Characterizing Disorder Within Atomic Environments Through Physics-preserving Graph Neural Networks

James Chapman*1, Tim Hsu†2, Xiao Chen2, Tae Wook Heo1, and Brandon C. Wood1

1Materials Science Division, Lawrence Livermore National Laboratory, Livermore, CA, USA
2Center for Applied Scientific Computing, Lawrence Livermore National Laboratory, Livermore, CA, USA

March 22, 2022

Abstract

Quantifying the level of atomic disorder within materials is critical to understanding how evolving local structural environments dictate performance and durability. Here, we leverage graph neural networks to define a physically interpretable metric for local disorder. This metric encodes the diversity of the local atomic configurations as a continuous spectrum between the solid and liquid phases, quantified against a distribution of thermal perturbations. We apply this novel methodology to three prototypical examples with varying levels of disorder: (1) solid-liquid interfaces, (2) polycrystalline microstructures, and (3) grain boundaries. Using elemental aluminum as a case study, we show how our paradigm can track the spatio-temporal evolution of interfaces, incorporating a mathematically defined description of the spatial boundary between order and disorder. We further show how to extract physics-preserved gradients from our continuous disorder fields, which may be used to understand and predict materials performance and failure. Overall, our framework provides an intuitive and generalizable pathway to quantify the relationship between complex local atomic structure and coarse-grained materials phenomena.

1 Introduction

Understanding how a material’s structure affects its properties is one of the most fundamental principles in materials science. At the center of this paradigm is the fact that macroscopic materials behavior begins at the atomic scale, with local atomic arrangements ultimately coming together to form structural features observed at larger length scales [1, 2]. Characterizing the nature and propagation of these local environments is therefore vital to understanding macroscale structure-property relationships and their evolution [3]. Complicating this endeavor is the fact that the long-range features often depend on structurally disordered atomic environments, which tend to dictate materials functionality [4]. For instance, transport, chemical reactivity, and phase nucleation are all profoundly affected by the presence of interfaces, interphases, and grain boundaries [5, 7]. These processes in turn are intricately connected to performance-durability trade-offs in both functional [8] and structural [9] materials. Examples include temperature-dependent microstructure evolution [10, 11], hotspot formation [12, 13], and the nucleation and growth of new material phases [14, 15].

However, quantifying local atomic disorder in a physically motivated way in practice is extraordinarily difficult [16]. Although a number of methods have been proposed to characterize local atomic environments, these methods are often not optimized to magnify the subtle differences
Figure 1: General workflow for calculating SODAS values. Atomic structures are converted into graph representations, which explicitly encode all necessary geometric information. These atomic graphs are then fed into a graph neural network, which has been trained to distinguish between the unique local geometries in different material phases. The graph neural network then gives each atomic environment a SODAS value, which classifies where in the phase space, between phases, that local structure is most likely to occur.

Existing methods can typically be grouped into three general classes, each of which carries distinct trade-offs: (1) *bond-angle functions* such as Adaptive Common Neighbor Analysis (CNA) [17], Steinhardt order parameters (SOP) [18], Ackland-Jones order parameters (AJ) [19], and atomic "Smoothness" metrics (ASM) [20], the local atomic environment metric (LAE) [21]; (2) *parameterized symmetry functions* such as the Smooth Overlap of Atomic Positions (SOAP) [22], Behler-Parinnello functions (BP) [23], Moment Tensor Representations (MTR) [24]; and the Adaptive Generalizable Neighborhood Informed functions (AGNI) [25]; and (3) *unsupervised machine learning methods* which include graph-based [26] and image-based [27] representations.

In general, it is highly desirable to develop a methodology that is by construction specifically designed to distinguish, quantify, and physically interpret regions with varying degrees of atomic disorder. Such a capability would enable more accurate predictions of how disordered atomic environments translate to higher-level features and functionality. For instance, mapping between discretized atomistic models and continuous field representations, such as phase-field [28] and finite-element [29] models, forces the use of ill-defined and arbitrary approximations [30], particularly when disorder is present. Moreover, continuous field representations propagate via local gradients [31], the evaluation of which amplifies inaccuracies associated with disordered regions. Addressing these shortcomings is therefore a critical priority.

To this end we introduce a new methodology composed of 2 stages: (1) explicit atomic structure representations with graph neural networks (GNN), and (2) stochastically mapping the local atomic structure to a thermodynamic order parameter using the GNN. Here, this mapping, henceforth referred to as the Stochastic Orderness Degree for Atomic Structures (SODAS), aims to describe an atom’s local geometry as the likelihood of that geometry existing within the abstract space between 2 unique material phases. Qualitatively one can imagine this space as a line drawn between a solid and a liquid, with any point along this line representing the "closeness" of the atomic environments at that point to either of the phases. Points closer to the crystal would be more ordered, while those closer to the liquid would be more disordered. The GNN implicitly learns where an atomic environment exists in this space during its optimization.
Importantly, the graph-based representations of atomic geometries encode the chemical and bond information directly onto the graph, alleviating the need to approximate the local geometry through highly-parameterized functions. As the graph itself contains only radial information, such as the scalar distance between atoms, it is inherently invariant under all transformation operations such as permutations, rotations, and translations. Since the GNN uses the graph-encoded information to establish the SODAS mapping, the underlying physics of the structure is preserved, as all knowledge of the structure is explicitly represented within the graph and propagated throughout the GNN.

The remainder of this paper is as follows. We begin by discussing several validation scenarios that showcase the accuracy and transferability of our proposed methodology. We show how the GNN SODAS predictions intuitively describe the gradual melting of homogeneous bulk structures. This is followed by an investigation into how well SODAS can classify chaotic solid-liquid interfaces, generated via classical molecular dynamics (CMD) simulations. We then showcase how the combination of SODAS and SGOP can be used to identify and characterize the time-dependant evolution of several microstructural environments during grain coarsening on systems composed of greater than 1.5 million atoms. Lastly, we provide a pathway of mapping the atomic configurations to a continuous scalar field to predict potential stress points within grain boundary regions. We culminate this work by discussing the impact of this work and its future use throughout the materials science domain.

2 Methods

2.1 Stochastic Orderness Degree for Atomic Systems

We define SODAS conceptually as the abstract point in phase space between two distinct atomic arrangements. Qualitatively, SODAS is defined as the stochastic mapping between local structure and temperature, with the assumption that the level of disorder present within the atomic environment will increase as a function of the system’s temperature. Specifically, each atom in the configuration is assigned a SODAS value, represented as \( \lambda_i \), via a GNN prediction, and is defined by:

\[
\lambda_i = f(\lambda_k, \lambda_{k+1}, \ldots, \lambda_{k+m})
\]  

(1)

where \( \{\lambda_k, \lambda_{k+1}, \ldots, \lambda_{k+m}\} \) represents the set of \( \lambda \) values associated with a given local atomic geometry, \( k \) represents a given temperature, \( m \) is defined as the size of the set, and \( i \) is an individual atom’s identifier. \( \lambda_i \) and local structure are connected stochastically (represented by \( f \)) via the GNN, with a single atomic environment being classified as a weighted combination of \( \lambda_k \) values (with \( k \) representing a given temperature), with the weights representing the likelihood of that particular environment occurring at different temperatures.

During training, all atoms at a given temperature are assigned the same value of \( \lambda_k \). In this way, the value of \( \lambda_k \) is not directly connected to an atom’s local structural geometry. However, at each temperature local atomic geometries bounce around an equilibrium point, which in this case can be thought of as the 0K structure. As the temperature increases the magnitude of the displacement from this structure increases, though some atomic neighborhoods may resemble low temperature structural motifs, even at higher overall temperatures.

\( \lambda_k \) is a continuous metric that ranges from 0 to 1, with 1 representing one phase and 0 representing the other. In this work we examine how this metric can be used to distinguish between a highly ordered crystal (\( \lambda_k = 1 \)) and a chaotic, disordered liquid (\( \lambda_k = 0 \)). \( \lambda_k \) is defined via the temperature of a configuration of atoms, and is represented as:

\[
\lambda_k = \frac{1}{1 + \exp\left(-\left(T_m/T_k\right)^s\right)}
\]  

(2)

where \( T_k \) the current temperature of the system, \( T_m \) is the temperature of the phase transition between solid and liquid, at a specific pressure, and \( s \) is a empirical scaling metric that determines
where in the temperature spectrum to begin the decay of $\lambda_k$ from ordered to disordered. The introduction of $s$ makes the definition of $\lambda_k$ universal to any material system, as one can simply tailor the value of $s$ for any unique material system. A plot of the relationship between $\lambda_k$ and temperature for a $s$ value of $\frac{3}{2}$, which was used in this work, can be found in the supplemental information.

It is important to note that this employing $T_m$ as the temperature of the phase transition between solid and liquid is not absolute, as one can define order and disorder in many unique ways. This ultimately allows the definition of $\lambda_k$ to be tailored as well to suite the needs of the user.

2.2 Training Data Preparation

Classical molecular dynamics (CMD), using the LAMMPS software package [32], was used to generate training data for the GNN model. Starting from bulk FCC aluminum (containing 1024 atoms), CMD was performed in the NVT ensemble for 1 nanosecond using the Zhou et al EAM potential [33], increasing the temperature linearly with respect to time. The initial and final temperatures for the CMD simulation was 100 and 1200, respectively. 1000 equidistant snapshots were then taken along the trajectory and labelled according to their respective temperatures. Further information regarding the training data preparation can be found in the supplemental information.

2.3 Graph Neural Network Implementation

2.3.1 Conversion to Graph

Prior to GNN operation, we converted the atomic systems into graphs by a simple cutoff radius-based neighbor list search (implemented using Atomic Simulation Environment [34]), with the cutoff $R_c = 3.5$ Å. Each node of the converted graph corresponds to the atom type $z$, and each edge the bond distance $d$.

2.3.2 GNN Operation

The GNN model used in this work consists of three components: the initial embedding, the atom-bond interactions, and the final output layers (Fig. 1).

In the initial embedding, each atom type $z$ is transformed into a feature vector by an Embedding layer (PyTorch [35]). Each bond distance $d$ is expanded into a $D$-dimensional feature vector by the Radial Bessel basis functions (RBF) [36]

$$\text{RBF}_n(d) = \sqrt{\frac{2}{R_c}} \sin\left(\frac{n\pi}{R_c}d\right),$$

(3)

where $n \in [1..D]$ and $R_c$ is the cutoff value. Both atom and bond feature vectors have the same length $D = 100$.

The atom-bond interactions are also known as GNN convolution, aggregation, or message-passing. There are many variants of GNN convolution operations that can be adopted from the literature. In this work, we choose the edge-gated graph convolution [37, 38]. The term atom-bond interaction is based on the fact that the nodes and the edges exchange information during the convolution operation. Specifically, the node features $h^{l+1}_i$ of node $i$ at the $(l+1)$th layer is updated as

$$h^{l+1}_i = h^l_i + \text{SiLU}\left(\text{LayerNorm}\left(W^l_i h^l_i + \sum_{j \in N(i)} \hat{e}^l_{ij} \odot W^l_d h^l_j\right)\right),$$

(4)

where SiLU is the Sigmoid Linear Unit activation function [39]; LayerNorm is the Layer Normalization operation [40]; $W_i$ and $W_d$ are weight matrices; the index $j$ denotes the neighbor node of node $i$; $\hat{e}^l_{ij}$ is the edge gate vector for the edge from node $i$ to node $j$; and $\odot$ denotes element-wise multiplication. The edge gate $\hat{e}^l_{ij}$ at the $l$th layer is defined as

$$\hat{e}^l_{ij} = \frac{\sigma(e^l_{ij})}{\sum_{j' \in N(i)} \sigma(e^l_{ij'}) + \epsilon},$$

(5)
where $\sigma$ is the sigmoid function, $e_{ij}^l$ is the original edge feature, and $\epsilon$ is a small constant for numerical stability. The edge features $e_{ij}^l$ is updated by

$$
e_{ij}^{l+1} = e_{ij}^l + \text{SiLU} \left( \text{LayerNorm} \left( W_g z_{ij} \right) \right),$$

where $W_g$ is a weight matrix, and $z_{ij}$ is the concatenated vector from the node features $h_i$, $h_j$, and the edge features $e_{ij}$:

$$z_{ij} = h_i \oplus h_j \oplus e_{ij}. \quad (7)$$

Lastly, in the final output layers, each node feature is eventually transformed into a scalar output $y$ ranging from 0 to 1. Effectively, the GNN predicts the SODAS metric for every atom.

Further details regarding model training are described in Supporting Information.

2.4 Atomistic Simulation Details

2.4.1 2 Phase Simulations

2 phase CMD simulations were performed on a orthogonal block of aluminum containing 23,040 atoms. All CMD simulations were performed in the NVT ensemble. Initially, half the box was superheated to 4000K to assure complete melting, while the remaining half of the box was set at 200K. This initial CMD simulation was performed for 1 nanosecond with separate NVT thermostats driving each region. The 4000K and 200K regions were then allowed to coexist using a single NVT thermostat to drive the CMD simulation. For the purpose of this analysis upon the introduction of the shared thermostat the temperature was fixed to 3 cases: one at 200K, another at 1000K, and a final case of 1500K. The purpose of these temperatures is to observe how our proposed methodology predicts the unique levels of atomic disorder present in each case. Further information regarding these simulations can be found in the Results section and in Fig. 3.

2.4.2 Grain Coarsening Simulations

CMD simulations in the NVT ensemble were performed for 6 polycrystalline cases, each with a varying number of initial grains. An initial bulk aluminum system containing roughly 1.6 million atoms was used to construct 6 polycrystalline structures, using the Atomsk software package [41], containing 5 and 250 initial grains. CMD simulations were performed on each case at 200K, 400K, and 600K. NVT simulations were run for approximately 1.5 nanoseconds for each combination of initial structure and temperature. All CMD simulations were done using LAMMPS with the Zhou et. al EAM potential [33]. Further details regarding the polycrystalline structures can be found in the Results section as well as the Supplemental Information.

2.5 Microstructure Characterization

Microstructure characterization occurs in 4 stages: (1) calculation of SODAS for all atoms in the system, (2) thresholding of the atomic configuration, based on an atom’s SODAS value and subsequent removal of all atoms below the threshold value, (3) conversion of the remaining atoms to a graph representation for the discovery of subgraphs within the graph, and (4) characterization of the grains through a graph order parameter. Fig. 4 (b) depicts this workflow visually. While step (1) requires little-to-no input from the user, step (2) requires one to define the level of disorder that needs to be captured when defining the interface regions. This choice highlights the intuitive nature of our proposed methodology, as the threshold value defines the structural properties of the interface region itself, with a near-zero threshold indicating grain boundaries which are extremely disordered and a value close to 1 representing highly crystalline boundaries. In principle, both classes of interfaces can exist within the same structure, which would require a more complex thresholding system, though for this work we assume a uniform local atomic environment amongst all grain boundaries.
For all microstructure characterization tests in this work we employ a $\lambda$ threshold of 0.7, implying that we are defining grain boundaries as local atomic environments showing disorder equivalent to the atomic perturbation observed at around 900K. $\lambda = 0.7$ was chosen due to the maximum temperatures described in the previous section. As we should expect the grain regions to experience perturbations no greater than those encountered at 600K, those observed at 900K should provide a good approximation for atoms that do not belong to the grains. Once the system has been thresholded, all atoms below the threshold value are removed, leaving only the atoms belonging to the grains.

This system is then mapped onto a graph, $G$, shown in Fig. 4(b), where edges are represented by $ij$ pairwise interactions within a 4Å cutoff radius. A recursive subgraph search algorithm is employed to discover all connected subgraphs, $S_G$, within the complete graph $G$. This algorithm is extremely efficient, discovering all subgraphs within a 1.6 million atom system in 1.2 seconds. As all interface atoms were removed prior to the graph construction, all subgraphs in $G$ represent the resulting grains contained within the structure. A 2D slice of the discovered grains in the 3D system is seen in Fig. 4(b). As each grain is represented by a subgraph in $G$, they can be characterized as graphs. This provides more detail into the grain’s shape and connectivity, rather than relying on metrics such as the number of atoms, diameter, etc, which are inherently not unique and do not provide a true quantitative measure of the underlying properties of the structure. Here we use a recently developed graph topology metric \cite{42} to characterize the subgraphs in $G$. Each edge in the subgraph is labelled according to:

$$\omega_{i,j} = \frac{1}{d_{i,j}} \quad \forall d_{i,j} \leq R_c$$ \hspace{1cm} (8)

$i$ and $j$ are the atomic indices of the atoms (note that self-interaction terms are not allowed). $d_{i,j}$ is defined as the $l^2$-norm between two atoms. $R_c$ is the cutoff radius, which was chosen as 4Å. $\omega_{i,j}$ represents the weight of a given edge for a specific pair of adjacent nodes in the graph. The degree of each node is finally defined as the sum of the elements in a node’s edge set, $d_i = \sum_j \omega_{i,j}$, where $J$ is the set of all neighbors of atom $i$. The degree sets are then fed into the scalar graph order parameter (SGOP) \cite{42} scheme for the final characterization of the subgraph. The SGOP functional form is defined as:

$$\theta_{S_G,R_c} = \sum_{m} D_s \log_b P(d_m) P(d_m) + d_m P(d_m)$$ \hspace{1cm} (9)

where $D_s$ is the set of unique node degrees in a subgraph, with $P_{dm}$ being the probability of
Figure 3: 2-phase simulation snapshots and their corresponding SODAS predictions. The interface structure is taken at the moment the 2 phases are allowed to coexist. The 200K, 1000K, and 1500K snapshots are taken at the end of each MD simulation.

a given degree, \( d_m \), occurring in the subgraph. As discussed in previous works, the SGOP value provides a measure of the shape and connectivity of the graph at a structural level. In this way we can observe, quantitatively, the difference between isotropic and anisotropic grains, even if they have the same number of atoms, diameter, circumference, density, etc.

### 3 Results

#### 3.1 Bulk Characterization

SODAS is used here to characterize the atomic environments encountered within the reference data used to train the GNN model. Fig. 2 provides a visual depiction of the atomic-level characterization. Here we see that the 0K structure is uniformly predicted to have a \( \lambda = 1 \), which is indicative of the perfect crystal. Subsequently, at 1200K, all atoms are uniformly predicted to have a \( \lambda = 0 \), which again is intuitive as the structure has reached the liquid phase. However, the structures that exist between these temperature extremes provide a more interesting characterization.

As shown in Fig. 2, structures that exist between 0K and 1200K yield a spectrum of atomic environments, often spread anisotropically throughout the structure. For example, the second structure in Fig. 2 which represents a structure at roughly 200K, has \( \lambda \) values ranging from 0.95 and 0.5. As described in the Methods section, this is due to similar atomic environments being present at several temperatures, resulting in the GNN averaging the \( \lambda \) values. Intuitively this makes sense, as the goal of the SODAS metric is to judge the likelihood of an atomic environment existing at an arbitrary point along the thermodynamic spectrum between 2 material phases. If a unique atomic environment occurs at multiple temperatures, one would expect its \( \lambda \) to be some weighted combination of the individual occurrences of the environment along the temperature spectrum.

#### 3.2 Disordered Solid-Liquid Interfaces

While crystalline interfaces are difficult to identify due to their structural similarities when compared to the crystal phase, disordered interfaces are complex because of their inherently chaotic nature. A multitude of unique local atomic environments can be present in disordered interfaces, making classification of both the ordered and disordered region’s morphology unintuitive. As was the case with the crystalline interfaces, bond-angle methods such as CNA and AJ often fail to distinguish between perturbed crystal and truly disordered regions, due to their corresponding functional
form’s overlap for anisotropic geometries. Methods such as SOAP will provide a more quantitative depiction of the system, but will ultimately again suffer from a clustering problem, where truly disordered atomic environments will be grouped together with highly perturbed crystalline environments, especially at the interface of the 2 phases.

To this end, we have performed 2-phase CMD simulations at several temperatures (200K, 1000K, and 1500K), to observe how SODAS classifies the unique structural environments present in each scenario. Further details regarding the simulation setup can be found in the Methods section. Fig. 3 shows the SODAS characterization of the atomic environments present for the three cases, along with two references (1) the perfect crystal and (2) the interface prior to phase coexistence.

For the CMD simulation at 200K SODAS correctly classifies all atomic geometries as being close to the perfect crystal upon equilibration of the 2 phases. It also identifies a small cluster of local disorder, similar to those found in the crystalline grain boundaries, indicating the presence of a point-defect like region. Interestingly, SODAS predicts this region within the previously melted phase, indicating that the defect was created as the liquid section was quenched down to 200K. At 1000K SODAS predicts the coexistence of the two phases, albeit with a slightly reduced crystalline region. Within the crystal region one can see that SODAS identifies the presence of moderate perturbations relative to the perfect crystal. We also see that the interface is not a perfect line between the 2 phases, and instead more closely resembles a jagged line. This is another example of how SODAS can be used to more accurately map the morphology of such structural domains.

At 1500K SODAS again correctly classifies all atomic environments as having significant disorder, relative to the crystal phase, indicating that the system has undergone a complete solid-to-liquid phase transformation. SODAS does however, predict varying levels of disorder as well as the presence of pockets within the liquid that show moderate-to-low disorder. Having the ability to predict structural disorder with this level of fidelity could make SODAS particularly useful in examining the structure-property relationships of highly complex atomic environments.

3.3 Autonomous Microstructure Characterization

Here, we showcase how a combination of GNN SODAS predictions and a graph topology algorithm can autonomously discover and characterize the complex structural environments encountered within a polycrystalline system as it evolves through both time and temperature domains. As described in the Computational Methods section, we employ three initial grain structures on an approximately 1.6 million atom FCC aluminum system containing 5, 50, and 250 initial grains respectively. Fig. 4 provides a visual depiction of SODAS ability to accurately map the local atomic environments in a physically-intuitive manner. In Fig. 4 one can see the clear differences in grain structure between the 5 and 250 grain cases, as well as how each system’s final structure changes as a function of temperature.

For the case of 5 initial grains, Fig. 4 clearly captures the coarsening of grains in the 600K case, where the two grains coalesce into a single grain across multiple planes. At 200K however, Fig. 4 indicates that grain coarsening does not occur, likely due to the kinetic energy present in the system being insufficient to overcome the potential barrier required for coalescence. Similarly, for the 250 grain case, we see strikingly different grain structures at 200K and 600K, as one would expect. At 200K, while many of the smaller grains coalesce, there is simply not enough kinetic energy to allow for larger grains to merge. At 600K, the final structures yields fewer and larger grains, as many of the medium-sized grains in the 200K’s finals structure were able to amalgamate. In all scenarios, it is clear from the color-coding that SODAS can provide a physically intuitive metric to determine the boundary of each grain, while also providing us with a way of quantifying the level of disorder present in both the grain and grain boundary regions.

Fig. 5 provides a quantitative understanding of how the grain structure changes as both a function of the initial grain morphology, as well as the temperature. Using Eq. 9 each grain discovered via the SODAS calculations was characterized to give a physical interpretation of both their respective shapes and connectivities. This provides an understanding of the grain structure with a higher level of fidelity than other metrics such as the number of atoms in the grain, or the grain radius. Here, each histogram shown in Fig. 5 is generated using the SGOP values calculated
Figure 4: (a) SODAS predictions on the initial and final configurations of the polycrystalline MD simulations. Each column block represents the initial grain structure present in the system, with the right-side representing the 5 initial grains, and the left side indicative of the 250 initial grains. Each row, color-coded in green and yellow, represents the temperature of the MD simulation, with green representing 200K and yellow being 600K. (b) Workflow of the unsupervised graph-based grain detection algorithm, visualized using simulated polycrystalline Al. (Left) SODAS values for each atomic environment present in the system. (Middle) Atoms-to-graph mapping (after SODAS thresholding), where node colors represent the connectivity of a given atom. (Right) Autonomous grain detection using recursive subgraph clustering.

Fig. 5 shows how the initial grain structure changes as a function of temperature. Fig. 5(a) showcases the temperature effects on the grain structure containing 250 randomly constructed grains. As is described in the Methods section, for every configuration a graph order parameter value is assigned for each grain in the structure. A normal distribution is then used to fit these values, whose mean and standard deviation can be used as microstructural feature vector. Fig. 5(a) provides insight into how the mean and standard deviation of this fitted distribution changes as a function of time.

Fig. 5(a,top-left) shows how the mean of the normal distribution changes over time. At 200K there is initially an increase in the mean, implying that larger, more connected grains exist throughout the structure. However, a plateau is reached within 2 ns, and no further increase in the mean can be seen. At 400K a similar trend can be seen, though some further growth in the grains can be seen between 10 and 14 ns. At 600K one can observe a constant trend, indicating that grain growth occurs over the course of the entire CMD simulation, with no signs of convergence.

Fig. 5(a,top-right) shows how the standard deviation of the normal distribution changes over time. Changes in the standard deviation can be attributed with grain coalescence, because as one grain grows in size, another neighboring grain must reduce in size as it is swallowed by the growing grain. This process ultimately causes a fluctuation in the spread of graph order parameter values, which is captured by the standard deviation of the fitted distribution.

At 200K there is initially an increase in the standard deviation, indicating the grain coalescence is occurring. However, in agreement with the mean values, a plateau is reached within 2 ns, and no further increase in the mean can be seen. At 400K, one can observe a constant increase in the standard deviation, indicating that grain coalescence is occurring throughout the coarse of the trajectory. However, the slope of this line can be qualitatively linked to the rate of coalescence, with a smaller slope being attributed to slower coalescence. At 600K one can observe a similar trend to 400K but with a larger slope, indicating that grain coalescence at 600K occurs more rapidly than it
Figure 5: Quantitative metrics characterizing microstructural changes over time. In each case the colors correspond to the temperature of the MD simulation. (a) Changes in the graph order parameter distribution’s (top, left) mean, (top, right) standard deviation, (bottom, left) multiplication of the mean and standard deviation, and (bottom, right) number of grains, all as a function of time, starting from the 250 grain structure. (b) SGOP microstructure characterization performed the SODAS predicted values for the case of (top) 5 initial grains and (bottom) 250 initial grains. Histograms are determined using the resulting SGOP values, calculated on the final structure from each MD simulation (the last point in each subplot in (a)). Inserted images, and corresponding arrow, show the final structure obtained from the 600K MD simulations. All histograms are normalized and show each x-value’s probability.

does at 400K.

Fig. 5 (a,bottom-left) combines the mean and standard deviation into a single metric which aims to capture the physical nature of both values. Here, the product of the two terms is used to gauge both the rate of grain growth, but also the magnitude of the grains themselves. At 200K we see a trend that confirms our previous observation that grain growth occurs rapidly, initially, but tapers off quickly. At 400K we observe a more complete picture, where slow grain growth occurs throughout the trajectory. For the case of 600K, faster growth can be seen, as well as the existence of larger grains.

This picture can also be complimented by observing how the number of grains changes as a function of time, as seen in Fig. 5 (a,bottom-right). We can see that at 200K, there is an initial reduction in the number of grains from 250 to approximately 75, where that number hold for the remainder of the simulation. At 400K there is a constant reduction in the number of grains from 250 roughly 30 by the end of the CMD simulation. At 600K the number of grains is reduced even further by the end of the simulation, going from 250 to only 12 grains. The picture painted in Fig. 5 (a,bottom-right) aligns well with the values predicted in Fig. 5 (a,bottom-left), indicating that our product metric can be used to quantify both the size and shape of the grains, but also the rate at which they grown under dynamic conditions.

Fig. 5 (b) provides insight into the final configuration’s microstructure for two scenarios: (top) initial microstructure of 5 random grains, (bottom) initial microstructure of 250 random grains. For the case of (b,top), at 200K there is an abundance of larger grains, as little-to-no grain coalescence occurred. At 400K, one can see a broadening of the distribution, indicating the existence of grains
both smaller and larger than those existing at 200K. This is due to a pair of two smaller grains merging into a larger grain, resulting in the larger grain increasing in size and the remaining components of the two smaller grains forming a single even smaller grain. The same trend continues at 600K, though happening more frequently, with even larger grains splitting and merging with other smaller grains, leaving more medium sized grains overall. Interestingly, this 600K phenomena leads to an overall decrease in the size of grains, also in part due to an increase in the size of the grain boundary regions, due to increased atomic perturbations near the edges of the crystalline regions.

A similar trend can be seen in Fig. 5(b, bottom) which showcases the temperature effects on the grain structure containing 250 randomly constructed grains. As the density of the grains has increased from the 5 grain system, one expects the size of the grains to be much smaller, as is predicted by SGOP. For this case however, an opposite trend can be observed when compared to the 5 grain case. Here, the overall size of the grains increases as a function of temperature, as many small grains are energetically unstable, and easily coalesce as the temperature is increased. Eventually, given enough time, the 250 grain scenario would converge to the observed with 5 initial grains once the grains grew to a large enough size. However, in all scenarios considered, it is likely that with a high enough temperature and a long enough simulation, a single crystal would be obtained, albeit with various defects scattered throughout. The important take-away from Fig. 5 is that SGOP can provide a quantitative and physically intuitive understanding of the microstructural morphology present within the system, which is characterized by SODAS, and provide a detailed description of how those features change under dynamic conditions.

### 3.4 Mapping Atoms to a Continuous Field

Lastly, since $\lambda$ is continuously valued over the tightly packed, discrete atoms, $\lambda$ can be converted to a full continuous field. We showcase this concept for the example of crystalline interface regions, such as those seen in symmetric tilt and twin boundaries, and edge/screw dislocations, which are often difficult to interpret due to their striking similarities to the crystalline phases of the underlying material. Further, grain boundary topology becomes non-trivial as the question of which atoms belong to the grain and which belong to the grain boundary is not intuitive due to the lack of significant structural disorder.

Existing methods such as CNA (as well as a-CNA), AJ, and SOAP provide descriptions of the local atomic geometries, though in most cases do not provide an intuitive metric for universally answering this question. For example, both CNA and AJ rely on strict and limited reference information to determine which "class" the local geometry belongs to. Due to significant overlap between classes, however, especially at non-zero temperatures, determining which class an environment belongs to is not obvious. While symmetry functions such as SOAP provide a more detailed description of the local geometries (though still yield non-unique representations), they still suffer from the same major drawback as CNA and AJ.

SODAS addresses this deficiency through its stochastic thermodynamic nature, relying on characterizing an atom’s local geometry in terms of its location in phase space, rather than its position within the structural space. Fig. 6 shows SODAS predictions on atoms for two highly-crystalline grain boundaries of aluminum. One can see the intuitive nature of SODAS, cleanly characterizing the "grain" regions with a $\lambda$ close to 1, and varying degrees of disorder present at the interface region. For boundaries that show higher degrees of crystallinity, such as those in $\Sigma 5(110)[120]$, the disorder present at the interface is minimal, as is expected, though is still clearly present. Likewise, for more disordered boundaries, such as those in $\Sigma 9(110)[110]$, a greater degree of disorder is detected within the interface region. These characterizations exemplify the notion of employing SODAS to provide a physics-informed prediction capability, where one can intuitively determine where the grain begins and ends without the need for complex reference frames and intrinsic uncertainty within the method itself.

One can then map the discrete information from the atomic system onto a continuous scalar field. Fig. 6 shows the continuous fields of the originally discrete, per-particle SODAS value $\lambda$. Additionally, the gradient norm $||\nabla \lambda||$ was calculated and visualized. This discrete-to-continuum conversion was done by interpolating the discrete $\lambda$ values onto a uniform grid using PyVista.
Figure 6: Continuous fields (and its gradient norm) of the originally discrete, per-particle SODAS value $\lambda$. The discrete-to-continuum conversion is done by interpolating the discrete $\lambda$ onto a uniform, fine grid. The gradient information can then be computed over the uniform grid.

From Fig. 6 one can observe how the original SODAS values are extrapolated into a continuous field. For the case of $\lambda$ we observe that the interface regions of the structures show a decrease in order due to the inherently disordered nature of the boundary region. When calculating the gradient of this field we observe areas of the structure where there are sharp changes in the SODAS values. The changes always occur near the grain boundaries, which is again expected, as these are locations within the structure where there is an abrupt change in the level of disorder present normal to the boundary plane.

Interestingly, the sensitivity of this detection can be seen in Fig. 6, where the gradient of the scalar field predicts two regions where there is an abrupt change in the SODAS values. As we move from the crystalline regions towards the interface along a vector normal to the boundary region, we first encounter a crystal-to-boundary region, followed by the boundary itself, and finally a boundary-to-crystal region as we move away from the interface. Therefore, the gradient predictions in Fig. 6 make intuitive sense, as the predicted abrupt changes in the SODAS values correspond to the crystal-to-boundary transition, with a clear separation between the two regions due to the similarities in the SODAS values of atoms within the boundary.

We also note the similarity between such a continuous field representation and the phase field representation used in phase field models. The continuous field representation, originally based on the atomistic calculations, can be used to inform the phase field order parameter, thereby establishing a connection between discrete, first-principles atomic calculations and continuous, phase field models.

4 Discussion

Understanding the relationship between atomic-level environments and long-range structural features is vital to understanding how atomic-level perturbations alter a material’s properties. Quantifying phenomena such as the nucleation and growth of interfaces in polycrystalline configurations relies on an accurate description of how local atomic geometries evolve over time, leading to differences in the long-range structure of the material. To this end, we have developed a characterization scheme which provides a description of the level of disorder present in an atomic geometry. By leveraging the explicit physics-preserving nature of graph representations, with the learning
power of machine learning, we establish a pipeline in which graph neural networks can learn the subtle differences in local structure between different material phases. This capability provides us with a continuous metric which scores a local atomic environment based on its location in the abstract phase space. We showcase how one can use this continuous metric to understand the long-range structural features present in a material by autonomously predicting grain nucleation during dynamic atomistic simulations, and also showing how to map the discrete point clouds of the atomistic picture onto a scalar field at grain boundaries.

We propose the use of our methodology to study a variety of complex materials characterization problems, particularly those which rely on the subtle and intricate relationship between atomic perturbations and long-range structure. Due to the explicit encoding of structural information via graph representations, our methodology has the ability to uniquely identify even subtle differences in local atomic geometries, making it ideal for predicting the differences between material phases. Rather than attempting to uniquely classify each atomic environment in an unsupervised manner, we instead characterize these local structures as a continuous value between two unique material phases, allowing for an intuitive and physics-aware way to identify even subtle structural differences both locally and over longer length scales. In principle, one can also employ more than two channels, allowing for the identification of more complex features within the material.

We also envision our proposed methodology as a vital tool in providing more physically reliable and accurate simulation models that rely on continuous field representations, such as those employed in the phase field and continuum models. Our ability to map the discrete atomic structure to a scalar field representation through our SODAS metric to predict properties such as stress points at interface regions, showcases that our computational framework can be reliably used to switch between representation schemes. Rather than approximating what the field representation should be, our workflow can provide an explicit atoms-to-field mapping, allowing for more accurate and physics-informed phase field models. This mapping also serves to provide access to an accurate and interpretable way to understand the gradients of these fields, a property of great importance in modelling how continuous representation models evolve through time. This capability provides phase field and continuum models with an unprecedented level of accuracy, as their foundations can be built from the explicit atomistic representation.

Supporting Information

This work contains supplemental information which can be found online.

Data Availability

All data required to reproduce this work can be requested by contacting the corresponding author.

Acknowledgements

J. Chapman, T. Hsu, X. Chen, T. W. Heo, and B. C. Wood are partially supported by the Laboratory Directed Research and Development (LDRD) program (20-SI-004) at Lawrence Livermore National Laboratory. This work was performed under the auspices of the US Department of Energy by Lawrence Livermore National Laboratory under contract No. DE-AC52-07NA27344.

Author Contributions

X. Chen and B. C. Wood supervised the research. J. Chapman performed all MD simulations, devised and implemented the autonomous microstructure characterization methodology. T. Hsu trained the GNN and performed all GNN-related predictions. J. Chapman and T. Hsu devised the
orderness concept. B. Wood, T. Hsu, and J. Chapman devised the atoms-to-field mapping, while T. Hsu implemented it. J. Chapman and T. Hsu wrote the manuscript with inputs from all authors.

Competing Interests

The authors declare no competing financial or non-financial interests.

References

1. Fish, J., Wagner, G. J. & Keten, S. Mesoscopic and multiscale modelling in materials. Nature Materials 20, 774–786 (June 2021).

2. Miller, R. E. & Tadmor, E. B. A unified framework and performance benchmark of fourteen multiscale atomistic/continuum coupling methods. Modelling and Simulation in Materials Science and Engineering 17, 053001. https://doi.org/10.1088/0965-0393/17/5/053001 (May 2009).

3. Karma, A. & Tourret, D. Atomistic to continuum modeling of solidification microstructures. Current Opinion in Solid State and Materials Science 20. Recent Advances in Solidification Microstructure- Experiments and Computational Analysis, 25–36. ISSN: 1359-0286. https://www.sciencedirect.com/science/article/pii/S1359028615300061 (2016).

4. Guo, S. et al. Fe-based amorphous coating for corrosion protection of magnesium alloy. Materials and Design 108, 624–631 (2016).

5. Lin, Y., Skaff, H., Emrick, T., Dinsmore, A. D. & Russell, T. P. Nanoparticle Assembly and Transport at Liquid-Liquid Interfaces. Science 299, 226–229 (2003).

6. Wacaser, B. A. et al. Preferential Interface Nucleation: An Expansion of the VLS Growth Mechanism for Nanowires. Advanced Materials 21, 153–165 (2009).

7. Willner, I. & Katz, E. Controlling Chemical Reactivity at Solid-Solution Interfaces by Means of Hydrophobic Magnetic Nanoparticles. Langmuir 22, 1409–1419 (2006).

8. Gittleman, C. S., Kongkanand, A., Masten, D. & Gu, W. Materials research and development focus areas for low cost automotive proton-exchange membrane fuel cells. Current Opinion in Electrochemistry 18. Energy Storage Energy Transformation, 81–89 (2019).

9. Abbasi, R. et al. A Roadmap to Low-Cost Hydrogen with Hydroxide Exchange Membrane Electrolyzers. Advanced Materials 31, 1805876 (2019).

10. Srinivasan, S. G., Baskes, M. I. & Wagner, G. J. Atomistic simulations of shock induced microstructural evolution and spallation in single crystal nickel. Journal of Applied Physics 101, 043504. eprint: https://doi.org/10.1063/1.2423084 https://doi.org/10.1063/1.2423084 (2007).

11. Liu, S. et al. Atomistic simulation of microstructure evolution of NiTi single crystals in bending deformation. Computational Materials Science 199, 110733. ISSN: 0927-025621004602 (2021).

12. Simon, M. & Meyer, E. L. Detection and analysis of hot-spot formation in solar cells. Solar Energy Materials and Solar Cells 94, 106–113. ISSN: 0927-0248. https://www.sciencedirect.com/science/article/pii/S0927024809003420 (2010).

13. Tokmakoff, A., Fayer, M. D. & Dlott, D. D. Chemical reaction initiation and hot-spot formation in shocked energetic molecular materials. The Journal of Physical Chemistry 97, 1901–1913. eprint: https://doi.org/10.1021/j100111a031 https://doi.org/10.1021/j100111a031 (1993).

14. Budevski, E., Staikov, G. & Lorenz, W. Electrocristallization: Nucleation and growth phenomena. Electrochimica Acta 45, 2559–2574. ISSN: 0013-4686. https://www.sciencedirect.com/science/article/pii/S0013468600003534 (2000).
15. Zhang, R., Khalizov, A., Wang, L., Hu, M. & Xu, W. Nucleation and Growth of Nanoparticles in the Atmosphere. *Chemical Reviews* **112**, PMID: 22044487, 1957–2011. eprint: https://doi.org/10.1021/cr2001756, https://doi.org/10.1021/cr2001756 (2012).

16. Bernstein, N. *et al.* Quantifying Chemical Structure and Machine-Learned Atomic Energies in Amorphous and Liquid Silicon. *Angewandte Chemie International Edition* **58**, 7057–7061 (2019).

17. Stukowski, A. Structure identification methods for atomistic simulations of crystalline materials. *Model Simul. Mat. Sci. Eng.* **20**, 045021 (2012).

18. Steinhardt, P. J. & Chaudhari, P. Point and line defects in glasses. *Philos. Mag. A* **44**, 1375–1381 (1981).

19. Ackland, G. J. & Jones, A. P. Applications of local crystal structure measures in experiment and simulation. *Phys. Rev. B* **73**, 054104 (2006).

20. Sharp, T. A. *et al.* Machine learning determination of atomic dynamics at grain boundaries. *Proceedings of the National Academy of Sciences* **115**, 10943–10947 (2018).

21. Rosenbrock, C. W., Homer, E. R., Csányi, G. & Hart, G. L. W. Discovering the building blocks of atomic systems using machine learning: application to grain boundaries. *npj Computational Materials* **3**, 29 (Aug. 2017).

22. De, S., Bartók, A. P., Csányi, G. & Ceriotti, M. Comparing molecules and solids across structural and alchemical space. *Phys. Chem. Chem. Phys.* **18**, 13754–13769 (2016).

23. Behler, J. Atom-centered symmetry functions for constructing high-dimensional neural network potentials. *J. Chem. Phys.* **134**, 074106 (2011).

24. Shapeev, A. V. Moment Tensor Potentials: A Class of Systematically Improvable Interatomic Potentials. *Multiscale Modeling & Simulation* **14**, 1153–1173 (2016).

25. Chapman, J., Batra, R. & Ramprasad, R. Machine learning models for the prediction of energy, forces, and stresses for Platinum. *Computational Materials Science* **174**, 109483. ISSN: 0927-0256 (2020).

26. Park, C. W. *et al.* Accurate and scalable graph neural network force field and molecular dynamics with direct force architecture. *npj Computational Materials* **7**, 73 (May 2021).

27. Chan, H., Cherukara, M., Loeffler, T. D., Narayanan, B. & Sankaranarayanan, S. K. R. S. Machine learning enabled autonomous microstructural characterization in 3D samples. *npj Computational Materials* **6**, 1 (Jan. 2020).

28. Steinbach, I. Phase-field models in materials science. *Modelling and Simulation in Materials Science and Engineering* **17**, 073001. https://doi.org/10.1088/0965-0393/17/7/073001 (July 2009).

29. Richmond, B. G. *et al.* Finite element analysis in functional morphology. *The Anatomical Record Part A: Discoveries in Molecular, Cellular, and Evolutionary Biology* **283A**, 259–274 (2005).

30. A temperature-dependent atomistic-informed phase-field model to study dendritic growth. *Journal of Crystal Growth* **579**, 126461. ISSN: 0022-0248 (2022).

31. Emmerich, H. Advances of and by phase-field modelling in condensed-matter physics. *Advances in Physics* **57**, 1–87. eprint: https://doi.org/10.1080/00018730701822522, https://doi.org/10.1080/00018730701822522 (2008).

32. Plimpton, S. Fast Parallel Algorithms for Short-Range Molecular Dynamics. *Journal of Computational Physics* **117**, 1–19. ISSN: 0021-9991. https://www.sciencedirect.com/science/article/pii/S002199918571039X (1995).

33. Zhou, X. *et al.* Atomic scale structure of sputtered metal multilayers. *Acta Materialia* **49**, 4005–4015. ISSN: 1359-6454 (2001).

34. Larsen, A. H. *et al.* The atomic simulation environment—a Python library for working with atoms. *Journal of Physics: Condensed Matter* **29**, 273002. http://stacks.iop.org/0953-8984/29/i=27/a=273002 (2017).
35. Paszke, A. et al. in Advances in Neural Information Processing Systems 32 (eds Wallach, H. et al.) 8024–8035 (Curran Associates, Inc., 2019).
36. Klicpera, J., Gross, J. & Gunnemann, S. Directional message passing for molecular graphs. arXiv preprint arXiv:2003.03123 (2020).
37. Bresson, X. & Laurent, T. Residual gated graph convnets. arXiv preprint arXiv:1711.07553 (2017).
38. Dwivedi, V. P., Joshi, C. K., Laurent, T., Bengio, Y. & Bresson, X. Benchmarking graph neural networks. arXiv preprint arXiv:2003.00982 (2020).
39. Elfwing, S., Uchibe, E. & Doya, K. Sigmoid-weighted linear units for neural network function approximation in reinforcement learning. Neural Networks 107. Special issue on deep reinforcement learning, 3–11. ISSN: 0893-6080 (2018).
40. Xu, J., Sun, X., Zhang, Z., Zhao, G. & Lin, J. Understanding and Improving Layer Normalization in Advances in Neural Information Processing Systems (eds Wallach, H. et al.) 32 (Curran Associates, Inc., 2019). https://proceedings.neurips.cc/paper/2019/file/2f4fe03d77724a7217006e5d16728874-Paper.pdf.
41. Hirel, P. Atomsk: A tool for manipulating and converting atomic data files. Computer Physics Communications 197, 212–219. ISSN: 0010-4655 (2015).
42. Chapman, J., Goldman, N. & Wood, B. C. Efficient and universal characterization of atomic structures through a topological graph order parameter. npj Computational Materials 8, 37. ISSN: 2057-3960. https://doi.org/10.1038/s41524-022-00717-7 (Mar. 2022).
43. Pozdnyakov, S. N., Zhang, L., Ortner, C., Csányi, G. & Ceriotti, M. Local invertibility and sensitivity of atomic structure-feature mappings. Preprint at http://arxiv.org/abs/2109.11440 (2021).
44. Sullivan, C. B. & Kaszynski, A. PyVista: 3D plotting and mesh analysis through a streamlined interface for the Visualization Toolkit (VTK). Journal of Open Source Software 4, 1450. https://doi.org/10.21105/joss.01450 (May 2019).