Probing configurational disorder in $ZnGeN_2$ using cluster-based Monte Carlo

Jacob J. Cordell [1,2,*] Jie Pan [2] Adele C. Tamboli [2,3] Garritt J. Tucker [1] and Stephan Lany [2,†]

[1]*Department of Mechanical Engineering, Colorado School of Mines, Golden, Colorado 80401, USA*

[2]*Materials Science Center, National Renewable Energy Laboratory, Golden, Colorado 80401, USA*

[3]*Physics Department, Colorado School of Mines, Golden, Colorado 80401, USA*

(Received 13 October 2020; accepted 10 December 2020; published 16 February 2021)

$ZnGeN_2$ is sought as a semiconductor with comparable lattice constant to GaN and tunable band gap for integration in optoelectronic devices. Configurational disorder on the cation sublattice of $ZnGeN_2$ can strongly modify the electronic structure compared to the ordered material, and both ordered and disordered forms of $ZnGeN_2$ are candidates for light-emitting diodes and other emitters. The nonisovalent character of the disordered species (Zn^{2+} and Ge^{4+}) subjects the cation ordering to strong short-range order effects. To model these effects, we use Monte Carlo (MC) simulations utilizing a cluster expansion to approximate formation enthalpy. Representative disordered configurations in 1024-atom supercells are relaxed in density functional theory calculations. From the MC structures, we extract a short-range order parameter (the N-cation coordination motif), and two long-range order parameters (Bragg-Williams and stretching parameters), and examine their correlations. We perform a thermodynamic integration to determine the mixing entropy and free energy. $ZnGeN_2$ exhibits a first-order phase transition with pronounced discontinuities in enthalpy and entropy, as well as in the structural order parameters. We discuss the relationship between the effective temperature used in the MC simulation and the growth temperatures in experiment in relation to the crossover from the nonequilibrium to the equilibrium growth regime. This work expands on current models of site disorder in $ZnGeN_2$ and provides atomic structure models with a systematic variation of the degree of cation disorder.

DOI: 10.1103/PhysRevMaterials.5.024604

I. INTRODUCTION

Site disorder in multinary materials is an area of growing interest for determining properties across diverse material fields from metal alloys [1–3] to electronics [4–7]. The ternary nitrides in particular exhibit a great deal of current interest in the literature [8–15]. Recent studies have highlighted the importance of disorder at short- (nearest neighbor), mid- (a few bond lengths), and long- (several periodic units) range in determining electronic properties in ternary nitride semiconductors [16–20], but significant challenges remain in deconvolving and characterizing site disorder on multiple scales. In this study we quantify site disorder in $ZnGeN_2$ and contrast correlations and limitations between order parameters obtained from theory and experiment. Transitions in short-range order (SRO) and long-range order (LRO) are rooted in changes in free energy of the system, where we extract the configurational entropy contribution from Monte Carlo (MC) simulation through integration of formation enthalpy over the temperature range of the simulation [21].

Disorder adds a significant entropic contribution to the free energy during synthesis and can be locked in due to limited kinetics [22–24], impacting electronic properties [25,26]. $ZnGeN_2$ specifically has displayed cation disorder experimentally and has been explored for use in light emitters and nonlinear optics [10,27–29]. $ZnGeN_2$ takes a wurtzite-derived crystal structure with the orthorhombic $Pna2_1$ space-group number 33 (SG33) in its ground-state ordered form as shown in Fig. 1. In contrast to several other wurtzite-like ternary nitrides, such as $ZnSnN_2$ [30], Zn_3MoN_4 [31], and Mg_2SbN_3 [32], cation ordering in $ZnGeN_2$ induces a significant distortion of the ideal wurtzite lattice and the lattice distortion causes peak splitting in x-ray diffraction (XRD) [10]. The XRD of disordered $ZnGeN_2$ resembles that of ideal wurtzite, space group $P6_3mc$ (SG186), with partial Zn/Ge occupancy on the cation site, but diffraction can not provide information on the degree of SRO in a sample. The similar scattering properties of Zn and Ge [10,27,33] further complicate the experimental analysis of disorder, making a complementary computational study even more valuable.

Previous reports have provided multiple models for disorder in wurtzite-derived Zn-IV-N_2 including the work of Quayle *et al.* into octet rule-conserving disorder based on differential stacking of known, low-energy ordered structures in $ZnSnN_2$ [16] and the investigation by Skachkov *et al.* on the direct impact of isolated and paired defects in $ZnGeN_2$ on band structure [34]. Quayle *et al.* showed that octet rule-conserving disorder yields XRD signatures indistinguishable from idealized wurtzite. Skachkov *et al.* investigated the impact of site disorder on electronic structure by comparing the band structure of $ZnGeN_2$ with different quantities and proximities of exchange defects, demonstrating cation antisites as the most dominant native defects in $ZnGeN_2$. In addition to these works, numerous defect studies have also looked at potential native and non-native doping effects in $ZnGeN_2$ [11,35–37]. This work expands on current models

*cordell@mines.edu

†stephan.lany@nrel.gov

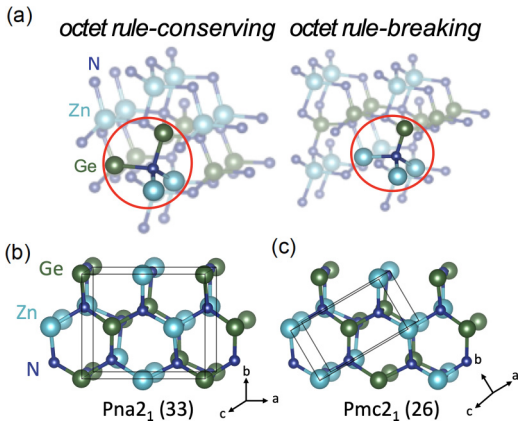

FIG. 1. (a) Octet rule-conserving (left) and breaking (right) motifs in wurtzite-derived $ZnGeN_2$. Orthorhombic primitive cells of (b) *Pna*2$_1$ (SG33), 16 atoms and (c) *Pmc*2$_1$ (SG26), 8 atoms are outlined for $ZnGeN_2$.

to include the systematic variation of the degree of disorder via Monte Carlo simulations. We also differentiate between SRO, in the sense of N-atom coordination motifs [30], and LRO in the sense of continuous periodicity within the crystal. The ground state of $ZnGeN_2$ consists exclusively of local octet rule-conserving motifs (N-Zn_2Ge_2); disorder introduces octet rule-breaking motifs, illustrated in Fig. 1.

Disordered distributions of different atom types on an underlying lattice or sublattice are often modeled using the method of special quasirandom structures (SQS) [38]. In the case of a fully randomized configuration, statistics provide the pair (or higher-order) correlation functions, and the SQS is then constructed as a small supercell of the underlying lattice with an atomic decoration that optimally approximates the ideal random correlation functions. The random approximation is often well justified for disorder between isovalent ions with little or moderate size mismatch, for example, in common III-V and II-VI semiconductor alloys [38]. However, it breaks down in the case of aliovalent mixing [26,30,39,40] or when highly size mismatched atoms are involved as in the case of quaternary $Ga_{1-x}In_xP_{1-y}N_y$ alloys [41]. In these cases, electrostatic interactions or attractive forces due to mutually compensating strain fields cause strong SRO effects at realistic temperatures, and the random mixing model describes only the hypothetical limit at infinite temperature.

For systems in which assumptions of complete randomness do not apply, MC-based strategies have been used to obtain configurations that are neither fully ordered nor completely random [26,41–43]. The MC approach determines the balance between configurational enthalpy and entropy via statistical sampling. For computational efficiency, the enthalpy evaluation is usually performed with a model Hamiltonian fitted to approximate first-principles energies, e.g., obtained from density functional theory (DFT). In $ZnSnN_2$, the local bonding environment in a good approximation predicts the configurational total energy, allowing the use of a very simple motif Hamiltonian [30]. In $ZnGeN_2$, however, we find that different structures consisting only of the ideal N-Zn_2Ge_2 motifs (Fig. 1) exhibit considerably different energies, hinting towards much stronger LRO effects. The cluster expansion approach [44–46] provides an efficient energy functional for the Zn/Ge decoration of the cation sublattice and captures contributions from both SRO and LRO. This work employs a cluster expansion-based MC approach implementing the clusters approach to statistical mechanics (CASM) provided by Puchala *et al.* and Thomas *et al.* [47,48] and implemented previously in studies of disorder in multinary semiconductors [49] and alloys [50].

II. METHODOLOGY

The method detailed below follows four steps to obtain structures with varying degree of cation order. The first three of these steps follow the workflow of the CASM software [47,48]. (1) First-principles calculations were carried out on an energetically representative set of $ZnGeN_2$ configurations to relax structure and obtain total energy. (2) A cluster expansion was fitted to these structures and validated through comparison to test configurations to obtain a Hamiltonian for approximating formation enthalpy of a generic ion configuration of the material of interest. (3) MC simulations were performed at selected temperatures using the cluster expansion to compare energy of ion configurations. (4) Supercells at temperatures of interest were relaxed in lattice vectors and atomic positions through DFT calculations to obtain structural order parameters and prepare these configurations for electronic structure calculations.

Our objective is to generate atomic structure models describing $ZnGeN_2$ in a disordered state as it is often experimentally observed [10,28]. Thin-film synthesis techniques such as sputtering or pulsed laser deposition often create materials in nonequilibrium states with energies higher than the expected ground state. The metastability manifests itself in uncrystallized amorphous structures [51,52], disorder in ternary or multinary compounds [53–55], or even materials grown with positive formation enthalpy [56] (thermochemically metastable). Fundamentally, the metastability results from kinetic limitations inhibiting the equilibration on different length scales, ranging from a few angstroms (crystallization) to several nanometers (atomic ordering) to the mesoscale and microscale (phase separation).

We investigate relationships among structure, order, and energy in $ZnGeN_2$ by comparing the experimentally obtainable stretching parameter and Bragg LRO parameter [57–59] to the fraction of short-range motifs and configurational formation enthalpy as a function of effective temperature. Mapping the bonding environment (motif composition) of cells, as a function of effective temperature, formation enthalpy, and order parameter, we expand the comparison of order metrics from previous work [17] to address implications for the tunability of order and order-parameter-dependent properties of $ZnGeN_2$.

To connect our MC simulations to nonequilibrium synthesis, we employ the concept of an effective temperature T_{eff}, allowing us to systematically vary the degree of atomic site disorder [60]. Whereas in equilibrium, thermal energy creates disorder, in nonequilibrium growth, disorder results from the randomness of atomic distributions and the inhibition of the thermodynamic driving forces towards ordered structures with lower enthalpy. The effective temperature corresponds

to the thermodynamic temperature that would result in the same degree of disorder as in the nonequilibrium system. The strong kinetic limitation at low deposition temperature T_{dep} causes a high degree of disorder (high T_{eff}), which gradually reduces as the increasing T_{dep} allows a partial equilibration. Consequently, T_{eff} decreases with increasing T_{dep} until equilibration on the respective length scale completes, at which point the two temperatures converge [60]. The nonmonotonic dependence of T_{eff} on T_{dep} will be further discussed under Sec. VII in the context of both the computational results of this work and available experimental data.

III. FITTING THE CLUSTER EXPANSION

In $ZnSnN_2$, the energy differences between cation arrangements were well described by the local motif Hamiltonian to an accuracy of 15 meV per cation (meV/cat) [30]. It was also observed that different LRO structures containing only ideal N-Zn_2Sn_2 motifs are essentially energetically degenerate. These findings highlight the dominant role of SRO in $ZnSnN_2$. However, in $ZnGeN_2$, the prevalence of energetic contributions beyond nearest-neighbor ions makes the motif Hamiltonian insufficient for approximating formation enthalpy. The two ideal motif (N-Zn_2Ge_2) structures realized in small unit cells of 8 and 16 atoms, i.e., the $Pmc2_1$ (SG26) and the $Pna2_1$ (SG33) structures [16], have an energy difference of $\Delta E_{26-33} = 43$ meV/cat (compared to 1 meV/cat in the case of $ZnSnN_2$ [61]). This additional energy contribution in $ZnGeN_2$ likely relates to the difference in ionic radii ($Zn^{2+} = 74$ pm, $Ge_4^+ = 53$ pm, $Sn^{4+} = 69$ pm) [62], and motivated the use of a cluster expansion instead of the simpler motif Hamiltonian in this study of $ZnGeN_2$.

The cluster expansion has the form

$$H = \sum_{\alpha} m_{\alpha} J_{\alpha} \left\langle \prod_{i \in \beta} \sigma_i \right\rangle, \tag{1}$$

where H is enthalpy of formation and m and J are the multiplicity (number of occurrences) and energy parameter (contribution from each cluster), respectively, of a given cluster α. σ is the configuration of occupation variables, σ_i on a lattice site of a given subset, where β is the set of clusters symmetrically equivalent to α and i is the index of the atoms in a cluster β [63].

To generate a training set for the cluster expansion (CE) from DFT calculations, all atomic structures of $ZnGeN_2$ were created by first generating supercells of a primitive ideal wurtzite cell with nitrogen as the anion and a generic cation site, where the wurtzite cell dimensions were normalized to the volume per atom of the $ZnGeN_2$ ground state (SG33). The total energies of the training and test set structures were calculated with DFT including relaxation of atomic positions and cell parameters, as described below in Sec. VIII. We use the formation enthalpies ΔH_f, obtained by subtracting the elemental reference energies [64,65]. For the SG33 ground state, we obtain $\Delta H_f = -704$ meV/cat. In the following, all energies are given on a per cation basis, the natural energy scale for Zn/Ge disorder on the cation sublattice where nitrogen is a spectator atom.

TABLE I. Comparison of cluster expansion (CE) and DFT enthalpy values with respect to basis set. CE-DFT indicates the difference in enthalpy of the SG33 structure as calculated by CE and DFT. SG26-SG33 provides the difference in enthalpy of SG26 and SG33 as calculated by CE; DFT calculations result in an enthalpy difference of 43 meV/cat between SG26 and SG33. All data are shown for the same training set of structures.

Basis set	CE-DFT (meV/cat)	SG26-SG33 (meV/cat)	CV score (meV/cat)	RMSE (meV/cat)
p10t6q6	0.68	41	13	10
p10t4	1.02	48	21	19
p10t4q4	0.80	49	21	19
p10t4q4q4	2.32	47	21	19

Supercell sizes between 8 and 32 atoms were used for cluster expansion fitting, and cells between 128 and 128 000 atoms were generated for MC simulations. The distributions of Zn and Ge atoms in a 1:1 ratio on the cation site were generated by exhaustive enumeration for the smaller 8- and 16-atom cells, thereby including the ordered SG33 and SG26 structures. In larger cells, the cation species (e.g., Zn^{2+} or Ge^{4+}) were randomly assigned. In addition to random seed structures, supercells of the SG33 structure as a fully ordered starting point for the MC simulations were created. For training and testing, we used a total of 4 8-atom structures, 85 16-atom, and 96 32-atom structures split into three overlapping sets. These cells were partitioned such that training sets contained ∼80 structures, including all 8-atom structures, both ordered structures (8-atom SG26 and 16-atom SG33), and a remaining mixture of 60% 16-atom and 30% 32-atom structures. Each training set has a corresponding test set comprised of the ∼100 remaining structures (16 and 32 atom).

The cluster expansion chosen for this work was determined via comparison of cross validation (CV) score and root-mean-square error using the CASM framework [47,48]. CASM automates the process of comparing different expansions given a basis set and outputs a hall of fame, which lists expansions in the order of increasing CV score beginning with the lowest CV score. The different possible basis sets are classified by the cluster sizes (number of atoms) and the respective cutoff distance for atoms in the cluster. Four basis sets were tested as listed in the left column of Table I. We use shorthand to indicate different cluster expansion schemes. In this shorthand, each letter denotes an interaction quantified by number of atoms followed by the cutoff distance in angstroms for that interaction. p denotes pairs (two atoms), t denotes triplets (three atoms), the first q denotes quadruplets and a second q, if present, denotes quintuplets, e.g., p10t4 indicates a basis set with pairs up to 10 Å and triplets up to 4 Å.

A genetic algorithm was used to iterate over expansions to sample many combinations of interactions within the infinite possible set constrained by the cutoff distance and ion quantities of the basis set and a linear regression was applied to refine this search in each space. This process was repeated using three different training sets of structures to provide comparison between potential expansions. The genetic algorithm used the "LeaveOneOut" scheme to include a subset of the

TABLE II. Mean average error for cluster expansions based on three separate fitting sets for basis set p10t4.

Fit Set	1	2	3
Low-energy MAE (meV/cat)	8.9	10.5	10.9
Total MAE (meV/cat)	18.0	17.6	15.7

structures in the training set and then calculate CV score and root-mean-square error (RMSE) using the full training set.

The best candidate cluster expansions were chosen from CASM's hall of fame and compared via their ability to predict the enthalpy of formation from DFT for two ordered structures (SG33 and SG26) and, more importantly, the difference in energy of these two ordered structures. Table I shows these enthalpy difference values per cation for the best cluster expansion from all fitting sets as determined by CV score for each basis set.

It is desirable to bias the cluster expansion such that low-energy structures, which occur more frequently in the MC sampling, are more accurately described than less probable, high-energy structures. Therefore, we employed a weighting algorithm based on the DFT reference energy. As described by Eq. (2), each training set structure receives a weight w, following an exponential distribution of the form

$$w = A \times \exp\left(\frac{-\Delta E}{kT}\right) + B, \tag{2}$$

where ΔE is the DFT energy difference between this structure and the SG33 ground state. Here, the kT parameter, corresponding to a fictitious temperature, is set to 100 meV/cat and defines the energy resolution for the weighting scheme. The remaining parameters are chosen as $A = 1.0$ and $B = 0.4$. The CV score was then evaluated using the K-fold scheme implemented in the SCIKIT-LEARN library [66].

Enthalpies predicted by cluster expansion are plotted against the DFT-calculated energies in Fig. 2 as a function of cell size and inclusion in the training or test set for the cluster expansion used in this work. The cluster expansion chosen (shown in Fig. 2) was comparable to other fits based on 12 unique basis sets and 3 sets of DFT-relaxed training structures. Figure 2 includes all available calculated energies (both within and outside of the given training set) for fitting set 1. The weighting scheme provided in Eq. (2) gives the plots of E_{clex} versus E_{DFT} in Fig. 2 their shape with lower-energy structures generally closer to the line of ideality than higher-energy structures.

Table II shows the mean absolute error (MAE) for these three separate cluster expansions using basis set p10t4. From Table I, the obvious choice of basis set appears to be p10t6q6, however, cluster expansions using this basis set identify an incorrect structure as the ground state. This qualitative error could indicate an overfitting issue, against which smaller basis sets are generally more robust. Additionally, the similar performance among the remaining basis sets indicates that adding quadruplets or larger clusters has limited benefits. Table II also shows the MAE for structures only considering those from −705 to −500 meV/cat, the lowest-energy structures in the set. This energy range includes the lower third of possible configuration energies. Error in this low-energy range is more significant for choosing a cluster expansion than at higher energies because the expansion must correctly identify the ground state and choose low-energy configurations over high-energy ones to be of value in MC.

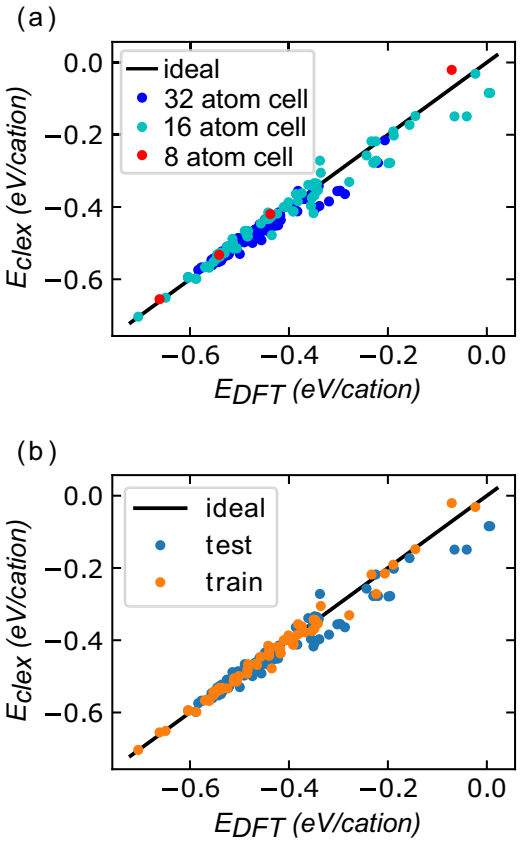

FIG. 2. Energies calculated via cluster expansion compared to those from DFT for basis set p10t4 and fitting set 1 (of 3) colored to show (a) distribution of supercell sizes and (b) test versus train data.

Variations of cluster expansions were calculated and compared using MAE and RMSE. From this analysis, a cluster expansion with pairs up to 10 Å and triplets up to 4 Å (basis set p10t4 from Table I) was chosen with MAE and RMSE per atom of 7.9 and 13 meV, respectively. We use the cluster expansion with the p10t4 basis set for the remainder of this work. Based on CV score and RMSE within a fitting set (Table I for fitting set 1) as well as MAE (Table II) and RMSE within a test set, basis set p10t4 closely matches larger, more inclusive basis sets p10t4q4 and p10t4q4q4. The cluster expansion using basis set p10t4 and fitting set 1 also correctly identifies relative energies of low-energy configurations in test supercells in contrast to p10t6q6.

Given the cluster expansion described above, a MC simulation progresses with acceptance criteria based on formation enthalpy. However, to properly interpret results, the simulation must be size converged; too small of a supercell leads to poor convergence at high temperatures while too large cells yield unfeasibly long equilibration periods. We tested convergence for supercells between 128 and 16 000 atoms in size with 1024 atom cells showing the best results at a range

of temperatures from 100 to 5000 K. For choosing a cell size and testing whether or not a structure converges at a given T_{eff}, cells were considered to be converged once seeds of both random and ordered starting configurations met in formation enthalpy. For MC simulations used in this work, convergence was then separately determined once the formation enthalpy of a supercell changed by less than 1 meV following 10 swaps. Given this convergence criterion, MC simulations were run from random seed configurations over a wider temperature range from 1 000 000 to 100 000 K in 10 000-K intervals, 100 000 to 10 000 K in 1000-K intervals, 10 000 to 5000 K in 100-K intervals, and 5000 to 100 K in 10-K intervals. For the cell size of 1024 atoms ($4 \times 4 \times 4$ supercell of the ground-state unit cell), we achieve convergence of the formation enthalpies for $T_{\text{eff}} = 2000$ K and higher as a function of number of MC passes where a pass is defined as a number of cation swaps equal to the number of cations in the system (see the Appendix for further detail on how size-converged supercells were chosen for this study).

IV. THERMODYNAMICS FROM MONTE CARLO SIMULATIONS

MC simulation has been used previously in a vast number of fields, such as to identify phase transitions in antiferromagnets, where free energy and entropy were obtained through numerical integration [21,67]. We employ this method to find configurational order transitions in $ZnGeN_2$ where these transitions play a role in electronic structure and properties. In this type of MC study, the Metropolis algorithm [68,69] samples the distribution of disordered configurations in relation to their formation enthalpies at a given effective temperature. According to the Metropolis method, the importance sampling of configurations X of an N-particle system is generated with probability $P \propto \frac{1}{Z}\exp(\frac{-H(X)}{k_B T})$ for temperature T and formation enthalpy $H(X)$, where Z is the partition function and k_B the Boltzmann constant. Here, $H(X)$ is calculated using the cluster expansion described in the previous section.

For a random atomic distribution, the configurational entropy is given by statistics,

$$S_{\text{rnd}}(T = \infty) = -k_B[x\ln(x) + (1-x)\ln(1-x)], \quad (3)$$

where x and $1-x$ are the compositions of the respective species, i.e., $x = 0.5$ as we consider only stoichiometric $ZnGeN_2$. In contrast, for a nonrandom, SRO distribution, the partition function Z, the temperature-dependent entropies and free energies are nontrivial. In MC calculations, the free-energy minimization is often implicit, and the entropy is not directly determined. However, the absolute free energy and entropy are obtained through thermodynamic integration [21] starting either from zero or infinite temperature where the entropy is known, i.e., $S = 0$ or as given by Eq. (3), respectively. We choose the high-temperature limit as the starting point, which is often more practical [21], and use the enthalpies of the random MC seed structures for the infinite-temperature limit. Equations (5)–(9) are adapted from Ref. [21] to illustrate the thermodynamic integration based on the data obtained from the CE-MC simulations.

We will consider $G = H - TS$ for disordered $ZnGeN_2$ with only configurational entropy (no vibrational effects). We start from the differential

$$d\left(\frac{1}{T}\right) = -\frac{1}{T^2}dT \quad (4)$$

and the Gibbs-Helmholtz equation

$$\left(\frac{\partial[G/T]}{\partial T}\right)_P = -\frac{1}{T^2}H. \quad (5)$$

Now, considering the integral of G/T from infinite T to a finite lower integration temperature T_i at constant pressure, by definition of the integral we find

$$[G/T](T = T_i) - [G/T](T = \infty) = \int_{\infty}^{T_i}\left(\frac{\partial[G/T]}{\partial T}\right)_P dT. \quad (6)$$

Using the Gibbs-Helmholtz equation

$$\frac{G(T_i)}{T_i} + S(T = \infty) = \int_{\infty}^{T_i} -\frac{1}{T^2}H(T)dT \quad (7)$$

and converting the integration variable to $1/T$ while multiplying by T_i yields

$$G(T_i) = T_i\left[-S(T = \infty) + \int_0^{\frac{1}{T_i}} H\left(\frac{1}{T}\right)d\left(\frac{1}{T}\right)\right]. \quad (8)$$

To calculate the free energy $\Delta G_{\text{MC}}(T)$ from Eq. (8), we use the enthalpy of formation $\Delta H_{\text{f}}(T)$ for $H(T)$ since the difference is a constant (sum of elemental reference energies) within our model for nonequilibrium disorder. The configurational entropy $\Delta S_{\text{MC}}(T)$ is obtained from

$$\Delta G_{\text{MC}}(T) = \Delta H_{\text{MC}}(T) - T\Delta S_{\text{MC}}. \quad (9)$$

As we identify the MC temperature with the effective temperature in nonequilibrium synthesis, we emphasize again that the temperature dependence of the free energy contains only configurational degrees of freedom, and excludes vibrational and ideal-gas contributions, which are associated with the much lower actual temperature. Therefore, $\Delta G_{\text{MC}}(T)$ does not represent the free energy of formation in the thermochemical sense. The experimentally observed ordering temperature around 1000 K, i.e., the point above which cation ordering equilibrates, suggests that only effective temperatures above 1000 K are relevant for the MC sampling. As mentioned above and discussed in more detail below under Sec. VII, T_{eff} is inversely correlated with deposition temperature T_{dep}, for T_{dep} below the equilibration point. The mapping between T_{eff} and T_{dep} is obtained in principle by comparison between measured and calculated order parameters [60], or perhaps via kinetic Monte Carlo calculations which, however, surpass the scope of this work.

Using effective temperature and formation enthalpies from MC, we followed Eq. (8) to obtain the change in configurational free energy, where Eq. (3) gives the configurational entropy at infinite temperature; i.e., for $x = 0.5$, $\Delta S_{\text{MC}} = 5.97 \times 10^{-5}$ eV K^{-1}. We used random seed configurations as a proxy for $T_{\text{eff}} = \infty$ for which we found $\Delta H_{\text{f}} = -440 \pm 20$ meV/cat, i.e., about 260 meV/cat above the ordered ground state. The formation enthalpy of these random seeds corresponds well to sample configurations equilibrated above 400 000 K which were found to exhibit

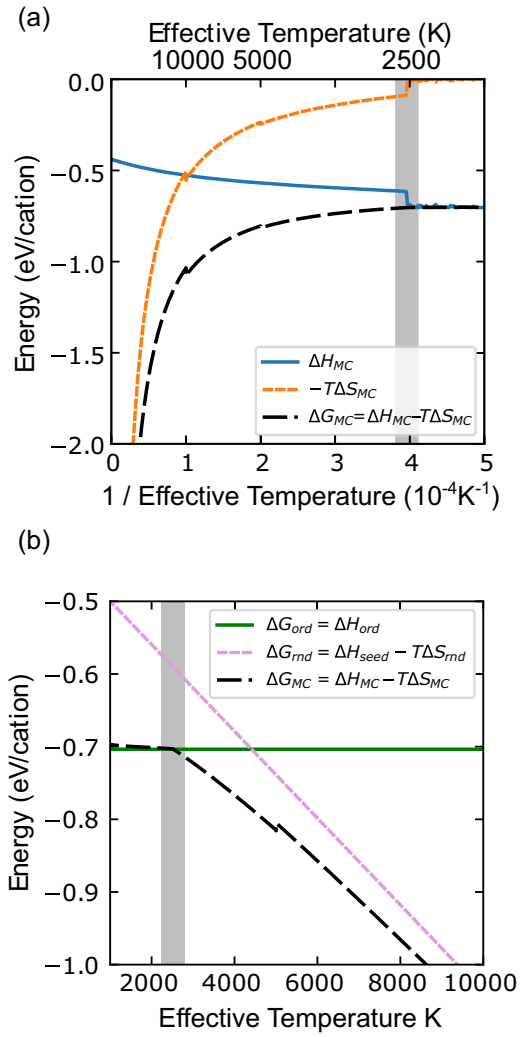

FIG. 3. (a) Contribution of entropy and formation enthalpy to configurational free energy in $ZnGeN_2$ as a function of the reciprocal effective temperature. (b) Free energies from MC calculations compared to those of ordered and fully random structures. The gray bars highlight the first-order transition.

$\Delta H_f = -440 \pm 10$ meV/cat. Following Eq. (8), integration over inverse temperature was carried out using the python package SCIPY's numerical trapezoidal rule solver [70]. The lowest inverse temperature sampled was 10^{-6} K^{-1}. Figure 3(a) shows ΔH_f obtained from the MC simulations alongside ΔG_{MC} and ΔS_{MC} obtained from the integration. Minor discontinuities in ΔG and ΔS occur at 10 000 and 5000 K ($1/T = 1 \times 10^{-4}$ and 2×10^{-4} K^{-1}, respectively). These discontinuities are artifacts of using multiple discrete temperature ranges for the MC integration, where discontinuities of <2 meV/cat in the underlying ΔH_f are amplified.

Clearly seen in Fig. 3(a) is a step in ΔH_f at 2500 K (4×10^{-4} K^{-1}) and a corresponding step in ΔS_{MC}, indicating latent heat associated with a first-order phase transition. In order to illustrate the phase transition behavior, we compare in Fig. 3 the numerical MC free energy with the model free energy of a random Zn/Ge distribution $\Delta G_{rnd}(T) = \Delta H_{rnd} - T\Delta S_{rnd}$ [cf. Eq. (3)], and with the ordered phase $\Delta G_{ord}(T) = \Delta H_{ord}(0)$, which is independent of temperature. From this analysis, we see that the ordered/random transition would be expected around 5000 K. The partially SRO structures obtained in the MC simulation reduce ΔG by up to 100 meV/cat relative to ordered or fully random structures and cut the transition temperature approximately in half. These effects illustrate strong nonrandom influence in disordered $ZnGeN_2$. At higher temperatures, ΔG_{MC} approaches ΔG_{rnd} continuously without further discontinuities, indicating a higher-order transition from SRO to random distributions.

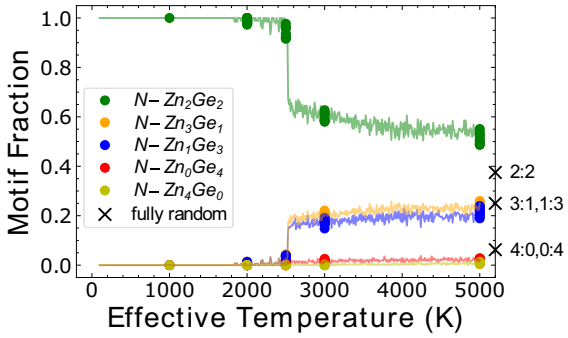

FIG. 4. Motif fractions as a function of effective temperature. Solid dots show data from constant-temperature MC equilibration; continuous lines show a simulated anneal. Black crosses (×) indicate concentration of motif fractions given a fully random distribution.

V. SHORT-RANGE ORDER-DISORDER TRANSITION

The term "order parameter" has been used in many material contexts historically with diverse meanings such as describing critical temperature in superconductors [71,72] and organization of Ising spin domains [73]. Here, we use order parameter to mean a metric for quantifying configurational cation disorder in $ZnGeN_2$ both theoretically and experimentally and relate these order parameters to the phase transitions described in the preceding section. Order parameters can depend largely on nearest-neighbor environment (SRO) or look at overall periodicity in a material (LRO) calculated with respect to the ground-state configuration as a reference. In this work we use a motif fraction to indicate the fraction of nitrogen in a supercell with a given coordination environment such as two Zn and two Ge, demonstrated in Fig. 1; this specific motif (N-Zn_2Ge_2) conserves the octet rule locally and is energetically favorable compared to the octet rule-breaking motifs.

Because each N has tetrahedral coordination, the proportion of each motif (plotted in Fig. 4) provides information on the bonding environment and SRO in the system. As mentioned in Sec. I, the differently charged cations Zn^{2+} and Ge^{4+} lead to nonrandom SRO even in the disordered phase. The black crosses (×) in Fig. 4 correspond to the concentration of each type of motif in a fully random distribution, $\frac{1}{16}$ for N-Zn_0Ge_4 and N-Zn_4Ge_0, $\frac{4}{16}$ for N-Zn_1Ge_3 and N-Zn_3Ge_1, and $\frac{6}{16}$ for N-Zn_2Ge_2.

Figure 4 uses data from MC for both simulations at constant temperature (dots) and simulated annealing (lines) with temperature steps of 10 K. The Appendix includes further

description of both of these types of MC including pass convergence criteria at each temperature and the Supplemental Material [74] includes the supercells equilibrated at discrete temperatures (dots, Fig. 4) in POSCAR format. From MC simulation, we identified a transition from ordered to disordered material between 2520 and 2530 K. The steep onset of the transition, vertical to within the 10-K temperature steps of the simulation, corresponds to the first-order transition in Fig. 3 and creates a step in accessible SRO parameters. This transition specifically designates a change from configurations of exclusively octet rule-conserving motifs to a mix of motif fractions. Even at temperatures above this transition, however, collective fractions of 4:0 motifs ($Zn_4Ge_0 + Zn_0Ge_4$) are below 0.07 at 5000 K, considerably lower than the value of 0.125 for a fully random system. The transition in SRO likely affects electronic properties as explored in the related $ZnSnN_2$ [30] and $ZnSnN_2 - ZnO$ [18] systems. However, we leave the investigation of order-electronic property relationships to future studies.

VI. LONG-RANGE ORDER PARAMETER η

In simulation, the Bragg and Williams LRO parameter η measures the fraction of ions on their defined sites for a given ordered configuration [57,58]. The LRO parameter is defined as

$$\eta = r_\alpha + r_\beta - 1, \tag{10}$$

where r_α is the fraction of site α occupied by atom α. Experimentally, η is determined from diffraction [59] where Rietveld refinement is used to obtain average site occupancies [75,76]. Given the relationship shown in Eq. (10), site occupancies are then used to extract the LRO parameter. However, obtaining XRD data suitable for Rietveld refinement in $ZnGeN_2$ is complicated by the small difference in scattering magnitude between Zn and Ge. In contrast to experimental analysis, it is possible in computational studies to know the exact quantity and position of ions of interest, making comparison of a given configuration and a reference structure trivial in principle, but other challenges arise. The ordered ground state can be realized in different symmetry-equivalent orientations within the MC supercell, necessitating calculation of an order parameter based on all possible symmetrically equivalent representations of the ground state. In order to calculate η for a supercell by comparing it to a reference, all possible sets of cation positions which correspond to the ordered structure of interest must also be compared, i.e., symmetry operations must be applied to find the symmetrically equivalent configurations. It is then appropriate to use the highest-order parameter obtained through these calculations as the "true" value of η, as the highest value most accurately reflects the extent of site ordering in the system.

There are 12 possible arrangements of cations on a wurtzite lattice that yield the orthorhombic ground-state structure in $ZnGeN_2$. The multiplicity of arrangements was determined by taking all possible combinations of 4 Zn and 4 Ge on 8 cation sites ($\binom{8}{4} = 70$ combinations) and quantifying the motif fractions in each of these possibilities. The calculation of motif fractions in the conventional ground-state unit cell of $ZnGeN_2$ yielded six structures with exclusively N-Zn_2Ge_2, of which DFT calculations showed four had the ground-state energy (the remaining two were SG26 instead of SG33). This number does not reflect possible external symmetry operations, of which three unique rotations were found. These external operations were applied to the permutations of octet rule-conserving reference configurations to obtain 12 unique references.

Figure 5(a) shows η for stoichiometric $ZnGeN_2$ as a function of T_{eff} with a color scale indicating ΔH_f of each supercell as calculated from DFT (see Sec. X). The sharp drop in η seen in Fig. 5(a) results from the order-disorder phase transition and coincides with the discontinuity in the short-range order parameter discussed above. Even above 5000 K, however, the system maintains a finite degree of LRO ($\eta > 0$), slowly decreasing between 5000 and 1 000 000 K. At that temperature, the highest T_{eff} sampled, it fluctuates around $\eta = 0.05 \pm 0.02$, a small nonzero value likely representing finite-size effects in the 1024-atom cell. Due to the sharp phase transition, our MC structures, which have been equilibrated at each temperature, do not exhibit intermediate values of either the short- or long-range order parameter, e.g., in the interval $0.3 < \eta < 0.9$.

We will now examine the correlation between SRO and LRO parameters. As shown in Fig. 5(b), small quantities of antisites correspond to small decreases in the fraction of N-Zn_2Ge_2 motifs in a near-linear fashion, implying a correlated loss of SRO and LRO as the system approaches the phase transition. At temperatures just above the phase transition, both order parameters are strongly reduced, but deviate significantly from the fully random system, with N-Zn_2Ge_2 fractions up to 0.6 and η up to 0.2. As they approach the random limit [N-Zn_2Ge_2 fraction of 0.375 and $\eta = 0$ as marked by $\times$ in Fig. 5(b)] with increasing effective temperature, the correlation becomes less defined, likely representing, again, residual finite-size effects. An experimental determination of these order parameters in the disordered phase would be interesting to quantify the effective temperatures and their variation with nonequilibrium growth parameters. However, given the experimental difficulties mentioned above, a measurement with the required resolution in the order parameters will be quite challenging.

Above, we discussed two transitions, an abrupt order-disorder transition around $T_{\text{eff}} = 2500$ K and a gradual higher-order transition from SRO to random. In similar systems such as $ZnSnN_2$, a third transition is also observed from a fully ordered system to a disordered system that preserves the octet rule as described in Ref. [16]. In the $ZnGeN_2$ configurations generated through MC, octet rule-conserving disorder is notably absent. The lack of this phase in $ZnGeN_2$, despite its presence in the structurally similar $ZnSnN_2$, is a result of the more pronounced energy contribution of LRO in $ZnGeN_2$. This difference is also visible in the larger formation enthalpy difference between the two octet-rule observing unit cells, i.e., 43 meV/cat (see Sec. II) in $ZnGeN_2$, compared to just 1 meV/cat [61] in $ZnSnN_2$. An important implication is that in $ZnGeN_2$, the loss of LRO and SRO is correlated, whereas in $ZnSnN_2$, SRO can persist in the presence of long-range disorder.

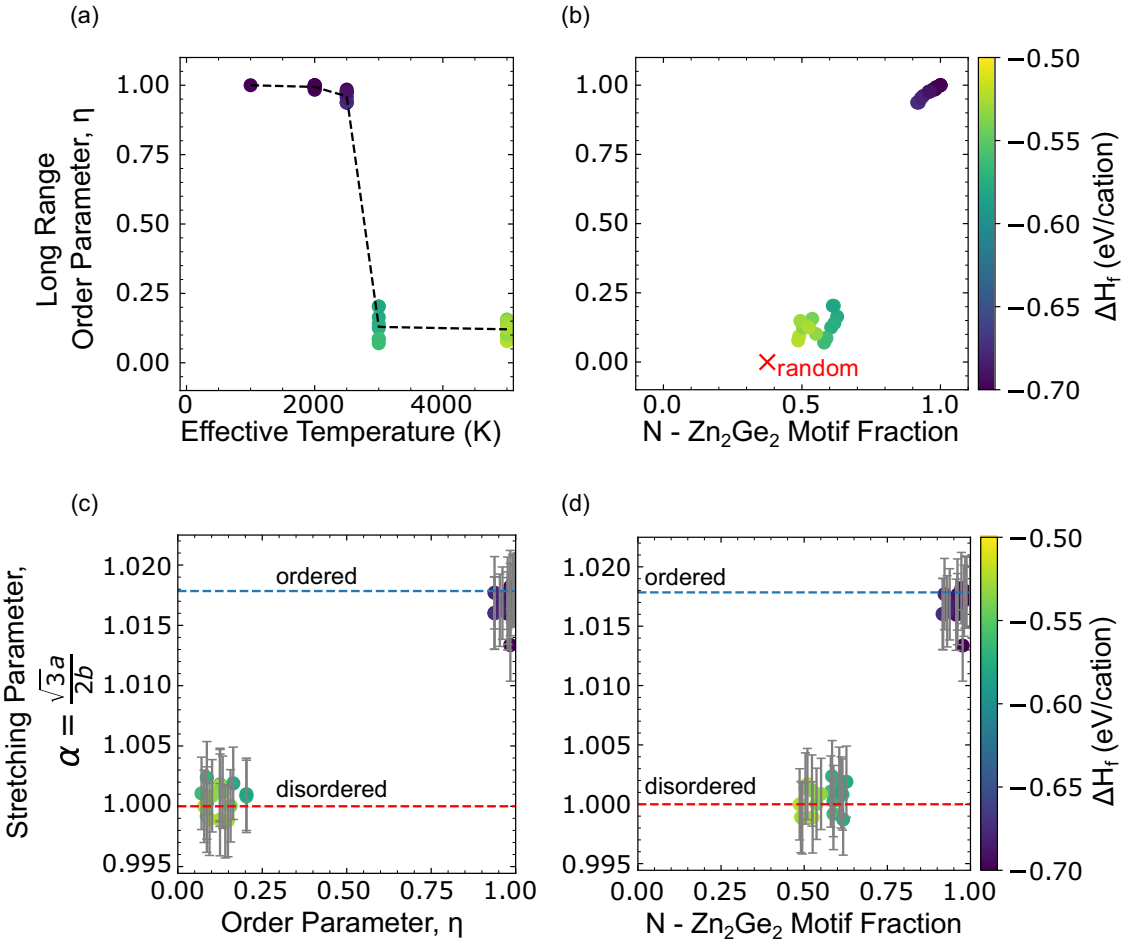

FIG. 5. Long-range order parameter (η) of supercells as a function of (a) T_{eff} with dashed line as a guide to the eye and (b) SRO parameter (motif fraction). The color scale indicates enthalpy ΔH_{f} of the individual configurations calculated from DFT. The red cross ($\times$) corresponds to a statistically random configuration. Stretching parameter as a function of (c) the LRO parameter η and (d) motif fraction of N-Zn_2Ge_2. Dashed lines show the stretching parameter of fully ordered ($\alpha = 1.018$) and fully disordered ($\alpha \equiv 1$) structures.

VII. STRETCHING PARAMETER

The LRO parameter η has not been measured and reported for $ZnGeN_2$ to date. However, a separate LRO parameter, the stretching parameter, which can be determined from Bragg angles without Rietveld refinement, has been measured for $ZnGeN_2$ in Refs. [10,77–79]. The stretching parameter measures lattice distortion of a structure mapped onto a specific crystallographic system such as a c/a ratio in a chalcopyrite-derived system, which in turn provides information on the ordering in the zinc-blende system [80,81]. In the present wurtzite-derived system, the orthorhombic primitive cell accommodates the cation ordering of the $ZnGeN_2$ ground state within the framework of the wurtzite lattice. The ordering breaks the symmetry of the wurtzite lattice, and atomic relaxation distorts the ratio of the orthorhombic lattice parameters a and b away from those of the ideal wurtzite lattice. On the other hand, in the disordered system, which strictly speaking has no symmetry, the average lattice recovers the ideal a/b ratio of wurtzite *as probed by diffraction measurements*.

In experimental characterization, the stretching parameter is also one of the most readily accessible order parameters to obtain, coming directly from a Pawley refinement [82] of a diffraction experiment. Reference [10] defines a stretching parameter for wurtzite-derived, orthorhombic systems based on the orthorhombic lattice parameters a and b and measures the stretching parameter for $ZnGeN_2$ grown at temperatures around the order-disorder transition. We use the stretching parameter from Ref. [10] here:

$$\alpha = \frac{\sqrt{3}a}{2b}, \tag{11}$$

where a and b are the orthorhombic lattice parameters of the system and $\frac{\sqrt{3}}{2}$ is the factor needed to map an orthorhombic system onto a hexagonal one given ideal hexagonal symmetry. $ZnGeN_2$ realizes this hexagonal structure for the fully disordered system, sometimes visualized as 1:1 shared occupancy of Zn and Ge on cation sites ($\eta = 0$). This disordered state therefore corresponds to $\alpha = 1$. For clarity, we use the convention $c_{or} = c_{hx}$ and $a_{or} > b_{or}$ for *or* orthorhombic and *hx* hexagonal lattice parameters. Our calculated value of the stretching parameter for the $ZnGeN_2$ ground-state structure is $\alpha = 1.018$, compared to measured values of $1.015 < \alpha < 1.025$ [10,77–79]. Calculated stretching parameters are taken

by fully relaxing (lattice vectors and atomic positions) the 1,024 atom supercells from MC in DFT. This relaxation also results in a volume difference between ordered and disordered cells where structures at $T_{\text{eff}} = 5,000$ K are up to 0.9% larger in volume than those at $T_{\text{eff}} = 2,000$ K. Figures 5(c) and 5(d) show the relationship between α and η and N-Zn_2Ge_2, respectively, with ΔH_{f} provided by the color scale. Partially ordered systems for which $0 < \eta < 1$ fall somewhere in-between the values of ordered and disordered stretching parameters. The uncertainty in stretching parameter (± 0.003) shown in Figs. 5(c) and 5(d) is attributed to the lack of a unique unit cell and orientation for disordered structures, an inherent limitation of the stretching parameter. This uncertainty reflects the potential change in stretching parameter due to rotating a cell to another possible reference orientation. Within this range, all supercells fall between the ordered $\alpha = 1.018$ and disordered $\alpha \equiv 1$ with the characteristic gap coinciding with the unsampled order-disorder transition in Fig. 4(a).

Comparing the merits of α to those of η and motif fraction, α is more straightforward to measure experimentally, but provides less detailed information on ordering. All three parameters identify the sharp order-disorder transition in $ZnGeN_2$. The motif fraction and η deviate slightly from the ideal structure before the transition, indicating that defectlike antisite formation precedes the phase transition. As discussed above, a certain sensitivity of these parameters to the degree of disorder (T_{eff}) exists also above the transition temperature. In contrast, α provides only a binary analysis of order, without systematic variation among the structures below or above the transition [see Figs. 5(c) and 5(d)]. Therefore, the stretching parameter α is useful in characterizing order versus disorder in a relevant material but, at least within our set of MC structures, it does not provide the resolution to determine the degree of ordering within a sample.

VIII. RELATIONSHIP BETWEEN T_{eff} AND T_{dep}

The effective temperature model relies on a mapping of nonequilibrium atomic configurations onto a pseudothermodynamic free-energy minimum, thereby implicitly assuming a Boltzmann distribution of microstates. Strictly speaking, there is no guarantee for such a distribution under nonequilibrium situations. However, according to the Bell-Evans-Polanyi principle [83,84], higher-energy states also have faster decay rates, so that an approximate Boltzmann-type distribution can result from partial equilibration. It is important to note, however, that a single T_{eff} describes at best one equilibration mechanism over a certain length scale, here, cation ordering involving short-range atomic diffusion on the order of a few nanometers.

Experimentally, the cation ordering of the SG33 ground state of $ZnGeN_2$ is not observed for typical growth temperatures up to 900 K for thin films grown by sputtering [28] or 800 K by molecular beam epitaxy [29], but it has been observed in growth and annealing experiments at higher temperatures. Specifically, the analysis of the stretching parameter after vapor-liquid-solid growth and from literature data in Ref. [10] demonstrated that the cation ordering occurs within an interval between 1050 and 1200 K. This transition is evidently a nonequilibrium effect and not a thermodynamic

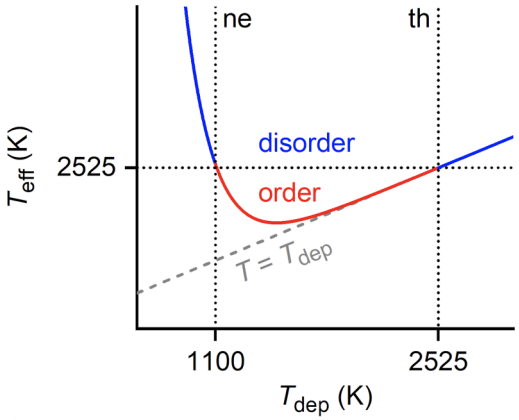

FIG. 6. Schematic illustration of the dependence of T_{eff} (measuring the degree of disorder) on the growth or deposition temperature T_{dep}. Experimentally, the transition proceeds from disorder to order at $T_{\text{dep}} \approx 1100$ K [10] and in Monte Carlo simulations from order to disorder at $T_{\text{eff}} = 2525$ K. Vertical lines indicate the nonequilibrium (ne) and the hypothetical thermodynamic (th) order-disorder transition.

transition because the material is disordered below and ordered above the transition temperature. On the other hand, in the MC simulations, the order-disorder phase transition occurs in a very narrow interval of effective temperatures, 2520 K $< T_{\text{eff}} <$ 2530 K. Considering the above-described nonmonotonic scaling between T_{eff} and the growth temperature T_{dep}, with an inverse relationship in the nonequilibrium regime below the transition, we can associate $T_{\text{eff}} = 2525$ K with the deposition temperature $T_{\text{dep}} = 1100$ K at which (on average between different experiments) ordering is observed [10].

Figure 6 schematically illustrates the relationship between T_{eff} and T_{dep}, with a nonequilibrium transition from disorder to order with increasing T_{dep} at 1100 K, and a hypothetical thermodynamic transition from order to disorder at $T_{\text{dep}} = T_{\text{eff}} = 2525$ K. Within our model, the Zn/Ge cations are randomly distributed in the low-temperature limit of T_{dep}, corresponding to $T_{\text{eff}} > 400\,000$ K from the Monte Carlo simulations as discussed above. The difference $T_{\text{eff}} - T_{\text{dep}}$, accounting for nonequilibrium effects, is gradually reduced with increasing T_{dep} as atomic diffusion allows for partial equilibration of the cation configuration. At the nonequilibrium transition temperature ($T_{\text{dep}} = 1100$ K in Fig. 6), the ordered phase occurs, but the equilibration of the cation distribution is not necessarily complete. For example, the concentration of antisite defects in the ordered phase [cf. Figs. 5(a) and 5(b)] is likely to exceed its thermodynamic value since T_{eff} remains larger than T_{dep} until T_{dep} significantly exceeds the transition temperature (cf. Fig. 6).

We emphasize again that the effective temperature concept applies only to free-energy contributions from configurational disorder in the cation arrangement. On the other hand, the actual temperature governs lattice vibrations and ideal-gas contributions to the chemical potentials, which determine the phase stability relative to melting and decomposition into the metals and molecular N_2. The decomposition is primarily driven by the ideal-gas contribution to the N chemical potential, offsetting the heat of formation from the elements. Our calculated formation enthalpy of $\Delta H_{\text{f}} = -0.70$ eV/cat

lies between that of InN (−0.30 eV/cat) and GaN (−1.63 eV/cat) [85], which is consistent with the observation that the decomposition of $ZnGeN_2$ becomes difficult to avoid above 1200 K [10]. These thermodynamic properties preclude the observation of the thermodynamic order-disorder transition in $ZnGeN_2$ ($T_{dep} = 2525$ K in Fig. 6).

While we did not observe intermediate values of the stretching parameter α in our simulation, the experimental data from [10] does show intermediate values, although only within a window less than 100 K wide, thereby seemingly indicating a more gradual transition. It should be noted that in nonequilibrium synthesis, such intermediate structures could occur, due to two reasons: first, the mapping of the nonequilibrium statistics of realized configurations onto an effective temperature model may not be perfect and the transition may occur less abruptly as a function of process parameters through kinetic trapping. Second, in polycrystalline films, not all grains may undergo the phase transition simultaneously, leading to an averaging of the order parameter between ordered and disordered grains. These caveats should be considered when formulating expectations for experimental results based on the present simulations. Additional insights could be gained from measurement of the evolution of the LRO parameter η or the SRO parameter around the experimental ordering temperature, although this would be a challenging experiment as mentioned previously.

Our effective temperature model is a first step toward the mapping between experimental nonequilibrium process parameters and a statistical, computational model of disorder. In $ZnGeN_2$, this mapping is currently based on three data points, i.e., the random disorder at $T_{eff} = 400{,}000$ K, the order-disorder transition at $T_{eff} = 2{,}525$ K, and the experimental ordering temperature around $T_{dep} = 1,100$ K depending on the synthesis method. The resulting relationship between T_{eff} and T_{dep}, as shown in Fig. 6, is therefore rather schematic, or perhaps semi-quantitative. Nevertheless, we obtain explicit atomic structure models that can be associated with different experimental situations. These structures will allow property predictions for disordered materials and hopefully enable the deliberate design of the materials properties utilizing the disorder degree of freedom.

IX. CONCLUSIONS

We presented a Monte Carlo study of cation disorder in $ZnGeN_2$ employing a cluster expansion for the total energy expression. The effective temperature used in the MC statistics connects to the deposition or growth temperature in experiment in a nonmonotonic fashion with an inverse relationship in the nonequilibrium regime and converging above the temperature where cation order equilibrates thermodynamically (above 1100 K in experiments for $ZnGeN_2$). The first-order order-disorder phase transition at $T_{eff} \approx 2500$ K lies in the nonequilibrium regime. We extracted SRO and LRO parameters from the MC-sampled configurations and examined their correlation. All order parameters drop simultaneously and sharply at the first-order transition. Significant nonrandomness persists in both SRO and LRO parameters above the transition, and complete randomness is approached only at very high effective temperatures in excess of 100 000 K.

While the SRO and LRO are closely intertwined, they have different merits and behaviors. While experimentally most straightforward and accessible, the stretching parameter α is conceptually somewhat limited because it relies on a mapping of the underlying wurtzite lattice onto an orthorhombic symmetry. This mapping is unambiguous only in the ordered ground state or the fully random limit, but not necessarily so in intermediate nonrandom disordered configurations. In our calculations, the stretching parameter offers only binary information, assuming either the value of the orthorhombic ordered ground state below the transition or unity (wurtzitelike) above the transition, without any meaningful resolution for intermediate configurations. The LRO (Bragg-Williams) parameter η and the SRO parameter (ideal N motif fraction) are correlated, exhibiting a small but significant reduction towards the transition, resulting from formation of Zn_{Ge} and Ge_{Zn} antisite defects in an otherwise largely ordered matrix. Despite the large drop at the transition, they still deviate considerably from the random limit, which is only gradually approached with increasing T_{eff}. While these effects are clearly observable in the simulation, they are nevertheless rather subtle.

Given experimental limitations such as limited sample quantities in thin films and the small scattering contrast between Zn and Ge, a direct experimental comparison will be challenging. Therefore, a particularly valuable conclusion of our work is the absence of octet rule-conserving disorder, which is a known feature of the related $ZnSnN_2$ nitride. In $ZnGeN_2$, in contrast, absence of LRO necessitates short-range disorder, which must be expected to affect the electronic structure, e.g., in the form of a band-gap reduction. This study provides a set of $ZnGeN_2$ supercell structures with systematic variation of the degree of disorder. These structures are available in the Supplemental Material [74] and facilitate the prediction of the properties of $ZnGeN_2$ as a function of disorder.

X. DETAILS FOR DFT CALCULATIONS

For each cluster expansion training set we used DFT with the generalized gradient approximation (GGA), Perdew-Burke-Ernzerhof (PBE) [86] type to relax structures and calculate formation energies. For the training of the cluster expansion, $ZnGeN_2$ unit cells and supercells were relaxed using a high-density k mesh with roughly 3000 k points per reciprocal atom. Kresse-Joubert projector augmented wave data sets (VASP version 4.6 pseudopotentials) were used for these calculations (i.e., Ge_d, N_s, and Zn) [87]. N_s is a soft type of pseudopotential that allows calculation with lower energy cutoff, which is beneficial for the feasibility of the large supercells used in this work. Extensive testing has shown that in most compounds, the N_s pseudopotential is fully suitable, unless bond distances are unusually short as in the case of molecular N_2 where even the standard N pseudopotential introduces significant errors and a hard (N_h) pseudopotential is needed. A total energy difference below 10^{-5} eV per supercell and total force tolerance of 0.02 eV $Å^{-1}$ on each atom served as convergence criteria for ionic relaxation. An energy cutoff of 380 eV was used along with a Coulomb

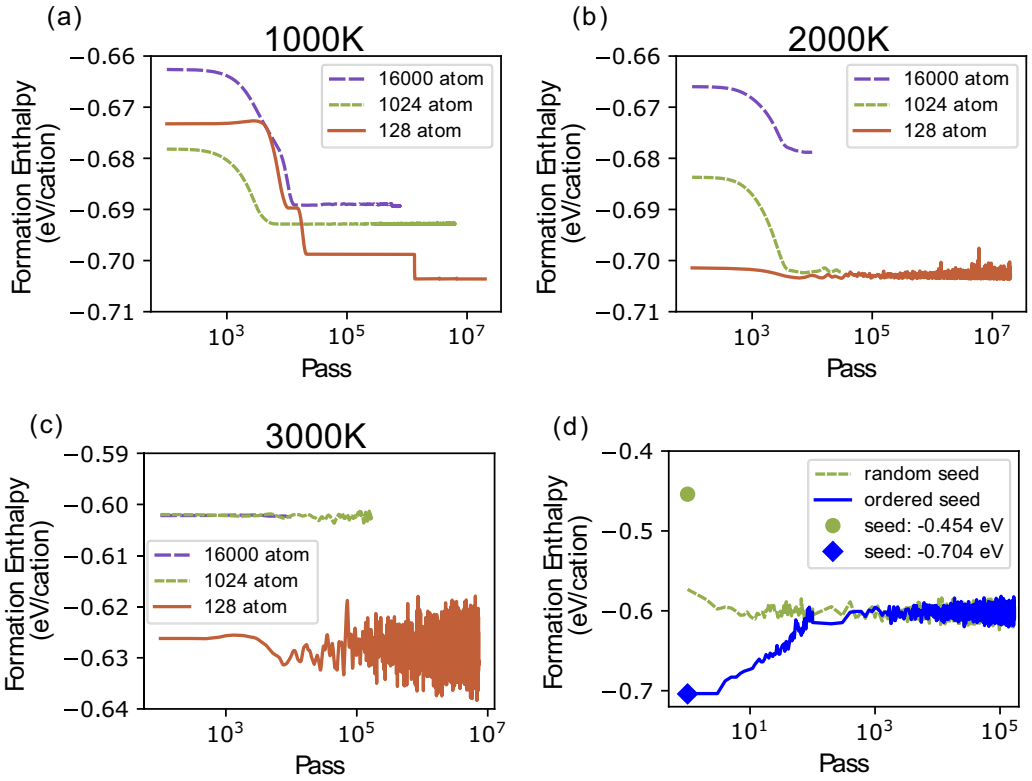

FIG. 7. Supercell size effects on convergence of formation enthalpy at (a) 1000 K, (b) 2000 K, and (c) 3000 K, comparing sizes of 16 000 (long dashed purple), 1024 (short dashed green), and 128 (solid orange) atoms. (d) Convergent behavior of formation enthalpy during equilibration at 3000 K for a 1024-atom simulation cell from random (dashed green) and ordered (solid blue) seed configurations.

potential $U - J = 6$ eV, applied to the Zn d orbital following the Dudarev approach [88]. Formation enthalpy was calculated using fitted elemental-phase reference energy (FERE) [64] chemical potentials of -0.50 eV for Zn, -4.14 eV for Ge, and -8.51 eV for N [18,65].

A set of supercells was selected at specific effective temperatures from the MC simulation to relax with DFT. These supercells contained 1024 atoms and DFT calculations followed the details above except that the gamma-point-only version of VASP was used along with a $1 \times 1 \times 1$ k mesh. These relaxed supercells were used for comparing formation enthalpies and stretching parameters described under Secs. V and VI. The unrelaxed POSCAR files for these calculations and used in the calculation of motif fractions and η can be found in the Supplemental Material [74].

ACKNOWLEDGMENTS

This work was supported by the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308 with the Alliance for Sustainable Energy, LLC, the manager and operator of the National Renewable Energy Laboratory. The funding was provided by the Office of Science, Basic Energy Sciences, Materials Sciences and Engineering Division. S.L. acknowledges support from the Energy Frontier Research Center Center for Next Generation of Materials Design. This work used high-performance computing resources located at NREL and sponsored by the Office of Energy Efficiency and Renewable Energy. The views expressed in the paper do not necessarily represent the views of the DOE or the U.S. government. The U.S. government retains and the publisher, by accepting the article for publication, acknowledges that the U. S. government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work, or allow others to do so, for government purposes.

APPENDIX: MONTE CARLO CONVERGENCE

Convergence tests for the MC simulation were run for effective temperatures of 5000, 3000, 2000, 1000, 500, 300, 200, and 100 K. Although the number of swaps in a pass increases linearly with cell size, the statistics of swapping cations to reach a converged state (reach a stable enthalpy as a function of passes) cause the number of passes needed to converge to grow exponentially with cell size. Too small of a supercell will never clearly equilibrate at high temperature due to the drastic enthalpy change associated with a single cation swap. Likewise, too large of a supercell prevents the structure from equilibrating at low temperatures due to the large number of swaps needed to reach a low-energy state and the low probability of these swaps occurring.

Supercells of 128, 1024, and 16 000 atoms were generated from the 16-atom primitive cell such that the ordered ground

state could be accommodated. Figure 7 shows a comparison of ΔH_f versus number of passes for the three cell sizes at the three different temperatures of 1000, 2000, and 3000 K. Passes were averaged over five, three, and one run for 128-, 1024-, and 16 000-atom supercells, respectively. Figure 7 shows convergence from random seeds with initial formation enthalpies of roughly −440 meV/cat. In the 1024 cell, MC fully equilibrates for 2000 K and higher. In Fig. 7(a), only the 128-atom cell reaches the ground state (700 meV/cat) at 1000 K. In Fig. 7(b), the 1024-atom cell also reaches the ground state at 2000 K, but a finite number of antisite defects occur and in Fig. 7(c), the temperature (3000 K) is above the phase transition; at 3000 K both 1024- and 16 000-atom cells converge, but 128-atom cells exhibit moderate (∼25 meV/cat) finite-size artifacts.

Figure 7(d) shows the equilibration period for MC simulations starting from ordered and random seed configurations consisting of 1024 atoms at 3000 K. For equilibration to be complete, the energies of structures initiated from the random and ordered seed configurations must converge. Apart from equilibrations shown in Fig. 7, some seeds were cooled from 1 000 000 to 100 000 K in 10 000-K intervals, 100 000 to 10 000 K in 1000-K intervals, 10000 to 5000 K in 100-K intervals, and 5000 to 100 K in 10-K intervals with each temperature set point equilibrated according to separate energy convergence criteria enabled within CASM. For these simulations with variable temperature (used in Figs. 3 and 4), configurations are assumed to be converged once the change in formation enthalpy remained below 1 meV after 10 swaps. This method of testing convergence resulted in hundreds of thousands of passes at high temperature where configurations change rapidly in formation enthalpy and less than 1000 passes at low temperature where cation swaps are less likely to occur. Selected supercells were relaxed and analyzed on a basis of SRO and LRO as well as effective temperature to compare a proxy for processing conditions to the resultant degree of order within a fully crystalline system.

[1] V. Ozoliš, C. Wolverton, and A. Zunger, First-principles theory of vibrational effects on the phase stability of Cu-Au compounds and alloys, Phys. Rev. B **58**, R5897 (1998).

[2] M. H. F. Sluiter, C. Colinet, and A. Pasturel, *Ab initio* calculation of the phase stability in Au-Pd and Ag-Pt alloys, Phys. Rev. B **73**, 174204 (2006).

[3] G. Ghosh, A. van de Walle, and 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**, 3202 (2008).

[4] A. Anedda, L. Garbato, F. Raga, and A. Serpi, Photoconductivity and trap distribution in $CdIn_2S_4$, Phys. Status Solidi A **50**, 643 (1978).

[5] H.-G. Brühl, L. Hildisch, H. Morwinski, W. Schmidt, and E. Schubert, Diffuse x-ray scattering and compositional disorder in $Ga_xIn_{1-x}P$, Phys. Status Solidi A **39**, 133 (1977).

[6] M. Quennet, A. Ritscher, M. Lerch, and B. Paulus, The order-disorder transition in Cu_2ZnSnS_4: A theoretical and experimental study, J. Solid State Chem. **250**, 140 (2017).

[7] S. M. Wasim, G. Marn, R. Marquez, and C. Rincn, On the effect of structural disorders on the urbachs tails of ternary chalcopyrite semiconductors and related ordered defect compounds, J. Appl. Phys. **127**, 035703 (2020).

[8] W. Sun, C. J. Bartel, E. Arca, S. R. Bauers, B. Matthews, B. Orvañanos, B.-R. Chen, M. F. Toney, L. T. Schelhas, W. Tumas *et al.*, A map of the inorganic ternary metal nitrides, Nat. Mater. **18**, 732 (2019).

[9] Y. Hinuma, T. Hatakeyama, Y. Kumagai, L. A. Burton, H. Sato, Y. Muraba, S. Iimura, H. Hiramatsu, I. Tanaka, H. Hosono *et al.*, Discovery of earth-abundant nitride semiconductors by computational screening and high-pressure synthesis, Nat. Commun. **7**, 11962 (2016).

[10] E. W. Blanton, K. He, J. Shan, and K. Kash, Characterization and control of $ZnGeN_2$ cation lattice ordering, J. Cryst. Growth **461**, 38 (2017).

[11] N. L. Adamski, Z. Zhu, D. Wickramaratne, and C. G. Van De Walle, Strategies for p-type doping of $ZnGeN_2$, Appl. Phys. Lett. **114**, 32101 (2019).

[12] X. Cao, F. Kawamura, T. Taniguchi, and N. Yamada, Electron-transport properties of degenerate $ZnSnN_2$ doped with oxygen, BMC Mater. **2**, 1 (2020).

[13] R. A. Makin, K. York, S. M. Durbin, N. Senabulya, J. Mathis, R. Clarke, N. Feldberg, P. Miska, C. M. Jones, Z. Deng *et al.*, Alloy-Free Band Gap Tuning Across the Visible Spectrum, Phys. Rev. Lett. **122**, 256403 (2019).

[14] E. Arca, J. D. Perkins, S. Lany, A. Mis, B.-R. Chen, P. Dippo, J. L. Partridge, W. Sun, A. Holder, A. C. Tamboli *et al.*, Zn_2SbN_3: Growth and characterization of a metastable photoactive semiconductor, Mater. Horiz. **6**, 1669 (2019).

[15] A. L. Greenaway, A. L. Loutris, K. N. Heinselman, C. L. Melamed, R. R. Schnepf, M. B. Tellekamp, R. Woods-Robinson, R. Sherbondy, D. Bardgett, S. Bauers *et al.*, Combinatorial synthesis of magnesium tin nitride semiconductors, J. Am. Chem. Soc. **142**, 8421 (2020).

[16] P. C. Quayle, E. W. Blanton, A. Punya, G. T. Junno, K. He, L. Han, H. Zhao, J. Shan, W. R. L. Lambrecht, and K. Kash, Charge-neutral disorder and polytypes in heterovalent wurtzite-based ternary semiconductors: The importance of the octet rule, Phys. Rev. B **91**, 205207 (2015).

[17] R. R. Schnepf, J. J. Cordell, M. B. Tellekamp, C. L. Melamed, A. L. Greenaway, A. Mis, G. L. Brennecka, S. Christensen, G. J. Tucker, E. S. Toberer *et al.*, Utilizing site disorder in the development of new energy-relevant semiconductors, ACS Energy Lett. **5**, 2027 (2020).

[18] J. Pan, J. J. Cordell, G. J. Tucker, A. Zakutayev, A. C. Tamboli, and S. Lany, Perfect short-range ordered alloy with line-compound-like properties in the $ZnSnN_2$: ZnO system, npj Comput. Mater. **6**, 63 (2020).

[19] S. R. Bauers, A. Holder, W. Sun, C. L. Melamed, R. Woods-Robinson, J. Mangum, J. Perkins, W. Tumas, B. Gorman, A. Tamboli *et al.*, Ternary nitride semiconductors in the

rocksalt crystal structure, Proc. Natl. Acad. Sci. USA **116**, 14829 (2019).

[20] T. D. Veal, N. Feldberg, N. F. Quackenbush, W. M. Linhart, D. O. Scanlon, L. F. Piper, and S. M. Durbin, Band gap dependence on cation disorder in $ZnSnN_2$ solar absorber, Adv. Energy Mater. **5**, 1501462 (2015).

[21] K. Binder, Monte Carlo study of entropy for face-centered cubic Ising antiferromagnets, Z. Phys. B: Condens. Matter **45**, 61 (1981).

[22] B. Hamilton, A. R. Peaker, and S. T. Pantelides, Entropy-Driven Metastabilities in Defects in Semiconductors, Phys. Rev. Lett. **61**, 1627 (1988).

[23] A. M. Holder, S. Siol, P. F. Ndione, H. Peng, A. M. Deml, B. E. Matthews, L. T. Schelhas, M. F. Toney, R. G. Gordon, W. Tumas *et al.*, Novel phase diagram behavior and materials design in heterostructural semiconductor alloys, Sci. Adv. **3**, e1700270 (2017).

[24] B. Liu, A. Seko, and I. Tanaka, Cluster expansion with controlled accuracy for the MgO/ZnO pseudobinary system via first-principles calculations, Phys. Rev. B **86**, 245202 (2012).

[25] L. C. de Carvalho, A. Schleife, J. Furthmüller, and F. Bechstedt, Ab initio calculation of optical properties with excitonic effects in wurtzite $In_xGa_{1-x}N$ and $In_xAl_{1-x}N$ alloys, Phys. Rev. B **87**, 195211 (2013).

[26] L. Li, J. T. Muckerman, M. S. Hybertsen, and P. B. Allen, Phase diagram, structure, and electronic properties of $(Ga_{1-x}Zn_x)(N_{1-x}O_x)$ solid solutions from dft-based simulations, Phys. Rev. B **83**, 134202 (2011).

[27] K. Du, C. Bekele, C. C. Hayman, J. C. Angus, P. Pirouz, K. Kash, and K. Kash, Synthesis and characterization of $ZnGeN_2$ grown from elemental Zn and Ge sources, J. Cryst. Growth **310**, 1057 (2008).

[28] C. L. Melamed, M. B. Tellekamp, J. S. Mangum, J. D. Perkins, P. Dippo, E. S. Toberer, and A. C. Tamboli, Blue-green emission from epitaxial yet cation-disordered $ZnGeN_{2-x}O_x$, Phys. Rev. Materials **3**, 051602 (2019).

[29] M. B. Tellekamp, C. L. Melamed, A. G. Norman, and A. Tamboli, Heteroepitaxial integration of $ZnGeN_2$ on GaN buffers using molecular beam epitaxy, Cryst. Growth Des. **20**, 1868 (2020).

[30] S. Lany, A. N. Fioretti, P. P. Zawadzki, L. T. Schelhas, E. S. Toberer, A. Zakutayev, and A. C. Tamboli, Monte carlo simulations of disorder in $ZnSnN_2$ and the effects on the electronic structure, Phys. Rev. Materials **1**, 035401 (2017).

[31] E. Arca, S. Lany, J. D. Perkins, C. Bartel, J. Mangum, W. Sun, A. Holder, G. Ceder, B. Gorman, G. Teeter *et al.*, Redox-mediated stabilization in zinc molybdenum nitrides, J. Am. Chem. Soc. **140**, 4293 (2018).

[32] K. N. Heinselman, S. Lany, J. D. Perkins, K. R. Talley, and A. Zakutayev, Thin film synthesis of semiconductors in the Mg-Sb-N materials system, Chem. Mater. **31**, 8717 (2019).

[33] L. Zhu, P. Maruska, P. Norris, P. Yip, and L. Bouthillette, Epitaxial growth and structural characterization of single crystalline $ZnGeN_2$, Mater. Res. Soc. Symp. Proc. **537**, G3.8 (1998).

[34] D. Skachkov, P. C. Quayle, K. Kash, and W. R. L. Lambrecht, Disorder effects on the band structure of $ZnGeN_2$: Role of exchange defects, Phys. Rev. B **94**, 205201 (2016).

[35] D. Skachkov, A. Punya Jaroenjittichai, L.-y. Huang, and W. R. L Lambrecht, Native point defects and doping in $ZnGeN_2$, Phys. Rev. B **93**, 155202 (2016).

[36] S. Lyu, D. Skachkov, K. Kash, E. W. Blanton, and W. R. L. Lambrecht, Band Gaps, Band-Offsets, Disorder, Stability Region, and Point Defects in II-IV-N_2, Phys. Status Solidi A **216**, 1800875 (2019).

[37] N. L. Adamski, Z. Zhu, D. Wickramaratne, and C. G. Van de Walle, Hybrid functional study of native point defects and impurities in $ZnGeN_2$, J. Appl. Phys. **122**, 195701 (2017).

[38] A. Zunger, S.-H. Wei, L. G. Ferreira, and J. E. Bernard, Special Quasirandom Structures, Phys. Rev. Lett. **65**, 353 (1990).

[39] S.-H. Wei, L. G. Ferreira, and A. Zunger, First-principles calculation of the order-disorder transition in chalcopyrite semiconductors, Phys. Rev. B **45**, 2533 (1992).

[40] P. Zawadzki, A. Zakutayev, and S. Lany, Entropy-Driven Clustering in Tetrahedrally Bonded Multinary Materials, Phys. Rev. Appl. **3**, 034007 (2015).

[41] K. Biswas and S. Lany, Energetics of quaternary III-V alloys described by incorporation and clustering of impurities, Phys. Rev. B **80**, 115206 (2009).

[42] K. Kim and A. Zunger, Spatial Correlations in GaInAsN Alloys and Their Effects on Band-Gap Enhancement and Electron Localization, Phys. Rev. Lett. **86**, 2609 (2001).

[43] N. Tsunoda, Y. Kumagai, A. Takahashi, and F. Oba, Electrically Benign Defect Behavior in Zinc Tin Nitride Revealed from First Principles, Phys. Rev. Appl. **10**, 011001 (2018).

[44] D. De Fontaine, Cluster approach to order-disorder transformations in alloys, Solid State Phys. **47**, 33 (1994).

[45] J. M. Sanchez, F. Ducastelle, and D. Gratias, Generalized cluster description of multicomponent systems, Phys. A (Amsterdam) **128**, 334 (1984).

[46] A. Van De Walle and G. Ceder, The effect of lattice vibrations on substitutional alloy thermodynamics, Rev. Mod. Phys. **74**, 11 (2002).

[47] B. Puchala and A. Van Der Ven, Thermodynamics of the Zr-O system from first-principles calculations, Phys. Rev. B **88**, 094108 (2013).

[48] J. C. Thomas and A. Van Der Ven, Finite-temperature properties of strongly anharmonic and mechanically unstable crystal phases from first principles, Phys. Rev. B **88**, 214111 (2013).

[49] S. Hadke, W. Chen, J. M. R. Tan, M. Guc, V. Izquierdo-Roca, G.-M. Rignanese, G. Hautier, and L. H. Wong, Effect of Cd on cation redistribution and order-disorder transition in $Cu_2(Zn,Cd)SnS_4$, J. Mater. Chem. A **7**, 26927 (2019).

[50] A. R. Natarajan, J. C. Thomas, B. Puchala, and A. Van der Ven, Symmetry-adapted order parameters and free energies for solids undergoing order-disorder phase transitions, Phys. Rev. B **96**, 134204 (2017).

[51] M.-A. Nicolet and P. Giauque, Highly metastable amorphous or near-amorphous ternary films (mictamict alloys), Microelectron. Eng. **55**, 357 (2001).

[52] C. C. Tsuei and P. Duwez, Metastable amorphous ferromagnetic phases in palladiumbase alloys, J. Appl. Phys. **37**, 435 (1966).

[53] S. Chen, A. Walsh, J.-H. Yang, X. G. Gong, L. Sun, P.-X. Yang, J.-H. Chu, and S.-H. Wei, Compositional dependence of structural and electronic properties of $Cu_2ZnSn(S, Se)_4$ alloys for thin film solar cells, Phys. Rev. B **83**, 125201 (2011).

[54] H. Von Bardeleben, The chemistry of structural defects in $CuInSe_2$, Sol. Cells **16**, 381 (1986).

[55] G. Seryogin, S. Nikishin, H. Temkin, A. Mintairov, J. Merz, and M. Holtz, Order–disorder transition in epitaxial znsnp 2, Appl. Phys. Lett. **74**, 2128 (1999).

[56] C. M. Caskey, R. M. Richards, D. S. Ginley, and A. Zakutayev, Thin film synthesis and properties of copper nitride, a metastable semiconductor, Mater. Horiz. **1**, 424 (2014).

[57] W. L. Bragg and E. J. Williams, The effect of thermal agitation on atomic arrangement in alloys, Proc. R. Soc. London A **145**, 699 (1934).

[58] W. L. Bragg and E. J. Williams, The effect of thermal agitation on atomic arrangement in alloys II, Proc. R. Soc. London A **151**, 540 (1935).

[59] F. C. Nix and W. Shockley, Order-disorder transformations in alloys, Rev. Mod. Phys. **10**, 1 (1938).

[60] P. F. Ndione, Y. Shi, V. Stevanovic, S. Lany, A. Zakutayev, P. A. Parilla, J. D. Perkins, J. J. Berry, D. S. Ginley, and M. F. Toney, Control of the electrical properties in spinel oxides by manipulating the cation disorder, Adv. Funct. Mater. **24**, 610 (2014).

[61] J. Pan, J. Cordell, G. Tucker, A. Tamboli, A. Zakutayev, and S. Lany, Interplay between composition, electronic structure, disorder, and doping due to dual sublattice mixing in nonequilibrium synthesis of $ZnSnN_2$: O, Adv. Mater. **31**, 1807406 (2019).

[62] R. D. Shannon, Revised effective ionic radii and systematic studies of interatomic distances in halides and chalcogenides, Acta Crystallogr. Sect. A: Cryst. Phys., Diffr., Theor. Gen. Crystallogr. **32**, 751 (1976).

[63] A. Van Der Ven, J. C. Thomas, Q. Xu, and J. Bhattacharya, Linking the electronic structure of solids to their thermodynamic and kinetic properties, Math. Comput. Simul. **80**, 1393 (2010).

[64] S. Lany, Semiconductor thermochemistry in density functional calculations, Phys. Rev. B **78**, 245207 (2008).

[65] V. Stevanovic, S. Lany, X. Zhang, and A. Zunger, Correcting density functional theory for accurate predictions of compound enthalpies of formation: Fitted elemental-phase reference energies, Phys. Rev. B **85**, 115104 (2012).

[66] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, Scikit-learn: Machine Learning in Python, J. Machine Learning Res. **12**, 2825 (2011).

[67] K. Binder, D. M. Ceperley, J.-P. Hansen, M. Kalos, D. Landau, D. Levesque, H. Mueller-Krumbhaar, D. Stauffer, and J.-J. Weis, *Monte Carlo Methods in Statistical Physics* (Springer, New York, 2012), Vol. 7.

[68] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, Equation of state calculations by fast computing machines, J. Chem. Phys. **21**, 1087 (1953).

[69] S. Chib and E. Greenberg, Understanding the metropolis-hastings algorithm, Am. Stat. **49**, 327 (1995).

[70] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. Jarrod Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. Carey *et al.*, SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python, Nat. Methods **17**, 261 (2020).

[71] C. Caroli and K. Maki, Fluctuations of the order parameter in type-II superconductors. II. Pure limit, Phys. Rev. **159**, 316 (1967).

[72] K. Maki, The critical fluctuation of the order parameter in type-II superconductors, Prog. Theor. Phys. **39**, 897 (1968).

[73] F. Bitter, A generalization of the theory of ferromagnetism, Phys. Rev. **54**, 79 (1938).

[74] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevMaterials.5.024604 for unrelaxed POSCAR files resultant from Monte Carlo simulations in this article.

[75] B. E. Warren, *X-Ray Diffraction* (Dover, New York, 1990).

[76] S. Nakatsuka and Y. Nose, Order–disorder phenomena and their effects on bandgap in $ZnSnP_2$, J. Phys. Chem. C **121**, 1040 (2017).

[77] T. Endo, Y. Sato, H. Takizawa, and M. Shimada, High-pressure synthesis of new compounds, $ZnSiN_2$ and $ZnGeN_2$ with distorted wurtzite structure, J. Mater. Sci. Lett. **11**, 424 (1992).

[78] R. Viennois, T. Taliercio, V. Potin, A. Errebbahi, B. Gil, S. Charar, A. Haidoux, and J.-C. Tédenac, Prospective investigations of orthorhombic $ZnGeN_2$: Synthesis, lattice dynamics and optical properties, Mater. Sci. Eng., B **82**, 45 (2001).

[79] M. Shang, J. Wang, J. Fan, H. Lian, Y. Zhang, and J. Lin, $ZnGeN_2$ and $ZnGeN_2$: Mn^{2+} phosphors: Hydrothermal-ammonolysis synthesis, structure and luminescence properties, J. Mater. Chem. C **3**, 9306 (2015).

[80] J. E. Jaffe and A. Zunger, Theory of the band-gap anomaly in AB C_2 chalcopyrite semiconductors, Phys. Rev. B **29**, 1882 (1984).

[81] R. R. Schnepf, B. L. Levy-Wendt, M. B. Tellekamp, B. R. Ortiz, C. L. Melamed, L. T. Schelhas, K. H. Stone, M. F. Toney, E. S. Toberer, and A. C. Tamboli, Using resonant energy x-ray diffraction to extract chemical order parameters in ternary semiconductors, J. Mater. Chem. C **8**, 4350 (2020).

[82] W. Cochran and G. S. Pawley, The theory of diffuse scattering of x-rays by a molecular crystal, Proc. R. Soc. London A **280**, 1 (1964).

[83] R. P. Bell, The theory of reactions involving proton transfers, Proc. R. Soc. London A **154**, 414 (1936).

[84] M. G. Evans and M. Polanyi, Some applications of the transition state method to the calculation of reaction velocities, especially in solution, Trans. Faraday Soc. **31**, 875 (1935).

[85] M. Ranade, F. Tessier, A. Navrotsky, and R. Marchand, Calorimetric determination of the enthalpy of formation of InN and comparison with AlN and GaN, J. Mater. Res. **16**, 2824 (2001).

[86] J. P. Perdew, K. Burke, and M. Ernzerhof, Generalized Gradient Approximation Made Simple, Phys. Rev. Lett. **77**, 3865 (1996).

[87] G. Kresse and D. Joubert, From ultrasoft pseudopotentials to the projector augmented-wave method, Phys. Rev. B **59**, 1758 (1999).

[88] S. L. Dudarev, G. A. Botton, S. Y. Savrasov, C. J. Humphreys, and A. P. Sutton, Electron-energy-loss spectra and the structural stability of nickel oxide: An LSDAU study, Phys. Rev. B **57**, 1505 (1998).