Efficient autonomous material search method combining \textit{ab initio} calculations, autoencoder, and multi-objective Bayesian optimization

Yuma Iwasaki, Hwang Jaekyun, Yuya Sakuraba, Masato Kotsugi and Yasuhiro Igarashi

\textbf{ABSTRACT}

Autonomous material search systems that combine \textit{ab initio} calculations and Bayesian optimization are very promising for exploring huge material spaces. Setting up an appropriate material search space is necessary for efficient autonomous material search. However, performing the autonomous search within the material space set up using the prepared descriptors is not sufficient to obtain an efficient search, which can be achieved by prioritizing specific descriptors and properties. Here, a material search system that can autonomously search the huge material space while performing multi-objective optimization that considers similarities among elements and emphasizes specific descriptors is proposed. This system has been used for a material exploration of Heusler alloys. The system has successfully proposed several candidate materials with half-metallic properties. The proposed system is very versatile and can be applied to various properties and material systems.

\textbf{ARTICLE HISTORY}

Received 14 July 2022
Revised 25 August 2022
Accepted 07 September 2022

\textbf{KEYWORDS}

Heusler alloy; \textit{ab initio} calculation; half metal; bayesian optimization

\section{Introduction}

In recent years, functional materials have gradually become multi-elemental, resulting in a significant increase in their numbers. In conventional material exploration methods, the search for these materials is conducted by humans through repeated material syntheses, measurements, and considerations. However, it has become extremely difficult to explore the huge material search space using only the conventional search for new materials. Therefore, autonomous material search methods have been developed by integrating data science with conventional material-development methods. These methods can be divided into two major categories. The first category is the robotic autonomous material search methods, which combine materials experiments, robotics, and machine learning [1–9]. The robot performs actual material synthesis and measurements, whereas machine learning determines the next candidate material and process conditions, which enables autonomous material synthesis. The other category is an in \textit{silico} autonomous material search method that combines \textit{ab initio} calculations and machine learning [10–15]. Using \textit{ab initio} calculations on compositions and structures determined by machine learning enables efficient autonomous material search.

Heusler alloys are an important application of the autonomous material search systems. They have attracted...
much attention in recent years because they exhibit various physical properties [16–19]. One of the most attractive physical properties is half-metallicity, which has one metallic spin band and one semiconducting spin band [20]. They have been extensively studied for applications in tunnel magnetoresistance [21–24], giant magnetoresistance [25–28], and other spintronics devices [29–31]. However, since the material search space is extremely vast due to their multi-elemental nature and their disordered phases, the development of Heusler alloys is very difficult.

For example, Sawada et al. developed an autonomous material search method to search for Heusler alloys with half-metallic properties [11]. In this method, an autonomous material search was conducted to maximize the spin polarization. However, maximizing the spin polarization alone is not sufficient for practical applications. Therefore, other material properties, such as the half-metallic gap and Curie temperature, should also be considered. In addition, material science scholars have developed new functional materials by considering the similarity among constituent elements and the influence of crystal structure. However, the previous method [11] had not utilized such ideas in their materials search; thus, a more efficient autonomous material search method should be developed.

In this study, we proposed a material search system that can autonomously search the material space using multi-objective optimization of multiple material properties while considering similarities among elements and emphasizing specific descriptors. This method is then used in the search for Heusler alloys with half-metallic properties.

### 2. The proposed method

Figure 1 shows an overview of the proposed method, in which the ab initio calculations and machine learning are sequentially repeated to autonomously search for materials. In the ab initio calculation part, the properties of the Heusler alloy are calculated according to the crystal structure and composition information determined in the machine learning part. The machine learning part derives information on the crystal structure and composition of the material from the accumulated Heusler alloy data to be used in the next ab initio calculations. The details of each part are described below.

#### 2.1. Ab initio calculations

In the ab initio calculations, Green’s function-based density functional theory calculations were conducted by the Korringa-Kohn-Rostoker coherent potential approximation (KKR-CPA) method using the AkaiKKR software [32]. Multi-elemental disordered phases can be calculated using the CPA, which can simulate them with high accuracy, especially in alloy systems [33–35]. Figure 2(a,b) show two major crystal structures of Heusler alloys: the full-Heusler structure (L21) and the half-Heusler structure (C15) illustrated by using VESTA [36]. Typically, sites A, B, and C have different elements. Here, based on the information on the crystal structures and composition of the materials, which are determined in the machine learning part described later, the KKR-CPA was conducted to calculate the spin polarization, $P$, and the half-metallic gap, $G$. Figure 2(c) shows an image of the density of state (DoS) for half-metallic materials. $P$ is calculated using Equation 1.

$$P = \frac{n^↑ - n^↓}{n^↑ + n^↓}$$  

(1)

where $n^↑$ and $n^↓$ are the up spin and down spin DoS at fermi energy, $E_F$, respectively. The half-metallic gap, $G$, is equivalent to the band gap of the down spin,
which is marked by the orange arrow in Figure 2(c). Larger G and P result in better half-metallic Heusler alloys. The lattice constants were determined to minimize the total energy. In the lattice constant optimization, the spin–orbit interactions and the relativistic effects were not considered. The DoS was then calculated using the optimized lattice constant. In these calculations, spin–orbit interaction and relativistic effects were taken into account. Thus, P and G are estimated based on the DoS calculated here. More details are given in the Supplementary Materials (S1).

2.2. Machine learning

In the machine learning part, Bayesian optimization [37], an autoencoder [38], and Pareto optimization [39] were used to determine the crystal structure and composition of the materials for the next KKR-CPA calculation on the basis of the accumulated data on P and G.

In this study, the material space of the Heusler alloys to be explored by the autonomous material search system was defined by Equations 2 and 3.

\[
\text{Full Heusler alloy} (S_3) : A_1^1A_2^2B_1^1B_2^2C_1^1C_2^2 \quad (2)
\]

\[
\text{Half Heusler alloy} (S_1) : A_1^1A_2^2B_1^1B_2^2C_1^1C_2^2 \quad (3)
\]

where

\[A^1, A^2 = \{Fe, Co, Ni, Cu, Zn, Ru, Rh, Pd, Ag, Cd, Ir, Pt, Au\}\]

\[B^1, B^2 = \{Ti, V, Cr, Mn, Y, Zr, Nb, Hf, Ta\}\]

\[C^1, C^2 = \{Al, Si, Ga, Ge, As, In, Sn, Sb, Tl, Pb, Bi\}\]

\[a = \{0, 0.2, 0.4, 0.6, 0.8, 1.0, 1.2, 1.4, 1.6, 1.8, 2.0\}\]

\[\hat{a} \hat{b} \hat{c} = \{0, 0.2, 0.4, 0.6, 0.8, 1.0\}\]

Figure 2(a) shows the periodic table illustrating compositional patterns of the material space. The A-site elements (A1, A2), B-site elements (B1, B2), and C-site elements (C1, C2), which are shown in Figure 2(a,b), were selected from the yellow, red, and blue regions, respectively, of the periodic table in Figure 2(a). The composition was set in increments of 0.2. The material was intentionally restricted to contain atoms of either Fe, Co, or Ni to reduce the total number of candidate materials. In the material space defined here, there are approximately 10 million candidate material patterns. It is quite difficult to carry out the KKR-CPA calculations for all these candidate materials. Thus, the KKR-CPA calculations were performed sequentially using machine learning to search the material space.

First, to create descriptors (explanatory variables) to represent this material space, a composition vector \(C\), a Magpie descriptor vector \(M\) and a crystal structure vector \(S\) were prepared. The composition vector \(C\) indicates the proportions of the elements in the material. Sites A, B, and C have 13, 9, and 11 elements, respectively, making the composition vector of 33 dimensions, as shown in the periodic table in Figure 3(a). \(M\) is composed of 28 values (e.g. average atomic number \(M_{\text{an}}\), average atomic weight \(M_{\text{aw}}\), and melting point \(M_{\text{mp}}\), etc.) that can be calculated from the composition vectors. The list of the Magpie descriptors is given in the Supplementary Materials (S2). By using the Magpie descriptors, it is possible to set up a material space that considers similarities between elements. For example, Fe and Ru, which are located near each other on the periodic table, have similar properties, while Fe and Bi, which are far apart, have dissimilar properties. It has been reported that the addition of a Magpie descriptor improved the accuracy of the machine learning model [41]. \(S\) is a two-dimensional one-hot vector that represents the two

Figure 2. Half-Metallic Heusler alloys: (a) Full-Heusler structure \(L_2\). (b) Half-Heusler structure \(C_{1b}\). (c) Density of state of the half metals with one metallic spin band and one semiconducting spin band.
crystal structures (i.e. the full-Heusler structure $S_F$, and the half-Heusler structure $S_H$).

Even if the material space is defined as a 63-dimensional vector combining $C$, $M$, and $S$, efficient autonomous search is not possible. This can be attributed to two reasons. First, the dimension of the material space is too large. General Bayesian optimization works properly in approximately 20 dimensions or less [42]. Second, from a materials science perspective, the crystal structure information ($S$) should be considered very important. Thus, to achieve an efficient autonomous search, a material space in which the information on the crystal structure is prioritized should be defined.

Therefore, an autoencoder, which is one of the dimensionality reduction methods, was used. Figure 3(b) shows an image of the dimensionality reduction of $C$ and $M$ by the autoencoder. The information on $C$ and $M$ was compressed in the middle layer of a neural network with the same input and output layers. Here, a 10-dimensional latent variable $Z$ was created. The $Z$ contains information on both $C$ and $M$. The importance of the crystal structure information in the autonomous material search can be adjusted by changing the number of the latent variables created by the autoencoder. This means reducing the number of latent variables to achieve an autonomous material search in which the crystal structure information is prioritized. More details on the calculation conditions and results are given in the Supplementary Materials (S3).

A 12-dimensional material space is defined by combining the latent variables $Z (Z_1, Z_2, \ldots, Z_{10})$ created by the autoencoder and the one-hot vector $S (S_F, S_H)$ representing the crystal structure. This material space is explored by the combination of the KKR-CPA method and the multi-objective Bayesian optimization (Figure 3(c)). Here, the following Gaussian process regression models were constructed for $P$ and $G$, respectively.

$$P = f(Z_1, Z_2, Z_3, \ldots, Z_{10}, S_F, S_H)$$

$$G = f(Z_1, Z_2, Z_3, \ldots, Z_{10}, S_F, S_H)$$

where $P$, $G$, $Z$, and $S$ are the spin polarization, half-metallic gap, latent variables created by the autoencoder, and one-hot-vectors of the crystal structure, respectively. The upper confidential bound (UCB) for each material is calculated as an acquisition function using Gaussian regression models [43]. The candidate material with the largest Pareto hypervolume is determined from the UCB value and the material data (training data), in which $P$ and $G$ are already observed. This is then used as the target for the next KKR-CPA calculation. In this way, it is possible to autonomously search for materials with both large $P$ and $G$. More details on the process are given in the Supplementary Materials (S4).
3. Results and discussion

The autonomous material search system kept running for approximately two months to search the material space of the Heusler alloys. Figure 4(a) shows the $P$ and $G$ values of the 1783 materials explored by the autonomous material search system. The blue circles and red squares represent the half-Heusler and full-Heusler structures, respectively. The solid black lines indicate the Pareto frontier found in the 1655th exploration. Figure 4(b) shows the DoS of Co$_{0.6}$Fe$_{0.4}$Cr$_{0.2}$Mn$_{0.2}$As$_{1.0}$ with a half-Heusler structure, which is one of the materials on the Pareto frontier. The DoS indicates that the material has large $P$ and $G$ values. Thus, this autonomous material search system can autonomously search the material space and automatically suggest materials with large $P$ and $G$.

Table 1 shows a list of candidates of new Heusler alloys on the Pareto frontier in Figure 4(a). The ternary Heusler alloys (i.e., Co$_{1.0}$Cr$_{1.0}$As$_{1.0}$ and Co$_{1.0}$Ti$_{1.0}$As$_{1.0}$) have been previously studied [44]. Meanwhile, other alloys (e.g., quaternary, quinary, and senary Heusler alloys) have not been investigated. All of them have half-Heusler structures. It is previously reported that the $G$ of half-Heusler alloys (CoMnSb, NiMnSb, etc.) is larger than that of full-Heusler alloys (Co$_2$MnSb, Co$_2$MnGa, etc.) because the half-Heusler structures have vacancies in the C sites (Figure 2(b)) [45]. The results obtained by examining the vast material space of Heusler alloys with the autonomous material search are shown in Figure 4(a). They indicated that the tendency was not only true for specific materials but also true globally. If the conventional autonomous search was performed without prioritizing crystal structure information, the number of times to search for full-Heusler alloys would have increased, and it would have taken more time to find the Pareto frontier materials.

The materials shown in Table 1 have the potential to exhibit half-metallic properties. However, the actual synthesis of these materials is expected to be challenging because it is extremely difficult to synthesize half-Heusler structures containing multiple elements. High-throughput experiments and autonomous robotic synthesis techniques may allow the synthesis of these materials possible in the future. In addition, most Heusler alloys shown in Table 1 contain As, which is a toxic element. This also increases the difficulty in their synthesis and applications. Therefore, an autonomous material search system that considers material toxicity is required. Therefore, a future autonomous search that considers material properties (e.g., $P$ and $G$) and toxicity can be developed based on the proposed autonomous material search system.

![Figure 4](image-url) Figure 4. Results of the autonomous material search in the Heusler alloy space. (a) Spin polarization, $P$, and half-metallic gap, $G$, of the materials explored by the autonomous material search. The black-solid line denotes the Pareto frontier. (b) the density of state of a half-Heusler alloy on the Pareto frontier, i.e., Co$_{0.6}$Fe$_{0.4}$Cr$_{0.2}$Mn$_{0.2}$As$_{1.0}$.
4. Conclusions

An autonomous material search system that combines ab initio calculations with an autoencoder and Bayesian optimization was developed. The system can autonomously explore the material space through multi-objective optimization of multiple material properties while considering the similarities among elements and emphasizing specific features (e.g. crystal structure). A new material with large $P$ and $G$ was predicted by using this system to explore the material space of Heusler alloys. The application of this system is not limited to Heusler alloys and can be used to explore various material spaces to achieve various physical properties.

Acknowledgments

We thank Dr. Y. Miura, Dr. K. Masuda, and Dr. K. Sodeyama in National Institute for Materials Science for valuable discussion. This study was supported by JST-CREST Grant No. JPMJCR21O1.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This study was supported by JST-CREST Grant No. JPMJCR21O1. This paper is partially based on results obtained from a project, [JPNP20003], commissioned by the New Energy and Industrial Technology Development Organization (NEDO).

ORCID

Yuma Iwasaki http://orcid.org/0000-0002-7117-277X
Hwang Jaekyun http://orcid.org/0000-0002-1792-0044
Yuka Sakuraba http://orcid.org/0000-0003-4618-9550
Masato Kotsugi http://orcid.org/0000-0002-4841-1808
Yasuhiro Igarashi http://orcid.org/0000-0003-1042-6657

Data availability statement

The data that support the results reported in this paper and other findings of this study are available from the corresponding author upon reasonable request.

References

[1] Coley CW, Thomas DA, Lummss JAM, et al. A robotic platform for flow synthesis of organic compounds informed by AI planning. Science. 2019;365(6453):DOI:10.1126/science.aax1566
[2] Burger B, Maffetone PM, Gusev VV, et al. A mobile robotic chemist. Nature. 2020;583(7815):237–241. DOI:10.1038/s41586-020-2442-2
[3] Nikolaev P, Hooper D, Webber F, et al. Autonomy in materials research: a case study in carbon nanotube growth. Npj Comput Mater. 2016;2:16031
[4] Shimizu R, Kobayashi S, Watanabe Y, et al. Autonomous materials synthesis by machine learning and robotics. APL Mater. 2020;8:111110.
[5] Li Z, Najeeb MA, Alves L, et al. Robot-Accelerated perovskite investigation and discovery. Chem Mater. 2020;32(13):5650–5663. (5650). DOI:10.1021/acs.chemmater.0c01153
[6] Roch AM, Hase F, Kreisbeck C, et al. ChemOS: orchestrating autonomous experimentation. Sci Rob. 2018;3(19):DOI:10.1126/scirobots.aat5559
[7] Attia PM, Grover A, Jin N, et al. Closed-Loop optimization of fast-charging protocols for batteries with machine learning. Nature. 2020;578:397–402.
[8] Granda JM, Donina L, Dragne V, et al. Controlling an organic synthesis robot with machine learning to search for new reactivity. Nature. 2018;559:377–381.
[9] Szymanski NJ, Zeng Y, Hoo H, et al. Toward autonomous design and synthesis of novel inorganic materials. Mater Horiz. 2021;8:2169–2198.
[10] Iwasaki Y, Sawada R, Saitoh E, et al. Machine learning automated identification of magnetic alloys beyond the Slater-Pauling limit. Commun Mater. 2021;2(31):DOI:10.1038/s43246-021-00135-0
[11] Sawada R, Iwasaki T, Ishida M. Boosting material modeling using game tree search. Phys Rev Mater. 2018;2:103802.
[12] Seko A, Togo A, Hayashi H, et al. Prediction of low-thermal-conductivity compounds with first-principles anharmonic lattice-dynamics calculations and bayesian optimization. Phys Rev Lett. 2015;115:205901.
[13] Jalem R, Kanamori K, Takeuchi I, et al. Bayesian-driven first-principles calculations for accelerating exploration of fast ion conductors for rechargeable battery application. Sci Rep. 2018;8(5845). DOI:10.1038/s41598-018-23852-y
[14] Kusaba A, Kangawa Y, Kuboyama T, et al. Exploration of a large-scale reconstructed structure on GaN(0001) surface by Bayesian optimization. Appl Phys Lett. 2022;120:021602.
[15] Furuya D, Miyashita T, Miura Y, et al. Autonomous synthesis system integrating theoretical, informatics, and experimental approaches for large-magnetic-anisotropy materials. Sci Technol Adv Mater: Methods. 2022;2(1):280–293.
[16] Chadow S, Qi X, Kubler J, et al. Tunable multifunctional topological insulators in ternary Heusler compounds. Nat Mater. 2010;9:541–545.
[17] Hirschberger M, Kushwaha S, Wang Z, et al. The chiral anomaly and thermopower of Weyl fermions in the half-Heusler GdPbBi. Nat Mater. 2016;15:1161–1165.
[18] Wollmann A, Nayak AK, Parkin SSP, et al. Heusler 4.0: tunable materials. Ann Rev Mater Res. 2017;47:247–470.
[19] Quinn RJ, Bos JW. Advances in half-Heusler alloys for thermoelectric power generation. Mater Adv. 2021;2:6246–6266.
[20] Elphick K, Frost W, Samiepour M, et al. Heusler alloys for spintronic devices: review on recent development and future perspectives. Sci Technol Adv Mater. 2021;22(1):235–271. DOI:10.1080/14686996.2020.1812364
[21] Sakuraba Y, Hattori M, Ougane M, et al. Giant tunneling magnetoresistance in Co$_2$MnSi/Al–O/Co$_2$
MnSi magnetic tunnel junctions. Appl Phys Lett. 2006;88:192508.

[22] Liu H, Honda Y, Taira T, et al. Giant tunneling magnetoresistance in epitaxial Co₂MnSi/MgO/Co₂Si magnetic tunnel junctions by half-metallicity of Co₂MnSi and coherent tunneling. Appl Phys Lett. 2012;101:132418.

[23] Moges K, Honda Y, Liu H, et al. Enhanced half-metallicity of off-stoichiometric quaternary Heusler alloy Co₂(Mn,Fe)Si investigated through saturation magnetization and tunneling magnetoresistance. Phys Rev B. 2016;93:134403.

[24] Shan R, Sukegawa H, Wang WH, et al. Demonstration of half-metallicity in fermi-level-tuned Heusler alloy Co₂FeAl₀.₅Sb₀.₅ at room temperature. Phys Rev Lett. 2009;102:246601.

[25] Iwase T, Sakuraba Y, Bosu S, et al. Large interface spin-asymmetry and magnetoresistance in fully epitaxial Co₂MnSi/Ag/Co₂MnSi current-perpendicular-to-plane magnetoresistive devices. Appl Phys Express. 2009;2:063003.

[26] Takahashi YK, Srinivasan A, Varaprasad B, et al. Large magnetoresistance in current-perpendicular-to-plane pseudospin valve using a Co₂Fe(Ge₀.₅Ga₀.₅)₂Heusler alloy. Appl Phys Lett. 2011;98:152501.

[27] Sakuraba Y, Ueda M, Miura Y, et al. Extensive study of giant magnetoresistance properties in half-metallic Co₂(Fe,Mn)Si-based devices. Appl Phys Lett. 2012;101:252408.

[28] Jung JW, Sakuraba Y, Sasaki TT, et al. Enhancement of magnetoresistance by inserting thin NiAl layers at the interfaces in Co₂FeGa₀.₅Ge₀.₅/Ag/Co₂FeGa₀.₅Ge₀.₅ current-perpendicular-to-plane pseudo spin valves. Appl Phys Lett. 2016;108:102408.

[29] Kimura T, Hashimoto N, Yamada S, et al. Room-Temperature generation of giant pure spin currents using epitaxial Co₂FeSi spin injectors. Npg Asia Mater. 2012;4:e9.

[30] Seki T, Sakuraba Y, Arai H, et al. High power all-metal spin torque oscillator using full Heusler Co₂(Fe,Mn)Si. Appl Phys Lett. 2014;105:092406.

[31] Bosu S, Sepehri-Amin H, Sakuraba Y, et al. High frequency out-of-plane oscillation with large cone angle in mag-flip spin torque oscillators for microwave assisted magnetic recording. Appl Phys Lett. 2017;110:142403.

[32] Akai H. Electronic structure Ni-Pd alloys calculated by the self-consistent KKR-CPA method. J Phys Soc Jpn. 1982;51:468–474.

[33] Khan SN, Staunton JB, Stocks GM. Statistical physics of multicomponent alloys using KKR-CPA. Phys Rev B. 2016;93:054206.

[34] Yang L, Liu B, Luo H, et al. Investigation of the site preference in Mn₂RuSn using KKR-CPA-LDA calculation. J Magn Mater. 2015;382(15):247–251. DOI:10.1016/j.jmmm.2015.01.081

[35] Jin K, Sales BC, Stocks GM, et al. Tailoring the physical properties of Ni-based single-phase equiatomic alloys by modifying the chemical complexity. Sci Rep. 2016;6:20159.

[36] Momma K, Izumi F. VESTA 3 for three-dimensional visualization of crystal, volumetric and morphology data. J Appl Crystallogr. 2011;44:1272–1276.

[37] Snoek J, Larochelle H, Adams RP. Practical Bayesian optimization of machine learning algorithms. Adv Neural Inf Process Syst. 2012;25:2960–2968.

[38] Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks. Science. 2006;313:504–507.

[39] Motoyama Y, Tamura R, Yoshimi K. Bayesian optimization package: pHYSBO. Comput Phys Commun. 2022;278:108405.

[40] Ward L, Agrawal A, Choudhary A, et al. A general-purpose machine learning framework for predicting properties of inorganic materials. Npj Comput Mater. 2016;2:16028.

[41] Zhou C, Yabo D, Zheng X, et al. Convolutional neural networks for crystal material property prediction using hybrid orbital-field matrix and Magpie descriptors. Crystals. 2019;9(4):191. DOI:10.3390/cryst9040191

[42] Wang Z, Zoghi M, Hutter F, et al. Bayesian optimization in high dimensions via random embeddings. Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence; 2013;1778–1784.

[43] Auer P. Using confidence bounds for exploitation-exploration trade-off. J Mach Learn Res. 2002;3:397–422.

[44] Jianhua M, Vinay I, Hegde K, et al. Computational investigation of half-Heusler compounds for spintronics applications. Phys Rev B. 2017;95:024411.

[45] Galanakis I, Movpropoulos PH, Dederichs PH. Electronic structure and Slater–Pauling behaviour in half-metallic Heusler alloys calculated from first principles. J Phys D Appl Phys. 2006;39:765–775.