Computational Materials Science 218 (2023) 112000

Contents lists available at ScienceDirect

Computational Materials Science

journal homepage: www.elsevier.com/locate/commatsci

Full length article

Semigrand-canonical Monte-Carlo simulation methods for charge-decorated cluster expansions

Fengyu Xie *, Peichen Zhong, Luis Barroso-Luque, Bin Ouyang, Gerbrand Ceder

Department of Materials Science and Engineering, University of California, Berkeley, 94720, CA, United States
Materials Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, 94720, CA, United States

ARTICLE INFO

Keywords:
Cluster expansion (CE)
Semigrand-canonical Monte-Carlo (sGCMC)
Ionic solid

ABSTRACT

Monte-Carlo sampling of lattice model Hamiltonians is a well-established technique in statistical mechanics for studying the configurational entropy of crystalline materials. When the species to be distributed on the lattice carry charge, the charge balance constraint on the overall system prohibits single-site Metropolis exchanges in MC. In this article, we propose two methods to perform MC sampling in the semigrand-canonical ensemble in the presence of a charge-balance constraint. The table-exchange method (TE) constructs small charge-conserving excitations, and the square-charge bias method (SCB) allows the system to temporarily drift away from charge neutrality. We illustrate the effect of internal hyper-parameters on the efficiency of these algorithms and suggest practical strategies on how to employ these algorithms in real applications.

1. Introduction

Configurational disorder is particularly important for understanding the thermodynamic properties of multi-component materials at finite temperatures. The cluster-expansion (CE) method has been a successful approach for studying the statistical mechanics of configurational disorder in solids [1–4]. The CE method has been used to calculate phase diagrams in alloys [5–8] and ionic solids [9–12], predict the short-range order related properties [13–16], find the ground-state ordering in alloys [17–22], and compute voltage profiles of battery electrode materials [23–27].

The CE method can be understood as a generalization of the Ising model. The micro-states of a solid solution are represented as sequences of occupancy variables $\boldsymbol{\sigma}$, which denote the chemical species occupying the crystallographic sites in a structure. The configuration energy is represented as a function of occupancy variables and is expanded as a sum of many-body interactions,

$$E(\boldsymbol{\sigma}) = \sum_{\beta} m_{\beta} J_{\beta} \left\langle \Phi_{\alpha}(\boldsymbol{\sigma}) \right\rangle_{\alpha \in \beta}, \tag{1}$$

where the *cluster basis functions* Φ_{α} take as input the occupancy values of different clusters of multiple sites. The cluster basis functions are then grouped and averaged over orbits β of symmetrically equivalent clusters to generate *correlation functions* $\langle \Phi_{\alpha} \rangle_{\alpha \in \beta}$; and m_{β} is the multiplicity of orbit β per crystallographic unit cell. The linear-expansion coefficients J_{β} are called *effective cluster interactions* (ECI). In a typical approach, ECIs are fitted to the first-principles calculated energy of a large number of ordered structures using (regularized) linear regression [28–36]. Thermodynamic quantities can be obtained by sampling the CE energy with Monte-Carlo simulations (CE-MC) [6,37–39]. The CE-MC workflow allows fast statistical mechanics computation of configurational disorder using only a relatively small amount of first-principles calculations. Further details of the CE-MC method can be found in various review papers [35,40–44].

CE-MC can be performed in a canonical ensemble or in a semigrand-canonical ensemble. In a canonical ensemble, the configuration states are sampled with a fixed composition of each species. Using the Metropolis–Hastings algorithm [45,46], a typical MC step involves swapping the species occupying two randomly chosen sites (canonical swap). In a semigrand-canonical ensemble, the states are sampled under fixed chemical potentials and allowing the relative amounts of each species to vary. An MC step in the semigrand-canonical ensemble usually replaces the occupying species on one randomly chosen site with another species (single-species exchange). Semigrand-canonical simulations are the preferred approach for studying phase transition with compositional changes in solids. In a semigrand-canonical simulation, the system is always in a single-phase state, and transitions between compositions are relatively easy to observe. In contrast, multiple phases with different compositions can coexist in canonical simulations, giving a disproportionate influence to the interfacial energy between phases.

Single-species exchanges can be applied without issue when the species are all charge-neutral atoms. However, in an ionic system in

* Corresponding author at: Department of Materials Science and Engineering, University of California, Berkeley, 94720, CA, United States.
E-mail addresses: fengyu_xie@berkeley.edu (F. Xie), gceder@berkeley.edu (G. Ceder).

https://doi.org/10.1016/j.commatsci.2022.112000
Received 28 September 2022; Received in revised form 6 December 2022; Accepted 21 December 2022
Available online 27 December 2022
0927-0256/Published by Elsevier B.V.

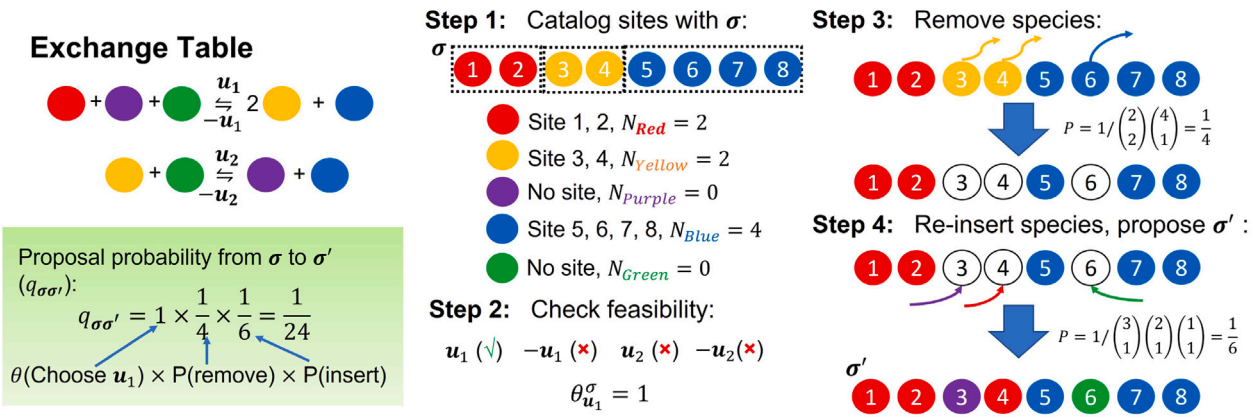

Fig. 1. The procedure for proposing a table exchange in a conceptual quinary system. The system contains five different species (colored circles) and eight sites in a single sub-lattice (labeled with indices). Four exchange directions $\mathbf{u}_1$, $-\mathbf{u}_1$, $\mathbf{u}_2$ and $-\mathbf{u}_2$ are included in the exchange table. In the green box to the lower-left, the probability for proposing a particular configuration σ' from σ ($q_{\sigma\sigma'}$) is calculated as the product of three probabilities: the probability for selecting an exchange direction, the probability for choosing removed species, and the probability for inserting new species to empty sites.

which species carry charge, net zero charge needs to be maintained, essentially coupling allowed species exchanges. For simulating ionic liquids, various methods have been proposed such as: inserting and removing only charge-neutral combinations of ions [47]; performing single insertion or deletion while controlling the statistical average of the net charge to be equal to zero [48,49]; or using an extended semigrand-canonical ensemble [50]. In practice, methods involving simple multi-species exchanges have been applied to specific battery materials [10,51]. However, charge-balanced MC methods in lattice-model CEs with arbitrary complexity have not been formally addressed in the literature yet.

In this study, we introduce two CE-MC sampling methods to handle the charge-balance constraint in the semigrand-canonical CE-MC for ionic systems with charge decoration. In the first method, which we refer to as the table-exchange (TE) method, MC samples are kept charge-neutral by using charge-conserving multi-species exchanges. The second method, labeled the square-charge bias (SCB) method, combines single-species exchanges which allows a non-zero charge with a penalty on the net charge to bias the simulation towards zero charge. We benchmark the computational efficiency of both methods over their respective hyper-parameters and demonstrate proper usage strategies using a complex rocksalt system with configurational disorder.

2. Methods

For simplicity, the formalism in the following discussion is limited to crystal structures with a single sub-lattice. However, the methodology can be easily extended to multiple sub-lattices. We also restrict our investigation to the application of a charge-balance constraint; although more generic integral constraints on the composition (e.g., fixing the atomic ratio between particular components to follow a specific hyper-plane in the composition space) can be addressed in the same manner.

2.1. Table-exchange method

In the semigrand-canonical ensemble with species carrying charge, every possible occupancy state must satisfy the following constraints:

$$\begin{aligned} &\sum_{s=1}^{S} C_s n_s = 0, \\ &\sum_{s=1}^{S} n_s = N, \\ &n_s \in \mathbb{N},\ \forall s \in \{1, 2, \dots, S\}, \end{aligned} \tag{2}$$

where s is the label of a species, n_s is the amount of species s in configuration $\boldsymbol{\sigma}$, and N is the total number of sites in the system. The first equation is a charge-balance constraint, where C_s is the charge of species s. The second equation requires the number of species to be equal to the number of sites. Eq. (2) is a system of linear Diophantine equations with natural number solutions. All integral solutions $\boldsymbol{n} = (n_1, \dots, n_S)$ to these Diophantine equations can be represented as a bounded fraction of a $(S-2)$-dimensional integer grid in $\mathbb{N}^S$ specified as follows, [52]

$$\begin{aligned} &\boldsymbol{n} = \boldsymbol{n}_0 + \sum_{i=1}^{S-2} x_i \boldsymbol{v}_i, \\ &\text{s.t.}\ \ x_i \in \mathbb{Z},\ \boldsymbol{v}_i \in \mathbb{Z}^S \\ &\quad\ n_s \in \mathbb{N},\ n_s \le N, \end{aligned} \tag{3}$$

where $\boldsymbol{n}_0$ is a base integer solution to Eq. (2), $\boldsymbol{v}_i$ are $S-2$ linearly independent basis vectors, and x_i are integer coordinates on the grid.

Any vector $\boldsymbol{u} = \boldsymbol{n}' - \boldsymbol{n}$ pointing from one solution $\boldsymbol{n}$ on the integer grid to another solution $\boldsymbol{n}'$ is called an *exchange direction*. An exchange direction physically represents a composition transfer under the charge-balance constraint. A selected set V among all possible exchange directions $\boldsymbol{u}$ is called an exchange table. Based on the exchange table V, we can define a random walk process between charge-balanced compositions as follows:

(1) Using the current composition $\boldsymbol{n}$, select one direction $\boldsymbol{u}$ from all feasible directions in the predefined exchange table V. The feasibility of a direction $\boldsymbol{u}$ is defined with the requirement that for all $u_s < 0$ (i.e. species s is being removed), we have $n_s > -u_s$, ensuring a move towards direction $\boldsymbol{u}$ would not result in a negative amount of any species.

(2) Perform the operation to the occupancy configuration according to the selected exchange $\boldsymbol{u}$, such that the composition $\boldsymbol{n}$ changes to $\boldsymbol{n} + \boldsymbol{u}$. Given $\boldsymbol{u} = (u_1, u_2, \dots, u_S)$. One such operation can be achieved by removing $-u_s$ of species s from the occupancy for all $u_s < 0$; then inserting u_s of species s into the empty sites for all $u_s > 0$. We call such an operation a *table exchange*. A table exchange results in a simultaneous exchange of species on multiple sites and is always charge-conserving. The number of sites U to be exchanged is called the *exchange size* in direction $\boldsymbol{u}$. Because any exchange should conserve the site number, $\sum_s u_s = 0$, it follows that $U = \sum_{u_s>0} u_s = \sum_{u_s<0} -u_s$.

A complete exchange table should have ergodicity, meaning that an MC simulation should be able to reach any charge-balanced composition starting from any arbitrary configuration. Once ergodicity is satisfied, the number of sites involved in the exchange directions should be minimal, as exchanging a large number of sites in an MC step can lead to a low acceptance ratio and thus inefficient sampling of the

configuration space. It is not necessary, nor practical, to include all possible directions $\boldsymbol{u}$ in the table. Usually, as a minimal setup, one can choose $S-2$ linearly independent basis vectors ($\{\boldsymbol{v}_i\}$) with minimal exchange size as well as their inverse vectors ($\{-\boldsymbol{v}_i\}$). The ergodicity of a table can be checked by enumerating charge-balanced compositions in a specific super-cell size as vertices of a graph, and checking graph connectivity between the compositions using vectors in the table as the edges of the graph. If ergodicity is not satisfied with the minimal setup, and the unreachable compositions are of interest, vectors linking the disconnected composition to other compositions should be added to the table, until the ergodicity is guaranteed.

According to the statements above, given an exchange table V, one can propose semigrand-canonical MC steps using the following procedure, which is illustrated schematically in Fig. 1,

1. Create a catalog of sites in the lattice. For a starting occupancy state $\boldsymbol{\sigma}$, indices j of sites are grouped by their current occupying species s to create sets of sites for each distinct species $J_s = \{j|\sigma_j = s\}$.
2. Select one feasible direction $\boldsymbol{u}$ from table V. The subset of table V with all feasible directions at occupancy $\boldsymbol{\sigma}$ is denoted as $V_{\boldsymbol{\sigma}}$. The probability of selecting direction $\boldsymbol{u}$ is denoted as $\theta^{\sigma}_{\boldsymbol{u}}$. In this work, we select all feasible directions with an equal probability ($\theta^{\sigma}_{\boldsymbol{u}} = 1/\mathrm{card}(V_{\boldsymbol{\sigma}}), \forall \boldsymbol{u} \in V_{\boldsymbol{\sigma}}$).
3. For all $u_s < 0$, randomly pick $-u_s$ sites from catalog J_s without replacement. Select all possible picking combinations at equal probability ($P = 1/\prod_{u_s<0} \binom{n_s}{-u_s}$) and remove the species from selected sites.
4. For all $u_s > 0$, randomly select u_s empty sites from the U empty sites created in Step 3 without replacement, and insert species s back to selected empty sites. All possible combinations of choices are selected with equal probability ($P = \prod_{u_s>0} u_s!/U!$). Propose the resulting occupancy state $\boldsymbol{\sigma}'$ as the next step in the Markov chain.

Note that the procedure above can result in an asymmetry between the exchange proposal probability from $\boldsymbol{\sigma}$ to $\boldsymbol{\sigma}'$ and the inverse proposal probability from $\boldsymbol{\sigma}'$ back to $\boldsymbol{\sigma}$. Such a proposal asymmetry can be balanced by multiplying with a composition dependent importance factor to adjust the acceptance probability as given by Eq. (4), such that detailed balance is ensured and the correct distribution is reached (see Supplementary Information for a detailed derivation),

$$p_{\boldsymbol{\sigma}\boldsymbol{\sigma}'} = \min\left\{1,\ \frac{\theta^{\sigma'}_{-\boldsymbol{u}} \prod_{u_s\neq 0} n_s!}{\theta^{\sigma}_{\boldsymbol{u}} \prod_{u_s\neq 0}(n_s+u_s)!} \exp\left[-\frac{1}{k_B T}\left(\Delta E_{\boldsymbol{\sigma}\boldsymbol{\sigma}'} - \sum_s \mu_s u_s\right)\right]\right\}. \tag{4}$$

In addition to table exchanges which change the composition, a portion ($0 \le w < 1$) of canonical swaps can also be mixed in the proposal. These canonical swaps can directly transfer between occupancies under the same composition with much less computational cost than table exchanges and are added to help explore occupancies with the same composition more efficiently, rather than having to do so with a combination of table exchanges. In the numerical results section, we will illustrate the importance of hyper-parameter w in the TE method.

2.2. Square-charge bias method

Proposing a table-exchange step and computing its energy change is more time-consuming compared to single-species exchanges, therefore it is desirable to find a method using single-species exchanges that still conserves charge balance. In the square-charge bias (SCB) method, we use single-species exchanges to span all occupancies regardless of charge balance. States in the Markov chain are allowed to leave charge-balance, however, a penalty proportional to the square of the net charge is added to the Hamiltonian to drive the sampled configurations back to charge balance. The acceptance probability of each single-species exchange step is evaluated using the following penalized Hamiltonian

$$H_{\mu,\lambda}(\boldsymbol{\sigma}) = E(\boldsymbol{\sigma}) - \sum_s \mu_s n_s + \lambda k_B T C(\boldsymbol{\sigma})^2, \tag{5}$$

where $E(\boldsymbol{\sigma})$ is the energy of occupancy $\boldsymbol{\sigma}$ computed from CE. The charge penalty factor $\lambda > 0$ is a hyper-parameter in the SCB method, and $k_B T$ is included explicitly in the penalty to keep λ dimensionless. $C(\boldsymbol{\sigma})$ is the net charge of occupancy $\boldsymbol{\sigma}$

$$C(\boldsymbol{\sigma}) = \sum_s C_s n_s. \tag{6}$$

In an SCB run, a simulation starts from a charge-balanced state. After reaching thermal equilibration, from all states in the equilibrated sample, we compute the average of physical quantities with only charge-balanced states (i.e., states with $C(\sigma) = 0$). In doing so, the charge-balance constraint is rigorously satisfied in the estimation of physical quantities. Furthermore, since when $C(\boldsymbol{\sigma}) = 0$ the penalized Hamiltonian will be equal to the physical Hamiltonian, $H_{\mu,\lambda}(\sigma) = E(\boldsymbol{\sigma}) - \sum_s \mu_s n_s$, the true semigrand-canonical distribution is recovered by using charge-balanced states only. The effect of the hyper-parameter λ on SCB is demonstrated in the numerical results section.

2.3. Comparing computational efficiency of sampling methods

If a CE-MC algorithm has hyper-parameters, it is desirable to optimize them such that the thermodynamic properties can be estimated accurately with minimal computational cost. To estimate the ensemble average $\overline{\theta}$ of a physical quantity θ, a Markov chain of states is generated using CE-MC, and at each step p, the value of θ for the current configuration is recorded as θ_p. We denote $\overline{\theta}_{[p,q]}$ as the mean of θ in a block from step p to step q. When using the SCB method, block means are computed using only charge-balanced states in each block. After thermal equilibration, we define the *block mean variance* at block length L, $\mathrm{Var}(\overline{\theta}_L)$, as the variance of the *block means* $\overline{\theta}_{[p,p+L]}, \overline{\theta}_{[p+L,p+2L]}, \cdots$ for blocks of L samples. The block mean variance can be used as a measure of uncertainty when estimating $\overline{\theta}$ using a block mean.

Suppose the true variance of θ in the ensemble is τ^2, then the *sampling efficiency* based on the property θ can be defined as follows [53],

$$\mathrm{eff}(\theta) = \frac{\tau^2}{L\mathrm{Var}(\overline{\theta}_L)}. \tag{7}$$

For ideal independent random sampling, one can expect $\mathrm{Var}(\overline{\theta}_L) = \tau^2/L$, such that $\mathrm{eff}(\theta) = 1$. In reality, Metropolis samples are always correlated and the efficiency is expected to be lower than 1, $\mathrm{eff}(\theta) < 1$. A CE-MC algorithm with higher sampling efficiency is less correlated and can thus reduce the uncertainty of estimation to an appropriate level with fewer sampling steps.

In the TE method, the time cost of a table exchange is significantly higher than a canonical swap, such that counting the number of Metropolis steps does not accurately reflect the computational cost. In this work, we use a modified version of Eq. (7) to evaluate the sampling efficiency. We replace the block length L with the average CPU time spent in each block $\overline{T}_L$. $\overline{T}_L$ is divided by the average CPU time $\overline{\delta t}$ spent in a single-species exchange, such that the efficiency value of algorithms is dimensionless, and can be compared regardless of the hardware used:

$$\mathrm{eff}_t(\theta) = \frac{\tau^2 \overline{\delta t}}{\overline{T}_L \mathrm{Var}(\overline{\theta}_L)}. \tag{8}$$

We define $\mathrm{eff}_t(\theta)$ in Eq. (8) as the *computational efficiency* for the property θ. We use $\mathrm{eff}_t(\theta)$ for benchmarking the proposed algorithms under varied hyper-parameters. Additionally, since the true ensemble variance τ^2 can never be accessed a-priori, we estimate τ^2 with the variance of θ in the whole thermally equilibrated sample.

2.4. Choosing chemical potentials

In the semigrand-canonical ensemble, the probability of a specific configuration $\boldsymbol{\sigma}$ at temperature T is proportional to the following Boltzmann factor

$$P(\boldsymbol{\sigma}) \propto \exp\left(-\frac{E-\sum_{s=1}^{S-1}\tilde{\mu}_s n_s}{k_B T}\right), \tag{9}$$

where $\tilde{\mu}_s = \mu_s - \mu_S$ is the relative chemical potential of species s with respect to a given species S, and μ_s is the chemical potential of species s.

When the occupying species carry charge, a charge-balance constraint (Eq. (2)) must be enforced, and the number of independent degrees of freedom in chemical potential values are decreased to $S-2$. Substituting Eq. (3) into Eq. (9), the probability of configuration $\boldsymbol{\sigma}$ under a charge-balance constraint is

$$P(\boldsymbol{\sigma}) \propto \exp\left(-\frac{E-\sum_{i=1}^{S-2}\hat{\mu}_i x_i}{k_B T}\right), \tag{10}$$

where $\boldsymbol{v}_i$ is an integral basis vector of compositions in Eq. (3), $\hat{\mu}_i = \sum_{s=1}^{S}\mu_s v_{is}$ is the total change of chemical potentials for moving the composition one step towards basis vector $\boldsymbol{v}_i$, and x_i is the number of steps moved towards direction $\boldsymbol{v}_i$. We refer to $\hat{\mu}_i$ as the *basis exchange potential* in direction $\boldsymbol{v}_i$.

In the TE method, once all $S-2$ basis exchange potentials $\hat{\boldsymbol{\mu}}$ are selected, one can arbitrarily set chemical potentials $(\mu_0, \ldots, \mu_s, \ldots)$ for each species as long as the relations

$$\hat{\mu}_i = \sum_{s=1}^{S}\mu_s v_{is} \ (i = 1, \ldots, S-2) \tag{11}$$

are satisfied. However, in the SCB method, an extended semigrand-canonical ensemble is sampled without charge-balance constraint, resulting in a penalized probability distribution

$$P(\boldsymbol{\sigma}) \propto \exp\left(-\frac{E+\lambda k_B T C^2 - \sum_{s=1}^{S-1}\tilde{\mu}_s n_s}{k_B T}\right), \tag{12}$$

in which the probability to sample a charge-balanced configuration implicitly depends on the $S-1$ relative chemical potentials $\tilde{\boldsymbol{\mu}}$, and the efficiency of the SCB method is thus affected. In order to maximize the efficiency of SCB, one should optimize the following objective function to determine the chemical potential of each species:

$$\begin{aligned} &\boldsymbol{\mu}^*, \lambda^* = \underset{\boldsymbol{\mu},\lambda}{\operatorname{argmax}}\, \mathrm{eff}_t(\boldsymbol{\mu}, \lambda) \\ &s.t.\ \hat{\mu}_i = \sum_{s=1}^{S}\mu_s v_{is} \ (i = 1, \ldots, S-2). \end{aligned} \tag{13}$$

In practice, it is possible to simplify the problem above with some compromise in optimality. For example, in Eq. (13), one can fix the penalty factor λ to an empirical value (e.g. $\lambda = 1.0$) and only search for optimal chemical potentials. One can also replace the objective function with transfer rates, as will be discussed in Section 3.2.

3. Numerical results

In this section, we demonstrate the influence of the hyper-parameters on the computational efficiency and thermal equilibration in the TE and SCB methods. We performed CE-MC simulations in a disordered rocksalt system. A rocksalt crystal structure is a basic prototype of ionic materials consisting of an FCC cation and an FCC anion sub-lattice, mimicking the basic chemistry of some novel Li-ion cathode systems which have been modeled with CE-MC in recent studies [54–56]. In the system used, Li^+, Mn^{3+}, Zr^{4+} are distributed on the cation sub-lattice, and O^{2-}, F^- are present on the anion sub-lattice. We refer to this system as LMZOF. The primitive cell of the LMZOF system is presented in Fig. 2(a) and the corresponding exchange directions are shown in Fig. 2(b).

We performed simulations for the TE and SCB methods under various hyper-parameters w and λ. After thermal equilibration, we calculated the computational efficiencies based on Eq. (8) for the following quantities: (1) E (configurational energy per super-cell), (2) x_{LiMnO_2} (atomic percentage of $LiMnO_2$) and (3) $x_{Li_2ZrO_3}$ (atomic percentage of Li_2ZrO_3). To discuss how the hyper-parameters w and λ affect the computational efficiency and thermal equilibration in the TE and SCB methods, we designed two simulation experiments: (1) one at $T = 5000$ K to simulate the system in a state of complete solubility and (2) another at $T = 2000$ K to simulate the system in a single phase (Li_2ZrO_3). In experiment (1), the sampling efficiencies are plotted as a function of hyper-parameters w and λ. In experiment (2), the thermal equilibration process was demonstrated with simulation trace plots, which showed the value of thermodynamic properties (such as the composition and the configuration energy) as a function of the simulation step. The chemical potentials used in experiments (1) and (2) are chosen using the strategies described in Section 2.4. The chemical potential values and additional details of these simulations are provided in the Supplementary Information.

3.1. Simulations with table exchange

In the TE method, the parameter w tunes the ratio of table exchanges to canonical swaps, where only table exchanges can explore different compositions. Most physical systems have a critical temperature T_c (or a series of critical temperatures) below which they phase separate into phases of distinct compositions (compounds or elemental states). Above T_c, complete solubility can be found. Under such circumstances, a low w will include more table exchanges to explore a wide distribution of compositions and consequently gives better sampling efficiencies. Fig. 3(a) and (b) shows the TE computational efficiency under 5000 K, where all the components in LMZOF are fully miscible (see Supplementary Information). The computational efficiencies for the configurational energy and compositions are both maximized at $w = 0\%$, indicating that no canonical swaps should be included.

Nevertheless, it is not always safe to fully exclude canonical swaps. Below the critical temperature, the semigrand-canonical ensemble distribution is usually concentrated near the composition of a single phase; thus, the ability to explore different occupancies within the same composition is more important (namely, the ability of canonical state transfers). It is still possible to achieve a canonical transfer with only table exchanges by performing multiple exchanges when the sum of all exchange directions equals to zero. However, besides being computationally more expensive than a canonical swap, table exchanges perturb many sites simultaneously and are therefore more likely to propose energetically unfavorable configurations, which can result in a lower acceptance ratio. As a result, having too low of a canonical swap percentage w can reduce the computational efficiencies, and lead to slow thermal equilibration, especially at a relatively low temperature. Such an example is illustrated in Fig. 4(a) and (b) in LMZOF at 2000 K. Even though the simulation was able to reach equilibrium at a single phase composition (Li_2ZrO_3, Fig. 4(a)) for $w = 5\%$ (red), compared to $w = 50\%$ (green), it failed to equilibrate to the correct ground-state configuration (the layered structure, shown in Fig. 4(b)) within a time limit of 3000 s.

3.2. Simulations with square-charge bias

In the SCB method, the penalty factor λ controls the trade-off between the fraction of charge-balanced states in the Markov chain and the Metropolis acceptance probability. Fig. 3(c) and (d) show the sampling efficiency in LMZOF at 5000 K. Optimal efficiency is found at an intermediate λ value ($\lambda = 0.2$). When the penalty λ is too small, the simulation can wander too far from charge-balanced states, and barely

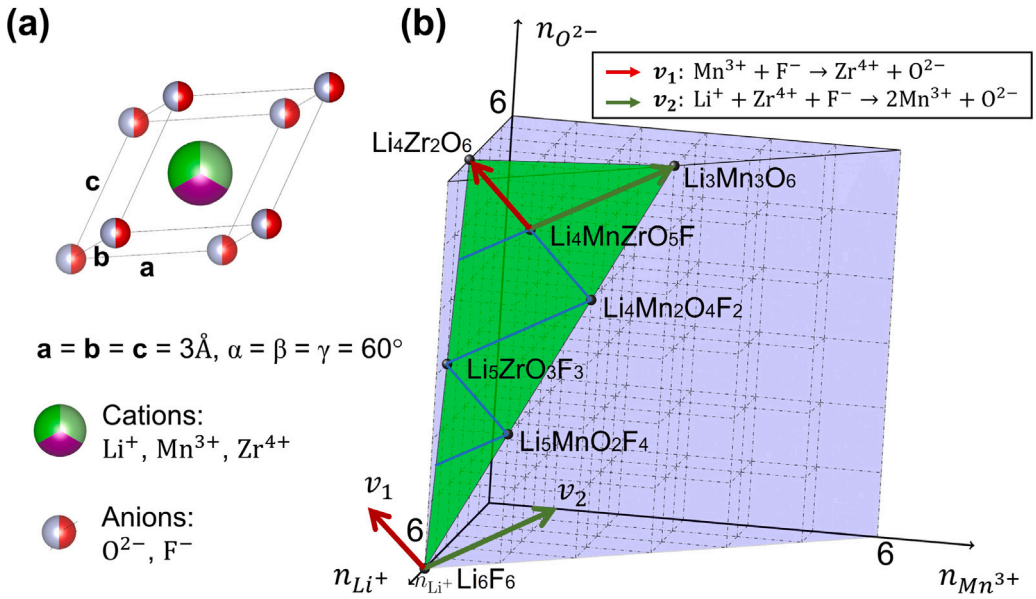

Fig. 2. Primitive cell and exchange directions in the LMZOF disordered rocksalt. (a) Rocksalt primitive cell of LMZOF, with partial occupancies of Li^+, Mn^{3+}, Zr^{4+} on the cation sub-lattice and O^{2-}, F^- on the anion sub-lattice. (b) Compositions of LMZOF in a super-cell of size 6 (6 cation sites and 6 anion sites). The x-, y-, and z-axis represent the amount of Li^+, Mn^{3+} and O^{2-}, respectively. The amount of Zr^{4+} and F^- can be computed by satisfying site number conservation on the cation and the anion sub-lattices. The purple dashed grid in three dimensions includes arbitrary compositions without enforcing charge balance. The solid grid on the green plane includes charge-balanced compositions only. Basis vectors $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ are marked with dark green and red arrows, respectively. The reaction formulas corresponding to $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ are listed on the top right. The inverse directions are not displayed.

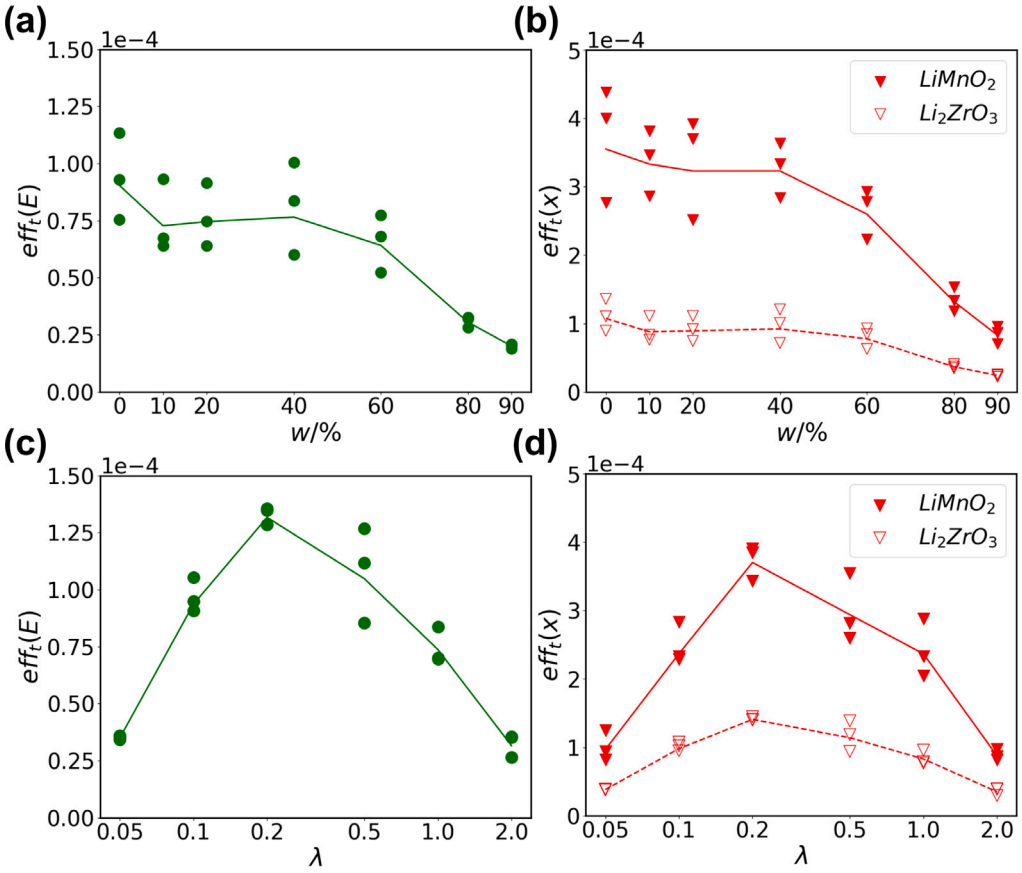

Fig. 3. (a)–(b) TE and (c)–(d) SCB computational efficiencies in the LMZOF system at $T = 5000$ K as a function of the canonical swap percentage w and the charge penalty factor λ. For each w and each λ, three simulations were run starting from different initial states. The average of three measurements for each w and each λ are connected with lines. (a) TE computational efficiencies for energy ($\mathrm{eff}_t(E)$, green dots and line) as a function of w. (b) TE computational efficiencies for sampling the $LiMnO_2$ composition ($\mathrm{eff}_t(x_{LiMnO_2})$, red solid triangles and solid line) and the Li_2ZrO_3 composition ($\mathrm{eff}_t(x_{Li_2ZrO_3})$, red hollow triangles and dashed line) as a function of w. (c) SCB computational efficiency for energy ($\mathrm{eff}_t(E)$, green dots and line) as a function of λ. (d) SCB computational efficiency for sampling the $LiMnO_2$ composition ($\mathrm{eff}_t(x_{LiMnO_2})$, red solid triangles and solid line) and the Li_2ZrO_3 composition ($\mathrm{eff}_t(x_{Li_2ZrO_3})$, red hollow triangles and dashed line) as a function of λ.

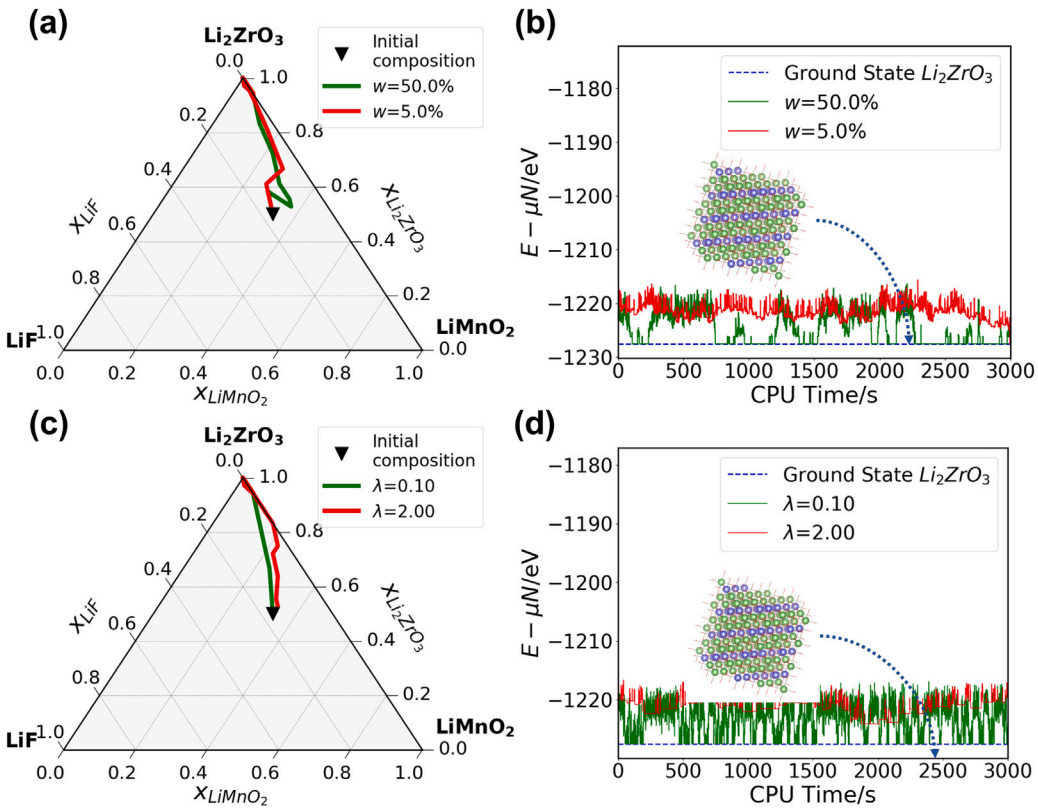

Fig. 4. (a)–(b) Trace plots of TE simulations in LMZOF system, at $T = 2000$ K, $w = 5\%$ and 50%; and (c)–(d) SCB simulations at $T = 2000$ K, $\lambda = 0.1$ and $\lambda = 2.0$. The simulations started from the same occupancy configuration. In (a) and (c), the simulated trajectory of compositions is plotted in the LMZOF phase space, and the initial state composition is marked with a black triangle. In (c) and (d), the simulated trajectories of the energy $(E - \mu N)$ are plotted as a function of the simulation time. The blue dashed baseline shows the energy of the Li_2ZrO_3 ground state. (a) Simulated trajectory of composition using $w = 5\%$(red) and 50%(green) in TE. (b) Simulated trajectory of energy with the chemical potential subtracted $(E - \mu N)$, using $w = 5\%$(red) and $w = 50\%$(green) in TE. (c) Simulated trajectory of compositions using $\lambda = 2.0$(red) and 0.1(green) in SCB. (d) Simulated trajectory of energy with the chemical potential subtracted $(E - \mu N)$, using $\lambda = 2.0$(red) and $\lambda = 0.1$(green) in SCB.

revisits charge-balanced configurations. Too large of a λ value limits its low-barrier pathways toward new charge-balanced configurations. Near either of these extreme circumstances, the sampling efficiency of SCB decreases. Fig. 4(c) and (d) show at $T = 2000$ K an overly large charge penalty $\lambda = 2.0$ (red) causes slow configurational equilibration to the layered Li_2ZrO_3 ground state due to the aforementioned limitation to access low-barrier pathways.

In the SCB approach, we can define the occupancy transfer rate (r_o) and the composition transfer rate (r_c) as follows:

$$r_o = \frac{\text{Count of occupancy transfers}}{\text{CPU time elapsed}}, \qquad r_c = \frac{\text{Count of composition transfers}}{\text{CPU time elapsed}}. \tag{14}$$

An occupancy transfer is counted when the Markov chain arrives at a new charge-balanced occupancy different from the last recorded charge-balanced state. A composition transfer is counted when a charge-balanced composition different from the last recorded composition is reached.

In Fig. 5, we computed the transfer rates in the SCB simulations at 5000 K in LMZOF. The maximum transfer rates occur at $\lambda = 0.5$. When compared to the sampling efficiency trend in Fig. 3(c) and (d), the efficiency at $\lambda = 0.5$ is only 20% lower than the optimal efficiency taken at $\lambda = 0.2$. Compared to the computational efficiency, the transfer rates can be tracked step by step without waiting for multiple blocks of the Markov chain to complete. They can also give a satisfactory estimation of the optimal λ. Therefore, when using the SCB method, one may instead choose an optimal λ to maximize the transfer rates as an alternative to maximizing the computational efficiency.

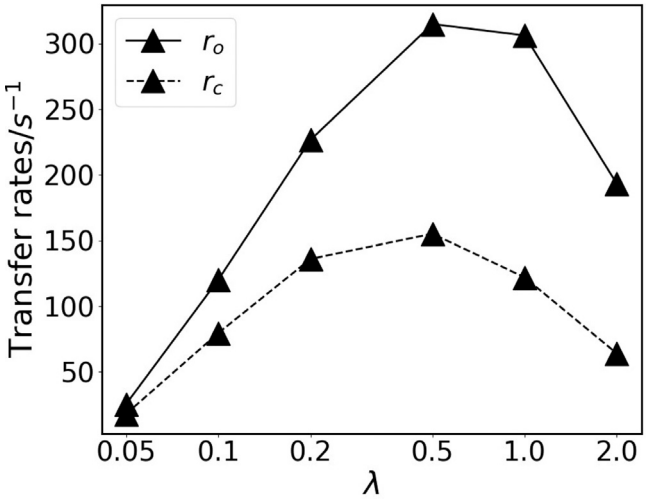

Fig. 5. Transfer rates of occupancy (r_o, solid line) and composition (r_c, dashed line) in SCB simulations of LMZOF at $T = 5000$ K, using varied λ.

3.3. The limitation of TE with a large table-exchange size

As illustrated in Fig. 3, the maximum computational efficiencies of the TE method and the SCB method are close in LMZOF ($\text{eff}_t(E) \approx 1.25 \times 10^{-4}$, $\text{eff}_t(x_{\text{LiMnO}_2}) \approx 4 \times 10^{-4}$ and $\text{eff}_t(x_{\text{Li}_2\text{ZrO}_3}) \approx 1.5 \times 10^{-4}$). Accordingly, the TE method is shown to have a similar performance as the SCB approach in a system with small table-exchange sizes (e.g., in

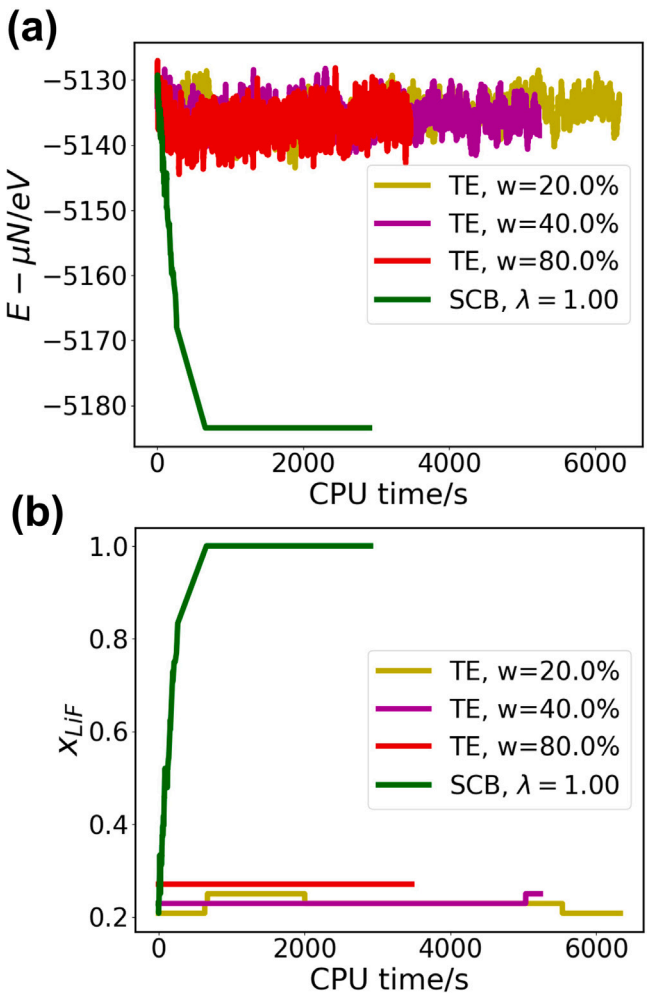

Fig. 6. Trace plots of TE simulations in LNMTOF system at $T = 1600$ K, $w = 20\%$ (yellow), 40% (purple) and 80% (red); and of an SCB simulation at the same temperature and $\lambda = 1.0$ (green). (a) Simulated trajectory of the energy subtracted by chemical potentials ($E - \mu N$) as a function of the simulation time. (b) Simulated trajectory of the LiF composition (x_{LiF}) as a function of the simulation time.

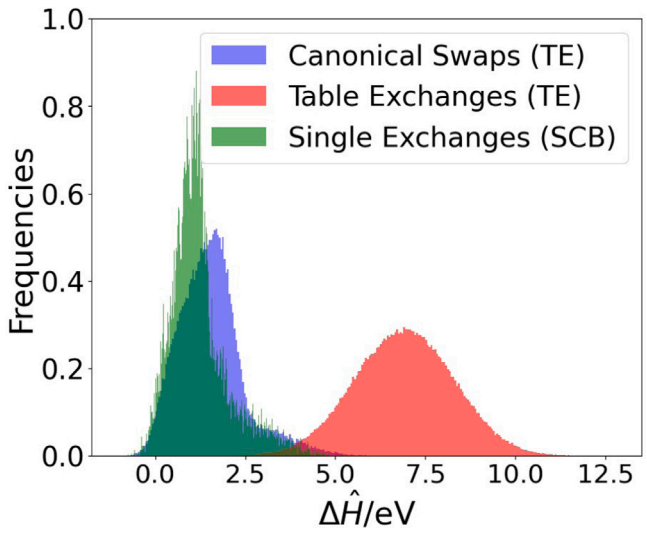

Fig. 7. Distribution of the effective perturbation energy ($\Delta\hat{H}$) of three types of Metropolis steps: canonical swaps (blue), table exchanges (red), and single exchanges (green). Metropolis steps were applied to 10 snapshot LNMTOF configurations drawn from the TE simulation at $T = 1600$ K, $w = 40\%$. In canonical swaps, $\Delta\hat{H} = \Delta E$. In table exchanges, $\Delta\hat{H} = \Delta(E - \mu N)$. In single-species exchanges (SCB), $\Delta\hat{H} = \Delta(E - \mu N + \lambda k_B T C^2)$, where $T = 1600$ K and $\lambda = 1.0$.

LMZOF, $U \leq 3$). However, when the exchange table involves exchanges of too many sites, the sampling efficiency of the TE method can be limited.

To illustrate this point, we further studied a disordered rocksalt-like system in the chemical space of $x \cdot \text{LiF} + (1-x) \cdot \text{LiNi}^{2+}_{1/3}\text{Mn}^{3+}_{1/3}\text{Ti}^{4+}_{1/3}\text{O}_2$ ($0 \leq x \leq 1$, referred as LNMTOF). The system consists of Li^+, Ni^{2+}, Mn^{3+}, Ti^{4+} on the cation sub-lattice, and O^{2-}, F^- on the anion sub-lattice, with an additional requirement that $n_{\text{Ni}^{2+}} = n_{\text{Mn}^{3+}} = n_{\text{Ti}^{4+}}$. To move the composition by adding/removing fluorine with charge-neutrality, the minimal basis exchange table in LNMTOF contains the following exchanges ($U = 9$):

$$6\ \text{LiF} \Longleftrightarrow \text{Li}_3\text{NiMnTiO}_6, \tag{15}$$

which is equivalent to the reaction:

$$3\text{Li}^+ + 6\text{F}^- \Longleftrightarrow \text{Ni}^{2+} + \text{Mn}^{3+} + \text{Ti}^{4+} + 6\text{O}^{2-}. \tag{16}$$

Fig. 6(a) and (b) shows the trajectories of energy subtracted by chemical potentials ($E - \mu N$) and LiF atomic percentage (x_{LiF}) simulated at $T = 1600$ K, using the TE method with varied w and the SCB method with $\lambda = 1.0$ (see details in Supplementary Information). Regardless of the value of w, all TE simulations are unable to reach the ground-state LiF as suggested by the SCB method. The transfers between compositions are nearly prohibited in TE, suggesting a very low acceptance ratio of table exchanges. Ten random configurations were drawn as snapshots from the Markov chain generated by the TE simulation at $w = 40\%$, to which three types of Metropolis steps (canonical swaps, table exchanges, single exchanges) were applied to calculate the effective perturbation energies ($\Delta\hat{H}$). The distributions of $\Delta\hat{H}$ with each type of Metropolis steps are shown in Fig. 7. The table exchanges (red) in LNMTOF show significantly higher perturbation energy compared to the canonical swaps (blue) and single exchanges in the SCB method (green). This is because many sites are required to exchange simultaneously. The large energy perturbation in TE prohibits effective transfer between different compositions and explains the slow thermal equilibration in the TE method. Therefore, we suggest using the SCB method instead of the TE method for the acceptable efficiency of thermal equilibration when large-sized exchanges are included.

4. Discussion & summary

We have introduced two general methods to perform semigrand-canonical CE-MC simulations with a charge-balance constraint and CE models of arbitrary complexity, enabling thermodynamic calculations for ionic materials with configurational disorder.

The effect of the fraction of canonical exchanges w mixed into the semigrand canonical trajectory, and the charge penalty factor λ in the TE and SCB methods are presented. In the TE method, using a proper w is essential to efficiently explore and equilibrate among same-composition configurations. In the SCB method, the penalty factor λ controls the trade-off between the ability to revisit charge balance and the ability to leave charge balance to explore new states. We show that the hyper-parameters w and λ can be optimized to improve computational efficiency.

In general, the SCB method is mathematically more straightforward, easier to implement, and less vulnerable to thermal equilibration problems at a large table exchange size (U) compared to the TE method. However, when the table exchange size is small, the TE method is a useful method due to its simplicity as the optimization of chemical potentials (Eq. (13)) is not necessary, and its competitive sampling efficiency to the SCB method. We recommend the following strategy to apply TE and SCB in practical CE-MC calculations:

1. Choose the proper method according to the size of table exchanges (based on the exchange size U). When the size of table exchange is large (for example, $U > 4$), TE should be used cautiously as it may lead to low sampling efficiency and slow thermal equilibration.
2. Set the chemical potentials as discussed in Section 2.4.
3. Scan a series of w or λ coarsely to benchmark the computational efficiency. For example, a series of $w = 90\%, 70\%, 50\%, 30\%$, and 10%; or a series of $\lambda = 0.1, 0.2, 0.5, 1.0$ and 2.0 may be sufficient.
4. Perform short trial simulations with each w or λ at the temperature and chemical potentials of interest. Record the trace of properties along with the CPU time elapsed. By inspecting the convergence of $E-\mu n$ and compositions, hyper-parameter values that result in slow thermal equilibration can be ruled out. Search among the remaining values of w or λ, in order to maximize the computational efficiency (eff_t).
5. Continue the simulation with the optimal hyper-parameter value and generate the formal MC samples.

CRediT authorship contribution statement

Fengyu Xie: Conceptualization, Methodology, Software, Investigation, Writing – original draft. **Peichen Zhong:** Validation, Resources, Reviewing and editing. **Luis Barroso-Luque:** Software, Resources, Reviewing and editing. **Bin Ouyang:** Validation, Resources, Reviewing and editing. **Gerbrand Ceder:** Supervision, Reviewing and editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This work was funded by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, Materials Sciences and Engineering Division under Contract No. DE-AC02-05-CH11231 (Materials Project program KC23MP). The work was also supported by the computational resources provided by the Extreme Science and Engineering Discovery Environment (XSEDE), supported by National Science Foundation grant number ACI1053575; the National Energy Research Scientific Computing Center (NERSC), and the Lawrencium computational cluster resource provided by the IT Division at the Lawrence Berkeley National Laboratory.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.commatsci.2022.112000.

References

[1] D. De Fontaine, Configurational thermodynamics of solid solutions, Solid State Phys. 34 (1979) 73–274.
[2] J.M. Sanchez, F. Ducastelle, D. Gratias, Generalized cluster description of multicomponent systems, Phys. A 128 (1–2) (1984) 334–350.
[3] D. De Fontaine, Cluster approach to order-disorder transformations in alloys, Solid State Phys. 47 (1994) 33–176.
[4] G. Ceder, A derivation of the Ising model for the computation of phase diagrams, Comput. Mater. Sci. 1 (2) (1993) 144–150.
[5] A. Kohan, P. Tepesch, G. Ceder, C. Wolverton, Computation of alloy phase diagrams at low temperatures, Comput. Mater. Sci. 9 (3–4) (1998) 389–396.
[6] A. van de Walle, G. Ceder, Automating first-principles phase diagram calculations, J. Phase Equilib. 23 (4) (2002) 348–359.
[7] G. Ghosh, A. Van de Walle, M. Asta, First-principles calculations of the structural and thermodynamic properties of bcc, fcc and hcp solid solutions in the Al–TM (TM=Ti, Zr and Hf) systems: a comparison of cluster expansion and supercell methods, Acta Mater. 56 (13) (2008) 3202–3221.
[8] C. Ravi, B. Panigrahi, M. Valsakumar, A. van de Walle, First-principles calculation of phase equilibrium of V-Nb, V-Ta, and Nb-Ta alloys, Phys. Rev. B 85 (5) (2012) 054202.
[9] P.D. Tepesch, A.F. Kohan, G.D. Garbulsky, G. Ceder, C. Coley, H.T. Stokes, L.L. Boyer, M.J. Mehl, B.P. Burton, K. Cho, et al., A model to compute phase diagrams in oxides with empirical or first-principles energy methods and application to the solubility limits in the CaO–MgO system, J. Am. Ceram. Soc. 79 (8) (1996) 2033–2040.
[10] F. Zhou, T. Maxisch, G. Ceder, Configurational electronic entropy and the phase diagram of mixed-valence oxides: the case of Li x FePO 4, Phys. Rev. Lett. 97 (15) (2006) 155704.
[11] C. Ravi, H. Sahu, M. Valsakumar, A. van de Walle, Cluster expansion Monte Carlo study of phase stability of vanadium nitrides, Phys. Rev. B 81 (10) (2010) 104111.
[12] W.D. Richards, S.T. Dacek, D.A. Kitchaev, G. Ceder, Fluorination of lithium-excess transition metal oxide cathode materials, Adv. Energy Mater. 8 (5) (2018) 1701533.
[13] C. Wolverton, A. Zunger, Z.-W. Lu, Long-versus short-range order in Ni 3 V and pd 3 V alloys, Phys. Rev. B 49 (22) (1994) 16058.
[14] C. Wolverton, V. Ozolins, A. Zunger, Short-range-order types in binary alloys: a reflection of coherent phase stability, J. Phys.: Condens. Matter 12 (12) (2000) 2749.
[15] A. Seko, K. Yuge, F. Oba, A. Kuwabara, I. Tanaka, T. Yamamoto, First-principles study of cation disordering in $MgAl_2O_4$ spinel with cluster expansion and Monte Carlo simulation, Phys. Rev. B 73 (2006) 094116.
[16] H. Ji, A. Urban, D.A. Kitchaev, D.-H. Kwon, N. Artrith, C. Ophus, W. Huang, Z. Cai, T. Shi, J.C. Kim, et al., Hidden structural and chemical order controls lithium transport in cation-disordered oxides for rechargeable batteries, Nature Commun. 10 (1) (2019) 1–9.
[17] S. Müller, A. Zunger, First-principles predictions of yet-unobserved ordered structures in the Ag-Pd phase diagram, Phys. Rev. Lett. 87 (16) (2001) 165502.
[18] A. Seko, K. Yuge, F. Oba, A. Kuwabara, I. Tanaka, Prediction of ground-state structures and order-disorder phase transitions in II-III spinel oxides: A combined cluster-expansion method and first-principles study, Phys. Rev. B 73 (18) (2006) 184117.
[19] S.V. Barabash, V. Blum, S. Müller, A. Zunger, Prediction of unusual stable ordered structures of Au-Pd alloys via a first-principles cluster expansion, Phys. Rev. B 74 (3) (2006) 035108.
[20] A. Seko, K. Shitara, I. Tanaka, Efficient determination of alloy ground-state structures, Phys. Rev. B 90 (2014) 174104.
[21] W. Huang, D.A. Kitchaev, S.T. Dacek, Z. Rong, A. Urban, S. Cao, C. Luo, G. Ceder, Finding and proving the exact ground state of a generalized Ising model by convex optimization and MAX-SAT, Phys. Rev. B 94 (13) (2016) 134424.
[22] P.M. Larsen, A.R. Kalidindi, S. Schmidt, C.A. Schuh, Alloy design as an inverse problem of cluster expansion models, Acta Mater. 139 (2017) 254–260.
[23] M. Aydinol, A. Kohan, G. Ceder, K. Cho, J. Joannopoulos, Ab initio study of lithium intercalation in metal oxides and metal dichalcogenides, Phys. Rev. B 56 (3) (1997) 1354.
[24] C. Wolverton, A. Zunger, First-principles prediction of vacancy order-disorder and intercalation battery voltages in Li x CoO 2, Phys. Rev. Lett. 81 (3) (1998) 606.
[25] A. Van der Ven, M.K. Aydinol, G. Ceder, First-principles evidence for stage ordering in Li x CoO2, J. Electrochem. Soc. 145 (6) (1998) 2149.
[26] M.A. y de Dompablo, A. Van der Ven, G. Ceder, First-principles calculations of lithium ordering and phase stability on Li x NiO 2, Phys. Rev. B 66 (6) (2002) 064112.
[27] R. Malik, F. Zhou, G. Ceder, et al., Phase diagram and electrochemical properties of mixed olivines from first-principles calculations, Phys. Rev. B 79 (21) (2009) 214201.
[28] D.B. Laks, L. Ferreira, S. Froyen, A. Zunger, Efficient cluster expansion for substitutional systems, Phys. Rev. B 46 (19) (1992) 12587.
[29] A. Seko, Y. Koyama, I. Tanaka, Cluster expansion method for multicomponent systems based on optimal selection of structures for density-functional theory calculations, Phys. Rev. B 80 (2009) 165122.
[30] E. Cockayne, A. van de Walle, Building effective models from sparse but precise data: Application to an alloy cluster expansion model, Phys. Rev. B 81 (1) (2010) 012104.
[31] L.J. Nelson, G.L. Hart, F. Zhou, V. Ozoliņš, et al., Compressive sensing as a paradigm for building physics models, Phys. Rev. B 87 (3) (2013) 035125.
[32] A. Seko, I. Tanaka, Cluster expansion of multicomponent ionic systems with controlled accuracy: importance of long-range interactions in heterovalent ionic systems, J. Phys.: Condens. Matter 26 (11) (2014) 115403.
[33] Z. Leong, T.L. Tan, Robust cluster expansion of multicomponent systems using structured sparsity, Phys. Rev. B 100 (13) (2019) 134108.

[34] L. Barroso-Luque, J.H. Yang, G. Ceder, Sparse expansions of multicomponent oxide configuration energy using coherency and redundancy, Phys. Rev. B 104 (2021) 224203, http://dx.doi.org/10.1103/PhysRevB.104.224203.
[35] J.H. Yang, T. Chen, L. Barroso-Luque, Z. Jadidi, G. Ceder, Approaches for handling high-dimensional cluster expansions of ionic systems, Npj Comput. Mater. 8 (1) (2022) 1–11.
[36] P. Zhong, T. Chen, L. Barroso-Luque, F. Xie, G. Ceder, An l0l2-norm regularized regression model for construction of robust cluster expansions in multicomponent systems, Phys. Rev. B 106 (2) (2022) 024203, http://dx.doi.org/10.1103/PhysRevB.106.024203.
[37] K. Binder, D.W. Heermann, Monte Carlo Simulation in Statistical Physics, Springer, New York, 1988.
[38] A.v. van de Walle, M. Asta, Self-driven lattice-model Monte Carlo simulations of alloy thermodynamic properties and phase diagrams, Modelling Simul. Mater. Sci. Eng. 10 (5) (2002) 521.
[39] A. Van De Walle, M. Asta, G. Ceder, The alloy theoretic automated toolkit: A user guide, CALPHAD 26 (4) (2002) 539–553.
[40] A. van de Walle, Methods for first-principles alloy thermodynamics, Jom 65 (11) (2013) 1523–1532.
[41] Q. Wu, B. He, T. Song, J. Gao, S. Shi, Cluster expansion method and its application in computational materials science, Comput. Mater. Sci. 125 (2016) 243–254.
[42] J. Sanchez, Foundations and practical implementations of the cluster expansion, J. Phase Equilib. Diffus. 38 (3) (2017) 238–251.
[43] S. Kadkhodaei, J.A. Muñoz, Cluster expansion of alloy theory: A review of historical development and modern innovations, JOM 73 (11) (2021) 3326–3346.
[44] J.-Z. Xie, X.-Y. Zhou, H. Jiang, Perspective on optimal strategies of building cluster expansion models for configurationally disordered materials, J. Chem. Phys. (2022).
[45] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, E. Teller, Equation of state calculations by fast computing machines, J. Chem. Phys. 21 (6) (1953) 1087–1092.
[46] W.K. Hastings, Monte Carlo sampling methods using Markov chains and their applications, Biometrika 57 (1) (1970) 97–109.
[47] J.P. Valleau, L.K. Cohen, D.N. Card, Primitive model electrolytes. II. The symmetrical electrolyte, J. Chem. Phys. 72 (11) (1980) 5942–5954.
[48] M.P. Allen, D.J. Tildesley, Computer Simulation of Liquids, Oxford, New York, 1987.
[49] D. Frenkel, B. Smit, Understanding Molecular Simulations, Academic, San Diego, 1996.
[50] H.W. Hatch, S.W. Hall, J.R. Errington, V.K. Shen, Improving the efficiency of Monte Carlo simulations of ions using expanded grand canonical ensembles, J. Chem. Phys. 151 (14) (2019) 144109.
[51] Z. Deng, G. Sai Gautam, S.K. Kolli, J.-N. Chotard, A.K. Cheetham, C. Masquelier, P. Canepa, Phase behavior in rhombohedral NaSiCON electrolytes and electrodes, Chem. Mater. 32 (18) (2020) 7908–7920.
[52] A.J. Robinson, A. Voronkov, Handbook of Automated Reasoning, Vol. 1, Elsevier, 2001.
[53] A. Gelman, G.O. Roberts, W.R. Gilks, et al., Efficient Metropolis jumping rules, Bayesian Stat. 5 (599–608) (1996) 42.
[54] Z. Lun, B. Ouyang, D.-H. Kwon, Y. Ha, E.E. Foley, T.-Y. Huang, Z. Cai, H. Kim, M. Balasubramanian, Y. Sun, et al., Cation-disordered rocksalt-type high-entropy cathodes for Li-ion batteries, Nature Mater. 20 (2) (2021) 214–221.
[55] K. McColl, R.A. House, G.J. Rees, A.G. Squires, S.W. Coles, P.G. Bruce, B.J. Morgan, M.S. Islam, Transition metal migration and O2 formation underpin voltage hysteresis in oxygen-redox disordered rocksalt cathodes, Nature Commun. 13 (1) (2022) 5275, http://dx.doi.org/10.1038/s41467-022-32983-w.
[56] X. Guo, C. Chen, S.P. Ong, The intercalation chemistry of the disordered RockSalt Li3V2O5 anode from cluster expansions and machine learning interatomic potentials, 2022, arXiv preprint arXiv:2208.14420.