Modeling Intercalation Chemistry with Multiredox Reactions by Sparse Lattice Models in Disordered Rocksalt Cathodes

Peichen Zhong^(D),^{1,2,†} Fengyu Xie^(D),^{1,2,†} Luis Barroso-Luque^(D),^{1,2} Liliang Huang^(D),² and Gerbrand Ceder^{1,2,*}

¹Department of Materials Science and Engineering, University of California, Berkeley, California 94720, United States

² Materials Sciences Division, Lawrence Berkeley National Laboratory, California 94720, United States

(Received 10 July 2023; revised 16 August 2023; accepted 30 August 2023; published 9 October 2023)

Modern battery materials can contain many elements with substantial site disorder, and their configurational state has been shown to be critical for their performance. The intercalation voltage profile is a critical parameter to evaluate the performance of energy storage. The application of commonly used cluster expansion techniques to model the intercalation thermodynamics of such systems *ab initio* is challenged by the combinatorial increase in configurational degrees of freedom as the number of species grows. Such challenges necessitate the efficient generation of lattice models without overfitting and proper sampling of the configurational space under the requirement of charge balance in ionic systems. In this work, we introduce a combined approach that addresses these challenges by (1) constructing a robust cluster expansion Hamiltonian using the sparse regression technique, including $\ell_0 \ell_2$ -norm regularization and structural hierarchy; and (2) implementing semigrand-canonical Monte Carlo to sample charge-balanced ionic configurations using the table-exchange method and an ensemble average approach. These techniques are applied to a disordered rocksalt oxyfluoride Li_{1,3-x}Mn_{0.4}Nb_{0.3}O_{1.6}F_{0.4} (LMNOF) that is part of a family of promising earth-abundant cathode materials. The simulated voltage profile is found to be in good agreement with experimental data and particularly provides a clear demonstration of the Mn and oxygen contributions to the redox potential as a function of Li content.

DOI: 10.1103/PRXEnergy.2.043005

I. INTRODUCTION

The market for electric-vehicle high-energy-density Liion batteries has witnessed tremendous growth over the past decade [1,2]. The increasing demand for electrical energy storage requires further development of highenergy-density-based cathode materials in rechargeable Li-ion batteries. Current cathode materials are mostly limited to layered Li(Ni,Mn,Co)O₂ (NMC) variants. The high cost of Ni and Co limits the large-scale expansion of Liion batteries with NMC-type cathodes. Recently developed disordered rocksalt with Li-excess (DRX) are promising earth-abundant cathode materials, which can enable scaling of Li-ion energy storage to several TWh/yr production. Although these compounds have no long-range cation order, the interactions between species generate

*gceder@berkeley.edu

short-range ordering (SRO), which critically affects the electrochemical performance [3]. The broad chemical flexibility and the wide variety of chemical environments that can be created for Li and the transition metals (TMs) that can be used by SRO and chemistry provide new opportunities to improve the cathode performance. Examples include improving cyclability via fluorine-vacancy doping of the anion sublattice [4,5], enhancing the rate capability by engineering the cation SRO [3], and achieving zero-strain cathodes for solid-state batteries [6–8].

Over the last two decades, many properties of Li-ion battery materials have been successfully predicted by firstprinciples calculations [9-11]. The equilibrium voltage is one of the fundamental quantities that help characterize the electrochemical performance of a particular material and is defined by the difference in Li chemical potentials between the cathode and anode:

$$V = -\frac{\mu_{\rm Li}^{\rm cathode} - \mu_{\rm Li}^{\rm anode}}{zF}.$$
 (1)

In Eq. (1), z is the charge transferred per ion, F is Faraday's constant, and μ_{Li} is the chemical potential of Li. For

[†]These two authors contributed equally.

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

example, considering a Li transition-metal oxide $\text{Li}_x \text{TMO}_2$ ($x_1 \le x \le x_2$) as the cathode and Li metal as the anode with the cell reaction

$$Li_{x_1}TMO_2 \rightarrow Li_{x_2}TMO_2 + (x_1 - x_2)Li,$$
 (2)

the approximated equilibrium voltage can be computed as [12]

$$\bar{V}(x_1, x_2) \approx -\frac{E_{\text{Li}_{x_1}\text{TMO}_2} - E_{\text{Li}_{x_2}\text{TMO}_2} - (x_1 - x_2)E_{\text{Li}}}{F(x_1 - x_2)}.$$
 (3)

The internal energy of the bcc Li metal E_{Li} , the lithiated structure $E_{\text{Li}_{x1}\text{TMO}_2}$, and the delithiated structure $E_{\text{Li}_{x2}\text{TMO}_2}$ can be obtained from first-principles density functional theory (DFT). In this approach, the entropic effect is assumed to be small at low temperatures, and the change in internal energy is used to approximate the chemical potential change. By computing the formation energy of Li_xTMO_2 structures with varied Li contents x, a convex hull can be constructed from the energy of ground-state structures at each content. A piecewise voltage profile can then be built from Eq. (3) via the ground states on the convex hull by using relevant constructive values of x on the hull.

Unlike simple Li transition-metal oxides (LiTMO₂), a typical DRX cathode $(Li_{1+x}M'_{a}M''_{b}O_{2-v}F_{v})$ is composed of three major components: (1) the redox-active species M', which provides electron redox; (2) the inert high-valent TM M", which charge compensates for the Li excess and stabilizes disordered structures [13]; (3) fluorine, which can improve the cyclability and allows more Li excess to be accommodated without losing TM redox by lowering the overall anion valence [14]. When the cathode is charged (discharged), Li is removed (inserted) into the cathode structure accompanied by oxidation (reduction), resulting in various oxidation states among the redox-active metal M' and oxygen atoms. Different oxidation states of a TM can exhibit very different local chemistry preferences (e.g., Mn³⁺ has a substantial Jahn-Teller effect compared to $Mn^{2+/4+}$). To capture these chemical differences in simulations, different valence states of the same elements must be treated as distinct species. This treatment is called *charge decoration*, which has been demonstrated to be essential in capturing the electronic entropy effect to construct the correct phase diagram in some compounds [15]. Charge decoration intrinsically increases the number of components and therefore the complexity of modeling the intercalation voltage profiles of DRXs.

A. Intractability of composition enumeration

To obtain the voltage profile of a DRX material, most previous studies have used the convex-hull construction

approach by finding the ground states (GSs), which we refer to as the GS algo (Fig. 1). In this approach, one tries to find the low-energy structures at varied Li contents x using a variety of algorithms [16,17]. These lowenergy configurations are calculated by DFT to construct the piecewise voltage profile following Eq. (3). The GS algo has shown reasonable predictions for voltages and redox mechanisms [16–21]. When a high number of components and valence states are present, the GS algo can become impractical as all possible valence combinations at each stage of delithiation must be enumerated. For example, when evaluating a delithiated supercell of composition $Li_{21}\square_{18}Mn_{12}Nb_9O_{48}F_{12}$, the combination of valence in Mn and O can take $Mn_6^{3+}Mn_6^{4+}$, $Mn_1^{2+}Mn_4^{3+}Mn_7^{4+}$, $Mn_2^{2+}Mn_2^{3+}Mn_8^{4+}$, and even $Mn_7^{3+}Mn_5^{4+}O_1^{-}$, etc. Enumerating all the possible compositions and searching for the possible ground states under each charge-decorated composition are NP-hard problems that become intractable, in particular when the supercell grows. To resolve the enumeration problem, Monte Carlo (MC) sampling is a better choice for studying configurational energetics in a high-dimensional space.

B. Fast growth of the cluster basis caused by charge decoration

To bridge the gap between 0-K ground states and sampling at finite temperatures, MC simulation with a cluster expansion (CE) as an effective Hamiltonian is typically used for intercalation chemistry in ordered cathode materials [22–25]. The CE casts the energy as a function of the occupancy of atoms on a set of predefined sites. For example, in a multicomponent system, the energy is expanded as

$$E(\boldsymbol{\sigma}) = \sum_{\beta} m_{\beta} J_{\beta} \left\langle \Phi_{\boldsymbol{\alpha} \in \beta} \right\rangle_{\beta} + \frac{E_0}{\varepsilon_r}, \qquad \Phi_{\boldsymbol{\alpha}} = \prod_{i=1}^{N} \phi_{\alpha_i}(\sigma_i).$$
(4)

A configuration σ represents a specific occupancy state of species on all the system sites, where σ_i describes which species sits on the *i*-th site of the lattice. The cluster basis functions $\Phi_{\alpha} = \prod_{i=1}^{N} \phi_{\alpha_i}(\sigma_i)$ are the product of site basis functions $\phi_{\alpha_i}(\sigma_i)$ across a collection α of multiple sites. They are taken as the average over the crystal symmetry orbits β , forming a complete basis to expand the scalar energy function on the configuration space. The expansion coefficients J_{β} are called effective cluster interactions (ECIs). The electrostatic energy (Ewald energy E_0/ε_r) is also included to capture long-range electrostatic interactions [26] $[E_0$ is the unscreened electrostatic energy, and $1/\varepsilon_r$ is fitted as one of the ECIs $(1/\varepsilon_r \ge 0)$]. We refer the reader to Ref. [27] for a more comprehensive description of the CE formalism in ionic systems.

Since the cluster site basis functions are defined by the number of components [28], the charge decoration can significantly increase the model complexity of the CE, resulting in a rapid growth in the number of clusters. Figure 2 illustrates how the number of cluster basis functions grows as the number of cation species included in the DRX increases, where the cutoff radius is fixed with pair, triplet, and quadruplet interactions up to 7, 4, and 4 Å. For the full description of Mn and O redox, more than 800 ECIs are predefined and need to be fitted, whereas the number of DFT calculations is typically much smaller than that of predefined ECIs. When building a CE, the ECIs $J^* = \{J_\beta\}$ can be obtained by fitting the DFT energy $E_{DFT,S}$ of training structures *S* to their correlation functions with

(regularized) linear regression:

$$\mathbf{J}^* = \arg\min_{\mathbf{J}} ||\mathbf{E}_{\text{DFT},S} - \mathbf{\Pi}_S \mathbf{J}||_2^2 + \rho(\mathbf{J})$$
(5)

with Π_S the feature matrix formed by the correlation functions and $\rho(\mathbf{J})$ a regularization term. High-component CE can easily be overfitted as the rank of the feature matrix (Π_S) is typically smaller than the dimension of ECIs dim(\mathbf{J}). The low-rank structure (referred to as *rank deficiency*) of the feature matrix Π_S in Eq. (5) requires selecting the most physically informative ECIs and avoiding overfitting in CE parameterization [29,30]. The rank deficiency can also arise from other physical



FIG. 1. An overview of reported methods for computing voltage profiles in (a) ordered and (b) disordered electrode materials. The green arrows represent the configurational samplings that can be accelerated by using cluster expansion as an effective Hamiltonian fitted from DFT calculations.



FIG. 2. An illustration of cluster basis growth: number of correlation functions versus number of components included in the CE with pair, triplet, and quadruplet cutoff radii of 7, 4, and 4 Å based on a rocksalt primitive cell with lattice parameter a = 3 Å.

constraints in ionic systems. For example, charge balance creates a linear dependency between the number of charge-decorated species and the corresponding correlation functions. Another challenge is that the training structures calculated with DFT are predominantly low in energy, and such low-energy structures often narrow the configurational space that can be represented in the training structures, in which, in principle, high-energy configurations could be included. DFT tends not to cover such configurations or relax them to lower-energy configurations by moving ions and electrons. The number of nonzero ECIs must be constrained $[||\mathbf{J}||_0 \leq \operatorname{rank}(\mathbf{\Pi}_S)]$ to prevent ECI fitting from being an underdetermined problem. This constraint can be achieved by properly introducing $\mu ||\mathbf{J}||_0$ as a regularization term in Eq. (5) to penalize the number of nonzero ECIs and impose sparsity [31].

C. Charge-neutrality constraint in MC sampling

After fitting the ECIs, MC simulations can be used to sample the energy of configurations under finite temperatures. Applying semigrand-canonical Monte Carlo (SGCMC) sampling is most suited for calculating voltage profiles [32,33]. The relation between the Li concentration x_{Li} and a Li chemical potential can be obtained by applying the Metropolis-Hastings algorithm with the Boltzmann distribution

$$f(E(\boldsymbol{\sigma}),\mu) \propto \exp\left(-\frac{E(\boldsymbol{\sigma})-\mu_{\mathrm{Li}}x_{\mathrm{Li}}N}{k_BT}\right).$$
 (6)

In Eq. (6), *E* is the energy of configuration σ given by the CE Hamiltonian, μ_{Li} is the Li chemical potential, x_{Li} is the Li concentration in configuration σ , *N* is the total number of Li and vacancy sites, k_B is the Boltzmann constant, and *T* is the simulation temperature. Computing voltage profiles using SGCMC has been successful in the study of several simple binary electrode materials, such as LiCoO₂/LiNiO₂ [33–35], MgTiS₂ [36], and Li₃V₂O₅ [37]. When charge-decorated CEs are used, the requirement of charge neutrality must be enforced in MC sampling. Because the training set only includes chargebalanced structures with no information about the chargeunbalanced structures, the energy predicted by the CE will be unphysical if configurations with nonzero net charge are assessed during SGCMC. Techniques for enforcing strict charge neutrality in SGCMC have been applied to several electrolyte systems [38,39]; however, few have been demonstrated in a system with complex redox reactions such as DRX [40].

To overcome all the abovementioned issues, we propose a voltage-calculation framework that combines several state-of-the-art methods in CE MC. With this framework, we demonstrate how to correctly model the intercalation voltage profile in DRX and, more generally, any complex ionic systems with redox-active ions and configurational disorder. In the following methodology sections, we introduce the construction of a robust and predictive cluster expansion Hamiltonian with $\ell_0\ell_2$ -norm regularized sparse regression [31], demonstrate an effective sampling strategy of the intercalation stages with SGCMC under charge balance using the table-exchange (TE) method [40], and illustrate an ensemble average method over representative structures to handle various chemical environments. In the results section, the equilibrium voltage profile of Li_{1,3-x}Mn_{0,4}Nb_{0,3}O_{1,6}F_{0,4} (LMNOF) is presented and compared with the experiments. To explain the redox mechanism, we analyze the proportion of multiple redoxactive species at varied Li contents. We find that the calculated voltage profile and redox mechanism agree well with experiments and argue that the ability of our method to describe oxygen redox in the Li-excess cathode accurately is particularly noticeable.

II. METHODOLOGY

A. Training structure generation for DFT

To describe DRX materials well, two configurational degrees of freedom need to be accurately represented: the Li-vacancy interactions and the different local chemical environments (i.e., the SRO of TMs and anions). We propose the following two-step procedure for generating the training set.

Pristine states. Generate several fully lithiated structures with different transition-metal and anion configurations in relatively small supercells (e.g., supercell structures with 10 times or 20 times the formula unit of Li_{1.3}Mn_{0.4}Nb_{0.3}O_{1.6}F_{0.4}). These structures can be generated from canonical MC samplings using a prefitted cluster expansion as the effective Hamiltonian or solely using the electrostatic energy for simplicity.



FIG. 3. (a) An illustration of the rank deficiency in the ECI fitting. The left empty bar represents the dimension of predefined ECIs with dim(\mathbf{J}) = 858. The middle empty bar indicates the number of structures sampled in the feature matrix with dim($\mathbf{\Pi}_S$) = (463 × 858). The blue shaded area overlapping the middle bar indicates a feature matrix of rank($\mathbf{\Pi}_S$) = 287. The right solid blue bar represents the number of nonzero ECIs after an $\ell_0\ell_2$ -norm regularized sparse regression, giving $||\mathbf{J}||_0 = 169$. (b) An illustration of the TEs used in charge-balanced SGCMC. (c),(d) The distribution of total on-site magnetizations of Mn (c) and O (d) atoms among all DFT- r^2 SCAN-calculated structures in the chemical space of Li_{1.3-x}Mn_{0.4}Nb_{0.3}O_{1.6}F_{0.4}. The valence of each Mn and O atom is determined by the on-site magnetization. From the histogram, the classification boundary between Mn^{4+/3+} and Mn^{3+/2+} is estimated to be $3.25\mu_B$ and $4.1\mu_B$, and the O⁻ classification is estimated to be $0.5\mu_B$. [As the percentage of O²⁻ with a low magnetization is too large compared to the percentage of O⁻, panel (d) is truncated on the *y* axis.]

(2) *Delithiated states*. Starting from the structures generated in step (1), fix the TM and O/F orderings and enumerate different Li-vacancy configurations at varied Li contents (e.g., x = 0.3, 0.5, 0.7 in Li_{1.3-x}). As the total number of enumerated structures can be large, one can further sort the structures at each Li content by their electrostatic energy and only keep the low-energy ones.

All the sampled structures are in the chemical space of $Li_{1,3-x}Mn_{0.4}Nb_{0.3}O_{1.6}F_{0.4}$, and this two-step procedure covers different Li-vacancy orderings in varied local chemical environments formed by TM and anion SRO to be calculated with DFT.

B. Sparse regression for charge-decorated CE

To obtain effective valence states of redox-active Mn and O species from a DFT-calculated configuration, the on-site magnetization can be used [27,41]. For example,

Figs. 3(c) and 3(d) show the distribution of magnetic moments representing $Mn^{2+/3+/4+}$ and $O^{2-/-}$ in our set of 463 DFT-calculated structures. The valence of each Mn and O atom is classified by the site magnetization using $3.25\mu_B$ for distinguishing $Mn^{4+/3+}$, $4.1\mu_B$ for separating $Mn^{3+/2+}$, and $0.5\mu_B$ to indicate O⁻ [42] (μ_B is the Bohr magneton). For the sparse regression of ECIs, we apply the $\ell_0\ell_2$ -norm regularization with hierarchy constraints [31]. The ECIs are optimized in the following mixed-integer quadratic programming problem:

$$\min_{\mathbf{J}} \mathbf{J}^{T} \mathbf{\Pi}_{S}^{T} \mathbf{\Pi}_{S} \mathbf{J}^{T} - 2\mathbf{E}_{\text{DFT}}^{T} \mathbf{\Pi}_{S} \mathbf{J} + \mu_{0} \sum_{c \in \mathbf{C}} z_{0,c} + \mu_{2} ||\mathbf{J}||_{2}^{2}$$
s.t. $Mz_{0,c} \geq J_{c}, \forall c \in \mathbf{C},$
 $Mz_{0,c} \geq -J_{c}, \forall c \in \mathbf{C},$
 $z_{0,b} \leq z_{0,a}, \forall a \subset b, \{a,b\} \in \mathbf{C},$
 $z_{0,c} \in \{0,1\}, \forall c \in \mathbf{C}.$ (7)

Here Π_S is the feature matrix, **J** are the ECIs, and $z_{0,c}$ is the slack variable representing $J_c = 0, z_{0,c} = 0$, and $J_c \neq 0, z_{0,c} = 1$. We set M = 100 to constrain the optimization boundaries, and $||\mathbf{J}||_2^2 = \mathbf{J}^T \mathbf{J}$ is a ridge regression term (ℓ_2 norm). We refer the reader to Ref. [31] for a detailed description of this approach, which we found to lead to a relatively sparse but accurate CE.

The CE Hamiltonian was constructed with pair interactions up to 7 Å, triplet interactions up to 4 Å, and quadruplet interactions up to 4 Å based on a rocksalt primitive cell with lattice parameter a = 3 Å leading to a possible 858 ECIs (including the constant term J_0). The ECIs were fitted using 463 training structures, forming a feature matrix of rank(Π_S) = 287. The resulting ECIs using the sparse regression in Eq. (7) contain 169 nonzero elements ($||\mathbf{J}||_0 = 169$). The relationship between dimension, rank, and the number of nonzero elements is illustrated in Fig. 3(a).

C. Charge-balanced Monte Carlo sampling

We applied SGCMC simulations on the Li-vacancy occupancy and the charge decoration degrees of freedom to obtain the voltage curve of the $\text{Li}_{1.3-x}\text{Mn}_{0.4}\text{Nb}_{0.3}\text{O}_{1.6}\text{F}_{0.4}$ composition. The delithiation is achieved by stepwise removal of Li atoms. In each MC step, the Li⁺ is removed (inserted), accompanied by the oxidation (reduction) of an Mn or O atom. The net charge of each configuration is maintained at zero by only executing a combination of site occupancy changes that are charge neutral. This type of MC step is referred to as table exchange (TE) [38,40]. In our calculations, we used the following three elemental classes of perturbations:

1.
$$\text{Li}^{+} + \text{Mn}^{2+} \rightarrow \text{Mn}^{3+} + \text{Vac},$$

2. $\text{Li}^{+} + \text{Mn}^{3+} \rightarrow \text{Mn}^{4+} + \text{Vac},$
3. $\text{Li}^{+} + \text{O}^{2-} \rightarrow \text{O}^{-} + \text{Vac}.$

Any other charge-conserving MC step can be expressed as a linear combination of these three classes and their inverses. For example, charge transfer such as $2Mn^{3+} \rightarrow Mn^{2+} + Mn^{4+}$ and $Mn^{4+} + O^{2-} \rightarrow Mn^{3+} + O^{-}$ can be achieved by a combination of elementary perturbations. As the intercalation is assumed to be topotactic, Mn, O, and F ions do not change sites. The acceptance probabilities of each MC step are scaled to ensure detailed balance (see Appendix B).

In a series of SGCMC simulations, the chemical potentials are scanned between two limiting values $\mu_{\text{Li}} \in [\mu_{\min}, \mu_{\max}]$ at finite temperature. For each SGCMC simulation with a given Li chemical potential μ , the concentration of each charge-decorated species is averaged over the sampled MC structures from the equilibrium. The Li chemical potential can be converted into the cathode voltage using $V = -(\mu - E_{\text{Li}})/e$. In this work, the numerical simulations were implemented with the open-source software package smol [43], and the ECI optimization was solved using the gurobi package [44].

D. The ensemble average method

Unlike ordered intercalation compounds that are characterized by a small number of local environments, the configurational disorder in DRX creates an abundance of local chemical environments, leading to a significant variation of Li extraction energy from different sites [45,46]. This multitude of environments is difficult to capture in a single small unit cell. Instead, we sample the delithiation from multiple distinct structures and obtain the true lithiation curve as the ensemble average of them.

We use a canonical MC simulation to generate an ensemble of fully lithiated (i.e., pristine) configurations $\{\sigma_i\}$, from which multiple representative structures are drawn (Fig. 4). The CE model used to generate the canonical MC structures is reported in Ref. [47]. To recover the actual SRO in DRX, we require that the average of physical quantities in the collection of representative structures be approximately equal to the actual ensemble average, assuming that infinitely many structures are drawn from the ensemble $(\langle \Pi \rangle_{\text{ens}} \approx \langle \Pi \rangle_{\infty} = \tilde{\Pi}_{\text{SRO}})$. For simplicity, in this work, we verify that the average energy $\langle E \rangle_{\text{ens}} = \langle \Pi \rangle_{\text{ens}} \cdot \mathbf{J} \approx \tilde{\Pi}_{\text{SRO}} \cdot \mathbf{J}$ converges to the ensemble average with an increasing number of selected structures.

Each simulated fully lithiated structure contains 120 atoms ($Li_{39}Mn_{12}Nb_9O_{48}F_{12}$) with a supercell lattice constant of about 10 Å. We rationalize the choice of such a supercell size because about 10 Å is a good cutoff to maintain enough distance between each atom and its periodic images and to encapsulate all cluster-interaction distances. Subsequently, SGCMC simulations are performed for every structure in the ensemble, and, for a given chemical potential, the species (e.g., Li) concentration is computed using the average

$$\langle x_{\rm Li}^{\mu} \rangle = \frac{1}{N_{\rm ens}} (\bar{x}_1 + \bar{x}_2 + \cdots),$$
 (8)

where $\{\bar{x}_1, \bar{x}_2, ...\}$ are the averaged Li concentrations given by the thermally equilibrated SGCMC sample starting from each structure. In this work, we used 30 disordered structures for the ensemble average.

III. SIMULATION RESULTS

We demonstrate the procedures presented above to model the intercalation thermodynamics of $Li_{1.3-x}Mn_{0.4}$ Nb_{0.3}O_{1.6}F_{0.4} (LMNOF). Figure 5(a) presents the simulated and the experimental voltage profiles [47]. The solid green line is the experimental charging profile under a low current density (20 mA/g) between 1.5 and 5.0



FIG. 4. An illustration of charge-balanced SGCMC on an ensemble of disordered structures (green, Li; white, vacancy; other colors, different TM species; anions are not displayed). An ensemble of fully lithiated disordered structures is first generated and referred to as the pristine state. The SGCMC is performed to topotactically delithiate each pristine structure at decreasing Li chemical potentials μ_{Li} (i.e., increasing voltages *V*). The concentration of the species is averaged over each SGCMC sampling and all the pristine structures.

V. The bottom of the discharge state (1.5 V) is aligned to be the fully lithiated state $(Li_{1,3}Mn_{0,4}Nb_{0,3}O_{1,6}F_{0,4})$. The green dots represent the computed voltage profile at T = 300 K. The slope and turning point in the slope are in good agreement with the experimental profile for 0.4 < x < 1.3, indicating that the Li-vacancy interaction and redox potentials of Mn and O are well described in our model. In the highly charged region (x < 0.7)specifically, the simulation shows remarkably good agreement with the experiment, reproducing the fact that the slope of the profile becomes flatter at $x \sim 0.7$ [marked by the gray dashed line in Fig. 5(a)]. The computed voltage profile is systematically lower than the experimental one, which is well known for most DFT functionals that are not augmented with a Hubbard U correction [48]. Even though the strongly constrained and appropriately normed (SCAN) functional removes more self-interaction than previous local density approximation (LDA) and generalized gradient approximation (GGA) [49], it does not fully remove it and underestimates the intercalation voltage. Recent studies have further addressed this issue by applying a Hubbard U correction to the SCAN and r^2 SCAN functions, which is found to reduce the prediction error in the formation energy of transition metal oxides [50,51].

The fraction of each redox-active species during intercalation is presented in Fig. 5(b). Mn^{2+} , Mn^{3+} , and Mn^{4+} are represented by orange, green, and blue dots, respectively, and O⁻ is represented by red dots. Comparing Figs. 5(a) and 5(b) makes it apparent that the Li content where the voltage profile flattens ($x \sim 0.7$) corresponds to the start of the oxidation of O^{2-} to O^{-} . The lowering of the voltage slope as the system changes from TM redox to O redox is consistent with the higher dilution of the oxygen charge compared to TM-redox centers. Figure 5(b) also reveals several key points about the redox mechanism in LMNOF. The oxidation of Mn does not appear consecutively from Mn^{2+} to Mn^{3+} to Mn^{4+} . The amount of Mn^{3+} and Mn⁴⁺ increase simultaneously as Mn²⁺ begins to be oxidized. The fact that different oxidation states of Mn coexist over a wide range of Li contents is likely due to the variety of local chemical environments induced by cation disorder. The coexistence is consistent with the marginal stability of Mn³⁺ and its propensity to disproportionate into Mn² and Mn⁴⁺ when it cannot exist in an environment where it can lower its energy significantly through a Jahn-Teller distortion [52]. O redox occurs after all the Mn^{2+} have been consumed but before all the Mn atoms are fully oxidized to +4 valence. The hybridized redox mechanism between Mn and O has been confirmed by previous synchrotron characterization experiments [47] and on related materials Li_{1.2}Mn_{0.4}Ti_{0.4}O_{2.0} by mass spectroscopy [53]. Our lattice model does not consider the formation of O₂ dimers that would lower the voltage by lowering the energy of the charged state. Its absence in our model may artificially raise the computed voltage somewhat, thereby reducing the discrepancy between experiments and simulation.



FIG. 5. (a) Calculated (circle) and experimental (solid line) voltage profiles of $\text{Li}_{1,3-x}\text{Mn}_{0,4}\text{Nb}_{0,3}\text{O}_{1.6}\text{F}_{0,4}$. (b) Calculated content of Mn^{2+} , Mn^{3+} , Mn^{4+} , and O^- per formula unit (f.u.) as a function of Li content (*x*). (c) Comparison of voltage profiles generated with a charge-decorated CE (green dots) and a CE without charge decoration (red dots). The shaded area represents the distribution of the derivative quantity |dx/dV|, which reflects the Li-site energy distribution during the intercalation process.

IV. DISCUSSION

Typical modeling of the intercalation energetics only includes the Li-vacancy degree of freedom, assuming that the electronic degrees of freedom can be integrated out. Integrating out degrees of freedom is based on two key assumptions [54]. (1) One assumes that the degree of freedom that is variationally removed is always optimized in the DFT calculation. More specifically, this would require that, for a given Li-vacancy configuration, the DFT calculation can easily find the charge decoration with the lowest energy. While this is a reasonable assumption for systems with a highly delocalized charge such as Li_xCoO_2 [55], it is unlikely to be the case for disordered systems where the variation of local environments and the highly localized charge easily lead to charge metastability in DFT. (2) The second assumption when degrees of freedom are variationally removed is that their entropy contribution can be neglected, as their variational optimization is supposed to find their ground state. The role of electronic entropy in intercalation systems is yet to be fully understood, but at least in Li_xFePO_4 it has been shown to be critical to reproduce the correct intercalation behavior [15]. Both assumptions are unlikely to be valid in DRX cathodes, given their multitude of possible redox couples and large variations of energetic environments.

To illustrate the significance of accurately capturing the electronic degree of freedom in DRX materials, we fitted another CE parameterizing only the Li-vacancy interactions. Mn and O atoms are considered single chargeless species regardless of their oxidation states in such a CE. The computed voltage profiles with the charge-decorated CE (in green) and the undecorated CE (in red) are compared in Fig. 5(c). The undecorated CE yields a feature-less voltage profile and virtually no change of slope as a function of Li content. The mechanism-related details are poorly portrayed in the profile. The shaded areas in Fig. 5(c) are the relative derivative of capacity with respect to voltage (dx/dV). Two distinct peaks are observed in the charge-decorated calculation (green), indicating the



FIG. 6. (a) Average energy (per atom) among pristine structures as a function of the number of pristine structures selected in an ensemble. The red shaded region indicates a variation of $\pm 1 \text{ meV/atom of } \langle E \rangle_{\infty}$. The orange and green lines represent the results of the supercell structure with 120 and 960 atoms, respectively. (b) Comparison of computational efficiency [defined in Eq. (C1)] between supercells with 120 and 960 atoms, computed and averaged over all the pristine structures in the ensembles.

contributions from Mn redox and O redox, whereas they cannot be adequately distinguished in the undecorated version (red).

We also highlight the importance of using the ensemble average method to capture the multitude of environments in disordered systems. In principle, one can approach the representation of disordered systems by using a single structure in a very large supercell (e.g., special quasirandom structure approach [56]) or by making an ensemble of multiple smaller-sized supercell structures and taking an average of the computed quantities. These two approaches are statistically equivalent, given that enough structures have been used with the smaller-sized supercells. Figure 6(a) shows the number of the structures required for the average energy per atom over structures to converge within ± 1 meV/atom of the equilibrium average $\langle E \rangle_{\infty}$ when choosing from an ensemble of representative structures in canonical MC [47]. The green line represents the results in a supercell of 960 atoms, whereas the orange line represents the results in a supercell of 120 atoms. The smaller-sized supercell requires drawing about 10 times more structures to converge, with the number of structures required roughly scaling with the supercell size $(960 = 120 \times 8)$. A large supercell approaches the actual distribution of SRO with fewer structures. However, a much longer MC sampling time is required in a large supercell, whereas in a smaller supercell, one can easily benefit from parallelizing multiple pristine structures to require much less total time consumption. To illustrate this factor, we introduce computational efficiency as a quantitative benchmark for SGCMC. Figure 6(b) shows the computational efficiency in SGCMC at varied voltages in supercells with 120 (orange) and 960 (green) atoms. Higher efficiency indicates that less computational time is required (see the details about the definition in the Appendix C). Approximating the computational efficiency, the sampling time required in each pristine structure scales as about $O(N^1)$ with the supercell size. Therefore, with the help of parallelization, using an ensemble of smaller structures is statistically as accurate but practically more efficient than using fewer large structures. An ensemble of smaller structures is also more tractable in DFT for computing other properties when necessary, such as electronic structure, whereas DFT is computationally prohibitive in large supercells even when the SRO can be well presented.

Finally, we discuss the necessity of using enough structure samples when smaller-sized supercells are used to model disordered systems. Figure 7 shows the variance of redox-active species with respect to the averaged Li content. As a result of the disorder, the variance in the content of each species among pristine structures is not negligible. This further discourages the methods based on



FIG. 7. Contents of each redox species per formula unit as a function of Li content x in $\text{Li}_x \text{Mn}_{0.4} \text{Nb}_{0.3} \text{O}_{1.6} \text{F}_{0.4}$ as computed from SGCMC simulations at T = 300 K. The statistics are computed based on an ensemble of disordered structures in supercells of $\text{Li}_{30x} \text{Mn}_{12} \text{Nb}_9 \text{O}_{48} \text{F}_{12}$. The average values over different disordered structures are marked with colored dots, the median values are marked with orange lines, and the variances are marked with error bars around the medians.

convex-hull and direct DFT evaluations (e.g., the GS algo) to accurately determine voltage profiles for materials with significant disorder such as DRX. When the size of the supercell in which one enumerates possible Li-vacancy and charge configurations is too small or the number of structure samples is too few, as is often the case in the GS algo, the variance resulting from different local chemical environments cannot be adequately captured.

V. SUMMARY AND OUTLOOK

In DRX materials, the abundance of available redoxactive transition metals creates new opportunities for cathode design, such as using the redox reactions between Cr^{3+}, Cr^{4+} , and Cr^{6+} [57,58], V^{3+}, V^{4+} , and V^{5+} [59], or even Fe^{2+}, Fe^{3+} , and Fe^{4+} [60]. However, the high number of chemical components, distinguished by their valence states in DRX materials renders the "curse of dimensionality" the main simulation obstacle [61]. As the complexity of the energy model grows with the number of components and charge decoration, the computational cost grows exponentially fast. For example, the curse of dimensionality has been an essential impediment for the computational design of high-entropy cathodes [6,58].

In this work, we propose the following procedure to obtain accurate intercalation voltage profiles in DRX with multiredox reactions: (1) build a training set containing different fully lithiated structures and enumerate the Livacancy orderings at varied delithiation levels calculated by DFT; (2) construct a charge-decorated CE Hamiltonian and fit the ECIs using the sparse regression technique to address the fast growth of the cluster basis and rank-deficiency issues (e.g., $\ell_0 \ell_2$ -norm regularized regression); and (3) run SGCMC in an ensemble of disordered structures under charge balance to obtain physically rational energetics and compute the voltage profiles.

As demonstrated in this work, the workflow above provides an effective way to study the intercalation chemistry of oxides with Mn and O redox activity. Looking forward, the recent development of machine-learning force fields (MLFFs) may provide new opportunities to accelerate the training structure generation process by accurately approximating DFT calculations [62-64]. The CE Hamiltonian can be further obtained by coarse-graining the MLFFpredicted configurational energy [65]. Constructing a valid MLFF for this purpose requires training data in a specific system with high-fidelity chemical accuracy. We suggest starting with fine-tuning the pretrained universal interatomic potential as a practical approach, especially using the charge-informed MLFF to include the heterovalent states of transition metals [64]. We believe that such an approach has the potential to bridge first-principles calculations, force fields, and cluster expansions and give a higher accuracy sampling at a lower computational cost to study intercalation chemistries in energy storage materials.

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-AC0205CH11231 (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 No. ACI1053575; the National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility located at Lawrence Berkeley National Laboratory; and the Lawrencium Computational Cluster resource provided by the IT Division at the Lawrence Berkeley National Laboratory.

APPENDIX A: DFT CALCULATIONS

DFT calculations were performed with the Vienna *ab* initio simulation package using the projector-augmented wave method [66,67], a plane-wave basis set with an energy cutoff of 520 eV, and a reciprocal space discretization of 25 k points per Å⁻¹. All the calculations were converged to 10^{-6} eV in total energy for electronic loops and 0.02 eV/Å in interatomic forces for ionic loops. To model the Li-Mn-Nb-O-F oxyfluoride system, we relied on the regularized strongly constrained and appropriately normed meta-GGA exchange-correlation functional $(r^2$ SCAN) [49,68], which is believed to better capture the cation-anion hybridization and Li coordination preference [69]. r^2 SCAN provides better computational efficiency than the earlier version of SCAN [70]. It is noted that the choice of different functionals has an impact on the anionic redox activities [71-73]. As a proof of concept in this study, r^2 SCAN is selected for both computational efficiency and accuracy.

APPENDIX B: TABLE EXCHANGE IN SGCMC

Because the TE Metropolis steps are proposed by randomly selecting and exchanging species in a configuration, the forward and backward proposal probabilities for a TE step can be imbalanced. To ensure detailed balance in the simulation, the acceptance probability of a TE step is adjusted as

$$p_{\sigma\sigma'} = \min\{1, r\},$$

$$r = \frac{\prod_{u_s \neq 0} n_s!}{\prod_{u_s \neq 0} (n_s + u_s)!} \exp\left[-\frac{(\Delta E_{\sigma\sigma'} - \sum_s \mu_s u_s)}{k_B T}\right],$$
(B1)

where σ and σ' are the configurations before and after applying the TE step, respectively; n_s is the number of species s in configuration σ before the step; u_s represents the change in the concentration of species *s* after the step; and the vector value $\mathbf{u} = (u_1, u_2, \dots, u_s)$ must fall into one of the three aforementioned TE categories or their inverses. All TE categories and their inverses were selected with equal probability.

Additionally, in the TE SGCMC, a percentage (w = 20%) of canonical swaps between the Li-vacancy and different oxidation states of the same elements are mixed with TE steps to improve the efficiency of exploring different Li and electronic configurations within the same compositions. For each chemical potential in SGCMC, we applied simulated annealing from T = 5000 to 100 K to approach the ground-state configurations of the Li-vacancy and oxidation states. Then, starting from such a ground state, the SGCMC was run at T = 300 K with 500 000 steps to thermally equilibrate the system, and another 500 000 steps were used to generate a sample of configurations for analysis [40].

APPENDIX C: COMPUTATIONAL EFFICIENCY

At each Metropolis step p in semigrand-canonical MC, the Li concentration (x_{Li}) at the current configuration is recorded as $x_{Li,p}$. We denote $\overline{x_{Li,[p,q]}}$ as the average of x_{Li} in a block from simulation step p to step q. After thermal equilibration, we define the *block-mean variance* at block length L, $Var(\overline{x_{Li,L}})$, as the variance of the *block means* $\overline{x_{Li,[p,p+L]}}, \overline{x_{Li,[p+L,p+2L]}}, \dots$ for each block containing L samples. The block-mean variance can be used as a measure of uncertainty when estimating $\overline{\theta}$ using a block mean. With the block-mean variance, we further define the *computational efficiency*,

$$\operatorname{eff}_{t}(x_{\mathrm{Li}}) = \frac{\tau^{2}}{\overline{T}_{L}\operatorname{Var}(\overline{x_{\mathrm{Li},L}})},$$
(C1)

where τ^2 is the ensemble variance of Li concentration x_{Li} approximated by the variance of all MC steps after thermal equilibration, and \overline{T}_L is the average CPU time spent in each *L*-step block. Given the same set of hardware used in MC simulations, the higher the computational efficiency, the less sampling time required for the uncertainty of the average Li concentration (i.e., the block-mean variance) to be decreased below the same threshold. In brief, higher sampling efficiency means less sampling time.

- E. A. Olivetti, G. Ceder, G. G. Gaustad, and X. Fu, Lithium-ion battery supply chain considerations: Analysis of potential bottlenecks in critical metals, Joule 1, 229 (2017).
- [2] J. B. Goodenough and Y. Kim, Challenges for rechargeable Li batteries, Chem. Mater. **22**, 587 (2010).
- [3] H. Ji, A. Urban, D. A. Kitchaev, D.-H. Kwon, N. Artrith, C. Ophus, W. Huang, Z. Cai, T. Shi, J. C. Kim, H. Kim, and

G. Ceder, Hidden structural and chemical order controls lithium transport in cation-disordered oxides for rechargeable batteries, Nat. Commun. **10**, 592 (2019).

- [4] W. D. Richards, S. T. Dacek, D. A. Kitchaev, and G. Ceder, Fluorination of lithium-excess transition metal oxide cathode materials, Adv. Energy Mater. 8, 1701533 (2018).
- [5] J. Huang, P. Zhong, Y. Ha, Z. Lun, Y. Tian, M. Balasubramanian, W. Yang, and G. Ceder, Oxygen vacancy introduction to increase the capacity and voltage retention in Li-excess cathode materials, Small Struct., 2200343 (2022)..
- [6] R. Zhang, et al., Compositionally complex doping for zerostrain zero-cobalt layered cathodes, Nature 610, 67 (2022).
- [7] K. Zhou, Y. Li, Y. Ha, M. Zhang, W. Dachraoui, H. Liu, C. Zhang, X. Liu, F. Liu, C. Battaglia, W. Yang, J. Liu, and Y. Yang, A nearly zero-strain Li-rich rock-salt oxide with multielectron redox reactions as a cathode for Li-ion batteries, Chem. Mater. 34, 9711 (2022).
- [8] X. Zhao, Y. Tian, Z. Lun, Z. Cai, T. Chen, B. Ouyang, and G. Ceder, Design principles for zero-strain Li-ion cathodes, Joule 6, 1654 (2022).
- [9] Y. S. Meng and M. E. Arroyo-De Dompablo, First principles computational materials design for energy storage materials in lithium ion batteries, Energy Environ. Sci. 2, 589 (2009).
- [10] A. Urban, D.-H. Seo, and G. Ceder, Computational understanding of Li-ion batteries, npj Comput. Mater. 2, 16002 (2016).
- [11] A. Van Der Ven, Z. Deng, S. Banerjee, and S. P. Ong, Rechargeable alkali-ion battery materials: Theory and computation, Chem. Rev. 120, 6977 (2020).
- [12] M. K. Aydinol, A. F. Kohan, G. Ceder, K. Cho, and J. Joannopoulos, Ab initio study of lithium intercalation in metal oxides and metal dichalcogenides, Phys. Rev. B 56, 1354 (1997).
- [13] A. Urban, A. Abdellahi, S. Dacek, N. Artrith, and G. Ceder, Electronic-structure origin of cation disorder in transitionmetal oxides, Phys. Rev. Lett. 119, 1 (2017).
- [14] Z. Lun, B. Ouyang, D. A. Kitchaev, R. J. Clément, J. K. Papp, M. Balasubramanian, Y. Tian, T. Lei, T. Shi, B. D. McCloskey, J. Lee, and G. Ceder, Improved cycling performance of Li-excess cation-disordered cathode materials upon fluorine substitution, Adv. Energy Mater. 9, 1802959 (2019).
- [15] F. Zhou, T. Maxisch, and G. Ceder, Configurational electronic entropy and the phase diagram of mixed-valence oxides: The case of Li_x FePO₄, Phys. Rev. Lett. **97**, 155704 (2006).
- [16] J. Lee, C. Wang, R. Malik, Y. Dong, Y. Huang, D. Seo, and J. Li, Determining the criticality of Li-excess for disordered-rocksalt Li-ion battery cathodes, Adv. Energy Mater. 11, 2100204 (2021).
- [17] B. Li, M. T. Sougrati, G. Rousse, A. V. Morozov, R. Dedryvère, A. Iadecola, A. Senyshyn, L. Zhang, A. M. Abakumov, M.-L. Doublet, and J.-M. Tarascon, Correlating ligand-to-metal charge transfer with voltage hysteresis in a Li-rich rock-salt compound exhibiting anionic redox, Nat. Chem. 13, 1070 (2021).
- [18] Z. Yao, S. Kim, J. He, V. I. Hegde, and C. Wolverton, Interplay of cation and anion redox in Li₄Mn₂O₅ cathode

material and prediction of improved $Li_4(Mn,M)_2O_5$ electrodes for Li-ion batteries, Sci. Adv. 4, eaao6754 (2018).

- [19] Q. Li, J. Wu, Z. Yao, Y. Xu, M. M. Thackeray, C. Wolverton, and V. P. Dravid, Dynamic imaging of metastable reaction pathways in lithiated cobalt oxide electrodes, Nano Energy 44, 15 (2018).
- [20] H. Ji, D. A. Kitchaev, Z. Lun, H. Kim, E. Foley, D. H. Kwon, Y. Tian, M. Balasubramanian, M. Bianchini, Z. Cai, R. J. Clément, J. C. Kim, and G. Ceder, Computational investigation and experimental realization of disordered high-capacity Li-ion cathodes based on Ni redox, Chem. Mater. **31**, 2431 (2019).
- [21] K. McColl, R. A. House, G. J. Rees, A. G. Squires, S. W. Coles, P. G. Bruce, B. J. Morgan, and M. S. Islam, Transition metal migration and O₂ formation underpin voltage hysteresis in oxygen-redox disordered rocksalt cathodes, Nat. Commun. 13, 5275 (2022).
- [22] M. Ångqvist, W. A. Muñoz, J. M. Rahm, E. Fransson, C. Durniak, P. Rozyczko, T. H. Rod, and P. Erhart, ICET–a Python library for constructing and sampling alloy cluster expansions, Adv. Theory Simul. 2, 1900015 (2019).
- [23] Q. Wu, B. He, T. Song, J. Gao, and S. Shi, Cluster expansion method and its application in computational materials science, Comput. Mater. Sci. 125, 243 (2016).
- [24] A. van de Walle, M. Asta, and G. Ceder, The alloy theoretic automated toolkit: A user guide, Calphad 26, 539 (2002).
- [25] J.-Z. Xie, X.-Y. Zhou, and H. Jiang, Perspective on optimal strategies of building cluster expansion models for configurationally disordered materials, J. Chem. Phys. 157, 200901 (2022).
- [26] A. Seko and 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, 115403 (2014).
- [27] L. Barroso-Luque, P. Zhong, J. H. Yang, F. Xie, T. Chen, B. Ouyang, and G. Ceder, Cluster expansions of multicomponent ionic materials: Formalism and methodology, Phys. Rev. B 106, 144202 (2022).
- [28] A. van de Walle, Multicomponent multisublattice alloys, nonconfigurational entropy and other additions to the alloy theoretic automated toolkit, Calphad 33, 266 (2009).
- [29] L. J. Nelson, G. L. W. Hart, F. Zhou, and V. Ozoliņš, Compressive sensing as a paradigm for building physics models, Phys. Rev. B 87, 035125 (2013).
- [30] A. Seko, A. Takahashi, and I. Tanaka, Sparse representation for a potential energy surface, Phys. Rev. B 90, 024101 (2014).
- [31] P. Zhong, T. Chen, L. Barroso-Luque, F. Xie, and G. Ceder, An $\ell_0\ell_2$ -norm regularized regression model for construction of robust cluster expansions in multicomponent systems, Phys. Rev. B **106**, 024203 (2022).
- [32] B. Puchala, J. C. Thomas, A. R. Natarajan, J. G. Goiri, S. S. Behara, J. L. Kaufman, and A. Van der Ven, CASM A software package for first-principles based study of multicomponent crystalline solids, Comput. Mater. Sci. 217, 111897 (2023).
- [33] C. Wolverton and A. Zunger, First-principles prediction of vacancy order-disorder and intercalation battery voltages in Li_xCoO_2 , Phys. Rev. Lett. **81**, 606 (1998).

- [34] A. Van der Ven, M. K. Aydinol, G. Ceder, G. Kresse, and J. Hafner, First-principles investigation of phase stability in Li_xCoO_2 , Phys. Rev. B **58**, 2975 (1998).
- [35] G. Houchins and V. Viswanathan, An accurate machinelearning calculator for optimization of Li-ion battery cathodes, J. Chem. Phys. 153, 054124 (2020).
- [36] S. K. Kolli and A. Van Der Ven, First-principles study of spinel MgTiS₂ as a cathode material, Chem. Mater. **30**, 2436 (2018).
- [37] X. Guo, C. Chen, and S. P. Ong, Intercalation chemistry of the disordered rocksalt Li₃V₂O₅ anode from cluster expansions and machine learning interatomic potentials, Chem. Mater. 35, 1537 (2023).
- [38] J. P. Valleau and L. K. Cohen, Primitive model electrolytes.
 I. Grand canonical Monte Carlo computations, J. Chem. Phys. 72, 5935 (1980).
- [39] Z. Deng, G. Sai Gautam, S. K. Kolli, J. N. Chotard, A. K. Cheetham, C. Masquelier, and P. Canepa, Phase behavior in rhombohedral NaSiCON electrolytes and electrodes, Chem. Mater. 32, 7908 (2020).
- [40] F. Xie, P. Zhong, L. Barroso-Luque, B. Ouyang, and G. Ceder, Semigrand-canonical Monte-Carlo simulation methods for charge-decorated cluster expansions, Comput. Mater. Sci. 218, 112000 (2023).
- [41] J. H. Yang, T. Chen, L. Barroso-Luque, Z. Jadidi, and G. Ceder, Approaches for handling high-dimensional cluster expansions of ionic systems, npj Comput. Mater. 8, 133 (2022).
- [42] D.-H. Seo, J. Lee, A. Urban, R. Malik, S. Kang, and G. Ceder, The structural and chemical origin of the oxygen redox activity in layered and cation-disordered Li-excess cathode materials, Nat. Chem. 8, 692 (2016).
- [43] L. Barroso-Luque, J. H. Yang, F. Xie, T. Chen, R. L. Kam, Z. Jadidi, P. Zhong, and G. Ceder, smol: A Python package for cluster expansions and beyond, J. Open Source Softw. 7, 4504 (2022).
- [44] Gurobi Optimization, LLC, Gurobi Optimizer Reference Manual (2021).
- [45] A. Abdellahi, A. Urban, S. Dacek, and G. Ceder, The effect of cation disorder on the average Li intercalation voltage of transition-metal oxides, Chem. Mater. 28, 3659 (2016).
- [46] A. G. Squires and D. O. Scanlon, Understanding the limits to short-range order suppression in many-component disordered rock salt lithium-ion cathode materials, J. Mater. Chem. A 11, 13765 (2023).
- [47] L. Huang, P. Zhong, Y. Ha, Z. Cai, Y. Byeon, T. Huang, Y. Sun, F. Xie, H. Hau, H. Kim, M. Balasubramanian, B. D. McCloskey, W. Yang, and G. Ceder, Optimizing Li-excess cation-disordered rocksalt cathode design through partial Li deficiency, Adv. Energy Mater. 13, 2202345 (2023).
- [48] F. Zhou, M. Cococcioni, C. A. Marianetti, D. Morgan, and G. Ceder, First-principles prediction of redox potentials in transition-metal compounds with LDA + U, Phys. Rev. B **70**, 235121 (2004). 0406382,
- [49] J. Sun, A. Ruzsinszky, and J. P. Perdew, Strongly constrained and appropriately normed semilocal density functional, Phys. Rev. Lett. 115, 036402 (2015).
- [50] N. Artrith, J. A. Garrido Torres, A. Urban, and M. S. Hybertsen, Data-driven approach to parameterize SCAN +

U for an accurate description of 3d transition metal oxide thermochemistry, Phys. Rev. Mater. **6**, 035003 (2022).

- [51] S. Swathilakshmi, R. Devi, and G. Sai Gautam, Performance of the *r*²SCAN functional in transition metal oxides, J. Chem. Theory Comput. **19**, 4202 (2023).
- [52] J. Reed and G. Ceder, Role of electronic structure in the susceptibility of metastable transition-metal oxide structures to transformation, Chem. Rev. 104, 4513 (2004).
- [53] T. Y. Huang, Z. Cai, M. J. Crafton, L. A. Kaufman, Z. M. Konz, H. K. Bergstrom, E. A. Kedzie, H. M. Hao, G. Ceder, and B. D. McCloskey, Quantitative decoupling of oxygenredox and manganese-redox voltage hysteresis in a cationdisordered rock salt cathode, Adv. Energy Mater. 2300241, 1 (2023).
- [54] G. Ceder, A derivation of the Ising model for the computation of phase diagrams, Comput. Mater. Sci. 1, 144 (1993).
- [55] M. Ménétrier, I. Saadoune, S. Levasseur, and C. Delmas, The insulator-metal transition upon lithium deintercalation from LiCoO₂: electronic properties and ⁷Li NMR study, J. Mater. Chem. 9, 1135 (1999).
- [56] A. Zunger, S.-H. Wei, L. G. Ferreira, and J. E. Bernard, Special quasirandom structures, Phys. Rev. Lett. 65, 353 (1990).
- [57] J. Huang, P. Zhong, Y. Ha, D.-h. Kwon, M. J. Crafton, Y. Tian, M. Balasubramanian, B. D. McCloskey, W. Yang, and G. Ceder, Non-topotactic reactions enable high rate capability in Li-rich cathode materials, Nat. Energy 6, 706 (2021).
- [58] Z. Lun, B. Ouyang, D.-h. Kwon, Y. Ha, E. E. Foley, T.-Y. Huang, Z. Cai, H. Kim, M. Balasubramanian, Y. Sun, J. Huang, Y. Tian, H. Kim, B. D. McCloskey, W. Yang, R. J. Clément, H. Ji, and G. Ceder, Cation-disordered rocksalttype high-entropy cathodes for Li-ion batteries, Nat. Mater. 20, 214 (2021).
- [59] M. Nakajima and N. Yabuuchi, Lithium-excess cationdisordered rocksalt-type oxide with nanoscale phase segregation: Li_{1.25}Nb_{0.25}V_{0.5}O₂, Chem. Mater. **29**, 6927 (2017).
- [60] Z. Lebens-Higgins, H. Chung, I. Temprano, M. Zuba, J. Wu, J. Rana, C. Mejia, M. A. Jones, L. Wang, C. P. Grey, and *et al.*, Electrochemical utilization of iron IV in the Li_{1.3}Fe_{0.4}Nb_{0.3}O₂ disordered rocksalt cathode, Batteries Supercaps 4, 771 (2021).
- [61] E. Weinan, Machine learning and computational mathematics, arXiv preprint ArXiv:2009.14596 (2020).
- [62] C. Chen and S. P. Ong, A universal graph deep learning interatomic potential for the periodic table, Nat. Comput. Sci. 2, 718 (2022).

- [63] S. Takamoto, *et al.*, Towards universal neural network potential for material discovery applicable to arbitrary combination of 45 elements, Nat. Commun. 13, 1 (2022).
- [64] B. Deng, P. Zhong, K. Jun, J. Riebesell, K. Han, C. J. Bartel, and G. Ceder, CHGNet as a pretrained universal neural network potential for charge-informed atomistic modelling, Nat. Mach. Intell. (2023).
- [65] J. Z. Xie, X. Y. Zhou, D. Luan, and H. Jiang, Machine learning force field aided cluster expansion approach to configurationally disordered materials: Critical assessment of training set selection and size convergence, J. Chem. Theory Comput. 18, 3795 (2022).
- [66] G. Kresse and J. Furthmüller, Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set, Comput. Mater. Sci. 6, 15 (1996).
- [67] G. Kresse and D. Joubert, From ultrasoft pseudopotentials to the projector augmented-wave method, Phys. Rev. B 59, 1758 (1999).
- [68] J. W. Furness, A. D. Kaplan, J. Ning, J. P. Perdew, and J. Sun, Accurate and numerically efficient r²SCAN metageneralized gradient approximation, J. Phys. Chem. Lett. 11, 8208 (2020). 2008.03374,
- [69] Y. Zhang, D. A. Kitchaev, J. Yang, T. Chen, S. T. Dacek, R. A. Sarmiento-Pérez, M. A. Marques, H. Peng, G. Ceder, and J. P. Perdew, *et al.*, Efficient first-principles prediction of solid stability: Towards chemical accuracy, npj Comput. Mater. 4, 1 (2018).
- [70] R. Kingsbury, A. S. Gupta, C. J. Bartel, J. M. Munro, S. Dwaraknath, M. Horton, and K. A. Persson, Performance comparison of r²SCAN and SCAN metaGGA density functionals for solid materials via an automated, highthroughput computational workflow, Phys. Rev. Mater. 6, 013801 (2022).
- [71] D.-H. Seo, A. Urban, and G. Ceder, Calibrating transitionmetal energy levels and oxygen bands in first-principles calculations: Accurate prediction of redox potentials and charge transfer in lithium transition-metal oxides, Phys. Rev. B 92, 115118 (2015).
- [72] A. S. Tygesen, J. H. Chang, T. Vegge, and J. M. García-Lastra, Computational framework for a systematic investigation of anionic redox process in Li-rich compounds, npj Comput. Mater. 6, 65 (2020).
- [73] J. Vinckeviciute, D. A. Kitchaev, and A. Van der Ven, A two-step oxidation mechanism controlled by Mn migration explains the first-cycle activation behavior of Li₂MnO₃based Li-excess materials, Chem. Mater. **33**, 1625 (2021).