Machine-learning accelerated annealing with fitting-search style for multicomponent alloy structure predictions

Chuannan Li [1,2] Hanpu Liang [3,4] Yifeng Duan [3,*] and Zijing Lin [1,2,†]

[1]*Department of Physics, University of Science and Technology of China, Hefei 230026, China*
[2]*Hefei National Laboratory, University of Science and Technology of China, Hefei 230088, China*
[3]*School of Materials and Physics, China University of Mining and Technology, Xuzhou, Jiangsu 221116, China*
[4]*Beijing Computational Science Research Center, Beijing 100193, China*

Structural prediction for the discovery of novel materials is a long sought-after goal of computational physics and materials sciences. The success is rather limited for methods such as the simulated annealing method (SA) that require expensive density functional theory (DFT) calculations and follow unintelligent search paths. Here a machine-learning based crystal combinatorial optimization program (CCOP) with a fitting-search style is proposed to drastically improve the efficiency of structural search in SA. CCOP uses a graph neural network energy prediction model to reduce the DFT cost and a deep reinforcement learning algorithm to direct the search path. Tests on six multicomponent alloys show the energy prediction model is capable of extracting the bonding characteristics of the complex alloys to achieve interpretability. It also achieves high accuracy with a tiny training set (an increment of 30 samples per iteration) by active learning in less than five iterations. Comparison with a few conventional methods shows that CCOP finds the lowest-energy structures with the smallest number of search steps. CCOP cuts the computing cost of SA by two orders of magnitude, while providing better search results than SA. CCOP is promising for serving as a broadly applicable tool for the efficient crystal structure predictions.

I. INTRODUCTION

Finding materials with the desired properties is an important task in the condensed matter physics and material sciences, and guides the direction in the structural search field [1,2]. Alloying is commonly used for obtaining the desired materials, as the structural, electronic, transport and optical properties of alloys can be tuned by varying the compositions [3–5]. Unfortunately, the complexity of alloy structural search grows exponentially with the number of atoms in a unit cell. Thus, reducing the computing costs while maintaining the reliability of the search results is of high importance for a structural search method.

To tackle the structural search problem, many algorithms have been developed, such as simulated annealing (SA) [6–8], *ab initio* random structure search [9], and the evolutionary algorithms, e.g., genetic algorithm (GA) [10–12] and particle swarm optimization [13,14]. The methods have shown success to various degrees, but there is still large room for improvement. For example, the GA search is extremely dependent on the quantity and diversity of chosen populations. SA and *ab initio* random structure search, however, require many thousands of iterations to reach the low-lying minima in the energy landscape and are expensive due to the CPU demanding density functional theory (DFT) [15,16] calculations. Fortunately,

SA can be easily parallelized [17], making it a good candidate for large-scale deployment to alleviate the bottleneck of DFT.

Machine learning (ML) is becoming increasingly popular in accelerating the discovery of new materials by encoding physical knowledge into property models [18,19]. For instance, the deep tensor neural network [20] unifies many-body Hamiltonians to design neural network. The crystal graph convolutional neural network (CGCNN) [21] considers the topology of crystal to build graph, providing a universal and interpretable representation of materials. These methods minimize the need of DFT calculations and have shown high performance in property predictions via the combination of ML and physical concepts. Nevertheless, training the ML models to gain the desired generalization capability still requires a large amount of labeled data that often means a large amount of expensive DFT computations. Therefore, reducing the number of data required by training the property prediction model (PPM) is a key issue to be resolved for the improved efficiency.

ML may also be used to design the strategy of exploring the potential energy surface (PES) to cut the computational cost of a structural search method. As is known, the reward-driven reinforcement learning (RL) focuses on the best policy of exploration in an interactive environment. RL has shown success in various fields. For instance, AlphaGo [22] showed its strong ability in combinatorial optimization to maximize the gain and defeated the world championship in Go game. In the fields of physics, chemistry, and biology, RL has been used to design nanophotonic devices [23] and drug molecules [24–27] by

---

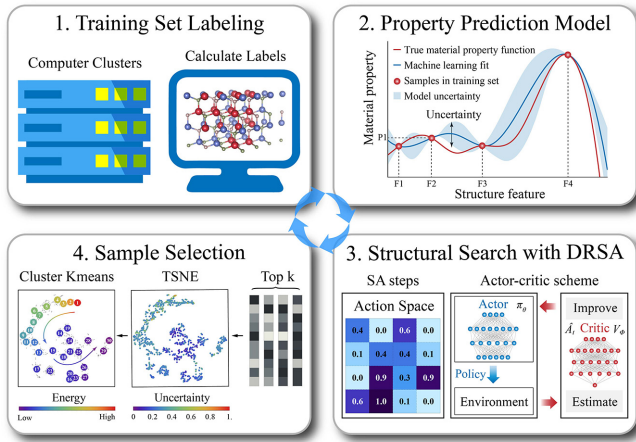*yifeng@cumt.edu.cn
†zjlin@ustc.edu.cn

FIG. 1. Workflow of the crystal combinatorial optimization program (CCOP). The program starts with a few initial structures as the samples. The samples are labeled with DFT energies and used as the training set (step 1). The labeled training set is used to train a machine-learning potential that fits the PES (step 2). It is followed by training an RL agent for the efficient sampling of the PES to find the low-energy structures (step 3). Then clustering analysis is performed on the sampled structures to select samples for the training set (step 4). The entire optimization program runs in a closed loop.

learning the policy to optimize the objective function. Naturally, RL is expected to be helpful for the structural search methods by learning and making decision on the favorable descent path on PES.

In this paper, we proposed a crystal combinatorial optimization program (CCOP), which uses a weighted crystal graph convolutional neural network (WCGCNN) as the PPM, and employs a deep reinforcement simulated annealing (DRSA) technique as the searching algorithm. The active learning [28–30] is applied with the selection of highly representative samples so that the training of PPM requires only a small data set. The DRSA is used to guide the search agent to further reduce the computational cost. The numerical efficiency of CCOP is illustrated by its applications to the search of the ordered structures of six testing multicomponent alloys: BN, $BeZnO_2$, AlNSiC, GaNZnO, $BeMgCaO_3$, and $Al_2Ga_2Zn_2N_3O_3$. The test also reveals that the PPM is interpretable, accurate and fast to compute. Meanwhile, the DRSA is shown to have the highest performance among the tested search methods, including the conventional RL algorithm of proximal policy optimization (PPO2) [31], SA, and the random search.

II. METHODS

The workflow of CCOP proposed here is illustrated in Fig. 1. It consists of four major parts: (i) training set labeling, (ii) property prediction model, (iii) structural search, and (iv) sample selection. Each part is explained as follows.

Training set labeling. To ensure the accuracy of the PPM in fitting the PES, the training set of structural samples is labeled with the single-point DFT energies that are calculated with the Vienna *Ab initio* Simulation Package (VASP) [32–34]. The generalized gradient approximation (GGA) with

Perdew-Burke-Ernzerhof (PBE) exchange and correlation functional [35] is used, and the ion-electron interactions are treated by projector-augmented-wave (PAW) [36,37] technique. The initial training structures are generated by sampling the PES with 30 SA steps. Thirty more structures selected by active learning are added to the training set after every fitting-search iteration of CCOP.

Property prediction model. Training a PPM to replace the expensive DFT calculations (step 2 in Fig. 1) is the second part. The PPM is built based on CGCNN, and modified under the architecture of message passing neural network [38]. At the kth message passing phase, atom vectors h_i^k, h_i^k and their bond feature vector e_{ij} form the message by function M_k. The aggregated message $m_i^{k+1} = \sum_{j \in N(i)} M_k(h_i^k, h_j^k, e_{ij})$ is the sum of all $N(i)$ neighbors of atom vector h_i^k in crystal graph $\mathcal{G}$. New representation of atom vector $h_i^{k+1} = U_k(h_i^k, m_i^{k+1})$ is obtained by updating function U_k. After K times of message passing, property can be predicted by $\hat{y} = R(\{h_i^K | i \in \mathcal{G}\})$, where R is a differentiable function. In order to focus on the property related messages in the model, we assign each message with a weight ω_j^k and modify M_k of CGCNN as

$$M_k = w_j^k \cdot \sigma\left(x_{ij}^k W_f^k + b_f^k\right) \odot g\left(x_{ij}^k W_s^k + b_s^k\right), \quad (1)$$

where $\odot$ denotes elementwise multiplication, W_f^k, W_s^k and b_f^k, b_s^k are weight matrices and bias vectors of the kth layer, respectively, and σ is a sigmoid function, g is a softplus function [39]. $x_{ij}^k = h_i^k \oplus h_j^k \oplus e_{ij}$ concatenates neighboring atom pair with their bond [40]. The 12 nearest-neighboring atoms are found by using a cutoff distance of 8 Å [41]. The message weights, ω_j^k, for the 12 neighboring atoms are initialized with the same value, since it is hard to tell which message is more important at first. This permits the PPM to reasonably allocate the relative strengths of the interactions between atoms. Moreover, we perform a gated structure [42] to control the update process,

$$U_k = z_i^k \odot h_i^k + (I - z_i^k) \odot m_i^{k+1}, \quad (2)$$

$$z_i^k = \sigma\left[W_u^k \cdot (h_i^k \oplus m_i^{k+1})\right], \quad (3)$$

where I is an all-ones vector, W_u^k is the weight matrix, σ is applied to scale each dimension of z_i^k in [0, 1], and weight vector z_i^k is used to determine the update ratio of h_i^k. The gated structure has shown a good performance in retaining and filtering information [43]. When the message passing process finish, atom i is embedded into its chemical environment by iteratively including the surroundings, thus h_i^K can be treated as the representation of atom i in the structure. As for the representation of crystal structure, we sum up $\mathcal{N}$ atom vectors and average them as the crystal vector $c = \sum_i h_i^K / \mathcal{N}$, which contains machine-learned structural features. A three-layer fully connected network [44] is set as the differentiable function for the property prediction,

$$\hat{y} = W_3(g(W_2(g(W_1 c + b_1)) + b_2)) + b_3, \quad (4)$$

where Ws are the weights, and bs are the biases. Eq. (4) is a universal approximator for nonlinear functions [45]. More details about the WCGCNN based PPM are given in the Sec. I of the Supplemental Material (SM) [46]. The PPM is

trained with the MXNet [47] framework using the Adam optimizer [48] for gradient descent optimization, using the mean square error (MSE) as the loss function (step 2 in Fig. 1). The training uses 150 samples obtained by SA and DFT calculations as the validation set. Since the size of training set is small and there are more parameters in WCGCNN than CGCNN, it is hard to determine the suitable model in the huge parameter space. Therefore, the weights of the nearest atoms and update ratios are frozen at the beginning of training (60 epochs), namely training the network in a smaller parameter space. The model with the lowest validation loss is then chosen, and the weights and update ratio are unfrozen to fine tune the model (60 epochs). The utilization of a two-step training technique can guarantee the reliability of the PPM during the search process [49]. Further information on the effects of network freezing can be found in Sec. II of SM [46].

Structural search with DRSA. Training an RL agent and use it to direct the SA path for the efficient determination of the lowest-energy structures (step 3 in Fig. 1) is the third part. The positions of atoms are encoded as a one-dimensional list, and a sequence of actions are applied to minimize the structural energy. An action is defined here as the exchange between atoms with the same sign of electricity, and the exchange between the same atoms is forbidden. The RL agent assigns weights to the allowed actions, and the masking of 0 is used to implement action suppression, thereby avoiding unreasonable structures.

At the tth SA step, the DRSA agent performs action a_t by the ϵ-greedy policy $\pi_\theta(a_t|s_t)$ to adjust structure s_t under the Metropolis criterion [6]. The value of energy descent $r_{t+1} = E_0 - \hat{E}_{t+1}$ is defined as the reward, where E_0 is the energy of the search starting structure calculated by DFT and $\hat{E}_{t+1}$ is the energy of sample predicted by PPM. The discounted sum of rewards is defined as $G(\tau) = \sum_{t=0}^{T-1} \gamma^t r_{t+1}$, where $\tau = \{s_0, a_0, s_1, r_1, a_1, \cdots, s_T, r_T\}$ is a trajectory of Markov decision process and γ is the discount factor determining the priority of short-term rewards. Minimizing the structural energy within T steps is equivalent to maximizing the expectation of G, thus the agent can learn the policy of energy descent.

The agent's policy $\pi_\theta(a_t|s_t)$ is a key for finding the optimal structure. The agent is trained by the clipped loss of PPO2 [31],

$$\mathcal{L}_A = -\hat{\mathbb{E}}_{\tau,t}[\min(p_t(\theta)\hat{A}_t, \text{clip}(p_t(\theta), 1-\epsilon, 1+\epsilon)\hat{A}_t)],$$
(5)

where $p_t(\theta) = \pi_\theta(a_t|s_t)/\pi_{\theta_{old}}(a_t|s_t)$ is the probability ratio of current and old policies, $\hat{A}_t$ is the estimator of advantage [50]. The clip function in Eq. (5) removes the incentive of p_t beyond the interval $[1-\epsilon, 1+\epsilon]$, and means a penalty for a large policy update. To reduce the variance of $\hat{A}_t$, the TD(0) [51] form of $\hat{A}_t$ is adopted, i.e., $\hat{A}_t = r_{t+1} + \gamma V_\pi(s_{t+1}) - V_\pi(s_t)$, where the state-value function $V_\pi(s_t) = \mathbb{E}_\tau[r_t|s_t]$ is the expected return from state s_t. $V_\pi(s_t)$ can be approximated by minimizing the loss [52]

$$\mathcal{L}_C = \hat{\mathbb{E}}_{\tau,t}[r_{t+1} + \gamma V_\pi(s_{t+1}) - V_\pi(s_t)]^2.$$
(6)

Through learning from the search paths, the agent's policy $\pi_\theta(a_t|s_t)$ generates a suitable weight for each action to

minimize energy [53], thus improving the search efficiency of SA actions. More information about the DRSA can be found in the Sec. III of SM [46].

Sample selection. Choosing the representative samples to add to the training set (step 4 in Fig. 1) is the last part of the workflow. To be most beneficial for reducing the mean absolute error (MAE) of PPM predictions, samples with the highest uncertainties should be considered. The uncertainty Ω is defined as the variance of predictions [28]

$$\Omega(\boldsymbol{x}) = \frac{1}{M-1} \sum_{m=1}^{M} (f_m(\boldsymbol{x}) - \frac{1}{M} \sum_{l=1}^{M} f_l(\boldsymbol{x}))^2,$$
(7)

where $\boldsymbol{x}$ is a searched sample, f_m denotes a trained PPM, and M is the number of PPMs. Specifically, 10% of the samples with the highest uncertainties and 10% of the samples with the lowest energies are selected. The crystal vectors of these samples are calculated by the PPM, followed by t-distributed stochastic neighbor embedding (TSNE) [54] to reduce the dimension of crystal vectors. The reduced vectors are grouped into 30 clusters by the Kmeans method [55]. The minimal energy sample in each cluster with its DFT energy computed is added into the training set. Clustering can contribute to enhancing the diversity of structures during the search process as detailed in Sec. IV of SM [46]. Moreover, the lowest-energy sample in the training set, with the energy referred as E_0 above, is used as the initial structure of the next fitting-search iteration.

III. RESULTS AND DISCUSSION

In this section, we first showcase the performance of the CCOP by applying it to three compositions with four distinct lattices. Next, we select the wurtzite lattice to compare the search ability of DRSA with SA, PPO2, and random search. Then, the following three parts delve deeper into the reasons for the success of CCOP: (i) role of RL and SA in DRSA, (ii) role of active learning in CCOP, and (iii) interpretability of PPM.

A. Performance on various lattices

To demonstrate the capability of CCOP in crystal structure prediction, we select three multicomponent alloys, BN, BeZnO$_2$, and GaNZnO, and conduct tests on four different lattices: perovskite, rocksalt, wurtzite, and zincblende. The grid points are fixed for each lattice, with 40 grid points in perovskite, 64 in rocksalt and zincblende, and 72 in wurtzite. The structural search is performed using the P1 space group symmetry, which means each atom is independently. The CCOP is initiated by DFT-SA, and each fitting-search iteration involves 30 samples selected by active learning. As illustrated in Fig. 2, CCOP effectively reduced the energy and identified low-energy ordered structures for almost all compositions and lattices.

More specifically, after searching, the BN composition exhibits configurations of alternating B and N atoms in the rocksalt, wurtzite, and zincblende lattices. This is due to the strong covalent bonds formed by the pairing of valence electrons between B and N, resulting in stable structures. Additionally, due to the composition not satisfying the perovskite
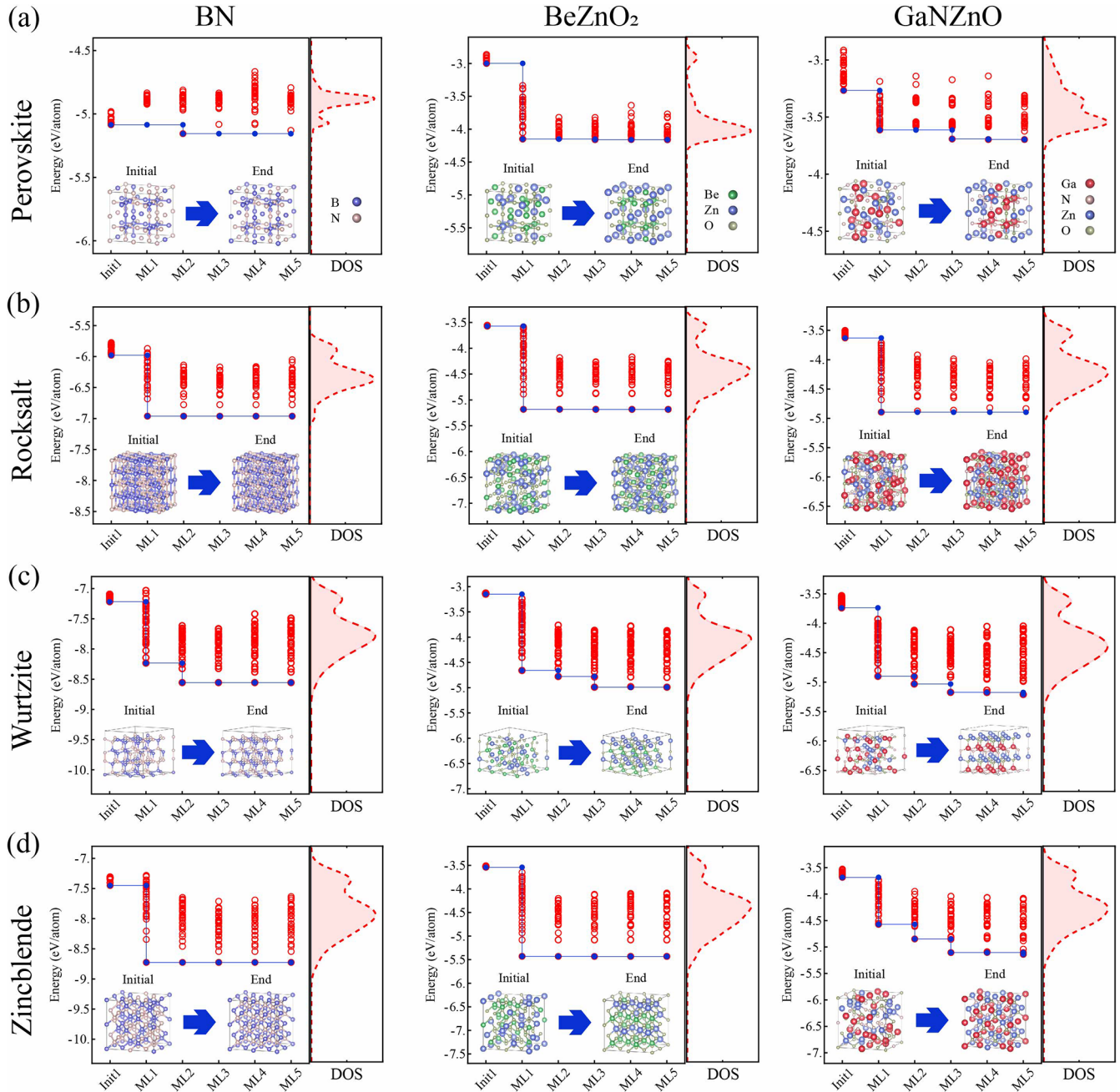
FIG. 2. Performance of CCOP on four different lattices (a) perovskite, (b) rocksalt, (c) wurtzite, and (d) zincblende with compositions of BN, $BeZnO_2$, and GaNZnO. The initial and final lowest-energy structures for each composition and lattice are displayed in figures, as well as the corresponding density of states (DOS) plots during the structural search process.

ABX_3 requirement, where A and B sites must be nonradioactive metals and X site is one or several elements from O, N, S, and F. As a result, it is challenging to find the lowest-energy ordered configuration in the perovskite lattice for the selected three compositions. For wurtzite lattice, the multicomponent alloys $BeZnO_2$ and GaNZnO show a noticeable transition from disordered to ordered layered structures, which is attributed to the their binary counterparts that can form stable wurtzite configurations. In other words, the electrons pairing between Be-O, Zn-O, and Ga-N reduces the energy and leads to the formation of the layered structures in $BeZnO_2$ and GaNZnO.

B. Comparison of search algorithms

The effectiveness of the search algorithm DRSA is tested by searching the lowest-energy structure of the ordered configurations of multicomponent alloy. Six alloys with compositions from simple to complex are used as the testing cases: BN, $BeZnO_2$, AlNSiC, GaNZnO, $BeMgCaO_3$, and $Al_2Ga_2Zn_2N_3O_3$. The search is restricted to that the alloys maintain the 72-atom wurtzite-like lattice configuration in any temperature environment [56–58]. The unit cell consists of eight lattice layers and there are nine atomic sites in each layer. Initially each layer is filled with nine cations or anions,
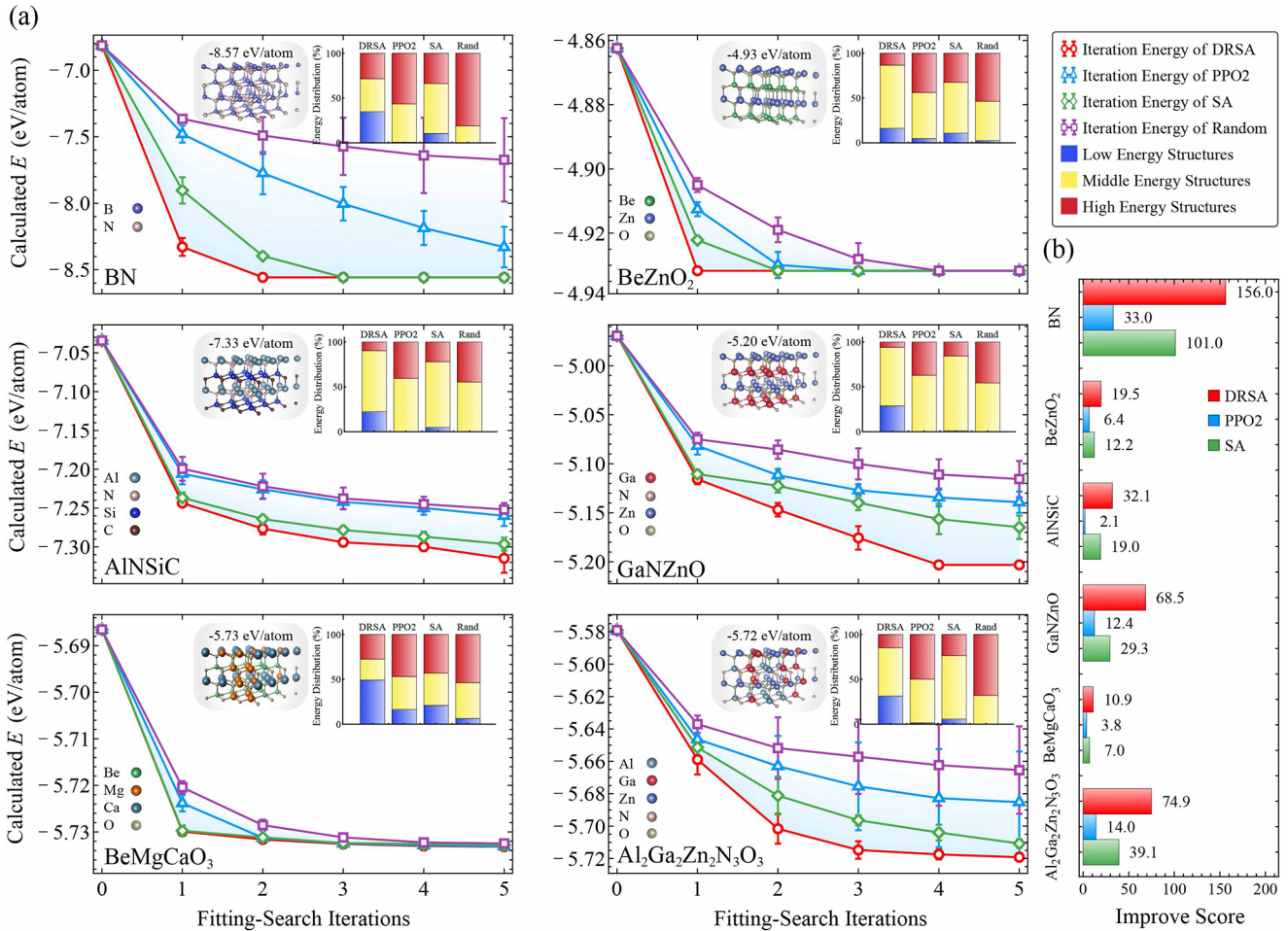
FIG. 3. (a) The calculated energy vs the fitting-search iteration for the six multicomponent alloys. The inner panels are the lowest-energy structures and the energy distribution percentage of the search results. (b) The improvement scores of DRSA, PPO2, and SA for the six multicomponent alloys.

and the cation-layer and the anion-layer alternate. The atomic arrangements are then changed in the search process to obtain lower-energy configurations. As there is only one type of anions (cations) in BN, the search action in BN refers to the exchange of the positions of anion and cation, instead of exchanging among cations or among anions for the structural searches of the other five alloys. These constraints on atom exchange effectively decrease the occurrence of implausible structures during the structural search.

Figure 3(a) displays the lowest energies and structural energy distributions of the six alloys versus the searching iterations for the DRSA, PPO2, SA, and random search methods. The performance of different methods may be measured by the improvement score, which is defined as $s = p \cdot \frac{1}{5} \sum_{i=1}^{5} |\Delta E_i| / |\Delta E_i^R| \cdot 100\%$. Here p is the weighted average concerning the structural energy distribution, with the weight ratios of 6:3:1 for the low-, middle-, and high-energy structures. $|\Delta E_i| = |E_i - E_0|$ is the energy difference between the lowest-energy structure of the ith iteration and the initial structure, and we use the random search $|\Delta E_i^R|$ as the benchmark. The improvement scores of DRSA, PPO2, and SA for the six alloys are shown in Fig. 3(b). On average, the improvement scores of DRSA, PPO2, and SA for the six

alloys are 60.37%, 11.93%, and 34.62%, respectively. DRSA has the highest score due to its combination of RL and SA, while PPO2 has the lowest performance due to its inefficiency in a large action space. More information about the performances of the 4 search methods can be found in Table S2 of SM [46]. As seen in Table S2, DRSA usually finds the lowest-energy structure with the smallest number of iterations. Moreover, all the six lowest-energy structures, one for each of the six alloys, are found by DRSA. However, four of them are missed by PPO2 and SA, while five of them are missed by the random search. Clearly, the newly proposed method of CCOP that combines the WCGCNN, SA, and the RL is capable of drastically reducing the computational cost, while maintaining the desirable accuracy. Compared to SA that is more efficient than PPO2 and the random search (Table S2), CCOP may reduce the computational cost from $7 \sim 9$ days for SA to $1 \sim 2$ hours (Section V of SM [46]), i.e., a reduction of two orders of magnitude.

C. Role of RL and SA in DRSA

To analyze the function of RL and SA in DRSA, we compare the accumulative rewards (ARs) and MAE of PPM
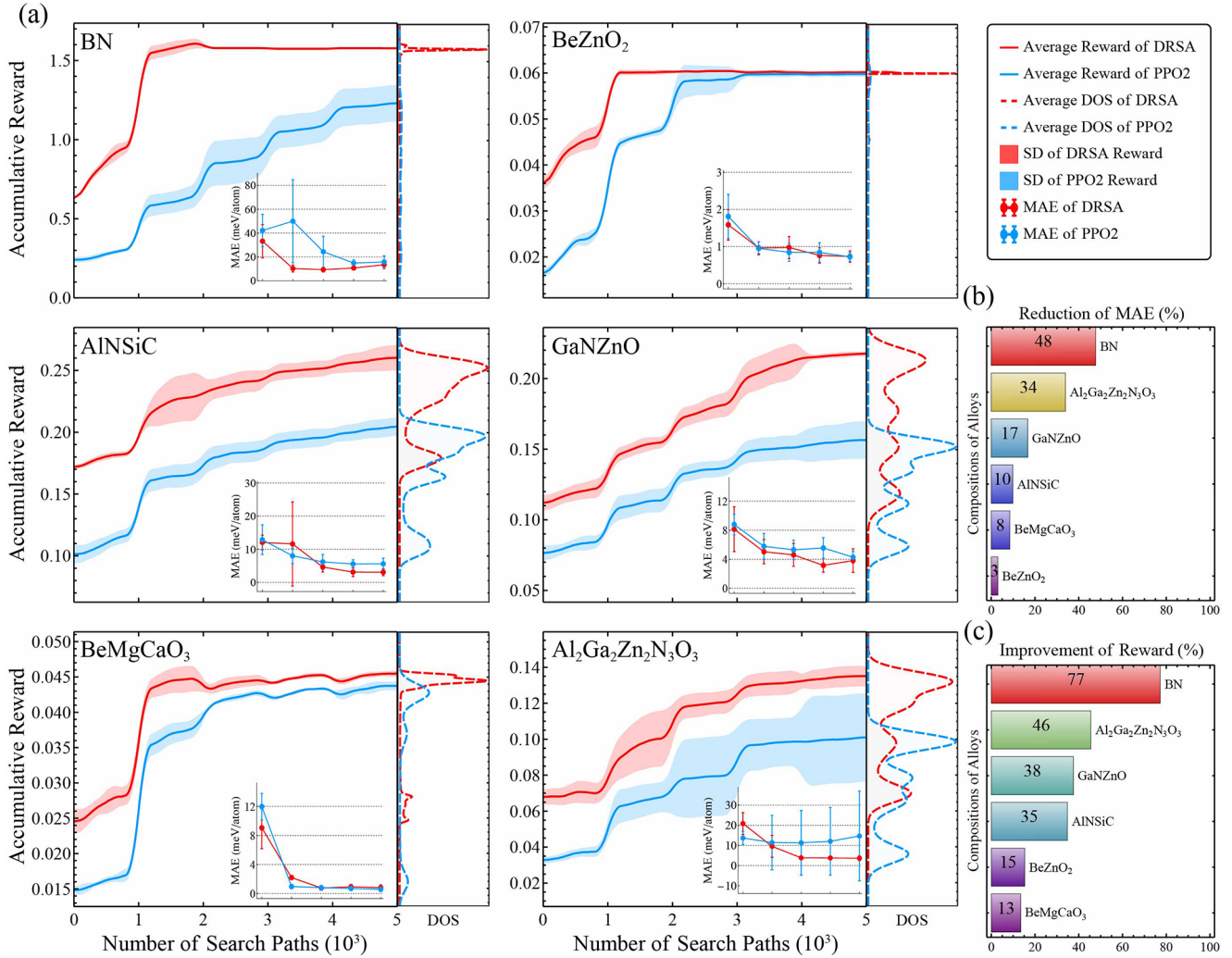
FIG. 4. Performances of DRSA and PPO2 on searching the low-energy structures of six multicomponent alloys. (a) The accumulative reward (AR, in eV) and the density of state (DOS) of DRSA and PPO2 vs the search paths. The shadow of the line is the standard deviation (SD) of the ARs. The inner panel shows the change of MAE with the fitting-search iteration. (b) Average reduction of MAE by DRSA relative to PPO2, and (c) Improvement on the AR by DRSA relative to PPO2.

during the structural search process. Figure 4 compares the performances of DRSA and PPO2 on searching the low-energy structures of six multicomponent alloys, averaged over five separate DRSA or PPO2 runs for every alloy. Each of the five DRSA or PPO2 runs consists of five fitting-search iterations and 1000 search paths per iteration, for a total of 5000 search paths per run. Figure 4(a) shows the accumulative rewards of DRSA and PPO2 for searching the low-energy structures of six alloys. AR is calculated as the total energy descent $E_0 - \hat{E}_n$, where $\hat{E}_n$ is the minimum predicted energy at the nth search path. As seen in Fig. 4(a), AR of DRSA is higher than that of PPO2 for every alloy examined. The results are understandable as DRSA is basically PPO2, but with the policy complexity reduced by the physical constraint of SA. With the help of the PPM agent that learns the weight of atomic exchange, which ensures the choice of atom exchanges is not as random as the SA, DRSA requires no more than 5000 search paths in reaching its maximum reward for most alloys. The fast convergence is achieved also because that the PPO2 agent learns the energy descent policy from the paths

of DRSA and the fitting-search style relieves the difficulty of training the agent, resulting in a reduced number of steps to reach the maximal reward.

DRSA uses fewer samples and performs better than PPO2. The improvement may be measured with the improvement of sample efficiency (ISE) defined as ISE = $N_{PPO2}/N_{DRSA} \times$ 100%, where N_{PPO2} and N_{DRSA} are the number of samples searched by PPO2 and DRSA, respectively. ISE, together with the number of executable actions (allowed exchanges of atom positions) and the number of possible structures for each of the six alloys are shown in Table I. The number of possible structures, or the possible permutations of the atom sites, is found to be from 10^9 to 10^{25} for the six alloys. As the number of possible structures is very large, certainly vastly larger than N_{PPO2} and N_{DRSA}, employing search methods such as DRSA and PPO2 is necessary in practice. Meanwhile, DRSA is superior with a smaller number of samples and a larger AR than PPO2.

Among the six alloys tested, the binary alloy BN shows the highest ISE of about 1370%. This may be attributed to

TABLE I. Number of executable actions (actions), number of possible structures (structures), and the improvement of sample efficiency (ISE) by DRSA relative to PPO2 for six multicomponent alloys.

Type of alloys	Actions	Structures	ISE (%)
BN	1296	10^{20}	1370
$BeZnO_2$	324	10^{9}	198
AlNSiC	648	10^{19}	349
GaNZnO	648	10^{19}	320
$BeMgCaO_3$	432	10^{15}	135
$Al_2Ga_2Zn_2N_3O_3$	756	10^{25}	192

that BN has the largest number of executable actions, 1296, at each search step (see Table I). In fact, without the constraint of SA, the random search at the beginning of PPO2 causes an inefficient exploration of the low-energy area. PPO2 also shows a slow convergence of the policy in a large action space. The lowest-energy structure is often missing even with five different runs of PPO2 (see Table S2 of SM for more testing results [46]). When the number of actions is relatively small, however, the difficulty of policy learning for the RL agent is much reduced and the performance of PPO2 is much improved. As a result, PPO2 may catch up DRSA on the accumulated reward, as shown in Fig. 4 for $BeZnO_2$ and $BeMgCaO_3$ with the number of actions of 324 and 432, respectively. The MAE of the predicted energies in each fitting-search iteration reflects the variation of the accuracy of PPM. As seen in the insets of Fig. 4(a), the MAE for either DRSA or PPO2 normally decreases at first due to the increase in the training data, and then converges because of the lack of diversity in the newly added representative samples. Overall, however, DRSA not only shows higher ARs, but also lower MAEs than PPO2 for the six tested alloys. To be more quantitative, Figs. 4(b) and 4(c) show respectively the reduction of MAE and the improvement of reward by DRSA relative to PPO2 for the tested alloys. Here, the reduction of MAE is calculated as $\sum_{i=1}^{5} = (MAE_i^{PPO2} - MAE_i^{DRSA})/MAE_i^{PPO2} \times 100\%$, where $MAE_i^{DRSA}(MAE_i^{PPO2})$ is the MAE of DRSA (PPO2) at the ith iteration. The improvement of reward is calculated as $\sum_{i=1}^{5000}(AR_i^{DRSA} - AR_i^{PPO2})/AR_i^{PPO2} * 100\%$, where $AR_i^{DRSA}(AR_i^{PPO2})$ is the AR of DRSA (PPO2) at the ith search path.

As can be seen in Figs. 4(b) and 4(c), there is a positive correlation between the magnitudes of the reduction of MAE and the improvement of reward, both are the largest for BN and the second largest for $Al_2Ga_2Zn_2N_3O_3$. While BN has the largest number of executable actions, $Al_2Ga_2Zn_2N_3O_3$ has the largest number of possible structures and the second largest number of executable actions (see Table I). Combined with a complicated chemical composition, $Al_2Ga_2Zn_2N_3O_3$ is hard for PPO2 to handle. Not only the SD of AR increases with the search, the SD of MAE (see the error bar of MAE) also increases, from 3.3 meV/atom for the first iteration to 22 meV/atom for the fifth iteration. Hence, PPO2 is unstable when applied to $Al_2Ga_2Zn_2N_3O_3$. Detailed data analysis shows that, due to a lack of Metropolis criterion to constrain the search direction, a significant number of

high-energy structures are produced by PPO2, causing difficulty in training PPM. It then leads to high MAE of PPM and improper exploration of PES. On the contrary, a positive feedback between fitting and search is formed in the constraint search of DRSA, leading to a continuous reduction of MAE and a stable search.

D. Role of Active Learning in CCOP

To illustrate the benefit of the active learning for the sample selection, the results of the clustering analysis at the last fitting-search iteration are displayed in Fig. 5. As shown in Fig. 5, the structural samples are grouped by Kmeans into 30 clusters, with the minimal sample energy decreases from cluster 1 to cluster 30 and the energies are similar for samples in the same cluster. The uncertainty of the PPM predictions [Eq. (7)] decreases with the decreased sample/cluster energy, indicating that the PPM is adaptive during the search, especially to the low-energy area of the PES. The correlation between the sample/cluster energies and uncertainties may be quantified with the Pearson correlation coefficient. As shown in Fig. S5 of the SM [46], the Pearson correlation coefficients are found to be higher than 0.9 for all six alloys, with the highest of 0.951 for $Al_2Ga_2Zn_2N_3O_3$, further illustrating the adaptability of PPM.

The proximity of the energies for structures in the same cluster and the strong correlation between the energies and uncertainties are the results of the feature classification of the PPM. Therefore, adding only the lowest-energy structure of each cluster to the training set is quite reasonable and has the benefit of minimizing the size of the training set. Meanwhile, adding the the high-energy structures with high PPM prediction uncertainties, as mentioned in the sample selection strategy above, can help the PPM to fit the entire PES better. Consequently, the active learning in the sample selection has a positive effect on the structural searching. It improves the accuracy of PPM by putting the most representative samples on the energy landscape in the training set. It also reduces the times of expensive DFT calculations [59,60] effectively by reducing the size of the training set. The DRSA predictions are also coherent with the physical intuition. The lowest-energy structures of AlNSiC, $BeZnO_2$, BN, and GaNZnO are predicted to be ordered structures with the obvious layered characteristics. Specifically, GaNZnO consists of the Ga-N and Zn-O layers placed alternatively along the z axis direction (see the insets in Fig. 5), providing the best match of the valence electrons for the stabilization of the sp^3 hybridization. For the quaternary alloy $BeMgCaO_3$, the Be and Ca atoms are placed in different layers due to the large difference in their atomic sizes, while the Mg atoms are evenly distributed in the Be- and Ca layer. Similar structural feature is observed in the lowest-energy structure of the quintary alloy $Al_2Ga_2Zn_2N_3O_3$.

E. Interpretability of PPM

Interpretability of ML model, e.g., the characteristic vectors extracted from the crystal information are consistent with the physical intuition, is desired in the fields of physics, chemistry, and material sciences. Here, AlNSiC is used as
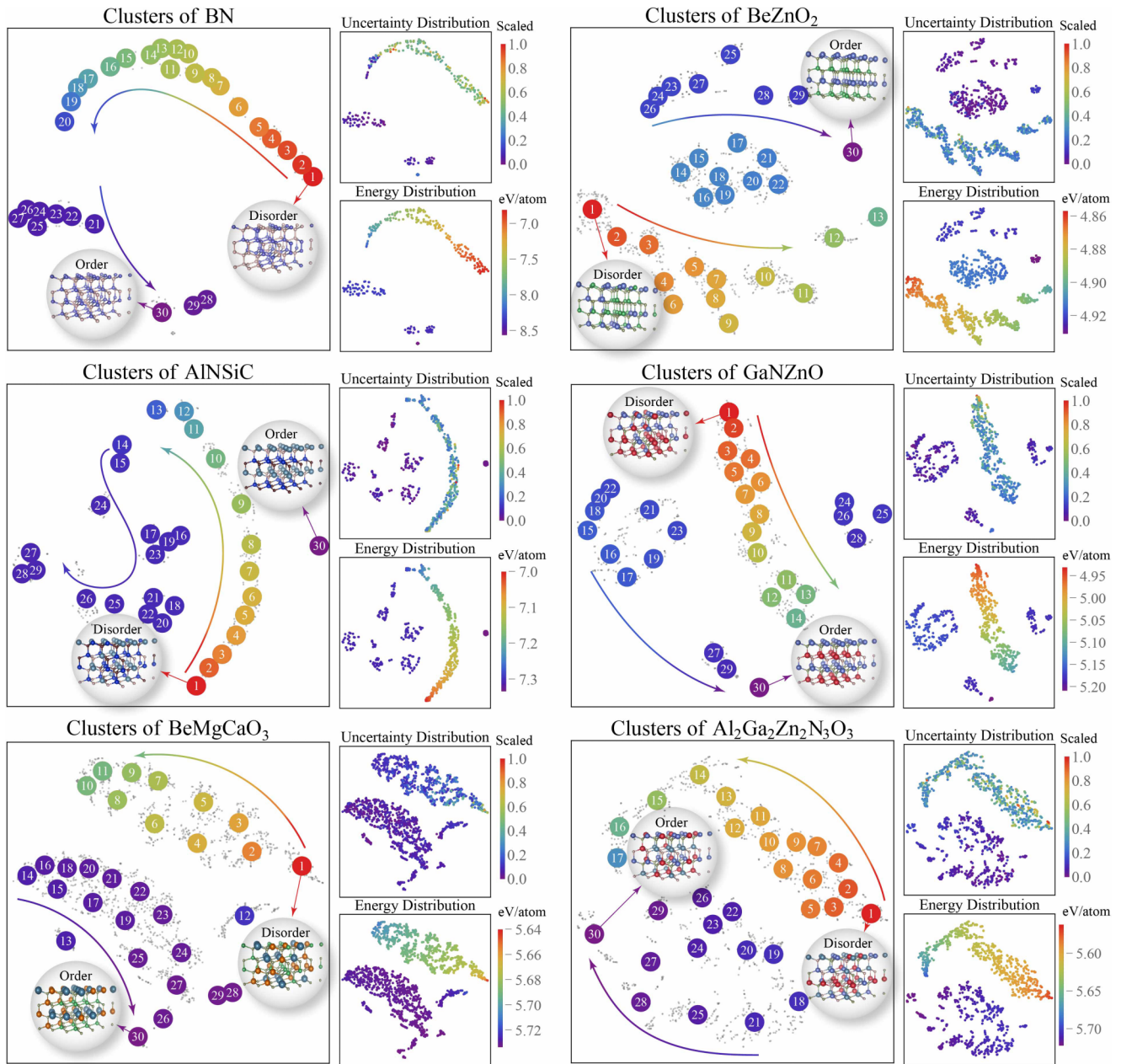
FIG. 5. The distribution of the clusters, the uncertainties, and the energies of selected samples for the six multicomponent alloys at the last fitting-search iteration. The *x* and *y* axis are the feature vectors generated by TSNE. Each cluster is labeled and colored by the rank of energy. The arrow curves show the direction of energy descent.

an example to show the interpretability of our PPM. We label each atom vector from 1 to 72 and each layer from 1 to 9, then construct the atom similarity matrix, with a size of 72×72, under different message passing phases. As seen in Fig. 6(a), the distribution of the atom similarity matrix before the PPM training mainly depends on the input atom features, e.g., electronegativity and valence electrons. The atom similarity matrix is almost unchanged after three message passing process, which indicates that the PPM cannot extract effective information from the input structures. After the training, atoms are gradually separated into two clusters, i.e., the Al-N and Si-C clusters. For example, the Al atoms ($1 \sim 9$) show a higher similarity with the N atoms ($10 \sim 18$) than the Si atoms ($19 \sim 27$) and C atoms

($28 \sim 36$), consistent with the fact that Al bonds with N, not Si or C.

The results are consistent with the structural characteristics of wurtzite structure and the matching of valence electrons. For instance, the binary compound AlN forms a stable wurtzite structure by the sp^3 hybridization for $3s^2 3p$ and $2s^2 2p^3$ electrons of Al and N atoms, respectively. Similarly, SiC forms a wurtzite structure by sp^3 hybridization for $3s^2 3p^2$ electrons of Si and $2s^2 2p^2$ electrons of C. Thereby the Al atoms at the lowest-energy order structure must be bonding with N first, while the high-energy structures have more random atomic distributions, as seen on the top of Fig. 6(c). Moreover, the Al-N and Si-C layers, which form the hexagonal rings with the neighbor atoms along the [001] direction,
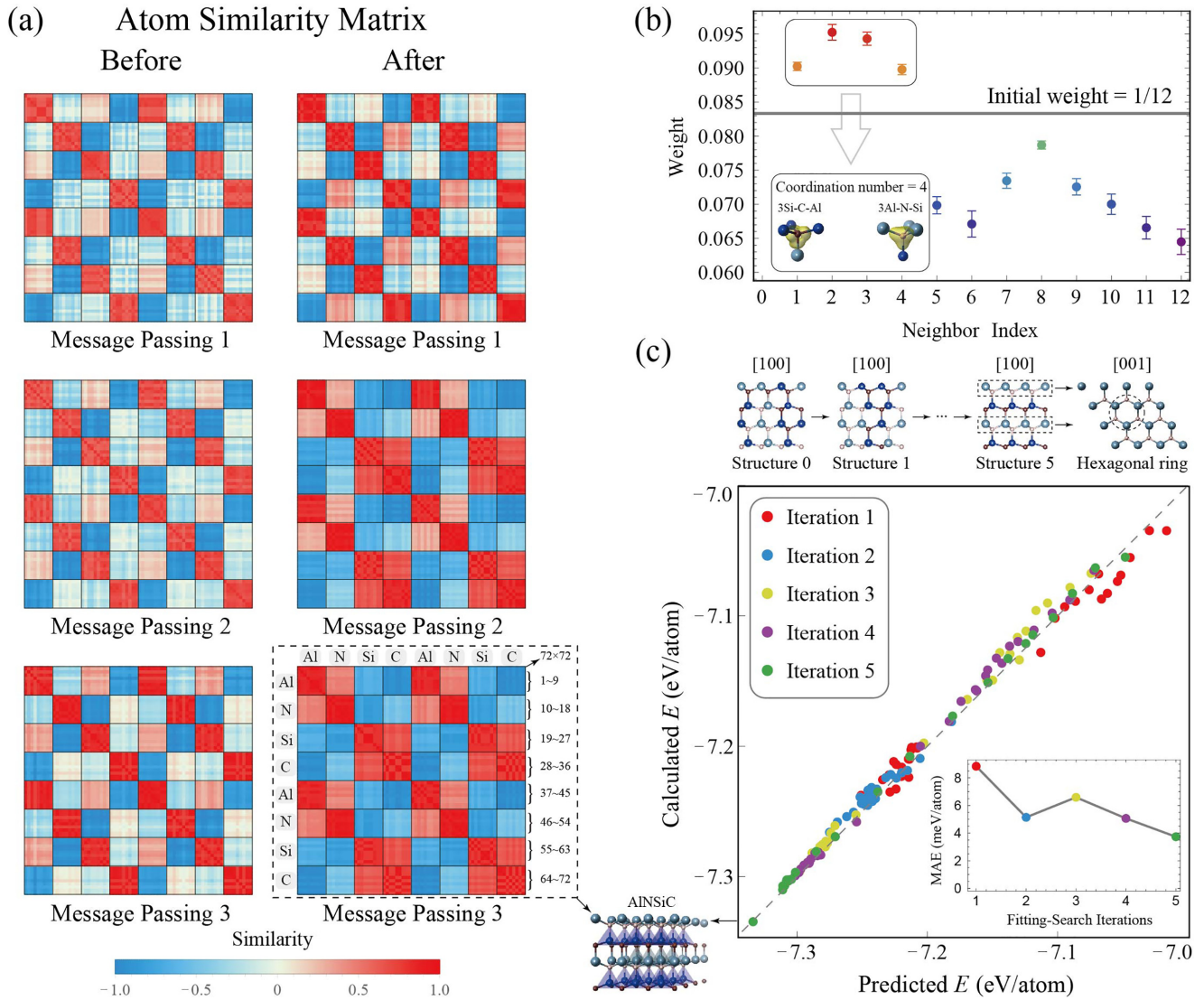
FIG. 6. (a) The atom similarity matrix of the ordered AlNSiC structure before and after PPM training. The similarity coefficient is defined as the cosine distance between the atom vectors. (b) Atomic weights of the 12 neighboring atoms by the trained PPM, notice that the nearest four atoms have higher weights than the others. (c) Comparison between the PPM predicted energies and the calculated DFT energies, with the inset showing the mean absolute error (MAE) of the predictions.

usually appear alternatively in the lowest-energy structure due to the crystal periodicity. Thus the Al and N atoms show high similarity in the atom similarity matrix, corresponding to the strong tendency of their bonding.

Figure 6(b) shows that, although the initial weights are uniformly set to 1/12, after the training, the weights of the four nearest-neighbor atoms become larger, while the weights for the other eight neighbors become smaller. That is, the energy prediction is predominately determined by the four nearest neighbor, and it matches perfectly with the four coordinated tetrahedrons, e.g., 3Si-C-Al and 3Al-N-Si, in the lowest-energy structure of AlNSiC. The atom similarity matrix and the weights verify that the PPM can effectively extract the structural characteristics from the training data. The learned weights ensures that the choice of atom exchanges is not as random as the SA, and reduces the cost of choosing the energy descent path in DRSA.

Figure 6(c) displays the PPM predicted energies against the DFT calculated values and the MAE for each fitting-search iteration. Based on the active learning, 30 most representative samples from the DRSA paths in each fitting-search iteration are selected to enhance as much as possible the prediction accuracy of PPM. As seen in Fig. 6(c), with the progress of the fitting-search iteration, the searching area gradually moves from the initial high-energy structures to the low-energy structures, and the corresponding MAE usually decreases. The results reflect that the search program is capable of effectively finding the low-energy area in the PES and finally obtains the ordered structure of AlNSiC with an energy of -7.33 eV/atom. Notice that all the energy values shown below refer to the energies per atom.

The interpretability of ML model as shown in Fig. 6 is a common feature of our PPM for all the alloys tested. Additional example can be found in Fig. S6 of the SM [46].

IV. CONCLUSIONS

A machine-learning assisted structural prediction method named as CCOP is proposed. The novel features of CCOP include: (i) Using a modified CGCNN model as the PPM to replace the expensive DFT calculations. (ii) Guiding the structural search paths by DRSA, a method combining the advantages of RL agent and Metropolis criterion, to accelerate the searching process. (iii) Employing an active learning based sample selection method to reduce the PPM prediction error and minimize the size of the training set.

Through testing applications concerning the structural searches of 4 different lattices and six multicomponent alloys, it is demonstrated that: (i) The CCOP can perform effective structural search across a diverse range of compositions and lattices. (ii) DRSA outperforms SA, PPO2, and the random search approach by finding the lowest-energy structures with the smallest number of steps. The benefit of DRSA is more pronounced when considering that SA, PPO2, and the random search miss most of the lowest-energy structures for six alloys. (iii) Selecting samples through active learning makes the PPM adaptive during the search, resulting in an efficient exploration of the low-energy area of the PES. (iv) The PPM has the desired feature of interpretability, since the results of the atom similarity matrix and the atom exchange weights for the nearest neighbors are both consistent with the physical intuition.

Overall, the integrated search framework of CCOP is found to cut the computational cost of a conventional SA by two orders of magnitude. Moreover, the CCOP can be integrated with the symmetry of space groups to enhance the efficiency of crystal structure prediction, thereby accelerating the pace of materials discovery.

[1] A. Jain, Y. Shin, and K. A. Persson, Nat. Rev. Mater. **1**, 15004 (2016).

[2] A. R. Oganov, C. J. Pickard, Q. Zhu, and R. J. Needs, Nat. Rev. Mater. **4**, 331 (2019).

[3] H. Liang, H. Zhong, S. Huang, and Y. Duan, J. Phys. Chem. Lett. **12**, 10975 (2021).

[4] H. Liang and Y. Duan, Nanoscale **13**, 11994 (2021).

[5] H. Liang and Y. Duan, Chin. Phys. B **31**, 076301 (2022).

[6] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, Science **220**, 671 (1983).

[7] L. T. Wille, Nature (London) **325**, 374 (1987).

[8] K. Doll, J. C. Schön, and M. Jansen, Phys. Chem. Chem. Phys. **9**, 6128 (2007).

[9] C. J. Pickard and R. J. Needs, J. Phys.: Condens. Matter **23**, 053201 (2011).

[10] D. M. Deaven and K. M. Ho, Phys. Rev. Lett. **75**, 288 (1995).

[11] S. M. Woodley, P. D. Battle, J. D. Gale, and C. Richard A. Catlow, Phys. Chem. Chem. Phys. **1**, 2535 (1999).

[12] A. O. Lyakhov, A. R. Oganov, H. T. Stokes, and Q. Zhu, Comput. Phys. Commun. **184**, 1172 (2013).

[13] Y. Wang, J. Lv, L. Zhu, and Y. Ma, Phys. Rev. B **82**, 094116 (2010).

[14] Y. Wang, J. Lv, L. Zhu, and Y. Ma, Comput. Phys. Commun. **183**, 2063 (2012).

[15] P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964).

[16] W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).

[17] E. Onbaşoğlu and L. Özdamar, J. Global Optim. **19**, 27 (2001).

[18] K. T. Schütt, H. E. Sauceda, P.-J. Kindermans, A. Tkatchenko, and K.-R. Müller, J. Chem. Phys. **148**, 241722 (2018).

[19] C. Chen, W. Ye, Y. Zuo, C. Zheng, and S. P. Ong, Chem. Mater. **31**, 3564 (2019).

[20] K. T. Schütt, F. Arbabzadah, S. Chmiela, K. R. Müller, and A. Tkatchenko, Nat. Commun. **8**, 13890 (2017).

[21] T. Xie and J. C. Grossman, Phys. Rev. Lett. **120**, 145301 (2018).

[22] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot *et al.*, Nature (London) **529**, 484 (2016).

[23] S. So, T. Badloe, J. Noh, J. Bravo-Abad, and J. Rho, Nanophotonics **9**, 1041 (2020).

[24] S. K. Gottipati, B. Sattarov, S. Niu, Y. Pathak, H. Wei, S. Liu, S. Liu, S. Blackburn, K. Thomas, C. Coley *et al.*, in *Proceedings of the 37th International Conference on Machine Learning* (PMLR, 2020), Vol. 119, pp. 3668–3679.

[25] M. Olivecrona, T. Blaschke, O. Engkvist, and H. Chen, J. Cheminformatics **9**, 48 (2017).

[26] D. C. Elton, Z. Boukouvalas, M. D. Fuge, and P. W. Chung, Mol. Syst. Des. Eng. **4**, 828 (2019).

[27] A. Zhavoronkov, Y. A. Ivanenkov, A. Aliper, M. S. Veselov, V. A. Aladinskiy, A. V. Aladinskaya, V. A. Terentiev, D. A. Polykovskiy, M. D. Kuznetsov, A. Asadulaev *et al.*, Nat. Biotechnol. **37**, 1038 (2019).

[28] D. W. Opitz and J. W. Shavlik, in *Advances in Neural Information Processing Systems* (MIT Press, Cambridge, MA, 1996), pp. 535–541.

[29] J. S. Smith, B. Nebgen, N. Lubbers, O. Isayev, and A. E. Roitberg, J. Chem. Phys. **148**, 241733 (2018).

[30] Z.-H. Zhou, J. Wu, and W. Tang, Artif. Intell. **137**, 239 (2002).

[31] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, arXiv:1707.06347.

[32] G. Kresse and J. Hafner, Phys. Rev. B **47**, 558 (1993).

[33] G. Kresse and J. Hafner, Phys. Rev. B **49**, 14251 (1994).

[34] G. Kresse and J. Furthmüller, Phys. Rev. B **54**, 11169 (1996).

[35] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).

[36] P. E. Blöchl, Phys. Rev. B **50**, 17953 (1994).

[37] G. Kresse and D. Joubert, Phys. Rev. B **59**, 1758 (1999).

[38] J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl, in *Proceedings of the 34th International Conference on Machine Learning* (PMLR 2017), Vol. 70, pp. 1263–1272.

[39] X. Glorot, A. Bordes, and Y. Bengio, in *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, Proceedings of Machine Learning Research, Vol. 15, edited by G. Gordon, D. Dunson, and M. Dudík (PMLR, Fort Lauderdale, FL, 2011), pp. 315–323.

[40] A. P. Bartók, M. C. Payne, R. Kondor, and G. Csányi, Phys. Rev. Lett. **104**, 136403 (2010).

[41] S. P. Ong, W. D. Richards, A. Jain, G. Hautier, M. Kocher, S. Cholia, D. Gunter, V. L. Chevrier, K. A. Persson, and G. Ceder, Comp. Mater. Sci. **68**, 314 (2013).

[42] K. Cho, B. van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (Association for Computational Linguistics, 2014), pp. 1724–1734.

[43] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, arXiv:1412.3555.

[44] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning* (MIT Press, Cambridge, MA, 2016).

[45] K. Hornik, M. Stinchcombe, and H. White, Neural Netw. **2**, 359 (1989).

[46] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevMaterials.7.033802 for further insights on the property prediction model, the impact of network freezing and clustering, additional details about the search algorithm, and the time estimation.

[47] T. Chen, M. Li, Y. Li, M. Lin, N. Wang, M. Wang, T. Xiao, B. Xu, C. Zhang, and Z. Zhang, in *Neural Information Processing Systems, Workshop on Machine Learning Systems* (Barcelona, Spain, 2015).

[48] D. P. Kingma and J. Ba, in 3rd International Conference for Learning Representations, San Diego, 2015.

[49] K. He, X. Zhang, S. Ren, and J. Sun, in *2015 IEEE International Conference on Computer Vision (ICCV)* (2015), pp. 1026–1034, https://ieeexplore.ieee.org/document/7410480.

[50] J. Schulman, P. Moritz, S. Levine, M. Jordan, and P. Abbeel, arXiv:1506.02438.

[51] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction* (MIT Press, Cambridge, MA, 2018).

[52] V. Mnih, A. P. Badia, M. Mirza, A. Graves, T. Lillicrap, T. Harley, D. Silver, and K. Kavukcuoglu, in *Proceedings of the 33rd International Conference on Machine Learning* (PMLR, 2016), Vol. 48, pp. 1928–1937.

[53] X. Glorot and Y. Bengio, in *Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2015* (IEEE, Piscataway, NY, 2015), pp. 1026–1034.

[54] L. van der Maaten and G. Hinton, J. Mach. Learn. Res. **9**, 2579 (2008).

[55] B. Bahmani, B. Moseley, A. Vattani, R. Kumar, and S. Vassilvitskii, arXiv:1203.6402.

[56] E. Tani, T. Sōma, A. Sawaoka, and S. Saito, Jpn. J. Appl. Phys. **14**, 1605 (1975).

[57] L. Dong and S. P. Alpay, Phys. Rev. B **84**, 035315 (2011).

[58] L. Li, J. T. Muckerman, M. S. Hybertsen, and P. B. Allen, Phys. Rev. B **83**, 134202 (2011).

[59] K. Gubaev, E. V. Podryabinkin, and A. V. Shapeev, J. Chem. Phys. **148**, 241727 (2018).

[60] K. Gubaev, E. V. Podryabinkin, G. L. Hart, and A. V. Shapeev, Comp. Mater. Sci. **156**, 148 (2019).