JCTC Journal of Chemical Theory and Computation

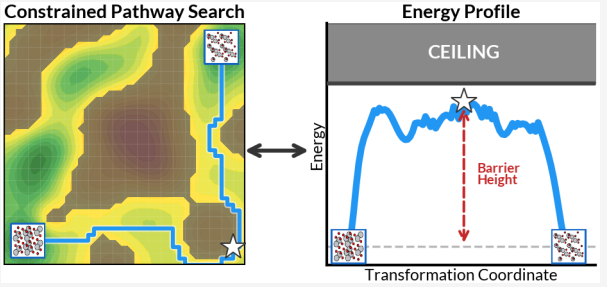
pubs.acs.org/JCTC Article

An Investigation into the Kinetic Persistence of TiO_2 Polymorphs Using Machine-Learning-Driven Pathfinding in Crystal Configuration Space

Max C. Gallant, David Mrdjenovich, and Kristin A. Persson*

Cite This: https://doi.org/10.1021/acs.jctc.6c01114 Read Online

ABSTRACT: As the number of theoretically predicted materials continues to expand, it is increasingly important to evaluate not only their thermodynamic stability but also their kinetic resilience against transformation to competing polymorphs. In this study, we investigate the hypothesis that the kinetic persistence of a metastable polymorph is related to the topography of the potential energy landscape separating it from the lower-energy phases. To accomplish this, we developed a new method for identifying diffusionless transformation pathways between metastable polymorphs and their ground-state counterparts and discussed the energetics of those pathways with respect to the experimental observations of each phase. This algorithm utilizes the recently developed crystal normal form, which provides a graph representation of the crystal configuration space and supplies the substrate for our pathfinding algorithm. We apply this method to the TiO_2 system, which contains the well-known anatase, rutile, and brookite phases, in addition to a number of hypothetical metastable polymorphs and high-pressure phases.



INTRODUCTION

Over the past decade, the field of synthesis science, in particular its theory and modeling aspects, has received increased attention due to its central role in developing novel inorganic functional materials. The relative recency of this focus is related to the dual rise of high-throughput simulation methods for determining material properties and the adoption of machine learning methods for computational materials science. Specifically, the data generated by high-throughput simulations have provided the fuel for the application of advanced generative deep learning methods to propose hypothetical new crystal structures.[1−3] These methods have made it relatively straightforward to generate proposals for crystal structures that satisfy a set of predefined performance targets. The resulting accumulation of candidate structures has given rise to the notion of a "synthesis bottleneck," referring to the gap between predicted hypothetical crystal structures and those that can be experimentally realized.

In response to this challenge, effort has been recently devoted to developing methods that help alleviate this bottleneck by guiding the synthesis route design. Among these methods are extensive natural language processing campaigns on published synthesis literature,[4−7] metrics of synthesizability based on thermodynamics,[8−12] algorithms for predicting synthesis outcomes,[13−15] and automated methods for designing synthesis routes.[6,16−20] Many of these approaches rely on thermodynamic features such as formation energies, phase diagrams, and convex hull geometry. While these features contain information about synthesizability,[11] thermodynamics alone do not provide sufficient predictive power to determine whether a metastable polymorph is kinetically stabilized as compared to other, more stable, polymorphs.[10] Furthermore, while dynamically unstable polymorphs can be excluded through full phonon calculations, this criterion alone does not guarantee the elimination of all kinetically nonpersistent structures. Even in well-studied binary systems such as Fe_2O_3 and TiO_2, a substantial fraction of unobserved, low-energy hypothetical polymorphs exhibit no imaginary phonon modes, indicating true metastability.[10] This raises a fundamental question: Why do many polymorphs that are seemingly thermodynamically viable and dynamically stable remain experimentally unrealized?

In previous investigations of kinetic effects on solid-state synthesis outcomes, Karan et al. identified diffusion in amorphous intermediates as a descriptor of product selectivity[15] and Aykol et al. developed a metric based on structural similarity to predict the ease of nucleation of a product phase.[21] Additionally, several methods use the

Received: June 4, 2026
Revised: August 12, 2026
Accepted: August 13, 2026

topography of the potential energy landscape (PES) to determine whether a given polymorph can be synthesized. One example of this approach connects the "basin hypervolume", or the extent of the region of configuration space for which a given polymorph is an attractor, to the synthesizability of that polymorph.[22,23] Another approach correlates the synthesizability of a polymorph with the absence of a facile transformation pathway from that phase to a lower-energy configuration.[24] It is this last mode of kinetic persistence that we investigated in this study.

Although phase transitions in periodic solids (e.g., martensitic transformations) are more often mediated by nucleation and growth than by diffusionless transformations, previous work has illustrated that the energy barriers associated with such transformations can provide descriptors regarding the kinetic stabilization of polymorphs.[24−26] The primary challenge in determining energy barriers for such transformations lies in efficiently exploring the vast potential energy surface to identify the relevant transformation pathways. Methods for addressing this problem range from sampling methods such as transition path sampling[27] and stochastic surface walking (SSW)[28] to methods based on structure matching and symmetry[29,30] to the nudged elastic band (NEB)[31] path relaxation technique. These methods each overcome the vastness of the search space using different pruning approaches: 1) transition path sampling methods iteratively perturb and reroute specific trajectories connecting two end points; 2) surface walking methods use bias potentials to drive a Monte Carlo PES exploration away from explored regions along soft modes; and 3) NEB locates the optimal form of an initial guess for the transformation trajectory.

Among these approaches, the nudged elastic band (NEB) method[31] and its extensions, climbing-image NEB (CI-NEB),[32] solid-state NEB (ssNEB),[33] and variable-cell NEB (VC-NEB),[34] stand out as the most widely used for identifying solid−solid transition pathways.[35−38] In this method, a number of intermediate configurations (images) are jointly relaxed in the presence of an added artificial spring force, which keeps them roughly evenly spread out in configuration space along the reaction coordinate. NEB and its variants are effective, but rely on the provisioning of an initial pathway guess. Stevanović et al. address this issue with an algorithm for identifying the unit cell correspondence between end point structures that minimizes the changes in coordination number (# of broken bonds) along the pathway. They show that this method yields high quality, low energy pathways between a number of polymorph pairs and provides evidence for the hypothesis that metastable polymorphs that are connected to a lower energy structure via a facile diffusionless transition pathway are less frequently observed experimentally.[24]

Here, we continue investigating this hypothesis and introduce an algorithm that identifies low energy pathways between polymorphs without relying on *any* prior information regarding the nature of the transition. In this respect, our method differs from each of the approaches above: unlike NEB (and the structure-mapping approach of Stevanović et al., which supplies such methods with a starting correspondence between end points), it requires neither an initial pathway guess nor a precomputed end point correspondence; and unlike stochastic surface walking, it performs a directed, deterministic search between two specified end points rather than a stochastic exploration of the potential energy surface.

This method is based on the Crystal Normal Form, which provides a graph representation of crystal configuration space, resolves the ambiguity of unit cell choice,[39] and enables the traversal of crystallographic phase space without the risk of revisiting the same structure. In their original work, Mrdjenovich and Persson utilize a water-filling algorithm that exhaustively searches the phase space graph to identify the lowest energy transition point separating two polymorph end points.[39] Such an exhaustive approach is only applicable to simple structures and suffers from combinatorially growing computational requirements as the number of atoms in the unit cell increases. Here, we combine the CNF graph with an algorithm for identifying low-energy transition pathways and apply the combined framework to the TiO_2 system. As previously noted, the TiO_2 system provides an excellent test case, as it includes several technologically important polymorphs (rutile, anatase, and brookite), high-pressure phases, and a range of dynamically stable yet experimentally unobserved ("theoretical") polymorphs within the same energetic window as the known structures.[10] We identify diffusionless transformation pathways and energy barriers separating the theoretical and known polymorphs and correlate the height of these energy barriers with experimental observations.

DESCRIPTION OF THE PATHFINDING ALGORITHM

The Crystal Normal Form (CNF) is an ordered list of integers that uniquely describes a crystal structure and resolves the ambiguities associated with the conventional unit cell representation. In the CNF, the lattice is represented by the lengths of seven "Voronoi vectors" (lattice generating vectors whose bisecting planes intersect the Wigner−Seitz cell) and the motif is represented by a list of relative atomic coordinates. By construction, if two unit cells (lattice/motif pairs) generate the same crystal structure, their CNF representations will be identical. In this sense, the CNF acts as a fingerprint for a crystal structure and as a method of identifying distinct structures.

In addition to this capacity for disambiguation, Mrdjenovich et al. developed a method for enumerating a set of neighbors for a given CNF point.[39] These neighbors are computed by incrementing or decrementing the integer values in the CNF coordinate list for a given structure. Adjusting the first seven coordinates of the list, which correspond to the lattice, applies a small strain to the structure. Adjusting the remaining coordinates applies a small displacement to one of the atoms in the structure. By repeatedly identifying neighboring structures (and neighbors-of-neighbors), this method enables an exhaustive enumeration of all crystal configurations and produces a graph ("the CNF graph") within which similar configurations are connected. Paths between points in this graph correspond to transition pathways composed of small strains or atomic displacements connecting polymorphs. These paths hold tremendous expressive power because these lattice and motif steps can be interleaved in whatever way is necessary to connect the end points. The magnitude of the structural differences between neighboring points in the graph is determined by two discretization parameters, ξ and δ, which determine the sizes of the lattice strain steps and atomic displacement steps, respectively.

As previously mentioned, for the simplest unit cells (<2 atoms per cell), it is possible to exhaustively search this CNF graph using a water-filling algorithm in conjunction with

https://doi.org/10.1021/acs.jctc.6c01114
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

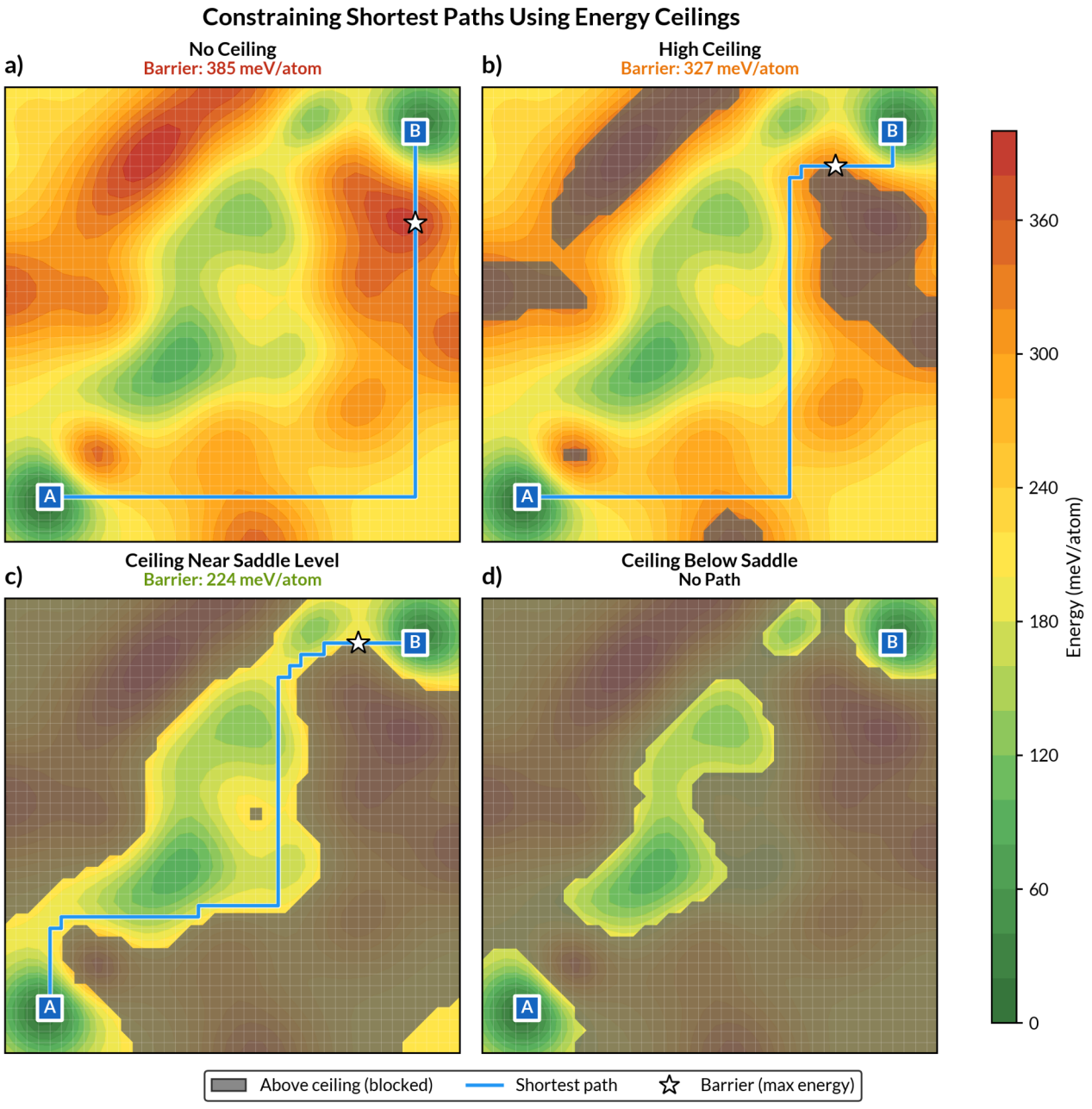


Figure 1. Ceiling lowering algorithm illustrated on a schematic energy landscape. The topography of the landscape in this figure was procedurally generated for illustration purposes but does not correspond to any real potential energy surface. The two square end points at local minima (A and B) illustrate how metastable polymorphs occupy local minima on the surface. In panels a)−d), gray areas correspond to configurations with energies above the energy ceiling for that panel. The blue paths are produced using the A* shortest-path finding algorithm on the accessible region of each space while only allowing steps to immediate neighboring squares. a) Without any energy ceiling limit imposed, the shortest path traverses an energy maximum, leading to a high "barrier". b), c) As the ceiling approaches the highest saddle separating the two end points, the shortest path gets closer to traversing the true saddle point. d) If the ceiling is lower than the true saddle point, no percolating pathway exists, and no transformation pathway is found. The stars mark the highest energy points encountered on each pathway.

density functional theory (DFT) calculations for assessing the energy of each point to first-principles accuracy. In this water-filling algorithm, the search begins at a specified initial configuration and systematically explores neighboring configurations in order of increasing energy. By exploring nodes in order of increasing energy, the algorithm ensures that a configuration is only explored after all accessible lower-energy configurations have been visited. As a result, when this algorithm reaches a saddle point, it will repeatedly explore the lowest energy point on the opposite side of the saddle until the local minimum for that basin is reached. If that basin is the end point, the algorithm will terminate. If that basin is not the end point, the algorithm will proceed as before: by exhaustively exploring points in order of ascending energy until yet another saddle point is reached and the algorithm's search frontier (the "water") expands ("flows") down into the next basin.

Such an approach is attractive because it is guaranteed to find the lowest-energy barrier separating two polymorphs (subject to the discretization parameters), but it quickly becomes infeasible due to the high compute cost of DFT

https://doi.org/10.1021/acs.jctc.6c01114
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

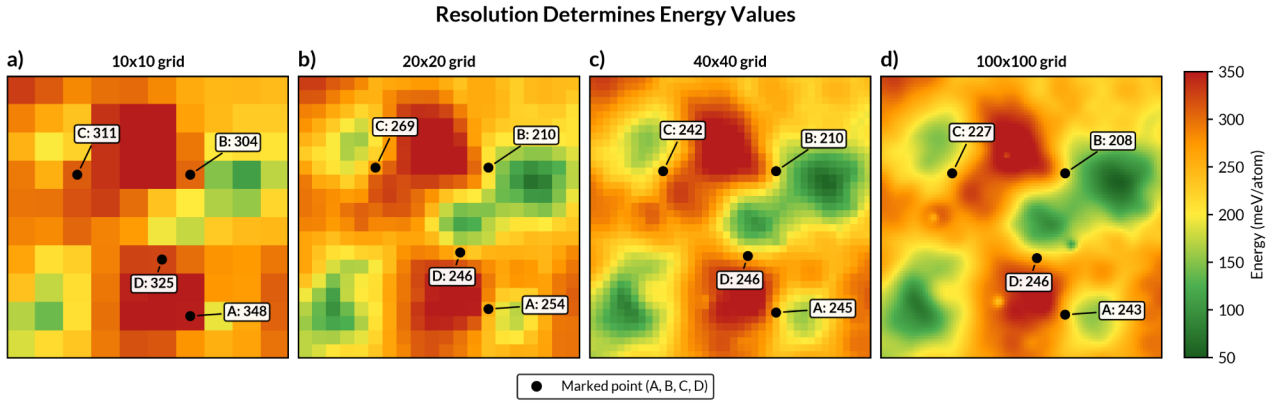


Figure 2. Resolution affects landscape features. Panels a) through d) show increasingly high resolution views of the same schematic potential energy landscape. At lower resolutions, fine details are not visible. Because configurations are mapped to their nearest points under CNF discretization, the assigned energy of a configuration depends on the chosen discretization parameters. This is illustrated here by the changing energy values associated with points labeled A, B, C, and D under different discretization levels.

calculations. This limitation can be partially mitigated by replacing DFT with an MLIP surrogate; however, the search space still grows combinatorially with system size, restricting practical applications to only modestly larger unit cells (typically <4−6 atoms, depending on the complexity of the problem).

An alternative method uses the A* shortest-pathfinding algorithm[40] to quickly (CPU seconds/mins) find *short* paths between polymorphs. The A* pathfinding algorithm is a classical graph search algorithm that identifies a shortest path between a set of source points and target points in a graph. Like all pathfinding algorithms, it iteratively explores points radiating out from the set of source points, keeping track of the distance required to reach each point from the source. The defining feature of A* is its use of a guiding heuristic to efficiently decide which points among those on the frontier (the unexplored points accessible from the explored points) should be explored next. Specifically, candidate points are ranked according to a score function, which is the sum of (a) the known length of the path required to reach the point from the source points and (b) a heuristic estimate of the remaining distance to the end points. The score function thus gives an estimate of the path length that might be achieved by exploring a particular point. In our use of the A* algorithm, we use the *minimum Manhattan distance between the point's CNF coordinate list and each of the end points CNF coordinate lists* as the guiding heuristic.

By prioritizing points for exploration that minimize this score function, A* selects points for exploration that appear *promising* (i.e., likely to yield a short path), increasing its efficiency over those of other path-finding algorithms. In contrast, for example, breadth-first search explores all points at a given distance before exploring any point at a greater distance. When breadth-first search is applied in the CNF case, the algorithm wastes immense computational time exploring points that are *farther* from the end points than the source itself, to such an extent that it fails to find any path (in a reasonable amount of time) in cases where the shortest path is long. It should be noted that all shortest-pathfinding algorithms are guaranteed to find a path (if one exists) given unlimited compute time, but in the case of the enormous, densely connected CNF graphs associated with large unit cells, most of them fail to terminate on time scales typical of computational materials science calculations.

As thus described, A* massively decreases the time required to find transformation pathways (compared to the water-filling approach described above), but it does not incorporate energetics and therefore cannot prefer pathways that traverse low-energy regions of configuration space. To address this, we modify the A* algorithm such that points can be excluded from the search space based on their energy. Specifically, when the list of candidate points for the next round of exploration is assembled from the neighbors of the previously explored points, an MLIP is used to quickly estimate the energy of each point. Points with energies above a predecided "ceiling" can be excluded from the candidate list, effectively limiting the A* algorithm to searching only low-energy points. In this way, the algorithm is diverted away from high-energy regions and forced to search for paths in regions where the energy stays under the ceiling value. Other filters can also be applied to divert the searcher. For example, in this work, we use a distance-based filter that excludes structures containing pairs of atoms that are much closer than their typical bond length. Such geometry-based filters are attractive when they can be evaluated more quickly than an MLIP energy calculation.

Finally, to search for *lowest-energy* pathways, we introduce an iterative refinement algorithm that applies A* pathfinding to progressively smaller subgraphs of the complete CNF graph by repeatedly lowering the energy ceiling mentioned above. Different energy ceilings leave different portions of the CNF graph accessible to the pathfinding algorithm. Specifically, for a given ceiling energy C, any path found must only visit nodes with a maximum energy value C. Once a path has been found under C achieving a maximum energy of $E_{max} < C$, the ceiling can be lowered to $E_{max} - \delta C$ (with $\delta C = 2\frac{\text{meV}}{\text{atom}}$, for example—though this is a tunable parameter), and the pathfinding process can be repeated. Thus, with each subsequent lowering of the ceiling and successful pathfinding attempt, a lower energy transition pathway is identified. A schematic showing how the shortest path varies with energy ceiling level on an artificial 2D potential energy landscape (generated procedurally for illustration purposes only) is shown in Figure 1. Pseudocode for the ceiling-constrained A* search and for the iterative ceiling-lowering procedure is provided in the Supporting Information (Algorithm S1 and Algorithm S2, respectively).

https://doi.org/10.1021/acs.jctc.6c01114
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Table 1. TiO_2 Polymorphs Selected for This Study[a]

MP ID	Phase Name	Other Names	Space Group	Atoms	$E_{hull}\left(\frac{meV}{atom}\right)$	Status
Naturally ubiquitous (ambient pressure)						
mp-390	Anatase	β-TiO_2	$I4_1/amd$	6	0	Observed
mp-1840	Brookite	γ-TiO_2	$Pbca$	24	3	Observed
mp-2657	Rutile	α-TiO_2	$P4_2/mnm$	12	44	Observed
Observed (ambient pressure)						
mp-554278	TiO_2–B	—	$C2/4m$	12	6	Observed
Observed (high pressure)						
mp-1439	Columbite	α-PbO_2-type, TiO_2–II	$Pbcn$	12	5	Observed
mp-430	Baddeleyite	TiO_2–III	$P2_1/c$	12	39	Observed
Theoretical						
mp-34688	—	—	$C2/c$	6	15	Theoretical
mp-2420244	—	—	$Pnma$-I	12	16	Theoretical
mp-754769	—	—	$Pnma$-II	24	36	Theoretical

[a]Experimental polymorphs are the three well-known ambient-pressure phases. "Observed" indicates structures with entries in the Inorganic Crystal Structure Database (ICSD), occasionally synthesized under high pressure conditions . Theoretical structures have no experimental reports in the ICSD.

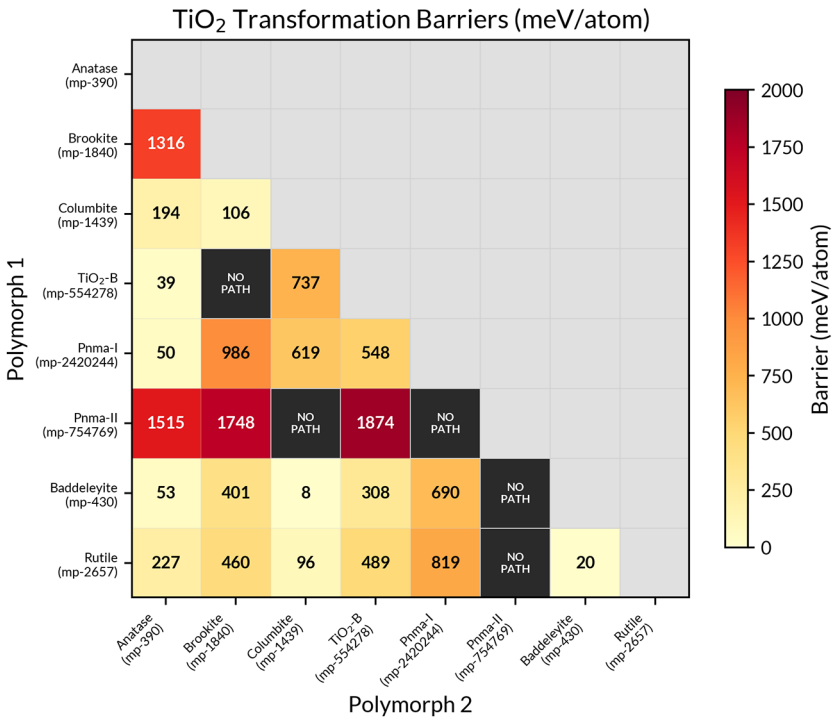


Figure 3. Energy barriers, as assessed by an MLIP trained on data at the DFT-GGA level of theory between all pairs of polymorphs considered in $\frac{meV}{atom}$. All values are in units of $\frac{meV}{atom}$.

It is important to remember that the CNF energy landscape is a discrete space. As a result, features of the landscape may appear or disappear as the coarseness of the discretization is adjusted. Typically, finer low-energy features (for example, narrow saddle points or low-energy valleys) become more accessible at higher resolutions, as shown schematically in Figure 2. If, after several failed pathfinding attempts at a given energy ceiling, no path is found, the resolution of the search space can be increased to reveal finer details of the energy surface, and a new round of attempts can commence. Ultimately, after a prescribed number of iterations, the algorithm terminates, and the lowest ceiling under which a pathway was found is treated as an upper bound on the true pathway barrier.

RESULTS AND DISCUSSION

At present, the Materials Project reports 46 polymorphs of TiO_2;[41] however, recognizing thermodynamic constraints, we restrict this study to polymorphs within a defined energy-above-hull window ($50\ \frac{meV}{atom}$) based on the three energies of the most common naturally occurring phases: anatase, brookite, and rutile. This filtering procedure yields 10 polymorphs, including the three most well-known experimental phases and seven other polymorphs (including columbite and baddeleyite, which have been reported in high-pressure (hp) experiments, as shown in Table 1). We filtered this set to polymorphs with 24 or fewer atoms (shown in Table 1), which led to the exclusion of the orthorhombic mp-775938 phase (36 atoms/

https://doi.org/10.1021/acs.jctc.6c01114
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

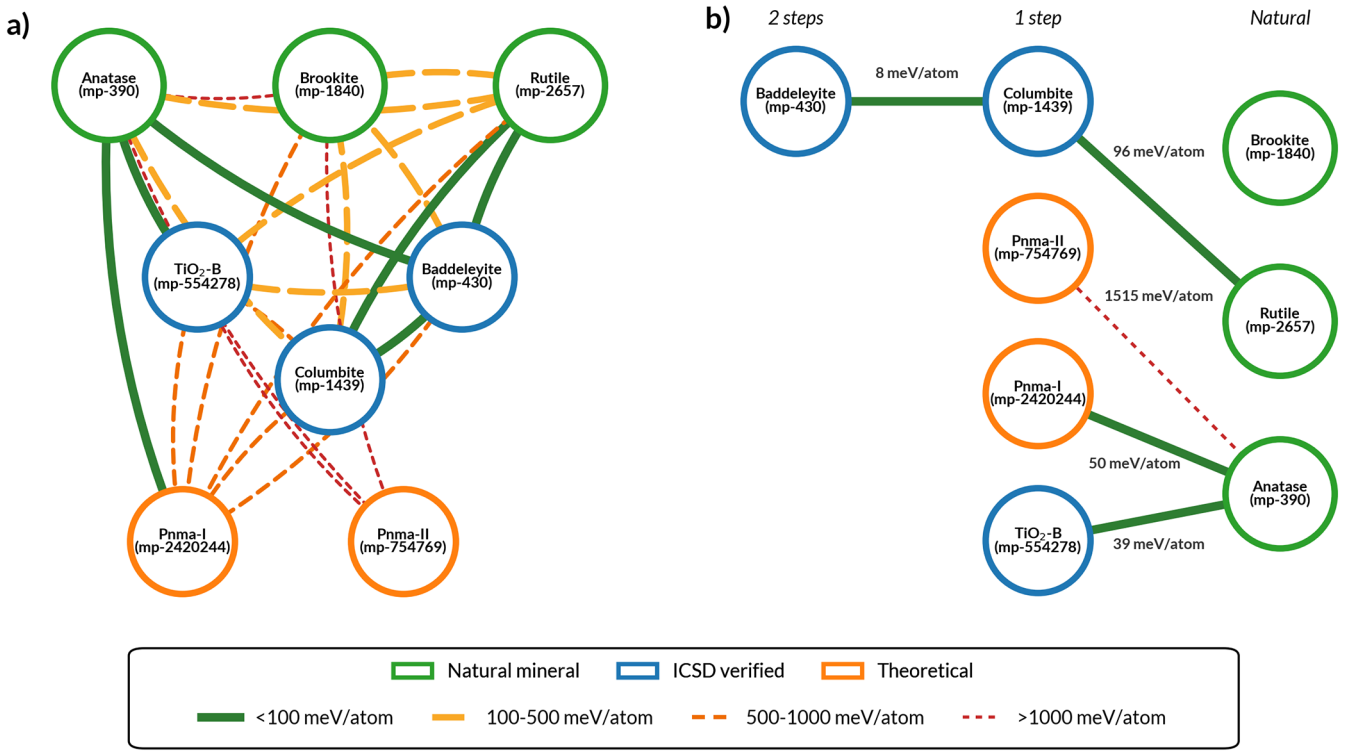


Figure 4. Qualitative pathway difficulty between polymorphs. The weight, color, and style of the edges connecting nodes reflect the height of the energy barrier between the phases they connect. a) All pathways identified between polymorphs. b) Depiction of the lowest energy pathways separating polymorphs from the naturally occurring phases. In the case of columbite, the lowest energy pathway is a 2-step pathway.

primitive cell). We also ultimately excluded mp-34688 from the polymorph list because the CNF revealed that this structure constitutes a minor distortion of anatase. Specifically, the transformation paths connecting this phase to anatase were <10 steps long, even at a fine discretization. This similarity was confirmed by the structure matching algorithm in pymatgen[42] which reported a match and the Earth Mover's Distance on the Pointwise Distance Distribution[43] representation (0.13 Å).

Any discussion of the energetics of TiO_2 polymorphs should address the well-known variations in the energetic ordering of these polymorphs computed using various DFT methods.[44−46] Specifically, under the generalized gradient approximation[47] (GGA), anatase is identified as the ground state, and brookite is lower in energy than rutile. Interestingly, more advanced methods, including diffusion Monte Carlo and finite-temperature approaches, also predict anatase as the ground state at 0 K and rutile as the stable phase at elevated temperatures, while brookite remains higher in energy than both. Notably, all three polymorphs are separated by only small 0 K energy differences, spanning a range of approximately $44\frac{\text{meV}}{\text{atom}}$ under GGA (and only $14\ \frac{\text{meV}}{\text{atom}}$ when computed by fixed-node diffusion quantum Monte Carlo). This close competition among rutile, brookite, and anatase is reflected in reports of successful synthesis and persistence of all three phases under ambient conditions.[48−50]

Because we utilize MLIPs trained and fine-tuned on GGA calculations for this work, the barrier heights that we compute are subject to a similar degree of error reflected in the incorrect ordering of these phases. We take this uncertainty to be roughly $50\ \frac{\text{meV}}{\text{atom}}$ because this is the energy range spanned by the anatase, brookite, and rutile phases (wherein the ordering is incorrect) under GGA. As a result, and particularly given the additional uncertainty associated with using a fine-tuned MLIP to compute these pathways, we interpret our results qualitatively rather than as precise quantitative estimates of the barrier heights.

The barrier heights computed between all pairs of polymorphs are listed in Figure 3. To map the specific barrier heights we measured to pathway feasibility, we refer to the oxygen-deficient columbite (mp-1439) phase, which is formed under high-pressure conditions. While it is possible to produce this phase using high-pressure synthesis conditions,[51] it tends to decompose under ambient conditions. Kinetic stabilization via nanocrystalline refinement can trap the metastable phase, but even optimized high-pressure sintering and torsion methods achieve only 85 wt% and 69 wt % columbite, respectively.[52−54] The marginal kinetic persistence of columbite is also reflected in the fact that it does occur naturally, but only very rarely and only in settings with extreme geological histories, such as asteroid or meteorite impacts.[55−57] As a result, we use the barrier that we identified between columbite and rutile ($96\ \frac{\text{meV}}{\text{atom}}$) as the basis for our choice of $100\ \frac{\text{meV}}{\text{atom}}$ as a rough upper bound for facile transformation. This value is used in Figure 4a,b, where pathways are expressed as connections in a network of transformations.

In Figure 4b, we show the lowest single pathways leading from the high-pressure or hypothetical phases to naturally occurring phases. We find facile pathways $\left(<100\frac{\text{meV}}{\text{atom}}\right)$ for all phases except *Pnma*-II (mp-754769, theoretical), which is only connected to brookite (mp-1840) via a very high-energy pathway. Notably, for baddeleyite (mp-430), there is a two-step transformation. Pathfinding attempts from *Pnma*-II to other polymorphs either failed to identify any pathway below

https://doi.org/10.1021/acs.jctc.6c01114
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

2000 $\frac{\text{meV}}{\text{atom}}$ or yielded only similarly high-energy routes $\left(>1500\frac{\text{meV}}{\text{atom}}\right)$.

Persistence of Anatase (mp-390), Brookite (mp-1840), and Rutile (mp-2657)

All three of anatase, brookite, and rutile TiO_2 are naturally occurring minerals.[58] According to a Quantum Monte Carlo study by Luo et al.,[44] the ground state TiO_2 phase at finite temperatures is rutile, and brookite is higher in energy than anatase at all temperatures. According to the central hypothesis of this work, we expect significant energy barriers to exist separating brookite from both anatase and rutile. Our results align with this expectation: we find high energy barriers separating brookite from rutile (460 $\frac{\text{meV}}{\text{atom}}$) and anatase (1316 $\frac{\text{meV}}{\text{atom}}$). Furthermore, experimental findings show that, upon heating to 500−600 °C, brookite transforms to rutile with little to no anatase byproduct.[59] This result suggests that the transformation pathway between brookite and rutile is more readily traversed than the one from brookite to anatase, and our results suggest the same: the 1316 $\frac{\text{meV}}{\text{atom}}$ barrier we find separating brookite from anatase is substantially higher than the 460 $\frac{\text{meV}}{\text{atom}}$ barrier separating brookite from rutile. Finally, using similar reasoning, we expect to find a barrier greater than 100 $\frac{\text{meV}}{\text{atom}}$ separating anatase from rutile, and our results also align with this expectation: we find a barrier of 227 $\frac{\text{meV}}{\text{atom}}$ between the two polymorphs. This result also agrees with previous computational studies, which report values of 233−333 $\frac{\text{meV}}{\text{atom}}$[60,61] at the GGA level of theory.

Transience of High-Pressure Columbite (mp-1439) and Baddeleyite (mp-430) Phases

The columbite and baddeleyite polymorphs of TiO_2 are well known as high-pressure phases that decay under ambient pressure to rutile or a combination of rutile and anatase.[53,62] Of these two phases, baddeleyite is the higher pressure phase,[63] and previous studies have shown that, upon decompression, it spontaneously decomposes to columbite[64] (before further decay to rutile occurs). Correspondingly, and according to our central hypothesis, one would expect to find only a low barrier ($<100\frac{\text{meV}}{\text{atom}}$ per our cutoff choice) separating baddeleyite from columbite. Our results indeed confirm this expectation, as our method computes a barrier of only 8 $\frac{\text{meV}}{\text{atom}}$ separating baddeleyite from columbite. This result also agrees with a previous study that found no barrier associated with the baddeleyite→columbite transformation.[62]

We also identify a facile pathway (barrier height 20 $\frac{\text{meV}}{\text{atom}}$) connecting baddeleyite to rutile. The higher barrier relative to the baddeleyite→columbite pathway is consistent with experimental observations that baddeleyite preferentially transforms to columbite rather than rutile. This result further confirms the findings of Wang et al.,[62] who also observed paths with barrier heights between 16 and 80$\frac{\text{meV}}{\text{atom}}$ for the baddeleyite ↔ rutile transformation. Finally, we note that, while previous work shows that columbite can be synthesized under high pressure and quenched to ambient pressure[53] it is difficult to obtain a phase pure product. In their recent work, Hidalgo-Jimenez et al. successfully synthesized columbite and achieved a result of 69 wt % columbite, and the remaining 31 wt % being anatase and rutile. The difficulty in achieving this, and the barrier of 96 $\frac{\text{meV}}{\text{atom}}$ we compute for this transformation, are consistent good agreement with a previous study of the same pathway, which reported a DFT-GGA barrier of 77 $\frac{\text{meV}}{\text{atom}}$[65] (though we note that this barrier was computed for compressed unit cells and may not be directly comparable).

Surprising Kinetic Persistence of TiO_2−B

The TiO_2−B polymorph has been successfully synthesized via topotactic and hydrothermal methods and explored as a candidate for photocatalysis and as an anode for Li-ion batteries.[61,66] The only experimentally recorded transformation pathway from TiO_2−B to another polymorph is the pathway from TiO_2−B to anatase.[61] This aligns well with our findings: we find a low-energy (39 $\frac{\text{meV}}{\text{atom}}$ barrier) transformation pathway connecting TiO_2−B to anatase, and pathways connecting TiO_2−B to all other polymorphs are high in energy (the next lowest energy transformation being the TiO_2−B ↔ baddeleyite transformation at 308 $\frac{\text{meV}}{\text{atom}}$). We do note, however, that the TiO_2−B→anatase barrier height computed by our method is somewhat lower than the value found by a DFT study, which reported a barrier height of 93 $\frac{\text{meV}}{\text{atom}}$ for the same transformation.[61] This finding appears to contradict our central hypothesis because, though the synthesis of TiO_2−B is somewhat difficult and sensitive to temperature (synthesis temperatures must be kept below 400−500 °C to avoid decomposition to anatase), it is possible to synthesize it. It therefore exhibits some degree of kinetic persistence despite the low barrier preventing its decomposition to anatase. It may be that other factors contribute to its stabilization, or that our estimated barrier of 39 $\frac{\text{meV}}{\text{atom}}$ is lower than the true value (as suggested by the previously mentioned DFT study of Lu et al.).[61]

Kinetic Persistence of *Pnma*-I and *Pnma*-II

We consider two, to the best of our knowledge, truly theoretical structures, *Pnma*-I (mp-2420244) and *Pnma*-II (mp-754769), in this study. We find a transformation pathway with a 50 $\frac{\text{meV}}{\text{atom}}$ barrier separating *Pnma*-I from anatase. If this low barrier height indeed inhibits the kinetic persistence of this phase, as our central hypothesis would suggest, the presence of this low energy pathway may explain the absence of any recorded observations of this phase. In contrast, the lowest energy barrier separating *Pnma*-II (mp-754769) from any other phase is 1515 $\frac{meV}{atom}$ (on the mp-754769 ↔ anatase pathway).

Kinetic Persistence of Brookite

It is notable that our method does not identify a low- or moderate-energy pathway between brookite and anatase, instead yielding a barrier of 1316 $\frac{\text{meV}}{\text{atom}}$. Since brookite is a persistent metastable polymorph, this is the expected result; however, it contrasts with prior studies that report substantially lower barriers, although those values themselves vary widely (73 $\frac{\text{meV}}{\text{atom}}$ in the stochastic surface walking study of Chen et al.[67] and 320 $\frac{\text{meV}}{\text{atom}}$ in Song et al.[60]). One possible explanation for this discrepancy is the increased complexity of the search

https://doi.org/10.1021/acs.jctc.6c01114
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

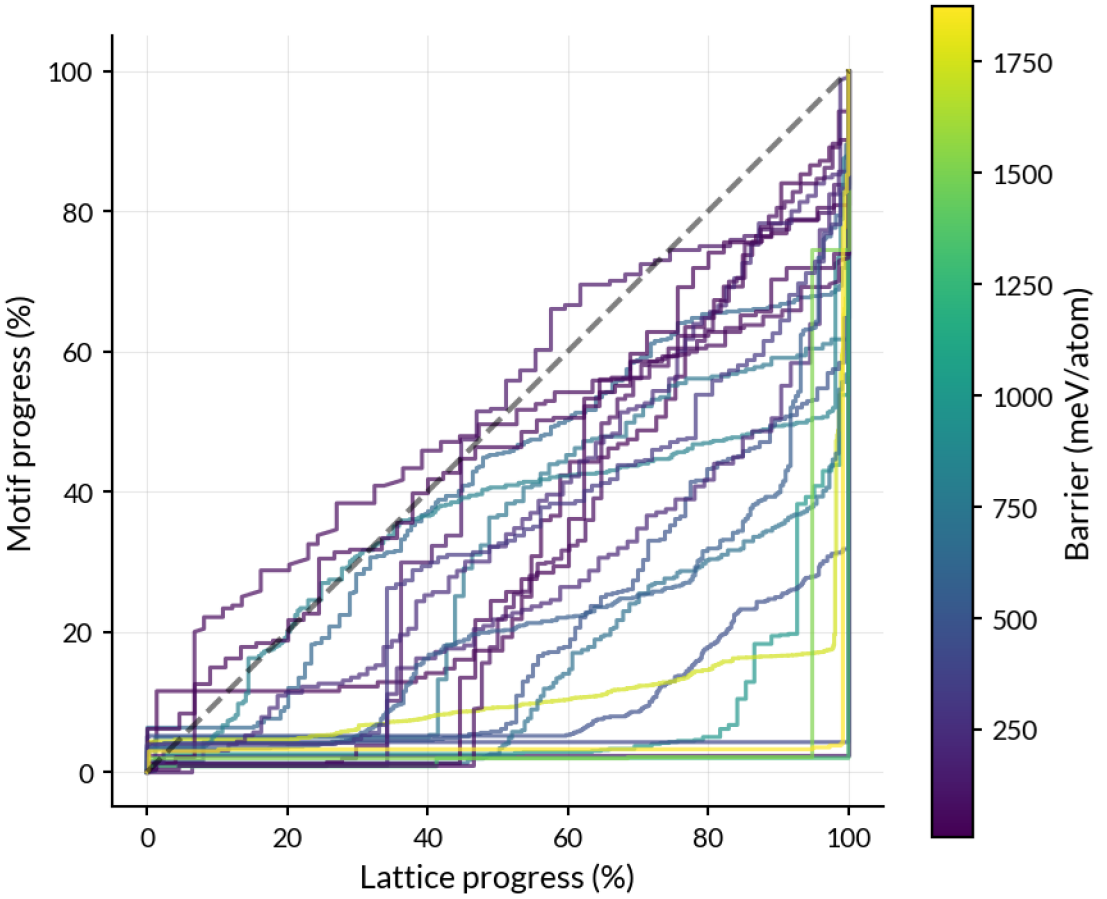


Figure 5. Pathway character and energy barrier height. One trace is shown for each polymorph pair for which a connecting path was identified. The height of the energy barrier associated with each path is given by the trace color. The percentage of motif and lattice progress is calculated by counting the number of steps in each category along each path.

associated with larger unit cells: we emphasize that the massive size and complexity of the discretized configuration space represented by the CNF graph for larger unit cells (both brookite and *Pnma*-II (mp-754769; theoretical) contain 24 atoms) increase the likelihood of missing low-energy pathways. This occurs due to several compounding factors: (i) path lengths are very large (>1000 steps at discretization levels that permit physically realistic configurations), (ii) the branching factor is extremely high (>120 neighbors per node), and (iii) the heuristic offers little to no discriminatory power among nodes in the search frontier, causing the A* algorithm to effectively degenerate into a uniform-cost search. That said, the algorithm is guaranteed to find a path given enough computational time, even at lower energy ceilings, so long as one exists. In light of this, while we elected to use a 24-h cutoff and a procedurally determined maximum number of iterations in the A* pathfinding algorithm efforts in this study, future valuable work will entail modifications of the A* algorithm used here, adjustments of the pathfinding parameters, and experiments using increased computational budgets.

A distinguishing feature of our pathfinding algorithm is that it produces paths composed of complex mixes of strains and atomic shuffles. It is straightforward to categorize each step as either a "lattice step" (a small strain) or a "motif step" (a small atomic shuffle). It is then possible to count the total number of steps of each type and measure the progress along each dimension over the course of the transformation. We illustrate this in Figure 5. Paths that stay near the diagonal have more evenly dispersed motif and lattice steps, while paths that stay near the *x*-axis begin with a distinct period of unit-cell deformation and end with a distinct period of atomic shuffle. Lower-energy pathways cluster near the diagonal, indicating that pathways with more balanced, concurrent strain and shuffle components are energetically favored.

Pathways involving 24-atom unit cells (including all paths with *Pnma*-II, as well as brookite ↔ *Pnma*-I, brookite ↔ anatase, and brookite ↔ TiO_2−B) consistently exhibit a two-stage character: an initial lattice distortion followed by motif distortion (Figure 5). This behavior likely reflects the interaction between the Manhattan-distance heuristic used in the A* algorithm and the high-dimensional CNF representation of larger unit cells. In particular, canonicalization can reorder motif coordinates during lattice changes, disrupting alignment with the target motif and increasing the apparent distance to the goal. This effect is amplified with system size, as motif coordinates scale as $3(n_{atoms}\text{-}1)$ while lattice degrees of freedom remain fixed, biasing the heuristic toward early lattice alignment and, consequently, higher-energy pathways. This bias may contribute to discrepancies with previously reported low-energy pathways.[67]

In light of these variations in path character, we attribute the discrepancies between our results and previous findings to differences in the identified paths themselves, rather than inaccuracies in the calculated energies along each path. If systematic energy errors were to blame, we would expect comparable discrepancies for large and small cells alike;

https://doi.org/10.1021/acs.jctc.6c01114
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

instead, our barriers agree with prior work for the structurally simpler phases and diverge only for the larger 24-atom cells, where the search is least likely to have located the true lowest-energy path.

Runtime Characteristics of the Algorithm

We note that the runtime required by our method is determined by three major contributions. In rough order of runtime requirements, these are (1) energy evaluations using the MLIP, (2) CNF neighbor computation, and (3) A*-heuristic evaluations. In general use, we expect that the energy evaluations will constitute the bulk of the runtime of our method. In the case study presented here, each individual A* search required on the order of 10^5 energy evaluations, a cost that grew with the dimensionality of the search: the 40-dimensional (12-atom) searches required roughly 10^4 to 10^5 evaluations each, and the 76-dimensional (24-atom) searches several times more, roughly 10^5 to 10^6. The runtime associated with these evaluations will depend on the type of MLIP used, and for the simplest systems, it may even be feasible to use DFT calculations (as in the original Mrdjenovich and Persson study of Zr[39]). We also note that, in this study, the Manhattan-distance heuristic required almost no computation. Since we assert that new heuristics may improve the ability of the method to identify low-energy paths, it is important to highlight that the runtime of such new candidate heuristics should be evaluated, since a slow heuristic can render the algorithm infeasible.

METHODS

MLIP Finetuning

Because millions of energy evaluations are required for these pathfinding efforts, a fast energy evaluator is needed. Explicit DFT calculations are orders of magnitude too slow (minutes to 1 h using GGA for configurations up to 24 atoms). Furthermore, we found that graph neural network MLIPs such as the message-passing atomic cluster expansion (MACE)[68] were 5−10 times slower than linear graph atomic cluster expansion (GrACE)[69] models. As a result, we elected to distill a GrACE model from a finetuned MACE model that was itself fine-tuned on a DFT dataset for configurations taken from the central regions of a first pass of transformation pathways identified using our method. This workflow is illustrated in Figure S3.

To produce DFT fine-tuning training data for this workflow, we executed the pathfinding algorithm on every pair of TiO_2 polymorphs enumerated in Table 1 and utilized a pretrained GRACE-FS-OAM model as the energy evaluator. From each of the resulting successful paths, we sampled 10 frames from the central regions of the path. We computed the energies of those configurations by using the Vienna Ab Initio Simulation Package (VASP) parameterized by the MPGGAStaticMaker[42,70] `atomate2` workflow which uses the generalized gradient approximation (GGA) functional of Perdew, Burke, and Ernzerhof.[47] We separated the resulting data set into a 90/10 train/test split and fine-tuned the MACE-MP-0 (small) foundation model using the training portion of this data set. Using the resulting model, we produced a large training set consisting of every frame from the initial pathfinding procedure and fine-tuned GRACE-FS-OAM on this data set, effectively distilling the finetuned MACE model into a linear GrACE model. The finetuned MACE model achieved high accuracy with an error of 12 $\frac{\text{meV}}{\text{atom}}$ on the held-out test set. The distillation process improved the accuracy of the pretrained GrACE model on the same DFT test set from 184 $\frac{\text{meV}}{\text{atom}}$ (pretrained) to 19.1 $\frac{\text{meV}}{\text{atom}}$ (finetuned). The parity plots vs the MACE-generated training set and the ground-truth DFT test set are shown in Figures S4 and S6.

Application of the Pathfinding Algorithm

To identify pathways between each pair of polymorphs from Table 1, we performed a parallel search procedure. Specifically, we selected a range of 32 ceiling energies evenly spaced between 0 $\frac{\text{meV}}{\text{atom}}$ and 2000 $\frac{\text{meV}}{\text{atom}}$, and assigned them to each of 32 worker processes, each performing its own refinement beginning at its own ceiling. Because the polymorphs under investigation do not all have the same number of atoms per unit cell, supercells were produced using the method of Mrdjenovich[71] such that the end point CNFs for each polymorph pair had matching dimensionalities. For example, the 6-atom anatase unit cell must be doubled to a 12-atom supercell for compatibility with polymorphs whose unit cells contain 12 atoms. This yielded dimensions of either 40 (for pairs with 12 atoms) or 76 (for pairs with 24 atoms). The dimensionalities used in the searches for each pair of polymorphs we investigated are reported in Table S1. We used a uniform lattice step size of ξ = 0.75 Å^2 and selected δ values such that the maximum atom step size with respect to the initial structure was 0.1 Å. We note that because the largest atom step size is given by the largest lattice parameter divided by δ, the size of an atom step changes throughout the pathway, as the lattice changes. Nevertheless, starting with an initial value of δ corresponding to a small atom step leads to the real step size remaining small throughout the pathfinding procedure. The parallel worker processes proceed in parallel, each iteratively reducing their own ceilings as pathways are found. The trajectories of these ceiling lowerings are shown in Figure S2. We report the lowest energy barrier pathway found for each pathway in Figure 3. We allowed each of these processes to run for 24 h. To manage the supercomputing resources used for this effort, we leveraged the `jobflow` and `jobflow-remote` Python packages.[72]

CONCLUSION

The algorithm introduced in this study extends the capabilities of the existing CNF path-finding approach and differs from other pathfinding methods in that it requires no prior input on the nature of the transition path. We illustrate its use in the rich TiO_2 system, within which we identify many low-energy transformation barriers. Our findings provide plausible mechanistic insight into the variations in the kinetic persistence of five metastable TiO_2 polymorphs.

This algorithm substantially expands the range of systems for which transformation pathway studies can be performed using the CNF, and we believe that it represents only a small portion of the potential utility of the CNF in the discovery of diffusionless transformations and in computational materials science, more generally. Future work should entail improvements to this pathfinding method such as unbiased search heuristics, intelligent neighbor pruning functions, fast data set generation for MLIP training, efficient methods for changing discretization resolution midsearch, and improvements to the pathfinding algorithm itself.

ASSOCIATED CONTENT

Data Availability Statement

Data and Software Availability Statement: The Crystal Normal Form pathfinding software used in this study is available at https://github.com/mcgalcode/crystal-normal-form. The transformation pathways, energy barriers, and associated data reported in this work are available at https://github.com/mcgalcode/tio2-paths-data.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.6c01114.

https://doi.org/10.1021/acs.jctc.6c01114
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Pseudocode for the ceiling-constrained A* search and the iterative refinement algorithm; table of CNF search-space dimensionalities for each polymorph pair; individual plots for lattice/motif pathway character; parity plots for the finetuned MLIP; and example of ceiling-lowering traces for parallel workers (PDF)

AUTHOR INFORMATION

Corresponding Author

Kristin A. Persson − *Department of Materials Science and Engineering, University of California, Berkeley, California 94720, United States; Materials Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, United States;* orcid.org/0000-0003-2495-5509; Email: kristinpersson@berkeley.edu

Authors

Max C. Gallant − *Department of Materials Science and Engineering, University of California, Berkeley, California 94720, United States; Materials Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, United States;* orcid.org/0009-0008-4099-6144

David Mrdjenovich − *Department of Materials Science and Engineering, University of California, Berkeley, California 94720, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jctc.6c01114

Author Contributions

M.C.G.: conceptualization (lead); software (lead); writing—original draft (lead). D.M.: conceptualization; software; writing—review and editing. K.A.P.: supervision; funding acquisition; writing—review and editing.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

M.C.G. acknowledges support from the Kavli Energy Nano-Science Institute (Kavli ENSI) Graduate Student Fellowship. K.P. acknowledges support from the Materials Project program KC23MP, supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, Materials Sciences and Engineering Division, under contract No. DE-AC02-05-CH11231. This research used resources of the National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility located at Lawrence Berkeley National Laboratory, operated under Contract No. DE-AC02-05CH11231. Additional computational resources were provided by the Lawrencium computational cluster at Lawrence Berkeley National Laboratory.

REFERENCES

(1) Xie, T.; Fu, X.; Ganea, O.-E.; Barzilay, R.; Jaakkola, T. S. Crystal diffusion variational autoencoder for periodic material generation. *arXiv* **2021**.

(2) Zeni, C.; et al. *Nature* **2025**, *639*, 624−632.

(3) Luo, X.; Wang, Z.; Wang, Q.; Shao, X.; Lv, J.; Wang, L.; Wang, Y.; Ma, Y. CrystalFlow: a flow-based generative model for crystalline materials. *Nat. Commun.* **2025**, *16*, 9267.

(4) Kononova, O.; Huo, H.; He, T.; Rong, Z.; Botari, T.; Sun, W.; Tshitoyan, V.; Ceder, G. Text-mined dataset of inorganic materials synthesis recipes. *Sci. Data* **2019**, *6*, 203.

(5) Cruse, K.; Trewartha, A.; Lee, S.; Wang, Z.; Huo, H.; He, T.; Kononova, O.; Jain, A.; Ceder, G. Text-mined dataset of gold nanoparticle synthesis procedures, morphologies, and size entities. *Sci. Data* **2022**, *9*, 234.

(6) He, T.; Huo, H.; Bartel, C. J.; Wang, Z.; Cruse, K.; Ceder, G. Precursor recommendation for inorganic synthesis by machine learning materials similarity from scientific literature. *Sci. Adv* **2023**, *9*, No. eadg8180.

(7) Zhang, J.; Zhang, L.; Sun, Y.; Li, W.; Quhe, R. *Comput. Mater. Sci.* **2024**, *240*, 113014.

(8) McDermott, M. J.; McBride, B. C.; Regier, C. E.; Tran, G. T.; Chen, Y.; Corrao, A. A.; Gallant, M. C.; Kamm, G. E.; Bartel, C. J.; Chapman, K. W.; Khalifah, P. G.; Ceder, G.; Neilson, J. R.; Persson, K. A. *ACS Cent. Sci.* **2023**, *9*, 1957−1975.

(9) Szymanski, N. J.; Byeon, Y.-W.; Sun, Y.; Zeng, Y.; Bai, J.; Kunz, M.; Kim, D.-M.; Helms, B. A.; Bartel, C. J.; Kim, H.; Ceder, G. Quantifying the regime of thermodynamic control for solid-state reactions during ternary metal oxide synthesis. *Sci. Adv.* **2024**, *10*, No. eadp3309.

(10) 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*, No. e1600225.

(11) Aykol, M.; Dwaraknath, S. S.; Sun, W.; Persson, K. A. Thermodynamic limit for synthesis of metastable inorganic materials. *Sci. Adv.* **2018**, *4*, No. eaaq0148.

(12) Tolborg, K.; Klarbring, J.; Ganose, A. M.; Walsh, A. *Digital Discovery* **2022**, *1*, 586−595.

(13) McDermott, M. J.; Dwaraknath, S. S.; Persson, K. A. A graph-based network for predicting chemical reaction pathways in solid-state materials synthesis. *Nat. Commun.* **2021**, *12*, 3097.

(14) Gallant, M. C.; McDermott, M. J.; Li, B.; Persson, K. A. *Chem. Mater.* **2025**, *37*, 210−223.

(15) Karan, V.; Gallant, M. C.; Fei, Y.; Ceder, G.; Persson, K. A. I. Ion correlations explain kinetic selectivity in diffusion-limited solid-state synthesis reactions. *Nat. Mater.* **2026**, *25*, 1385−1392.

(16) Szymanski, N. J.; et al. *Nature* **2023**, *624*, 86−91.

(17) Chen, J.; Cross, S. R.; Miara, L. J.; Cho, J.-J.; Wang, Y.; Sun, W. *Nat. Synth* **2024**, *3*, 606−614.

(18) Wang, Z.; Kononova, O.; Cruse, K.; He, T.; Huo, H.; Fei, Y.; Zeng, Y.; Sun, Y.; Cai, Z.; Sun, W.; Ceder, G. Dataset of solution-based inorganic materials synthesis procedures extracted from the scientific literature. *Sci. Data* **2022**, *9*, 231.

(19) Wang, Z.; Sun, Y.; Cruse, K.; Zeng, Y.; Fei, Y.; Liu, Z.; Shangguan, J.; Byeon, Y.-W.; Jun, K.; He, T.; Sun, W.; Ceder, G. *Nat. Synth* **2024**, *3*, 527−536.

(20) Ding, H.; Dwaraknath, S. S.; Garten, L.; Ndione, P.; Ginley, D.; Persson, K. A. *ACS Appl. Mater. Interfaces* **2016**, *8*, 13086−13093.

(21) Aykol, M.; Montoya, J. H.; Hummelshøj, J. *J. Am. Chem. Soc.* **2021**, *143*, 9244−9259.

(22) Novick, A.; Nguyen, Q.; Jankousky, M.; Tellekamp, M. B.; Toberer, E. S.; Stevanović, V. *J. Am. Chem. Soc.* **2025**, *147*, 4419−4429.

(23) Stevanović, V. *Phys. Rev. Lett.* **2016**, *116*, 075503.

(24) Stevanović, V.; Trottier, R.; Musgrave, C.; Therrien, F.; Holder, A.; Graf, P. *Phys. Rev. Mater.* **2018**, *2*, 033802.

(25) Schön, J. C.; Putz, H.; Jansen, M. *J. Phys.: condens. Matter* **1996**, *8*, 143.

(26) Therrien, F.; Jones, E. B.; Stevanović, V. Metastable materials discovery in the age of large-scale computation. *Appl. Phys. Rev.* **2021**, *8*, 031310.

(27) Zhu, L.; Cohen, R. E.; Strobel, T. A. *J. Phys. Chem. Lett.* **2019**, *10*, 5019−5026.

(28) Shang, C.; Liu, Z.-P. *J. Chem. Theory Comput.* **2013**, *9*, 1838−1845.

(29) Wang, F.-C.; Ye, Q.-J.; Zhu, Y.-C.; Li, X.-Z. *Phys. Rev. Lett.* **2024**, *132*, 086101.

(30) Capillas, C.; Perez-Mato, J. M.; Aroyo, M. I. *J. Phys.: condens. Matter* **2007**, *19*, 275203.

(31) Jonsson, H.; Mills, G.; Jacobsen, K. W. *Classical and Quantum Dynamics in Condensed Phase Simulations*; WORLD SCIENTIFIC, 1998; pp. 385−404.
(32) Henkelman, G.; Uberuaga, B. P.; Jónsson, H. *J. Chem. Phys.* **2000**, *113*, 9901−9904.
(33) Sheppard, D.; Xiao, P.; Chemelewski, W.; Johnson, D. D.; Henkelman, G. A generalized solid-state nudged elastic band method. *J. Chem. Phys.* **2012**, *136*, 074103.
(34) Qian, G.-R.; Dong, X.; Zhou, X.-F.; Tian, Y.; Oganov, A. R.; Wang, H.-T. *Comput. Phys. Commun.* **2013**, *184*, 2111−2118.
(35) Chakraborty, T.; Rogal, J.; Drautz, R. *J. Phys.: condens. Matter* **2015**, *27*, 115401.
(36) Zarkevich, N. A.; Johnson, D. D. *Phys. Rev. B* **2018**, *97*, 014202.
(37) Eyméoud, P.; Huang, L.; Maugis, P. *Scr. Mater.* **2021**, *205*, 114182.
(38) Trinkle, D. R.; Hennig, R. G.; Srinivasan, S. G.; Hatch, D. M.; Jones, M. D.; Stokes, H. T.; Albers, R. C.; Wilkins, J. W. *Phys. Rev. Lett.* **2003**, *91*, 025701.
(39) Mrdjenovich, D.; Persson, K. A. *Phys. Rev. Mater.* **2024**, *8*, 033401.
(40) Hart, P. E.; Nilsson, N. J.; Raphael, B. A Formal Basis for the Heuristic Determination of Minimum Cost Paths. *IEEE Trans. Syst. Sci. Cybernetics* **1968**, *4*, 100−107.
(41) Horton, M. K.; et al. *Nat. Mater.* **2025**, *24*, 1522−1532.
(42) Ong, S. P.; Richards, W. D.; Jain, A.; Hautier, G.; Kocher, M.; Cholia, S.; Gunter, D.; Chevrier, V. L.; Persson, K. A.; Ceder, G. *Comput. Mater. Sci.* **2013**, *68*, 314−319.
(43) Widdowson, D. E.; Kurlin, V. A. *SIAM J. Appl. Math.* **2026**, *86*, 898−918.
(44) Luo, Y.; Benali, A.; Shulenburger, L.; Krogel, J. T.; Heinonen, O.; Kent, P. R. C. *New J. Phys.* **2016**, *18*, 113049.
(45) Muscat, J.; Swamy, V.; Harrison, N. M. *Phys. Rev. B* **2002**, *65*, 224112.
(46) Zhu, T.; Gao. S.-P. *J. Phys. Chem. C* **2014**, *118*, 11385−11396.
(47) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865−3868.
(48) Castro, A. L.; Nunes, M. R.; Carvalho, A. P.; Costa, F. M.; Florêncio, M. H. *Solid State Sci.* **2008**, *10*, 602−606.
(49) Chen, G.; Chen, J.; Song, Z.; Srinivasakannan, C.; Peng, J. *J. Alloys Compd.* **2014**, *585*, 75−77.
(50) Lee, B. I.; Wang, X.; Bhave, R.; Hu, M. *Mater. Lett.* **2006**, *60*, 1179−1183.
(51) Spektor, K.; Tran, D. T.; Leinenweber, K.; Häussermann, U. *J. Solid State Chem.* **2013**, *206*, 209−216.
(52) Liu, B.; Sun, L.; Wu, Y.; Zhang, Y.; Li, Z.; Chen, J.; Xie, C.; Huang, Q.; Ma, M.; Luo, K.; Gao, Y.; He, J.; Yu, D.; Xu, B.; Zhao, Z. *J. Eur. Ceram. Soc.* **2021**, *41*, 4951−4957.
(53) Hidalgo-Jiménez, J.; Akbay, T.; Ishihara, T.; Edalati, K. *J. Mater. Chem. A* **2023**, *11*, 23523−23535.
(54) Razavi-Khosroshahi, H.; Edalati, K.; Hirayama, M.; Emami, H.; Arita, M.; Yamauchi, M.; Hagiwara, H.; Ida, S.; Ishihara, T.; Akiba, E.; Horita, Z.; Fuji, M. *ACS Catal.* **2016**, *6*, 5103−5107.
(55) Xie, X.; Gu, X.; Chen, M. *Acta Geochimica* **2023**, *42*, 1−8.
(56) El Goresy, A.; Chen, M.; Gillet, P.; Dubrovinsky, L.; Graup, G.; Ahuja, R. *Earth Planet. Sci. Lett.* **2001**, *192*, 485−495.
(57) Wu, X.; Meng, D.; Han, Y. *Am. Mineral* **2005**, *90*, 1458−1461.
(58) Triebold, S.; Luvizotto, G. L.; Tolosana-Delgado, R.; Zack, T.; von Eynatten, H. *Contrib. Mineral. Petrol.* **2011**, *161*, 581−596.
(59) Li, J.-G.; Ishigaki, T. *Acta Mater.* **2004**, *52*, 5143−5150.
(60) Song, M.; Lu, Z.; Li, D. *Nanoscale* **2020**, *12*, 23183−23190.
(61) Lu, Z.; Sushko, P. V.; Li, D. *Cryst. Growth Des.* **2025**, *25*, 10637−10650.
(62) Wang, W.-X.; Jiang, Z.-Y.; Bo-Zhou Zhang, B.-Z.; Zhang, Z.-Y. Theoretical study on martensitic-type transformation of rutile, columbite, and baddeleyite phase of TiO 2. *Europhys. Lett.* **2021**, *134*, 46001.
(63) Sato, H.; Endo, S.; Sugiyama, M.; Kikegawa, T.; Shimomura, O.; Kusaba, K. *Science* **1991**, *251*, 786−788.
(64) Murata, H.; Taniguchi, T.; Tanaka, I. First-principles calculations of high-pressure phase transition of TiO2 during decompression: From baddeleyite-type TiO2 to α-PbO2-type TiO2. *J. Appl. Phys.* **2016**, *120*, 142108.
(65) Wang, W.-X.; Jiang, Z.-Y.; Lin, Y.-M.; Zheng, J.-M.; Zhang, Z.-Y. *Chin. Phys. B* **2020**, *29*, 076101.
(66) Liang, S.; Wang, X.; Qi, R.; Cheng, Y.-J.; Xia, Y.; Müller-Buschbaum, P.; Hu, X. *Adv. Funct. Mater.* **2022**, *32*, 2201675.
(67) Chen, Z.-Y.; Chen, G.-W.; Zhu, S.-C. *J. Phys. Chem. C* **2022**, *126*, 11846−11854.
(68) Batatia, I.; Benner, P.; Chiang, Y.; Elena, A. M.; Kovács, D. P.; Riebesell, J.; Advincula, X. R.; Asta, M.; Avaylon, M.; Baldwin, W. J.; Berger, F.; et al. A foundation model for atomistic materials chemistry. *J. Chem. Phys.* **2025**, *163*, 184110.
(69) Lysogorskiy, Y.; Bochkarev, A.; Drautz, R. Graph atomic cluster expansion for foundational machine learning interatomic potentials. *Npj Comput. Mater.* **2026**, *12*, 114.
(70) Ganose, A. M.; et al. *Digital Discovery* **2025**, *4*, 1944−1973.
(71) Mrdjenovich, D. J. *An Algorithm for Exploring Crystallographic Configuration Space* Ph.D. Thesis; UC Berkeley, 2022.
(72) Rosen, A. S.; Gallant, M.; George, J.; Riebesell, J.; Sahasrabuddhe, H.; Shen, J.-X.; Wen, M.; Evans, M. L.; Petretto, G.; Waroquiers, D.; Rignanese, G.-M.; Persson, K. A.; Jain, A.; Ganose, A. M. *J. Open Source Softw.* **2024**, *9*, 5995.

https://doi.org/10.1021/acs.jctc.6c01114
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX