Ranking Pareto optimal solutions based on projection free energy

Ryo Tamura [1,2,3,4,*] Kei Terayama [5] Masato Sumita [4,1] and Koji Tsuda [3,2,4,†]

[1]International Center for Materials Nanoarchitectonics, National Institute for Materials Science, Tsukuba 305-0044, Japan
[2]Research and Services Division of Materials Data and Integrated System, National Institute for Materials Science, Tsukuba 305-0044, Japan
[3]Graduate School of Frontier Sciences, The University of Tokyo, Chiba 277-8561, Japan
[4]RIKEN Center for Advanced Intelligence Project, Tokyo 103-0027, Japan
[5]Graduate School of Medical Life Science, Yokohama City University, Kanagawa 230-0045, Japan

(Received 22 February 2023; revised 22 June 2023; accepted 2 August 2023; published 19 September 2023)

Based on available datasets prepared by numerical simulations and machine learning, maps of properties for materials that have not yet been synthesized can be developed. These maps can be used to select promising materials for synthetic experiments. With a single objective function, the ranking of the optimal solutions can be simply obtained based on the values of the target property. However, applications with multiple target properties require the calculation of Pareto optimal solutions to visualize trade-offs. These solutions are generally ranked manually, selecting the weight of the multiple objectives based on prior knowledge. In this study, to provide an automated ranking of Pareto solutions, we introduced the most-isolated Pareto solution (MIPS) score, which is defined by a projection free energy. Using the MIPS ranking, it is possible to appropriately select the most isolated materials predicted in the property space. To verify the effectiveness of the proposed method, we used a database of semiconductors created by density-functional theory. Our method was able to correctly select and rank the most isolated solutions in both convex and concave two-dimensional Pareto frontiers, outperforming the most relevant outlier detection methods. We also demonstrated that our approach can be easily extended to three-dimensional property spaces.

DOI: 10.1103/PhysRevMaterials.7.093804

I. INTRODUCTION

The increasing computational power of high-performance computers and the development of automated techniques for numerical simulations have facilitated the calculation of the properties of a growing number of materials [1]. As a result, many accessible computational material databases have been developed for inorganic materials [2,3], organic materials [4], and molecules [5,6]. In addition, the use of machine learning in materials science has contributed to remarkable results [7–11]. By applying machine learning models to experimental data, the properties of unknown materials can be predicted with high accuracy. Predicted properties can then be used as a reference to select promising new materials, which will be the target of forthcoming experiments.

If we consider a single target property, it is sufficient to select the material with the highest (or lowest, depending on the context) predicted property value. However, in the design process, materials are generally selected considering multiple desired properties at the same time. Moreover, the selection of materials often leads to trade-offs (i.e., the enhancement of a property implies the worsening of another one). To treat such trade-off problems, multiobjective optimization methods can be used [12–14]. Because a single solution does not exist for this type of problem, these methods find multiple Pareto optimal solutions (see Fig. 1), each of which is an *optimal solution* obtained by varying the balance of the objective functions. In general, the Pareto solutions are ranked using an arbitrary weighted sum [15–19]. The weights of the objectives are adjusted manually, thus requiring previous field knowledge. A systematic method for determining the balance of objective functions was discussed in Refs. [20,21], where the central part of the Pareto frontier was regarded as that including the best solutions. This means that the Pareto solutions are ranked based on the proximity to the center of the frontier, which is too simple. In general, without the direct use of human knowledge, it is difficult to automatically select promising materials when multiple objective properties are considered.

In this study, we develop a method to automatically provide the ranking of Pareto solutions based on projection free energy. Considering a material from the Pareto solutions, if there are few others with similar properties in the property space (i.e., isolatability is high), then the material can be regarded as having unique properties (i.e., it is a curious material). To evaluate the isolation of each Pareto solution, we introduce the most-isolated Pareto solution (MIPS) score, inspired by the definition of free energy (see Fig. 1). Using the MIPS score, the ranking of Pareto solutions is provided. Note that the ranking of the Pareto solutions considered in this study is

*tamura.ryo@nims.go.jp
†tsuda@k.u-tokyo.ac.jp

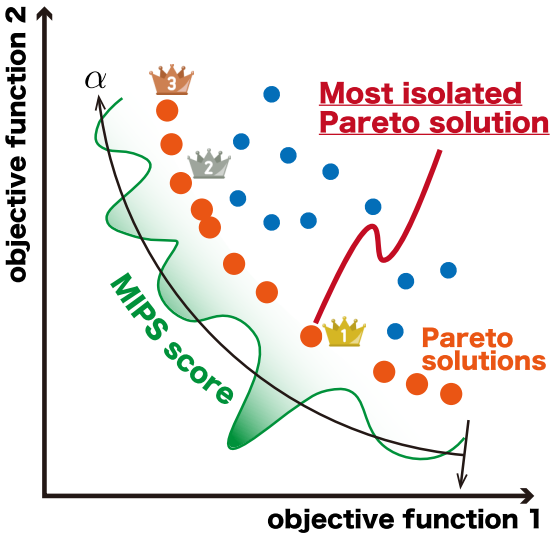

FIG. 1. Schematics of Pareto solutions minimizing two objective functions and MIPS score. Based on the MIPS score, the ranking of Pareto solutions (orange) that are isolated from other datapoints are determined automatically.

different from the *Pareto ranking* used by genetic algorithms [22]. In the *Pareto ranking*, all the solutions are ranked and all the Pareto solutions are located at the highest rank.

Let us explain the physical background of the idea that the isolation of each Pareto solution can be evaluated in terms of free energy. First, we consider two simple physical systems A and B whose ground states are not degenerated, and they have the same number of discrete states. We also assume that system A has few low-energy excited states and system B has many low-energy excited states. Examples of a density of states (DoS) representing these systems are shown in the left panels of Fig. 2(a). Focusing on the ground state, system A has fewer low-energy excited states, so it can be assumed that this ground state is more isolated from the other solutions if the value of the energy is used as a measure of distance. System B, on the other hand, has many excited states in the neighborhood of the ground state, and the ground state can be considered as nonisolated. Using the DoS, the temperature dependence of the Helmholtz free energy can be calculated for each system and is shown in the right panel of Fig. 2(a). System A has a larger free energy, indicating that it is more unstable than System B, and has an isolated ground state. In this way, ground state isolation can be measured in terms of free energy.

Next, we consider the relationship between the isolation of the Pareto solution and the *projection* free energy. To apply the above idea, it is necessary to consider a projection into a one-dimensional space from the space in which multiple objective functions are defined. This projection is defined so that each Pareto solution becomes an optimal solution with the smallest value in the projected one-dimensional space, i.e., the ground state. The simplest projection is to focus on only one of the objective functions. For example, let us consider the dataset with two objective functions as shown in Fig. 2(b). In this case, if we make a projection considering only one objective function, the projected DoS is shown in Fig. 2(b). Note that

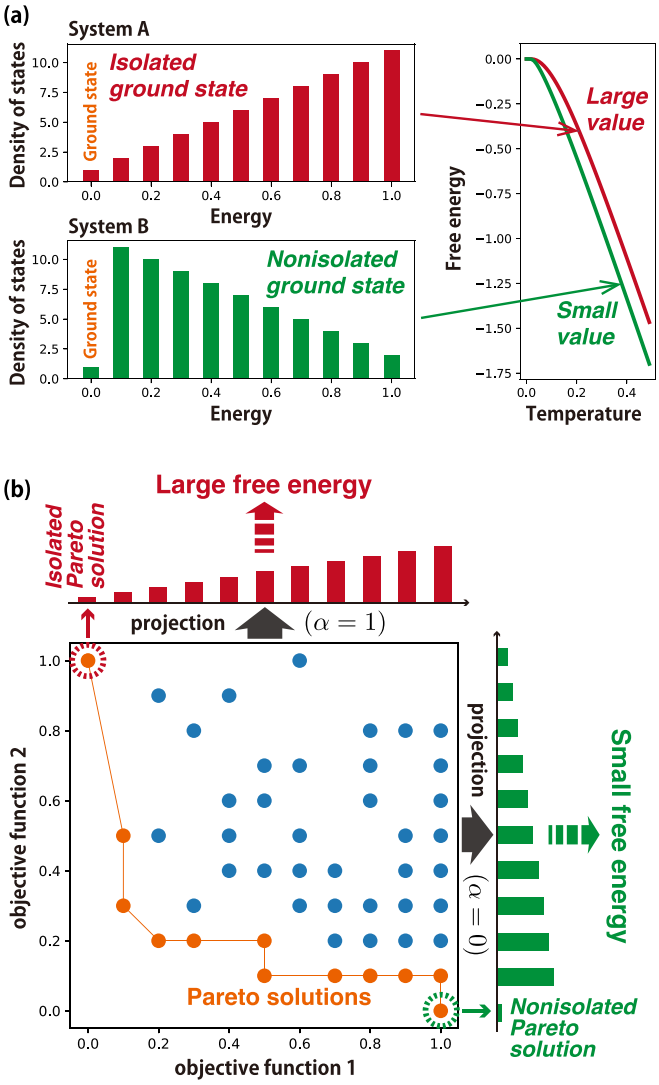

FIG. 2. (a) Examples of density of states for system A having few low-energy excited states and system B having many low-energy excited states. The temperature dependence on free energy results are shown for systems A and B. (b) Examples of two-dimensional Pareto frontier and projected one-dimensional DoS. For top and right DoS, objective functions 1 and 2 are only considered, respectively. These cases correspond to the Tchebycheff decomposition defined by Eq. (2) with $\alpha = 1$ and 0, respectively. The Pareto solutions highlighted by red and green dotted circles correspond to the ground state (optimal solution) for top and right DoS cases, respectively. If a projection with different α is performed, different Pareto solution becomes optimal solution.

these are the same DoS considered above, that is, for objective functions 1 and 2, the DoS is the same as the systems A and B, respectively. For each projection, the Pareto solution that lies in the optimal solution is different. Thus, by estimating the free energy in each case and comparing its magnitude [free energy results are shown in Fig. 2(a)], we can compare the isolation for different Pareto solutions. Here, the free energy defined in this projected space is called the *projection* free energy. In this example, via the projection free energy, we can conclude that the Pareto solution considering only objective

function 1, highlighted by the red circle, is more isolating than the Pareto solution considering only objective function 2, highlighted by green circle. For other Pareto solutions, we search for a one-dimensional space in which each Pareto solution becomes an optimal solution and calculate the projection free energy. By comparing the projection free energy, the ranking of Pareto solutions based on the isolation can be created. Here, the projection to one-dimensional space is performed by Tchebycheff decomposition. Note that the relationship between the difficulty in the optimization problems and the shape of DoS has been discussed [23,24].

A detailed definition based on projection free energy of the MIPS is provided in Sec. II. In Sec. III, we report the results of the material selection based on the MIPS using a dataset of semiconductors obtained by density functional theory (DFT). In Sec. IV, we present the discussion and summary of our findings.

II. METHODS

We define the MIPS based on the concept of projection free energy. We consider a multiobjective optimization approach based on the Tchebycheff decomposition [25–27] with two objective functions. Here, we assume that these are minimized simultaneously. It follows that the optimization problem in the d-dimensional space defined by $\mathbf{x} \in \mathbb{R}^d$ can be described as follows:

$$\min_{\mathbf{x}} g_\alpha(\mathbf{x}), \tag{1}$$

where $g_\alpha(\mathbf{x})$ is constructed using the two objective functions $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$ so that

$$g_\alpha(\mathbf{x}) = \max[\alpha f_1(\mathbf{x}), (1-\alpha) f_2(\mathbf{x})]. \tag{2}$$

The value of α ($0 \leqslant \alpha \leqslant 1$) determines the balance of the two objectives. Note that the role of α is similar to that of the coefficient of a weighted sum [i.e., $\alpha f_1(\mathbf{x}) + (1-\alpha) f_2(\mathbf{x})$]. When α is small, $f_2(\mathbf{x})$ has more relevance than $f_1(\mathbf{x})$; when α is large, the situation is reversed (see Fig. 1). When α is fixed, all data points are projected to one-dimensional data via $g_\alpha(\mathbf{x})$. Thus, a one-dimensional histogram can be drawn depending on α, and one of the Pareto solutions is located at the optimal solution depending on α.

For the analysis target of materials, the dataset of material descriptors $D = \{\mathbf{x}_i\}_{i=1,\ldots,N}$ is given. Here, index i indicates the materials and D includes N types of materials. We consider the case in which $\forall i$ the values of the objective functions $\{f_1(\mathbf{x}_i), f_2(\mathbf{x}_i)\}$ are predicted by numerical simulations or machine learning. Thus, in our problem setting, the materials in the Pareto frontier can be easily determined. Note that a material with $\mathbf{x}^*$ belonging to the Pareto optimal is one that satisfies the following condition: there is no $\mathbf{x}_i$ for which $f_k(\mathbf{x}_i) < f_k(\mathbf{x}^*)$ for $\forall k$. We impose a min-max normalization for $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$ defined as follows:

$$f_1'(\mathbf{x}) = \frac{f_1(\mathbf{x}) - \min\{f_1(\mathbf{x}_i)\}_{i=1,\ldots,N}}{\max\{f_1(\mathbf{x}_i)\}_{i=1,\ldots,N} - \min\{f_1(\mathbf{x}_i)\}_{i=1,\ldots,N}}, \tag{3}$$

$$f_2'(\mathbf{x}) = \frac{f_2(\mathbf{x}) - \min\{f_2(\mathbf{x}_i)\}_{i=1,\ldots,N}}{\max\{f_2(\mathbf{x}_i)\}_{i=1,\ldots,N} - \min\{f_2(\mathbf{x}_i)\}_{i=1,\ldots,N}}. \tag{4}$$

The normalization ensures that the two objective functions have the same range of values, thus contributing equally to the target function that can be redefined as follows:

$$g_\alpha'(\mathbf{x}) = \max[\alpha f_1'(\mathbf{x}), (1-\alpha) f_2'(\mathbf{x})]. \tag{5}$$

We perform an additional normalization so that the values of the target function are in the range [0, 1] for each α:

$$\mathcal{H}_\alpha(\mathbf{x}) = \frac{g_\alpha'(\mathbf{x}) - \min\{g_\alpha'(\mathbf{x}_i)\}_{i=1,\ldots,N}}{\max\{g_\alpha'(\mathbf{x}_i)\}_{i=1,\ldots,N} - \min\{g_\alpha'(\mathbf{x}_i)\}_{i=1,\ldots,N}}. \tag{6}$$

Here, $\mathcal{H}_\alpha(\mathbf{x})$ is regarded as the Hamiltonian function depending on α from which we evaluate the projection free energy. The partition function depending on the temperature T is defined as follows:

$$Z_\alpha(T) = \sum_{i=1}^{N} \exp[-\mathcal{H}_\alpha(\mathbf{x}_i)/T]. \tag{7}$$

Thus, the projection free energy depending on α is given by

$$F_\alpha(T) = -T \log Z_\alpha(T) \tag{8}$$

$$= E_\alpha(T) - T S_\alpha(T), \tag{9}$$

where $E_\alpha(T)$ and $S_\alpha(T)$ represent the energy and entropy terms in the original definition of free energy. When $T = 0$, only the ground state with $\mathcal{H}_\alpha(\mathbf{x}) = 0$ contributes to the free energy calculation and $F_\alpha(0) = 0$. As T is increased, the contribution of excited states is augmented in $F_\alpha(T)$. In particular, for small T, the contribution of the excited states close to the ground state is important. When there is a limited number of other data points near the ground state, the free energy $F_\alpha(T)$ with small T becomes large, corresponding to a low entropy value (see Fig. 2). Since the ground state is one of the Pareto solutions, we can use the free energy with small T when considering the isolation from other solutions in the neighborhood of the Pareto solution. Here, a sufficiently small temperature T^* is predetermined, and the MIPS score is defined as follows:

$$\mathrm{MIPS}(\alpha) = F_\alpha(T^*). \tag{10}$$

A Pareto solution with a large MIPS(α) can be regarded as isolated from the other data points. Note that, in this paper, $T^* = 0.01$ is utilized, and this value was determined by calculating the T^*-dependence of the ranking (see Fig. S1 [28]), which will be discussed later.

The following is a procedure to determine a ranking of Pareto solutions using the MIPS score.

(1) For a given α, project all data into one-dimensional space.

(2) Calculate the MIPS score according to Eq. (10). The MIPS score represents the performance measure of the Pareto solution located in the optimal solution within the projected one-dimensional space.

(3) Repeat steps 1 and 2 while varying the value of α. In this case, the same Pareto solution may be the optimal solution for different values of α. Therefore, each Pareto solution can have multiple MIPS scores corresponding to different α values. In such cases, the highest MIPS score among the multiple scores is chosen as the score for corresponding Pareto solution.

(4) Arrange the Pareto solutions in ascending order of their MIPS scores.

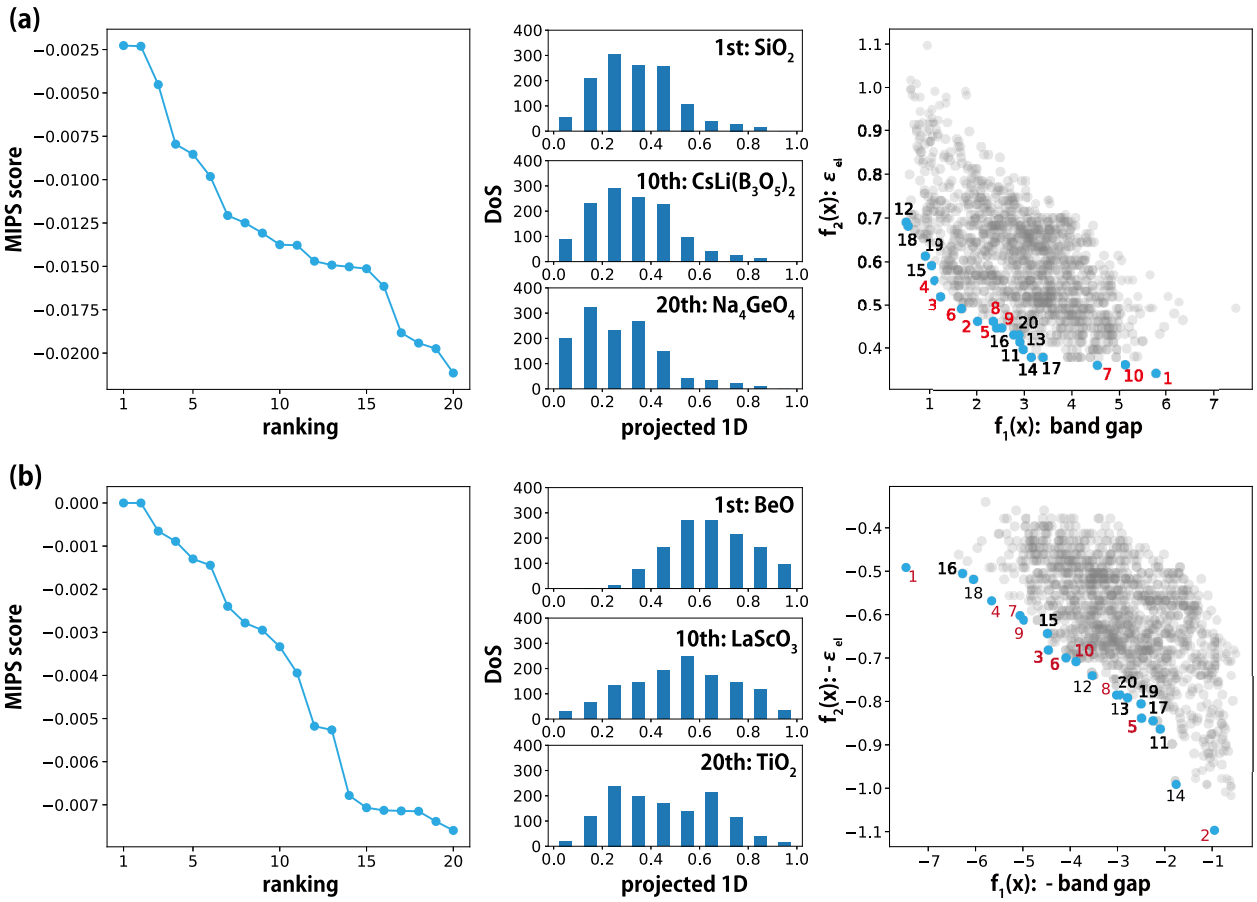

FIG. 3. MIPS values (left column), representative DoS (middle column), and distribution of datapoints (right column) with a convex- or concave-shaped Pareto frontier in the property space spanned by the band gap and electric dielectric constant. The first 20 solutions ranked by MIPS are represented by blue datapoints. All DoS are summarized in Figs. S2 and S3 [28].

However, in our present implementation, there might be some Pareto solutions for which the MIPS score cannot be defined. Here, α is varied from 0 to 1 using discrete increments, not continuous ones. Then, for considering all discrete α, there might be some Pareto solutions that do not become an optimal (minimum) solution in the projected one-dimensional space. In such cases, these particular Pareto solutions should be ranked as the lowest in the ranking.

III. RESULTS

A. MIPS ranking in a two-dimensional property space

We test our ranking model based on the MIPS using a dataset of 1278 semiconductors obtained by density functional theory (DFT) calculations [29]. We address the well-known trade-off between the band gap and dielectric constant of semiconductors. First, we consider the case in which both the band gap and electric dielectric constant ($\epsilon_{\rm el}$) are simultaneously minimized, which gives 33 Pareto solutions. The MIPS are calculated by varying α from 0 to 1 using 0.001 increments. For seven Pareto solutions, the MIPS score cannot be defined. In Fig. S1 [28], the transition of MIPS ranking depending on T^* is shown. It can be seen that the Pareto solutions in the top ranking do not change much when T^* is changed from 0.001 and 0.01. Thus, we use $T^* = 0.01$, which achieves that the ranking does not change significantly as decreasing T^* and includes sufficient information around Pareto solution. Note that from T^* dependency, the robustness of ranking can be discussed (see Fig. S1 [28]). In Table I, the MIPS ranking is summarized when $T^* = 0.01$, and for the materials with 27th, an appropriate α cannot be found. The top 20 results with the largest MIPS are shown in Fig. 3(a). The left column shows the MIPS values depending on the ranking. In the middle column, some representative DoS (first, tenth, and 20th in the ranking) are shown. Clearly, in the first DoS, the number of materials with low values in the projected one-dimensional space is small, indicating SiO_2 is most isolated. In addition, we can see that the number of materials with low values on the projected one-dimensional space increases as we focus on the lower ranking. Thus, there are only a small number of materials around a Pareto solution with a high MIPS. These results show that the ranking based on the MIPS is in accordance with our purpose of providing a ranking of Pareto solutions in terms of isolation. The distributions of the selected Pareto solutions and other datapoints in the property space are shown in the right column of Fig. 3(a), where it can be seen that a variety of isolated solutions were selected from a convex-shaped Pareto frontier using the MIPS ranking.

We also consider a similar problem with a concave-shaped Pareto frontier. In this case, the band gap and electric dielectric

TABLE I. MIPS ranking for a convex-shaped Pareto frontier when band gap and electronic dielectric constant (ϵ_{el}) are minimized.

Ranking	Materials	MIPS score	Band gap	ϵ_{el}
1	SiO_2	−0.002270662	5.79	2.2
2	$RbMnO_4$	−0.002309888	2.02	2.9
3	K_3BiO_4	−0.004522329	1.24	3.3
4	Na_6PbO_5	−0.007955985	1.1	3.6
5	$Rb_2Be_2O_3$	−0.00853923	2.42	2.8
6	K_4PbO_4	−0.009820247	1.68	3.1
7	$Na_2Si_2O_5$	−0.012068143	4.54	2.3
8	K_4SnO_4	−0.012493567	2.35	2.9
9	$K_2Na_4Be_2O_5$	−0.013082062	2.53	2.8
10	$CsLi(B_3O_5)_2$	−0.013750656	5.15	2.3
11	KNa_2BO_3	−0.013784018	2.98	2.5
12	K_3AuO	−0.014702593	0.5	4.9
13	$Na_7Al_3O_8$	−0.014925278	2.91	2.6
14	Na_3BO_3	−0.015025572	3.15	2.4
15	Na_5BiO_5	−0.015135497	1.03	3.9
16	K_4HfO_4	−0.01615541	2.78	2.7
17	Na_4SiO_4	−0.018832454	3.4	2.4
18	$Cd_2Sb_2O_7$	−0.019424779	0.54	4.8
19	$Rb_5Au_3O_2$	−0.019745257	0.9	4.1
20	Na_4GeO_4	−0.021150197	2.89	2.7
21	Na_2SiO_3	−0.023911928	3.99	2.4
22	KBO_2	−0.024760391	4.12	2.4
23	$Rb_7Au_5O_2$	−0.024863074	0.68	4.7
24	$NaBO_2$	−0.025005608	4.19	2.4
25	$NaCuO_2$	−0.025564282	0.79	4.6
26	SnO_2	−0.026644142	0.86	4.5
27	$AlAsO_4$	-	4.39	2.4
27	K_4GeO_4	-	2.78	2.8
27	$Na_4B_2O_5$	-	3.92	2.4
27	Rb_3NaPbO_4	-	1.63	3.3
27	Na_3AsO_4	-	3.38	2.4
27	Rb_4PbO_4	-	1.6	3.3
27	$CsCdBO_3$	-	1.69	3.1

TABLE II. MIPS ranking for a concave-shaped Pareto frontier when band gap and electronic dielectric constant (ϵ_{el}) are maximized.

Ranking	Materials	MIPS score	Band gap	ϵ_{el}
1	BeO	−1.48E-09	7.47	3.1
2	PtO_2	−3.91E-07	0.95	12.5
3	ThO_2	−0.00065343	4.45	4.8
4	$HfSiO_4$	−0.000891161	5.66	3.7
5	$In_2Pt_2O_7$	−0.001298629	2.48	6.9
6	$ScTaO_4$	−0.00144392	4.08	5
7	$ZrSiO_4$	−0.002396839	5.06	4
8	$TaBiO_4$	−0.002783542	3.02	6.1
9	$CaHfO_3$	−0.002949965	4.98	4.1
10	$LaScO_3$	−0.003333902	3.87	5.1
11	V_2HgO_6	−0.003943528	2.09	7.3
12	$ScNbO_4$	−0.00517729	3.53	5.5
13	$Ta_2Bi_4O_{11}$	−0.005263738	2.77	6.2
14	$TlPd_3O_4$	−0.006784382	1.76	9.8
15	$CaZrO_3$	−0.007067831	4.47	4.4
16	$BeAl_2O_4$	−0.007128697	6.27	3.2
17	$Sc_2Pt_2O_7$	−0.007141839	2.24	7
18	Al_2O_3	−0.007150093	6.04	3.3
19	$NaNb_4Bi_5O_{18}$	−0.007385994	2.49	6.4
20	TiO_2	−0.007594438	2.94	6.1
21	Nb_2CdO_6	−0.007769357	3.54	5.4
22	$TiCdO_3$	−0.007860211	3.03	5.8
23	$SrHfO_3$	−0.008652802	4.68	4.1
24	$TlPt_3O_4$	−0.009422992	1.78	9.6
25	$LaTaTiO_6$	−0.009535061	3.08	5.6
26	$TiPbO_3$	−0.009952586	1.84	7.9
27	Bi_2O_3	−0.011299502	2.15	7
28	YZnAsO	-	1.11	9.8
28	$CaTa_2Bi_2O_9$	-	3.03	5.6
28	Ta_4PbO_{11}	-	3.5	5.5
28	Ta_2CdO_6	-	4.08	4.8

constant are simultaneously maximized and 31 Pareto solutions are defined. Since the definition of the MIPS is based on the minimization of the objective functions, we consider the corresponding negative values of the two properties of the dielectric materials, so that the original maximization is transformed into a minimization problem that can be addressed by the proposed method. In this case, for four Pareto solutions, MIPS score cannot be defined. In Table II, the MIPS ranking is summarized when $T^* = 0.01$, and the results are presented in Fig. 3(b). Note that in Fig. S1 [28], MIPS ranking depending on T^* is shown. The first two materials (BeO and PtO_2) with the highest rank appear highly isolated. This result is correctly reflected by the associated MIPSs. We conclude that the MIPS ranking can be also used for problems with a concave-shaped Pareto frontier.

B. Comparison with outlier detection techniques

Our ranking method can be used to extract isolated datapoints from Pareto solutions. Outlier detection techniques are used to perform the same task using machine learning. To investigate the possible advantages of our ranking method, we apply the outlier detection techniques, namely, isolation forest [30], local outlier factor [31], and one-class support vector machine (SVM) to solve the maximization problem of band gap and electronic dielectric constant. The results are presented in Fig. 4. We consider the following two cases: (1) all data, including Pareto solutions, are used as the training dataset; (2) only Pareto solutions are used as the training dataset. Considering case (1), the isolation forest and one-class SVM recognize a datapoint/solution as an outlier when at least one of the objective functions is particularly small. This means that these techniques cannot be used to define the isolated Pareto solutions. Because the local outlier factor correctly selects all the Pareto solutions, it is not suitable for the ranking process. Considering case (2), only the solutions located at the extremities of the Pareto frontier are selected as outliers by the considered methods. These results indicate that it is impossible to create the ranking of Pareto solutions by using outlier detection techniques.

C. MIPS ranking in a three-dimensional property space

By extending the target function domain, we can apply our method to a property space of more than two dimensions. In

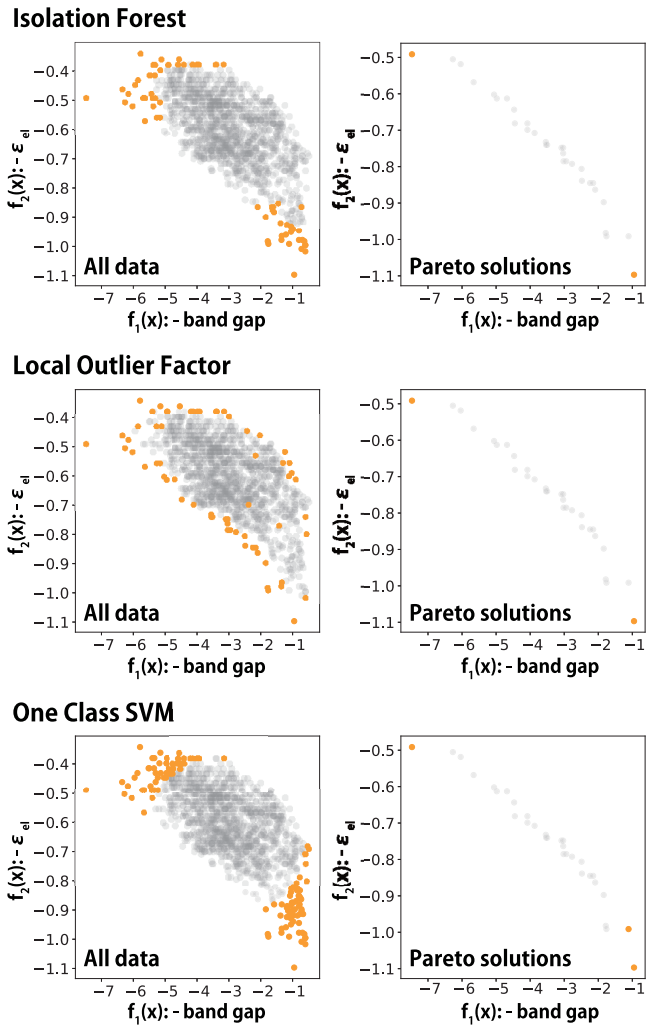

FIG. 4. Distributions of outliers (orange) and other datapoints (gray) obtained using three types of outlier detection techniques (isolation forest, local outlier factor, and one-class SVM) when all the datapoints (left column) or the Pareto solutions only (right column) are used in the training phase.

particular, in this section, we address the three-dimensional case. In addition to the band gap and the electronic dielectric constant, we consider the ionic dielectric constant as the third property. We also consider the case in which these properties are maximized and 70 Pareto solutions are determined. Here, the target function inspired by the Tchebycheff decomposition is as follows:

$$g_\alpha(\mathbf{x}) = \max[\alpha_1 f_1(\mathbf{x}), \alpha_2 f_2(\mathbf{x}), \alpha_3 f_3(\mathbf{x})], \qquad (11)$$

where $f_1(\mathbf{x})$, $f_2(\mathbf{x})$, and $f_3(\mathbf{x})$ are the three objective functions. Choosing the values of α_1, α_2, and α_3, we apply two constraints: $\alpha_i > 0$ $(i = 1, 2, 3)$ and $\alpha_1 + \alpha_2 + \alpha_3 = 1$. Under these constraints, the MIPS are calculated by varying α_i from 0 to 1 using a 0.01 increments. In this case, for 15 Pareto solutions, the MIPS score cannot be defined. The top 20 results with the largest MIPS are shown in Table III and Fig. 5. From the distributions, it can be observed that a wide variety of Pareto solutions are chosen.

TABLE III. MIPS ranking of top 20 for a three-dimensional Pareto frontier when the band gap, and the electric (ϵ_{el}) and ionic (ϵ_{ion}) dielectric constants are maximized. The remaining ranking is shown in Table S1 [28].

Ranking	Materials	MIPS score	Band gap	ϵ_{el}	ϵ_{ion}
1	BeO	−1.48E-09	7.47	3.1	3.9
2	PtO_2	−3.91E-07	0.95	12.5	1.2
3	$TiPbO_3$	−1.05E-06	1.84	7.9	661.4
4	La_3TaO_7	−7.48E-05	3.63	4.8	73.4
5	$Sr_4Ta_2O_9$	−0.000120481	4.4	3.7	32.7
6	$CaHfO_3$	−0.000177427	4.98	4.1	17.2
7	Sr_4As_2O	−0.000247042	1.06	8.3	62.8
8	$CaZrO_3$	−0.000429054	4.47	4.4	22.5
9	$Rb_2Ti(WO_4)_3$	−0.000444705	2.83	5.3	98.3
10	$TaBiO_4$	−0.000463315	3.02	6.1	36
11	Ta_4PbO_{11}	−0.000495176	3.5	5.5	27.2
12	ThO_2	−0.00065306	4.45	4.8	11.9
13	$HfSiO_4$	−0.00089102	5.66	3.7	7
14	$NaNb_4Bi_5O_{18}$	−0.000948134	2.49	6.4	48.2
15	Bi_2O_3	−0.001084116	2.15	7	24.9
16	$TlPd_3O_4$	−0.001142092	1.76	9.8	25.6
17	$La_3HfGa_5O_{14}$	−0.00128972	3.8	4.1	53.4
18	$In_2Pt_2O_7$	−0.00130258	2.48	6.9	5.4
19	$ScTaO_4$	−0.001393008	4.08	5	16.9
20	Tl_4O_3	−0.001521282	0.72	10.2	32.5

High-k dielectrics are materials that have wide band gaps and high dielectric constants [32,33]. For example, these properties are crucial for gate dielectrics and capacitors in semiconductor technology [34]. In this section, we focus on a Pareto frontier to maximize both the band gap and dielectric constants. This helps in searching for suitable materials for high-k dielectrics. Among the top ten materials in the MIPS ranking, there are some materials that have not been thoroughly investigated through experiments. These materials include $Sr_4Ta_2O_9$, $CaHfO_3$, Sr_4As_2O, $Rb_2Ti(WO_4)_3$, and $TaBiO_4$. The MIPS ranking suggests that it would be interesting to experimentally verify these materials. $Sr_4Ta_2O_9$ and $CaHfO_3$ are particularly promising materials as they possess wide band gaps and high dielectric constants and some previous studies were conducted on these materials [35,36]. $Rb_2Ti(WO_4)_3$ and $TaBiO_4$ are also considered promising because they exhibit significantly high dielectric constants and their band gaps are comparable to those of $BaTiO_3$ and $SrTiO_3$, which are commonly used in ceramic capacitor applications [37,38]. On the other hand, Sr_4As_2O has a narrower band gap but a larger dielectric constant. Its inclusion in the MIPS ranking adds diversity to the material choices. Ideally, it would be beneficial to investigate all the materials that are Pareto solutions, but in this case, there are 70 Pareto solutions. The MIPS ranking provides a strategy to prioritize the order in which these materials should be considered. In recent years, the amount of data in computational materials databases have increased drastically. Consequently, the number of Pareto solutions in such databases is also expected to increase. By utilizing these databases, the MIPS ranking is anticipated to effectively facilitate the exploration of new materials.

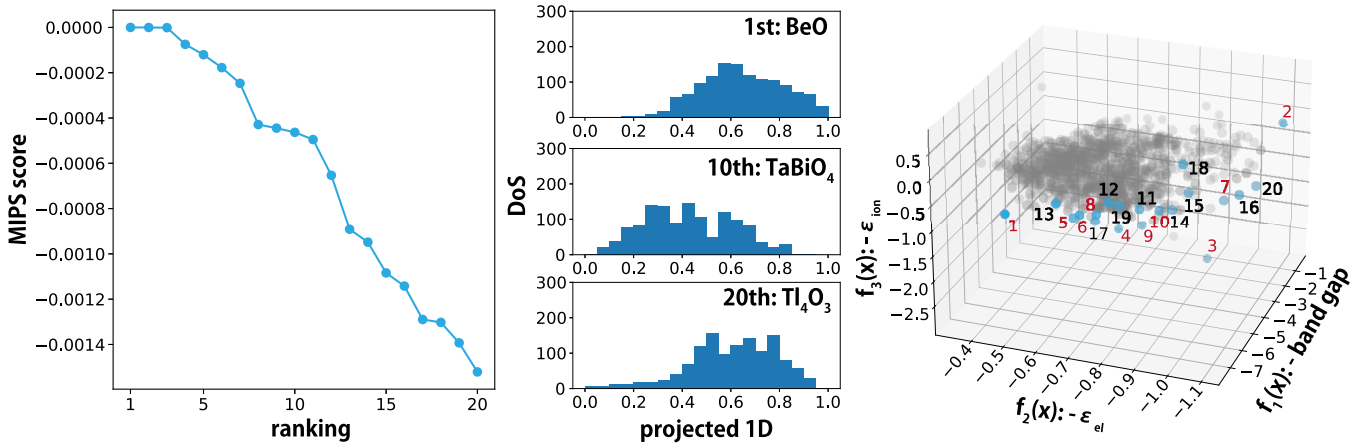

FIG. 5. MIPS values (left column), representative DoS (middle column), and distribution of datapoints (right column) with a three-dimensional Pareto frontier in the property space spanned by the band gap, and the electric and ionic dielectric constants. The first 20 solutions ranked by MIPS are represented by blue datapoints. All DoS are summarized in Fig. S4 [28]. Note that, since our model is developed to minimize objectives, when a problem requires a maximization of the target properties we simply consider their values with the opposite sign.

IV. DISCUSSION AND SUMMARY

We proposed a method to automatically provide the ranking of Pareto solutions in order of decreasing isolation. To measure the isolation of each Pareto solution, we defined the most-isolated Pareto solution (MIPS), which was defined by the free energy. We used a multiobjective optimization approach based on Tchebycheff decomposition to project multiple objective functions to one-dimensional function. To test the proposed method, we used a dataset of semiconductor and a two- and three-dimensional property spaces were considered. Our results show that the most-isolated Pareto solutions can be selected automatically using our ranking. The implementation of our method can be found on GitHub [39] in the PYTHON package called Ranking of Pareto solutions based on Projection Free energy (RPPF).

The ranking of Pareto solutions based on the MIPS can be used in several applications, for instance, in the analysis of simulation results and in association with existing optimization methods. Numerical simulations for materials design can be performed using large datasets of materials, resulting in several Pareto solutions. Using a ranking method for these Pareto solutions, materials with interesting properties can be selected among them. Subsequent research on the selected solutions may deepen our understanding of the unusual properties of materials and contribute to the establishment of guiding principles for the development of new ones.

Moreover, our method can be used in cooperation with black-box optimization [40], which uses machine-learning predictions to select appropriate materials that have not yet been synthesized or simulated [41–44]. Using the MIPS ranking, it would be possible to select a wide variety of materials from Pareto solutions predicted by machine learning. The selected materials will be evaluated by conducting experiments or simulations. Repeating the selection and evaluation processes of the black-box optimization, desired materials will be discovered. Using the MIPS would prevent the selection of materials with similar properties. This could improve the performance of batch experiments where robotic systems for materials development are employed [45–49].

In this study, we focused only on trade-off problems of semiconductors. However, several disciplines, including materials science, physics, and chemistry, aim at solving this type of problem. Some relevant examples of trade-offs that could be addressed by the MIPS ranking based on the projection free energy are the efficiency and power trade-off in thermodynamic systems [50], the permeability and selectivity trade-off characterizing polymer membranes [51], and the cost and performance trade-off in scientific research.

ACKNOWLEDGMENTS

We thank Akira Takahashi, Fumiyasu Oba, Yohei Akimoto, and Koji Hukushima for the useful discussions. Numerical calculations were performed using a supercomputer made available by the Institute for Solid State Physics, University of Tokyo. The research presented in this article was supported by the JSPS Grant-in-Aid for Scientific Research (Grant No. 21H01008), JST-Mirai Program (Grant No. JPMJMI21EA), and MEXT Program: Data Creation and Utilization-Type Material Research and Development Project (Grants No. JPMXP112271469, No. JPMXP1122683430, and No. JPMXP1122712807).

[1] G. Pizzi, A. Cepellotti, R. Sabatini, N. Marzari, and B. Kozinsky, AiiDA: automated interactive infrastructure and database for computational science, Comput. Mater. Sci. **111**, 218 (2016).

[2] A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, and K. A. Persson, Commentary: The materials project: A materials genome approach to accelerating materials innovation, APL Mater. **1**, 011002 (2013).

[3] J. Zhou, L. Shen, M. D. Costa, K. A. Persson, S. P. Ong, P. Huck, Y. Lu, X. Ma, Y. Chen, H. Tang, and Y. P. Feng, 2DMatPedia, an open computational database of two-dimensional materials from top-down and bottom-up approaches, Scientific Data **6**, 86 (2019).

[4] S. S. Borysov, R. M. Geilhufe, and A. V. Balatsky, Organic materials database: An open-access online database for data mining, PLOS ONE **12**, e0171501 (2017).

[5] P. Morgante and R. Peverati, ACCDB: A collection of chemistry databases for broad computational purposes, J. Comput. Chem. **40**, 839 (2019).

[6] M. Nakata and T. Shimazaki, PubChemQC Project: A large-scale first-principles electronic structure database for data-driven chemistry, J. Chem. Inf. Model. **57**, 1300 (2017).

[7] K. Rajan, Materials informatics, Mater. Today **8**, 38 (2005).

[8] G. Pilania, C. Wang, X. Jiang, S. Rajasekaran, and R. Ramprasad, Accelerating materials property predictions using machine learning, Sci. Rep. **3**, 2810 (2013).

[9] R. Gómez-Bombarelli, J. N. Wei, D. Duvenaud, J. M. Hernández-Lobato, B. Sánchez-Lengeling, D. Sheberla, J. Aguilera-Iparraguirre, T. D. Hirzel, R. P. Adams, and A. Aspuru-Guzik, Automatic chemical design using a data-driven continuous representation of molecules, ACS Cent. Sci. **4**, 268 (2018).

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

[11] R. Ramprasad, R. Batra, G. Pilania, A. Mannodi-Kanakkithodi, and C. Kim, Machine learning in materials informatics: recent applications and prospects, npj Comput. Mater. **3**, 54 (2017).

[12] H. Tamaki, H. Kita, and S. Kobayashi, Multi-objective optimization by genetic algorithms: a review, in *Proceedings of IEEE International Conference on Evolutionary Computation, Nagoya, Japan* (IEEE, New York, 1996), pp. 517–522.

[13] A. Konak, D. W. Coit, and A. E. Smith, Multi-objective optimization using genetic algorithms: A tutorial, Reliability Engineering & System Safety **91**, 992 (2006).

[14] C. Coello Coello, Evolutionary multi-objective optimization: a historical view of the field, IEEE Computational Intelligence Magazine **1**, 28 (2006).

[15] J. L. Leyten and S. R. Kurvers, Robustness of buildings and HVAC systems as a hypothetical construct explaining differences in building related health and comfort symptoms and complaint rates, Energy and Buildings **38**, 701 (2006).

[16] P. Hoes, M. Trcka, J. L. M. Hensen, and B. H. Bonnema, Optimizing building designs using a robustness indicator with respect to user behavior, in *Proceedings of Building Simulation 2011: 12th Conference of International Building Performance Simulation Association, Sydney* (IBPSA Australasia and AIRAH, Melbourne, 2011).

[17] Q. Long, C. Wu, T. Huang, and X. Wang, A genetic algorithm for unconstrained multi-objective optimization, Swarm and Evolutionary Computation **22**, 1 (2015).

[18] S. Monghasemi, M. R. Nikoo, M. A. Khaksar Fasaee, and J. Adamowski, A novel multi criteria decision making model for optimizing time–cost–quality trade-off problems in construction projects, Expert Systems with Applications **42**, 3089 (2015).

[19] C. Xu, Y. Ke, Y. Li, H. Chu, and Y. Wu, Data-driven configuration optimization of an off-grid wind/PV/hydrogen system based on modified NSGA-II and CRITIC-TOPSIS, Energy Convers. Manage. **215**, 112892 (2020).

[20] D. Gaudrie, R. L. Riche, V. Picheny, B. Enaux, and V. Herbert, Budgeted multi-objective optimization with a focus on the central part of the pareto front – extended version, arXiv:1809.10482.

[21] D. Gaudrie, R. Le Riche, V. Picheny, B. Enaux, and V. Herbert, Targeting solutions in Bayesian multi-objective optimization: sequential and batch versions, Annals of Mathematics and Artificial Intelligence **88**, 187 (2020).

[22] C. M. Fonseca and P. J. Fleming, Genetic algorithms for multiobjective optimization: Formulation, discussion and generalization, in *Proceedings of the ICGA-93: Fifth International Conference on Genetic Algorithms* (IBPSA Australasia and AIRAH, Melbourne, 1993), pp. 416–423.

[23] T. Asselmeyer, W. Ebeling, and H. Rosé, Smoothing representation of fitness landscapes — the genotypephenotype map of evolution, Biosystems **39**, 63 (1996).

[24] H. Rosé, W. Ebeling, and T. Asselmeyer, The density of states — a measure of the difficulty of optimisation problems, in *Parallel Problem Solving from Nature — PPSN IV*, edited by H.-M. Voigt, W. Ebeling, I. Rechenberg, and H.-P. Schwefel, (Springer, Berlin, 1996), pp. 208–217.

[25] K. Miettinen, *Nonlinear Multiobjective Optimization* (Springer, New York, 1998).

[26] Q. Zhang and H. Li, MOEA/D: A multiobjective evolutionary algorithm based on decomposition, IEEE Transactions on Evolutionary Computation **11**, 712 (2007).

[27] X. Ma, Q. Zhang, G. Tian, J. Yang, and Z. Zhu, On Tchebycheff decomposition approaches for multiobjective evolutionary optimization, IEEE Transactions on Evolutionary Computation **22**, 226 (2018).

[28] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevMaterials.7.093804 for Fig. S1: MIPS ranking depending on T^*; Fig. S2: All DoS for a convex two-dimensional Pareto frontier; Fig. S3: All DoS for a concave two-dimensional Pareto frontier; Fig. S4: All DoS for a three-dimensional Pareto frontier; Table S1: MIPS ranking for a three-dimensional Pareto frontier.

[29] A. Takahashi, Y. Kumagai, J. Miyamoto, Y. Mochizuki, and F. Oba, Machine learning models for predicting the dielectric constants of oxides based on high-throughput first-principles calculations, Phys. Rev. Mater. **4**, 103801 (2020).

[30] F. T. Liu, K. M. Ting, and Z.-H. Zhou, Isolation forest, in *2008 Eighth IEEE International Conference on Data Mining* (IEEE, New York, 2008), pp. 413–422.

[31] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, LOF: identifying density-based local outliers, in *ACM Sigmod Record*, Vol. 29 (ACM, New York, 2000), pp. 93–104.

[32] G. Ribes, J. Mitard, M. Denais, S. Bruyere, F. Monsieur, C. Parthasarathy, E. Vincent, and G. Ghibaudo, Review on high-k dielectrics reliability issues, IEEE Trans. Device Mater. Reliab. **5**, 5 (2005).

[33] J. Kittl, K. Opsomer, M. Popovici, N. Menou, B. Kaczer, X. Wang, C. Adelmann, M. Pawlak, K. Tomida, A. Rothschild, B.

Govoreanu, R. Degraeve, M. Schaekers, M. Zahid, A. Delabie, J. Meersschaut, W. Polspoel, S. Clima, G. Pourtois, W. Knaepen *et al.*, High-k dielectrics for future generation memory devices, Microelectron. Eng. **86**, 1789 (2009).

[34] J. Robertson, Band offsets, Schottky barrier heights, and their effects on electronic devices, J. Vac. Sci. Technol., A **31**, 050821 (2013).

[35] M. Lukosius, C. Wenger, T. Schroeder, J. Dabrowski, R. Sorge, I. Costina, H.-J. Müssig, S. Pasko, and C. Lohe, Atomic - vapour - deposited HfO_2 and $Sr_4Ta_2O_9$ layers for metal-insulator-metal applications, Microelectron. Eng. **84**, 2165 (2007).

[36] S. Lee, J. Lim, and Y. S. Lee, Violet-blue emission property of the cerium-ion doped $CaHfO_3$: Doping and temperature dependence, Luminescence **33**, 1257 (2018).

[37] J.-H. Jeon, Effect of $SrTiO_3$ concentration and sintering temperature on microstructure and dielectric constant of $Ba_{1-x}Sr_xTiO_3$, J. Eur. Ceram. Soc. **24**, 1045 (2004).

[38] H. Tang, Y. Lin, and H. A. Sodano, Synthesis of high aspect ratio $BaTiO_3$ nanowires for high energy density nanocomposite capacitors, Adv. Energy Mater. **3**, 451 (2013).

[39] https://github.com/tsudalab/RPPF.

[40] K. Terayama, M. Sumita, R. Tamura, and K. Tsuda, Black-box optimization for automated discovery, Acc. Chem. Res. **54**, 1334 (2021).

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

[42] K. Kitai, J. Guo, S. Ju, S. Tanaka, K. Tsuda, J. Shiomi, and R. Tamura, Designing metamaterials with quantum annealing and factorization machines, Phys. Rev. Res. **2**, 013319 (2020).

[43] R. Tamura, T. Osada, K. Minagawa, T. Kohata, M. Hirosawa, K. Tsuda, and K. Kawagishi, Machine learning-driven optimization in powder manufacturing of Ni-Co based superalloy, Mater. Des. **198**, 109290 (2021).

[44] Y. Motoyama, R. Tamura, K. Yoshimi, K. Terayama, T. Ueno, and K. Tsuda, Bayesian optimization package: PHYSBO, Comput. Phys. Commun. **278**, 108405 (2022).

[45] B. P. MacLeod, F. G. L. Parlane, T. D. Morrissey, F. Häse, L. M. Roch, K. E. Dettelbach, R. Moreira, L. P. E. Yunker, M. B. Rooney, J. R. Deeth, V. Lai, G. J. Ng, H. Situ, R. H. Zhang, M. S. Elliott, T. H. Haley, D. J. Dvorak, A. Aspuru-Guzik, J. E. Hein, and C. P. Berlinguette, Self-driving laboratory for accelerated discovery of thin-film materials, Sci. Adv. **6**, eaaz8867 (2020).

[46] R. Shimizu, S. Kobayashi, Y. Watanabe, Y. Ando, and T. Hitosugi, Autonomous materials synthesis by machine learning and robotics, APL Mater. **8**, 111110 (2020).

[47] L. M. Roch, F. Häse, C. Kreisbeck, T. Tamayo-Mendoza, L. P. E. Yunker, J. E. Hein, and A. Aspuru-Guzik, ChemOS: An orchestration software to democratize autonomous discovery, PLOS ONE **15**, e0229862 (2020).

[48] S. Matsuda, G. Lambard, and K. Sodeyama, Data-driven automated robotic experiments accelerate discovery of multicomponent electrolyte for rechargeable Li-O_2 batteries, Cell Reports Physical Science **3**, 100832 (2022).

[49] Y. Li, L. Xia, Y. Fan, Q. Wang, and M. Hu, Recent advances in autonomous synthesis of materials, ChemPhysMater **1**, 77 (2022).

[50] N. Shiraishi, K. Saito, and H. Tasaki, Universal Trade-Off Relation between Power and Efficiency for Heat Engines, Phys. Rev. Lett. **117**, 190601 (2016).

[51] H. B. Park, J. Kamcev, L. M. Robeson, M. Elimelech, and B. D. Freeman, Maximizing the right stuff: The trade-off between membrane permeability and selectivity, Science **365**, 1137 (2017).