ARTICLES SCIENCE CHINA Materials

mater.scichina.com link.springer.com

Published online 16 March 2020 | https://doi.org/10.1007/s40843-019-1255-4
*Sci China Mater* 2020, 63(6): 1024–1035

# Bayesian optimization based on a unified figure of merit for accelerated materials screening: A case study of halide perovskites

Xiwen Chen[1], Chen Wang[1], Zhenzhu Li[1], Zhufeng Hou[2] and Wan-Jian Yin[1,3*]

**ABSTRACT** **The figure of merit is of crucial importance in materials design to search for candidates with optimal functionality. In the field of photovoltaics, the bandgap ($E_g$) is a well-recognized figure of merit for screening solar cell absorbers subject to the Shockley-Queisser limit. In this paper, the bandgap as the figure of merit is challenged since an ideal solar cell absorber requires multiple criteria such as stability, optical absorption, and carrier lifetime. Multiple criteria make the quantitative description of material candidates difficult and computationally time-consuming. Taking halide perovskites as an example, we combine thermodynamic stability ($\Delta H_d$) and $E_g$ into a unified figure of merit and use Bayesian optimization (BO) to accelerate materials screening. We have found that, in comparison to an exhaustive search *via* multiple parameters, BO based on the unified figure of merit can screen optimal candidates ($E_{g,PBE}$ between 0.6–1.2 eV, $\Delta H_d$>−29 meV per atom) more efficiently. Therefore, the proposed method opens a viable route for the search of optimal solar cell absorbers from a large amount of material candidates with less computational cost.**

**Keywords:** descriptor, machine learning, Bayesian optimization, DFT calculations

## INTRODUCTION

Design of novel materials in the fields of catalysis [1–3] and photovoltaics [4–10] is highly dependent on the figure of merit, which relates material functionality to material parameters. For a long time, the bandgap ($E_g$) has been considered as a crucial material parameter in the search of novel semiconductors for photovoltaic applications subject to the Shockely-Queisser limit [11]. Accordingly, a number of novel materials have been considered as promising semiconductor alternatives to mature solar cell absorbers, such as $Cu_2S$ ($E_g$= 1.2 eV) [12,13], $FeS_2$ ($E_g$=1.0 eV) [14,15], $Zn_3P_2$ ($E_g$= 1.5 eV) [16–18], $Cs_2CuInCl_6$ ($E_g$=1.4 eV) [4], and $Cs_2AgInBr_6$ ($E_g$=1.5 eV) [4,7]. However, these materials did not show promising performance, which highlights the importance of other material parameters such as stability, optical absorption, and carrier lifetime for the identification of solar cell absorbers.

By considering multiple material parameters, the hierarchical structure for material design of solar cell absorbers is shown in Fig. 1. The key parameters may include bandgap, decomposition energy, effective mass, exciton binding energy, carrier lifetime, etc. These material parameters are not easily accessed and are usually derived by state-of-the-art methods such as density functional theory (DFT) calculations. As a result, an exhaustive search based on DFT, where all material parameters for each material candidate are calculated (Fig. 2a), becomes the prevalent screening strategy. For example, Zhao *et al.* [8] considered five parameters: decomposition energy, bandgap, electron and hole effective mass, and exciton binding energy, for sixty-four perovskites, finally selecting eleven perovskites for solar cell applications. Huo *et al.* [9] used high-throughput computing based on phase stability, formation energy, bandgap, effective mass, and optical absorption to select five compounds among 168

[1] College of Energy, Soochow Institute for Energy and Materials InnovationS (SIEMIS), and Jiangsu Provincial Key Laboratory for Advanced Carbon Materials and Wearable Energy Technologies, Soochow University, Suzhou 215006, China
[2] State Key Laboratory of Structural Chemistry, Fujian Institute of Research on the Structure of Matter, Chinese Academy of Sciences, Fuzhou 350002, China
[3] Key Lab of Advanced Optical Manufacturing Technologies of Jiangsu Province & Key Lab of Modern Optical Technologies of Education Ministry of China, Soochow University, Suzhou 215006, China
* Corresponding author (email: wjyin@suda.edu.cn)

 © Science China Press and Springer-Verlag GmbH Germany, part of Springer Nature 2020 

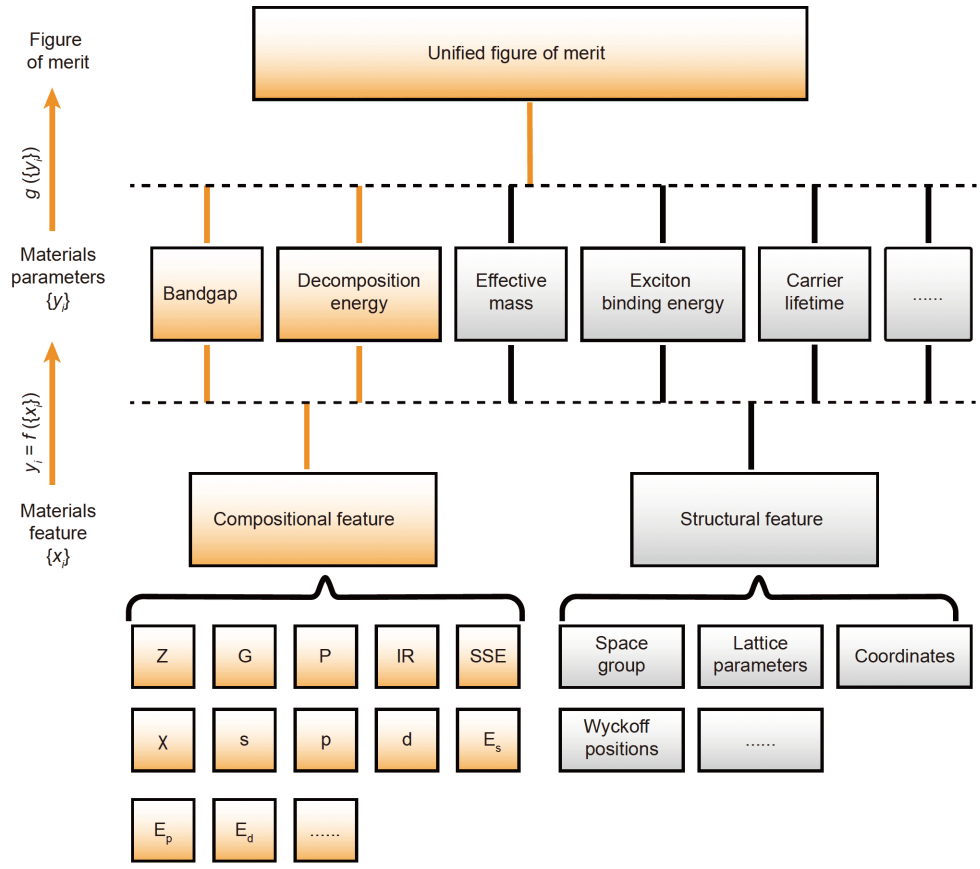



**Figure 1** The hierarchical structure of materials design for solar cell absorber. The color-highlighted part are considered in this work.

types of chalcogenide single perovskites. Jacobs *et al.* [10] sequentially considered energy above the convex hull, the Perdew-Burke-Ernzerhof (PBE)-bandgap, the Heyd-Scuseria-Ernzerhof (HSE)-bandgap, and spectroscopic limited maximum efficiency (SLME) calculations, and finally selected 15 perovskites for potential solar cell applications.

When the number of material candidates is huge, direct DFT calculation on material parameters is computationally heavy. Machine learning has become an alternative and effective approach for materials screening [5,19–25]. Based on existing data, machine learning can learn the correlation between material parameters $\{y_i\}$ and material features $\{x_i\}$ and infer a relation $y_i=f(\{x_i\})$ (Fig. 1) [26]. The features should be easily accessed and can be quantified as a dataset $\{x_i\}$, which can be used as numerical representation of material parameters $\{y_i\}$. Then, for new materials with given $\{x_i\}$, material parameters can be obtained in a much faster way than with direct DFT calculations. Therefore, the machine learning approach can be applied to a huge number of candidates when direct DFT calculation for material parameters is unaffordable. For example, Lu *et al.* [19] predicted the bandgaps of 5,158 unexplored possible hybrid organic-inorganic perovskites (HOIPs) based on a database of 212 reported HOIP bandgaps. They performed an exhaustive search to select 218 stable perovskites with the proper bandgap for solar cells. Weston *et al.* [20] used machine learning to predict the band gaps of 1,568 kesterite $I_2$-II-IV-$VI_4$ compounds based on a bandgap database of 200 kesterite compounds. They found 242 materials with a bandgap in the optimum range of 1.2–1.8 eV, whose stability was assessed by calculating the energy above the convex hull, and identified 25 compounds for photovoltaic devices.

Despite the successful attempts of these methodologies, there are still unsolved issues and challenges in the practical applications of machine learning. First, the exhaustive search approach can waste computational time since it is not able to avoid unnecessary calculations. An

 © Science China Press and Springer-Verlag GmbH Germany, part of Springer Nature 2020 

example is the case of bandgap calculations for extremely unstable compounds. It would become extremely computationally expensive when the number of candidate materials is large. Second, multiple parameters are not able to quantitatively describe the potential of a material as a solar cell absorber, which would make direct comparison between two compounds difficult. For example, it is difficult to compare an extremely stable compound with a bandgap of 1.61 eV to an intermediately stable compound with a bandgap of 1.50 eV. A common bandgap criterion of 1.0–1.6 eV excludes a stable compound with a bandgap of 1.61 eV. Third, current machine learning models are trained based on a small data set and make predictions for a large amount of data, therefore, the validity of the machine learning model is still a problem.

As shown in Fig. 2b, Bayesian optimization (BO) is a powerful technique for finding optimal values of function $y_i=f(\{x_i\})$ within a given $\{x_i\}$ set [26,27]. Different from one-shot machine-learning approaches, BO can be initialized with a small group of data and evolve into a reasonably good model during sampling-modeling iterations. Due to its dynamical and efficient nature, it is also called the adaptive machine-learning method [28–30].

In this paper, using halide perovskites as a case study, we use a unified figure of merit that combines the material parameters of bandgap and thermodynamic stability, and use BO to screen materials for solar cell absorbers. Perovskite $ABX_3$ is a typical crystal structure with chemical flexibility on all A, B, and X sites and, therefore, provides a treasure trove of material candidates for high-throughput and machine-learning studies in material design. The prototype phase of perovskite ($ABX_3$) is the cubic $Pm\bar{3}m$ phase, which has $O_h$ symmetry. Double perovskite ($A_2B(I)B(III)X_6$) replaces two divalent $B^{2+}$ ions with one monovalent ion $B(I)^{1+}$ and one trivalent $B(III)^{3+}$ ion on the basis of perovskite ($ABX_3$), with the space group becoming $Fm\bar{3}m$. Based on experimental reports [31–44], we chose elements in B(I) and B(III) sites that are adjacent to elements from known perovskite materials. We considered 438 perovskites in total, including 18 single perovskites $ABX_3$ (A=K, Rb, Cs; B=Pb, Sn; X=Cl, Br, I), and 420 double perovskites $A_2B(I)B(III)X_6$ (A=Na, K, Rb, Cs; B(I)=Na, K, Rb, Cu, Ag, In, Tl; B(III)=Sb, Bi, Ga, In, Tl; X=Cl, Br, I) as shown in Fig. 3.

As a demonstration of the algorithm, we only considered decomposition energies and bandgap, as shown in Fig. 1. The algorithm can be extended to other material parameters such as effective mass, optical absorption, and carrier lifetime in the future if these data are available. In

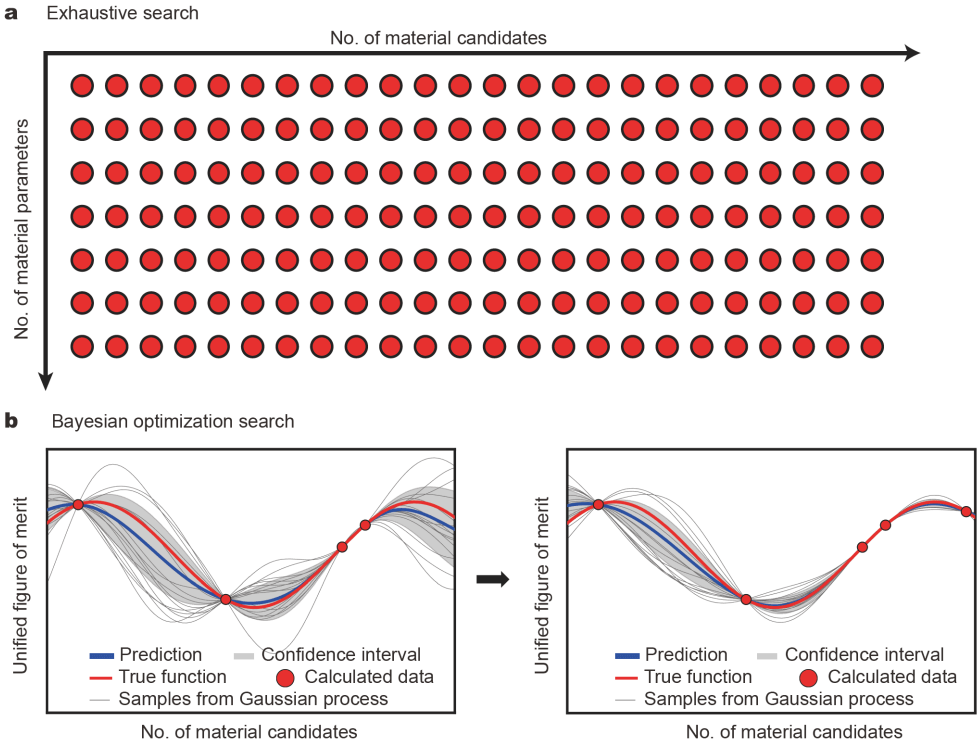



**Figure 2** Schematic approach for (a) exhaustive search, and (b) Bayesian optimization.


© Science China Press and Springer-Verlag GmbH Germany, part of Springer Nature 2020

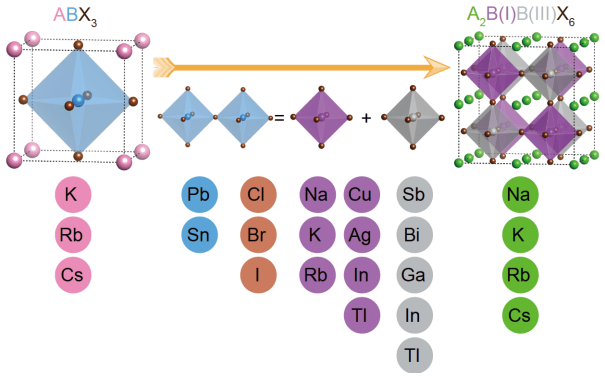



**Figure 3** Material systems considered in this work, including 18 single perovskites and 420 double perovskites.

the current work, an exhaustive search requires 438 (number of materials) × 2 (number of parameters)=876 calculations to identify the optimal 23 materials within the defined criteria. We have found that BO is three (two) times more efficient than an exhaustive search for identifying 80% (60%) of stable perovskites (with optimal band gaps). Taking the stability and bandgap as a unified figure of merit, BO is twice as fast as an exhaustive search for selecting 80% of promising solar cell absorbers. Since BO is a type of recommendation algorithm, which usually starts from a small amount of data and adaptively corrects its model, it is more appropriate for experimental implementation. The methods proposed here demonstrate more efficient searching power for the selection of candidates with the proper bandgap and decomposition energies than a conventional exhaustive search.

## METHODS

### First-principles calculations

First-principles calculations were performed based on DFT [45,46] as implemented using the VASP 5.4 code [47]. For the exchange-correlation functional, we used the generalized gradient approximation (GGA) [48] within the PBE. The plane wave cutoff energy was 300 eV. The total energies of perovskites and their secondary phases were calculated. The same k-mesh density was chosen for all the structures by setting KSPACING=0.12. The crystal structure was fully relaxed until all atomic forces were less than 0.01 eV $Å^{-1}$ and the convergence criterion for the total energy was 0.5 meV at each self-consistent field iteration.

For the calculation of band gaps, the strings of k-points were set to Γ-Z-T-Y-S with 40 k-points between two neighboring high symmetry k-points. In the current work, fundamental bandgaps, i.e., the energy level difference between the valence band maximum (VBM) and conduction band minimum (CBM), were considered without distinguishing between direct or indirect bandgap. Because the error of the HSE bandgap relative to the experimental bandgap is 0.3 eV, the HSE bandgap criterion value around the ideal 1.4 eV bandgap for a single junction solar cell was taken as 1.1–1.7 eV. GGA-calculated bandgaps were corrected according to the correlation function $E_g^{HSE} = 1.2607 \times E_g^{PBE} + 0.2246$ [10]. Thus, the PBE bandgap should be 0.6–1.2 eV.

### Linear programming (LP) for thermodynamic stability

Thermodynamic stability can be quantified by the decomposition energy, which is defined as the total energy difference between the competing secondary phases and the material of interest. As shown in Fig. 4a, the thermodynamic stability of double halide perovskites $A_2B(I)B(III)X_6$ was conventionally evaluated through a particular decomposition pathway, for example, to binary competing phases, i.e., $A_2B(I)B(III)X_6 \rightarrow 2AX + B(I)X + B(III)X_3$ [7,8]. In principle, thermodynamic stability can only be quantified by the decomposition pathways of binary competing secondary phases. However, Xiao *et al.* [42] proposed that the decomposition pathways to binary phases are not optimal pathways for many halide double perovskites, so they may overestimate their stability. In order to consider more decomposition pathways, Filip *et al.* [49] employed the quaternary phase diagram that contains more than 5,000 decomposition pathways. They considered 17 decomposition pathways from a plane of the quaternary phase diagram to study the stability of 18 double perovskites. Despite of considering multiple decomposition pathways, the calculations still cannot guarantee the optimal decomposition pathway, which has a minimum decomposition energy among all the possible pathways.

In order to determine the optimal decomposition pathway, we have calculated the total energies of all possible competing phases that exist in the inorganic crystal structure database (ICSD) and the powder diffraction file (PDF) card. We have found that there are important competing phases missing in the ICSD, for example, $Cs_3In_2Br_9$, which would result in unphysical results. An additional five secondary phases, $K_3BiCl_6$, $K_3Bi_2Cl_9$, $KCu_2Cl_3$, $Rb_3BiCl_6$ and $Rb_3Sb_2Cl_9$, were considered *via* atomic replacement since we have found their close counterparts do exist, although they do not exist in either the ICSD or PDF. There are 235 secondary phases

 © Science China Press and Springer-Verlag GmbH Germany, part of Springer Nature 2020 

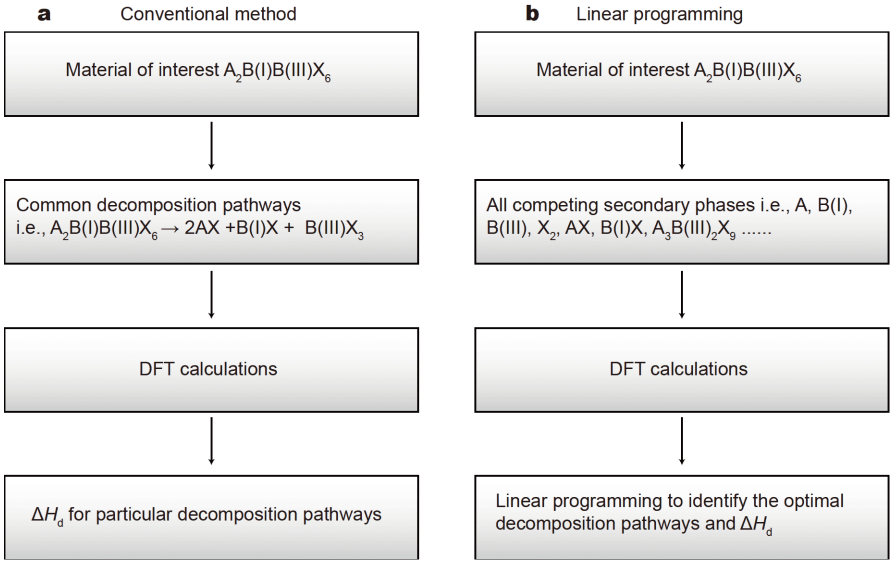



**Figure 4** The comparison of (a) conventional methods and (b) linear programming to calculate the decomposition energies of materials.

considered in total (Table S1).

LP is an efficient approach for determining the optimal decomposition pathway and consists of linear constraints and an objective function [50,51]. Assuming Equation (1) is the decomposition pathway, Equation (2) is the corresponding decomposition energy. We choose Equations (3–7) as linear constraints on the basis of mass conservation and Equation (2) as the objective function. If the objective function is the lowest number, then $n_i$ is available. $A_2B(I)B(III)X_6$ and $A_{li}B(I)_{mi}B(III)_{si}X_{ti}$ represent halide double perovskite and secondary phases, respectively. Halide single perovskites are treated as the same as halide double perovskites (Fig. 4b).

$$A_2B(I)B(III)X_6 \rightarrow \sum_{i=1}^{i} n_i \cdot A_{l_i}B(I)_{m_i}B(III)_{s_i}X_{t_i}, \quad (1)$$

$$\Delta H_d = \sum_{i=1}^{i} n_i \cdot E\left(A_{l_i}B(I)_{m_i}B(III)_{s_i}X_{t_i}\right) - E\left(A_2B(I)B(III)X_6\right), \quad (2)$$

$$\sum_{i=1}^{i} n_i \cdot l_i = 2, \quad (3)$$

$$\sum_{i=1}^{i} n_i \cdot m_i = 1, \quad (4)$$

$$\sum_{i=1}^{i} n_i \cdot s_i = 1, \quad (5)$$

$$\sum_{i=1}^{i} n_i \cdot t_i = 6, \quad (6)$$

$$0 \le n_i \le \min\left\{\frac{2}{l_i}, \frac{1}{m_i}, \frac{1}{s_i}, \frac{6}{t_i}\right\}. \quad (7)$$

We calculated the lowest decomposition energies ($\Delta H_d$) of perovskites that have been reported, as shown in Table S2. According to the results, we concluded that double perovskites could been synthesized when $\Delta H_d$ is greater than 0 meV per atom, [32,37,40,52–59] and there is no synthesis in the experiment when $\Delta H_d$ is lower than −29 meV per atom. It is worth noting that $\Delta H_d$ is between −29–0 meV per atom, that is, $Cs_2AgSbCl_6$ [44,60], $Rb_2NaInCl_6$ [53], $Cs_2AgSbBr_6$ [44], $Cs_2NaInBr_6$ [53], $Cs_2KBiCl_6$ [61], $Cs_2NaSbBr_6$ [54], $Cs_2NaBiI_6$ [43] and $Cs_2AgBiI_6$ [62] are experimentally synthesized by synthesizing nanocrystals or controlling the experimental temperature. We listed the bandgap, decomposition energies and optimal decomposition pathways in Table S3. The comparison of our decomposition energies and optimal decomposition pathways with others is shown in Table S4. Since some results contain the particular decomposition pathway to binary competing secondary phases, they are not listed in Table S4.

### BO method

The BO is a well-established technique for black-box optimization [63–65]. Bayesian prediction models were employed to predict the black-box function, where the uncertainty of the predicted function was also evaluated as predictive variance. The next candidates for experiments or calculations were chosen based on the predicted

 © Science China Press and Springer-Verlag GmbH Germany, part of Springer Nature 2020 

values and variances (Fig. 2b). In materials science, we are aware of at least four successful applications of BO: elastic properties, melting temperature, lattice thermal conductivity, and the design of a grain boundary interface [65]. In this work, we used the open source library termed COMmon Bayesian Optimization library (COMBO) to perform the optimization process. COMBO has been successfully used in other material search areas [27,66]. For the BO portion, we used Gaussian processes to predict the black-box function. A fundamental characteristic of Gaussian processes [67] is that the uncertainty of the predicted function is also evaluated as predictive variance. Using the predicted values and variance, the acquisition function was calculated based on Thompson sampling [68].

In the black-box function, decomposition energy and bandgaps are considered to be merit properties of materials. We are more interested in the application of halide perovskite materials in photovoltaics, therefore, our concern is the maximum $\Delta H_d$ and a PBE bandgap between 0.6 and 1.2 eV. $\Delta H_d$ and $-|E_g-0.9|$ are the targets of the black-box function. We combined the two targets and because of the difference in the order of magnitude between them, took the target as a unified figure of merit. These targets could be regarded as the dependent variables of the black-box function and the independent variables of the black-box function were the physical quantities related to $\Delta H_d$ and $E_g$ of halide perovskite materials. We considered the properties of elements that affect the thermodynamic stability and bandgap of halide perovskites. Therefore, we selected 14 features, as listed in Table 1 [69]. This black-box function gradually becomes clear after the cyclic iteration of BO.

Since BO is very effective in finding at least one of the top-k points, it is necessary to construct a function of $\Delta H_d$ and $E_g$ that is the maximum value for a suitable photovoltaic material. As shown in Fig. 5, twenty perovskites were randomly selected to calculate their $\Delta H_d$ and $E_g$. Through BO, a suitable material was recommended and at the same time, its $\Delta H_d$ and $E_g$ were calculated. The properties of the material were added to the model again, optimized, and then one suitable material was recommended. This process was repeated until all suitable materials were selected.

## RESULTS AND DISCUSSION

For BO, we first choose proper features. Fourteen elemental features listed in Table 1 [69] have been considered. The features can be classified into three groups: location in the periodic table of elements (Nos. 1–3), size (No. 4), and electronic structure (Nos. 5–14). Use of all fourteen features may lead to a waste of resources and over-fitting. Therefore, the first task is to select the proper feature(s). Ionic radii are considered as a key feature to determine the thermodynamic stability of perovskites [70]. To improve model precision, other features are added according to the principle that a good feature set should include features that are minimally correlated. According to the feature correlation map (Fig. 6), five sets of features were selected, as shown in Table 2, to test the model performance.

To show the efficiency of BO for materials screening, we first only focus on the stability and set the target as the number of calculations (referred as "step") to get the most stable perovskite among a total of 438. In principle, an

**Table 1** The 14 features used in BO

| No. | Feature symbol | Feature introduction |
|---|---|---|
| 1 | Z | Atomic number |
| 2 | G | The group of atom in periodic table of elements |
| 3 | P | The period of atom in periodic table of elements |
| 4 | IR | Ionic radius |
| 5 | s | The number of electrons in the s orbit |
| 6 | p | The number of electrons in the p orbit |
| 7 | d | The number of electrons in the d orbit |
| 8 | $E_s$ | s orbital energy |
| 9 | $E_p$ | p orbital energy |
| 10 | $E_d$ | d orbital energy |
| 11 | h | Highest occupied atomic level |
| 12 | l | Lowest unoccupied atomic level |
| 13 | SSE | The atomic solid state energy [69] |
| 14 | $\chi$ | Pauling's electronegativity |

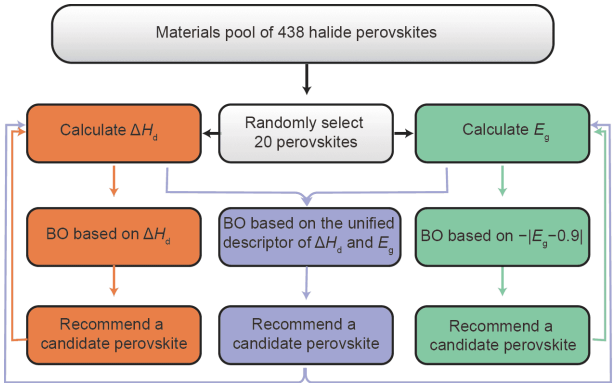



**Figure 5** Schematics of the searching for solar cell absorbers combing density of function (DFT) and BO.


© Science China Press and Springer-Verlag GmbH Germany, part of Springer Nature 2020

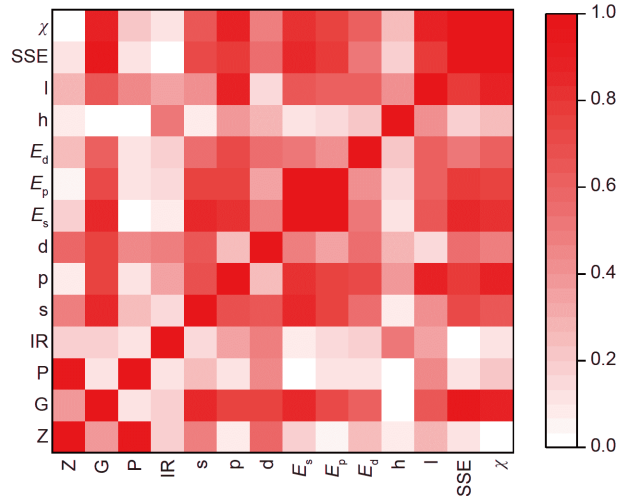



**Figure 6** Pearson correlation map for features. The color varies with the absolute value of the associated Pearson correlation coefficient. The lighter the tone used, the less significant is the corresponding correlation.

**Table 2** The 5 feature sets used in the BO

| Nos. | Feature sets |
|---|---|
| 1 | IR |
| 2 | IR, SSE |
| 3 | P, IR, SSE |
| 4 | P, IR, $E_s$ |
| 5 | P, IR, $E_p$, $E_d$ |

exhaustive search requires the calculation of 438 perovskites to identify the most stable, shown in Fig. 7a. Based on probability statistics, the average number of calculations (step) for an exhaustive search to find the most stable perovskite is 219, half of the total number of possibilities.

The average numbers of steps to find the most stable perovskite using BO with five sets of features (Table 2) are shown in Table 3. The results firstly show that the BO has much improved performance by reducing the average number of steps from 211 of the 100 exhaustive searches to below 100. The results are consistent with the probability of finding the most stable perovskite within a fixed number of steps (for example, 80 in Fig. 8). The results next show that the feature set No. 5 performs the best with an average number of steps of 53, four times faster than an exhaustive search. One example of the BO process and its comparison with an exhaustive search are shown in Fig. 7. A clear trend is observed that BO is able to select the most stable perovskites with higher decomposition energy. In addition, the remaining perovskites are more and more unstable which have lower decomposition energy. It clearly demonstrates that BO can effectively select the most stable compounds with fewer calculation steps.

BO is usually used to identify a batch of materials with optimal properties, not only the most suitable one. Therefore, another way to show the performance of BO is to see how many steps are required to select a fraction of targeted materials. According to a systematic comparison of theory and experiment, a decomposition energy ($\Delta H_d$) value of −29 meV per atom has been set as the lower limit for stable perovskites [62,71]. Our goal is to select several stable perovskites using fewer calculation steps. Fig. 9a shows the step number for an exhaustive search and BO to select a fraction of stable perovskites. BO exhibits much better performance than the exhaustive search. There are 107 stable perovskites using the exhaustive search and BO can identify 80% of the stable perovskites within 104 calculations, which is three times more efficient than the exhaustive search.

According to calculations by Jacobs *et al.* [10], the optimal range for the PBE-calculated bandgap is within 0.6–1.2 eV, with the optimal value being around 0.9 eV. Therefore, we established the formula as the objective

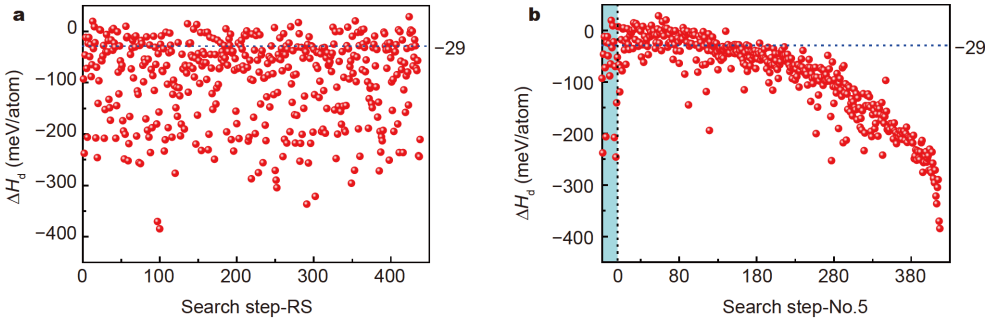



**Figure 7** Screening for stable perovskites through (a) random search (RS), and (b) BO with No. 5 feature set in Table 2. Materials before 0 are used as training dataset for BO.


© Science China Press and Springer-Verlag GmbH Germany, part of Springer Nature 2020

**Table 3** For 100 times random selection of 20 initial perovskites, the average steps were needed by using exhaustive search or BO to find the most stable perovskite.

| Feature sets | The average step |
|---|---|
| RS | 211 |
| No. 1 | 71 |
| No. 2 | 91 |
| No. 3 | 66 |
| No. 4 | 63 |
| No. 5 | 53 |

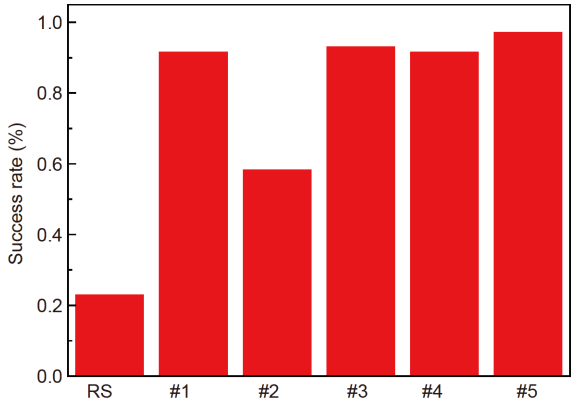



**Figure 8** Success rate (within 400 times of BO calculations) for identifying the most stable perovskites within 80 steps. The features used in the search are shown along the horizontal axis: the numbers denote the features listed in Table 2.

function for BO:

$$y = -\left|E_g - 0.9\right|.$$

A similar procedure for finding a perovskite with high decomposition energy has been applied to select a perovskite with a PBE bandgap close to 0.9 eV. The comparison of performance between an exhaustive search and BO is shown in Fig. 9b. The BO shows better performance than the exhaustive search. For example, there is a total of 84 perovskites with a PBE bandgap ranging from 0.6 to 1.2 eV. By using the exhaustive search, 234 calculations are required to find 50% (60% of total) suitable perovskite materials. In contrast, BO only requires 133 calculations. BO-aided materials screening is able to accelerate the discovery of stable perovskites and perovskites with the proper bandgap.

When selecting good solar cell absorbers, one should consider both properties of the bandgap and thermodynamic stability. If an exhaustive search is used, 438 stability calculations and 438 bandgap calculations are required to find 23 suitable materials (Fig. 10a and Table 4 [37,40,42,44,62,72]). To accelerate materials screening with both criteria of bandgap and decomposition energy, we combine both parameters into one figure of merit. Constructing the analytical form of this figure of merit is important since it directly impacts the performance of BO. At this time, there is no specific approach for constructing the most efficient figure of merit. In view of the advantages of BO in searching the maximum value, and that the perovskites with the largest decomposition energy and the bandgap closest to 0.9 eV are the most suitable, a linear model ($\Delta H_d - 100\times|E_g - 0.9|$) was established to search these perovskite materials in the highlight of Fig. 10a as soon as possible. However, this model can easily find the materials that are particularly stable, but the bandgap is particularly inappropriate. Thus, we built the exponential model ($e^{-\left(\frac{\Delta H_d}{29}\right)^2-(E_g-0.9)^2}$), where $\Delta H_d$ and $E_g$ are in units of meV and eV, respectively. This model represents the distance from a point to the decomposition energy of 0 meV per atom and a bandgap of 0.9 eV in Fig. 10a.

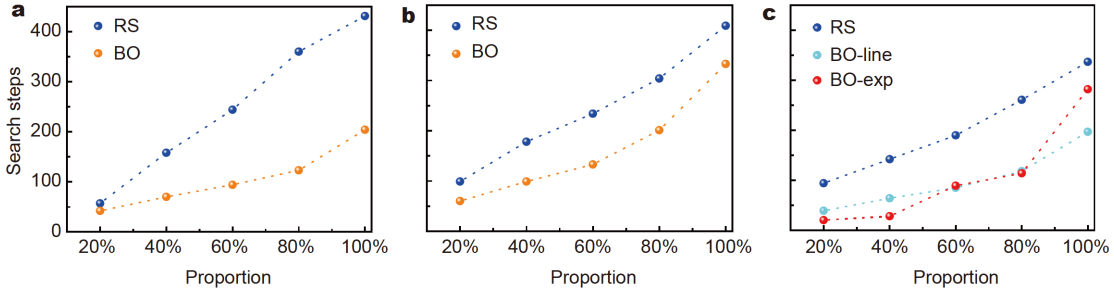



**Figure 9** The comparisons of number of steps for RS and BO to select a fraction of perovskites that (a) have $\Delta H_d$ greater than −29 meV per atom; (b) have $E_g$ between 0.6–1.2 eV; (c) meet both criteria in (a) and (b). In (c), the performances of two unified figures of merit (line: $\Delta H_d - 100\times|E_g - 0.9|$, exp: $e^{-\left(\frac{\Delta H_d}{29}\right)^2-(E_g-0.9)^2}$) are shown.


© Science China Press and Springer-Verlag GmbH Germany, part of Springer Nature 2020

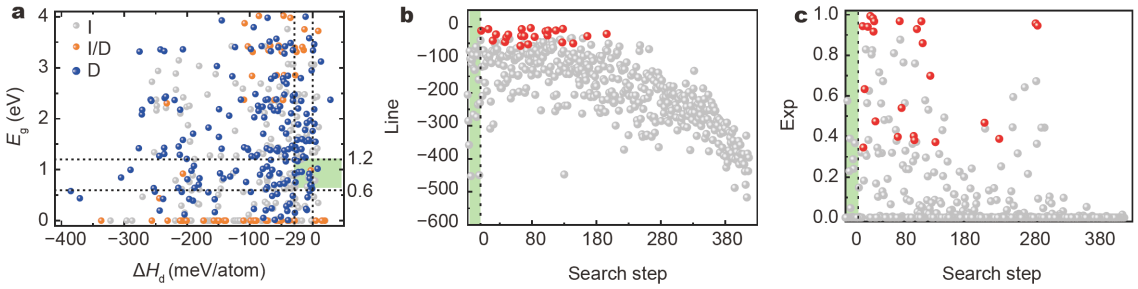



**Figure 10** The perovskites for solar cell absorbers screened by (a) exhaustive search, (b) BO based on the linearly combined figure of merit and (c) BO based on the exponentially combined figure of merit. D, I and I/D represent direct bandgap, indirect bandgap, and quasi-direct bandgap respectively. Materials before 0 are used as training dataset for BO.

**Table 4** The promising perovskite compounds hunted by BO based on thermodynamic stability and PBE-bandgap. The results of the synthesis or unsynthesis on the experiment are also shown.

| Nos. | Compound | $\Delta H_d$ (meV per atom) | $E_g$ (eV) | D or I | Exp | Refs. |
|---|---|---|---|---|---|---|
| 1 | $Cs_2AgInCl_6$ | 7.8806 | 1.017 | D | y | [40] |
| 2 | $Cs_2AgSbBr_6$ | −2.0142 | 0.855 | I | y | [44] |
| 3 | $Cs_2TlTlCl_6$ | 4.1792 | 0.787 | I | y | [37] |
| 4 | $CsSnCl_3$ | −5.3077 | 0.919 | D | y | [72] |
| 5 | $Rb_2AgSbBr_6$ | −6.9130 | 0.856 | I | | |
| 6 | $Rb_2AgInCl_6$ | −2.9178 | 0.985 | I/D | | |
| 7 | $Cs_2InBiBr_6$ | −4.8178 | 0.828 | D | n | [42] |
| 8 | $K_2AgSbBr_6$ | −6.9195 | 0.830 | I | | |
| 9 | $Cs_2NaTlBr_6$ | −2.0768 | 0.702 | D | | |
| 10 | $Cs_2TlSbI_6$ | −17.2836 | 0.958 | D | | |
| 11 | $Rb_2NaTlBr_6$ | −3.5934 | 0.699 | D | | |
| 12 | $Cs_2CuBiCl_6$ | −25.2816 | 0.913 | I | | |
| 13 | $CsSnBr_3$ | −0.7284 | 0.629 | D | y | [72] |
| 14 | $K_2AgInCl_6$ | −22.7029 | 0.953 | D | | |
| 15 | $Cs_2AgBiI_6$ | −28.4603 | 0.900 | I | y | [62] |
| 16 | $Rb_2TlSbI_6$ | −28.2046 | 0.926 | D | | |
| 17 | $Rb_2InInBr_6$ | −9.0265 | 0.666 | I | | |
| 18 | $Rb_2InBiBr_6$ | −27.3527 | 0.753 | D | | |
| 19 | $Rb_2TlTlCl_6$ | −27.2889 | 0.710 | I | | |
| 20 | $Rb_2AgBiBr_6$ | −17.5889 | 1.199 | I | | |
| 21 | $Cs_2InBiCl_6$ | −23.8300 | 1.167 | D | n | [42] |
| 22 | $Rb_2CuSbCl_6$ | −27.8889 | 0.646 | I | | |
| 23 | $Rb_2TlSbBr_6$ | −28.6685 | 1.193 | D | | |

The performances of the two unified figures of merit are shown in Figs 9c and 10. For the exhaustive search, the optimal range ($\Delta H_d$ >−29 meV and 0.6 eV < $E_g$ < 1.2 eV) is highlighted in green in Fig. 10. All 438 × 2 calculations must be performed to identify all materials in the green area, which is an exhaustive process with lack of prior material knowledge. The screening efficiency of the exhaustive search is shown in Fig. 9c. For BO using the linear combination metric, the screening efficiency is significantly enhanced. It only takes 39 steps to screen out 20% of targeted perovskites, while it requires 94 steps for the exhaustive search. BO using the exponential combi-

 © Science China Press and Springer-Verlag GmbH Germany, part of Springer Nature 2020 

nation metric shows much better performance than the linear combination metric at the initial steps. For example, it only requires 20 (28) steps to screen 20% (40%) of targeted perovskites, which is ~2 (5) times faster than BO using the linear combination (exhaustive search). BO using the linear combination becomes efficient when the search proportion is high.

Our results suggest that the exponential combination is appropriate for the screening case when one needs to mine potential candidates as fast as possible but the completeness of the search is not important. For the linear combination, it is initially slow but becomes efficient for a complete search of possibilities. It is important to note that these two figures of merit are constructed intuitively and an algorithm for constructing an optimal figure of merit is necessary.

## CONCLUSIONS

In conclusion, we have demonstrated the efficiency of BO for screening stable perovskite and perovskites with optimal bandgaps. We further combined two material parameters, bandgap and decomposition energy, as one unified figure of merit. BO based on this figure of merit has demonstrated twice the screening efficiency of optimal materials within the appropriate range ($\Delta H_d > -29$ meV and 0.6 eV < $E_g$ < 1.2 eV) compared with an exhaustive search. BO can search for 80% of promising solar cell absorbers within 110 steps, which is twice as fast as an exhaustive search. In this work, we developed an algorithm to consider multiple descriptors and demonstrated the applicability of algorithm by using two descriptors. There is no problem to extend the algorithm to more descriptor in future when the values of related descriptors can be provided. This work demonstrates the potential of BO for screening promising solar cell absorbers, in particular for the case of a huge number of material candidates with only a handful of available data.

**Received 24 December 2019; accepted 15 January 2020; published online 16 March 2020**


1 Nørskov JK, Bligaard T, Rossmeisl J, *et al.* Towards the computational design of solid catalysts. Nat Chem, 2009, 1: 37–46
2 Greeley J. Theoretical heterogeneous catalysis: Scaling relationships and computational catalyst design. Annu Rev Chem Biomol Eng, 2016, 7: 605–635
3 Bernales V, Ortuño MA, Truhlar DG, *et al.* Computational design of functionalized metal-organic framework nodes for catalysis. ACS Cent Sci, 2018, 4: 5–19
4 Dai J, Ma L, Ju M, *et al.* In- and Ga-based inorganic double perovskites with direct bandgaps for photovoltaic applications. Phys Chem Chem Phys, 2017, 19: 21691–21695
5 Jain A, Voznyy O, Sargent EH. High-throughput screening of lead-free perovskite-like materials for optoelectronic applications. J Phys Chem C, 2017, 121: 7183–7187
6 Meng W, Wang X, Xiao Z, *et al.* Parity-forbidden transitions and their impact on the optical absorption properties of lead-free metal halide perovskites and double perovskites. J Phys Chem Lett, 2017, 8: 2999–3007
7 Zhao XG, Yang D, Sun Y, *et al.* Cu-In halide perovskite solar absorbers. J Am Chem Soc, 2017, 139: 6718–6725
8 Zhao XG, Yang JH, Fu Y, *et al.* Design of lead-free inorganic halide perovskites for solar cells *via* cation-transmutation. J Am Chem Soc, 2017, 139: 2630–2638
9 Huo Z, Wei SH, Yin WJ. High-throughput screening of chalcogenide single perovskites by first-principles calculations for photovoltaics. J Phys D-Appl Phys, 2018, 51: 474003
10 Jacobs R, Luo G, Morgan D. Materials discovery of stable and nontoxic halide perovskite materials for high-efficiency solar cells. Adv Funct Mater, 2019, 29: 1804354
11 Shockley W, Queisser HJ. Detailed balance limit of efficiency of p-n junction solar cells. J Appl Phys, 1961, 32: 510–519
12 Hall RB, Birkmire RW, Phillips JE, *et al.* Thin-film polycrystalline $Cu_2S/Cd_{1-x}Zn_xS$ solar cells of 10% efficiency. Appl Phys Lett, 1981, 38: 925–926
13 Wang LW. High chalcocite $Cu_2S$: A solid-liquid hybrid phase. Phys Rev Lett, 2012, 108: 085703
14 Bi Y, Yuan Y, Exstrom CL, *et al.* Air stable, photosensitive, phase pure iron pyrite nanocrystal thin films for photovoltaic application. Nano Lett, 2011, 11: 4953–4957
15 Hu J, Zhang Y, Law M, *et al.* Increasing the band gap of iron pyrite by alloying with oxygen. J Am Chem Soc, 2012, 134: 13216–13219
16 Bhushan M, Catalano A. Polycrystalline $Zn_3P_2$ Schottky barrier solar cells. Appl Phys Lett, 1981, 38: 39–41
17 Yin WJ, Yan Y. The electronic properties of point defects in earth-abundant photovoltaic material $Zn_3P_2$: A hybrid functional method study. J Appl Phys, 2013, 113: 013708
18 Fagen EA. Optical properties of $Zn_3P_2$. J Appl Phys, 1979, 50: 6505–6515
19 Lu S, Zhou Q, Ouyang Y, *et al.* Accelerated discovery of stable lead-free hybrid organic-inorganic perovskites *via* machine learning. Nat Commun, 2018, 9: 3405
20 Weston L, Stampfl C. Machine learning the band gap properties of kesterite I2-II-IV-V4 quaternary compounds for photovoltaics applications. Phys Rev Mater, 2018, 2: 085407
21 Pilania G, Mannodi-Kanakkithodi A, Uberuaga BP, *et al.* Machine learning bandgaps of double perovskites. Sci Rep, 2016, 6: 19375
22 Shi S, Gao J, Liu Y, *et al.* Multi-scale computation methods: Their applications in lithium-ion battery research and development. Chin Phys B, 2016, 25: 018212
23 Liu Y, Zhao T, Ju W, *et al.* Materials discovery and design using machine learning. J Materiomics, 2017, 3: 159–177
24 Liu Y, Zhao T, Yang G, *et al.* The onset temperature ($T_g$) of $As_xSe_{1-x}$ glasses transition prediction: A comparison of topological and regression analysis methods. Comput Mater Sci, 2017, 140: 315–321
25 Liu Y, Wu J, Yang G, *et al.* Predicting the onset temperature ($T_g$) of $Ge_xSe_{1-x}$ glass transition: A feature selection based two-stage support vector regression method. Sci Bull, 2019, 64: 1195–1203
26 Fukazawa T, Harashima Y, Hou Z, *et al.* Bayesian optimization of chemical composition: A comprehensive framework and its application to $RFe_{12}$-type magnet compounds. Phys Rev Mater, 2019,


 © Science China Press and Springer-Verlag GmbH Germany, part of Springer Nature 2020 

3: 053807
27 Ju S, Shiga T, Feng L, *et al.* Designing nanostructures for phonon transport *via* Bayesian optimization. Phys Rev X, 2017, 7: 021024
28 Xue D, Balachandran PV, Hogden J, *et al.* Accelerated search for materials with targeted properties by adaptive design. Nat Commun, 2016, 7: 11241
29 Lookman T, Balachandran PV, Xue D, *et al.* Statistical inference and adaptive design for materials discovery. Curr Opin Solid State Mater Sci, 2017, 21: 121–128
30 Balachandran PV, Xue D, Theiler J, *et al.* Adaptive strategies for materials design using uncertainties. Sci Rep, 2016, 6: 19660
31 Kuku T. Electrical conductivity of $CuSnI_3$, $CuPbI_3$ and $KPbI_3$. Solid State Ion, 1987, 25: 1–7
32 Smit WMA, Dirksen GJ, Stufkens DJ. Infrared and Raman spectra of the elpasolites $Cs_2NaSbCl_6$ and $Cs_2NaBiCl_6$: Evidence for a pseudo Jahn-Teller distorted ground state. J Phys Chem Solids, 1990, 51: 189–196
33 Thiele G, Richter R. Crystal structure of trirubidium hexachlorothallate, $Rb_3TlCl_6$. Z für Kristallographie-Crystline Mater, 1993, 205: 129–130
34 Thiele G, Serr BR. Crystal structure of rubidium triiodostannate (II), $RbSnI_3$. Z für Kristallographie-Crystline Mater, 1995, 210: 64
35 Kuok MH, Tan LS, Shen ZX, *et al.* A Raman study of $RbSnBr_3$. Solid State Commun, 1996, 97: 497–501
36 Yamada K, Kumano K, Okuda T. Conduction path of the sodium ion in $Na_3InCl_6$ studied by X-ray diffraction and $^{23}Na$ and $^{115}In$ NMR. Solid State Ion, 2005, 176: 823–829
37 Retuerto M, Emge T, Hadermann J, *et al.* Synthesis and properties of charge-ordered thallium halide perovskites, $CsTl^{+}_{0.5}Tl^{3+}_{0.5}X_3$ (X=F or Cl): Theoretical precursors for superconductivity? Chem Mater, 2013, 25: 4071–4079
38 Linaburg MR. Studies of Halide Perovskites $CsPbX_3$, $RbPbX_3$ (X= Cl-, Br-, I-), and Their Solid Solutions. Dissertation for Master Degree. Columbus: The Ohio State University, 2015
39 Protesescu L, Yakunin S, Bodnarchuk MI, *et al.* Nanocrystals of cesium lead halide perovskites ($CsPbX_3$, X=Cl, Br, and I): Novel optoelectronic materials showing bright emission with wide color gamut. Nano Lett, 2015, 15: 3692–3696
40 Volonakis G, Haghighirad AA, Milot RL, *et al.* $Cs_2InAgCl_6$: A new lead-free halide double perovskite with direct band gap. J Phys Chem Lett, 2017, 8: 772–778
41 Xiao Z, Du KZ, Meng W, *et al.* Chemical origin of the stability difference between copper(I)- and silver(I)-based halide double perovskites. Angew Chem, 2017, 129: 12275–12279
42 Xiao Z, Du KZ, Meng W, *et al.* Intrinsic instability of $Cs_2In(I)$-$M(III)X_6$ (M=Bi, Sb; X=Halogen) double perovskites: A combined density functional theory and experimental study. J Am Chem Soc, 2017, 139: 6054–6057
43 Zhang C, Gao L, Teo S, *et al.* Design of a novel and highly stable lead-free $Cs_2NaBiI_6$ double perovskite for photovoltaic application. Sustain Energy Fuels, 2018, 2: 2419–2428
44 Yang B, Hong F, Chen J, *et al.* Colloidal synthesis and charge-carrier dynamics of $Cs_2AgSb_{1-y}Bi_yX_6$ (X: Br, Cl; $0 \le y \le 1$) double perovskite nanocrystals. Angew Chem Int Ed, 2019, 58: 2278–2283
45 Hohenberg P, Kohn W. Inhomogeneous electron gas. Phys Rev, 1964, 136: B864–B871
46 Kohn W, Sham LJ. Self-consistent equations including exchange and correlation effects. Phys Rev, 1965, 140: A1133–A1138
47 Kresse G, Furthmüller J. Efficiency of *ab-initio* total energy calculations for metals and semiconductors using a plane-wave basis set. Comput Mater Sci, 1996, 6: 15–50
48 Perdew JP, Wang Y. Accurate and simple analytic representation of the electron-gas correlation energy. Phys Rev B, 1992, 45: 13244–13249
49 Filip MR, Liu X, Miglio A, *et al.* Phase diagrams and stability of lead-free halide double perovskites $Cs_2BB'X_6$: B=Sb and Bi, B'=Cu, Ag, and Au, and X=Cl, Br, and I. J Phys Chem C, 2017, 122: 158–170
50 Kirklin S, Meredig B, Wolverton C. High-throughput computational screening of new Li-ion battery anode materials. Adv Energy Mater, 2013, 3: 252–262
51 Saal JE, Kirklin S, Aykol M, *et al.* Materials design and discovery with high-throughput density functional theory: The open quantum materials database (OQMD). JOM, 2013, 65: 1501–1509
52 Morss LR, Siegal M, Stenger L, *et al.* Preparation of cubic chloro complex compounds of trivalent metals: $Cs_2NaMCl_6$. Inorg Chem, 1970, 9: 1771–1775
53 Knop O, Cameron TS, Adhikesavalu D, *et al.* Crystal chemistry of complex indium(III) and other M(III) halides, with a discussion of M–Cl bond lengths in complex M(III) chlorides and of the structures of and hydrogen bonding in $(NH_4)_2[InCl_5(H_2O)]$, $K_3InCl_6{\cdot}nH_2O$, $(MeNH_3)_4[InCl_6]Cl$, and $(Me_2NH_2)_4[InCl_6]Cl$. Can J Chem, 1987, 65: 1527–1556
54 Oomen EWJL, Smit WMA, Blasse G. The luminescence of $Cs_2NaSbCl_6$ and $Cs_2NaSbBr_6$: A transition from a localized to a delocalized excited state. Chem Phys Lett, 1987, 138: 23–28
55 Slavney AH, Hu T, Lindenberg AM, *et al.* A bismuth-halide double perovskite with long carrier recombination lifetime for photovoltaic applications. J Am Chem Soc, 2016, 138: 2138–2141
56 Volonakis G, Filip MR, Haghighirad AA, *et al.* Lead-free halide double perovskites *via* heterovalent substitution of noble metals. J Phys Chem Lett, 2016, 7: 1254–1259
57 Pantaler M, Cho KT, Queloz VIE, *et al.* Hysteresis-free lead-free double-perovskite solar cells by interface engineering. ACS Energy Lett, 2018, 3: 1781–1786
58 Slavney AH, Leppert L, Saldivar Val-des A, *et al.* Small-band-gap halide double perovskites. Angew Chem Int Ed, 2018, 57: 12765–12770
59 Zhou L, Xu YF, Chen BX, *et al.* Synthesis and photocatalytic application of stable lead-free $Cs_2AgBiBr_6$ perovskite nanocrystals. Small, 2018, 14: 1703762
60 Deng W, Deng ZY, He J, *et al.* Synthesis of $Cs_2AgSbCl_6$ and improved optoelectronic prop-erties of $Cs_2AgSbCl_6/TiO_2$ heterostructure driven by the interface effect for lead-free double perovskites solar cells. Appl Phys Lett, 2017, 111: 151602
61 Barbier P, Drache M, Mairesse G, *et al.* Phase transitions in a $Cs_{2-x}K_{1+x}BiCl_6$ solid solution. J Solid State Chem, 1982, 42: 130–135
62 Creutz SE, Crites EN, De Siena MC, *et al.* Colloidal nanocrystals of lead-free double-perovskite (elpasolite) semicon-ductors: Synthesis and anion exchange to access new materials. Nano Lett, 2018, 18: 1118–1123
63 Jones DR, Schonlau M, Welch WJ. Efficient global optimization of expensive black-box functions. J Glob Optim, 1998, 13: 455–492
64 Snoek J, Larochelle H, Adams RP. Practical Bayesian optimization of machine learning algorithms. Adv Neural Infor Process Syst, 2012, 2: 2951–2959
65 Ueno T, Rhone TD, Hou Z, *et al.* COMBO: An efficient Bayesian optimization library for materials science. Mater Discovery, 2016, 4: 18–21
66 Li X, Hou Z, Gao S, *et al.* Efficient optimization of the performance

 © Science China Press and Springer-Verlag GmbH Germany, part of Springer Nature 2020 

of $Mn^{2+}$-doped kesterite solar cell: Machine learning aided synthesis of high efficient $Cu_2(Mn,Zn)Sn(S,Se)_4$ solar cells. Sol RRL, 2018, 2: 1800198
67 Rasmussen CE, Williams CKI. Gaussian Processes for Machine Learning. Cambridge: MIT Press, 2006
68 Chapelle O, Li L. An empirical evaluation of thompson sampling. In: Proceedings of the 24th International Conference on Neural Information Processing Systems. 2011. 2249–2257
69 Pelatt BD, Kokenyesi RS, Ravichandran R, *et al.* Atomic solid state energy scale: Universality and periodic trends in oxidation state. J Solid State Chem, 2015, 231: 138–144
70 Sun Q, Yin WJ. Thermodynamic stability trend of cubic perovskites. J Am Chem Soc, 2017, 139: 14905–14908
71 Sun S, Hartono NTP, Ren ZD, *et al.* Accelerated development of perovskite-inspired materials *via* high-throughput synthesis and machine-learning diagnosis. Joule, 2019, 3: 1437–1451
72 Peedikakkandy L, Bhargava P. Composition dependent optical, structural and photoluminescence characteristics of cesium tin halide perovskites. RSC Adv, 2016, 6: 19857–19860

**Acknowledgements** Yin WJ acknowledges funding support from the National Key Research and Development Program of China (2016YFB0700700), the National Natural Science Foundation of China (11974257, 11674237 and 51602211), the Natural Science Foundation of Jiangsu Province of China (BK20160299), and the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD). The theoretical work was carried out at the National Supercomputer Center in Tianjin and the calculations were performed on TianHe-1(A).

**Author contributions** Yin WJ supervised the project. Chen X, Wang C, and Li Z performed the calculations. Hou Z provided guidance for algorithm. Chen X and Yin WJ wrote the paper. All authors attended on discussion.

**Conflict of interest** The authors declare that they have no conflict of interest.

**Supplementary information** Supporting data are available in the online version.



**Xiwen Chen** is a master student at the College of Energy, Soochow Institute for Energy and Materials InnovationS (SIEMIS), Soochow University. Her interest focuses on the search and design of perovskite photovoltaic materials.



**Wan-Jian Yin** is a professor in SIEMIS, Soochow University, China. He received his BS degree (2004) and PhD degree (2009) from Fudan University, China. He worked in the National Renewable Energy Laboratory (NREL) and University of Toledo, USA from 2009 to 2015. His research interests include computational study of solar energy materials, defect physics in semiconductors and machine-learning on materials design.

## 基于统一品质因子的贝叶斯优化方法加速材料筛选: 以卤化物钙钛矿为例

陈曦雯[1], 王陈[1], 李珍珠[1], 侯柱峰[2], 尹万健[1,3*]

**摘要** 品质因子在材料筛选和新材料设计中至关重要. 比如, 根据Shockley-Queisser极限, 虽然带隙是筛选光伏材料的重要品质因子, 但将带隙作为唯一品质因子却又过于简单, 因为理想的太阳能电池吸收材料需要满足多种条件, 如高稳定性、高光吸收系数和长载流子寿命等. 然而, 同时考虑多种品质因子大大增加了材料筛选的难度. 本工作以卤化物钙钛矿为例, 将热力学稳定性和带隙结合成一个统一的品质因子, 采用贝叶斯优化方法来加速材料筛选. 我们发现, 与传统的穷举法相比, 基于统一品质因子的贝叶斯方法可以更加有效地筛选出目标材料. 该工作为从大量材料中搜索出理想的目标材料提供了一个计算成本较低的方法.

 © Science China Press and Springer-Verlag GmbH Germany, part of Springer Nature 2020 