Graph Theory Based Approach to Characterize Self Interstitial Cluster Morphologies

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

\textsuperscript{a}Computational Analysis Division, BARC, Visakhapatnam, Andhra Pradesh, India - 530012
\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

Morphology of self interstitial atom (SIA) clusters formed after a collision cascade is an important aspect of radiation damage. We present a method to characterize the morphology of a cluster by precisely identifying its constituent homogeneous components. The constituent components are identified as parallel bundles of SIAs, rings and other configurations based on the properties of alignment of the SIA lines and their neighborhood relationships. We reduce the problem of decomposition of a cluster into components and characterizing them into graph theory problems of finding connected components and finding cycles in a graph representation of a cluster.

The method is used to study over 1000 clusters formed in W collision cascades for energies ranging from 50 keV to 200 keV. We show the typical cluster shapes for each morphology type identified using the method and compare the structural description with the results from dislocation analysis. The description is found to be in agreement for components with big parallel bundle of SIA. We demonstrate with examples that for other cases such as a mixed cluster, the presented method provides a better description of the structural details. The study gives statistical distribution of different morphologies across energies and their properties such as component sizes and orientations.

keywords: Collision cascades, Radiation damage, Molecular dynamics, Cluster shapes, Defect morphology; 2000 MSC: 68U06, 82D08, 05C90, 68W06, 92E10

1 INTRODUCTION

Particle irradiation, such as by ions or neutrons, gives rise to defects in materials by displacing atoms, potentially affecting multiple atoms in sequences of consecutive collisions, known as collision cascades. In an energetic collision cascade, defect clusters of self-interstitial atoms (SIAs) and vacancies are formed. Vacancy cluster formation is relatively rare in the MD cascade simulations and they show lesser morphological variety than those of SIA clusters \cite{1}. The morphology of a SIA cluster is an important aspect for its diffusion profile, thermal stability and interaction with other defects which in turn decide the long-term radiation damage \cite{2, 3, 4, 5, 6, 7}. Molecular dynamics (MD) simulations provide a useful tool for studying the complex processes in collision cascades, and the resulting damage structures. Analysis of cluster morphologies predicted by MD collision cascades simulations can help in giving important insights into the radiation damage at atomistic scale and facilitating systematic study of different morphological transitions, interactions and migrations of the defect clusters produced. The morphologies produced with different inter-atomic potentials can be compared with results from Density Functional Theory (DFT) or experiments to validate the inter-atomic potentials.

The stable ground state configuration for a single SIA defect is a \(001\) dumbbell in bcc-Fe, and a \(\langle 111 \rangle\) crowdion in bcc W. According to DFT calculations \cite{8, 9}, the ground state in bcc-Fe for clusters of fewer than five SIA is a collection of \(\langle 110 \rangle\) dumbbells, while for larger clusters the ground state is a collection of \(\langle 111 \rangle\) crowdions. Parallel bundles of SIA dumbbells or crowdions (here after termed as only SIA) are generally the predominant type of clusters that form in a cascade. Larger parallel clusters can be identified as small dislocation loops, with a well-defined Burgers vector. These parallel clusters can be highly glissile depending on their orientation e.g. \(\langle 111 \rangle\) oriented clusters in W. The diffusion profile also depends on the size of these clusters \cite{10, 7}. Other SIA cluster morphologies include planar rings and 3D-rings such as C\(15\), multiple parallel bundles of SIAs oriented in different directions, collection
of SIAs oriented disorderly in different lines and combinations of these configurations [9, 1, 11, 12, 13, 14, 15, 6]. These cluster configurations are generally sessile [16] but many of these are meta-stable and change to glissile form even at low to moderate temperatures while others may remain sessile with considerable lifetimes [7].

The MD study for W by Wahyu et. al. [13] groups all the SIA clusters bigger than or equal to size 30 into five categories viz. homogeneously oriented parallel SIAs in (i) ⟨111⟩ and (ii) ⟨100⟩ directions, (iii) mixture of ⟨111⟩ and ⟨100⟩ loops, (iv) 3D cluster of ⟨111⟩ SIAs and (v) mixed 3D cluster composed of ⟨111⟩ and ⟨100⟩ SIAs. Another MD study of clusters in W by Sand et. al. [16] reports some of the sessile clusters having complex configurations that often have partial parallel oriented dumbbells. A similar study for Fe [15] classifies the clusters into C15-like structures and dislocation loops.

Our recent work on the classification of cluster shapes for bcc W and Fe [1] uses unsupervised machine learning with histogram of angle and distances of neighboring defects as features. The method has the advantage of using a general machine learning approach without any domain specific input. The classes can be broadly summarized as parallel SIAs, some other non-parallel but specific structures such as tripod-like arrangements of dumbbells, a pair of dumbbells arranged in T shape, 3D-rings, and also some non-specific arrangement of dumbbells and crowdions in random orientations. The classification however fails to distinguish between the clusters where different small components are attached to a similar bigger component due to its use of global histogram of angles and distances as features. For example, clusters having a single big parallel bundle of SIAs along with different smaller components e.g. big parallel bundle of SIAs with a ring, or non-parallel random oriented component, or parallel bundle of SIAs in different orientations are all grouped together in a single class.

To differentiate between the different cluster morphologies we can look at the arrangement of SIAs or SIA lines (axis of SIA dumbbell / crowdion) with respect to their neighborhood. Figure 1 (a) shows a cluster where all the SIAs are parallel to each other. In Figure 1 (b) SIAs are oriented in three different directions and every SIA has parallel SIAs in its neighborhood. There are essentially three parallel components, differing in direction of orientation. Figure 1 (e) shows another common pattern for parallel clusters where the cluster is augmented with a few non-parallel dumbbells on the fringes. We will show later in the results that such clusters are preferentially oriented in ⟨100⟩ direction while clusters with all parallel SIAs (Figure 1 (a)) are oriented in ⟨111⟩ direction. In Figure 1 (d), the arrangement of each SIA is very specific with respect to its neighbors. It can be seen that the hexagonal ring shape implies that every dumbbell is a part of a triad such that the relationship for the three neighboring dumbbells with each other is symmetric and transitive. We will further discuss this in methodology section. This specific pattern looks similar to Laves Phase C15 structure [17] which is of much interest because of its very high stability and sessile nature. Figure 1 (e) shows a configuration which is neither parallel nor it resembles any specific recurring order. Such clusters are generally metastable. Figure 1 (f) shows a cluster with one parallel component augmented with a hexagonal ring. The dislocation analysis of such a cluster might show only a blob of defects or only show a single parallel loop with or even without a connected blob of defects. A method which characterizes the morphology of a cluster using only the global properties of all the defects might not be able to identify the two different components present in such a mixed cluster. However, a bottom-up approach that looks at each SIA and its neighborhood to identify the local arrangement can distinguish between the homogeneous components within a mixed cluster.

We present a method to define the morphology of a cluster based on its constituent homogeneous components. A separate component in a cluster is a group of defects that are all arranged in a specific order. For example the cluster in Figure 1 (f) has one component of parallel dumbbells and another a hexagonal ring. The method identifies the composition of a mixed cluster into constituent components of various kinds in addition to analyzing single homogeneous clusters. The components are characterized as parallel SIAs (further distinguished with their orientations), planar rings, 3D-ring or other configuration of dumbbells and crowdions. For a component, we can study the orientations of SIA lines and the details of relative arrangement of SIA lines to form a specific morphology.

We explore morphologies of over 1000 SIA clusters formed in 149 W collision cascades with primary knock-on (PKA) energies ranging from 50 keV to 200 keV. The new method gives an automated way to identify various components including Laves phase structures and mixed clusters that are used to
classify the cluster morphologies. We show the typical morphologies identified and draw a comparison with the analysis using dislocation loops or only the orientations of SIAs. We show the statistical distribution of various morphologies and their properties.

Section 2 describes the methodology to find components of a cluster. The implementation of the method is incorporated as part of an open-source application for analyzing database of cascades, Csaran\textsh[18] available at \url{https://github.com/haptork/csaran\textsh}. Section 3 presents results of the method in a database of W collision cascades. The section presents some typical morphologies, trends of their properties and their statistical distribution across energies. We compare the results and also look at the challenges and possible limitations of the methodology. Conclusion is given in Section 4.

2 METHODOLOGY

We present a method to characterize an interstitial cluster by finding its constituent components. The steps for the method include (i) identification of defects, (ii) defining lines along SIAs (dumbbells / crowdions axes) and merging coinciding lines, and (iii) defining graph adjacency matrix to find connected components. This section gives an overview of each step before we present individual steps in detail, in further sub-sections.

In a Wigner–Seitz based method, an SIA is identified if a lattice site is occupied by more than a single atom \([19, 20]\). For most of the SIAs, a single lattice site is occupied by two interstitial atoms. We use an efficient modification of Wigner–Seitz method that uses modular arithmetic to associate each atom to its nearest lattice site \([1]\). The method also includes an option to include interstitial-vacancy pairs that are more than a threshold distance apart (hereafter termed as i–v pairs). These non-surviving extra pairs of defects, when included in the defect clusters, help in defining the shape of a cluster.

For a single SIA defect, the two atoms form a dumbbell shape around the lattice site and all the three points (two atoms and a lattice site) are approximately collinear. In addition, it is likely to find one or sometimes even two nearby i–v pairs that are more than a threshold distance apart which too are collinear \([1, 21]\). This configuration is termed as crowdion (Figure 2) \([19]\). Bigger defect clusters can be seen as composed of group of dumbbells / crowdions. Most clusters have only parallel group of these, while a few are such that the dumbbells or crowdions can be ordered in a specific ring kind of formation or even can be seen as totally random arrangements of crowdions and dumbbells. In a non-parallel arrangement and sometimes in a parallel arrangement as well, the lattice site is not always collinear to the pair of atoms that occupy it depending on the arrangement of SIA defects.

We first define a line for every SIA defect as the line formed by the two atoms and a lattice site) are approximately collinear. In addition, it is likely to find one or sometimes even two nearby i–v pairs that are more than a threshold distance apart which too are collinear \([1, 21]\). This configuration is termed as crowdion (Figure 2) \([19]\). Bigger defect clusters can be seen as composed of group of dumbbells / crowdions. Most clusters have only parallel group of these, while a few are such that the dumbbells or crowdions can be ordered in a specific ring kind of formation or even can be seen as totally random arrangements of crowdions and dumbbells. In a non-parallel arrangement and sometimes in a parallel arrangement as well, the lattice site is not always collinear to the pair of atoms that occupy it depending on the arrangement of SIA defects.

We first define a line for every SIA defect as the line formed by the two atoms occupying the same lattice site (SIA-line). The lattice site itself may or may not be collinear. We also define lines for the i–v pairs (Fig. 2(b)). We then merge the coinciding lines. We now look for the relationship of every line with its neighboring lines. All the neighboring lines that have a specified relationship (such as parallel to each other) are grouped into one component. This is done by defining an adjacency matrix, with
Figure 2: The figure schematically shows the output for a cluster after each step. The initial inputs from defect identification step are defect coordinates of the (i) lattice site and pair of atoms that occupy the lattice site (shown with group of three points with gray background), and coordinates of (ii) interstitial-vacancy pairs (shown with group of two points). In the first step the lines are defined by joining each pair of atoms and interstitial-vacancy pairs to get a number of lines (shown in b). The coinciding lines are then merged together (c). The angle and distance relationship between these lines are then used to find structurally homogeneous components that constitute the cluster (d).

lines as nodes and connectivity decided by the angle and distance between the lines. Then, by finding the connected components \[22, 23\] in the adjacency matrix we transitively group together all the lines that hold the defined relationship.

We can see that the overall method builds the solution in multiple layers of abstractions going from position of defects to lines, to graphs with lines as nodes and their connected components. Following are the three steps along with the information needed to proceed to the next step.

1. Finding defects.
   - In addition to the coordinates of defects, we need to have information about which atoms occupy the same lattice site. We also need i−v pairs grouped together.

2. Defining and merging coinciding lines.
   - For each line we define relationship with other lines using different distance metrics viz. shortest distance between lines, distance between line-segments and angle between lines.

3. Defining graph adjacency matrix and finding connected components.
   - We employ domain based knowledge to define rules for connections in the adjacency matrix and then find the connected components. The ring components are further verified by checking for cycles in the graph.

The algorithm for the first step (finding defects) is described in our previous work \[1\]. The second and third steps are discussed in the following sub-sections.

2.1 Defining Lines

1. Define line equations for surviving SIAs and interstitial-vacancy pairs (Fig. 2(b)). The parametric equation can be found for line passing through the two atoms in former case and interstitial and vacancy pair in the latter case.
    - For each SIA, the closest lattice site and another atom that occupies the same site are also found while identifying the defects. We find the equation of line that passes through these two atoms. The lattice site itself may or may not be collinear to this line. We also define lines passing through the i−v pairs.

2. Find and merge neighboring coincident lines and collinear points (Fig. 2(c)).
• For every point in the cluster, we find neighboring points and merge their lines if they are coincident. This operation can be made efficient using kd-tree data-structure \([24]\) to look for neighbors.

• In crowdions or their parallel bundles, it is very common to have many interstitial-vacancy lines coincident to SIA lines e.g. in Fig. 2(c), a single line \(L_1\) is defined by merging the coincident i–v pair lines with the SIA line.

• As shown by \(L_2\) and \(L_3\) lines in Fig. 2(c), for some crowdions in a bundle of crowdions, the lattice points can be seen as falling in a separate sub-line possibly due to local stresses due to nearby non-parallel structures or nearby clusters. The line of lattice points can be parallel as in \(L_2\) or at an angle as in \(L_3\) with the line defined by SIA. For such cases, the atoms are merged together in the SIA line while the vacancies are merged together in a sub-line with a separate line equation.

The lines constructed as described above when visualized help in qualitative assessment of different structures such as parallel crowdions (Figure 1 (a)), hexagonal ring(Figure 1 (f)) and C\(_{15}\) 3D-rings (Figure 1 (d)) and mixed clusters.

We use parametric equations of lines for efficiently calculating different properties such as 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 can be defined as: 
\[
\cos(\langle \vec{v}_1, \vec{v}_2 \rangle) = \frac{\langle \vec{a}_1 − \vec{a}_2, \vec{v}_1 \times \vec{v}_2 \rangle}{\|\vec{v}_1 \times \vec{v}_2\|}
\]

where \(\vec{v}_1\) and \(\vec{v}_2\) are direction vectors of the lines and \(\vec{a}_1\) and \(\vec{a}_2\) are any two points in the lines. The equation for shortest distance between two lines can be defined as:

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

The unit direction vector can be directly used to find the orientation of the line. We can also add line properties like defect count for each line, offset of non-collinear lattice sites, angle and distance of sub-lines and deviation of the line from the perfect orientations.

The lines and their neighborhood relationship are used in finding and characterizing the components. The distance between the lines can be defined as shortest of the distances between any point in the two lines or the shortest distance between the lines themselves using the equation shown above. Both of these metrics have their usage in defining adjacency matrix as we will see in the next section.

2.2 Defining Graphs and Finding Connected Components

In this step we first define the rules for connectivity between pairs of neighboring lines and then check for the connected components given the adjacency matrix built with the connectivity rules. The connectivity rules are defined differently for bundles of parallel dumbbells and planar rings or C\(_{15}\)-like 3D-rings.

After merging the coinciding 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.

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

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

where \(\theta\) is the angle between the lines and \(d\) is the shortest distance between the two lines found using the Equation (1).

All the lines that are parallel to each other and are 1NN or lesser distance apart are considered connected. The nearest neighbor lines are 1NN apart for almost all clustered defects in W. Taking the
distance tolerance as 2NN does not change the results and might work better for other materials like Fe.

A 3D-ring or C15-like structure can be defined as a composite of di-interstitial tripod structure (Figure 3 (ii)) and tri-interstitial hexagonal (Figure 3 (iii)) shape [17, 9]. Both of these structures are stable by themselves and also occur as a stand-alone sessile cluster. To understand the adjacency rules for rings we will now look at the structural details of these two.

![Figure 3](image)

**Figure 3**: The schematic figure shows that for some of the clusters, the cluster shape might appear to be different in a different snapshot of time due to thermal vibrations. A dumbbell may appear as a crowdion with extra interstitial-vacancy pair. For the parallel pair of crowdions in (i), with thermal vibrations, the SIA can be found at lattice site of ‘a’ or ‘b’ pair while the other pair left may appear as an interstitial-vacancy pair found based on a threshold. Similarly in (ii), the main triangular di-interstitial may appear as a tripod if all ‘a’, ‘b’ and ‘c’ are there, if two of them are there then the shape will look similar to tripod with one leg short while if only one appears it will look like a T-shaped cluster. For simplification the non-planar tripod is shown as a planar structure. A hexagonal tri-interstitial ring appears in only one form (iii). (iv) shows the combination of (ii) and (iii) forming a 3D ring shape. Figure 4 shows some of these clusters as seen in cascades.

The hexagonal structure is a planar arrangement of three dumbbells whose lattice sites are not collinear. The lines drawn by joining the dumbbell atoms appear as alternate sides of hexagon, forming a sixty degree angle with each other. All the three lines are oriented in ⟨110⟩ direction and the lattice points of these lines are 3NN distance apart from each other. The clusters that have this arrangement also have another crowdion / dumbbell that is orthogonal to the plane of hexagon and is oriented along ⟨111⟩ direction.

The di-interstitial tripod when appearing as a stand-alone cluster can appear in four different forms in different time snapshots (Figure 3(iii)) due to thermal vibrations. The forms are as given below.

- tripod like slightly non-planar arrangement of three dumbbells all 3NN apart and nearly at 90° angles with each other. The orientations of the three dumbbells changes gradually from ⟨110⟩ to ⟨111⟩ (Figure 4 (a)). It can be said that two dumbbells are approximately ⟨111⟩ while the third one is oriented in ⟨110⟩.

- tripod like slightly non-planar arrangement of two dumbbells and one i–v pair. Two dumbbells are 3NN apart and at 90° from each other and 1NN apart and 60° from the i–v pair. The orientations of the dumbbells is approximately ⟨221⟩ or between ⟨111⟩ and ⟨110⟩ while the i–v pair is oriented approximately in ⟨211⟩ or even more roughly in ⟨100⟩ direction (Figure 4 (b)).

- a pair of orthogonal dumbbells that are 1NN apart and oriented in ⟨100⟩ direction. One of the dumbbell is almost perfectly in ⟨110⟩ direction while the other one is slightly off towards ⟨11x⟩, where x is approximately 0.3 (Figure 4 (c)).

- triangle form where three atoms share a single lattice site and no i–v pair appears. The line drawn through any two points is in ⟨110⟩ direction (Figure 4 (d)).
A tripod or hexagonal cluster is sometimes augmented with a single or a couple of crowdions or dumbbells. The tripods can even appear in dumbbells. We show these composites and related clusters in the results section (Section 3). A more detailed discussion of the di-interstitial clusters can be a separate study in itself. However based on the above structural information for the two basis structure of rings, the values for adjacency matrix of rings are defined by the following relation:

$$ a_{i,j} = \begin{cases} 1, & \text{if } \theta \approx 60^\circ \text{ or } 90^\circ \text{ and } d' = 1\text{NN or } 3\text{NN} \\ 0, & \text{otherwise} \end{cases} \quad (3) $$

where $\theta$ is the angle between the lines and $d'$ is the shortest distance between any lattice points associated with the line. Both $60^\circ$ or $90^\circ$ and $1\text{NN}$ or $3\text{NN}$ are valid values for different time snapshots of the di-interstitial tripod like arrangement while for tri-interstitial hexagon, $3\text{NN}$ and $60^\circ$ suffice.

We define the connections by the above equations and find the connected components. For a ring structure, in addition to the binary relation defined by the Equation (3), it is also essential that the three 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 planar or 3D-ring (C15). These are called triangle graphs or C3 (cycles of size 3) in a graph [25]. 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 cycle contains 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 lines that hold the relationship defined in Equation (3).

A hexagon and C15 both look like rings, while tripods or their variations have ring like appearance only when appearing as a triangle. Moreover, they have a cycle in their adjacency matrix and are a basis shape for C15 ring. Due to these reasons we can categorize them in the rings category. It must be noted that we can classify tripods, hexagons, 3D-rings and variations of tripod all separately by defining separate adjacency matrix rules and cycle conditions according to the structural information.

A cluster with parallel or ring component, can have separate SIA lines that do not form part of the main bigger component e.g. crowdion tail of a ring or $\langle 111 \rangle$ crowdions on the fringes of a parallel component oriented in $\langle 100 \rangle$ direction. After we have found parallel components and ring like components, the components formed of neighboring non-parallel and non-ring lines can be identified as components formed of SIAs arranged in no-specific configuration. These are mostly non-recurring transient configurations. The rare occurrence implies that they might be less stable and change to more stable configuration. We will look at some of these clusters in the following results section.
3 RESULTS

The dataset used for the analysis has collision cascades in bulk W simulated with an initial temperature at 0 K, and evolved for 40ps. Electronic stopping is applied to atoms with energies above 10 eV \cite{26}. The database contains total 139 cascades at 50 keV, 100 keV, 150 keV and 200 keV simulated with Derlet et al. and Bjorkas et al. potential \cite{27, 28}. The cascades contain 1170 clusters of different sizes and morphologies. The algorithm takes approximately a minute to process the database on a regular desktop computer once defects and clusters have been identified. The defect and clusters identification using \textit{Csaransh} takes around ten minutes for the whole database. Algorithms at both the steps are memory efficient i.e. they do not allocate memory for the whole lattice.

In this section, we first show some of the noticeable clusters from the database having different morphologies of components and their combinations. We discuss their structural details, and challenges and limitations in identifying them. We also show comparison with cluster description using dislocation loops for some typical clusters that bring out the similarities and differences between the two results. We then look at some statistical trends for the properties and distribution of different morphologies.

3.1 Morphologies and Typical Clusters

Many of the clusters in the database are composed of fully parallel SIA lines with \(\langle 111 \rangle\) orientation (symbolized as \(\parallel\)). However, there are also clusters that have a single parallel component along with a few non-parallel SIA lines (symbolized as \(\parallel\mid\)). The typical clusters of this class are shown in Figure 5(a). Figure 5(b) shows a bigger cluster having this morphology and Figure 5(c) shows a smaller one. In all of these clusters the SIAs in main parallel component are always in \(\langle 100 \rangle\) orientation, while the other SIAs are mostly in \(\langle 111 \rangle\) and less often in \(\langle 110 \rangle\) orientation. The main component in Figure 5(d) and Figure 5(e) is in \(\langle 110 \rangle\) and \(\langle 111 \rangle\) direction which is quite rare for this morphology (one parallel component with a few non-parallel SIAs). These latter clusters are expected to be short lived metastable structures given their rare occurrence.

![Figure 5: Clusters with a component comprised of a bundle of parallel SIAs, along with a few non-parallel ones. Different colors represent orientations.](image)

Figure 6 shows some of the clusters found to have multiple parallel components (\(||/\)/). We observe that the components are in either \(\langle 111 \rangle\) or \(\langle 100 \rangle\) orientations. Figure 6(a) and (b) show clusters where there are three components, two in \(\langle 111 \rangle\) orientation and one having \(\langle 100 \rangle\) orientation sandwiched in between. Figure 6(c) and (d) show clusters with two components. In former the components are in \(\langle 100 \rangle\) and \(\langle 111 \rangle\) orientation, while in latter both have \(\langle 100 \rangle\) orientation.
Figure 6: Clusters with multiple parallel components. Different colors represent orientations.

Figure 7 shows some clusters that are identified as having three-dimensional ring structures with symmetry corresponding to the $C_{15}$ Laves phase (symbolized as @). A single 3D ring can be seen as either made of four planar hexagon rings sharing common edges (Figure 3 (b)) or four tripods (Figure 3 (c)) sharing common edges or one hexagon and one tripod without any shared edges. Figure 7 (a), (b) and (d) show multiple intertwined rings. There is a $\langle 111 \rangle$ tail in Figure 7(a) and one ring connected to one hexagon by $\langle 111 \rangle$ SIA lines in Figure 7(b).

The rings have $\langle 110 \rangle$ SIAs except for some deviations such as the ones shown in Figure 7 (c) and (d). In Figure 7 (c) the deviation can be viewed as the presence of tripod with one i–v pair (Figure 7 (b)) instead of all three dumbbells (Figure 7 (a)). These slight differences can happen due to thermal vibrations. It is interesting to note that when a tripod appears as an independent cluster the orientation of one SIA is $\langle 110 \rangle$ and the other two gradually deviate towards $\langle 111 \rangle$ (Figure 7 (a)).

Figure 7: Clusters with 3D rings. Different colors represent orientations. The vacancy sites that are not part of any SIA line and the lattice sites that are not collinear with their SIA line are represented as separate gray circles.

Figure 8 shows some of the clusters that are identified as composed of tripods and hexagons (also symbolized as @). Figure 8 (a) and (e) show a tripod and a hexagon with an extra SIA dumbbell, respectively. Figure 8 (d) shows two tripods arranged back to back. Figure 8 (b) and (c) shows structures that are almost like a tetra-pod. In Figure 8 (c) all the dumbbells of the tetra-pod are in $\langle 111 \rangle$ orientation. These structures are rare to recur.

Figure 8: Clusters composed of tripods and hexagons. The vacancy sites that are not part of any SIA line and the lattice sites that are not collinear with their SIA line are represented as separate gray circles.
Figure 9 shows some of the clusters composed of ring and parallel component both (symbolized as @||). Figure 9 (c) and (e) have multiple rings with (111) parallel component that seems like an extension of a single tail. Figure 9 (a) has both (111) and (100) parallel components. The 3D looking ring component in Figure 9 (b) and (c) contains some i-v pairs which change their usual C\textsubscript{15} like appearance. However, it is very likely that they are not very different and the deviations are due to the nearby parallel component and thermal vibrations. A method that uses orientations of SIA lines alone or uses some global feature is likely to be less effective in identifying morphologies in these mixed clusters especially in presence of deviations.

![Figure 9](image_url)

Figure 9: Clusters composed of both rings and parallel components.

Figure 10 shows the clusters that have neither a parallel component nor a ring component (#). Figure 10 (a) shows one such arrangement that occurs often exactly with same specifications. The di-interstitial cluster with orthogonal pair of dumbbells looks similar to the T-shaped variation of a tripod or triangle Figure 3 (c). However, here the dumbbells are 2NN distance apart instead of 1NN and the orientations are ⟨110⟩ and ⟨111⟩. We found that these clusters are less thermally stable compared to the tripod having T-shaped variation. Figure 10 (d) shows an arrangement which looks close to a pair of (a).

![Figure 10](image_url)

Figure 10: Clusters that have neither parallel nor ring or their basis components.

Figure 10 (b), (c) and (e) show some arrangements of SIAs where it is difficult to glean any specific order. Moreover, these arrangements rarely ever recur. Hence, it is very likely that these are very short living transients with stable counterpart in either ring like arrangement or the short living T-shaped arrangement shown in (a). We can classify them into one of these based on their similarity but it will not be fair unless we get some more idea on the mechanism of their transition to stable counterpart, possibly by studying their evolution in MD at different temperatures.

Figure 11 and Figure 12 shows dislocation analysis results found using the DXA algorithm [29] with Ovito [30]. Figure 11 (a), (b) and (c) correspond to Figure 5 (a), Figure 6 (c) and (b), respectively. The dislocation loops correspond well with the parallel components identified by the presented algorithm. The orientations shown also match.

Figure 12 shows some cases where results of dislocation loops lack the complete description of the cluster morphology. Figure 12 (a) and (b) correspond to Figure 6 (a) (||/) and Figure 9 (d) (@||). Figure 12 (a) shows only one parallel ⟨111⟩ component possibly because of smaller sizes of other ⟨111⟩ and ⟨100⟩ constituent components. Similarly, the hexagonal ring of figure Figure 9 (d) is not identified.
and only a ⟨111⟩ loop is shown. A cluster with only a single ⟨111⟩ loop in W is glissile as opposed to the sessile nature of these mixed morphologies. The missing information about the structure can have implications on conclusions drawn based on the structure-property relationship. Figure 12 (c) and (d) show another set of mixed clusters of type ||// and @||. In both, a dislocation loop is shown along with a blob representing defects with no dislocation loop.

Figure 13 shows the total number of defects and fraction of defects in clusters for different energies. With increase in energy both the number of defects and variability in them increases. The fraction of defects in clusters increase sharply initially and then gradually seems to saturate at around 90 percent.

Figure 14 shows the distribution of defects among different cluster morphologies at each energy. Figure 14 (a) shows the defects in each morphology as a fraction of the total defects in clusters. It also shows fraction of defects that are in parallel component without regard to orientation and the composition of cluster i.e. sum of all the parallel components appearing in ||, ||−||, ||// and @||. Figure 14 (b) shows the mean number of clusters appearing per cascade for each morphology.

3.2 Statistical Results

Figure 13 shows the total number of defects and fraction of defects in clusters for different energies. With increase in energy both the number of defects and variability in them increases. The fraction of defects in clusters increase sharply initially and then gradually seems to saturate at around 90 percent.

Figure 14 shows the distribution of defects among different cluster morphologies at each energy. Figure 14 (a) shows the defects in each morphology as a fraction of the total defects in clusters. It also shows fraction of defects that are in parallel component without regard to orientation and the composition of cluster i.e. sum of all the parallel components appearing in ||, ||−||, ||// and @||. Figure 14 (b) shows the mean number of clusters appearing per cascade for each morphology.
Figure 13: (a) Point estimates for number of defects in a cascade at different energies. (b) Fraction of defects that form a cluster, for different energies.

Figure 14: Distribution of defects across different morphologies, (a) as a fraction of the total defects in clusters, (b) as the average number of clusters of that morphology in a single cascade. The fraction of defects appearing in parallel components in (a) represent defects in parallel components of $\|$, $\|-\|$, $\|--\|$, and $\bowtie$.

Figure 15 shows the size distribution of each morphology at different energies. The fraction of defects in non-parallel clusters decreases with energy (Figure 14 (a)) while number of clusters increase (Figure 14 (b)), implying a decrease in the sizes of these clusters as evident from Figure 15. Bigger sized clusters are mostly of type $\|--\|$ with multiple parallel components. The clusters in $\|--\|$ increase in both size and number with energy.

Figure 16 shows the distribution of sizes of components for multiple parallel component clusters ($\|--\|$). The components are arranged with descending order of their sizes. For many of the clusters, the size of the smallest component is also considerable. The number of components go up to six.
Figure 15: Size distribution of each cluster morphology at different energies.

Figure 16: Sizes of different components in clusters with multiple parallel components. Size of the clusters and different constituent components are represented with different colors.

Figure 17 shows the joint distribution of ring size and number of non-coinciding SIA and i–v pair lines forming the ring component. It shows that a very high number of clusters are of size 2 formed of 2 or 3 lines which implies a tripod shaped cluster. There are also clusters having very large number of lines and comparatively small size such as 6 lines and size 2 which is indicative of a 3D ring.

Figure 18 shows the distribution of SIA orientations for different morphologies and components. For || the orientations are shown separately for main component and blocking SIAs and for // orientations are shown separately for biggest, second biggest and all other components. It can be observed that almost all the ⟨111⟩ oriented clusters fall in category || and ⟨100⟩ in category ||. Most of the lines in rings (@) are ⟨110⟩ while in #, both ⟨110⟩ and ⟨111⟩ are almost equal.
Figure 17: Joint distribution of ring size and number of lines (non-coinciding SIA lines and i–v pairs) for clusters with rings (both @ and |||). Different shapes represent the different frequency of occurrences. The histogram bars on the margins (top and right) show marginal distributions for the ring size and number of lines, separately. Di-interstitial formed of 2 or 3 lines are most common while there are also 3D rings having a large number of lines and a comparatively small size e.g. (size 2 with 6 lines).

Figure 18: Distribution of SIA dumbbell / crowdion orientations for different morphologies. Different orientations are represented by different shapes / colors. For ||−! and ||/, the different sizes of glyphs represent the SIAs appearing in different components within the cluster. X-axis value @, contains all the ring components irrespective of whether appearing in @ or |||.

Figure 19 shows the number of atoms in a SIA line for ⟨111⟩ and ⟨100⟩ oriented parallel components of size 8 or more as a function of neighboring SIA lines within 1NN (distance between lines found using Equation (1)). It has been postulated that the ⟨111⟩ SIAs in the central part of a cluster are more extended forming longer crowdions than the SIAs on the surface [31]. The number of SIA lines within 1NN is a good indicator of whether a SIA is in the surface of the cluster or towards the center in the bulk. For a ⟨111⟩ parallel cluster the maximum number of neighbors for a central SIA are six while for ⟨100⟩ this value is four.
Figure 19: Number of atoms in a SIA line as a function of its nearest neighboring SIAs. The number of neighbors indicate whether a SIA is in central part of the cluster or on the surface. The plot shows that the extents of a SIA that is 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 SIAs with four neighbors become longer on average but still a good fraction has only two atoms. The extent of SIAs in ⟨100⟩ rarely goes beyond three atoms while ⟨111⟩ crowdions are longer.

We find only dumbbells (SIAs with two atoms) in the ⟨110⟩ direction. Figure 19 shows that the extents of SIAs in ⟨100⟩ rarely goes beyond three atoms while in ⟨111⟩ longer crowdions are being formed. For ⟨111⟩ crowdions the central SIAs are longer. The correlation value between the two variables (extent of a SIA and its number of neighbors) is 0.78. While for ⟨110⟩ the correlation value is just 0.44. It can be observed from the plot that although the extent of SIAs increase on average with increase in neighbors for ⟨100⟩ but still a good fraction of SIAs that are in the central region of a cluster (neighbor count of four), have only two atoms. While for ⟨111⟩, a central SIA almost always has more than two atoms.

Figure 20 shows the deviation in SIA orientation from primary ideal directions (⟨111⟩, ⟨110⟩ and ⟨100⟩) and deviation in collinearity of lattice point from the two atoms co-occupying it. These deviations can occur due to thermal vibrations and presence of non-parallel arrangement of SIA in a cluster. The lattice-point non-collinearity value represents the distance of the lattice point (in Å) from the line defined by the two dumbbells occupying it. The angular deviation from perfect SIA orientation is normalized by dividing the angle with the maximum value of deviation observed i.e. 30°.
4 CONCLUSION

We have described a bottom-up approach to characterize the cluster morphologies based on their constituent components that are composed of SIA lines sharing a specific neighborhood relationship. The reduction of the problem to well-established problems from graph theory viz. connected components and cycles in a graph, have made the solutions efficient and easy to extend to a new arrangement of SIAs by defining rules for the adjacency matrix. The implementation of the algorithm is fast making it ideal to automatically characterize clusters in a big dataset.

The method is used to study a dataset of 1170 clusters formed in 149 collision cascades in W from 50 keV to 200 keV. We show the example clusters from different morphologies identified that include parallel loops in different orientations, clusters with multiple parallel loops, 2D and 3D rings, combination of rings and parallel components and other specific and random arrangement of SIAs. We validate the method by showing same results as dislocation analysis for clusters with big parallel components. For other clusters, the presented method can give a more accurate description. We show examples for cluster morphologies where dislocation analysis alone may not be sufficient for the accurate prediction of cluster structure and its implied behavior. The graph adjacency matrix and other derivative features like cycles in graph, graph clustering coefficient etc. can also be used with automated classification techniques like graph neural networks.

The statistical results show the presence of parallel bundles of SIAs in ⟨111⟩ and ⟨100⟩ orientations which is in agreement with the prior simulations [16, 13] and experiments [11]. The presence of 3D rings and mixed parallel components is also in agreement with simulations [16, 13]. The correlation found between the extent of a SIA line and it being in the central part of a cluster shows agreement with the prior observation [31] and quantifies it. The results also show structural details and statistics of rings, mixed clusters and other non-specific structure of clusters. The structural details and statistical distributions give a comprehensive way to summarize the micro-structural changes due to primary radiation damage. With this understanding a systematic study to establish structure-property relationships for the morphologies can be carried out such as their diffusion, stability and interaction.

REFERENCES

[1] U. Bhardwaj, A. E. Sand, M. Warrier, Classification of clusters in collision cascades, Computational Materials Science 172 (2020) 109364. doi:https://doi.org/10.1016/j.commatsci.2019.109364.

[2] D. Bacon, F. Gao, Y. Osetsky, The primary damage state in fcc, bcc and hcp metals as seen in molecular dynamics simulations, Journal of Nuclear Materials 276 (1) (2000) 1 – 12. doi:https://doi.org/10.1016/S0022-3115(99)00165-8.

[3] B. Singh, S. Golubov, H. Trinkaus, A. Serra, Y. Osetsky, A. Barashev, Aspects of microstructure evolution under cascade damage conditions, Journal of Nuclear Materials 251 (1997) 107 – 122, proceedings of the International Workshop on Defect Production, Accumulation and Materials Performance in an Irradiation Environment. doi:https://doi.org/10.1016/S0022-3115(97)00244-4.

[4] Y. Osetsky, D. Bacon, A. Serra, B. Singh, S. Golubov, Stability and mobility of defect clusters and dislocation loops in metals, Journal of Nuclear Materials 276 (1) (2000) 65 – 77. doi:https://doi.org/10.1016/S0022-3115(99)00170-1.

[5] C. Becquart, A. Souidi, C. Domain, M. Hou, L. Malerba, R. Stoller, Effect of displacement cascade structure and defect mobility on the growth of point defect clusters under irradiation, Journal of Nuclear Materials 351 (1) (2006) 39 – 46, proceedings of the Symposium on Microstructural Processes in Irradiated Materials. doi:https://doi.org/10.1016/j.jnucmat.2006.02.022.

[6] Y. Osetsky, D. Bacon, B. Singh, B. Wirth, Atomistic study of the generation, interaction, accumulation and annihilation of cascade-induced defect clusters, Journal of Nuclear Materials 307-311 (2002) 852 – 861. doi:https://doi.org/10.1016/S0022-3115(02)01094-2.
[7] F. Gao, D. Bacon, Y. Osetsky, P. Flewitt, T. Lewis, Properties and evolution of sessile interstitial clusters produced by displacement cascades in alpha-iron, Journal of Nuclear Materials 276 (1) (2000) 213 – 220. doi:https://doi.org/10.1016/S0022-3115(99)00180-4.

[8] C.-C. Fu, J. Dalla Torre, F. Willaime, J.-L. Bocquet, A. Barbu, Multiscale modelling of defect kinetics in irradiated iron, Nature materials 4 (1) (2005) 68–74.

[9] M.-C. Marinica, F. Willaime, J.-P. Crocombette, Irradiation-induced formation of nanocrystallites with c15 laves phase structure in bcc iron, Phys. Rev. Lett. 108 (2012) 025501. doi:10.1103/PhysRevLett.108.025501.

[10] C. Domain, C. Becquart, L. Malerba, Simulation of radiation damage in fe alloys: an object kinetic monte carlo approach, Journal of Nuclear Materials 335 (1) (2004) 121 – 145. doi:https://doi.org/10.1016/j.jnucmat.2004.07.037.

[11] X. Yi, M. Jenkins, M. Briceno, S. Roberts, Z. Zhou, M. Kirk, In situ study of self-ion irradiation damage in W and W–5re at 500c, Philosophical Magazine 93 (14) (2013) 1715–1738. doi:10.1080/14786435.2012.754110.

[12] A. Sand, J. Dequeker, C. Becquart, C. Domain, K. Nordlund, Non-equilibrium properties of interatomic potentials in cascade simulations in tungsten, Journal of Nuclear Materials 470 (2016) 119 – 127. doi:https://doi.org/10.1016/j.jnucmat.2015.12.012.

[13] W. Setyawan, G. Nandipati, K. J. Roche, H. L. Heinisch, B. D. Wirth, R. J. Kurtz, Displacement cascades and defects annealing in tungsten, part i: Defect database from molecular dynamics simulations, Journal of Nuclear Materials 462 (2015) 329 – 337. doi:https://doi.org/10.1016/j.jnucmat.2014.12.056.

[14] R. Alexander, M. C. Marinica, L. Proville, F. Willaime, K. Arakawa, M. R. Gilbert, S. L. Dudarev, Ab initio scaling laws for the formation energy of nanosized interstitial defect clusters in iron, tungsten, and vanadium, Physics Rev. B 94 (2) (2016) 024103. doi:10.1103/PhysRevB.94.024103.

[15] J. Byggmästar, F. Granberg, K. Nordlund, Effects of the short-range repulsive potential on cascade damage in iron, Journal of Nuclear Materials 508 (2018) 530 – 539. doi:https://doi.org/10.1016/j.jnucmat.2018.06.005.

[16] A. E. Sand, S. L. Dudarev, K. Nordlund, High-energy collision cascades in tungsten: Dislocation loops structure and clustering scaling laws, EPL (Europhysics Letters) 103 (4) (2013) 46003. doi:10.1209/0295-5075/103/46003.

[17] L. Dézerald, M.-C. Marinica, L. Ventelon, D. Rodney, F. Willaime, Stability of self-interstitial clusters with c15 laves phase structure in iron, Journal of Nuclear Materials 449 (1) (2014) 219 – 224. doi:https://doi.org/10.1016/j.jnucmat.2014.02.012.

[18] U. Bhardwaj, H. Hemani, M. Warrier, N. Semwal, K. Ali, A. Arya, Csaransh: Software suite to study molecular dynamics simulations of collision cascades, Journal of Open Source Software (Sep 2019). doi:10.21105/joss.01461.

[19] J. Gibson, A. N. Goland, M. Milgram, G. Vineyard, Dynamics of radiation damage, Physical Review 120 (4) (1960) 1229.

[20] K. Nordlund, R. Averback, Point defect movement and annealing in collision cascades, Physical Review B 56 (5) (1997) 2421.

[21] U. Bhardwaj, S. Bukkuru, M. Warrier, Post-processing interstitialcy diffusion from molecular dynamics simulations, Journal of Computational Physics 305 (2016) 263 – 275. doi:https://doi.org/10.1016/j.jcp.2015.10.034.

[22] R. Tarjan, Depth-first search and linear graph algorithms, SIAM Journal on Computing 1 (2) (1972) 146–160. doi:10.1137/0201010.
[23] D. J. Pearce, An improved algorithm for finding the strongly connected components of a directed graph, Tech. rep. (2005).

[24] J. L. Bentley, Multidimensional binary search trees used for associative searching, Commun. ACM 18 (9) (1975) 509–517. doi:10.1145/361002.361007.

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

[26] A. E. Sand, K. Nordlund, On the lower energy limit of electronic stopping in simulated collision cascades in Ni, Pd and Pt, J. Nucl. Mater. 456 (0) (2015) 99–105. doi:10.1016/j.jnucmat.2014.09.029.

[27] P. M. Derlet, D. Nguyen-Manh, S. L. Dudarev, Multiscale modeling of crowdion and vacancy defects in body-centered-cubic transition metals, Phys. Rev. B 76 (2007) 054107. doi:10.1103/PhysRevB.76.054107.

[28] C. Björkas, K. Nordlund, S. Dudarev, Modelling radiation effects using the ab-initio based tungsten and vanadium potentials, Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms 267 (18) (2009) 3204 – 3208. doi:https://doi.org/10.1016/j.nimb.2009.06.123.

[29] A. Stukowski, V. V. Bulatov, A. Arsenlis, Automated identification and indexing of dislocations in crystal interfaces, Modelling and Simulation in Materials Science and Engineering 20 (8) (2012) 085007. doi:10.1088/0965-0393/20/8/085007.

[30] A. Stukowski, Visualization and analysis of atomistic simulation data with OVITO-the Open Visualization Tool, Modelling and simulation in materials scient and engineering 18 (1) (JAN 2010). doi:{10.1088/0965-0393/18/1/015012}.

[31] S. L. Dudarev†, Coherent motion of interstitial defects in a crystalline material, Philosophical Magazine 83 (31-34) (2003) 3577–3597. doi:10.1080/14786430310001599388.