Bhardwaj, Utkarsh; Sand, Andrea E.; Warrier, Manoj

Graph theory based approach to characterize self interstitial defect morphology

Published in:
Computational Materials Science

DOI:
10.1016/j.commatsci.2021.110474

Published: 01/07/2021

Document Version
Peer-reviewed accepted author manuscript, also known as Final accepted manuscript or Post-print

Published under the following license:
CC BY-NC-ND

Please cite the original version:
Graph Theory Based Approach to Characterize Self Interstitial Defect Morphology

Utkarsh Bhardwaj\textsuperscript{a}, Andrea E. Sand\textsuperscript{b}, Manoj Warrier\textsuperscript{a,c}

\textsuperscript{a}Computational Analysis Division, BARC, Vizag, AP, India-530 012
\textsuperscript{b}Department of Physics, P.O. Box 43, FI-00014 University of Helsinki, Finland
\textsuperscript{c}Homi Bhabha National Institute, Anushaktinagar, Mumbai, Maharashtra, India - 400 094

Abstract

The defect morphology is an essential aspect of the evolution of crystal microstructure and its response to stress. While reliable and efficient standard computational algorithms exist for finding defect concentration and size distribution in a crystal, defect morphology identification is still nascent. The need for an efficient and comprehensive algorithm to study defects is becoming more evident with the increase in the amount of simulation data and improvements in data-driven algorithms.

We present a method to characterize a defect’s morphology precisely by reducing the problem into graph theoretical concepts of finding connected components and cycles. The algorithm can identify the different homogenous components within a defect cluster having mixed morphology. We apply the method to classify morphologies of over a thousand point defect clusters formed in high energy W collision cascades. We highlight our method’s comparative advantage for its completeness, computational speed, and quantitative details.

Keywords: Defect morphology, Defects in crystal, Collision cascades, Radiation damage, Molecular dynamics, Graph applications

2000 MSC: 68U06, 82D08, 05C90, 68W06, 05C90

1. Introduction

Defects in materials can have different morphologies and sizes. High stress conditions like collision cascades caused by high energy irradiation can result
in formation of peculiar defect morphologies such as rings in Fe [1] and W [2] or \(\langle 100 \rangle\) dislocations in W [3] in addition to ground state \(\langle 111 \rangle\) configuration. The morphology of a defect decides its thermal stability, migration properties, and interactions with other defects [4, 5, 6, 7, 8]. These properties govern the microstructural evolution of the crystal and its mechanical and physical properties. For this reason, the morphology of defects has been of wide interest in the study of the effects of irradiation on materials and designing materials with desired properties.

With continuous improvement in computational power, the simulation data for defects in crystals has also increased. It is now possible to carry out Molecular Dynamics (MD) simulations of high energy collision cascades with more statistics. While the defect concentration and size distribution can be analyzed reliably and efficiently, it is not yet the case with defect morphology identification. For this reason, the defect morphology studies have been mostly limited to either particular defects of interest [9,10], or qualitative observations [11,3]. For small datasets of clusters, it is possible to inspect the defect morphology visually. However, the qualitative and visual assessment becomes intractable and unreliable for bigger databases. The promising applications of data-driven methods in various scientific fields have garnered broad interest to create databases of atomistic simulations of materials such as DefectDB [11,12] and CascadesDB [12,13,14]. An efficient automatic tool that can reliably extract defect concentration and defect morphologies accompanied by an extensive database like CascadesDB can open avenues for new data-driven algorithms to find new patterns and trends.

The traditional methods for finding point defects, such as Wigner-Seitz (W-S) [15,16] and effective-sphere (ES) for finding displaced atoms [1,17], do not differentiate the point defect clusters based on morphology. Dislocation loops have been identified in several ways, with the DXA algorithm [12] lately establishing itself as a useful tool in the field. However, it does not specify the morphology of non-dislocation defects and small point defect clusters. Moreover, it gets memory intensive and slow as the system size grows. The traditional
geometrical methods, such as common neighbor analysis, centrosymmetric parameter, etc., indicate defective regions in the crystal but fail to describe the defect morphology and concentration. There are newer geometrical feature vectors designed with a focus on supervised learning to identify defects from lattice atoms and visualize them [18] [19]. The accuracy of machine learning that uses the traditional, hand-crafted approach to feature designing is limited by the relevant information captured in the feature vector. The advantage of supervised learning for the tasks of defect identification and characterization is not yet clear, especially when the crystal structure and defect morphology have well-defined descriptions amenable to efficient deterministic methods. However, machine learning explorations can guide the way for establishing deterministic definitions. An unsupervised learning method [2] used for defect classification performs well in differentiating edge dislocations, C15 Laves phase like rings [10] [20], etc. The C15 Laves phase ring defect is a 3D ring-like geometrical configuration of atoms with distinct properties such as high thermal stability and sessile nature that can serve as a nucleation site. The unsupervised algorithm outlines all the possible defect shapes observed in Fe and W cascades. However, it fails to distinguish mixed clusters and provide details of morphologies such as orientations of dislocation loops and dumbbells constituting a defect.

Several studies indicate the presence of complex morphology in radiation-induced defects. An early MD study of clusters formed in energetic cascades in W by Sand et al. [3] uses W-S and individual inspection guided by potential energy analysis to report dislocation loops. The study also reports some of the clusters having complex configurations that often have partial parallel oriented dumbbells. A subsequent MD study of radiation damage in W across different temperatures and energy ranges by Wahyu Setyawan et al. [21] presents a detailed categorization of all SIA clusters using a combination of W-S, ES, and SIA dumbbell/crowdion orientations. However, it groups the clusters of less than size 30 and significant sized clusters that do not form a dislocation loop under the ambiguous category of 3D clusters, which is subcategorized based on the constituent SIA orientations. A study on the effect of the repulsive part
of interatomic potential on cascade damage in Fe\cite{22} uses similar methods. Besides, it labels non-dislocation clusters with dumbbells in specific orientations as C15-like, which are of much interest due to their sessile nature and high stability in Fe. The method used to label C15 is verified by visual inspection of positive cases found with the technique. Since the structure of the primary radiation damage has a significant impact on the mobility, interaction, and annealing behavior of the defects, longer time scale predictions of the microstructural evolution would benefit from a more detailed description of each defect. Furthermore, a recent study by Mason et al.\cite{23} reports the formation of intermediate complex configurations, as initially separate defects meet and merge during the process of annealing. The identification of the defect morphology at these stages of annealing can potentially provide insights and aid predictions of the time scale and final product of the transformation.

We present a method to define the morphology of a cluster based on its homogeneous constituent components. A separate component in a cluster is defined as composed of SIA dumbbells and crowdions that all hold a specific relationship with their neighbors. We represent a defect as a computational graph, with SIA dumbbells/crowdions as nodes and the relationship between them as edges. The problem of identifying the morphology is reduced to finding connected components and cycles in the graph. The method gives structural details of defects such as the orientation and magnitude of the Burgers vector for an edge dislocation, number and proportions of different morphologies in a multi-component defect, and degree of disorientation and extent of individual SIAs that constitute a defect. We explore the defect morphologies of over a thousand SIA clusters formed in high energy collision cascades in W. We ascertain the presence of 3D rings corresponding to the C15 Laves phase geometry in W and show new insights about other morphologies.
2. Methods

2.1. Molecular Dynamics Simulation Dataset

We apply our defect morphology identification method to the defects formed in MD simulations of high energy collision cascades in W. The simulations were carried out at an initial temperature of 0 Kelvin and evolved for 40ps. Electronic stopping was applied to atoms with energies above 10 eV \[24\]. A PKA was selected from the lattice atoms of the cubic simulation cell. The desired kinetic energy was given in a random initial direction. Periodic boundaries were used for each cascade. It was ensured that simulation cell size and PKA chosen were such that no atom reaches the boundaries. Temperature control at 0K was applied to all atoms within three atomic layers from the cell borders using a Berendsen thermostat \[25\]. MD simulations of high energy collision cascades with the Derlet potential \[26\] stiffened by Bjorkas \[27\] were carried out.

2.2. Overview of the developed method

The initial inputs to the algorithm developed for morphology identification are the atomic coordinates of point defects in the cluster and their nearest lattice points. In this work, an efficient implementation of W-S and ES based method \[2\] is used to find the point defect clusters from the given coordinates of the atoms. The point defects identification algorithm associates each atomic coordinate to its nearest lattice site using modular arithmetic. If a lattice site has more than one atom associated, all the associated atomic coordinates along with the lattice site are added to the defects list while labeling the closest atom and lattice site as annihilated. The annihilated atoms are not counted in defect count or cluster size. However, these are useful in defining morphology. More displaced atoms and their closest lattice sites are then added to the defects list as annihilated pairs using the ES method with a cut-off threshold of approximately 0.3 times the lattice constant that can vary slightly with simulation temperature \[1, 28\].

Figure 1 summarizes the steps in our method consisting of (i) defining lines along the displaced atoms and associated lattice points, (ii) merging coinciding
lines, and (iii) finding connected components in the graph representation of a defect.

![Figure 1](image)

**Figure 1**: The figure schematically summarises the output for a cluster after each step in our methodology. The final algorithm outputs the constituent morphologies of the components forming the input defect. (a) Initial inputs from the defect identification: coordinates of the displaced atoms and their nearest lattice site. (b) Define lines by (i) joining the pair of atoms that occupy the same lattice site, (ii) displaced atom and lattice site if only a single atom occupies the lattice site. (c) Merge the coinciding lines. (d) Find structurally homogeneous components that constitute the defect based on angle and distance relationship between the neighboring lines.

We first find interatomic line equations for each string of displaced atoms that form a dumbbell/crowdion. These lines correspond to the nodes in the graph representation of a defect. Whether to join two nodes in the graph with an edge is predicated upon rules defined separately for parallel and non-parallel ring-like configurations. For a parallel component, a connecting edge between two neighboring nodes is added if the lines they represent are parallel. In contrast, for a ring, an edge is added if the angle between 1NN neighboring interatomic lines is approximately 60 or 90 degrees. The connected components algorithm from graph theory is used [29] to find distinct homogeneous components. To further verify ring morphology, the graph representation must also exhibit cycles of length three or more. This verification step adds robustness to the algorithm against errors that can arise due to thermal vibrations. A parallel component forms an edge dislocation by introducing an extra plane of atoms in the crystal. The direction of the Burgers vector is the same as the direction of the lines formed of the strings of displaced atoms, and the number of extra atoms in lines decides the magnitude of the Burgers vector. The morphologies that are neither parallel nor form rings are transient, non-specific configurations. These are meta-stable configurations that quickly transition to a stable parallel
or ring form when relaxed at around room temperature. A defect can be composed of multiple components, such as a combination of multiple dislocations or a glissile dislocation trapped with a sessile ring.

The computational cost of our method to define morphology given a defect is limited by the operation of finding neighboring dumbbells/crowdions to define lines by merging all those that are collinear. The operation can be performed by using kd-tree [30] to achieve \( O(n \log n) \) computational complexity where \( n \) is number of dumbbells/crowdions in a defect.

2.3. Defining and Merging Lines

1. Define parametric line equations for lattice sites associated with a pair of atoms as the line passing through the two atoms. The lattice site itself may not lie on the interatomic line. For the lattice site associated with only a single displaced atom, the line is defined as passing through the lattice site and the displaced atom (Fig. 2(b)).

2. Find and merge neighboring coincident interatomic lines and collinear points (Figure 1 (c)). This operation can employ k-d tree data-structure [30] to look for neighboring lines efficiently.

In Figure 1 (c), a single line \( L_1 \) is defined by merging the coincident lines. As shown by \( L_2 \) and \( L_3 \) lines in Figure 1 (c), for some crowdions, the lattice points can be seen as falling in a separate sub-line, possibly due to local stresses caused by nearby non-parallel structures or other clusters. When visualized, the lines help in the qualitative assessment of different structures such as parallel crowdions, hexagonal rings, 3D-rings, and mixed clusters.

We use parametric equations of lines for efficiently calculating different properties such as the shortest distance between two lines, angle between two lines and orientation of a line. For the parametric equation of a line passing through two points \( a \) and \( b \), we define a unit direction vector \( \vec{v} = \vec{a} - \vec{b}/\|\vec{a} - \vec{b}\| \). The equations for angle between two lines (\( \theta \)) and their shortest distance (\( d \)) can be defined as:
\[
\theta = \arccos \langle \vec{v}_1, \vec{v}_2 \rangle 
\]

\[
d = \frac{\langle \vec{a}_1 - \vec{a}_2, \vec{v}_1 \times \vec{v}_2 \rangle}{\| \vec{v}_1 \times \vec{v}_2 \|} 
\]

\(\vec{v}_1\) and \(\vec{v}_2\) are direction vectors of the lines and \(\vec{a}_1\) and \(\vec{a}_2\) are the position vectors of any two points in the lines. The second term in the numerator, \(\vec{v}_1 \times \vec{v}_2\) gives a vector orthogonal to both the lines. The dot product of the normalized value of this orthogonal vector with the vector defined by two points on lines \(\vec{a}_1\) and \(\vec{a}_2\) gives the projection of the latter on the former. This perpendicular distance is also the shortest distance between the lines.

For coincident lines, \(\theta\) and \(d\) both should be close to zero. We will use these metrics for graph representation also. The threshold for angle and distance that we chose is 20 degrees and 1.0 Å, respectively. The results are not very sensitive to the values of the thresholds.

The unit direction vector also represents the orientation of the line. We also add extra attributes like defect count for each interatomic line, their extent, deviation from the standard orientations, and offset of lattice sites that do not lie on the interatomic line. The defect counts for the constituent interatomic lines in an edge dislocation indicate the magnitude of the Burgers vector for the dislocation.

### 2.4. Graph Representation

After merging the coinciding interatomic lines, we construct the adjacency matrix \(A\), which is an \(n \times n\) sized matrix for a cluster with \(n\) number of lines. Each value \(a_{ij}\) of the matrix is either 1 if the \(i^{th}\) and \(j^{th}\) lines are marked as connected or 0 if they are not connected. The connectivity rules are defined differently for bundles of parallel dumbbells and planar rings or C15-like 3D-rings.
2.4.1. Edge Predicate for Parallel component

For components consisting of parallel dumbbells/crowdions the adjacency matrix values are defined by the following relation:

\[ a_{i,j} = \begin{cases} 
1, & \text{if } \theta \approx 0 \text{ and } d \leq 1NN \\
0, & \text{otherwise} 
\end{cases} \]  

(3)

\( \theta \) is the angle between the interatomic lines, and \( d \) is the shortest distance between the two interatomic lines found using the Equation (2).

2.4.2. Edge Predicate for Ring component

A 3D-ring or C15-like structure can be defined as a composite of di-interstitial tripod structure and tri-interstitial hexagonal shape [10, 20] (Figure 2 (a)). Both of these structures are stable by themselves and also occur as a stand-alone sessile cluster. To define the adjacency rules, we require the structural details of these two basic shapes.

Figure 2: The figure shows a schematic representation of basic cluster shapes and the effect of thermal vibrations on their different possible appearances. Green circles represent interstitials, and red hollow circles represent lattice points. (a) shows the tripod/triangle and hexagon as a basis that forms the 3D ring shape. The triangular di-interstitial appears in four forms depending on how many extra pairs of displaced atoms and corresponding lattice sites (labeled with i, ii, iii) appear. The extra pairs may or may not appear as displaced depending on the thermal vibration at a simulation frame. If all the three extra pairs appear displaced, then the defect appears as a tripod with a vacancy at the center. The non-planar tripod appears planar in the schematic. (b) shows that one of the dumbbells in a size two parallel cluster may appear either as a crowdion or dumbbell, depending on the atomic vibration.

The hexagonal structure is a planar arrangement of three dumbbells whose lattice sites are not collinear. The interatomic lines drawn by joining the dumbbell atoms appear as alternate sides of a hexagon, forming a sixty-degree angle with each other. All the three lines are oriented in \( \langle 110 \rangle \) direction, and the
lattice points of lines are 3NN distance apart. The clusters that have this arrangement also have at least a single crowdion/dumbbell that is orthogonal to the plane of the hexagon and is oriented along \(\langle111\rangle\) direction.

The di-interstitial tripod, when appearing as a stand-alone cluster, can have four different forms depending on how many displaced atoms appear (Figure 2 (a)). If we look at the same defect at different time frames in an MD simulation carried out at a non-zero temperature, the thermal disturbances can vibrate and cause differences in how many displaced atoms appear. Since we apply our analysis to a single instance of simulation with one set of atomic coordinates, we will find different defects of this morphology in different forms. We need to define a relationship that identifies all these apparent forms as the same morphology. A planar symmetrical tripod structure formed of three dumbbells will have sixty degrees angle between the dumbbells. However, the actual defect is slightly non-planar, and the angle between the dumbbells is observed to be ninety degrees. When a pair of displaced atom and lattice site is not present in the tripod, making it asymmetric, the angle subtended is then observed to be around sixty degrees. The following relation defines the values for the adjacency matrix of rings.

\[
a_{i,j} = \begin{cases} 
1, & \text{if } \theta \approx 60 \text{ or } 90 \text{ and } d' = 1\text{NN or } 3\text{NN} \\
0, & \text{otherwise}
\end{cases}
\]  

(4)

\(\theta\) is the angle between the lines, and \(d'\) is the shortest distance between any lattice points associated with the displaced atoms in the line. Both 60° or 90° and 1NN or 3NN are valid values for different appearances of the di-interstitial tripod-like arrangement, while for tri-interstitial hexagon, 3NN and 60° suffice.

2.5. Finding the components

Using the connected components algorithm, we group all the nodes (representing SIA interatomic lines) that are connected by the same type of edges (defined in Equation (3) and Equation (4)). A parallel component forms an edge dislocation with the Burgers vector having direction along the constituent
interatomic lines and magnitude determined by the number of extra atoms in the lines.

New specific configurations other than parallel and rings can be identified by defining rules based on the distance and angle relationship. The orientation of dumbbells/crowdions can also be used if required. When we apply this method on W collision cascades, we find one specific configuration, which is neither ring nor parallel. It consists of an orthogonal pair of 2NN separated dumbbells (shown in Figure 2(e)).

After identifying specific configurations based on rules such as parallel and ring components, a small fraction of small components having no specific configuration may be found. These are primarily non-recurring transient configurations that may appear different from specific configurations due to thermal vibrations. All the non-specific configurations can be grouped into one single category. In W, they mostly change to a more stable glissile form.

2.6. Check for cycles in ring components

For a ring structure, in addition to the binary relation defined by the Equation (4), it is also essential that the three interatomic lines in a hexagon, tripod, or their 3D composite mutually hold the same relationship. More specifically, if \( a_{ij} = 1 \) and \( a_{jk} = 1 \), \( i, j, k \) will be grouped together into one component, however for a ring it is also essential that \( a_{ik} = 1 \) otherwise it is a random non-parallel arrangement of dumbbells and not a ring. In the graph representation the edges \( ij, jk \) and \( ki \) form a cyclic path connecting the three vertices \( i, j, k \). Such a cycle in a graph is known as a triangle graph, or C3 cycle [31]. After finding the connected components, we look for these cycles of three to verify a ring-like structure. There are also bigger cycles in a 3D-ring, and the biggest cyclic path will traverse the complete ring shape. Another way of differentiating between planar and 3D rings is that in a 3D-ring, there are multiple C3 cycles, while in a planar ring (tripod or hexagonal ring), there is only a single cycle as there are only three interatomic lines that hold the relationship defined in Equation (4). The probable ring components that do not form such cycles
are classified as unordered transient configurations that are neither parallel nor
ring. This condition takes care of spurious thermal noise that may appear at
higher temperatures.

3. Results

The database contains 139 high energy collision cascades simulated with the
Derlet potential [26] stiffened by Bjorkas [27] for bcc W. The cascades contain
1170 clusters of different sizes and morphologies.

The defect morphology identification takes approximately a minute to pro-
cess the database on a regular desktop computer once defects and clusters have
been identified. The defect and cluster identification using an efficient imple-
mentation of W-S and ES [2] take less than ten minutes for the whole database.
In comparison, the dislocation analysis with DXA algorithm [32] as currently
implemented in Ovito [33] takes more than two minutes to process a single 150
keV cascade on the same desktop computer, while a 200 keV cascade having a
simulation size of 190 unit cells could not be processed due to insufficient system
memory.

3.1. Defect Morphologies in W collision cascades

The morphology of SIA defects in the database of W collision cascades is
found to be composed of parallel bundles of SIA strings forming edge dislo-
cations, 3D rings corresponding to C15 structure, their planar basis shapes
(triangle/tripod and hexagon [10, 20, 2]), 2NN separated orthogonal pair of
dumbbells and a few transient unordered configurations. Figure 3 shows typical
defects for each morphology.

The ⟨111⟩ dislocations are known to be most stable configuration. The
dislocations having ⟨100⟩ orientation are observed to have a few ⟨111⟩ crowdions
on the fringes. The ratio of ⟨100⟩ dumbbells and ⟨111⟩ dumbbells/crowdions is
very likely to affect the defect behavior, such as thermal stability and energy
required for dislocation glide. The direction of dumbbells in 3D rings is mostly
⟨110⟩, while for their planar basis, it can be anywhere between ⟨110⟩ and ⟨111⟩.
Figure 3: Different defect morphology in W collision cascades with the symbols that we use to represent them. The ASCII symbols are chosen such that they indicate the morphology. (a) parallel bundle of ⟨111⟩ directed SIA that form 1/2⟨111⟩ edge dislocations, (b) defects composed of multiple edge dislocations, (c) parallel group of ⟨100⟩ directed SIA that form ⟨100⟩ edge dislocations, (d) 3D rings and their planar ring basis, (e) meta-stable defects with no particular order, (f) defects with rings and edge-dislocations. The first two defects in (d) show planar rings that are basic shapes for 3D rings. Lines are drawn along the dumbbells/crowdions and are colored according to their orientation.

The orientation and angles of the 3D rings correspond to the C15 Laves phase geometry. The rings may have a few ⟨111⟩ crowdions extending as tails. Slight disorientation from ⟨110⟩ direction can also occur if the ring is adjoined with an edge dislocation forming a composite cluster. The planar ring with tripod or triangle di-interstitial form can also compose into structures other than the 3D C15 like rings, such as two of the tripods stacked one after the other. Another specific configuration observed is composed of two orthogonal dumbbells that are 2NN distance apart (the first defect in Figure 3(e)). Since this specific configuration of size two does not form any bigger morphology and the defects of this kind are relatively less in number, we group the unordered arrangements and this specific morphology in a single category for statistics and plots. Table 1
lists the different morphologies along with the symbol we use to represent them. The symbols are chosen such that they give an idea of the morphology.

Table 1: Description and ASCII symbol for different cluster morphologies.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>@</td>
<td>Rings, (C15-like or its basis shapes)</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>#</td>
<td>Non-parallel &amp; non-ring configurations</td>
</tr>
</tbody>
</table>

The composite morphologies (Figure 3(b) and (f)) render different behavior from the individual constituent components, e.g., the movement of an otherwise glissile dislocation can be restricted by being trapped with a ring (||@) or in conjunction with another dislocation (||//). Figure 4 shows some cases where dislocation analysis carried with DXA algorithm [32] can result in the omission of such distinctions in composite clusters.

DXA is an algorithm to find dislocations of significant sizes. The dislocations found by DXA are always identified with our method (Figure 4(a, b, c)). However, the morphological details such as presence of (111) crowdions around the (100) dislocation loop in (a) can not be obtained with dislocation analysis alone. Such a detail may affect a defect’s properties like thermal stability and diffusion. Other example of such details include length/extent of crowdions, their orientation, and relationship with neighbors. Moreover, The DXA algorithm can sometimes ignore the smaller defect components of around size ten or less. For composite clusters such as Figure 4(d) and (e), the omission of a defect component may incorrectly suggest that the defect is a glissile dislocation. The DXA algorithm creates a mesh or blob around the smaller identified defects and other dislocations (Figure 4(f) and (g)). In contrast, our algorithm clearly distinguishes between ring morphology, smaller point defect clusters forming a parallel bundle of dumbbells as dislocations, and other specific configurations such as orthogonal pair of dumbbells. It can also be extended to identify any
configuration composed of dumbbells and crowdions arranged in a specific order.

Figure 4: Comparison of dislocation loops found using DXA dislocation analysis as currently implemented in Ovito. The bigger defect plots with loops and blobs show DXA results, while the insets show plots for the corresponding defects analyzed using our method. The DXA algorithm draws loops around the dislocations and shows the non-dislocation as blobs. The color of the atoms in the DXA plot represents whether an atom is embedded in the bcc neighborhood (blue color if it is, white if it is not) using structural analysis. In addition to the dislocations found by the DXA algorithm, our method identifies smaller dislocations and non-dislocation morphologies. The first row shows example cases where DXA and our method agree. The DXA dislocation loops in (d) and (e) only show a single \langle 111 \rangle loop and miss the other smaller components that include both edge dislocations and hexagonal ring. In (g) the non-dislocation part is shown as a blob connected with the dislocation while (f) shows an example where top part is shown as blob while lower \langle 111 \rangle loop is omitted from DXA analysis. Our method shows the morphology of the non-dislocation part, too.

3.2. Statistical distribution of defect morphology

We use the developed method to identify the basic morphology for each of the 1170 point defect clusters in the dataset. Figure 5 shows the comparison of defect concentration over different energies and the distribution of the defects in different morphologies listed in Table 1. The majority of defects arrange in parallel bundles of dumbbells and crowdions, especially in smaller, glissile \langle 111 \rangle orientations and big mixed, sessile composites of \langle 111 \rangle - \langle 100 \rangle multi-component
dislocation loops. Figure 5 (b) shows the fraction of defects forming a parallel bundle of dumbbells/crowdions (black inverted triangle), which includes $\langle 111 \rangle$ oriented bundles (blue circles), $\langle 100 \rangle$ bundles (green triangle), and combinations of $\langle 111 \rangle - \langle 100 \rangle$ (orange square). The number of the multi-component edge dislocations and their sizes (right inset in Figure 5 (c)) increase with energy. The number of distinct loops in a single multi-component dislocation goes up to six (left inset in Figure 5 (c)).

Figure 5: Statistics of defects formed in W collision cascades. The ASCII symbols on axis labels and legend (||, @ etc.) correspond to the morphologies as shown in Figure 3 and Table 1. (a) Shows an average number of point defects per cascade and fraction of point defects in a cluster. (b) Shows the relative distribution of in-cluster defects among morphologies across PKA energies. The second plot shows the average number of times a defect morphology appears in a cascade. (c) Shows the size distribution of each morphology. The size for multi-component loops (||/||/) shows a definite increasing trend with energy and is shown separately on the right inset. Other morphologies do not show a clear trend with energy. The number of components and their sizes are shown on the left.

The fraction of point defects in non-parallel defect morphologies decreases
with energy while the number of defects per cascade increase (Figure 5 (b)),
implying a decrease in the sizes of these defects. Compared to dislocations,
these defects are fewer; however, their role in the evolution of a cascade might
be significant depending on their thermal stability and interactions. Table 2
lists the distribution of the absolute number of defects and number of point
defect clusters into different morphologies for each energy.

Table 2: Distribution of absolute number of point defects in each morphology and the number
of clusters found for each morphology. $E_{pka}$ is in keV.

| $E_{pka}$ | cascades | kind     |   ||   |   ||   | @ | #   | total |
|-----------|----------|----------|---|---|---|---|---|---|---|---|
| 50        | 70       | defects  | 1457 | 46 | 436 | 90 | 85 | 76 | 381 |
|           |          | clusters | 292  | 4  | 20  | 9  | 26 | 30 | 381 |
| 100       | 20       | defects  | 1223 | 86 | 622 | 22 | 43 | 49 | 2045 |
|           |          | clusters | 126  | 2  | 15  | 3  | 11 | 13 | 170 |
| 150       | 40       | defects  | 4352 | 201| 2895| 78 | 63 | 106| 7700 |
|           |          | clusters | 381  | 8  | 35  | 9  | 22 | 36 | 491 |
| 200       | 19       | defects  | 859  | 12 | 1627| 17 | 36 | 16 | 2567 |
|           |          | clusters | 94   | 1  | 12  | 1  | 14 | 6  | 128 |

3.3. Exploring internal morphological details

The method gives insights into internal details of defect morphologies. For
example, it has been postulated that the strings of displaced atoms in the central
part of a 1/2⟨111⟩ dislocation are more extended, forming longer crowdions
than the strings on the surface[34]. In a bundle of parallel crowdions forming
a dislocation, some of the crowdions are at the boundary or interface, while
others are surrounded by neighboring crowdions. The number of crowdions
within 1NN is a good indicator of whether a crowdion is on the cluster’s surface
or towards the center in bulk. For a ⟨111⟩ parallel cluster, the maximum number
of neighbors for a central crowdion is six, while for ⟨100⟩, this value is four.

We find only dumbbells (SIA lines with only two displaced atoms) in the
⟨110⟩ direction. The extents of displaced atoms in ⟨100⟩ rarely go beyond three
atoms, while in ⟨111⟩, longer strings of displaced atoms form (see Figure 6).
For ⟨111⟩ oriented dislocations, there are long strings of displaced atoms in the
central SIA lines. The correlation value between the extent of an SIA line and
Figure 6: The number of atoms in an SIA interatomic line as a function of the number of interatomic lines within 1NN distance. An SIA interatomic line may have only two atoms forming a dumbbell if the extent of the displacement is limited, or it may have more than two atoms forming a crowdion if the displacement along that orientation is extended. The number of 1NN neighbors indicates whether an SIA line is in the central part of the cluster or on the surface. The plot shows that the extent of an SIA line in the central part having six neighbors is longer and is almost always more than the base value of two. For ⟨100⟩ SIAs, the central SIA lines with four neighbors become longer on average, but a good fraction has only two atoms. The extent of SIA lines in ⟨100⟩ rarely goes beyond three atoms, while ⟨111⟩ crowdions are longer.

its number of neighbors is 0.78. While for ⟨110⟩, the correlation value is just 0.44.

The details used in this analysis, such as the extent or length of crowdions, orientation, and the number of neighbors for each crowdion/dumbbell, are not provided by existing methods.

4. Discussion

The method developed in this study identifies in-cascade defects in W, and their different morphologies, such as single dislocations along ⟨111⟩ and ⟨100⟩ orientations, their multi-component composites, 3D rings corresponding to the C15 Laves phase, its planar basis rings, and 2NN separated orthogonal pair of dumbbells. The method also identifies mixed defects composed of dislocations and rings. The statistics on fraction and defect size of different morphologies show that glissile ⟨111⟩ edge dislocations of small size are predominant. However, as the PKA energy increases, the number and size of sessile multi-component dislocations also increase. Although rarer than dislocations, the
sessile rings can play an essential role in the cascade’s thermal evolution by acting as trapping sites for dislocations due to their sessile nature. However, the extent of their role will depend on the morphological stability and nature of interactions with other defects. The distribution of in-cascade defect morphology is an essential metric for validation and comparison of interatomic potentials and different treatments of electronic stopping used in the MD simulations. While recent experiments have validated the presence of ⟨100⟩ dislocations in W [35, 36], in addition to 1/2⟨111⟩ dislocations, the experimental investigation of smaller defect morphologies found in simulations is still intractable with conventional transmission electron microscopy [37]. Higher scale models that use the defect morphology distribution and their properties can help in such validations.

The presented method provides structural details such as the orientation and extent of individual SIA strings and their relationship with neighbors in addition to the morphology of the constituent components of a defect. We use the information to quantify the postulated correlation [34] between the extent of an SIA line and whether it is in the central part of the cluster. The information also shows that the ⟨100⟩ dislocation loops almost always have ⟨111⟩ crowdions and dumbbells on the fringes and the 3D rings in W correspond to the C15-like structure. It gives the number of defects in each component of a multi-component dislocation and a composite cluster with both ring and parallel components.

These morphological details can be used to study the mechanisms of morphological transitions and migration of defects, providing insights on the relationship between structural details, size of the defect, and defect properties like transition energy, recombination radius, and migration energy. These can be input to higher scale models such as Monte Carlo methods, which currently use approximate functions for such essential relationships [38]. These higher-scale predictions of the microstructure can be validated with the current experiments.

The dataset used in this work has a single temperature and uses the same interatomic potential. The distribution of defects over different morphologies
might change depending on the temperature or interatomic potential. The com-
parative study of defect morphologies across different interatomic potentials or
temperatures is out of the scope of the paper. The morphologies can also be
used as inputs to the density functional theory (DFT) studies for checking the
defect’s energetic stability.

The results are presented for the bcc W, however the method is not limited
to bcc lattice. The algorithms are effective for defining defect morphologies of
SIA defects that form dumbbells or crowdions in any crystal. The algorithms
will work for polycrystals as well; however, labeling the orientation of individual
interatomic lines and, in turn, dislocation loops such as $\langle 111 \rangle$ or $\langle 100 \rangle$ will have
to consider the orientation of the particular crystal where the defect is formed.
If a defect is present in the grain boundary or across it then the orientations
can be defined based on global coordinates. Different crystals and materials
might have various other defect morphologies. Identifying a defect morphology
with the presented method requires defining basic rules that are characte-
ristic of that morphology. To identify a new defect morphology, a characteris-

tic property such as specific distance or angle relationship between neighboring
dumbbells/crowdions or their orientations must be found, apriori. This can be
a weakness of the method if unknown and intricate morphologies that are diffi-
cult to define by rules exist in the application domain. However, we find that
the kinds of stable defect morphologies found in crystals have a distinct, definite
structure that can be easily defined and incorporated into our method.

Another limitation of the method arises from the thermal noise inherent in
dynamic simulations at finite temperatures. While a reasonable degree of ther-
mal fluctuations poses no problems to W-S or ES analysis, at the level of sen-
sitivity of the method presented here, these transient displacements may rarely
induce errors, potentially changing the angle of a line relative to its neighbors.
In a large defect, these spurious errors do not affect the overall morphology,
however small defects with noise may appear as random arrangements. This
issue can be trivially circumvented by energy minimization of the defect struc-
tures. For large databases, such a step would entail a high computational cost.
The extra check for cycles in the graph representing a probable ring component increases robustness against misclassification to rings due to spurious thermal noise while also incurring a low computational cost.

5. Conclusions

The approach developed in this work identifies defect morphologies and also resolves constituent uniform morphologies in a multi-component defect. A morphologically homogenous component is composed of SIA dumbbells/crowdions that share a characteristic neighborhood relationship. The reduction of the problem to well-established problems from graph theory viz. connected components and cycles in a graph make the solution efficient and easy to extend to a new morphology by defining its fundamental characteristics as a decision rule for connecting graph nodes. The implementation of the algorithm is fast, making it ideal for use in big datasets.

The method developed is applied to a dataset of 149 high energy MD collision cascades in W consisting of over a thousand SIA defect clusters. The analysis identifies the morphology of every defect. The morphologies found in W include single dislocations along \( \langle 111 \rangle \) and \( \langle 100 \rangle \) orientations, their multi-component composites, 3D rings corresponding to the C15 Laves phase, its planar basis rings, and 2NN separated orthogonal pair of dumbbells. The method also identifies mixed defects composed of dislocations and rings. The method can be easily extended for other materials and new morphologies that show a characteristic structure.

Our method can efficiently find all the dislocations that the DXA dislocation analysis algorithm identifies. Moreover, it distinguishes between various morphologies that are either missed by DXA or marked as non-dislocation, such as the smaller parallel bundle of crowdions/dumbbells, rings and their basis, mixed composite dislocation with rings. We show that the omission of defect components by dislocation analysis can lead to inaccurate prediction of multi-component morphology into single glissile dislocation.
We show the statistical distribution of defect morphologies found in W collision cascades. We also provide size distribution for each of the morphology and for different components of multi-component dislocations. The morphological details that our method provides show that ⟨100⟩ dislocations in W preferably have ⟨111⟩ crowdions/dumbbells around the fringes, and the 3D rings are found to correspond to the C15 structure. The internal details are further used to show the earlier observed correlation between extents of crowdions and their relative position in a dislocation. Any of the existing methods does not provide these details. These details can be used to study morphological transition mechanisms and the relationship between structural details and defect properties.

Data availability

The raw data required to reproduce these findings are partially available to download from the open database https://cascadesdb.org/. The processed data and code required to reproduce these findings are available to download from https://github.com/haptork/csaransh.

References


[4] D. Bacon, F. Gao, Y. Osetsky, The primary damage state in fcc, bcc and hcp metals as seen in molecular dynamics simulations, Journal of Nu-


URL https://db-amdis.org/defectdb

URL https://www-nds.iaea.org/publications/indc/indc-nds-0803

URL https://cascadesdb.iaea.org/


[31] E. W. Weisstein, Triangle graph
URL https://mathworld.wolfram.com/TriangleGraph.html


[33] A. Stukowski, Visualization and analysis of atomistic simulation data with OVITO-the Open Visualization Tool, MSMSE 18 (1) (JAN 2010). doi:10.1088/0965-0393/18/1/015012

[34] S. L. Dudarev, Coherent motion of interstitial defects in a crystalline

