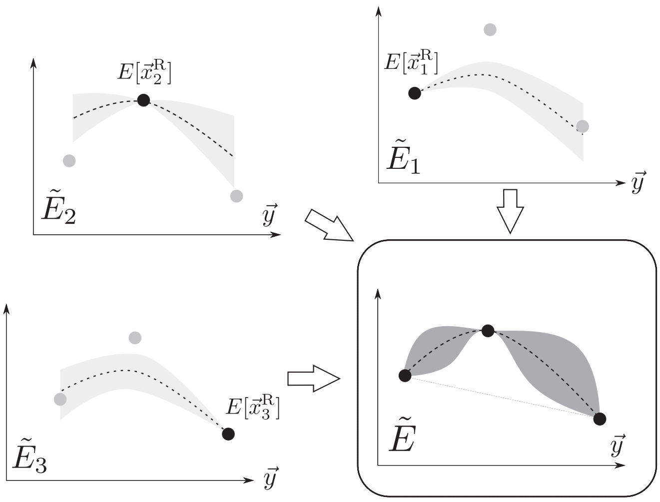

FIG. 4. A schematic diagram of the procedure for constructing an integrated model. The three graphs outside the box denote the models defined by Eq. (3). They are integrated into the model described by Eq. (10).

$(\vec{x}_1^R, \vec{x}_2^R, \vec{x}_3^R)$, and the error ε_i in each model is large when the distance of the reference point from $\vec{y}$ is large.

We then integrate these models, $\tilde{E}_1, \dots, \tilde{E}_M$, into a single model $\tilde{E}$ by imposing the following condition to the distribution of ε_i:

$$\tilde{E}_1[\vec{y}] = \tilde{E}_2[\vec{y}] = \cdots = \tilde{E}_M[\vec{y}] \equiv \tilde{E}[\vec{y}]. \tag{4}$$

This condition can be rewritten as follows:

$$\varepsilon_i = \tilde{E}[\vec{y}] - E\big[\vec{x}_i^R\big] - E'[\vec{y}] + E'\big[\vec{x}_i^R\big]. \tag{5}$$

Therefore, the conditional probability distribution of $\tilde{E}$ is

$$\Pr(\tilde{E}[\vec{y}] = t) = \frac{\prod_i P_i\big(t - E\big[\vec{x}_i^R\big] - E'[\vec{y}] + E'\big[\vec{x}_i^R\big]\big)}{\int dt' \prod_i P_i\big(t' - E\big[\vec{x}_i^R\big] - E'[\vec{y}] + E'\big[\vec{x}_i^R\big]\big)}, \tag{6}$$

where P_i denotes the probability distribution function of $\mathcal{N}(0, S_i^2)$. It is straightforward to see that this conditional distribution is the normal distribution whose mean $\tilde{\mu}$ and variance $\tilde{S}^2$ are

$$\tilde{\mu} = E'[\vec{y}] + \frac{1}{\Omega}\sum_i \omega_i\big\{E\big[\vec{x}_i^R\big] - E'\big[\vec{x}_i^R\big]\big\}, \tag{7}$$

$$\tilde{S}^2 = \frac{\sigma^2}{\Omega}, \tag{8}$$

where ω_i is a weight and Ω is a normalization factor that are defined as

$$\omega_i \equiv \frac{1}{\big|\vec{y} - \vec{x}_i^R\big|}, \qquad \Omega \equiv \sum_i \omega_i. \tag{9}$$

In another expression, our integrated model is

$$\tilde{E}[\vec{y}] = E'[\vec{y}] + \frac{1}{\Omega}\sum_i \omega_i\big\{E\big[\vec{x}_i^R\big] - E'\big[\vec{x}_i^R\big]\big\} + \tilde{\varepsilon}, \tag{10}$$

where $\tilde{\varepsilon}$ is a random variable whose distribution is $\mathcal{N}(0, \tilde{S}^2)$. Algorithm 1 summarizes the construction of the model explained so far in the form of a pseudocode. Note that the value for the input variable σ^2 has not yet been determined. The characteristics of $\tilde{E}[\vec{y}]$ is illustrated in the right-bottom panel in Fig. 4. Although $\tilde{E}[\vec{y}]$ is singular at $\vec{y} = \vec{x}_i^R$, it is easy to see that this is removable and $\lim_{\vec{y}\to\vec{x}_i^R} \tilde{E}[\vec{y}] = E[\vec{x}_i^R]$.

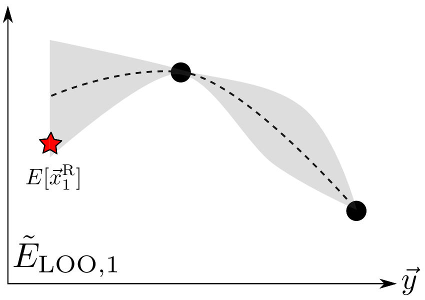

FIG. 5. A schematic plot of $\tilde{E}_{\mathrm{LOO},i}$ for $i = 1$ and $M = 3$.

Algorithm 1 A pseudocode of the algorithm for determination of $\tilde{\mu}$ in Eq. (7) and $\tilde{S}^2$ in Eq. (8). This needs σ^2 and Δ_i as inputs, which are calculated by the algorithm shown in Alg. 2.

```
Input: $\vec{x}_i^R$ $(i = 1, \dots, M)$, $\vec{y} \notin \{\vec{x}_i^R\}$ : descriptor; $\sigma^2$;
  $E'[\vec{y}]$: energy estimation; $\Delta_i \equiv E[\vec{x}_i^R] - E'[\vec{x}_i^R]$
Output: $\tilde{\mu}$, $\tilde{S}^2$
  Initialize $\tilde{\mu} \leftarrow 0$, $\Omega \leftarrow 0$
  for $i = 1$ to $M$ do
    $\Omega \leftarrow \Omega + 1/|\vec{y} - \vec{x}_i^R|$
    $\tilde{\mu} \leftarrow \tilde{\mu} + \Delta_i/|\vec{y} - \vec{x}_i^R|$
  end for
  Average: $\tilde{\mu} \leftarrow \tilde{\mu}/\Omega + E'[\vec{y}]$ [Eq. (7)]
  Variance: $\tilde{S}^2 \leftarrow \sigma^2/\Omega$ [Eq. (8)]
```

The formation energy from the unary systems can be calculated as $\Delta E \simeq \tilde{E}[\vec{y}] - E[(\text{the unary systems})]$, where $E[(\text{the unary systems})]$ is calculated by an accurate method.

To complete the formulation, we discuss estimation of σ^2 based on the data for the reference systems. We use the maximum likelihood estimation. However, it cannot be directly applied to our model because the distribution of $\tilde{\varepsilon}$ becomes the δ function in the limit of $\vec{y} \to \vec{x}_i^R$, regardless of the value of σ^2. To avoid this singularity, we consider a model $\tilde{E}_{\mathrm{LOO},i}$ that is constructed from all reference systems but $\vec{x}_i^R$. Figure 5 depicts the construction of such a model. Now, we consider the probability of $\tilde{E}_{\mathrm{LOO},i}[\vec{y}]$ at $\vec{y} = \vec{x}_i^R$. Regarding the probability for $\tilde{E}_{\mathrm{LOO},i}[\vec{y} = \vec{x}_i^R] = E[\vec{x}_i^R]$ as a likelihood L_i, we select the value of σ^2 that maximizes $\mathcal{L} = \prod_i L_i$. We then obtain

$$\sigma^2 = \frac{1}{M}\sum_i \Omega_{\mathrm{LOO},i}\big(E\big[\vec{x}_i^R\big] - E'\big[\vec{x}_i^R\big]\big)^2, \tag{11}$$

where

$$\Omega_{\mathrm{LOO},i} \equiv \sum_{j\neq i} \frac{1}{\big|\vec{x}_i^R - \vec{x}_j^R\big|}. \tag{12}$$

The determination of σ^2 is summarized in a form of a pseudocode in Alg. 2. The output values of σ^2 and Δ_i correspond to those in the algorithm described in Alg. 1.

Algorithm 2 A pseudocode of the algorithm for determination of σ^2 in Eq. (8). The energy difference Δ_i is also output to be used in the algorithm described in Alg. 1.

```
Input: $\vec{x}_i^R$ $(i = 1, \ldots, M)$: descriptor;
    $E[\vec{x}_i^R]$, $E'[\vec{x}_i^R]$: energy estimation
Output: $\Delta_i (i = 1, \ldots, M)$, $\sigma^2$
    Initialize $\sigma^2 \leftarrow 0$
    for $i = 1$ to $M$ do
        $\Delta_i \leftarrow E[\vec{x}_i^R] - E'[\vec{x}_i^R]$
        $\sigma^2 \leftarrow \sigma^2 + \Delta_i^2 \sum_{j \neq i} 1/|\vec{x}_i^R - \vec{x}_j^R|$
    end for
    $\sigma^2 \leftarrow \sigma^2/M$ [Eq. (11)]
```

In the actual application, we calculate $E'[\vec{x}_i^R]$ with the KKR-LDA method and $E[\vec{x}_i^R]$ with the PAW-GGA method. We need to calibrate $E[\vec{x}_i^R]$ with a linear term because the difference in the treatment of the core electrons is another major source of error in the total energy, which is described well by a linear function. We deal with it by extending our model to include an adjustable linear term, and we use it in the actual calculation. We discuss this extension of the model in Appendix B.

IV. RESULTS AND DISCUSSION

A. Integration method

We show how the integration model explained in Sec. III C works before we present the benchmark of the whole scheme in the next subsection. Let us take a simple example first: we consider a one-dimensional function $E[x] = \sin(\pi x)$ as a true model. Assume that we have many data about $E'[x]$, which is an approximate function and actually obeys $E'[x] = \sin(\pi x) + 2x$. Because their derivatives differ from each other by 2, the difference $E'[x] - E'[x_i^R]$ deviates from $E[x] - E[x_i^R]$ roughly by $2(x - x_i^R)$. The assumption of the model described in Sec. III C was that the error in $E'[x]$ is large when $|x - x_i^R|$ is large. This is correct in the magnitude of the errors, but there is a bias in its sign. In this example, we prepare a dataset of $E'[x]$ for $x = -0.5$ to 0.5 with an interval of 0.02. We choose five reference points x_i^R $(i = 1, \ldots, 5)$ from the x-values, and we prepare a dataset of $E[x_i^R]$.

The result of the prediction is shown in Fig. 6. The original points of $E'[x]$ are shown as points in cyan. The purple points show the mean value of the model; the light purple region shows the values of the standard deviation. The green curve shows $E[x]$, the true model, and the green circles are the reference points used in the prediction.

The mean values, which may be used as values of a point estimation, are improved from the original data in most of the region. The error region also covers the line of the true model except in the region $x > 0.26$.

It is noteworthy that the bias in the sign of the error mentioned above makes the uncertainty of the model large. This would be improved if we assumed an asymmetric distribution for ε_i in Eq. (3). However, the resultant form of the integrated model becomes more complicated.

Next, let us see the results of the formation energy obtained by this method. Figure 7 shows the formation energy of $(\mathrm{Sm}_{1-\alpha}\mathrm{Zr}_\alpha)\,(\mathrm{Fe}_{1-\beta}\mathrm{Co}_\beta)_{12-\gamma}\mathrm{Ti}_\gamma$ from the unary systems for $\gamma = 0$ (left) and $\gamma = 0.5$ (right). The mean values are shown in the top two figures, and the standard deviations are shown in the bottom two figures. In each figure, the horizontal axes show values of the Co/(Fe+Co) ratio (β), and the vertical axes show values of the Zr concentration (α).

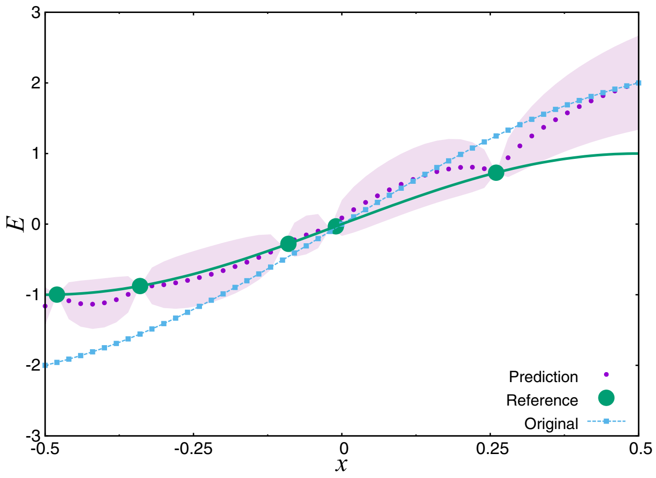

FIG. 6. Results of the integration scheme when $E[x] = \sin(\pi x)$ (green line) and $E'[x] = \sin(\pi x) + 2x$ (cyan points). The green circles denote the points of reference, $E[x_i^R]$. The mean value of the improved model, $\tilde{E}[x]$, is denoted by the purple points. The standard deviation is denoted by the region in light purple.

The standard deviation of the model is zero at the corners of the figure in $\gamma = 0$ (left bottom) because they correspond to the reference points, namely SmFe_{12}, ZrFe_{12}, SmCo_{12}, and ZrCo_{12}. Therefore, the mean values at the corners of the left top figure are identical with the reference data (obtained by PAW-GGA). Because the other reference points are $\mathrm{SmFe}_{11}\mathrm{Ti}$ and $\mathrm{ZrFe}_{11}\mathrm{Ti}$ (both at $\beta = 0$, $\gamma = 1$), the contours are placed slightly to the right in $\gamma = 0$. This is more obvious in the plot for $\gamma = 0.5$ (the right bottom panel).

Although we used the mean values ($\tilde{\mu}$) in the Bayesian optimization in Sec. IV B, it is important to check how the model has uncertainty in its prediction, which is typically represented by the standard deviation $\tilde{S}$: the model needs more reference points when $\tilde{S}$ is too large compared with the value of $\tilde{\mu}$. One can also make use of the upper (lower) confidence bound $\tilde{\mu} + k\tilde{S}$ ($\tilde{\mu} - k\tilde{S}$)—where k is a positive adjustable parameter—instead of $\tilde{\mu}$ to take account of the uncertainty in a pessimistic (optimistic) manner.

B. Bayesian optimization

In this subsection, we show the performance of the Bayesian optimization and discuss the results. Figure 8 shows one of the optimization processes with respect to magnetization using the descriptor no. 8. The highest magnetization found in the first i iterations, $\max_{j \leqslant i} \mu_0 M(j)$, is plotted against the number of iterations, i. In this run, the highest $\mu_0 M$ in all 3630 systems was found at the 30th iteration, where we define the zeroth iteration as that with the first sample. This process depends on a random sequence that is used in the sampling from the candidate lists. To take statistical profiles, we repeat the optimization scheme (which we call a session) 1000 times.

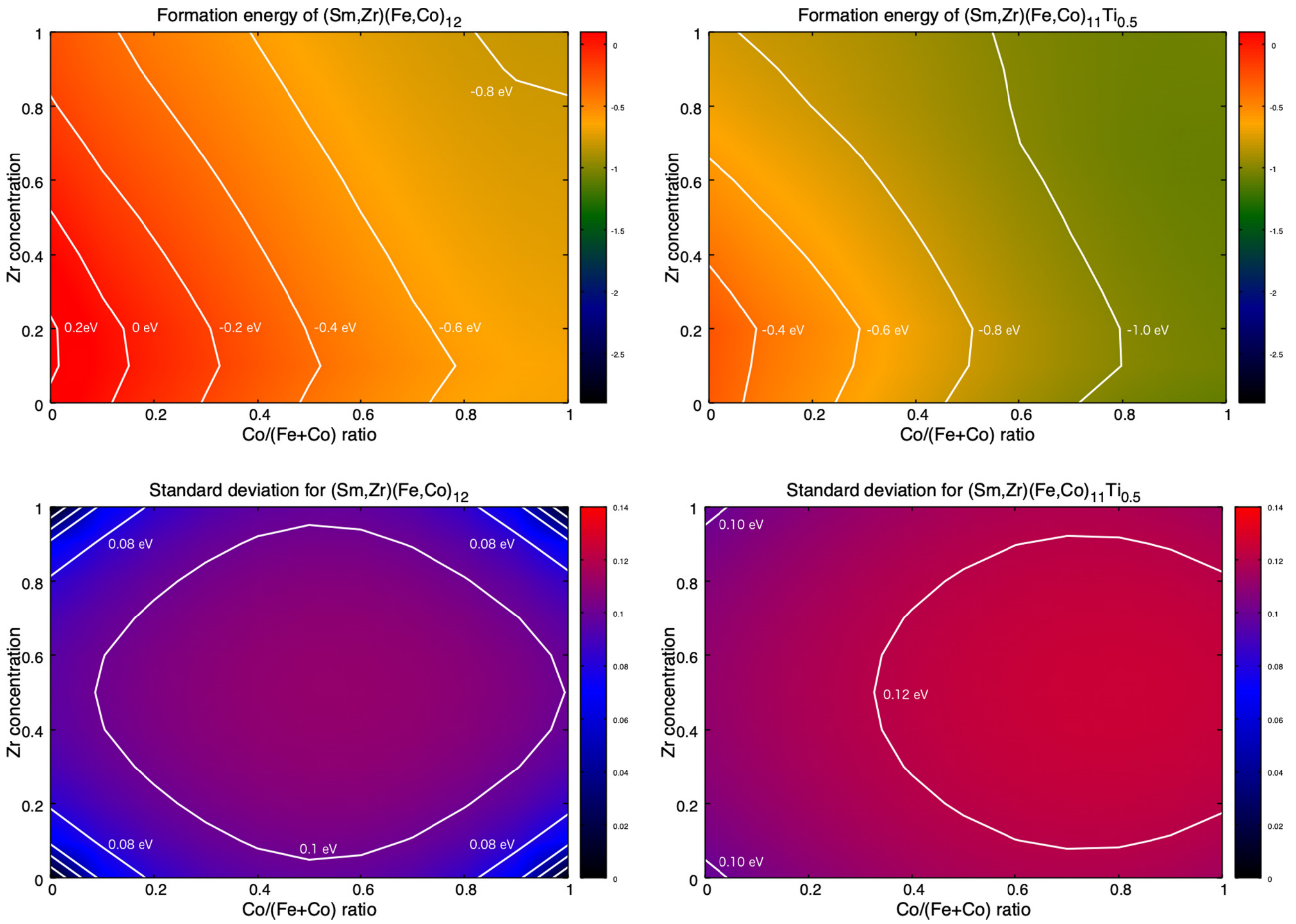

FIG. 7. The mean values (top) and the standard deviations (bottom) of the integrated model for Sm(Fe, Co)$_{12}$ (left) and Sm(Fe, Co)$_{11.5}$Ti$_{0.5}$ (right). The intervals of the contours are 0.2 eV for the mean values and 0.02 eV for the standard deviations.

To analyze the efficiency as a function of the number of iterations, we consider a cumulative distribution $\mathcal{D}_i(s)$ that is defined as the number of sessions in which a system with a higher score than s is found in the first i iterations. We show a plot of $\mathcal{D}_i(s)$ in Bayesian optimization of magnetization using the descriptor no. 8 as the left figure in Fig. 9, where the horizontal axis shows the number of iterations, i, and the vertical axis shows the score variable, s. We also show a plot of the cumulative probability $\mathcal{P}_i(s)$ in the random sampling that is analytically obtained at the right-hand side. Because we took a sufficient number of sessions, there is a negligible difference between the two figures for the first 10 iterations. The efficiency in the left figure suddenly improves when the Bayesian optimization is switched on.

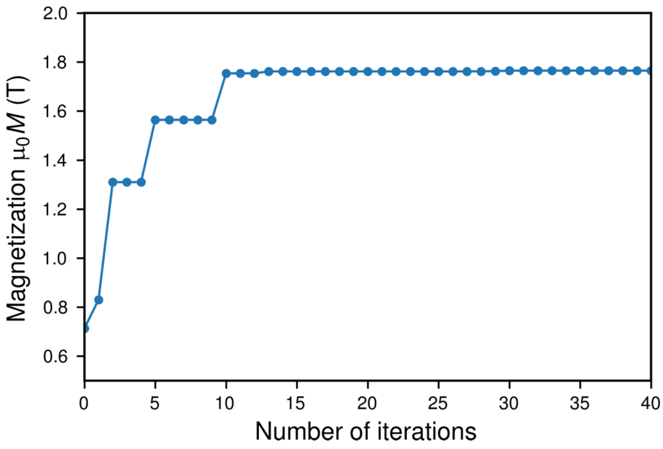

FIG. 8. The best magnetization found in the loop of a session as a function of the number of iterations. The first result in a session corresponds to the point at the zeroth iteration.

Figure 10 shows the success rate of finding the systems with the top 10 values of the target properties—magnetization ($\mu_0 M$), Curie temperature (T_C), and formation energy from the unary systems (ΔE)—within 50 steps. The results with Bayesian optimization (BO) are compared with the search by the random sampling (RS). The numbers with "#" in the figure denote the descriptors listed in Table I. We find that the efficiency depends significantly on the choice of the descriptor. It is obvious from the figure that the descriptors nos. 7–11 are very efficient in Bayesian optimization and much superior to those with nos. 1–6 and the random sampling. This clearly shows that β and γ are important factors because the descriptors nos. 7–11 differ from nos. 2–6 only by β and γ used instead of N_T.

This example demonstrates how we can incorporate domain knowledge into machine learning. It is known that

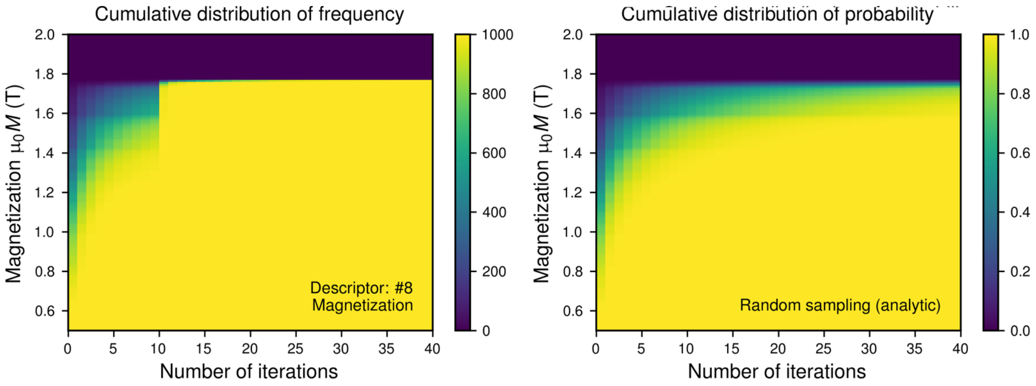

FIG. 9. Left: The cumulative distribution of frequency $\mathcal{D}_i(s)$ in the optimization of magnetization using the Bayesian optimization with the descriptor no. 8. Right: The cumulative probability of probability $\mathcal{P}_i(s)$ that is analytically calculated for the optimization of magnetization with the random sampling.

the magnetization of ferromagnetic random alloys of Fe and another transition metal is usually a well-behaved function of the number of valence electrons. It is called the Slater-Pauling curve. This curve can be reproduced well by first-principles calculation with CPA [68], and so it is the Slater-Pauling-like curve for the Curie temperature [69]. These effects have also been observed in ThMn$_{12}$-type compounds experimentally [27,44], and explained theoretically [38,43,49,61,70]. On the basis of these previous studies, we were able to expect that including β [the Co/(Fe+Co) ratio] and γ (the Ti content) in the descriptor would improve the efficiency of the search in advance.

However, we also find that β and γ alone do not work as an efficient descriptor. The results of the Bayesian optimization using β, γ, and the pair of them as descriptors are also shown in Fig. 10. Those success rates are significantly lower than the rates with the descriptors nos. 7–11. This is because there are 66 [=3 (for R) $\times$ 2 (for Z) $\times$ 11 (for α)] candidates that have common values of β and γ on the list, and 50 steps are not enough to obtain an adequate Bayesian model and draw

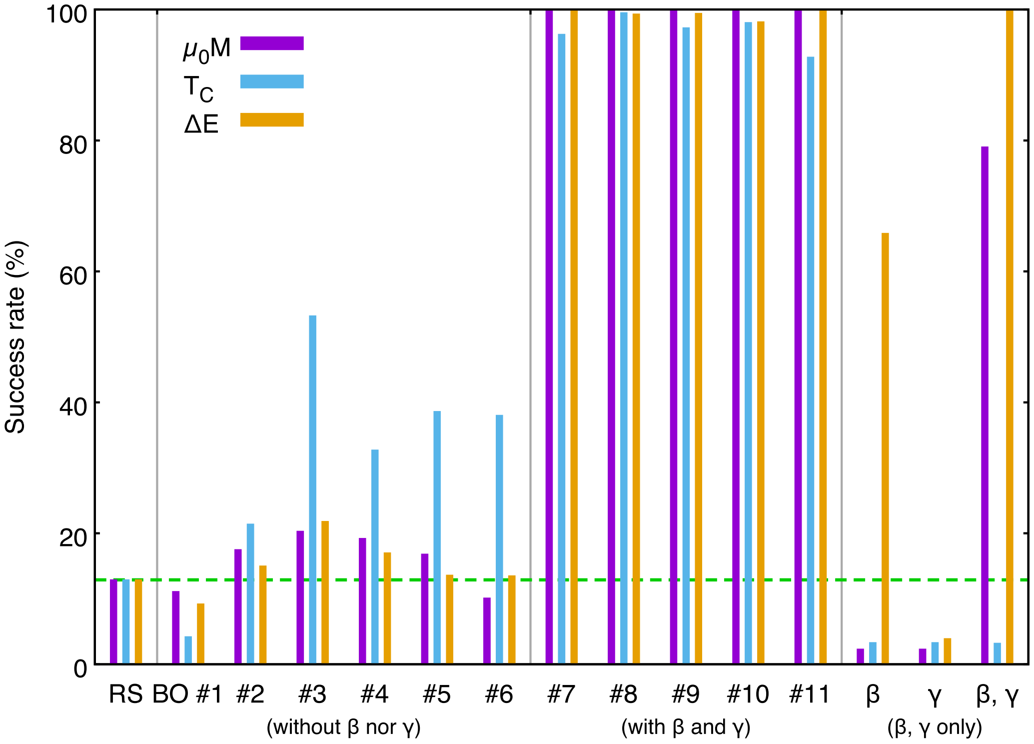

FIG. 10. Success rate of finding the systems with the top 10 values of magnetization ($\mu_0 M$), Curie temperature ($T_{\rm C}$), and formation energy from the unary systems (ΔE) among the 3630 candidates within 50 steps. RS stands for the random search, and BO stands for the Bayesian optimization. The descriptors used in the search are shown along the horizontal axis: the numbers denote the descriptors listed in Table I; the label β, γ, and β, γ denote the results with using β, γ, and β, γ as descriptors, respectively. The horizontal dashed line (green) is a guide to the eye to show the height of the bars for the random sampling.

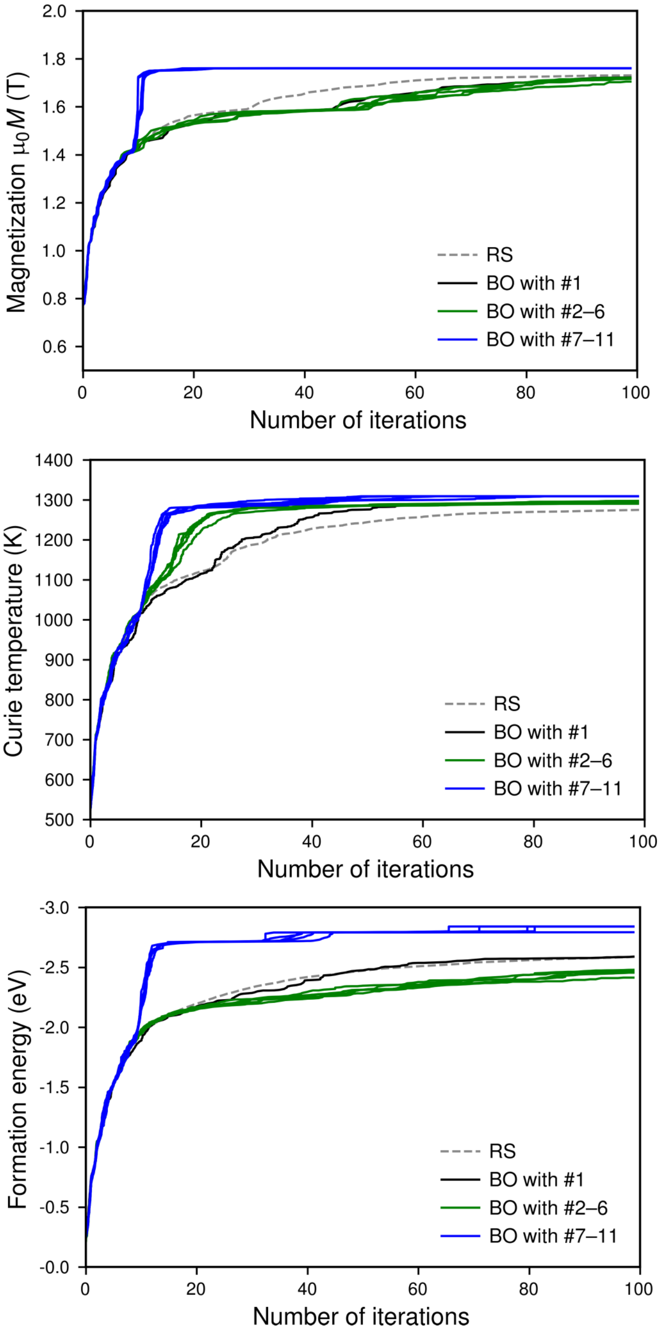

FIG. 11. The 95% contours of the cumulative distribution $\mathcal{D}_i(s)$ in the random sampling (gray dashed) and in the Bayesian optimizations with the descriptor no. 1 (black), the descriptors nos. 2–6 (green), and the descriptors nos. 7–11 (blue). The horizontal axis shows i; the vertical axis shows s. The top, middle, and bottom figures show results in the optimization of the magnetization, the Curie temperature within the mean-field approximation, and the formation energy from the unary systems, respectively.

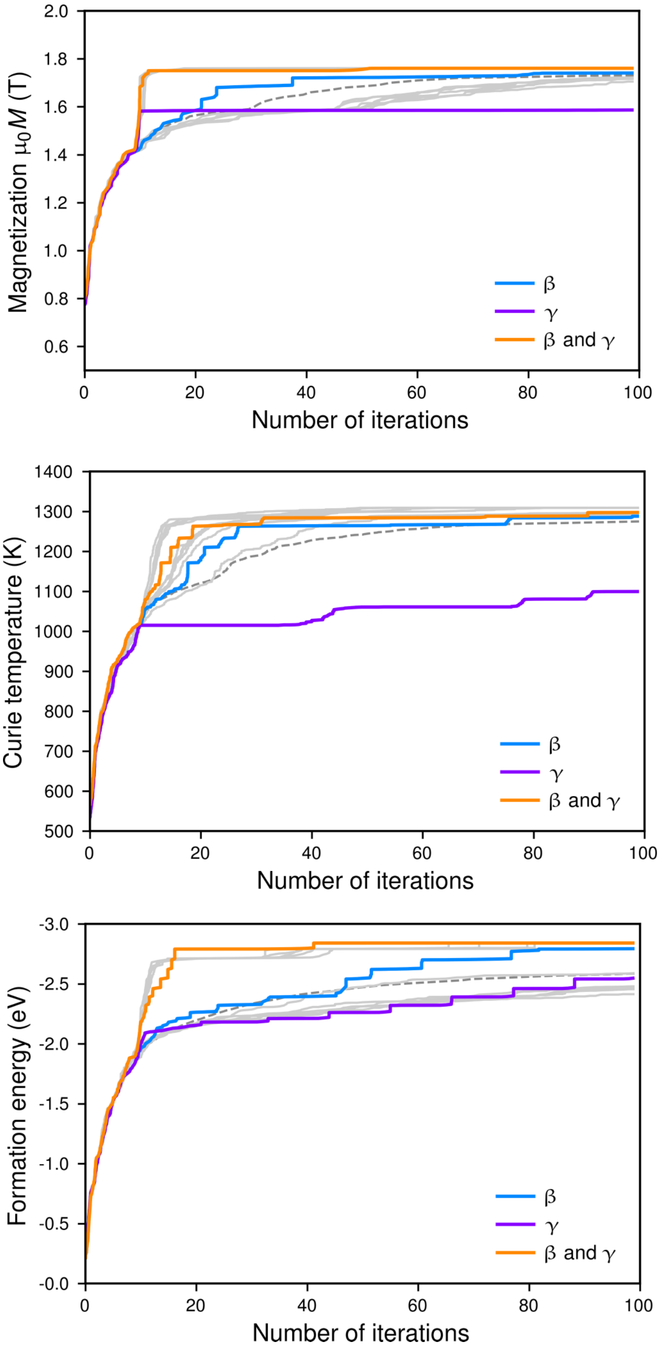

FIG. 12. The 95% contours of the cumulative distribution $\mathcal{D}_i(s)$ in the Bayesian optimizations with the descriptor β (cyan), the descriptor γ (purple), and the descriptor (β, γ) (orange). Those in the random sampling and in the Bayesian optimization with the other descriptors are also shown as gray lines. The horizontal axis shows i; the vertical axis shows s. The top, middle, and bottom figures show results in the optimization of the magnetization, the Curie temperature within the mean-field approximation, and the formation energy from the unary systems, respectively.

one of the top 10 systems by chance. It is noteworthy that the efficiency of the Bayesian optimization is largely improved by adding only N_{2a}, as in the descriptor no. 7.

Figure 11 shows the 95% (950 sessions) contour of $\mathcal{D}_i(s)$ obtained with the descriptors nos. 1–11. The best 95% of the data points are laid above those curves. The panels show results with the random search and the Bayesian optimization with the descriptors nos. 1–11. As shown in these graphs, the efficiency of the search depends much on the target property to optimize. The descriptors nos. 7–11, which are with β and γ, have a satisfying efficiency, with which even 20 iterations are enough to obtain a nearly best score regardless of the target property. The situation is quite different with the descriptors nos. 1–6. In the optimization of the Curie temperature, these descriptors work efficiently. However, the optimizations of the magnetization and the formation energy progress even more slowly than the random sampling.

This difference between the Curie temperature and the other targets is also seen in Fig. 12, where the pair of β and γ are used as descriptors. The pair descriptor works as efficiently as the descriptors nos. 7–11 in the optimization of the magnetization and the formation energy. However, a discernible difference in efficiency exists in the optimization of the Curie temperature. This suggests the importance of information about elements at the 2a site (R and Z), which is consistent with Dam *et al.*'s observation that the concentration of rare-earth elements is important in explaining the Curie temperature of binary alloys composed of a rare-earth element and a 3*d* transition metal [6].

The dependence of the search efficiency on the dimension of the descriptor is also noteworthy. When the dimension of a descriptor is large, the descriptor can accommodate a large search space on the one hand. However, modeling tends to be difficult with a higher dimensional space, which is referred to as "curse of dimensionality," on the other hand. Note that we have descriptors with four different dimensions in the groups of the descriptors nos. 1–6 and nos. 7–11. Figure 11 shows that the dimension has only a minor effect on the efficiency. This dependence is magnified by the more stringent criterion of the top 10 benchmark as shown in Fig. 10, especially in the results with the descriptors nos. 1–6. However, the difference among the descriptors nos. 7–11 is still subtle. Therefore, we expect that it would be safe to include six or a little more variables in the descriptor when we optimize for another target property.

V. CONCLUSION

In this paper, we presented a machine-learning scheme for searching high-performance magnetic compounds. Our scheme is based on Bayesian optimization, and it has a much higher efficiency than random sampling. We demonstrated its efficiency by taking the example of optimization of magnetization, Curie temperature, and formation energy for the search space of magnet compounds having the $ThMn_{12}$ structure. One of the typical results is the success rate of finding the top 10 systems with the highest properties when 50 systems are sampled from a candidate list of 3630 systems (Fig. 10). The success rate is more than 90% with our scheme when the descriptor is appropriately chosen, while it is approximately 10% in the random sampling.

The efficiency is maximized when we include the Co content (β), the Ti content (γ), and the information of the R and Z elements (e.g., N_{2a}) in the descriptor. This improvement is what we could expect from the previous studies of magnet compounds. We stressed that it is important to incorporate domain knowledge into the choice of a descriptor. We also discussed how many variables a descriptor can accommodate without deteriorating the search efficiency. Although an excessive addition of variables to the descriptor can lose the efficiency of the search, there was not a significant loss when we doubled the dimension of the optimal descriptor.

TABLE II. Optimized lattice parameters for $R\mathrm{Co}_{12}$ ($R =$ Y, Nd, Sm). See Table III for definitions of $p_{8\mathrm{i}}$ and $p_{8\mathrm{j}}$.

R	a (Å)	c (Å)	$p_{8\mathrm{i}}$	$p_{8\mathrm{j}}$
Y	8.282	4.659	0.3585	0.2738
Nd	8.336	4.677	0.3590	0.2695
Sm	8.309	4.669	0.3587	0.2715

We also proposed an integration scheme of two datasets to improve the accuracy of an inexpensive large-sized dataset with use of an accurate and small-sized dataset (reference dataset). The algorithm (Algorithms 1 and 2) is easy to implement and fast. Prediction with a confidence bound (or the standard deviation $\tilde{S}$) is another feature of the scheme. We have also shown how it worked in the calculation of the formation energy (Fig. 7).

ACKNOWLEDGMENTS

We acknowledge the support from the Elements Strategy Initiative Project under the auspices of MEXT. This work was also supported by MEXT as a social and scientific priority issue (Creation of new functional Devices and high-performance Materials to Support next-generation Industries; CDMSI) to be tackled by using the post-K computer, by the "Materials research by Information Integration" Initiative (MI^{2}I) project of the Support Program for Starting Up Innovation Hub from the Japan Science and Technology Agency (JST). The computation was partly conducted using the facilities of the Supercomputer Center, the Institute for Solid State Physics, the University of Tokyo, and the supercomputer of ACCMS, Kyoto University. This research also used computational resources of the K computer provided by the RIKEN Advanced Institute for Computational Science through the HPCI System Research project (Project ID:hp170100).

APPENDIX A: LATTICE PARAMETERS

Table II lists the lattice parameters for $R\mathrm{Co}_{12}$ that we used in the calculations. We assumed the $ThMn_{12}$ structure [space group $I4/mmm$ (no. 139)] for the system. The definitions of $p_{8\mathrm{i}}$ and $p_{8\mathrm{j}}$ are summarized in Table III with representable atomic positions of the atoms.

In our KKR-CPA calculations for $R\mathrm{Fe}_{11}\mathrm{Ti}$ ($R =$ Y, Nd, Sm) and $Z\mathrm{Fe}_{11}\mathrm{Ti}$ ($Z =$ Zr, Dy), we assumed that they also have the $ThMn_{12}$ structure. The lattice parameters are reduced from the structure obtained in Ref. [62] and the

TABLE III. Representable atomic positions in the $ThMn_{12}$ structure. The variables x, y, and z denote the point (ax, ay, cz) in Cartesian coordinates.

Element	Site	x	y	z
Th	2a	0	0	0
Mn	8f	0.25	0.25	0.25
Mn	8i	$p_{8\mathrm{i}}$	0	0
Mn	8j	$p_{8\mathrm{j}}$	0.5	0

TABLE IV. Inner coordinates (x, y, z) of Fe and Ti in $DyFe_{11}Ti$ and $ZrFe_{11}Ti$ where Dy and Zr are placed at (0, 0, 0). The position of Ti is denoted by a parenthesis. These values correspond to the point (ax, by, cz) in Cartesian coordinates where $a = 8.455$, $b = 8.262$, and $c = 4.715$ in $ZrFe_{11}Ti$, and $a = 8.518$, $b = 8.410$, and $c = 4.727$ in $DyFe_{11}Ti$.

	Fe(8f)			Fe(8i)			Fe(8j)		
	x	y	z	x	y	z	x	y	z
$DyFe_{11}Ti$	0.255	0.251	0.251	(0.375	0.000	0.000)	0.272	0.500	0.000
	0.255	0.749	0.749	−0.348	0.000	0.000	−0.266	0.500	0.000
	0.755	0.249	0.751	0.005	0.355	0.000	0.506	0.286	0.000
	0.755	0.751	0.249	0.005	−0.355	0.000	0.506	−0.286	0.000
$ZrFe_{11}Ti$	0.256	0.250	0.251	(0.381	0.000	0.000)	0.276	0.500	0.000
	0.256	0.750	0.750	−0.342	0.000	0.000	−0.264	0.500	0.000
	0.756	0.250	0.751	0.007	0.352	0.000	0.507	0.300	0.000
	0.756	0.750	0.250	0.007	−0.352	0.000	0.507	−0.300	0.000

structure given in Table IV. The value of $\sqrt{ab}$ is used as the a parameter for the reduced cell to keep the cell volume. The inner parameters are determined so that they minimize the deviation in the space of the coefficients (x, y, z) of the unit vectors, which corresponds to the point (ax, by, cz) in the Cartesian coordinates. The reduced values are listed in Table V.

APPENDIX B: INTEGRATION MODEL WITH A LINEAR TERM

In this Appendix, we incorporate an adjustable linear term into the relation of $E[\vec{x}]$ and $E'[\vec{x}]$, which appear in Sec. III C: we consider $E'[\vec{x}]$ as an approximate function of $E[\vec{x}] + \vec{a} \cdot \vec{x} + b$ in this Appendix.

This changes Eq. (3) to

$$\tilde{E}_i[\vec{y}] - E\left[\vec{x}_i^R\right] = E'[\vec{y}] - E'\left[\vec{x}_i^R\right] + \vec{a} \cdot \left(\vec{y} - \vec{x}_i^R\right) + \varepsilon_i. \quad \text{(B1)}$$

Therefore, our model does not depend on the variable b.

Equation (7) is modified as follows:

$$\tilde{\mu} = E'[\vec{y}] + \frac{1}{\Omega} \sum_i \omega_i \left\{E\left[\vec{x}_i^R\right] - E'\left[\vec{x}_i^R\right] + \vec{a} \cdot \left(\vec{y} - \vec{x}_i^R\right)\right\}, \quad \text{(B2)}$$

while Eq. (8) is left unchanged.

TABLE V. Reduced values of the lattice parameters for $RFe_{11}Ti$ (R = Y, Nd, Sm) and $ZFe_{11}Ti$ (Z = Zr, Dy). See Table III for definitions of p_{8i} and p_{8j}.

Z	a (Å)	c (Å)	p_{8i}	p_{8j}
Y	8.476	4.730	0.3606	0.2764
Nd	8.560	4.701	0.3596	0.2703
Sm	8.523	4.713	0.3590	0.2728
Zr	8.358	4.715	0.3565	0.2850
Dy	8.464	4.727	0.3584	0.2776

The estimation by the maximum likelihood method with $\tilde{E}_{\mathrm{LOO},i}$ described in Sec. III C is applicable to the variable $\vec{a}$ and σ^2. The equation for $\vec{a}$ is

$$\left(\sum_i \frac{1}{\Omega_{\mathrm{LOO},i}} \vec{\xi}_i \otimes \vec{\xi}_i\right)\vec{a} = \sum_i \left(E\left[\vec{x}_i^R\right] - E'\left[\vec{x}_i^R\right] - \frac{1}{\Omega_{\mathrm{LOO},i}} \sum_{k\neq i} \omega_k \left\{E\left[\vec{x}_k^R\right] - E'\left[\vec{x}_k^R\right]\right\}\right), \quad \text{(B3)}$$

where $\vec{\xi}_i$ is defined as

$$\vec{\xi}_i \equiv \sum_{k\neq i} \omega_k \left\{\vec{x}_i^R - \vec{x}_k^R\right\} \quad \text{(B4)}$$

and the symbol $\otimes$ denotes the dyadic product, with which $(\alpha_1, \alpha_2, \alpha_3) \otimes (\beta_1, \beta_2, \beta_3)$ is defined as the matrix whose (i, j) component is $\alpha_i \beta_j$. This equation can be solved for $\vec{a}$ without determining σ^2.

The equation for σ^2 is modified as

$$\sigma^2 = \frac{1}{M} \sum_i \Omega_{\mathrm{LOO},i} \left(E\left[\vec{x}_i^R\right] - E'\left[\vec{x}_i^R\right] - \frac{\sum_{k\neq i} \omega_k \left\{E\left[\vec{x}_k^R\right] - E'\left[\vec{x}_k^R\right] + \vec{a} \cdot \left(\vec{x}_i^R - \vec{x}_k^R\right)\right\}}{\Omega_{\mathrm{LOO},i}}\right)^2.$$

APPENDIX C: DIMENSIONS FOR *R* AND *Z*

We prepared dimensions to include information of the R and Z elements in our design of the descriptors in Table I. Because the corresponding choice is only 6 in combination (R = Y, Nd, Sm; Z = Zr, Du) while it is known that high-dimensionality often causes problems in modeling, readers may doubt its necessity. We compare the efficiency of the search for the smaller space in which those elements are fixed to R = Y and Z = Zr in Fig. 13. The solid line along the cross symbols in the figures denote the frequency of finding the system with the best score out of 1000 sessions in the restricted space when we use the set of α, β, and γ as a descriptor. We also show the curve elongated six times along the x-axis (the dotted line) because one has to optimize also for the other combinations of R and Z to obtain the optimal

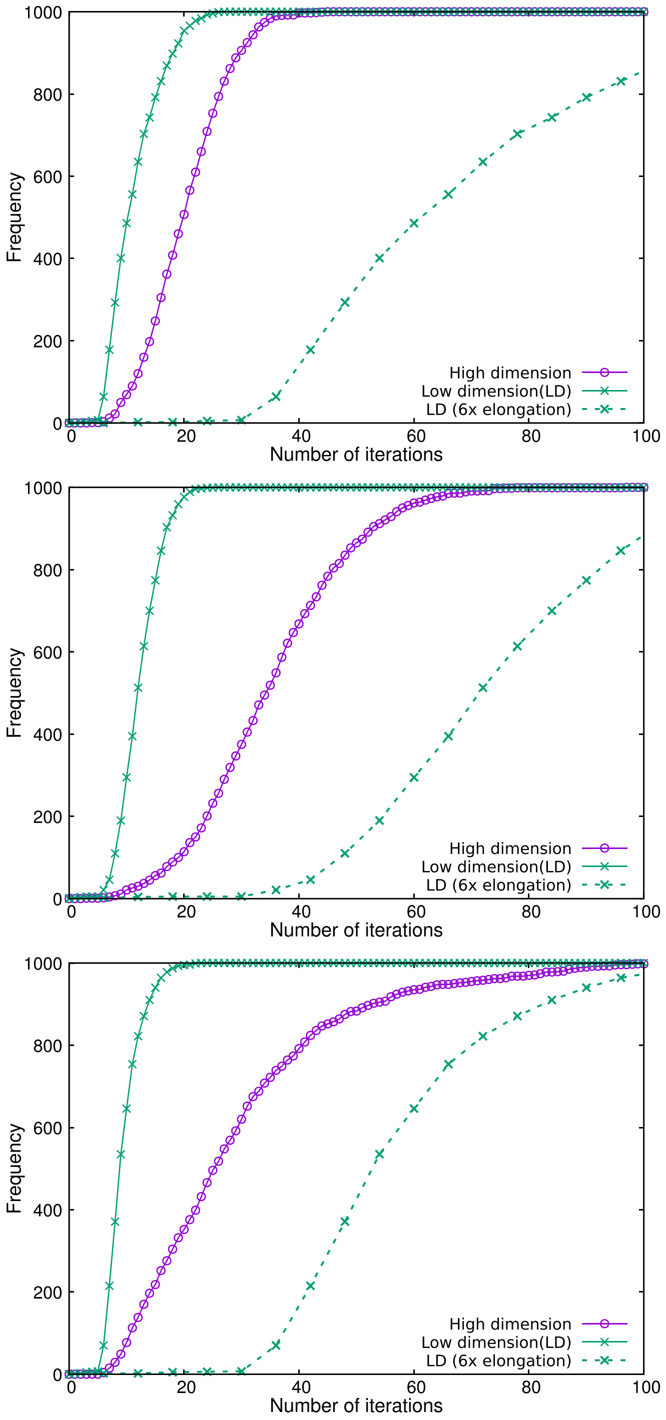

FIG. 13. Frequency of finding the system with the highest magnetization (top), Curie temperature (middle), and the lowest formation energy (bottom) among 1000 sessions. Those in the search for the small space where R and Z are fixed to $R =$ Y and $Z =$ Zr are denoted by the solid line along the cross symbols. The dotted line along the cross symbols shows the curve elongated six times along the x-axis. The line along the circle symbols shows the result of the search in the full space using the descriptor no. 9 in Table I.

TABLE VI. The top 10 systems with the highest values of the magnetization. The magnetic moment from the f-electrons is assumed to be 3.27 μ_B for Nd, 0.71 μ_B for Sm, and $-10\,\mu_B$ for Dy. The variable $\mu_0 M$ denotes the magnetization, T_C is the Curie temperature, and ΔE is the formation energy from the unary systems. The values for $Nd(Fe_{0.7}Co_{0.3})_{12}$, which has the highest magnetization among the systems with a negative formation energy, and the values for $(Sm_{0.7}Zr_{0.3})(Fe_{0.9}Co_{0.1})_{12}$, which has the highest magnetization among the systems having Sm in them with a negative formation energy, are also shown.

Formula	$\mu_0 M$ (T)	T_C (K)	ΔE (eV)
$NdFe_{12}$	1.95	844	0.405
$Nd(Fe_{0.9}Co_{0.1})_{12}$	1.94	1012	0.230
$Nd(Fe_{0.8}Co_{0.2})_{12}$	1.93	1111	0.095
$(Nd_{0.9}Zr_{0.1})Fe_{12}$	1.93	835	0.443
$(Nd_{0.9}Zr_{0.1})(Fe_{0.9}Co_{0.1})_{12}$	1.92	1011	0.272
$(Nd_{0.9}Zr_{0.1})(Fe_{0.8}Co_{0.2})_{12}$	1.91	1098	0.140
$(Nd_{0.8}Zr_{0.2})Fe_{12}$	1.91	841	0.446
$(Nd_{0.8}Zr_{0.2})(Fe_{0.9}Co_{0.1})_{12}$	1.91	1009	0.272
$(Nd_{0.7}Zr_{0.3})Fe_{12}$	1.89	845	0.385
$(Nd_{0.8}Zr_{0.2})(Fe_{0.8}Co_{0.2})_{12}$	1.89	1086	0.143
$Nd(Fe_{0.7}Co_{0.3})_{12}$	1.89	1143	−0.142
$(Sm_{0.7}Zr_{0.3})(Fe_{0.9}Co_{0.1})_{12}$	1.77	1002	−0.008

system for the full space, which takes approximately six times larger. For comparison, we show the result of optimization for the full space using the descriptor no. 9 by the line along the circle points. We set the number of iterations before Bayesian optimization to 5, which is smaller than the number we use above (=10), because we know that the full-space search is so fast that the search in the smaller space cannot catch up if it starts 60 (=10 × 6) iterations behind (Figs. 10 and 11). We see from Fig. 13 that the search with the full space is more efficient, even with this setup, than repeating the search for the small space six times. Therefore, the dimensions for R and Z actually contribute to the search efficiency.

APPENDIX D: DATA FROM THE FIRST-PRINCIPLES CALCULATIONS

In this Appendix, we list systems with the predicted highest values of magnetization and Curie temperature in order to make a comparison with available experimental data

TABLE VII. The top 10 systems with the highest values of the Curie temperature. The variable $\mu_0 M$ denotes the magnetization, T_C is the Curie temperature, and ΔE is the formation energy from the unary systems.

Formula	T_C (K)	$\mu_0 M$ (T)	ΔE (eV)
$Sm(Fe_{0.2}Co_{0.8})_{12}$	1310	1.47	−0.631
$(Sm_{0.9}Dy_{0.1})(Fe_{0.2}Co_{0.8})_{12}$	1309	1.39	−0.654
$(Sm_{0.8}Dy_{0.2})(Fe_{0.2}Co_{0.8})_{12}$	1307	1.32	−0.676
$(Sm_{0.7}Dy_{0.3})(Fe_{0.2}Co_{0.8})_{12}$	1305	1.24	−0.696
$(Sm_{0.6}Dy_{0.4})(Fe_{0.2}Co_{0.8})_{12}$	1304	1.17	−0.715
$(Sm_{0.5}Dy_{0.5})(Fe_{0.2}Co_{0.8})_{12}$	1302	1.09	−0.733
$(Sm_{0.4}Dy_{0.6})(Fe_{0.2}Co_{0.8})_{12}$	1300	1.01	−0.752
$(Nd_{0.7}Dy_{0.3})(Fe_{0.2}Co_{0.8})_{12}$	1300	1.37	−0.553
$(Nd_{0.6}Dy_{0.4})(Fe_{0.2}Co_{0.8})_{12}$	1299	1.27	−0.594
$(Nd_{0.8}Dy_{0.2})(Fe_{0.2}Co_{0.8})_{12}$	1299	1.46	−0.510

and to serve as guiding information for future experimental synthesis.

In Table VI, we show the top 10 systems of the highest magnetization. Here we add the contribution of the f-electrons of Nd, Sm, and Dy with the assumption that those have $g_J J$ of $3.27\mu_B$, $0.71\mu_B$, and $-10\mu_B$. Due to this additional magnetic moment, the systems with Nd have an advantage over the other systems, and all the top 10 systems have Nd in them.

It should also be noted that the values for the formation energy are positive for all the top 10 systems, and violate a necessary condition for the thermodynamic stability. The highest magnetization among the systems with a negative value of the formation energy is obtained by $\mathrm{Nd(Fe_{0.7}Co_{0.3})_{12}}$. Note also that this does not ensure the stability against the other phases. We also show the values for $\mathrm{(Sm_{0.7}Zr_{0.3})(Fe_{0.9}Co_{0.1})_{12}}$ because the magnetic anisotropy of Nd tends not to be uniaxial in $\mathrm{ThMn_{12}}$-type systems in the absence of a third element.

The best system, $\mathrm{NdFe_{12}}$, has already been synthesized by Hirayama *et al.* [39]. From the results of our calculations, this system seems to be near the upper limit of magnetization at absolute zero.

In Table VII, we show the top 10 systems with the highest values of Curie temperature. Although the formation energy is negative for all the systems in the table, it should be noted again that those incorporate only the competition with the unary phases, and it does not ensure stability against the other phases.

The best system in the list is $\mathrm{Sm(Fe_{0.2}Co_{0.8})_{12}}$. Although Hirayama *et al.* have reported that they could synthesize $\mathrm{Sm(Fe_{0.8}Co_{0.2})_{12}}$ as a film, to our knowledge there is no experimental report for a successful synthesis of compounds with a higher concentration of Co.

[1] M. Rupp, A. Tkatchenko, K.-R. Müller, and O. A. Von Lilienfeld, Fast and Accurate Modeling of Molecular Atomization Energies with Machine Learning, Phys. Rev. Lett. **108**, 058301 (2012).

[2] L. M. Ghiringhelli, J. Vybiral, S. V. Levchenko, C. Draxl, and M. Scheffler, Big Data of Materials Science: Critical Role of the Descriptor, Phys. Rev. Lett. **114**, 105503 (2015).

[3] A. Seko, A. Togo, H. Hayashi, K. Tsuda, L. Chaput, and I. Tanaka, Prediction of Low-Thermal-Conductivity Compounds with First-Principles Anharmonic Lattice-Dynamics Calculations and Bayesian Optimization, Phys. Rev. Lett. **115**, 205901 (2015).

[4] T. L. Pham, H. Kino, K. Terakura, T. Miyake, and H. C. Dam, Novel mixture model for the representation of potential energy surfaces, J. Chem. Phys. **145**, 154103 (2016).

[5] T. Lam Pham, H. Kino, K. Terakura, T. Miyake, K. Tsuda, I. Takigawa, and H. Chi Dam, Machine learning reveals orbital interaction in materials, Sci. Technol. Adv. Mater. **18**, 756 (2017).

[6] H. C. Dam, V. C. Nguyen, T. L. Pham, A. T. Nguyen, K. Terakura, T. Miyake, and H. Kino, Important descriptors and descriptor groups of Curie temperatures of rare-earth transition-metal binary alloys, J. Phys. Soc. Jpn. **87**, 113801 (2018).

[7] T.-L. Pham, N.-D. Nguyen, V.-D. Nguyen, H. Kino, T. Miyake, and H.-C. Dam, Learning structure-property relationship in crystalline materials: A study of lanthanide–transition metal alloys, J. Chem. Phys. **148**, 204106 (2018).

[8] R. Ouyang, S. Curtarolo, E. Ahmetcik, M. Scheffler, and L. M. Ghiringhelli, SISSO: A compressed-sensing method for identifying the best low-dimensional descriptor in an immensity of offered candidates, Phys. Rev. Mater. **2**, 083802 (2018).

[9] A. Seko, H. Hayashi, H. Kashima, and I. Tanaka, Matrix- and tensor-based recommender systems for the discovery of currently unknown inorganic compounds, Phys. Rev. Mater. **2**, 013805 (2018).

[10] T. Ueno, T. D. Rhone, Z. Hou, T. Mizoguchi, and K. Tsuda, COMBO: An efficient Bayesian optimization library for materials science, Mater. Discovery **4**, 18 (2016).

[11] S. Ju, T. Shiga, L. Feng, Z. Hou, K. Tsuda, and J. Shiomi, Designing Nanostructures for Phonon Transport Via Bayesian Optimization, Phys. Rev. X **7**, 021024 (2017).

[12] S. Kikuchi, H. Oda, S. Kiyohara, and T. Mizoguchi, Bayesian optimization for efficient determination of metal oxide grain boundary structures, Physica B **532**, 24 (2018).

[13] R. Yuan, Z. Liu, P. V. Balachandran, D. Xue, Y. Zhou, X. Ding, J. Sun, D. Xue, and T. Lookman, Accelerated discovery of large electrostrains in $\mathrm{BaTiO_3}$-based piezoelectrics using active learning, Adv. Mater. **30**, 1702884 (2018).

[14] A. R. Oganov and M. Valle, How to quantify energy landscapes of solids, J. Chem. Phys. **130**, 104504 (2009).

[15] M. C. Kennedy and A. O'Hagan, Predicting the output from a complex computer code when fast approximations are available, Biometrika **87**, 1 (2000).

[16] G. Pilania, J. E. Gubernatis, and T. Lookman, Multi-fidelity machine learning models for accurate bandgap predictions of solids, Comput. Mater. Sci. **129**, 156 (2017).

[17] S. J. Pan, Q. Yang *et al.*, A survey on transfer learning, IEEE Trans. Knowl. Data Eng. **22**, 1345 (2010).

[18] M. L. Hutchinson, A. Erin, B. M. Gibbons, S. Paradiso, J. Ling, and B. Meredig, Overcoming data scarcity with transfer learning, arXiv:1711.05099, http://www.quantum-machine.org/workshops/nips2017/.

[19] K. Ohashi, Y. Tawara, R. Osugi, J. Sakurai, and Y. Komura, Identification of the intermetallic compound consisting of Sm, Ti, Fe, J. Less-Common Met. **139**, L1 (1988).

[20] K. Ohashi, Y. Tawara, R. Osugi, and M. Shimao, Magnetic properties of Fe-rich rare-earth intermetallic compounds with a $\mathrm{ThMn_{12}}$ structure, J. Appl. Phys. **64**, 5714 (1988).

[21] Y.-C. Yang, L.-S. Kong, S.-H. Sun, D.-M. Gu, and B.-P. Cheng, Intrinsic magnetic properties of $\mathrm{SmTiFe_{10}}$, J. Appl. Phys. **63**, 3702 (1988).

[22] R. Verhoef, F. R. De Boer, Z.-D. Zhang, and K. H. J. Buschow, Moment reduction in $\mathrm{RFe}_{12-x}\mathrm{T}_x$ compounds (R = Gd, Y and T= Ti, Cr, V, Mo, W), J. Magn. Magn. Mater. **75**, 319 (1988).

[23] D. B. De Mooij and K. H. J. Buschow, Some novel ternary $\mathrm{ThMn_{12}}$-type compounds, J. Less-Common Met. **136**, 207 (1988).

[24] K. H. J. Buschow, Structure and properties of some novel ternary Fe-rich rare-earth intermetallics (invited), J. Appl. Phys. **63**, 3130 (1988).

[25] S. S. Jaswal, Y. G. Ren, and D. J. Sellmyer, Electronic structure and magnetic properties of Fe-rich ternary compounds: $YFe_{10}V_2$ and $YFe_{10}Cr_2$, J. Appl. Phys. **67**, 4564 (1990).

[26] R. Coehoorn, Electronic structure and magnetism of transition-metal-stabilized $YFe_{12-x}M_x$ intermetallic compounds, Phys. Rev. B **41**, 11790 (1990).

[27] K. H. J. Buschow, Permanent magnet materials based on tetragonal rare earth compounds of the type $RFe_{12-x}M_x$, J. Magn. Magn. Mater. **100**, 79 (1991).

[28] S. Sakurada, A. Tsutai, and M. Sahashi, A study on the formation of $ThMn_{12}$ and $NaZn_{13}$ structures in $RFe_{10}Si_2$, J. Alloys Compd. **187**, 67 (1992).

[29] A. Sakuma, Self-consistent band calculations for $YFe_{11}Ti$ and $YFe_{11}TiN$, J. Phys. Soc. Jpn. **61**, 4119 (1992).

[30] S. Asano, S. Ishida, and S. Fujii, Electronic structures and improvement of magnetic properties of $RFe_{12}X$ (R = Y, Ce, Gd; X = N, C), Physica B **190**, 155 (1993).

[31] M. Akayama, H. Fujii, K. Yamamoto, and K. Tatami, Physical properties of nitrogenated $RFe_{11}Ti$ intermetallic compounds (R=Ce, Pr and Nd) with $ThMn_{12}$-type structure, J. Magn. Magn. Mater. **130**, 99 (1994).

[32] M. D. Kuz'min, M. Richter, and K. H. J. Buschow, Iron-rich versus cobalt-rich $ThMn_{12}$-type intermetallics: A comparative study of the crystal fields, Solid State Commun. **113**, 47 (1999).

[33] A. M. Gabay, A. Martín-Cid, J. M. Barandiaran, D. Salazar, and G. C. Hadjipanayis, Low-cost $Ce_{1-x}Sm_x(Fe, Co, Ti)_{12}$ alloys for permanent magnets, AIP Adv. **6**, 056015 (2016).

[34] W. Körner, G. Krugel, and C. Elsässer, Theoretical screening of intermetallic $ThMn_{12}$-type phases for new hard-magnetic compounds with low rare earth content, Sci. Rep. **6**, 24686 (2016).

[35] L. Ke and D. D. Johnson, Intrinsic magnetic properties in $R(Fe_{1-x}Co_x)_{11}TiZ$ (R = Y and Ce; Z = H, C, and N), Phys. Rev. B **94**, 024423 (2016).

[36] T. Fukazawa, H. Akai, Y. Harashima, and T. Miyake, First-principles study of intersite magnetic couplings in $NdFe_{12}$ and $NdFe_{12}X$ (X= B, C, N, O, F), J. Appl. Phys. **122**, 053901 (2017).

[37] T. Fukazawa, H. Akai, Y. Harashima, and T. Miyake, First-principles study of spin-wave dispersion in $Sm(Fe_{1-x}Co_x)_{12}$, J. Magn. Magn. Mater. **469**, 296 (2019).

[38] T. Miyake, K. Terakura, Y. Harashima, H. Kino, and S. Ishibashi, First-principles study of magnetocrystalline anisotropy and magnetization in $NdFe_{12}$, $NdFe_{11}Ti$, and $NdFe_{11}TiN$, J. Phys. Soc. Jpn. **83**, 043702 (2014).

[39] Y. Hirayama, Y. K. Takahashi, S. Hirosawa, and K. Hono, $NdFe_{12}N_x$ hard-magnetic compound with high magnetization and anisotropy field, Scr. Mater. **95**, 70 (2015).

[40] Y. Hirayama, T. Miyake, and K. Hono, Rare-earth lean hard magnet compound $NdFe_{12}N$, JOM **67**, 1344 (2015).

[41] Y.-C. Yang, X.-D. Zhang, L.-S. Kong, Q. Pan, and S.-L. Ge, New potential hard magnetic material—$NdTiFe_{11}N_x$, Solid State Commun. **78**, 317 (1991).

[42] Y.-C. Yang, X.-D. Zhang, S.-L. Ge, Q. Pan, L.-S. Kong, H. Li, J.-L. Yang, B.-S. Zhang, Y.-F. Ding, and C.-T. Ye, Magnetic and crystallographic properties of novel Fe-rich rare-earth nitrides of the type $RTiFe_{11}N_{1-\delta}$, J. Appl. Phys. **70**, 6001 (1991).

[43] Y. Harashima, K. Terakura, H. Kino, S. Ishibashi, and T. Miyake, First-principles study on stability and magnetism of $NdFe_{11}M$ and $NdFe_{11}MN$ for M= Ti, V, Cr, Mn, Fe, Co, Ni, Cu, Zn, J. Appl. Phys. **120**, 203904 (2016).

[44] Y. Hirayama, Y. K. Takahashi, S. Hirosawa, and K. Hono, Intrinsic hard magnetic properties of $Sm(Fe_{1-x}Co_x)_{12}$ compound with the $ThMn_{12}$ structure, Scr. Mater. **138**, 62 (2017).

[45] S. Suzuki, T. Kuno, K. Urushibata, K. Kobayashi, N. Sakuma, K. Washio, H. Kishimoto, A. Kato, and A. Manabe, A $(Nd, Zr)(Fe, Co)_{11.5}Ti_{0.5}N_x$ compound as a permanent magnet material, AIP Adv. **4**, 117131 (2014).

[46] N. Sakuma, S. Suzuki, T. Kuno, K. Urushibata, K. Kobayashi, M. Yano, A. Kato, and A. Manabe, Influence of Zr substitution on the stabilization of $ThMn_{12}$-type $(Nd_{1-\alpha}Zr_\alpha)(Fe_{0.75}Co_{0.25})_{11.25}Ti_{0.75}N_{1.2\text{–}1.4}$ ($\alpha = 0\text{–}0.3$), AIP Adv. **6**, 056023 (2016).

[47] T. Kuno, S. Suzuki, K. Urushibata, K. Kobayashi, N. Sakuma, M. Yano, A. Kato, and A. Manabe, $(Sm, Zr)(Fe, Co)_{11.0\text{–}11.5}Ti_{1.0\text{–}0.5}$ compounds as new permanent magnet materials, AIP Adv. **6**, 025221 (2016).

[48] S. Suzuki, T. Kuno, K. Urushibata, K. Kobayashi, N. Sakuma, K. Washio, M. Yano, A. Kato, and A. Manabe, A new magnet material with $ThMn_{12}$ structure: $(Nd_{1-x}Zr_x)(Fe_{1-y}Co_y)_{11+z}Ti_{1-z}N_\alpha(\alpha = 0.6\text{–}1.3)$, J. Magn. Magn. Mater. **401**, 259 (2016).

[49] Y. Harashima, T. Fukazawa, H. Kino, and T. Miyake, Effect of R-site substitution and pressure on stability of RFe_{12}: A first-principles study, J. Appl. Phys. **124**, 163902 (2018).

[50] COMmon Bayesian Optimization Library (COMBO), https://github.com/tsudalab/combo.

[51] P. Hohenberg and W. Kohn, Inhomogeneous electron gas, Phys. Rev. **136**, B864 (1964).

[52] W. Kohn and L. J. Sham, Self-consistent equations including exchange and correlation effects, Phys. Rev. **140**, A1133 (1965).

[53] J. Jensen and A. R. Mackintosh, *Rare Earth Magnetism* (Clarendon, Oxford, 1991).

[54] M. Richter, Band structure theory of magnetism in 3d-4f compounds, J. Phys. D **31**, 1017 (1998).

[55] I. L. M. Locht, Y. O. Kvashnin, D. C. M. Rodrigues, M. Pereiro, A. Bergman, L. Bergqvist, A. I. Lichtenstein, M. I. Katsnelson, A. Delin, A. B. Klautau, B. Johansson, I. Di Marco, and O. Eriksson, Standard model of the rare earths analyzed from the Hubbard I approximation, Phys. Rev. B **94**, 085137 (2016).

[56] J. P. Perdew and A. Zunger, Self-interaction correction to density-functional approximations for many-electron systems, Phys. Rev. B **23**, 5048 (1981).

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

[58] G. Kresse and D. Joubert, From ultrasoft pseudopotentials to the projector augmented-wave method, Phys. Rev. B **59**, 1758 (1999).

[59] QMAS—Quantum MAterials Simulator Official Site, http://qmas.jp.

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

[61] T. Fukazawa, H. Akai, Y. Harashima, and T. Miyake, First-principles study of intersite magnetic couplings and Curie temperature in $RFe_{12-x}Cr_x$ (R = Y, Nd, Sm), J. Phys. Soc. Jpn. **87**, 044706 (2018).

[62] Y. Harashima, K. Terakura, H. Kino, S. Ishibashi, and T. Miyake, First-principles study of structural and magnetic properties of R(Fe, Ti)$_{12}$ and R(Fe, Ti)$_{12}$N (R = Nd, Sm, Y), in Proceedings of Computational Science Workshop 2014 (CSW2014), JPS Conf. Proc. **5**, 011021 (2015).

[63] P. Soven, Coherent-potential model of substitutional disordered alloys, Phys. Rev. **156**, 809 (1967).

[64] P. Soven, Application of the coherent potential approximation to a system of Muffin-Tin potentials, Phys. Rev. B **2**, 4715 (1970).

[65] H. Shiba, A reformulation of the coherent potential approximation and its applications, Prog. Theor. Phys. **46**, 77 (1971).

[66] A. I. Liechtenstein, M. I. Katsnelson, V. P. Antropov, and V. A. Gubanov, Local spin density functional approach to the theory of exchange interactions in ferromagnetic metals and alloys, J. Magn. Magn. Mater. **67**, 65 (1987).

[67] T. Fukazawa, H. Akai, Y. Harashima, and T. Miyake, Curie temperature of Sm_2Fe_{17} and $Nd_2Fe_{14}B$: A first-principles study, IEEE Trans. Magn., 1 (2019), doi: 10.1109/TMAG.2019.2895669.

[68] P. H. Dederichs, R. Zeller, H. Akai, and H. Ebert, Ab-initio calculations of the electronic structure of impurities and alloys of ferromagnetic transition metals, J. Magn. Magn. Mater. **100**, 241 (1991).

[69] C. Takahashi, M. Ogura, and H. Akai, First-principles calculation of the Curie temperature Slater–Pauling curve, J. Phys.: Condens. Matter **19**, 365233 (2007).

[70] Y. Harashima, K. Terakura, H. Kino, S. Ishibashi, and T. Miyake, Nitrogen as the best interstitial dopant among X=B, C, N, O, and F for strong permanent magnet $NdFe_{11}TiX$: First-principles study, Phys. Rev. B **92**, 184426 (2015).