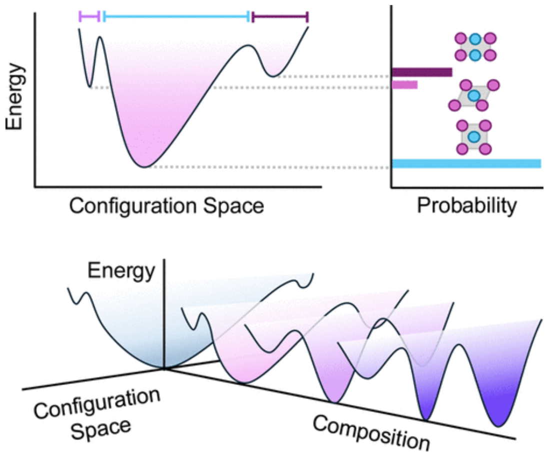
Basin-Size Mapping: Prediction of Metastable Polymorph Synthesizability Across TaC–TaN Alloys

Andrew Novick, Matthew Jankousky, Eric Toberer, and Vladan Stevanovic
Colorado School of Mines, Golden, Colorado 80401, USA
Quan Nguyen
Washington University in St. Louis, St. Louis, Missouri 63130, USA
Brooks Tellekamp
National Renewable Energy Laboratory, Golden, Colorado 80401, USA

Abstract

The sizes of the basins of attraction on the potential energy surface are helpful indicators in determining the experimental synthesizability of metastable phases. In principle, these basins can be controlled with changes in thermodynamic conditions such as composition, pressure, and surface energy. Herein, we use random structure sampling to computationally study how alloying smoothly perturbs basin of attraction sizes. The $TaC_{1-x}N_x$ pseudobinary is an ideal test system given the structural and polymorphic contrast of its parent compounds and their technological relevance as epitaxial substrates for $Al_{1-x}Ga_xN$. While we find limited thermodynamic stability across all computationally observed phases, random structure sampling shows a significant composition region where the rocksalt basin dominates. As such, we predict the potential for the nonequilibrium synthesis of metastable rocksalt $TaC_{1-x}N_x$ alloys as substrates for $Al_{1-x}Ga_xN$. At higher nitrogen concentrations, other low-energy metastable polymorphs emerge that continue to retain the hexagonal close packing suitable for III–N growth. Confidence in these trends was established through uncertainty quantification of the basin sizes and energy distributions; such analysis utilized the Beta and Dirichlet distributions. We also find (a) polymorph basin sizes can be rationalized in terms of energetic preferences for different coordination environments; and (b) basin sizes universally shrink with increasing nitrogen content, making the system more prone to amorphous growth.



Introduction

Many critical synthetic processes are conducted far from thermodynamic equilibrium and produce metastable materials. (1,2) Most commonly, these syntheses involve either irreversible bond formation or rapid quenching, as seen in physical vapor deposition techniques. (3) Computational approaches to understanding metastability often focus on the energy with respect to the convex hull. (4) While energetics can provide useful bounds on the limits of metastability, differentiating between structures that fall within such bounds can be challenging. (1) Indeed, low-energy structures are not necessarily realizable, (5) and high-energy structures cannot be automatically discarded. (6)

From soft matter to astronomy, basins of attraction serve as a natural compromise for succinctly describing potential energy surfaces without reducing them solely to the energy of their local minima. (7–11) Recent work has highlighted the basin of attraction size as a factor in determining polymorph metastability. (1,12–18) The boundaries of each basin are defined such that any atomic configuration within the basin will fall into the energetic minimum of that basin upon structural relaxation. As such, the relative basin sizes represent the probability for a random configuration to relax into a given structure. These probabilities can be used to rank the experimental realizability of different metastable polymorphs. (1,12,18) The basins of attraction and their corresponding probabilities are illustrated for a hypothetical system in Figure 1a,b.

Figure 1

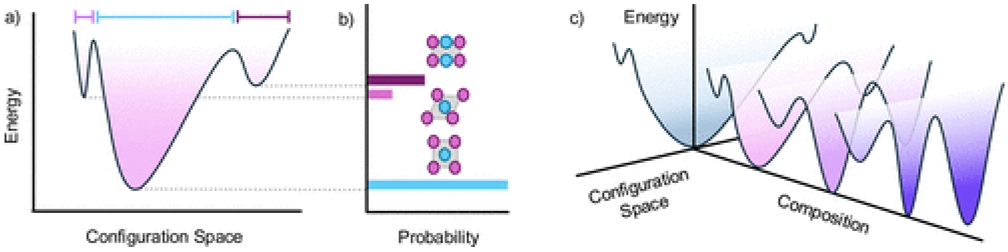


Figure 1. (a) Basins of attraction within configuration space lead to distinct, dynamically stable phases. The relative size of each basin determines (b) the probability for a random state to structurally relax into that phase. (c) This work investigates how basins of attraction can be controlled through composition modulation.

Still, there remains an open scientific question of what determines the size of the basin of attraction and how such sizes can be rationally controlled. Success would provide critical access to synthetic navigation of metastable polymorphs. Figure 1c sketches the hypothesized change in basin sizes with composition. Herein, we use alloying to computationally study the effects of chemical composition on basin sizes, and thus polymorph selection, in the TaCxN1–x pseudobinary alloy.

Calculating basin sizes is nontrivial since they are relative quantities–the proportion of configuration space taken up by a single basin is dependent on the other competing basins. As such, an unbiased sampling of the potential energy surface is required to estimate basin sizes. Our recent work employs random structure generation and repeated relaxations (19) to explore the potential energy surface and the basins therein. (12) Of particular relevance is the study of ZnZrN2, which compared the efficacy of various computational techniques including Special Quasi-random Structures, vibrational free energies, and random structure sampling. (16) Ultimately, only random structure sampling was able to explain the growth of a metastable phase due to its relatively large basin size.

While previous studies gave insight into thermodynamic basins and phase selection, they focused on individual chemistries, or a couple of distinct compounds. (1) A systematic characterization of basin size under gradual condition changes is necessary if we are to understand and control basin size. Here, we use alloying as an ideal tool for such a systematic study; by finely varying composition, we perturb the potential energy surface and finally observe the resulting trends in basin sizes.

TaCxN1–x is an ideal system to study metastability and polymorphism due to the differences in carbon and nitrogen bonding. In particular, the contrast in the observed polymorphism of the parent compounds provides an interesting path to explore basin sizes; TaN exhibits rich polymorphism, (20−22) while TaC has only been realized in the rocksalt structure. (23,24) Technologically, a single-phase TaCxN1–x rocksalt alloy would be of use as a substrate for Al1–xGaxN growth. (24) Finally, TaCxN1–x serves as a useful case example, providing insight into the broader material classes of carbo-nitride alloys and metallic nitrides, both of which are often grown through nonequilibrium methods. These classes have a variety of applications including opteoelectronics, ultrawide bandgap devices, and electrochemical energy storage. (2,25) As one might expect, understanding their synthesis remains an ongoing area of research.

Results and Discussion

Parent-Compound Basins

We will start our discussion with TaC since it is the simplest member of the TaCxN1–x system. In Figure 2a, we show the observed energies collected from ∼1000 random structure samples. Most of the randomly initialized structures relax into high-energy local minima with little structural symmetry. Only 7% of the samples finished with a spacegroup larger than 10, and 9% of samples relaxed into minima less than 0.2 eV/atom from the ground state. These results are typical in random structure sampling (17) and motivate the need for robust uncertainty quantification when drawing conclusions about basin sizes.

Figure 2

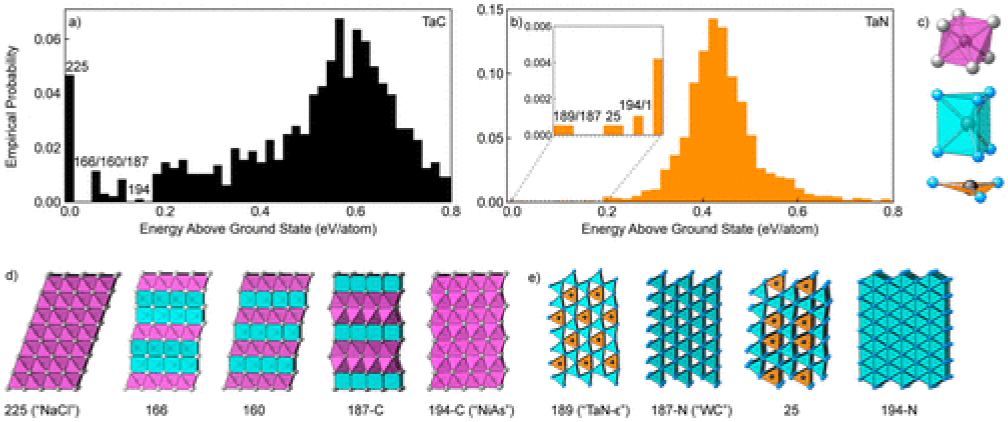


Figure 2. (a) The relaxation of Ta12C12 random structures yields diverse array of atomic configurations. Within 200 meV of the ground state, five TaC phases are repeatedly found and identified by space group here. (b) For the nitride analog, far fewer structures are found near the ground state. (c) The low-energy structures of TaC and TaN are comprised of octahedral, prismatic, and trigonal planar polyhedra. (d) For TaC, the five lowest-energy structures are built from slabs of octahedral and trigonal prismatic polyhedra. (e) Low-energy TaN structures share these prismatic polyhedra slabs building blocks; for some phases, a minority of sites have been replaced with trigonal planes. All polyhedra are represented with Ta as the central atom.

Despite the large number of local minima, random structure sampling finds five low-energy, high-symmetry structures, with rocksalt being the ground state. Moreover, the rocksalt structure is found to have the largest basin, occurring 45 times. For context, the next most popular structure, spacegroup 160 (SG 160), only occurs 10 times. The combination of rocksalt being the ground state and having the largest basin size is consistent with TaC only being experimentally realized in the rocksalt structure. (23,24) The absence of polymorphism for compounds with a rocksalt ground state fits within a broader trend as well. (26)

In Figure 2d, we visualize the five lowest energy TaC polymorphs found by random structure sampling. Interestingly, all five polymorphs are 6-fold coordinated and made up the same two underlying polyhedra: trigonal prisms (turquoise) and octahedra (magneta), as illustrated in Figure 2c. Both trigonal prisms and octahedra have equilateral triangles of carbon; the shared structural motif presents itself as an ideal interface for stacking the two polyhedra. It is therefore not surprising that three of the five low-energy structures exhibit alternating layers of trigonal prisms and octahedra. Furthermore, the basin sizes of structures other than rocksalt follow the following trend: the structures mixing octahedra and trigonal prisms (SG 166, 160, 187-C) occur next most frequently, while the structure with only octahedra arranged hexagonally (SG 194-C) occurs infrequently. As such, basin size is not only determined by the local polyhedra, but how they are arranged.

The random structure sampling outcome for TaN differs significantly from that of TaC. In Figure 2b, the ground state and other low-energy structures (SG 189, 187, 25, 194) are barely visible in the energy distribution, each occurring only once. In fact, of the ~2000 random structure samples, only 0.3% relaxed into a structure within 0.2 eV/atom of the ground state. The experimental ground state of TaN is the SG 189 structure (TaN-ϵ), (27,28) and SG 187 has been produced under high pressures. (21,29) Even with narrow basins, both structure-types were independently found through random structure sampling and their energetic ordering is in alignment with experiment.

In Figure 2e, the four lowest energy polymorphs are shown. Like TaC, these structures are all a combination of just two underlying polyhedra, except here we have trigonal planes and trigonal prisms (panel c). TaN-ϵ structure is found to be the ground state and adopts a mixture of trigonal prisms and trigonal planes such that the average coordination of Ta and N is 5-fold. However, the SG 187 structure serves as a helpful counterexample, showing that 6-fold coordination only results in a relatively small energetic penalty (0.033 eV/atom). Looking across these two parent compounds, we see building blocks that are both distinct and shared. Trigonal prisms are shared across parent compounds, indicating a possibility for energetic favoring in the alloys.

Typically, random structure sampling shows the observed ground state has the largest basin of attraction, as with TaC. Indeed, the inverse correlation between energy and basin size has been noted. (19,30–32) However, TaN serves as a useful counterexample where its ground state has an exceptionally narrow basin. If the lack of large basins were due to the structure generation method, then we would have observed similarly small basins in TaC, as well as other systems where the structure generation code was used (e.g., PbTe, SnO2). (12,26) However, since we see low-energy structures corresponding to large basins in those systems but not in TaN, we conclude that the TaN basin sizes are indicative of its underlying potential energy surface. Furthermore, the TaN result is in agreement with a broader trend of metallic nitrides having narrow ground state basins. (16,18)

The large number of observed low-symmetry structures can be interpreted as TaN having high a propensity to form a glass. (33,34) Moreover, a common indicator of glass forming ability is in the narrow energetic range of configurational states, (35) which is the case for low-symmetry structures exhibited by TaN, but not TaC. The existence of both the low-energy minimum and a large set of energetically similar low-symmetry states helps to explain why TaN is commonly found in the literature as either an amorphous material (36,37) or one of three ordered compounds. (20,21) In contrast, the ordered rocksalt structure is the only TaC phase we find in the literature. (23,24)

Alloy Energetics and Structures

Having established the basin sizes and structures found for TaC and TaN, we now have the requisite context to study their alloys. In Figure 3a, we show the empirical probability for a random structure to relax into a given energy range, all as a function of composition. Each distribution is built from a significant number of 24-atom cell relaxations, ranging from ~1000 to 2000. The number of random structures is determined by the need to produce sufficiently small uncertainties in the basin sizes. All energies are expressed relative to the convex hull, which in this case is simply a linear combination of the parent ground state energies.

Figure 3

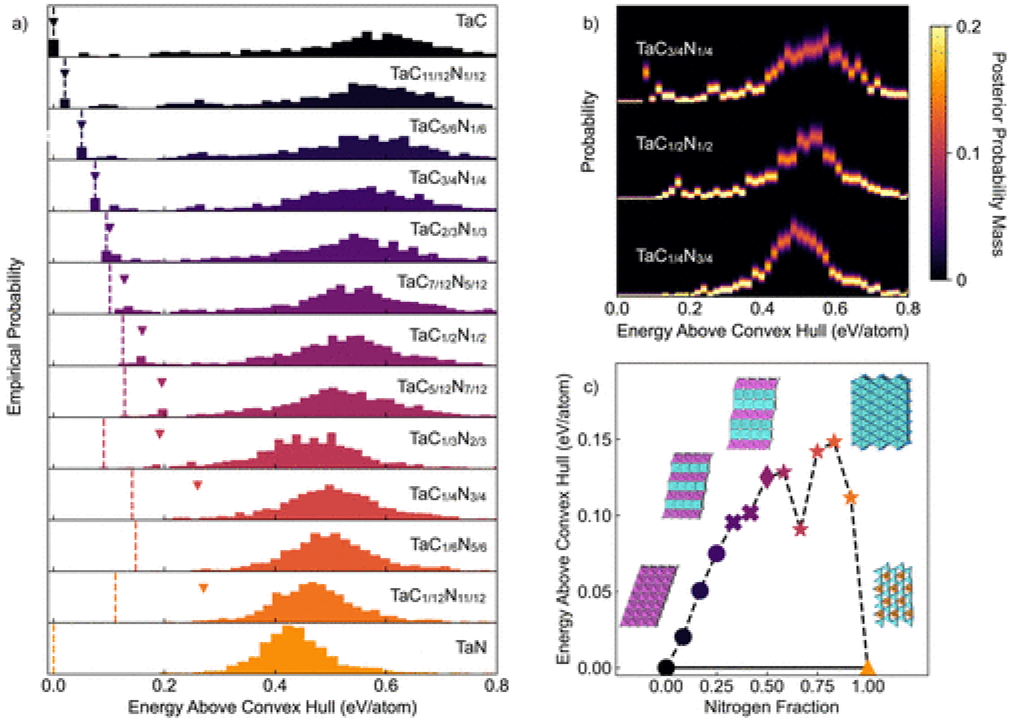


Figure 3. (a) Alloys between TaC and TaN show smooth transitions in the empirical energy distributions. By adding in nitrogen, the probability density at low energies continues to decrease resulting in a corrugated potential energy surface. Despite the addition of nitrogen, the rocksalt structure-type is found at nearly every composition; the minimum energy rocksalt decorations are highlighted with triangles. The observed ground state energies are marked with dashed lines. (b) The Dirichlet distribution allows for an estimate of the underlying probability distribution over energies given a finite number of samples. The uncertainties are shown to be sufficiently small. Here, the convex hull bounds the lowest energy considered in calculating the Dirichet distribution. (c) The polyhedra content of the lowest energy structures changes with composition. Trigonal prisms are incorporated with increasing nitrogen content in the alloys. All of the alloyed structures have positive energies of mixing, and are therefore above the convex hull (black line).

We find a series of interesting trends when gathering statistics on both high and low-energy structures through random structure sampling. Like most material attributes, the distribution of observed energies varies smoothly with composition. Furthermore, with increasing nitrogen concentration, the distributions become narrower, and the modes shift closer to the observed ground states. The proportion of states within 0.2 eV/atom of the ground state decreases with increasing nitrogen concentration as well (these trends are further illustrated in Figure S2). Both trends indicate an increase in glass forming ability with nitrogen content. The trend in glass forming ability could only have been studied by expanding our consideration to energy ranges far above the convex hull.

Much of the change in energetics seen in Figure 3a maps on to structural changes of the observed random structure samples. We find that the energy for structure samples is linearly correlated with the average coordination number of Ta (Figure S2). Following chemical intuition, the lowest energy structures tend to have the highest average coordination numbers (i.e., between 5 and 6). With increasing nitrogen content, there is an associated drop in the proportion of structures with an average coordination above 5 (illustrated in Figure S3). Synthesizing these trends, we conclude that the drop in low-energy structures for nitrogen-rich compositions is explained by the uptick in poorly bonded structures.

One could always ask whether the empirical probability distribution of energies closely matches the true underlying distribution. In Figure 3b we use the Dirichlet distribution to quantify the uncertainty in the empirical probability distribution given the number of observed random structure samples. All three probability distributions exhibit small degrees of smearing, associated with low residual uncertainties. We also include the possibility that a random structure sample could have an energy lower than any structure observed from sampling. The flat lines on the left end of each distribution show that this probability is low and unlikely to significantly affect our results. As such we conclude that our description of the energy distribution is sufficiently precise for discussing both high-energy states and trends in the observed ground states. For more details on the Dirichlet distribution, please see the Methods.

While we are largely concerned with basin sizes and the overall distribution of energies, it is worth pausing to appreciate the intriguing structural changes that emerge for the observed ground states as a function of composition (Figure 3c). Interestingly, all observed alloy ground state structures were also observed as low-energy structures for at least one of the two parents. From TaC up to $TaC_{3/4}N_{1/4}$, rocksalt is the ground state structure-type. However, past $x = 0.25$, the ground state structure begins to incorporate trigonal prisms. As seen in Figure 3c, the ground state structure shifts to SG 166, 160, and 194, associated with an increase in the proportion of trigonal prisms. Finally, only pure TaN exhibits trigonal planar environments. Again, this result is chemically intuitive; trigonal prism environments are shared by both TaC and TaN, while trigonal planar environments are specific to TaN.

When alloying, the ground state structures find a structural compromise by incorporating polyhedra that are preferable for both parent compounds.

Tuning Basin Size with Composition

Having looked more broadly at the energetics and structure, we can now focus specifically on how basin sizes change with composition. In Figure 4a, we plot the probability for a random structure to relax into various basins of attraction. Again, these probabilities are proportional to basin sizes. All basins are labeled by the "parent spacegroup", which refers to the assigned spacegroup when carbon and nitrogen are considered to be the same element. The carbon-rich compositions are dominated by the rocksalt basin, and other high-symmetry structure-types tend to fluctuate under 1%. In brown, we show the aggregate probability for falling into any basin corresponding to a structure-type with a parent spacegroup larger than 10. Even with such an inclusive symmetry cutoff, we see that the total probability for finding such structures decreases with nitrogen content.

Figure 4

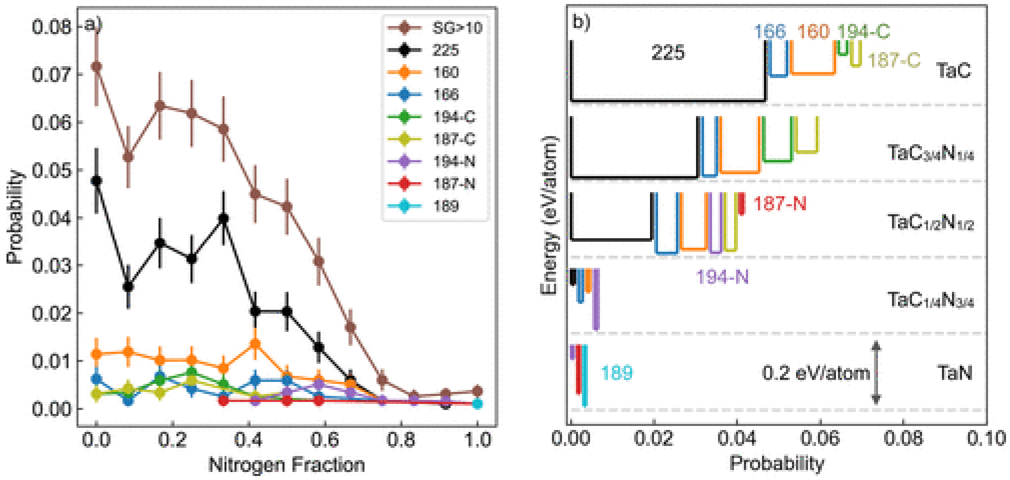


Figure 4. (a) The probability for a random structure to fall into a given basin is shown. The uncertainty in these probabilities is determined using the Beta distribution. Each error bar represents the standard deviation of the corresponding Beta distribution. Most random structure samples relax into high-energy, low-symmetry structures. The proportion of samples that relax into a spacegroup greater than 10 peaks at pure TaC and decreases with increasing nitrogen concentration. (b) Basins are represented as flat wells, with their widths and depths corresponding to their basin sizes and energies, respectively. Of the high-symmetry structures, rocksalt is the most prevalent up until TaC1/4N3/4, indicating a possibility for growing rocksalt metastably.

The uncertainty in the basin probabilities is quantified using a series of Beta distributions. For each basin at a given composition, a Beta distribution is established using only two inputs: the number of samples that fell into the basin and the number of samples that did not. The mode of the Beta distribution (empirical probability) is plotted as a point, while the error bars designate one standard deviation from the mode. While standard deviations provide a useful compression of the breadth of the Beta distributions, it is worth emphasizing that these distributions are not Gaussian. In Figure 6, we visualize the Beta distributions for all rocksalt basins across the composition space. From illustrating these distributions, it becomes clear that the dip in basin size at TaC11/12N1/12 is representative of statistical uncertainty rather than a chemical trend. Overall, we find the uncertainty to be sufficiently small given the number of random structure samples to conclude that the rocksalt basin is the largest, and symmetric structures decrease with nitrogen content.

A pictorial representation of basin size is shown in Figure 4b, where the width of the basin is proportional to its probability of occurrence. The basins are arbitrarily drawn with flat bottoms since knowledge about the curvature of the basin is unknown. The depth of these basin minima are expressed relative to the lowest energy structure discovered in random structure sampling for that composition. For TaC and TaC3/4N1/4, we see all the same basins occur, but the energetics shift as the other structure-types decrease in energy relative to the rocksalt. Moving from TaC3/4N1/4 to TaC1/2N1/2, all basins continue to shrink, and the lowest energy structures incorporate trigonal prisms. Interestingly, for TaC1/4N3/4 and TaN, all basins are incredibly narrow, signifying that much of the potential energy surface is dominated by high-energy structures with low symmetry (as supported by Figure 3a). Finally, it is only for pure TaN that SG 189 is observed. The absence of the 189 SG from other alloy compositions can be rationalized due to its unique trigonal planar coordination environments that are unfavorable for carbon. In particular, the SG 189 requires a 5-fold coordination of anions, and all of the discovered low-energy polymorphs of TaC exhibit 6-fold carbon coordination.

Mixing Thermodynamics

Basins of attraction give an indication of whether a phase will form from nonequilibrium growth, but they do not predict whether the material will undergo phase separation upon thermodynamic equilibration. As such, one could ask the following questions: (i) if a metastable thin film were annealed, would it undergo phase separation; and (ii) how strongly do basin sizes correlate with free energies of mixing? To complete our analysis of TaCxN1–x, we evaluate the thermodynamics of the observed polymorphs across composition space. The independent cell approximation is utilized for both its computational efficiency and ability to estimate configurational entropies at finite temperatures. (38−42) For a given composition, all polymorphs showing up in random structure sampling more than 1% of the time are considered. Additional polymorph–composition pairs are added for clarity, including the 189 SG, which does not occur for any of the alloy random structure sampling. For more information on the independent cell approximation, please see the Methods.

In Figure 5a, we highlight the thermodynamic density of states (TDOS) for SG 225 (“NaCl”) and 187-N (“WC”) structure-types since they have a strong structural contrast. All energies are expressed relative to the parent compounds, as enthalpies of mixing (ΔHmix). The SG 225 energies increase with nitrogen concentration, while the 187-N energies decrease. Such a finding provides a direct demonstration that nitrogen-rich compositions favor prismatic structures. Furthermore, the spread in the SG 225 TDOS is wider than that of 187-N, signaling a higher preference for anion disorder in the 187-N structure.

Figure 5

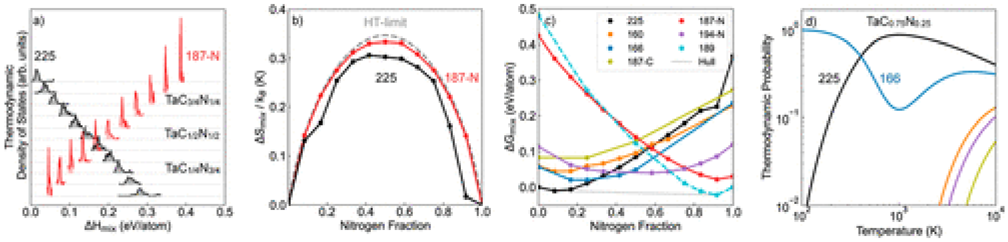


Figure 5. Thermodynamic analysis is conducted on the discovered low-energy polymorphs. (a) The TDOS is shown for two contrasting structure-types, SG 225 (black) and 187-N (red). With increasing nitrogen content, the energetic penalty for adopting octahedra over trigonal prisms increases. (b) The smaller variance in the 187-N TDOS translates to a higher entropy of mixing at 2000 K. (c) With enthalpies and entropies calculated, we show the free energy of mixing for all discovered low-energy polymorphs at 2000 K. There is limited solubility for both TaC and TaN. Many of the polymorphs are similar in free energy, as expected due to their similar structures. (d) In the limit where interfacial energies between polymorphs are negligible, the polymorphs can be treated as states within an ensemble. For instance, at TaC0.75N0.25, one can expect to see mostly SG 225, with a mix of the SG 166 polymorph as well.

As a result, in Figure 5b the configurational entropy of SG 187-N is far closer to the ideal entropy of mixing (gray) than the SG 225 structure. In the high-temperature limit, the entropy of mixing for both phases converges to the ideal entropy of mixing. Herein, we set the effective temperature to 2000 K since previous work has shown that high effective temperatures correspond well to sputtering on low-temperature substrates. (43)

With the TDOS in hand, we calculate the free energies of mixing for all eight observed polymorphs (Figure 5c). As consistent with our prior findings, octahedra-rich polymorphs are favored by carbon while the trigonal-prism and trigonal-planar motifs are favored by nitrogen. The free energy convex hull (gray) suggests that at elevated temperatures the SG 225 phase is stable to ~x = 1/12 nitrogen; more nitrogen-rich compositions will decompose to TaC11/12N1/12 and TaC1/12N11/12. Surprisingly, the two structures built exclusively from trigonal prisms (SG 194 and 187) exhibit startlingly different free energy curves.

Throughout the TaC1–xNx alloy space, the lowest energy polymorphs have several competing phases that are close in energy (<0.1 eV/atom). One resulting outcome may involve competing nucleation when the basin sizes are similar. However, the chemical structures of these phases offer a more complex scenario: multiple polymorphs can coexist as states within a single phase through stacking disorder. In other words, partial ergodicity may extend beyond individual basins of attraction. Such intergrowths are enabled by the common octahedral and prismatic polyhedra, which all share triangular faces and hexagonal anion packing.

In the limit where interfacial energies are low between polymorphs, the independent cell approximation can be extended across multiple polymorphs. From Zint in eq 10, we estimate the free energy of the phase with intergrowths using the standard equation: $\Delta \mathrm{Gint} = -\mathrm{kBT} \ln \mathrm{Zint}$. Overall, the ΔGint that emerges from a phase with intergrowths largely reflects the lowest energy polymorph. The probability for various polymorph contributions at a given temperature and composition are given by eq 9. Despite these modest changes to stability, the ergodic approximation does suggest a nontrivial concentration of stacking faults associated with mixing polymorphs. In Figure 5d, we illustrate the potential intergrowth concentration for TaC3/4N1/4. Under the ergodic approximation at 2000 K, an intergrowth involving ∼70% SG 225 and ∼30% SG 166 is predicted.

It is worth noting that the energetics differ slightly from that of Figure 3c. Namely, the observed lowest energy structure found through random structure sampling has a parent spacegroup of 225, whereas a single SG 166 structure was found, and it is 4 meV/atom higher in energy. When conducting the independent cell approximation, 50 configurations of each structure-type are initialized, and with this increased, focused sampling, the lowest energy 166 SG configuration observed is 2 meV/atom below the lowest energy 225 SG configuration. As such, at 0 K the intergrowth would be entirely SG 166, as seen in Figure 5d. Ultimately, these small differences in energetic ordering are small, and most likely inconsequential at relevant growth temperatures.

Implications for the Synthesis of TaC1–xNx

Beyond serving as a model system to understand how composition can adjust basin size, TaC and TaC1–xNx alloys have potential as epitaxial substrates. (24) As one might expect based on the different crystal structures and local motifs, there is limited solubility between TaC and TaN under thermodynamic equilibrium. However, the SG 225 structure-type is the dominant basin up to TaC0.25N0.75. As such, there is a possibility for stabilizing the rocksalt structure out to much more N-rich compositions with the use of a nonequilibrium growth technique like physical vapor deposition. To access these compositions, one might try conducting physical vapor deposition under carbon-rich compositions initially, and then layering on relatively nitrogen-rich layers. Previously,

$Al_{0.7}Ga_{0.3}N$ has been grown on binary TaC. (24) Successful growth of metastable $TaC_{1-x}N_x$ rocksalt films would provide access to a wider range of lattice parameters (Figure S7).

From $TaC_{0.25}N_{0.75}$ to TaN, the potential energy surface becomes increasingly glassy, with many high-energy minima and narrow basins. It thus may be harder to form metastable, crystalline thin films at those compositions; rather, amorphous films are generally expected. Potentially, one could modulate the film composition during growth to ultimately reach more N-rich compositions or use an appropriate epitaxial substrate.

Potential for Accelerated Basin-Size Mapping

For complex systems where the relevant structures have large unit cells, the basin-size mapping method would still work. However, larger simulation cells would be required, adding to the computational cost. Furthermore, a general trend is that basin sizes tend to shrink with the number of atoms in the simulation cell, since the number of minima on the potential energy surface is expected to grow exponentially. (44) As such, more random structure samples would be required in order to discover the relevant basins and obtain sufficient statistics to precisely determine their relative sizes.

The use of machine-learned interatomic potentials (MLIP) could increase the efficiency of basin-size mapping, allowing for better statistics and larger simulation cells. Still, leveraging MLIPs for basin-size mapping currently poses its own set of challenges. MLIPs would likely require large amounts of training data to handle the incredibly broad array of configurations that is inherent to random structure sampling. Incorrectly predicting the forces on atoms could result in significantly different relaxation trajectories and qualitatively different basin sizes. Given the sensitivity and difficulty in training such a model, uncertainty quantification for MLIP predictions would be desirable. However, quantifying the uncertainty in the energies and forces predicted by MLIPs is already a difficult task and an ongoing area of research. (45−47) Evaluating uncertainties over relaxation trajectories would only compound the difficulty. There is, however, reason for optimism given the promising recent efforts in producing MLIPs for random structure search. (48,49) For instance, with the use of a newly developed ephemeral data derived potential, Pickard has been able to use 56-atom cells to conduct random search for high-pressure phases of boron. (50) If MLIPs become sufficiently accurate for basin-size mapping, the combination of MLIPs with the uncertainty quantification shown herein would provide two complementary strategies for effectively mapping basin sizes.

Conclusions

Herein, we demonstrate that alloying tunes the potential energy surface, thus altering basin sizes and metastable polymorph synthesizability. Random structure sampling provides a method for probing the distribution of states that occur far from equilibrium, where random structures locally relax into various basins of attraction. For TaCxN1–x, the change in basin sizes with nitrogen content can be rationalized in terms of two effects. First, there is a gradual shift toward a corrugated potential energy surface with smaller basins, as seen by the observed energy distributions. Second, the preference for nitrogen to adopt trigonal prism environments results in growing the basins of structures rich in trigonal prisms and shrinking the basins associated with octahedra-rich structures. With respect to growing TaCxN1–x for epitaxial purposes, we note that adding nitrogen to TaC does not immediately reduce the rocksalt basin size. As such, there is a significant composition range where the rocksalt basin dominates in size, despite it being thermodynamically unstable. Within that composition window, there is an opportunity for leveraging nonequilibrium syntheses to grow metastable TaCxN1–x alloys with variable lattice parameters.

The generally small basin sizes found across the TaCxN1–x composition space posed a challenge in robustly establishing trends. Uncertainty quantification was conducted for both basin sizes and energy distributions of local minima. Looking forward, we expect that the uncertainty quantification presented herein will allow for a more efficient use of random structure sampling for basin-size mapping, allowing for searches over broader thermodynamic conditions.

Methods

Structure Generation

Lattices are constructed such that lattice vector angles are between 60 and 140°. Lattice vectors are set to have lengths between 0.8 and 1.4 of the overall “scale” of the structure. Atoms are then placed within the lattice. The scale is used to adjust the volume of the cell such that the random structures are initialized from physically reasonable starting points. To be clear, these physically reasonable structures are often multiple eV/atom above their local minima. We calibrate the scale such that the shortest bond length is 90–100% of the bond lengths found in experimentally determined ground state structures. The shortest Ta–C and Ta–N bond lengths, as found from the experimentally reported ground state structures (SG 225 and SG 189) in the Inorganic Crystal Structure Database (ICSD) are 2.2 and 2.0 Å, respectively. The shortest starting bond length in the produced random structures for all compositions was 1.9 Å.

Atom placement is biased toward configurations that are dominantly cation–anion coordinated, by distributing cations and anions over two interpenetrating grids within the simulation cell, as detailed elsewhere. (12) Random structure sampling has been demonstrated to be effective in finding both the ground state structures as well as the experimentally realizable metastable polymorphs for systems such as C, Si, MgO, ZnO, SnO2, group-IV carbides, and ternary nitrides. (12,14,17,18)

Simulation-Cell Size

Each random structure is populated with 24 atoms. Generally, we find that 24-atom cells are sufficient for modeling disordered systems. In our previous work on IV–VI alloys, we showed that an ensemble of 24-atom supercells produced high-temperature ensemble averages of thermodynamics and the bulk modulus that were sufficiently close to a 128-atom Special Quasi-random Structure. (41) Even smaller cell sizes have been successfully used for modeling disorder within the independent cell approximation. (38,51) Moreover, in our work on elemental silicon (33) and SiO2, (52) we showed that 24-atom cells are sufficient to accurately model structural properties of amorphous and glassy states. However, running tests for cell sizes with less than 16 atoms resulted in finite-size effects for both alloy and glassy ensembles. (33,41) Regarding basin sizes, using structures with too few atoms has also resulted in an over representation of ordered structures that did not match experimental polymorphism in SnO2, while 24-atom cells did. (12) Finally, nucleation processes that result in the growth of a certain phase start at the nanometer scale. Hence, cells on the order of 10s of atoms are commensurate with the potential energy surface that nucleation seeds would encounter.

Composition Sampling

24-atom simulation cells include 12 cation sites and 12 anion sites. As such, 13 different compositions are available, including the two parent compounds. The number of random structures used for each composition is determined through convergence testing. For TaC, only 1000 structures was necessary, and for both TaC0.916N0.084 and TaN, 2000 structures were required. For all other compositions, 1200 structures were found to be sufficient for the converged results to be obtained.

Structure Classification

To determine basin sizes, relaxed random structure samples need to be grouped. When constructing basins for pure compounds, the choices for grouping are fairly straightforward, but with alloys there are nuances that need explaining. Here we use the average coordination number of the Ta atoms and the “parent spacegroup” as classification criteria. Structures are considered to be within the same basin if they have the same parent spacegroup and their average coordination number is within 0.1 of each other.

The coordination shell for each Ta atom was calculated in the following way. Space is searched radially outward from the central Ta atom, and atoms that are encountered are added to the coordination shell. The search radius increases until one of two criteria are met: (1) the radial distance has exceeded 0.5 Å of the shortest bond length found; or (2) a Ta atom is found. If a second Ta atom is encountered, it is not added to the coordination shell. Disorder can result in structural distortions that break local symmetry. As a result, the coordination environment of each Ta atom will be slightly different, even within a given structure. We use the average coordination number such that our classification methods are not too sensitive to anion disorder, but rather pick up on the underlying structure-types.

Similarly, we evaluate the “parent spacegroup” by assigning all C and N atoms the same label (C). The spacegroup assignment is then evaluated using a tolerance factor of 0.5 Å to accommodate the differences in the bond distances and bond angles due to the varying atom identities. The tolerance of 0.5 Å for both the spacegroup and coordination shell accommodates the expected differences in the Ta–C and Ta–N bond lengths. The magnitude was also justified by extensive visual inspection of the grouped structures and both their polyhedra and bond arrangements therein. All spacegroups were calculated using spglib. (53)

Quantifying Basin-Size Uncertainty with the Beta Distribution

The relative basin sizes are expressed as probabilities; namely, we estimate the probability that a randomly generated structure will relax into a given basin of attraction. However, for a finite number of samples, there is a chance that the empirical probability distribution (the one measured) differs from the true underlying distribution. The same argument can be made for the empirical energy distributions: an observed distributions can differ from the true distribution.

In this work, we take a Bayesian approach and employ the related Beta and Dirichlet distributions to quantify the uncertainty in the measured probabilities and energy distributions. (54,55) The goal in using these distributions is to produce a probability distribution over possible outcomes given the observed data (i.e., a posterior distribution). The breadths of the resulting posterior distributions are used to demonstrate the uncertainty in our measurements. In Figure 4a, we quantify the uncertainty as the standard deviation in the posterior Beta distribution for each basin size. In Figure 3b, we visually show the breadth in the posterior Dirichlet distribution over energy distributions is small.

We will begin with our treatment of the Beta distribution since it is the one-dimensional analog of a Dirichlet. The Beta distribution is useful for quantifying the uncertainty in the probability for an event with two possible outcomes. In our case, the two possible outcomes will be whether a random structure relaxes into a given basin. Since there are only two possibilities, the probability distribution is determined by a single parameter, θ, which expresses the probability for one of the categories to occur (the probability for the other outcome is 1 – θ). A Beta distribution is specified by two parameters α and β, and is given by

$$P(\theta|\alpha,\beta)=\theta\alpha-1(1-\theta)\beta-1\int 10\theta\alpha-1(1-\theta)\beta-1\mathrm{d}\theta \qquad (1)$$

Assuming a uniform prior (α = β = 1), then the posterior distribution over possible θ values (conditioned on observed data) is given by a Beta distribution, where α – 1 is the number of times that the first outcome has been observed, and β – 1 is number of times the second outcome has been observed. In Figure 6, we show the Beta distributions for the rocksalt basin probabilities as a function of composition. The exact α and β parameters are included in the Supporting Information (Table S1).

Figure 6

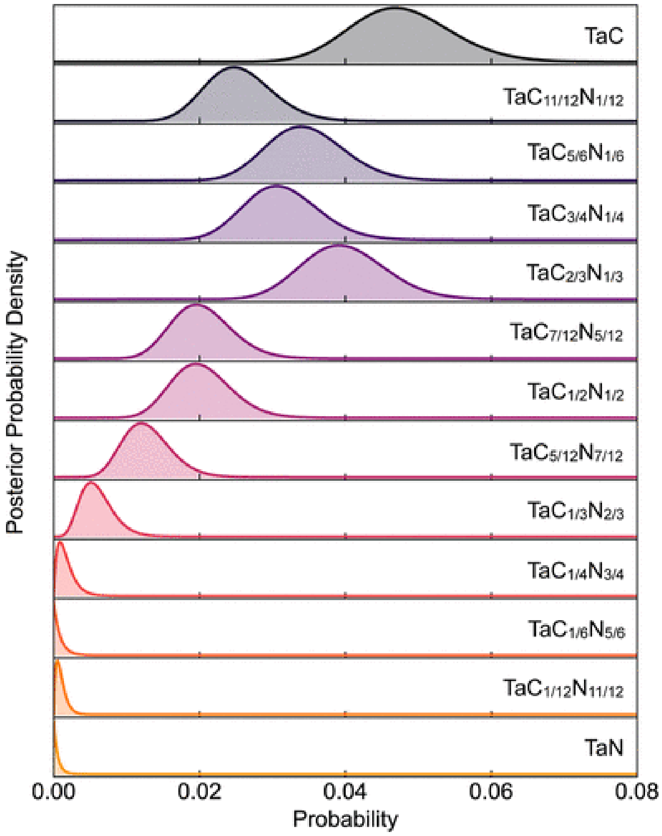


Figure 6. The Beta distribution is employed to quantify the uncertainty in the probability to fall into a given structure-type. While this procedure is conducted for all structure-types, here we highlight rocksalt. The posterior probability densities for rocksalt probabilities are calculated from the values in Table S1 and are shown on a linear scale for each composition. The integral of the probability density functions is, by definition, unity. The error bars shown in the Figure 4a correspond to the standard deviation of these probability density functions. Broadly, the probability of falling into the rocksalt basin decreases with nitrogen content.

Quantifying Energy-Distribution Uncertainty with the Dirichlet Distribution

Through random structure sampling, we seek to obtain relative basin sizes and the distribution of energetic states. The Beta distribution is useful for quantifying the uncertainty in basin sizes, but the Dirichlet distribution is more suited for handling the uncertainty in energy distributions. As such, we employ the Dirichlet distribution, the n-dimensional analog of Beta, to quantify the uncertainty in events that have multiple categorical outcomes.

In the Dirichlet distribution, Θ represents the underlying probability distribution, and it is a vector that adds up to 1. Each component of Θ represents the probability of a given outcome (the multidimensional analog to θ). K is the number of assumed categories and α – 1 is a vector of length K. If we assume a uniform prior (α = 1), then α – 1 corresponds to the observed counts for all categories. As such, the posterior Dirichlet distribution represents the distribution over possible categorical distributions, Θ, with the following form

$$P(\Theta|\alpha)=1B(\alpha)\prod i=1K\Theta\alpha i-1i \qquad (2)$$

where B(α) is the normalizing constant. As can be seen above, the only two inputs to the Dirichlet distribution are the number of polymorphs, K, and the number of times each polymorph has been observed in random structure sampling, α – 1.

The above treatment is extended to producing uncertainty in the observed energy distributions. To do so, we divide each energy spectrum into 50 bins (the same number of bins as is shown in the energy histograms in Figure 3). The lower bound of the energy spectrum is taken to be a linear combination of the parent energies for that composition. In other words, we assume that the enthalpy of mixing will be positive for all observed structures. The upper bound is arbitrarily set to an enthalpy of mixing of 0.8 eV/atom. The Dirichlet distribution is then produced from the absolute counts for each energetic bin, as seen in Figure 7.

Figure 7

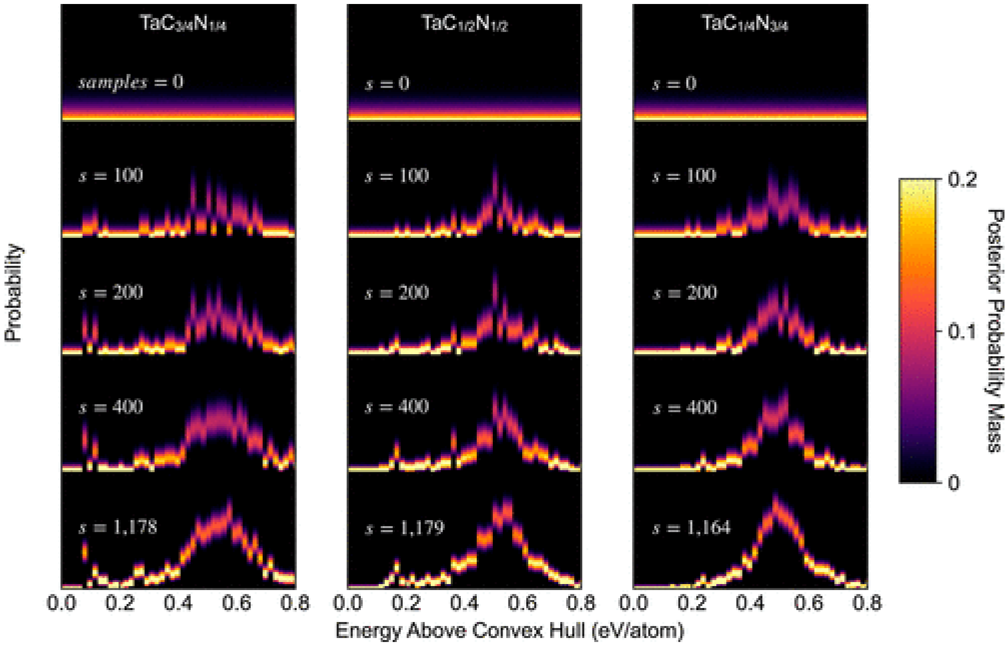


Figure 7. For each composition, a Dirichlet distribution is defined, and the samples from the distributions are visualized with a heat map. For each energy bin, the heat map values sum to one, corresponding to a probability over possible energy-bin probabilities. With an increasing number of random structure samples, the energy distributions take form. Minimal smearing, and thus small uncertainties, are produced with ~1000–2000 random structures.

Here, the Dirichlet distribution is a distribution over possible energy histograms, and each Dirichlet sample corresponds to a histogram. It can be difficult to ascertain much knowledge by viewing many histogram samples superimposed on top of each other. Instead, we take these histograms and plot their overall density. The heat map values are normalized such that, for a given energy window, the distribution over heat map values sums to one. As such, the heat map value (i.e., the Dirichlet probability) corresponds to the probability that, for a given energy bin, the histogram will be a given value.

In Figure 7, we show how the posterior distribution over energy distributions develops with an increasing number of observed random structures (samples). Our prior distribution is the weakest possible uniform distribution. As such, our α vector begins with all ones. Most samples from the uniform distribution tend to result in uniform energy distributions as well. After 100 observations, the energy distributions still have significant uncertainty, corresponding to the smearing in the y-values. By 400 observations, the general shape of the distributions is produced. We continue to sample random structures until the distributions are sufficiently converged. Notably, the Dirichlet distribution shows that relatively small uncertainties can be produced with roughly 1000–2000 random structures.

The Dirichlet distribution sampling was conducted with scipy. (56) It is worth noting that the distribution over possible heights for a single energy bin is a Beta distribution (Figure S7).

Independent Cell Approximation

The use of the Independent Cell Approximation is the same as described before. (41) Structures are generated on a fixed lattice and relaxed. The enthalpy of mixing, $\Delta H_{mix,i}$, for configuration i, is defined in the following equation

$$\Delta H_{mix,i}=(E_i+pV_i)-\sum_j^k x(E_j+pV_j) \tag{3}$$

Here, the second term sums over k parent compounds with indices j. p is the pressure and V is the volume per atom. All results presented correspond to the low pressure ($p \approx 0$) case for which internal energy and enthalpy are equal, as are the Helmholtz and Gibbs free energies. The probability for the ith configuration is given by

$$p_i=\frac{e^{-E_i/k_BT}}{Z} \tag{4}$$

where k_B is Boltzmann's constant and T is temperature. The partition function, Z, serves as a normalization constant, resulting from the sum over all n configurations

$$Z=\sum_i^n e^{-E_i/k_BT} \tag{5}$$

The ensemble averaged enthalpy of mixing is thus

$$\Delta H_{mix}=\mathbb{E}_i[\Delta H_{mix,i}]=\sum_i^n p_i\Delta H_{mix,i} \tag{6}$$

The entropy of mixing is calculated as

$$\Delta S_{mix}=-k_B\sum_i^n p_i\ln p_i \tag{7}$$

Since exhaustive sampling is not employed, the entropy needs to be scaled; here we scale the entropy by a multiplicative constant such that in the high-temperature limit, the entropy approaches $\Delta S = -k_Bx\ln(x) - (1-x)k_B\ln(1-x)$, as was done previously. (41) Again, x is the fraction of nitrogen. The free energy of mixing, ΔG_{mix}, is evaluated as

$$\Delta G_{mix}=\Delta H_{mix}-T\Delta S_{mix} \tag{8}$$

Composition-polymorph pairs that are on the lower bound of the convex hull are considered to be thermodynamically stable. The convex hull was generated using the qhull algorithm (57) within the scipy library. (56) Custom code was built to isolate the lower bound of the hull.

To evaluate the concentration of intergrowths, we assume full ergodicity between structure-types. As such, the probability for the ath structure-type to occur is given by

$pa=\mathrm{e}-\Delta G\mathrm{mix},a/k\mathrm{B}TZ\mathrm{int}$ (9)

where Zint is defined as the sum over all structure-types

$Z\mathrm{int}=\sum a\mathrm{e}-\Delta G\mathrm{mix},a/k\mathrm{B}T$ (10)

DFT Relaxations

All structures are fully relaxed using density functional theory. All degrees of freedom are relaxed including the volume, cell shape, and atomic positions. We employ the PBE functional (58) within the projector-augmented wave method (59) as implemented in the VASP code. (60) The total energies are calculated using a plane-wave cutoff of 340 eV and the Monkhorst–Pack k-point grid (61) is chosen such that the total energies are converged to within 3 meV/atom. In addition, an on-site Coulomb interaction U = 3 eV term is applied to Ta-d states following the rotationally invariant DFT + U formalism. (62) The choice of the U-value follows the fitted elemental reference energy (FERE) procedure (63) developed to provide accurate compound enthalpies of formation. All structures are relaxed to total final pressure of less than 3 kbar. Management of calculations and the analysis of results is carried out using the pylada software. (64)

Notes

The authors declare no competing financial interest.

Acknowledgments

This work was supported under NSF OAC 2118201. M.J. and V.S acknowledge support from NSF Grant no. DMR-1945010. Additional support came from NSF OAC 1940199 (A.N. and E.T.) and OAC-1940224 (Q.N.). This work was authored in part by the National Renewable Energy Laboratory (NREL), operated by Alliance for Sustainable Energy, LLC, for the U.S. Department of Energy (DOE) under Contract no. DE-AC36-08GO28308. This work was supported by the Laboratory Directed Research and Development (LDRD) Program at NREL. The views expressed in the article do not necessarily represent the views of the DOE or the U.S. Government.

References

1. Therrien, F.; Jones, E. B.; Stevanović, V. Metastable materials discovery in the age of large-scale computation. Applied Physics Reviews 2021, 8, 031310, DOI: 10.1063/5.0049453
2. Greenaway, A. L.; Melamed, C. L.; Tellekamp, M. B.; Woods-Robinson, R.; Toberer, E. S.; Neilson, J. R.; Tamboli, A. C. Ternary Nitride Materials: Fundamentals and Emerging

Device Applications. Annu. Rev. Mater. Res. 2021, 51, 591– 618, DOI: 10.1146/annurev-matsci-080819-012444

3. Schnepf, R. R.; Cordell, J. J.; Tellekamp, M. B.; Melamed, C. L.; Greenaway, A. L.; Mis, A.; Brennecka, G. L.; Christensen, S.; Tucker, G. J.; Toberer, E. S.; Lany, S.; Tamboli, A. C. Utilizing Site Disorder in the Development of New Energy-Relevant Semiconductors. ACS Energy Lett. 2020, 5, 2027– 2041, DOI: 10.1021/acsenergylett.0c00576
4. Aykol, M.; Dwaraknath, S. S.; Sun, W.; Persson, K. A. Thermodynamic limit for synthesis of metastable inorganic materials. Sci. Adv. 2018, 4, eaaq0148 DOI: 10.1126/sciadv.aaq0148
5. Sun, W.; Dacek, S. T.; Ong, S. P.; Hautier, G.; Jain, A.; Richards, W. D.; Gamst, A. C.; Persson, K. A.; Ceder, G. The thermodynamic scale of inorganic crystalline metastability. Sci. Adv. 2016, 2, e1600225 DOI: 10.1126/sciadv.1600225
6. Pöhls, J.-H.; Heyberger, M.; Mar, A. Comparison of computational and experimental inorganic crystal structures. J. Solid State Chem. 2020, 290, 121557, DOI: 10.1016/j.jssc.2020.121557
7. Stillinger, F. H.; Weber, T. A. Packing structures and transitions in liquids and solids. Science 1984, 225, 983– 989, DOI: 10.1126/science.225.4666.983
8. Asenjo, D.; Paillusson, F.; Frenkel, D. Numerical calculation of granular entropy. Phys. Rev. Lett. 2014, 112, 098002, DOI: 10.1103/PhysRevLett.112.098002
9. Martiniani, S.; Schrenk, K. J.; Stevenson, J. D.; Wales, D. J.; Frenkel, D. Structural analysis of high-dimensional basins of attraction. Phys. Rev. E 2016, 94, 031301, DOI: 10.1103/PhysRevE.94.031301
10. Martiniani, S.; Casiulis, M. When you ’t count, sample! Computable entropies beyond equilibrium from basin volumes. Pap. Phys. 2023, 15, 150001, DOI: 10.4279/pip.150001
11. Valade, A.; Libeskind, N. I.; Pomarède, D.; Tully, R. B.; Hoffman, Y.; Pfeifer, S.; Kourkchi, E. Identification of basins of attraction in the local Universe. Nat. Astron. 2024, 8, 1610–1616, DOI: 10.1038/s41550-024-02370-0
12. Stevanović, V. Sampling Polymorphs of Ionic Solids using Random Superlattices. Phys. Rev. Lett. 2016, 116, 075503, DOI: 10.1103/PhysRevLett.116.075503
13. Caskey, C. M.; Holder, A.; Shulda, S.; Christensen, S. T.; Diercks, D.; Schwartz, C. P.; Biagioni, D.; Nordlund, D.; Kukliansky, A.; Natan, A. Synthesis of a mixed-valent tin nitride and considerations of its possible crystal structures. J. Chem. Phys. 2016, 144, 144201, DOI: 10.1063/1.4945561
14. Jones, E. B.; Stevanović, V. Polymorphism in elemental silicon: Probabilistic interpretation of the realizability of metastable structures. Phys. Rev. B 2017, 96, 184101, DOI: 10.1103/PhysRevB.96.184101
15. Fertitta, E.; Das, S.; Banerjee, D.; Ebrahimi, F.; Barraud, C.; Du, K.; Tian, H.; Pickard, C. J.; Weber, C.; Ramesh, R.; Littlewood, P.; Dubbink, D. Study of disorder in pulsed laser

deposited double perovskite oxides by first-principle structure prediction. npj Comput. Mater. 2021, 7, 92, DOI: 10.1038/s41524-021-00561-1

16. Woods-Robinson, R.; Stevanović, V.; Lany, S.; Heinselman, K. N.; Horton, M. K.; Persson, K. A.; Zakutayev, A. Role of disorder in the synthesis of metastable zinc zirconium nitrides. Phys. Rev. Mater. 2022, 6, 043804, DOI: 10.1103/PhysRevMaterials.6.043804
17. Jankousky, M.; Garrity, E. M.; Stevanović, V. Polymorphism of group-IV carbides: Structures,(meta) stability, electronic, and transport properties. Phys. Rev. Mater. 2023, 7, 053606, DOI: 10.1103/PhysRevMaterials.7.053606
18. Zakutayev, A.; Jankousky, M.; Wolf, L.; Feng, Y.; Rom, C. L.; Bauers, S. R.; Borkiewicz, O.; LaVan, D. A.; Smaha, R. W.; Stevanovic, V. Synthesis pathways to thin films of stable layered nitrides. Nat. Synth. 2024, 3, 1471– 1480, DOI: 10.1038/s44160-024-00643-0
19. Pickard, C. J.; Needs, R. J. Ab initiorandom structure searching. J. Phys.: Condens. Matter 2011, 23, 053201, DOI: 10.1088/0953-8984/23/5/053201
20. Friedrich, A.; Winkler, B.; Bayarjargal, L.; Juarez Arellano, E. A.; Morgenroth, W.; Biehler, J.; Schröder, F.; Yan, J.; Clark, S. M. In situ observation of the reaction of tantalum with nitrogen in a laser heated diamond anvil cell. J. Alloys Compd. 2010, 502, 5– 12, DOI: 10.1016/j.jallcom.2010.04.113
21. Friedrich, A.; Morgenroth, W.; Bayarjargal, L.; Juarez-Arellano, E. A.; Winkler, B.; Konôpková, Z. In situ study of the high pressure high-temperature stability field of TaN and of the compressibilities of -TaN and TaON. High Pressure Res. 2013, 33, 633– 641, DOI: 10.1080/08957959.2013.813943
22. Chaudhuri, S.; Maasilta, I. J.; Chandernagor, L.; Ging, M.; Lahtinen, M. Fabrication of superconducting tantalum nitride thin films using infrared pulsed laser deposition. J. Vac. Sci. Technol., A 2013, 31, 061502, DOI: 10.1116/1.4812698
23. Nakamura, K.; Yashima, M. Crystal structure of NaCl-type transition metal monocarbides MC (M= V, Ti, Nb, Ta, Hf, Zr), a neutron powder diffraction study. Mater. Sci. Eng., C 2008, 148, 69– 72, DOI: 10.1016/j.mseb.2007.09.040
24. Roberts, D. M.; Hachtel, J. A.; Haegel, N. M.; Miller, M. K.; Rice, A. D.; Tellekamp, M. B. Designing TaC Virtual Substrates for Vertical AlxGa1–xN Power Electronics Devices. PRX Energy 2024, 3, 033007, DOI: 10.1103/PRXEnergy.3.033007
25. Nehate, S.; Saikumar, A.; Prakash, A.; Sundaram, K. A review of boron carbon nitride thin films and progress in nanomaterials. Mater. Today Adv. 2020, 8, 100106, DOI: 10.1016/j.mtadv.2020.100106
26. Jankousky, M.; Chen, H.; Novick, A.; Stevanović, V. All “roads” lead to rocksalt structure. J. Am. Chem. Soc. 2024, 146, 23004– 23011, DOI: 10.1021/jacs.4c02974
27. Christensen, A. N.; Lebech, B. A reinvestigation of the structure of ϵ-tantalum nitride. Acta Crystallogr., Sect. B 1978, 34, 261– 263, DOI: 10.1107/S0567740878002733

28. Lehmann, T. S.; Niewa, R. Electrochemical synthesis of transition metal oxide nitrides with -TaN, -NbN and -Mo2N structure type in a molten salt system. Z. Naturforsch., B:J. Chem. Sci. 2020, 75, 33– 40, DOI: 10.1515/znb-2019-0022
29. Brauer, G.; Mohr, E.; Neuhaus, A.; Skokan, A. -TaN. eine Hochdruckform von Tantalnitrid. Monatsh. Chem. 1972, 103, 794– 798, DOI: 10.1007/bf00905439
30. Doye, J. P.; Wales, D. J.; Miller, M. A. Thermodynamics and the global optimization of Lennard-Jones clusters. J. Chem. Phys. 1998, 109, 8143– 8153, DOI: 10.1063/1.477477
31. Doye, J. P. K.; Massen, C. P. Characterizing the network topology of the energy landscapes of atomic clusters. J. Chem. Phys. 2005, 122, 084105, DOI: 10.1063/1.1850468
32. Massen, C. P.; Doye, J. P. Power-law distributions for the areas of the basins of attraction on a potential energy landscape. Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys. 2007, 75, 037101, DOI: 10.1103/physreve.75.037101
33. Jones, E. B.; Stevanović, V. The glassy solid as a statistical ensemble of crystalline microstates. npj Comput. Mater. 2020, 6, 56, DOI: 10.1038/s41524-020-0329-2
34. Wu, M.; Tang, G.; Qian, G.; Qian, Q.; Chen, D.; Dong, G.; Yang, Z. The multicomponent oxide glass as a statistical ensemble of neighboring glassy compounds in the composition space. J. Am. Ceram. Soc. 2023, 106, 306– 316, DOI: 10.1111/jace.18763
35. Perim, E.; Lee, D.; Liu, Y.; Toher, C.; Gong, P.; Li, Y.; Simmons, W. N.; Levy, O.; Vlassak, J. J.; Schroers, J.; Curtarolo, S. Spectral descriptors for bulk metallic glasses based on the thermodynamics of competing crystalline phases. Nat. Commun. 2016, 7, 12315, DOI: 10.1038/ncomms12315
36. Tsukimoto, S.; Moriyama, M.; Murakami, M. Microstructure of amorphous tantalum nitride thin films. Thin Solid Films 2004, 460, 222– 226, DOI: 10.1016/j.tsf.2004.01.073
37. Dastan, D.; Shan, K.; Jafari, A.; Gity, F.; Yin, X.-T.; Shi, Z.; Alharbi, N. D.; Reshi, B. A.; Fu, W.; Ţălu, Ş. Influence of nitrogen concentration on electrical, mechanical, and structural properties of tantalum nitride thin films prepared via DC magnetron sputtering. Appl. Phys. A: Mater. Sci. Process. 2022, 128, 400, DOI: 10.1007/s00339-022-05501-4
38. Jiang, C.; Uberuaga, B. P. Efficient ab initio modeling of random multicomponent alloys. Phys. Rev. Lett. 2016, 116, 105501, DOI: 10.1103/PhysRevLett.116.105501
39. Yang, K.; Oses, C.; Curtarolo, S. Modeling off-stoichiometry materials with a high-throughput ab-initio approach. Chem. Mater. 2016, 28, 6484– 6492, DOI: 10.1021/acs.chemmater.6b01449
40. Sorkin, V.; Tan, T. L.; Yu, Z.; Zhang, Y. Generalized small set of ordered structures method for the solid-solution phase of high-entropy alloys. Phys. Rev. B 2020, 102, 174209, DOI: 10.1103/PhysRevB.102.174209

41. Novick, A.; Nguyen, Q.; Garnett, R.; Toberer, E.; Stevanović, V. Simulating high-entropy alloys at finite temperatures: An uncertainty-based approach. Phys. Rev. Mater. 2023, 7, 063801, DOI: 10.1103/PhysRevMaterials.7.063801
42. Kuner, M. C.; Rothchild, E.; Asta, M. D.; Chrzan, D. C. Ab initio property predictions of quinary solid solutions using small binary cells. Comput. Mater. Sci. 2024, 238, 112924, DOI: 10.1016/j.commatsci.2024.112924
43. Ndione, P. F.; Shi, Y.; Stevanovic, V.; Lany, S.; Zakutayev, A.; Parilla, P. A.; Perkins, J. D.; Berry, J. J.; Ginley, D. S.; Toney, M. F. Control of the electrical properties in spinel oxides by manipulating the cation disorder. Adv. Funct. Mater. 2014, 24, 610– 618, DOI: 10.1002/adfm.201302535
44. Stillinger, F. H. Energy Landscapes, Inherent Structures, and Condensed-Matter Phenomena; Princeton University Press, 2015.
45. Zhu, A.; Batzner, S.; Musaelian, A.; Kozinsky, B. Fast uncertainty estimates in deep learning interatomic potentials. J. Chem. Phys. 2023, 158, 164111, DOI: 10.1063/5.0136574
46. Tan, A. R.; Urata, S.; Goldman, S.; Dietschreit, J. C.; Gómez-Bombarelli, R. Single-model uncertainty quantification in neural network potentials does not consistently outperform model ensembles. npj Comput. Mater. 2023, 9, 225, DOI: 10.1038/s41524-023-01180-8
47. Carrete, J.; Montes-Campos, H.; Wanzenböck, R.; Heid, E.; Madsen, G. K. H. Deep ensembles vs committees for uncertainty estimation in neural-network force fields: Comparison and application to active learning. J. Chem. Phys. 2023, 158, 204801, DOI: 10.1063/5.0146905
48. Pickard, C. J. Ephemeral data derived potentials for random structure search. Phys. Rev. B 2022, 106, 014102, DOI: 10.1103/PhysRevB.106.014102
49. Salzbrenner, P. T.; Joo, S. H.; Conway, L. J.; Cooke, P. I. C.; Zhu, B.; Matraszek, M. P.; Witt, W. C.; Pickard, C. J. Developments and further applications of ephemeral data derived potentials. J. Chem. Phys. 2023, 159, 144801, DOI: 10.1063/5.0158710
50. Pickard, C. J. Beyond theory-driven discovery: introducing hot random search and datum-derived structures. Faraday Discuss. 2025, DOI: 10.1039/d4fd00134f
51. Sarker, P.; Harrington, T.; Toher, C.; Oses, C.; Samiee, M.; Maria, J.-P.; Brenner, D. W.; Vecchio, K. S.; Curtarolo, S. High-entropy high-hardness metal carbides discovered by entropy descriptors. Nat. Commun. 2018, 9, 4980, DOI: 10.1038/s41467-018-07160-7
52. Wolf, L.; Novick, A.; Stevanović, V. Modeling glasses from first-principles using random structure sampling. arXiv 2024, arXiv:2410.09726
53. Togo, A.; Shinohara, K.; Tanaka, I. Spglib: a software library for crystal symmetry search. Sci. Technol. Adv. Mater.:Methods 2024, 4, 2384822, DOI: 10.1080/27660400.2024.2384822

54. Gelman, A.; Carlin, J. B.; Stern, H. S.; Rubin, D. B. Bayesian Data Analysis; Chapman and Hall/CRC, 1995.
55. Ross, S. M. Introduction to Probability Models; Academic Press, 2014.
56. Virtanen, P. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. Nat. Methods 2020, 17, 261– 272, DOI: 10.1038/s41592-019-0686-2
57. Barber, C. B.; Dobkin, D. P.; Huhdanpaa, H. The quickhull algorithm for convex hulls. ACM Trans. Math Software 1996, 22, 469– 483, DOI: 10.1145/235815.235821
58. Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. Phys. Rev. Lett. 1996, 77, 3865– 3868, DOI: 10.1103/PhysRevLett.77.3865
59. Blöchl, P. E. Projector augmented-wave method. Phys. Rev. B: Condens. Matter Mater. Phys. 1994, 50, 17953– 17979, DOI: 10.1103/PhysRevB.50.17953
60. Kresse, G.; Furthmüller, J. Comput. Mater. Sci. 1996, 6, 15, DOI: 10.1016/0927-0256(96)00008-0
61. Monkhorst, H. J.; Pack, J. D. Special points for Brillouin-zone integrations. Phys. Rev. B: Solid State 1976, 13, 5188– 5192, DOI: 10.1103/PhysRevB.13.5188
62. Liechtenstein, A. I.; Anisimov, V. I.; Zaanen, J. Density-functional theory and strong interactions: Orbital ordering in Mott-Hubbard insulators. Phys. Rev. B: Condens. Matter Mater. Phys. 1995, 52, R5467– R5470, DOI: 10.1103/PhysRevB.52.R5467
63. Stevanović, V.; Lany, S.; Zhang, X.; Zunger, A. Correcting density functional theory for accurate predictions of compound enthalpies of formation: Fitted elemental-phase reference energies. Phys. Rev. B: Condens. Matter Mater. Phys. 2012, 85, 115104, DOI: 10.1103/PhysRevB.85.115104
64. pylada–A python computational physics framework. https://github.com/pylada (accessed January, 2023).