








RESEARCH ARTICLE | SEPTEMBER 01 2023

Universal and interpretable classification of atomistic structural transitions via unsupervised graph learning FREE

Special Collection: [Accelerate Materials Discovery and Phenomena](#)

Bamidele Aroboto ; Shaohua Chen ; Tim Hsu ; Brandon C. Wood ; Yang Jiao ; James Chapman  

 Check for updates

Appl. Phys. Lett. 123, 094103 (2023)
<https://doi.org/10.1063/5.0156682>


View
Online


Export
Citation

CrossMark

Articles You May Be Interested In

Protein folding intermediates on the dimensionality reduced landscape with UMAP and native contact likelihood

J. Chem. Phys. (August 2022)

Automatic graph representation algorithm for heterogeneous catalysis

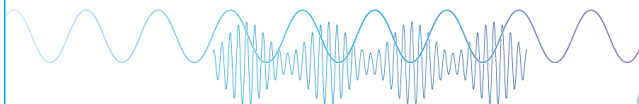
APL Machine Learning (July 2023)

Analysis of braid manifolds using uniform manifold approximation and projection

J Acoust Soc Am (March 2023)

Webinar

Boost Your Signal-to-Noise Ratio with Lock-in Detection



Sep. 7th – Register now



Zurich
Instruments

Universal and interpretable classification of atomistic structural transitions via unsupervised graph learning

Cite as: Appl. Phys. Lett. **123**, 094103 (2023); doi: [10.1063/5.0156682](https://doi.org/10.1063/5.0156682)

Submitted: 2 May 2023 · Accepted: 21 August 2023 ·

Published Online: 1 September 2023



View Online



Export Citation



CrossMark

Bamidele Aroboto,¹ Shaohua Chen,^{2,3} Tim Hsu,⁴ Brandon C. Wood,⁵ Yang Jiao,² and James Chapman^{1,a)}

AFFILIATIONS

¹Department of Mechanical Engineering, Boston University, Boston, Massachusetts 02215, USA

²Materials Science and Engineering, Arizona State University, Tempe, Arizona 85281, USA

³College of Energy and Power Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China

⁴Center for Applied Scientific Computing, Lawrence Livermore National Laboratory, Livermore, California 94550, USA

⁵Materials Science Division, Lawrence Livermore National Laboratory, Livermore, California 94550, USA

Note: This paper is part of the APL Special Collection on Accelerate Materials Discovery and Phenomena.

a) Author to whom correspondence should be addressed: jc112358@bu.edu

ABSTRACT

Materials processing often occurs under extreme dynamic conditions leading to a multitude of unique structural environments. These structural environments generally occur at high temperatures and/or high pressures, often under non-equilibrium conditions, which results in drastic changes in the material's structure over time. Computational techniques, such as molecular dynamics simulations, can probe the atomic regime under these extreme conditions. However, characterizing the resulting diverse atomistic structures as a material undergoes extreme changes in its structure has proved challenging due to the inherently non-linear relationship between structures as large-scale changes occur. Here, we introduce SODAS++, a universal graph neural network framework, that can accurately and intuitively quantify the atomistic structural evolution corresponding to the transition between any two arbitrary phases. We showcase SODAS++ for both solid–solid and solid–liquid transitions for systems of increasing geometric and chemical complexity, such as colloidal systems, elemental Al, rutile and amorphous TiO₂, and the non-stoichiometric ternary alloy Ag₂₆Au₅Cu₁₉. We show that SODAS++ can accurately quantify all transitions in a physically interpretable manner, showcasing the power of unsupervised graph neural network encodings for capturing the complex and non-linear pathway, a material's structure takes as it evolves.

Published under an exclusive license by AIP Publishing. <https://doi.org/10.1063/5.0156682>

Materials under extreme thermal, mechanical, and chemical conditions often experience drastic changes in their structure.^{1–5} These structural changes begin at the atomic scale and propagate upward in length scale to ultimately affect larger-scale features, such as the material's microstructure.^{6–8} Due to conditions, such as high temperature and pressure, these changes in the atomic structure are often non-trivial and highly correlated, with complex atomic diffusion events acting as a driving force for changes in local atomic geometries.^{9,10} As these structural changes often lead to significant differences in material properties, a thorough understanding of how atomic-level configurations evolve over time is necessary. Therefore, if one aims to quantify the changes in materials properties under such extreme conditions, algorithmic tools that can accurately map out both subtle and significant differences in atomic geometries is vital.

Over the last few decades, numerous developments on this front have occurred, leading to a plethora of methods with excellent atomic environment characterization capabilities.^{11–26} However, having the ability to provide a unique geometric descriptor for a particular atomic structure does not inherently mean that an interpretable pathway between two random structures exists.²⁷ Such a pathway is critical to understanding the spatiotemporal behavior of materials under extreme conditions because materials often undergo a cascade of subtle changes that culminate in significant differences in their atomic structures over time.^{27–29} Quantifying this process from start-to-finish, on both collective and local atomic scales, is critical to understanding the question of why materials behave the way they do when placed in extreme external conditions, laying the foundation for predictive modeling of such systems and their optimization.

Recent work has shown that answering this question is non-trivial for the majority of atomistic descriptors due to numerous challenges, such as the uniqueness of descriptors,³⁰ their computational cost,³¹ and their generality to any arbitrary material system.³² In particular, many methods employ a “bottom-up” approach when characterizing structures as a whole, aiming to accurately quantify local atomic sites and then project the collection of atomic features into a global descriptor.^{32–34} This bottom-up approach often results in a loss of information due to the inherently disconnected nature of treating atomic environments as independent entities.³⁰

Here, we build upon our previous work, the Structural Orderness Degree for Atomic Systems (SODAS),³⁵ and present SODAS++, which overcomes the limitations of the original SODAS algorithm by addressing two major components: (1) extension to multicomponent systems and (2) universal applicability to any atomic structural transformation, not just those that depend on temperature. We achieve (1) by interfacing SODAS with ALIGNN,^{23,36} a line-graph approach that incorporates bonds and bond angles into the graph representation. (2) is accomplished by mapping the ALIGNN^{23,36} latent space to an intuitive lower-dimensional manifold using a dimensionality reduction technique called UMAP.³⁷ Importantly, SODAS++ does not require a training stage, instead relying purely on the unsupervised GNN encoding. This allows for SODAS++ to be generalized to any arbitrary material system.

The SODAS formalism relies on a scalar order parameter, γ , to determine how far along a given structural transition a particular atomic system is. In the original SODAS implementation,³⁵ γ was explicitly linked to temperature, limiting its applicability to temperature-dependent structural transitions. SODAS++ alleviates this dependency, instead relying on the GNN latent space encoding to produce an accurate phase space. By combining (1) and (2),

SODAS++ can quantify any arbitrary structural transformation, under dynamic external conditions, in a universal and efficient manner. Importantly, SODAS++ naturally encodes collective correlations among local environments as it takes a “top-down” approach by building global graph of the physical system. We showcase the accuracy and generality of SODAS++ by characterizing a multitude of structural transformations, such as the density-dependent phase changes in colloidal particles, the compression and melting of FCC aluminum, the amorphization of rutile TiO₂, and finally the melting of the non-stoichiometric ternary alloy Ag₂₆Au₅Cu₁₉.

A detailed description of the SODAS++ workflow is described in the following paragraphs, and a visual depiction can be seen in Fig. 1. We begin by leveraging the ALIGNN formulation,^{23,36} where two graphs are used to encode one atomic structure: an original atomic graph G and its corresponding line graph $L(G)$. The nodes and edges in G represent atoms and bonds, respectively. The nodes and edges in $L(G)$, on the other hand, represent bonds and bond angles, respectively. Note that the edges in G and the nodes in $L(G)$ are identical entities and share the same embedding during GNN operation. In this work, we extended the ALIGNN encoding to also explicitly represent dihedral angles. We used Atomic Simulation Environment³⁸ and PyTorch Geometric³⁹ to construct the graph representations and to calculate all bond angles.

In the initial encoding, the atom type, the bond distance, and the angular values are converted from scalars to feature vectors for subsequent neural network operations. The atom type z is transformed by a learnable embedding matrix.⁴⁰ The bond distance d is expanded into the Radial Bessel basis proposed by Klicpera *et al.*⁴¹ The angles α are expanded into the “standard” Gaussian basis implemented by Schnet.⁴²

The interaction operations are also known as graph convolution, aggregation, or message-passing. Following the ALIGNN paper,³⁶ we

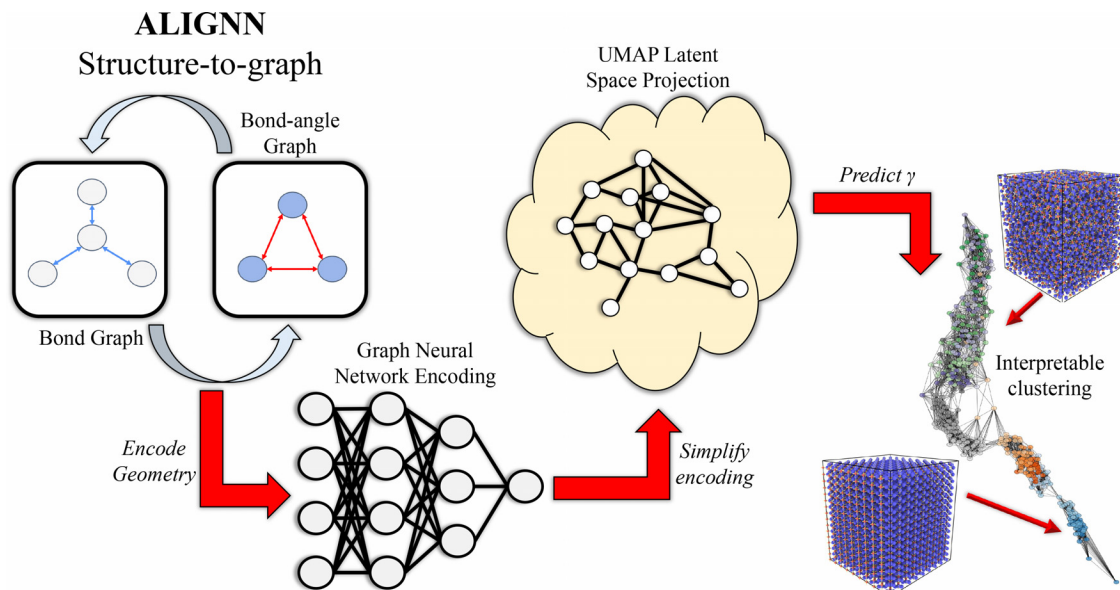


FIG. 1. Visual workflow of the SODAS++ methodology. Atomic structures are represented as graphs and then encoding with the GNN ALIGNN framework. This encoding is then projected onto a low-dimensional manifold using UMAP. Planar graphs are then constructed from the low-dimensional manifold, and a graph pathfinding algorithm is employed to learn the phase transition.

also adopted the edge-gated graph convolution^{43,44} for the interaction operations. The node features \vec{h}_i^{l+1} of node i at the $(l + 1)$ th layer are updated as

$$\vec{h}_i^{l+1} = \vec{h}_i^l + \text{SiLU} \left(\text{LayerNorm} \left(\vec{W}_s^l \vec{h}_i^l + \sum_{j \in \mathcal{N}(i)} \hat{e}_{ij}^l \odot \vec{W}_d^l \vec{h}_j^l \right) \right), \quad (1)$$

where SiLU is the sigmoid linear unit activation function;⁴⁵ LayerNorm is the layer normalization operation;⁴⁶ \vec{W}_s and \vec{W}_d are weight matrices; the index j denotes the neighbor node of node i ; \hat{e}_{ij} is the edge gate vector for the edge from node i to node j ; and \odot denotes element-wise multiplication. The edge gate \hat{e}_{ij}^l at the l th layer is defined as

$$\hat{e}_{ij}^l = \frac{\sigma(\vec{e}_{ij}^l)}{\sum_{j \in \mathcal{N}(i)} \sigma(\vec{e}_{ij}^l) + \epsilon}, \quad (2)$$

where σ is the sigmoid function, \vec{e}_{ij}^l is the original edge feature, and ϵ is a small constant for numerical stability. The edge features \vec{e}_{ij}^l are updated by

$$\vec{e}_{ij}^{l+1} = \vec{e}_{ij}^l + \text{SiLU} \left(\text{LayerNorm} \left(\vec{W}_g^l \vec{z}_{ij}^l \right) \right), \quad (3)$$

where \vec{W}_g is a weight matrix and \vec{z}_{ij} is the concatenated vector from the node features \vec{h}_i , \vec{h}_j , and the edge features \vec{e}_{ij} ,

$$\vec{z}_{ij} = \vec{h}_i \oplus \vec{h}_j \oplus \vec{e}_{ij}. \quad (4)$$

We applied the same edge-gated convolution scheme [Eqs. (1)–(3)] to operate on both the atomic graph G and the line graphs $L(G)$. In the case of G , the edge-gated convolution updates nodes, which represent atoms, and edges, which represent bonds, while exchanging information between the two. In the case of $L(G)$, the convolution updates nodes, which represent bonds, and edges, which represent angles. Note that by iteratively applying the convolution operation on both the original graph and the line graph, the angular information stored in $L(G)$ can propagate to G . Due to the nature of the edge-gated convolution, all the feature/embedding vectors for atoms, bonds, and angles during the interaction layers have the same length, or the same number of channels D . Finally, the final output layers pool (by summation) the node features of G and transform the pooled embedding into an output vector, which preserves the dimensionality of the original input layer. All model parameters can be found in Table I. We emphasize here that the GNN encoding requires no training process and is purely unsupervised. The GNN is simply learning the information present in the ALIGNN representation and projecting it onto a higher dimensional space.

The resulting GNN latent space is used to determine a pathway between the initial and final configurations of a given trajectory. This is accomplished in a four step process: (1) normalization of the original latent space, (2) projection of the normalized latent space onto a manifold using UMAP,³⁷ (3) conversion of the manifold space into a planar graph, and (4) recursive graph search to determine an optimal pathway. The following paragraphs provide more detail into each step.

We apply a normalization algorithm \mathcal{N} to the initial GNN latent space to ensure that all dimensions are weighted equally. Here,

TABLE I. Model parameters.

Name	Notation	Value
Number of interaction layers	L	6
Radial Bessel basis cutoff (bond distance)	c_d	5 Å
Gaussian basis range (cosine and sine angles)	c_x	(−1, 1)
Number of channels	D	100

$\mathcal{N} = \sigma_X(\max_X - \min_X) + \min_X$, where σ_X is the standard deviation for a given dimension X , while \min_X and \max_X are the minimum and maximum values along dimension X . We then project the normalized latent space onto a lower-dimensional manifold (LD) using UMAP.³⁷ UMAP is chosen due to its algorithm being rooted in topology, making the resulting space ideal for the construction of a planar graph. The planar graph is generated from the UMAP manifold space via a k -nearest neighbor (kNN) approach. For all systems studied in this work, k was set to 15. The nearest neighbors were identified using a Minkowski distance metric⁴⁷ in the UMAP space.

From the kNN graph, a recursive graph search algorithm is employed to discover a pathway between the initial and final configurations in the original trajectory. Upon reaching any given node, we rank its neighboring nodes and chose the closest neighbor that has a possible pathway to the final node. No node can be visited twice during the pathway generation. If at any point a pathway cannot be created due to lack of suitable neighbors, the algorithm goes back one node in the path and moves to the next suitable neighbor. This iterative process results in a realistic pathway generation by mimicking the true oscillations in phase space during a molecular dynamics trajectory and is a graph-analog for damped potential energy surface exploration.⁴⁸

Finally, we define a phase space order parameter, γ , as the point between two endpoints within the phase space. The purpose of γ is to project the inherently non-linear transition pathway onto an order parameter that varies linearly within the space. Conceptually, γ is the location along a given transition, a proxy for a phase-space reaction coordinate. For example, when considering a solid-to-liquid transformation, one can think of γ as a proxy for the configurational entropy at any arbitrary point along the trajectory. Using our graph-pathing scheme, we can break the path into iterative segments, where γ can be calculated as the distance along the total path. Therefore, we define γ as

$$\gamma_k = \frac{\sum_i^k d_i}{\sum_j^N d_j}, \quad (5)$$

where k defines a specific configuration of atoms, i indexes the iterative path segments before node k , while j indexes the segments along the total pathway, and N defines the total number of segments in the path. Therefore, for any arbitrary phase transition, γ defines how “far along” the transition a given configuration of atoms exists at. When a new point p is placed in the LD, its γ value is determined by finding the closest node belonging to the generated path and assigning that node’s γ to p .

To validate SODAS++, we performed MD simulations on several systems ranging from fictitious toy problems to the melting of oxides and ternary alloys. All MD simulations were performed using LAMMPS.⁴⁹ Interatomic potentials for Al,⁵⁰ TiO₂ (Ref. 51), and Ag₂₆Au₅Cu₁₉ (Ref. 52) were used. NVT simulations were performed for both melting and compression trajectories. For melting, temperatures were initialized to 100 K and culminated at 3000 K, though only snapshots within 2 ps of melting were taken from the high temperature regime. Compression simulations were performed at 300 K.

The utility of SODAS++ is also demonstrated in quantifying entropy-driven disorder-order phase transition in 3D hard-sphere systems,^{53,54} which are excellent models for many colloidal systems dominated by strong repulsive interactions. The transition is induced by increasing density or packing fraction ϕ (via compression), and the equilibrium phase diagram possesses three branches: disordered liquid ($0 < \phi < 0.494$), co-existence ($0.494 < \phi < 0.545$), and crystalline solid ($0.545 < \phi < 0.74$). Here, we use event-driven molecular dynamics⁵⁵ to capture the disorder-order transition, starting from a metastable liquid state with $\phi = 0.54$. Once the density increases beyond the freezing point ($\phi_F = 0.545$), the liquid becomes unstable and undergoes a rapid collective structural re-arrangement to the crystalline state (with defects), which is then continuously compressed to the maximal density.

It can be seen in Figs. 2(a) and 3 that the disorder-to-order transition of the colloidal system is clearly captured by SODAS++. The metastable liquid states are clustered and separated from the crystalline states (connected via a continuous path); and a sharp, discontinuous-like transition between the two phases can be seen in both Figs. 2(a) and 3(a). Particularly, from Fig. 2(a), one can see two distinct subgraphs connected via a single edge. The two subgraphs represent the

disordered and ordered states with the single graph edge representing the transition from order-to-disorder. One can also see from Fig. 2(a) that the subgraph representing the ordered states can be broken into two further subgraphs by examining the average edge connections to each node. This further breakdown represents the transition of the ordered region from one phase to a second ordered phase. This transition can also be seen in Fig. 3(a) with a second discrete jump around (0, 10) in (L1, L2) coordinates. The coloring of nodes in Figs. 3(b) and 2(a), which again is determined via Fig. 5, follows the trend observed in Fig. 3(a), indicating that the unsupervised determination of γ is accurate.

Figure 4 depicts the SODAS++ characterization of (a, b) the compression of Al at 100 K and (c, d) the melting of Al via superheating. From Fig. 4(a), the continuous compression of Al is captured well via the latent space projection, with a continuous pathway from low-to-high density observed in Fig. 4(b). One important note regarding Fig. 4(b) is that the blue and orange points seem to overlap in the 2D view; however, in 3D, they do not overlap. From Fig. 4(c), one can see that SODAS++ also captures the continuous heating of Al from perfect crystal to liquid phase extremely well. Similar to the compression case, an intuitive pathway is found between the initial and final configurations in Fig. 4(d). It is important to note that the compression and melting scenarios present wildly different transition types, with one depending on a continuous disordering of the structure and another the slow change in interatomic distances but without breaking the symmetry of the crystal. SODAS++ quantitatively captures both of these structural transitions in a way that is interpretable.

Figure 5 focuses on order-to-disorder transitions but for more complex structural systems, with (a)–(d) capturing the amorphization of TiO₂ and (e)–(i) predicting the melting of a randomly distributed

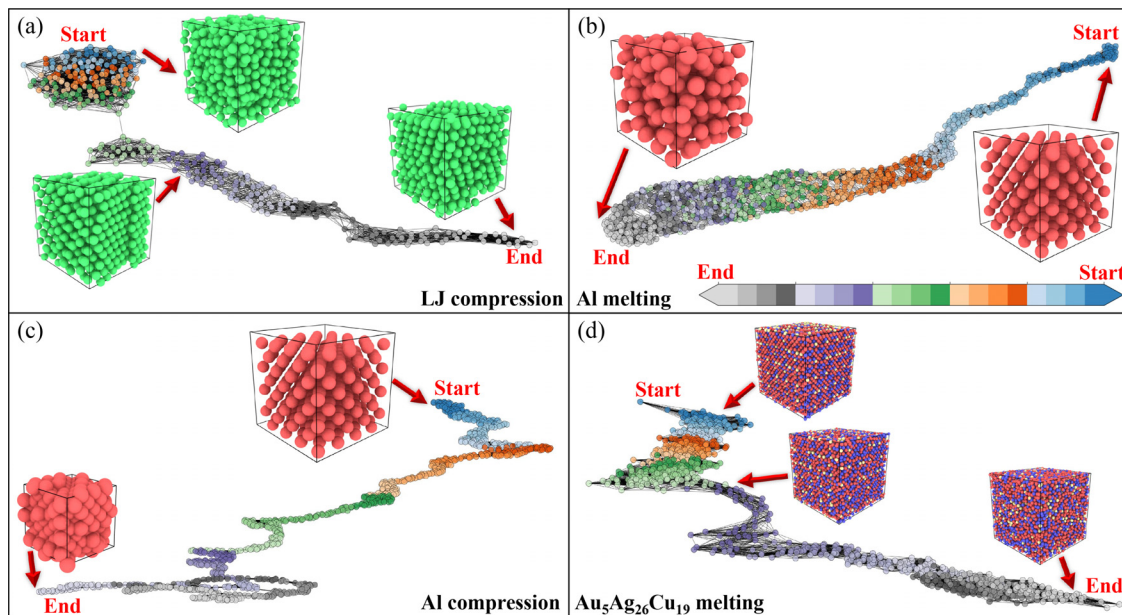


FIG. 2. Graph representation of the SODAS++ learned phase transition. Colors represent the γ calculated from the graph pathfinding algorithm, with a colorbar found in (b). (a) The density-based transition of the colloidal system. (b) The melting of Al. (c) The compression of bulk Al. (d) The melting of Ag₂₆Au₅Cu₁₉. Inserted images show representative snapshots of the atomic structures during specific points along the trajectory.

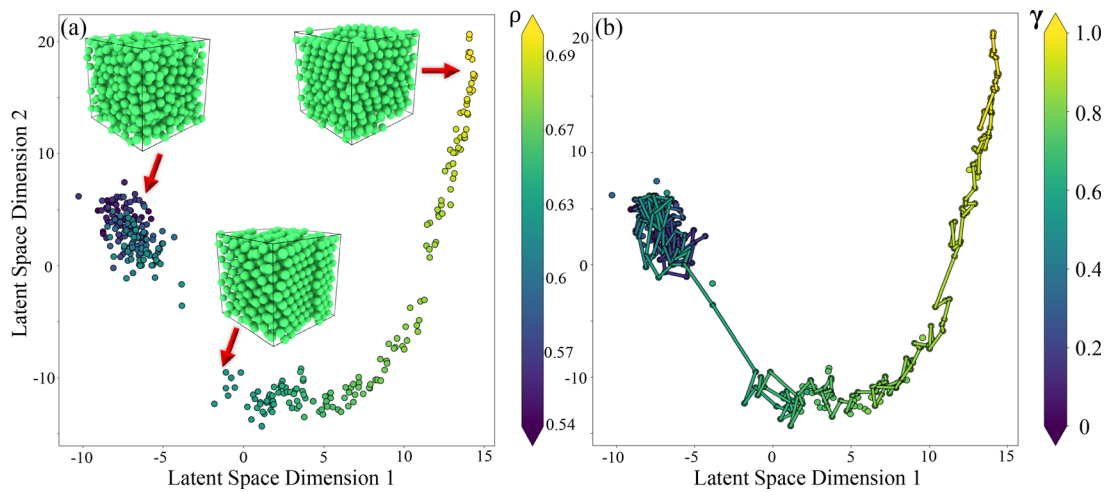


FIG. 3. SODAS++ UMAP latent space projection for a colloidal system's density-based structural transition. (a) The latent space projection color-coded based on the density of the structure. Inserted images show representative points along the trajectory. (b) The latent space color-coded based on the predicted γ values. A path through the graph shows the result of the graph pathfinding algorithm, color coded based on the edge's γ .

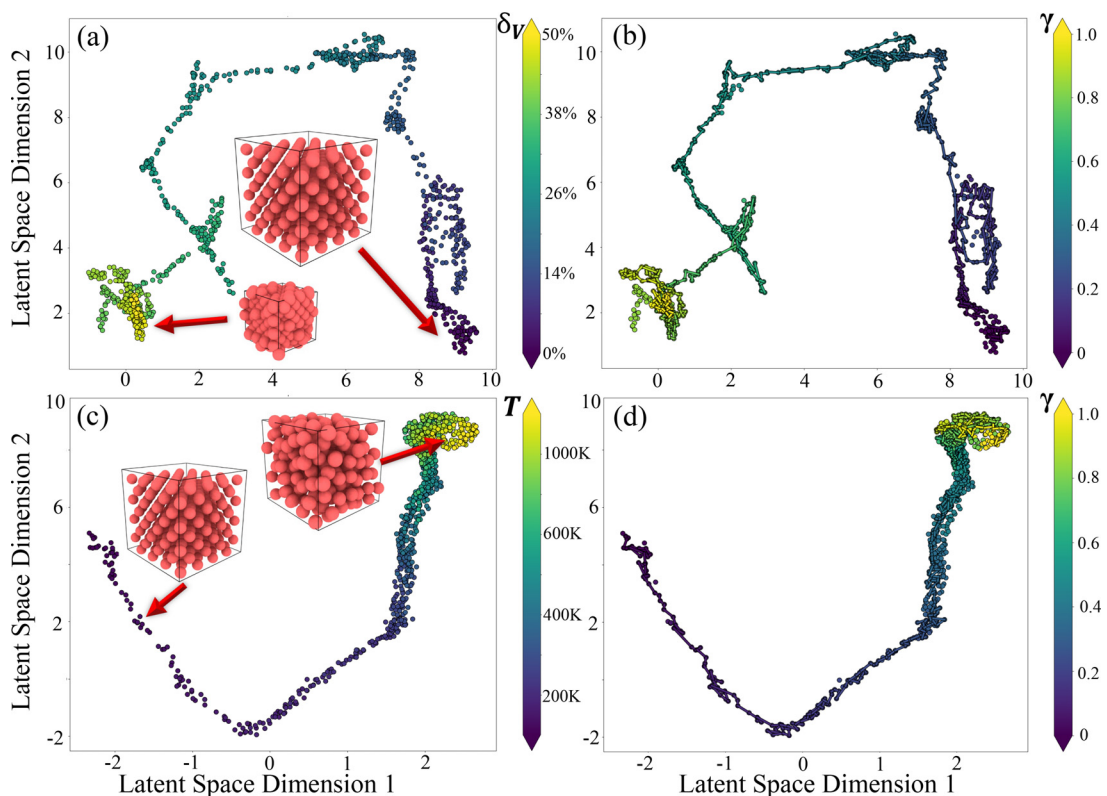


FIG. 4. SODAS++ UMAP latent space projection for [(a) and (b)] compression and [(c) and (d)] melting of Al. (a) The latent space projection color-coded based on the change in volume relative to the perfect ground-state crystal volume. (c) The latent space projection color-coded based on the temperature of the system. Inset images show representative points along the trajectory. (b) and (d) The latent space color-coded based on the predicted γ values. A path through the graph shows the result of the graph pathfinding algorithm, color coded based on the edge's γ .

non-stoichiometric ternary alloy. For the amorphization of TiO_2 in Figs. 5(a)–5(d), we again see that SODAS++ provides an intuitive characterization a function of temperature. The amorphization of TiO_2 proceeds in two stages: (1) disordering the crystal into an amorphous state and (2) transition of the amorphous system to a liquid. We can see from Fig. 5(d) that γ predicts a continuous transition pathway from the crystal to the liquid. From the radial distribution functions (RDF) shown in Figs. 5(a)–5(c), one can see the two-stage transition process as a function of γ . At small γ values, crystalline rutile is still observed. At approximately the halfway point ($\gamma = 0.5$), we can see the first two peaks of Ti–Ti combine into a single peak, along with a broadening of the peaks after 4 Å. This long-range broadening is also present in the O–O distribution after 3 Å. These RDF changes correspond to the transition from crystalline rutile to amorphous TiO_2 . Finally, we observe further broadening of the peaks, especially the first peak, which corresponds to the transition from an amorphous state to a liquid. Again, we emphasize the ability of SODAS++ to not only uniquely characterize these three atomic systems but also connect them in a way that is physically justified and intuitive.

Figures 5(e)–5(i) show the melting of $\text{Ag}_{26}\text{Au}_5\text{Cu}_{19}$ via continuous heating simulations. Again, SODAS++ does an excellent job of capturing the continuous order-to-disorder transition as the thermostat temperature is increasingly ramped up. Here, an interesting

solid–solid transition occurs prior to the melting of the system. At around 1000 K, the phase space breaks into two distinct regions, as evidenced by Fig. 5(i). Interestingly, this transition can be explained by observing the RDF as a function of temperature, as shown in Figs. 5(e)–5(h). At lower temperatures, Cu–Au distances (h) are the second most common peak at interatomic distances smaller than 2.5 Å. However, at approximately 1000 K Ag–Cu distances become equivalent in likelihood at those distances, with Au–Au distances becoming more likely as temperature increases as well. This indicates that the atomic structure changes from a Cu–Au dominant geometry to one in which Ag plays a larger role as the system becomes more disordered. Figure 5(i) captures this transition well by predicting a “bridge” region between the two distinct clusters, one representing low-temperature motifs and the other higher-temperature ones. The γ characterization in Fig. 5(i) quantifies this transition extremely well, further signifying that the SODAS++ latent space can quantify the structural changes in a material in a unique and interpretable manner.

In this work, we showcased an improvement over the SODAS methodology, SODAS++, that alleviates the pitfalls of SODAS by generalizing the framework to any class of structural transition and any class of material. Using SODAS++, we quantified structural transitions of colloidal systems of increasing density, the low-temperature compression and melting of bulk Al, the amorphization of TiO_2 , and

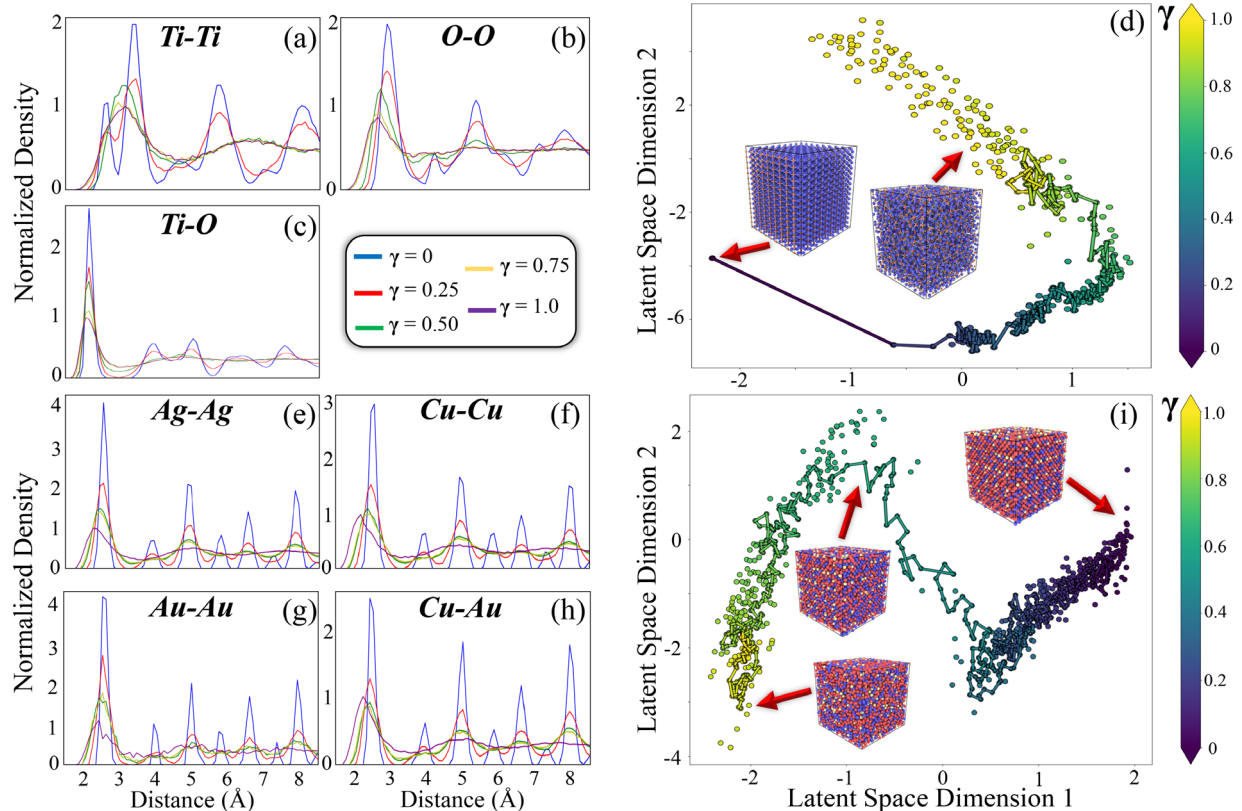


FIG. 5. SODAS++ UMAP latent space projection for [(a) and (d)] amorphization of TiO_2 and [(e)–(i)] melting of $\text{Ag}_{26}\text{Au}_5\text{Cu}_{19}$. Legend to the right of (c) pertains to all subplots [(a)–(c), (e)–(h)]. (a)–(c) The radial distribution function for TiO_2 , decomposed into the three pairwise chemical interactions. (d) The latent space projection color-coded based on the predicted γ value. (e)–(h) The radial distribution function for $\text{Ag}_{26}\text{Au}_5\text{Cu}_{19}$, decomposed into four chosen pairwise chemical interactions. (i) The latent space color-coded based on the predicted γ values. A path through the graphs in (d) and (i) shows the result of the graph pathfinding algorithm, color coded based on the edge’s γ .

the melting of $\text{Ag}_{26}\text{Au}_5\text{Cu}_{19}$. These systems represent a wide spectrum of materials classes and changes in atomic structure. In all cases, SODAS++ quantified these structural transitions in a manner that was not only accurate but was, more importantly, interpretable. This interpretability is a critical piece of machine learning methodologies for physical systems, as it is not sufficient to simply apply black-box models to systems without understanding the underlying mechanisms at play.

SODAS++ also does not require training a model on reference data, as the characterizations shown in this work are created via the GNN encoding, which simply operates on the information contained within the initial graph. In this way, SODAS++ is an unsupervised technique that can encode any structural transition in a physically interpretable manner. This process makes SODAS++ generalizable to any material system that can be appropriately encoded as a graph. Based on this, we speculate that SODAS++ could be used as a foundation for generative modeling of atomistic trajectories, in which a path can be generated between known endpoints, alleviating the need for complex and expensive simulations. We do note the inherently non-linear pathways found between endpoints in this work, however, implying that further work is needed to interpolate between points within the latent space using sparse datasets. We hope that methods, such as SODAS++, can help to alleviate the current bottlenecks with characterizing atomistic structures and provide a pathway to accurate, unique, and interpretable atomistic descriptors.

J.C. thanks the College of Engineering and the Department of Mechanical Engineering at Boston University. J.C. also thanks the BU's Shared Computing Cluster and the Massachusetts Green High Performance Computing Center. Y.J. was partially supported by the National Science Foundation (NSF) under Grant No. 2020277. This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract No. DE-AC52-07NA27344.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Bamidele Aroboto: Formal analysis (equal); Methodology (equal); Validation (equal); Writing – original draft (supporting). **Shaohua Chen:** Data curation (equal); Formal analysis (equal). **Tim Hsu:** Conceptualization (equal); Writing – review & editing (equal). **Brandon wood:** Conceptualization (equal); Writing – review & editing (equal). **Yang Jiao:** Conceptualization (equal); Project administration (equal); Supervision (equal); Writing – original draft (equal). **James Chapman:** Conceptualization (equal); Project administration (equal); Software (lead); Supervision (equal); Writing – original draft (equal).

DATA AVAILABILITY

All data required to reproduce this work will be given upon request by contacting the corresponding author. The source code for this work, along with working examples, can be found at <https://github.com/Materials-Informatics-Laboratory/SODAS>.

REFERENCES

- B. A. Remington, G. Bazan, J. Belak, E. Bringa, J. D. Colvin, M. J. Edwards, S. G. Glendinning, D. H. Kalantar, M. Kumar, B. F. Lasinski, K. T. Lorenz, J. M. McNaney, S. M. Pollaine, D. Rowley, J. S. Stölkén, S. V. Weber, W. G. Wolfer, M. Caturla, D. S. Ivanov, L. V. Zhigilei, B. Kad, M. A. Meyers, M. Schneider, D. D. Meyerhofer, B. Yaakobi, and J. S. Wark, "Materials science under extreme conditions of pressure and strain rate," *Metall. Mater. Trans. A* **35**, 2587–2607 (2004).
- S. Eswarappa Prameela, T. M. Pollock, D. Raabe, M. A. Meyers, A. Aitkaliyeva, K.-L. Chintersingh, Z. C. Cordero, and L. Graham-Brady, "Materials for extreme environments," *Nat. Rev. Mater.* **8**, 81–88 (2022).
- C.-S. Yoo, "Chemistry under extreme conditions: Pressure evolution of chemical bonding and structure in dense solids," *Matter Radiat. Extremes* **5**, 018202 (2020).
- M. L. Martin, M. Dadfarnia, A. Nagao, S. Wang, and P. Sofronis, "Enumeration of the hydrogen-enhanced localized plasticity mechanism for hydrogen embrittlement in structural materials," *Acta Mater.* **165**, 734–750 (2019).
- W. G. Fahrenholtz and G. E. Hilmas, "Ultra-high temperature ceramics: Materials for extreme environments," *Scr. Mater.* **129**, 94–99 (2017).
- O. Hod, E. Meyer, Q. Zheng, and M. Urbakh, "Structural superlubricity and ultralow friction across the length scales," *Nature* **563**, 485–492 (2018).
- J. Shamblin, M. Feyngenson, J. Neufeind, C. L. Tracy, F. Zhang, S. Finkeldei, D. Bosbach, H. Zhou, R. C. Ewing, and M. Lang, "Probing disorder in isometric pyrochlore and related complex oxides," *Nat. Mater.* **15**, 507–511 (2016).
- Y. Li, Y. Li, A. Pei, K. Yan, Y. Sun, C.-L. Wu, L.-M. Joubert, R. Chin, A. L. Koh, Y. Yu, J. Perrino, B. Butz, S. Chu, and Y. Cui, "Atomic structure of sensitive battery materials and interfaces revealed by cryo-electron microscopy," *Science* **358**, 506–510 (2017).
- M. Vaidya, S. Trubel, B. Murty, G. Wilde, and S. Divinski, "Ni tracer diffusion in CoCrFeNi and CoCrFeMnNi high entropy alloys," *J. Alloys Compd.* **688**, 994–1001 (2016).
- E. Castle, T. Csanádi, S. Grasso, J. Dusza, and M. Reece, "Processing and properties of high-entropy ultra-high temperature carbides," *Sci. Rep.* **8**, 8609 (2018).
- A. Stukowski, "Structure identification methods for atomistic simulations of crystalline materials," *Model. Simul. Mater. Sci. Eng.* **20**, 045021 (2012).
- P. J. Steinhardt and P. Chaudhari, "Point and line defects in glasses," *Philos. Mag. A* **44**, 1375–1381 (1981).
- G. J. Ackland and A. P. Jones, "Applications of local crystal structure measures in experiment and simulation," *Phys. Rev. B* **73**, 054104 (2006).
- C. L. Kelchner, S. J. Plimpton, and J. C. Hamilton, "Dislocation nucleation and defect structure during surface indentation," *Phys. Rev. B* **58**, 11085–11088 (1998).
- J. Chapman, N. Goldman, and B. C. Wood, "Efficient and universal characterization of atomic structures through a topological graph order parameter," *npj Comput. Mater.* **8**, 37 (2022).
- S. De, A. P. Bartók, G. Csányi, and M. Ceriotti, "Comparing molecules and solids across structural and alchemical space," *Phys. Chem. Chem. Phys.* **18**, 13754–13769 (2016).
- J. Behler, "Atom-centered symmetry functions for constructing high-dimensional neural network potentials," *J. Chem. Phys.* **134**, 074106 (2011).
- A. V. Shapeev, "Moment tensor potentials: A class of systematically improvable interatomic potentials," *Multiscale Model. Simul.* **14**, 1153–1173 (2016).
- P. M. Larsen, S. Schmidt, and J. Schiøtz, "Robust structural identification via polyhedral template matching," *Modell. Simul. Mater. Sci. Eng.* **24**, 055007 (2016).
- J. Chapman, R. Batra, and R. Ramprasad, "Machine learning models for the prediction of energy, forces, and stresses for platinum," *Comput. Mater. Sci.* **174**, 109483 (2020).
- C. W. Park, M. Kornbluth, J. Vandermause, C. Wolverton, B. Kozinsky, and J. P. Mailoa, "Accurate and scalable graph neural network force field and molecular dynamics with direct force architecture," *npj Comput. Mater.* **7**, 73 (2021).
- Z. Yang and M. J. Buehler, "Linking atomic structural defects to mesoscale properties in crystalline solids using graph neural networks," *npj Comput. Mater.* **8**(1), 198 (2022).

- ²³T. Hsu, T. A. Pham, N. Keilbart, S. Weitzner, J. Chapman, P. Xiao, S. R. Qiu, X. Chen, and B. C. Wood, "Efficient and interpretable graph network representation for angle-dependent properties applied to optical spectroscopy," *npj Comput. Mater.* **8**, 151 (2022).
- ²⁴V. Bapst, T. Keck, A. Grabska-Barwińska, C. Donner, E. D. Cubuk, S. S. Schoenholz, A. Obika, A. W. R. Nelson, T. Back, D. Hassabis, and P. Kohli, "Unveiling the predictive power of static structure in glassy systems," *Nat. Phys.* **16**, 448 (2020).
- ²⁵G. M. Coli and M. Dijkstra, "An artificial neural network reveals the nucleation mechanism of a binary colloidal AB₁₃ crystal," *ACS Nano* **15**, 4335–4346 (2021).
- ²⁶E. Boattini, M. Dijkstra, and L. Filion, "Unsupervised learning for local structure detection in colloidal systems," *J. Chem. Phys.* **151**, 154901 (2019).
- ²⁷C. Li, J. Liu, J. Chen, Y. Yuan, J. Yu, Q. Gou, Y. Guo, and X. Pu, "An interpretable convolutional neural network framework for analyzing molecular dynamics trajectories: A case study on functional states for g-protein-coupled receptors," *J. Chem. Inf. Model.* **62**, 1399–1410 (2022).
- ²⁸A. R. Oganov, C. J. Pickard, Q. Zhu, and R. J. Needs, "Structure prediction drives materials discovery," *Nat. Rev. Mater.* **4**, 331–348 (2019).
- ²⁹Y. Shin, J. Berry, N. Pannucci, M. P. Haataja, J. E. Toettcher, and C. P. Brangwynne, "Spatiotemporal control of intracellular phase transitions using light-activated optodroplets," *Cell* **168**, 159–171.e14 (2017).
- ³⁰S. N. Pozdnyakov, L. Zhang, C. Ortner, G. Csányi, and M. Ceriotti, "Local invertibility and sensitivity of atomic structure-feature mappings," *Open Res. Europe* **1**, 126 (2021).
- ³¹Y. Zuo, C. Chen, X. Li, Z. Deng, Y. Chen, J. Behler, G. Csányi, A. V. Shapeev, A. P. Thompson, M. A. Wood, and S. P. Ong, "Performance and cost assessment of machine learning interatomic potentials," *J. Phys. Chem. A* **124**, 731–745 (2020).
- ³²L. Himanen, M. O. Jäger, E. V. Morooka, F. Federici Canova, Y. S. Ranawat, D. Z. Gao, P. Rinke, and A. S. Foster, "Dscribe: Library of descriptors for machine learning in materials science," *Comput. Phys. Commun.* **247**, 106949 (2020).
- ³³F. Musil, A. Grisafi, A. P. Bartók, C. Ortner, G. Csányi, and M. Ceriotti, "Physics-inspired structural representations for molecules and materials," *Chem. Rev.* **121**, 9759–9815 (2021).
- ³⁴R. Batra, H. D. Tran, C. Kim, J. Chapman, L. Chen, A. Chandrasekaran, and R. Ramprasad, "General atomic neighborhood fingerprint for machine learning-based methods," *J. Phys. Chem. C* **123**, 15859–15866 (2019).
- ³⁵J. Chapman, T. Hsu, X. Chen, T. W. Heo, and B. C. Wood, "Quantifying disorder one atom at a time using an interpretable graph neural network paradigm," *Nat. Commun.* **14**, 4030 (2023).
- ³⁶K. Choudhary and B. DeCost, "Atomistic line graph neural network for improved materials property predictions," *npj Comput. Mater.* **7**, 185 (2021).
- ³⁷L. McInnes, J. Healy, N. Saul, and L. Großberger, "UMAP: Uniform manifold approximation and projection," *J. Open Source Softw.* **3**, 861 (2018).
- ³⁸A. H. Larsen, J. J. Mortensen, J. Blomqvist, I. E. Castelli, R. Christensen, M. Dułak, J. Friis, M. N. Groves, B. Hammer, C. Hargus *et al.*, "The atomic simulation environment—a python library for working with atoms," *J. Phys.* **29**, 273002 (2017).
- ³⁹M. Fey and J. E. Lenssen, "Fast graph representation learning with PyTorch geometric," *arXiv:1903.02428* (2019).
- ⁴⁰A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga *et al.*, "PyTorch: An imperative style, high-performance deep learning library," *Adv. Neural Inf. Process. Syst.* **32**, 8026–8037 (2019).
- ⁴¹J. Klicpera, J. Groß, and S. Günnemann, "Directional message passing for molecular graphs," *arXiv:2003.03123* (2020).
- ⁴²K. T. Schütt, H. E. Sauceda, P.-J. Kindermans, A. Tkatchenko, and K.-R. Müller, "SchNet—A deep learning architecture for molecules and materials," *J. Chem. Phys.* **148**, 241722 (2018).
- ⁴³X. Bresson and T. Laurent, "Residual gated graph convnets," *arXiv:1711.07553* (2017).
- ⁴⁴V. P. Dwivedi, C. K. Joshi, T. Laurent, Y. Bengio, and X. Bresson, "Benchmarking graph neural networks," *arXiv:2003.00982* (2020).
- ⁴⁵S. Elfving, E. Uchibe, and K. Doya, "Sigmoid-weighted linear units for neural network function approximation in reinforcement learning," *Neural Netw.* **107**, 3–11 (2018).
- ⁴⁶J. L. Ba, J. R. Kiros, and G. E. Hinton, "Layer normalization," *arXiv:1607.06450* (2016).
- ⁴⁷Z. Li, Q. Ding, and W. Zhang, "A comparative study of different distances for similarity estimation," in *Intelligent Computing and Information Science*, edited by R. Chen (Springer Berlin Heidelberg, Berlin, Heidelberg, 2011), pp. 483–488.
- ⁴⁸A. Samanta, M. A. Morales, and E. Schwegler, "Exploring the free energy surface using ab initio molecular dynamics," *J. Chem. Phys.* **144**, 164101 (2016).
- ⁴⁹A. P. Thompson, H. M. Aktulga, R. Berger, D. S. Bolintineanu, W. M. Brown, P. S. Crozier, P. J. in 't Veld, A. Kohlmeyer, S. G. Moore, T. D. Nguyen, R. Shan, M. J. Stevens, J. Tranchida, C. Trott, and S. J. Plimpton, "LAMMPS—A flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales," *Comput. Phys. Commun.* **271**, 108171 (2022).
- ⁵⁰X. Zhou, H. Wadley, R. Johnson, D. Larson, N. Tabat, A. Cerezo, A. Petford-Long, G. Smith, P. Clifton, R. Martens, and T. Kelly, "Atomic scale structure of sputtered metal multilayers," *Acta Mater.* **49**, 4005–4015 (2001).
- ⁵¹J. Chapman, K. E. Kweon, Y. Zhu, K. Bushick, L. B. Bayu Aji, C. A. Colla, H. Mason, N. Goldman, N. Keilbart, S. R. Qiu, T. W. Heo, J. Rodriguez, and B. C. Wood, "Hydrogen in disordered titania: Connecting local chemistry, structure, and stoichiometry through accelerated exploration," *J. Mater. Chem. A* **11**, 8670 (2023).
- ⁵²X. W. Zhou, R. A. Johnson, and H. N. G. Wadley, "Misfit-energy-increasing dislocations in vapor-deposited CoFe/NiFe multilayers," *Phys. Rev. B* **69**, 144113 (2004).
- ⁵³B. J. Alder and T. E. Wainwright, "Phase transition for a hard sphere system," *J. Chem. Phys.* **27**, 1208–1209 (1957).
- ⁵⁴M. Rintoul and S. Torquato, "Metastability and crystallization in hard-sphere systems," *Phys. Rev. Lett.* **77**, 4198 (1996).
- ⁵⁵Y. Jiao, F. H. Stillinger, and S. Torquato, "Nonuniversality of density and disorder in jammed sphere packings," *J. Appl. Phys.* **109**, 013508 (2011).