



**HAL**  
open science

## Coping with the stochasticity of collision cascades in Molecular Dynamics simulations

Thomas Jarrin, Antoine Jay, Nicolas Richard, Anne Hémeryck

► **To cite this version:**

Thomas Jarrin, Antoine Jay, Nicolas Richard, Anne Hémeryck. Coping with the stochasticity of collision cascades in Molecular Dynamics simulations. Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms, 2021, 500-501, pp.1-9. 10.1016/j.nimb.2021.02.015 . hal-03287083

**HAL Id: hal-03287083**

**<https://laas.hal.science/hal-03287083>**

Submitted on 15 Jul 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Coping with the stochasticity of collision cascades in Molecular Dynamics simulations

Thomas Jarrin<sup>a,b,\*</sup>, Antoine Jay<sup>b</sup>, Nicolas Richard<sup>a</sup>, Anne Hémercyck<sup>b</sup>

<sup>a</sup>CEA, DAM, DIF, F-91297 Arpajon, France

<sup>b</sup>LAAS-CNRS, Université de Toulouse, CNRS, Toulouse, France

---

## Abstract

The high stochasticity of collision cascades makes running large sets of molecular dynamics simulations mandatory to obtain meaningful statistics. The convergence of the number of defects and of clusters and of the Primary Knock-On Atom (PKA) penetration depth with respect to the number of simulations in the sets is investigated for PKAs of 1 keV and 5 keV in Si. Two methods for setting the initial directions of the PKAs are compared: one is entirely based on randomness and the other integrates symmetry considerations. The latter method eases the convergence of the results. Larger sets are needed to converge the PKA depth than the number of clusters and defects. We observe that the higher the energy of the PKA, the harder it gets to reach the convergence. We find that large sets of simulations enables to get rid of the influence of the initial position of the PKA.

*Keywords:* displacement damage, collision cascades, molecular dynamics, statistics, stochasticity, silicon

---

## 1. Introduction

Over the past 30 years, Molecular Dynamics (MD) has become the most popular simulation tool for Displacement Damage (DD) simulations [1, 2]. The atomistic nature of the technique associated to the good compromise between computational cost and accuracy of the physical description allowed by the interatomic potentials are true assets over Binary Collision Approximation (BCA) techniques [3]. However, the stochastic nature of the collision cascades is generally badly treated within MD whereas the very little computational cost of collision cascades simulations with BCA codes enables to perform very large number of simulations. Depending on the initial direction of the Primary Knock-on Atom (PKA), the sequences of collisions and therefore the resulting cascades outputs can be drastically different. If one wants to assess the response of a material to particle irradiation with MD, underestimating the high stochasticity of the cascades and thus the need to perform large number of simulations is a fundamental issue. To provide with the best picture of the response of a material, it is necessary to work with results (*i.e.* mean values and distributions) obtained with meaningful statistical ensembles. Ensuring the convergence of the results with respect to the number of simulations performed is therefore of prime importance.

In spite of this, looking at recently published articles, no consensus on a satisfactory number of simulations to

perform, on a strategy to employ to have converged results or even on the need for a preliminary convergence study appears. Below is a non-exhaustive list of the chosen methods for setting both the initial directions of the PKA and the size of the statistical ensembles in recent MD collision cascades simulations articles: Trung *et al.* constructed sets of 10 simulations in NiAl [4], He *et al.* based their statistical ensembles on 30 simulations initiated in random directions distinct by a little angle from a defined direction in GaAs [5], Buchan *et al.* chose to construct their mean values in graphite over 25 simulations in definite directions [6], Zarkadoula *et al.*'s statistics for Ni and NiFe are based on 12 simulations in random directions [7], Gao *et al.* chose to initiate the cascades with random PKA directions for the 20 simulations performed in GaAs [8] and Christie *et al.* chose to do 20 simulations in well defined directions in graphite [9].

This lack of coherence in the methods employed probably comes from a lack of studies focused on the specific aspect of the stochasticity in MD simulations of collision cascades. To our knowledge, there exists only two papers dedicated to this subject:

- In 2015, Warriier *et al.* proved in [10] that for cascades initiated with PKAs of energies in the range of 1 keV to 5 keV whose initial directions are distinct and randomly chosen, 50-60 simulations are needed to make the number of displaced atoms stabilize in W and 60-80 simulations in Cu.
- More recently, Voskoboinikov studied the influence of the number of simulations performed on the estimated number of defects and clusters in surface

---

\*Corresponding author

Email address: thomas.jarrin@cea.fr (Thomas Jarrin)

collision cascades in Ni, Al, Ti and their alloys [11] and provided a minimum number of simulations to be performed to reach the convergence depending on the system of interest, the temperature and the energy of the PKA.

In our research group, Jay *et al.* [12] and Jarrin *et al.* [13, 14] based their statistical ensembles on the large number of 100 simulations in Si, Ge and Si-Ge alloys. Moreover, like some others did [6, 9] but contrary to what Warrior *et al.* did in [10], Jay *et al.* [12] and Jarrin *et al.* [13, 14] took into account the symmetry of the system when setting the initial directions of the PKAs.

Because the number of simulations needed to have converged results is highly dependent on the system of interest, on the energy of the PKA and on the chosen convergence criterion, it is not possible to give a definite number of simulations to perform which would be valid for every possible case. Nevertheless, by studying the convergence of collision cascades outputs initiated with PKAs of 1 keV and 5 keV in Si, this article aims at providing the proofs of the necessity to conduct a preliminary study on the convergence of cascades properties with respect to the number of simulations performed to assess the truthfulness of the results. It also demonstrates the necessity to treat the cascades results in a statistically rigorous way and shows the assets of setting the initial directions of the PKAs considering the symmetry of the material. In this article, we demonstrate that the results obtained when carefully choosing the initial directions of the PKAs using a method based on the symmetry of the system (described in section 2.2 for diamond-like structures) are better converged than results obtained by randomly setting the initial directions of the PKAs. Finally, it studies the influence of the initial position of the PKA on the results in Si. This last aspect is relatively straightforward to check in simple structures like Si but can become very complex depending on the crystal structure of the material under investigation. In section 2 the simulation details and the methods employed to study the convergence of the results are presented. The results of the convergence study are presented and discussed in section 3.

## 2. Methods

### 2.1. Molecular Dynamics simulations

The MD simulations have been run using the LAMMPS code [15]. The cascades are initiated with 1 keV or 5 keV PKAs in Si. Boxes of 1 000 000 atoms made of  $50 \times 50 \times 50$  diamond-like unit cells are employed in both cases. The simulation boxes are divided into two areas: in the outer cells the velocities are rescaled to maintain the temperature at a desired value (thus absorb the thermal wave), and the inner cells form an NVE ensemble in which the atoms evolve freely to simulate the collision cascade.

The initialization of the simulation is made by scaling the velocities of all atoms, such that the overall temperature is the desired one (300 K in the current work). The system is equilibrated during 20 ps with a timestep of 1 fs.

As the velocity of the atoms drastically changes throughout the cascade, the integration timestep has to be changed during the simulations. Calculations initiated with both 1 keV and 5 keV PKAs are run for 200 ps. We imposed the condition that no atom moves more than 0.02 Å between two steps of the simulation, with a timestep varying between 0.01 fs and 1 fs.

The potential employed is the Stillinger-Weber (SW) one developed in [16]. To better describe short interatomic distances, the SW potential must be combined to a repulsive two-body potential. The chosen repulsive potential is the Ziegler Biersack Littmark one (ZBL) [17]. SW and ZBL potentials are combined together through a Fermi function as in [18]:

$$V_{tot}(r) = (1 - F(r))V_{ZBL}(r) + F(r)V_{SW}(r) \quad (1)$$

where  $r$  is the distance between two atoms,  $V_{tot}$  is the total potential,  $V_{ZBL}$  the repulsive ZBL potential,  $V_{SW}$  the SW potential and  $F$  the Fermi function used to link the two potentials. The expression of the Fermi function as well as its parameters values used for Si can be found in [13].

To account for the electronic effects, the Two Temperature Model is used in combination with MD [19]. The chosen parameters are detailed in [13] and [14].

### 2.2. Settings for the convergence study

Some vocabulary needs to be defined for the proper understanding of this study. The term "simulation" refers to a single collision cascade simulation. The term "set" refers to a given number of simulations performed to calculate the statistical quantities of the set. In each simulation of a set, the initial direction of the PKA is different and its energy is the same. The size of the set is the number of simulations in the set.

Two distinct methods are employed to study the convergence:

(i) In the first method, called **Random**, the initial directions of the PKAs are randomly chosen.

(ii) In the second method, called **Symmetry**, a set of distinct unit vectors (*i.e.* directions) uniformly distributed on the surface of a sphere is defined. Only the directions which are in the minimal space domain inequivalent to the others by crystalline symmetries have been kept, as depicted in Fig. 1. The unit sphere surrounding an atom is therefore divided into 6 symmetric parts. The minimal space domain is one of these parts. Thus, for each direction in the minimal space domain, there exists 5 others perfectly equivalent in the unit sphere. The weight of all the chosen directions is therefore the same. The python script for the construction of the directions of Fig. 1 (*i.e.* for a diamond structure) is available as a supplementary material.

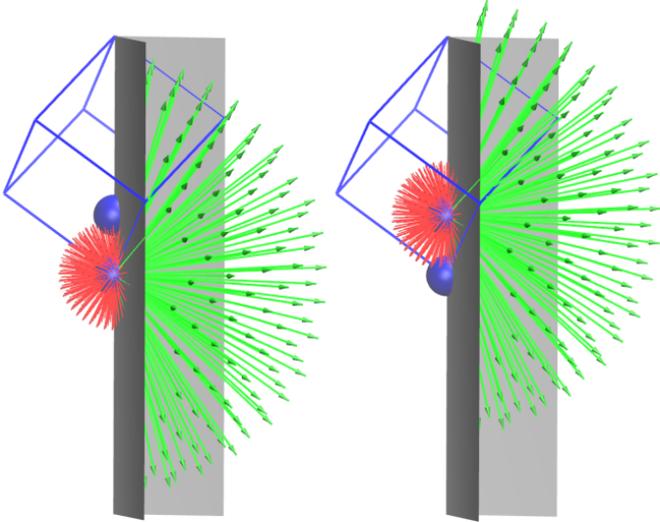


Figure 1: Scheme of the minimal space domain in which 100 inequivalent directions are constructed. The two inequivalent atomic sites of the face-centred cubic diamond primitive unit cell are represented in blue. The inequivalent directions are depicted by long green arrows and the minimal space domain is delimited by the grey boundaries. The red small arrows represent the directions symmetrically equivalent to the green arrows of the minimal space domain.

The convergence is studied on: the distributions and the mean values of the number of defects and of the number of clusters of defects at the end of the cascades and of the PKA penetration depth. Depending on the initial direction of the PKA, the size of the cascades can be very different. In some cases, the PKA or a Secondary Knock-On Atom (SKA) can find itself in what is called a channeling direction and travel over long distances without undergoing any collision. Thus, the size of the simulation box has to be carefully chosen so that no atom exits the box, even in the case of channeling. The size of the box is therefore highly dependent on the energy initially imparted to the PKA. During a simulation, if one atom or more exits the box, the simulation should not be deleted from the statistics. Actually, the channeling simulations are important cases which must be part of the statistics to have a complete picture of the material response. As a solution, the simulation can be relaunched from its last non-exited step with a bigger box.

Throughout this article, the number of defects refers to the number of vacancies or interstitials, not the sum of the two. The Lindemann criterion [20] is employed for counting the defects in the damage structures. A discussion on the different existing methods for counting defects can be found in [1]. It was observed that the ratio of the number of defects obtained with all the methods, including Wigner-Seitz and Lindemann, is nearly constant for every considered cascade. Two defects are in the same cluster when they are not distant by more than  $a \times \sqrt{3}/2$ ,  $a$  being the lattice parameter of the system (5.431 Å in Si [21]) so that a cluster of defects is defined as a group of two or

more defects. For a single collision cascade, the PKA penetration depth corresponds to the maximum value of the projection of the vector linking the initial position of the PKA to its current position, on the initial velocity vector of the PKA. It is detailed with a scheme in [13].

The convergence is here considered as a combination of reproducibility and accuracy. To study the reproducibility of the simulations, statistical quantities for sets of different sizes are compared. Those quantities of interest are the mean value ( $\mu$ ), the minimum value (Min), the first quartile value (Q1), the median or second quartile value (Q2), the third quartile (Q3) and the maximum value (Max). The quartile values are defined as follows: 25% of the values of the statistical distribution lies below the first quartile value (Q1), 50% below the median value (Q2) and 75% below the third quartile (Q3) value. Those statistical quantities are useful to describe distributions as they allow us to investigate the stabilization (*i.e.* reproducibility) of the distributions with respect to the number of simulations performed. The accuracy is quantified through the standard error of the mean (SEM). This quantity measures the uncertainty in the calculation of a mean value. Rather than the brute standard error of the mean we employ the ratio of the standard error of the mean to the mean value  $SEM/\mu$  (called error coefficient).

### 3. Results and discussions

For both the Random and Symmetry methods, the distributions (*i.e.* box plots) and mean values of the number of defects, number of clusters and of the PKA penetration depth for Si PKAs of 1 keV in Si and sets of size 10, 20, 30, 40, 50, 60, 70, 80, 90 and 100 are presented in sections 3.1, 3.2 and 3.3. For the same set sizes, with the Symmetry method and with Si PKAs of 5 keV in Si, the distributions and mean values of the number of defects, number of clusters and of the PKA penetration depth are presented in section 3.4. The influence of the lattice site initially chosen as the position of the PKA is investigated in section 3.5.

#### 3.1. Convergence of the number of defects for 1 keV PKAs

Results provided in this section are obtained for PKAs of 1 keV in Si. The mean values of the number of defects in the Random and Symmetry cases for PKAs of 1 keV in Si are respectively shown in Fig.2(a) and Fig.2(b). The box plots of the number of defects in the Random and Symmetry cases for PKAs of 1 keV in Si are respectively displayed in Fig.2(c) and Fig.2(d).

Both the box plots and mean values of Fig.2 are subject to strong variations for simulation sets of size inferior or equal to 30: the mean values range from 58 to 70 in the Random case and from 63 to 69 in the Symmetry case. Between 40 and 70 simulations in the sets, the discrepancies between the mean values are smaller, except in the Symmetry case (*i.e.* Fig.2(b)) where a noticeable difference can be spotted between the sets of size 50 (67 defects)

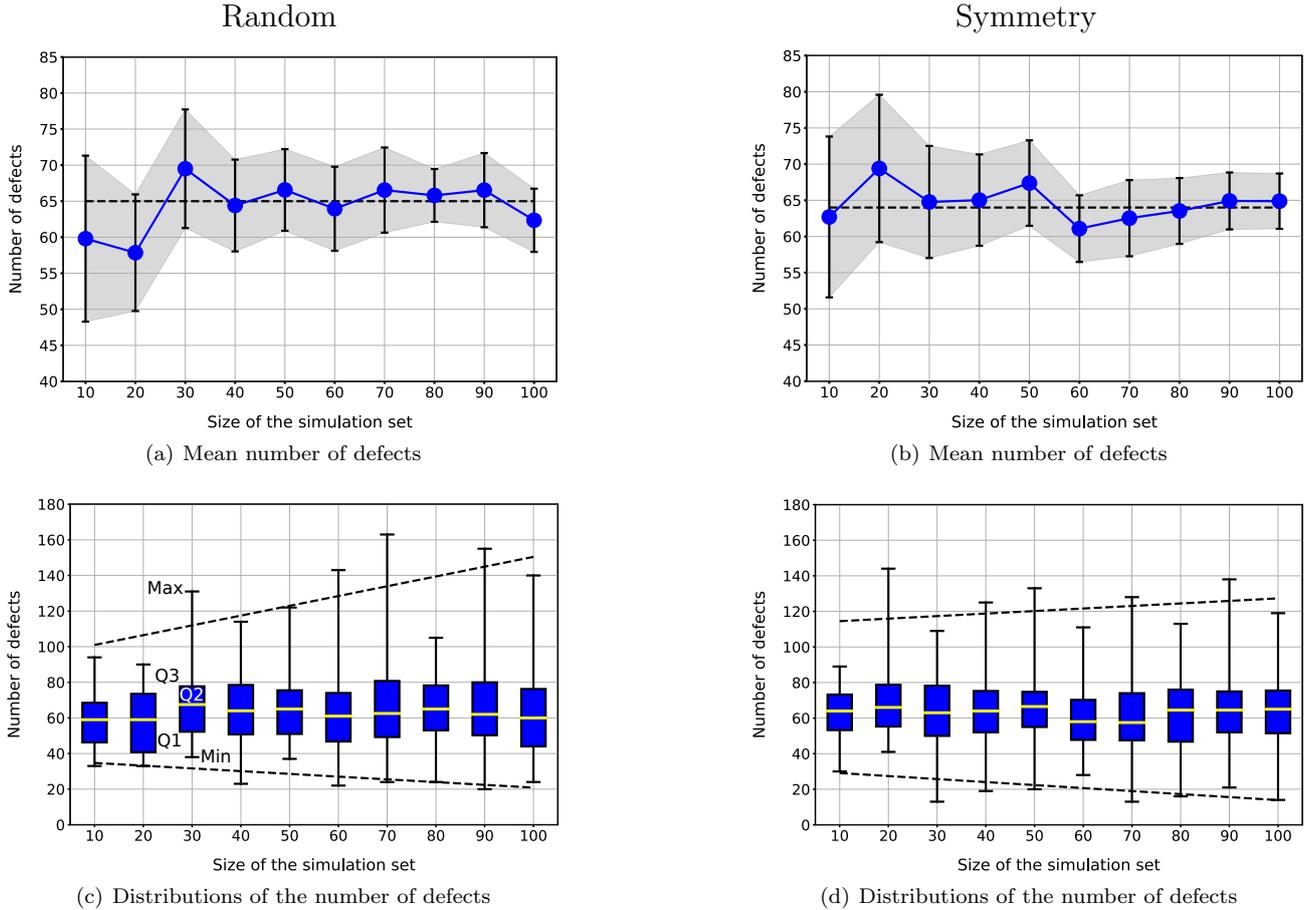


Figure 2: Mean number of defects in the (a) Random and (b) Symmetry cases and box plots of the number of defects in the Random (c) and Symmetry (d) cases for 1 keV Si PKAs in Si with respect to the size of the simulation sets (*i.e.* the number of simulations in the sets). In (a) and (b), for each size of set, the error bars correspond to  $-1.96 \times \text{SEM}$  and  $1.96 \times \text{SEM}$ . The probability that the real mean value sits within the error bars range is equal to 0.95. The region of the graphs delimited by the error bars of each set size is shaded. Solid blue lines are guides to the eye. In (a) and (b), dashed lines correspond to the mean values calculated with all the simulations performed, in each case (550). In (c) and (d), dashed lines represent linear regressions of the Min and Max values.

and 60 (61 defects). From 70 simulations in the sets, the mean values of the number of defects are well stabilized in the Symmetry case (*i.e.* Fig.2(b)): the mean values range from 62 to 65 defects. It seems to be the case as well for the Random case of Fig.2(a) but the mean value for the set size 100 suffers from a non negligible decrease compared to the previous value (change of 5 defects between the set of size 90 and the set of size 100). From sets of size 30, the distributions of Fig.2(c) and Fig.2(d) seem to be consistent one with each other. The distributions of the sets made of 80, 90 and 100 simulations for the Symmetry case have almost equal median (64.5, 64.5 and 65.0 respectively) and third quartile value (76.0, 75.0 and 75.0 respectively) whereas for the same set sizes for the Random case, the box plots still show some more important discrepancies.

Both in the Random and Symmetry case, the Min and Max values of the box plots are respectively smaller and greater for large sets than for small sets (see linear regressions *i.e.* dashed lines in Fig.2(c) and Fig.2(d)). The

simulations having their Min and Max values far from the central box not being very frequent, it is more likely to observe such simulations with larger sets of simulations. Also, the extreme cases depicted by the Min and Max simulations corresponding to worst cases (PKA and SKAs channelling) scenarios or best cases (PKA and SKAs shock very close neighbor atoms) scenarios, they contain important physical information. Indeed, the worst case scenario corresponds to a cascade going deep into the material with defects distributed all along the trajectories of the PKA and SKAs and the best case scenario depicts a cascade with a small but dense region of defects. To correctly sample the Min and Max cases, it is needed to resort to large sets of simulations.

The evolution of the error coefficient ( $\text{SEM}/\mu$ ) of the number of defects of Fig.3 gives valuable information regarding the convergence of the sets. The error coefficient is very similar for both the Random and Symmetry cases and decreases from 0.09-0.10 to 0.03-0.04, with a faster decrease for smaller sets. Thus, the uncertainty in the cal-

calculation of the mean value decreases by a factor of about three when going from a set of size 10 to a set of size 100.

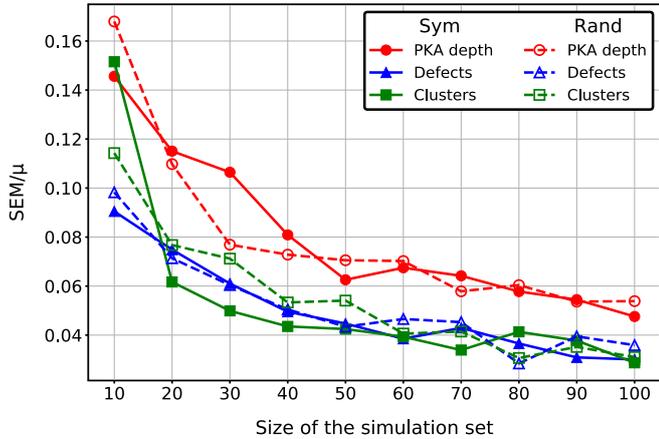


Figure 3: Ratio of the Standard Error of the Mean SEM to the mean value  $\mu$  of the number of defects, clusters and PKA penetration depth for collision cascades initiated by PKAs of 1 keV in Si in the Random (dashed lines) and Symmetry (full lines) cases with respect to the size of the simulation set. Lines are only guide to the eye.

Considering the better stability of the box plots in the Symmetry case (Fig.2(d)) than in the random case (Fig.2(c)) and the good consistency of the mean values of Fig.2(b) for large simulation sets, the Symmetry method appears to be more efficient than the Random one to make the number of defects converge for 1 keV cascades in Si. However, even with the Symmetry method, sets made of a minimum of 70 simulations need to be employed in order to stabilize both the mean values and distributions.

### 3.2. Convergence of the number of clusters for 1 keV PKAs

Results provided in this section are obtained for PKAs of 1 keV in Si. The mean values of the number of clusters in the Random and Symmetry cases for PKAs of 1 keV in Si are respectively displayed in Fig.4(a) and Fig.4(b). The box plots of the number of clusters in the Random and Symmetry cases for PKAs of 1 keV in Si are respectively shown in Fig.4(c) and Fig.4(d).

The variations in the mean number of clusters are more visible in the Random case (Fig.4(a)) than in the Symmetry case (Fig.4(b)): the mean number of clusters goes from 5.5 to 6.5 and from 6.0 to 6.5 respectively. The distributions of the Symmetry case of Fig.4(d) show the striking feature of all having the same median value (6), eight out of ten sharing the same first quartile value (5) and the same third quartile value (7) for seven out of ten. In comparison the distributions of Fig.4(c) show much less consistency, with median values ranging from 5 to 7 and only six out of ten showing the same median value of 6 for the Random case. Just like for the defects, the Min and Max value of the box plots are respectively smaller and greater for large sets than for small sets.

The evolution of the error coefficient for the number of clusters in Fig.3 shows the same trend than for the previously commented error coefficient for the number of defects: it decreases quickly for low values of set size and slower and slower as the set size increases. It means that the accuracy in the calculation of the mean values of the number of clusters substantially increases with the size of the set increasing.

Considering the very consistent box plots for the number of clusters with the Symmetry method of Fig.4(d) and the very small variations in the mean values of the number of clusters with the Symmetry method (Fig.4(b)), this method eases the convergence, in terms of reproducibility, of the number of clusters compared to the Random one. The number of clusters appears to need a smaller set size to be already converged properly than the number of defects. From 60 simulations with the Symmetry method, for collision cascades of 1 keV in Si, the number of clusters already shows excellent stability in terms of distributions and mean values.

### 3.3. Convergence of the PKA depth for 1 keV PKAs in Si

Results provided within this section are obtained with 1 keV PKAs in Si. Fig.5(a) and Fig.5(b) respectively show the mean values of the penetration depth of 1 keV Si PKAs in Si in the Random and Symmetry cases with respect to the size of the simulation sets. Fig.5(c) and Fig.5(d) display the box plots of the PKA penetration depth in the Random and Symmetry cases respectively, with respect to the size of the simulation sets.

In the Random case, the mean values of Fig.5(a) and distributions of Fig.5(c) show significant variations up to set sizes of 50 simulations: mean values range from 22.6 Å to 33.4 Å. From 50 simulations, the mean values stabilize (but still show some discrepancies) between 24.1 Å and 28.5 Å and the distributions show a lot less variations.

In the Symmetry case, the mean values of Fig.5(b) show great consistency for set sizes of 10 to 40. Within the same range the distributions of Fig.5(d) also show remarkable stability for such small set sizes. However, this consistency in the results at small set sizes cannot be attributed to an early reproducibility of the results. Indeed, from 50 simulations, the mean values are all shifted up compared to the mean values obtained from smaller set sizes. The same happens for the distributions which are all shifted up. To us, this artifact of convergence observed for sets of 40 simulations and less results from the stochasticity of the cascades and is therefore coincidental. It also proves that assessing the reproducibility and thus the convergence of the results is not an easy task. From set sizes of 50 simulations, the mean values show excellent consistency: they are all comprised between 24.3 Å and 26.2 Å.

Concerning the box plots, the median values show some variations, but the overall shapes are consistent. This combined with the good stability of the mean values of Fig.5(b) point towards the fact that reproducibility is attained. Also, once again, both in the Random and Sym-

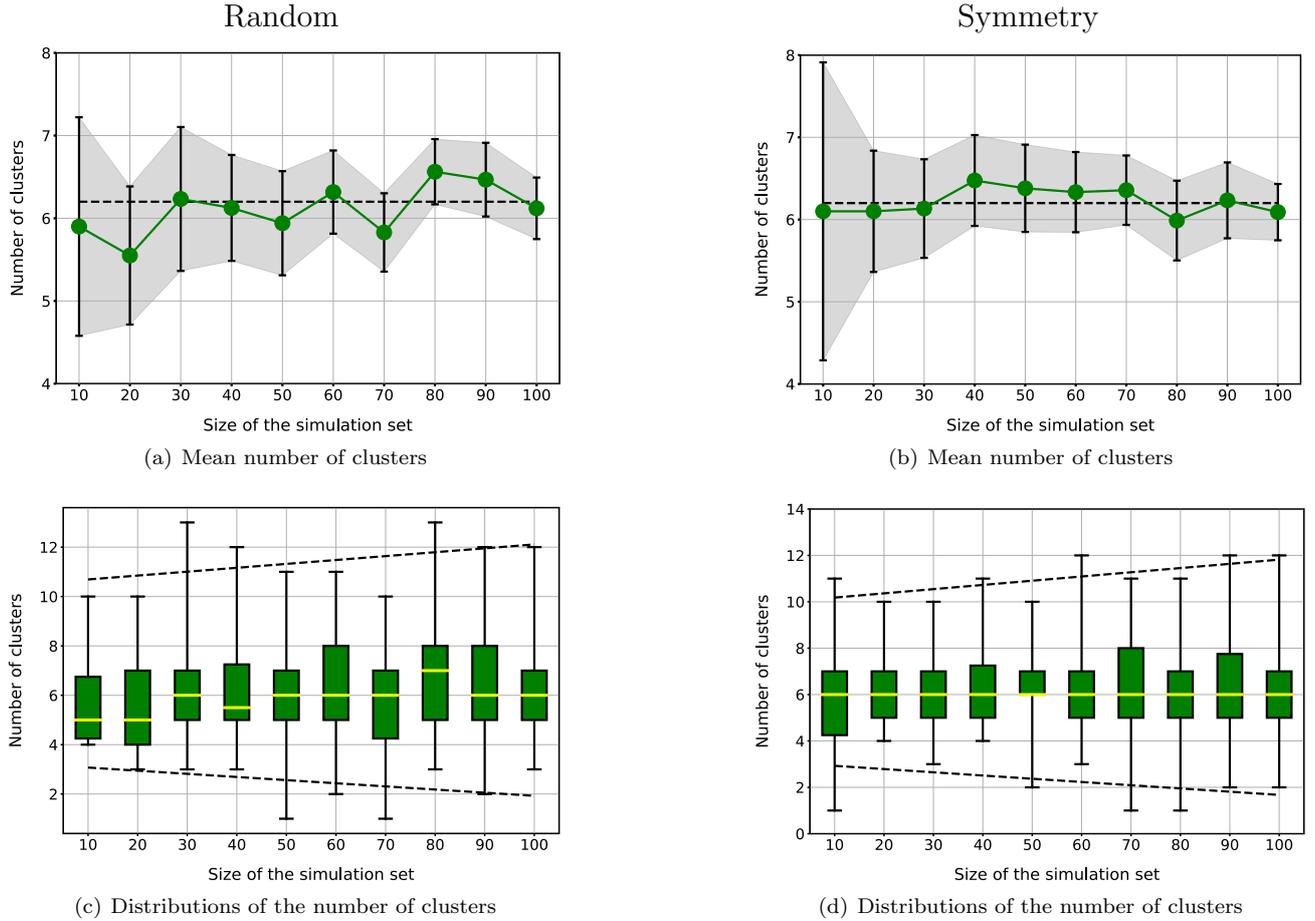


Figure 4: Mean number of clusters in the (a) Random and (b) Symmetry cases and box plots of the number of defects in the Random (c) and Symmetry (d) cases for 1 keV Si PKAs in Si with respect to the size of the simulation sets (*i.e.* the number of simulations in the sets). In (a) and (b), for each size of set, the error bars correspond to  $-1.96 \times \text{SEM}$  and  $1.96 \times \text{SEM}$ . The probability that the real mean value sits within the error bars range is equal to 0.95. The region of the graphs delimited by the error bars of each set size is shaded. Solid green lines are guides to the eye. In (a) and (b), dashed lines correspond to the mean values calculated with all the simulations performed, in each case (550). In (c) and (d), dashed lines represent linear regressions of the Min and Max values.

metry case, the Min and Max value of the box plots are respectively smaller and greater for large sets than for small sets.

The analysis of Fig.3, which displays the evolution of the error coefficient of all the properties of interest in the Random and Symmetry cases reveals that the error coefficient is comparable in the Symmetry and Random case. However, even if showing the same decrease rate, the error coefficient of the PKA depth (Fig.3) is noticeably higher than the ones of the number of defects and clusters. It is approximately of 0.15 when the error coefficient for the other properties is around 0.10 (set size of 10) and of 0.05 when the error coefficient for the other properties is around 0.03 (set size of 100). It reveals that the PKA penetration depth mean values are more subject to uncertainties and thus more difficult to converge than the number of clusters and defects: if one wants to reach an error coefficient of 0.06, about 30 simulations are needed for the number of clusters and defects whereas about 80 simulations are needed for the PKA penetration depth. Another argu-

ment in favor of this statement is the comparison of the mean values obtained with all the simulations performed in the Random and Symmetry cases (550 simulations in each case). To our knowledge, no collision cascades simulations mean values have ever been calculated on ensembles that large. For the number of defects we obtain 65 and 64 in the Random and Symmetry case respectively (see 2) and for the number of clusters 6.2 (see 4) in both cases. Those values, obtained with two completely independent sets of simulations, are very close and indicate excellent convergence for very large sets. In the case of the PKA depth, the mean values obtained are 25.8 Å and 24.4 Å in the Random and Symmetry case respectively (see 5). Even if close, the discrepancy between those two values is by far more important than for the number of defects and clusters. Even with sets as large as 550 simulations, the mean PKA depth values show some variations. Additionally to the reduced accuracy of the PKA depth property compared to the number of defects and of clusters observed in Fig.3, PKA depth mean values are more difficult to repro-

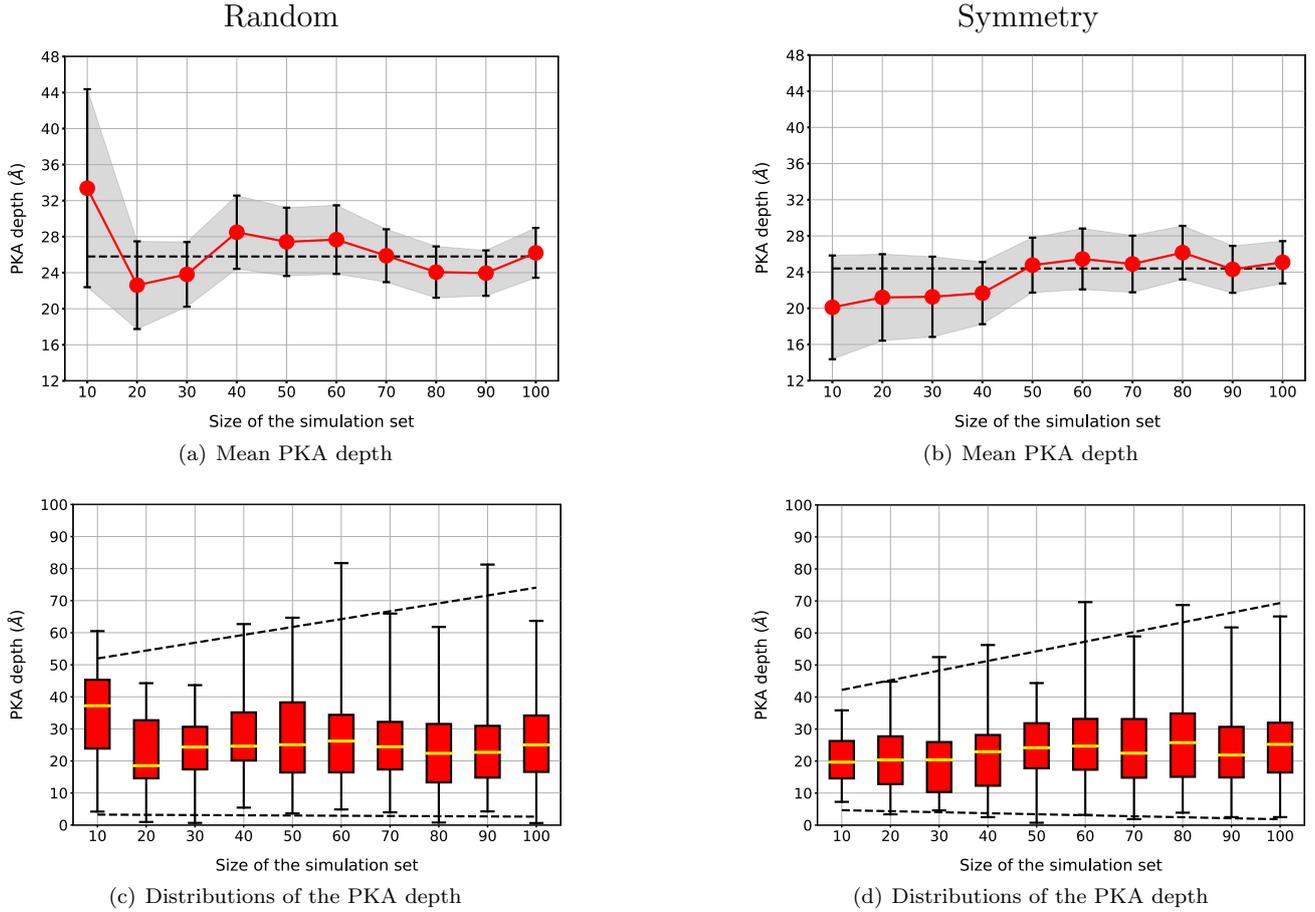


Figure 5: Mean PKA depths in the (a) Random and (b) Symmetry cases and box plots of the number of defects in the Random (c) and Symmetry (d) cases for 1 keV Si PKAs in Si with respect to the size of the simulation sets (*i.e.* the number of simulations in the sets). In (a) and (b), for each size of set, the error bars correspond to  $-1.96 \times \text{SEM}$  and  $1.96 \times \text{SEM}$ . The probability that the real mean value sits within the error bars range is equal to 0.95. The region of the graphs delimited by the error bars of each set size is shaded. Solid red lines are guides to the eye. In (a) and (b), dashed lines correspond to the mean values calculated with all the simulations performed, in each case (550). In (c) and (d), dashed lines represent linear regressions of the Min and Max values.

duce even with very large sets.

Considering the box plots and mean values of Fig.5, no method seems to provide with better convergence results than the other method. The mean values and distributions are stabilized from sets of size 70, even if the consistency is less obvious than with the number of defects or clusters. If the question of a compromise is not relevant for the number of clusters and defects as the error coefficient is already very low and the distributions and mean values stabilized for set sizes of 70 simulations and above for both properties, it is an interesting question for the PKA depth. Indeed, an error coefficient on the PKA depth of 0.04 is not even attainable with a set of 100 simulations. By choosing the number of simulations to be performed, one somehow chooses the accuracy of the mean values conclusions are built from.

On the whole, the analysis of the results of Fig.2, Fig.3, Fig.4 and Fig.5 allows us to draw some conclusions on the convergence of the outputs of cascades initiated with 1 keV PKAs in Si:

(i) Cascades results constructed from sets of made of less than 60 simulations show too much variations and too big error coefficients to draw conclusions.

(ii) The results obtained with the Symmetry method are subject to less variations in both the mean values and distributions, especially for small set sizes.

(iii) Accuracy on the estimation of mean PKA depth values is reduced compared to the accuracy of estimated mean values of other properties. The production of defects and clusters being global effects, the distributions are well centred around their mean values (variation coefficient  $\sigma/\mu$  about 0.3). As an illustration, a PKA in a channeling direction usually produces less defects than the mean (it loses more energy to the electrons through the TTM) but the production of the defects being global, *i.e.* also due to SKAs and other atoms, a large part of the cascade behaves "as usual". On the contrary, the PKA depth is drastically affected by a slight change in the PKA direction. The PKA depth distributions are therefore very dispersed (variation coefficient  $\sigma/\mu$  about 0.5): the distributions tend to show

more variations and the errors in the mean value ( $SEM/\mu$ ) greater than for the number of defects and clusters.

(iv) Large sets are needed to correctly sample the Min and Max cases, *i.e.* the best and worst cases scenarios. Indeed, channeling and PKAs shocking very close atoms being "rare events", a large number of simulations are needed to observe a sufficient number of those events for the Min and Max values to be consistent between all the tested set sizes.

### 3.4. Convergence of collision cascades results with 5 keV PKAs in Si

Considering the results of sections 3.1, 3.2 and 3.3 which proved that the Symmetry method is more adequate, within this section, only the Symmetry method is employed. Fig.6 shows the evolution of the error coefficient for all the properties studied in the Symmetry case for 1 keV and 5 keV PKAs. Fig.7(a), Fig.7(c) and Fig.7(e) respectively display the mean values of the number of defects, of the number of clusters and of the PKA penetration depth for cascades of 5 keV in Si. Fig.7(b), Fig.7(d) and Fig.7(f) respectively display the box plots of the number of defects, of the number of clusters and of the PKA penetration depth for cascades of 5 keV in Si.

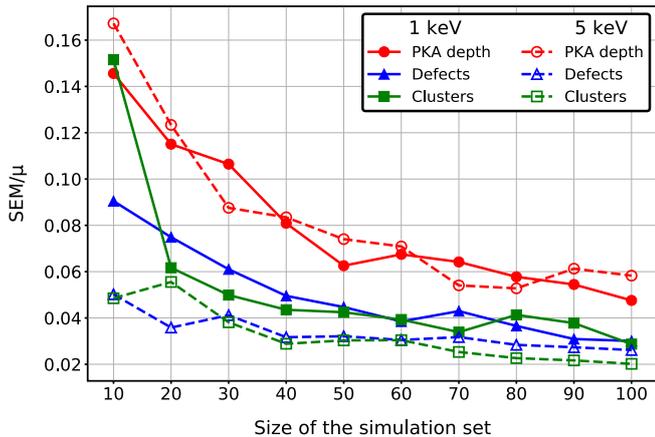


Figure 6: Ratio of the Standard Error of the Mean (SEM) to the mean value  $\mu$  of the number of defects, clusters and PKA penetration depth for collision cascades initiated by PKAs of 1 keV and 5 keV in Si with respect to the size of the simulation set. Lines are only guide to the eye.

The mean numbers of defects of Fig.7(a) show significant variations for set sizes up to 60: it ranges from 287 to 306 defects. For larger sets, the data still do not show clear sign of convergence: variations of about 3-5 defects between all adjacent mean values from 60 simulations in the sets to 100 simulations in the sets are observed. However, with respect to the number of defects at stake (around 300) the variations between each set size are in the order of 2% only: if no perfect consistency is reached, it does not jeopardize the interpretation of the results. The distributions of the defects in Fig.7(b) show great consistency

already from a set size of 60. It supports our thinking that the variations in the mean values of Fig.7(a) are therefore not the result of a bad convergence but simply are due to slight variations in the distributions.

The evolution of the mean PKA penetration depth of Fig.7(f) is very similar to the one of the mean number of defects commented above: strong variations for small set sizes, smaller variations but still no obvious convergence at larger set sizes. Just like in the case of the mean number of defects the variations are very small for set sizes above 70 (about 4% at most) and does not prejudice the interpretation of the results. Moreover, the range of the error bars is larger than the variations in the mean values, thus reducing the risk in the misinterpretation of the results. The distributions of the PKA depth of Fig.7(f) show significant variations in the min, max, first quartile and third quartile values except for the sets of sizes 90 and 100 which are very similar. For all the set sizes the median values are very consistent and do not show large variations.

The evolution of the mean number of clusters of Fig.7(c) shows small variations with mean values ranging from 19 to 21 clusters. The three last values (for set sizes of 80, 90 and 100) seem pretty well converged: the mean values are of 20.3, 20.0 and 19.6. The box plots of Fig.7(c) show significant variations for small sizes of sets but great consistency for large set sizes (above 60 simulations per sets).

For all the properties investigated, the Min and Max values of the box plots (*i.e.* right column of Fig.7) are respectively always smaller and greater with large sets of simulations than small sets of simulations. The Min and Max value appears to be consistent from 50 simulations for the number of defects, 60 simulations for the number of clusters and 90 simulations for the PKA depth. Again, it highlights that the PKA depth is harder to converge than the number of defects and number of clusters.

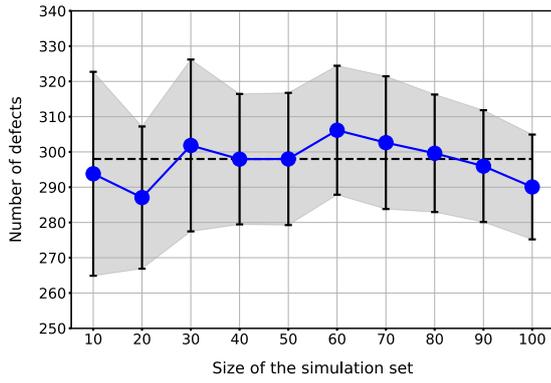
The error coefficient of the PKA depth of Fig.6 are similar in the 1 keV and 5 keV cases: decrease from around 0.15 to approximately 0.05. However, the error coefficients of the number of defects and clusters in the 5 keV case are lower than in the 1 keV case.

On the whole, for 5 keV cascades in Si, it appears that:

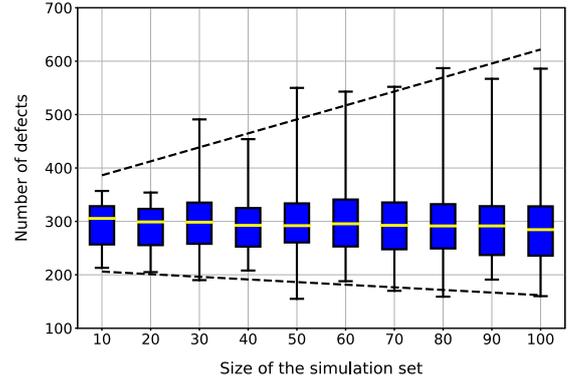
- (i) Stability in the distributions can be attained when no obvious stability in the mean values is reached.
- (ii) The mean values should always come with the SEM value in order not to misinterpret the results.
- (iii) Reproducibility between the sets is harder to attain with 5 keV cascades than with 1 keV cascades.

### 3.5. Influence of the initial position of the PKA

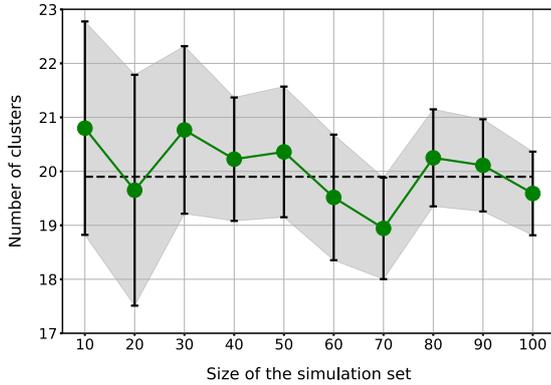
After having studied the convergence of the number of defects and of clusters and of the PKA penetration depth for 1 keV and 5 keV PKAs in Si, one may ask if the initial position of the PKA has an influence. We seek to answer two questions in this section: (i) does the initial site of the PKA have an influence on the cascades outputs? (ii)



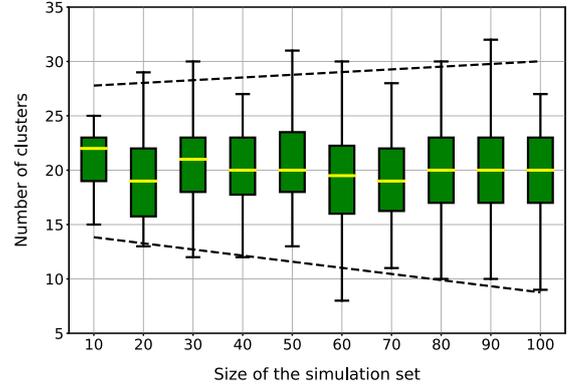
(a) Mean number of defects at the end of the collision cascades.



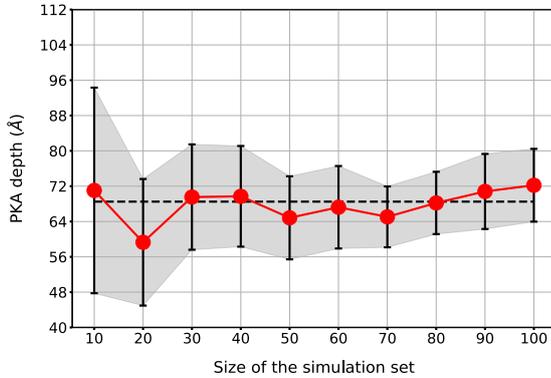
(b) Distributions of the number of defects at the end of the collision cascades.



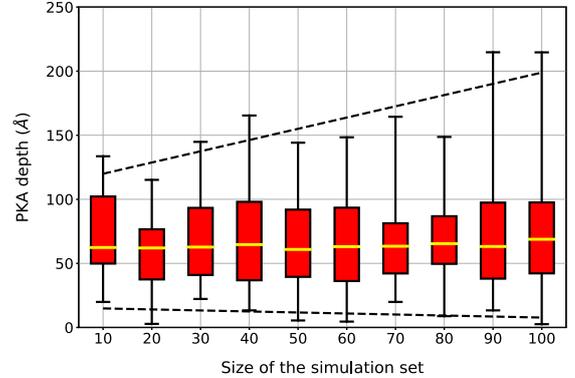
(c) Mean number of clusters at the end of the collision cascades.



(d) Distributions of the number of clusters at the end of the collision cascades.



(e) Mean penetration depths attained during the cascades by the PKAs.



(f) Distributions of the penetration depths attained during the cascades by the PKAs.

Figure 7: Mean (a) and box plots (b) of the number of defects, (c), (d) number of clusters and (e), (f) PKA penetration depth in collision cascades initiated by 5 keV Si PKAs in Si with respect to the size of the simulation set. For (a), (b) and (c), solid lines are guides to the eye. Dashed lines in (a), (c) and (e) correspond to the mean values obtained from all the simulations performed (550). Dashed lines in (b), (d) and (f) correspond to linear regressions of the Min and Max values.

do the two lattice sites of Fig.1 need to be sampled with the Symmetry method? We highlight the fact that drawn conclusions of this part are only valid for mono-atomic materials in the diamond-like structure.

We ran four sets of 10 simulations and four sets of 100 simulations with 1 keV PKAs in Si with the Symmetry method. Four distinct initial PKA positions are sampled.

The first two positions (named site #1 and site #1 bis) correspond to the first site of the primitive diamond cell (lower site of Fig.1), the two other positions (named site #2 and site #2 bis) correspond to the second site of the primitive diamond cell (upper site of Fig.1). Site #1 and Site #1 bis (respectively Site #2 and Site #2 bis) occupy the same site in their own primitive unit cell, but were

Table 1: Mean values of the number of defects, number of clusters and PKA depth for 1 keV cascades in Si from sets of sizes 10 and 100. To assess the influence of the initial position of the PKA, both sites of the primitive diamond cell (see Fig.1) are sampled twice: for each of the two sites, two initial positions are chosen in the simulation box. In parenthesis are given the SEM values.

	PKA position	Mean defects	Mean clusters	Mean PKA depth (Å)
10 simulations	Site #1	63 (6)	6 (0.9)	20.1 (2.9)
	Site #1 bis	62 (5)	6 (0.6)	16.7 (2.0)
	Site #2	59 (5)	6 (0.6)	26.8 (4.4)
	Site #2 bis	60 (7)	5 (0.6)	28.7 (4.7)
100 simulations	Site #1	65 (2)	6 (0.2)	25.1 (1.2)
	Site #1 bis	63 (2)	6 (0.2)	25.9 (1.4)
	Site #2	64 (2)	6 (0.2)	24.4 (1.2)
	Site #2 bis	66 (2)	6 (0.2)	24.6 (1.2)

chosen in in the following simulation in different primitive cells. Table 1 displays the mean values of the number of defects, number of clusters and PKA depth for the sets of 10 and 100 simulations in the four sampled positions (Site #1, Site #1 bis, Site #2 and Site #2 bis).

Considering the four sets made of 10 simulations, the mean number of defects and clusters are coherent between all the sampled PKA positions: the mean number of defects ranges from 59 to 63 and the mean number of clusters from 5 to 6. However, the SEM values are quite large (expected knowing the error coefficient graph of Fig.3) which reveals the mean values suffer from big uncertainties. The most striking feature is the significant variations in the mean PKA depth. It ranges from 16.7 Å to 28.7 Å. The initial PKA position has a drastic influence on the mean PKA depth. Moreover, it appears that the primitive cell site on which the PKA is initially located plays an important role: mean PKA depth of 16.7 Å and 20.1 Å for sites #1 and of 26.8 Å and 28.7 Å for sites #2.

Considering the four sets made of 100 simulations, the mean number of defects and clusters and, contrary to the sets of size 10, the mean PKA depth show excellent coherence between the four PKA positions sampled: the mean number of defects ranges from 63 to 66, the mean number of clusters is equal to 6 for all the simulations and the mean PKA depth ranges from 24.4 Å to 25.9 Å. The large variations observed with sets of size 10 disappear with sets of 100 simulations. The site of the primitive cell of the diamond structure does not influence the results as it does with sets of size 10. The SEM values are also a lot smaller than in the 10 simulations cases and show great consistency. Therefore, with sets of size 100, the influence of the initial position of the PKA is not an issue anymore.

#### 4. Conclusion

The present study is the result of more than 2 000 MD simulations of collision cascades in Si. We find that a method based on symmetry considerations for the determination of the initial directions of the PKA improves the convergence and the consistency of the results in Si. It is also shown that in Si, it is harder to converge the PKA penetration depth than the number of defects and clusters.

To obtain satisfactory convergence of the PKA depth and of the mean number of defects and clusters in Si, sets of simulations made of at least 70 simulations are needed even for PKA energies as low as 1 keV. At 5 keV, convergence as satisfactory as what is obtained at 1 keV is not yet obtained with sets of 100 simulations but the consistency of the results is enough to be able to draw meaningful conclusions on the data obtained with sets of around 80 simulations. It appears that the higher the energy of the PKAs, the bigger the sets must be. The simulations corresponding to best case and worst case scenarios being consistent between the size of sets starting from 60 simulations for 1 keV PKAs and 80 simulations for 5 keV PKAs in Si, large sets are also necessary to correctly sample the range of possible scenarios. Additionally, the initial position of the PKA is not found to have any influence on the results, at least at 1 keV, if the simulation sets are large enough. Finally, taking as example 1 keV and 5 keV PKAs in Si, more than only providing with a method to cope with the stochasticity of the cascades, this article proves the need to treat the data related to cascades in a statistically rigorous way: a convergence study is mandatory, the mean values should always come with the SEM value and the distributions be graphically represented.

#### Acknowledgments

The authors would like to thank Layla Martin-Samos from CNR-IOM for fruitful discussions. This article is based upon work from COST Action TUMIEE CA17126, supported by COST (European Cooperation in Science and Technology). Calculations have been performed using HPC resources from GENCI-CCRT supercomputer at CEA, DAM, DIF, HPC resources from GENCI (Grant A0030907474) and HPC resources from CALMIP (Grant 1555). T. Jarrin, A. Jay, A. Hémercyck and N. Richard are active members of the Multiscale and Multi-Model Approach for Materials in Applied Science consortium (MAM-MASMIAS consortium), and acknowledge the efforts of the consortium in fostering scientific collaboration.

## References

- [1] K. Nordlund, M. Ghaly, R. S. Averback, M. Caturla, T. Diaz de la Rubia, and J. Tarus, *Phys. Rev. B* **57**, 7556 (1998).
- [2] K. Nordlund, S. Zinkle, S. A.E, F. Granberg, A. R.S, R. Stoller, T. Suzudo, L. Malerba, F. Banhart, W. Weber, F. Willaime, S. Dudarev, and D. Simeone, *Phys. Rev. B* **512**, 450 (2018).
- [3] M. Robinson, *Rad. Eff. Def. Sol.* **null**, 3 (1994).
- [4] N. Trung, H. Phuong, and M. Starostenkov, *Lett. Mat.* **9**, 168 (2019).
- [5] H. He, C. He, J. Zhang, W. Liao, H. Zang, Y. Li, and W. Liu, *Nucl; Eng. Tech.* (2019), 10.1016/j.net.2019.12.027.
- [6] J. Buchan, M. Robinson, H. Christie, D. Roach, D. Ross, and N. Marks, *J. App. Phys.* **117**, 245901 (2015).
- [7] E. Zarkadoula, G. Samolyuk, H. Xuen, H. Bei, and W. J. Weberab, *Scr. Mat.* **124**, 6 (2016).
- [8] F. Gao, N. Chen, E. Hernandez-Rivera, D. Huang, and P. Levan, *J. App. Phys.* **121**, 095104 (2017).
- [9] H. Christie, M. Robinson, D. Roach, D. Ross, I. Suarez-Martinez, and N. Marks, *Carbon* **81**, 105 (2015).
- [10] M. Warrior, U. Bhardwaj, H. Hemani, R. Schneider, A. Mutzke, and M. Valsakumar, *J. Nucl. Mat.* **467**, 457 (2015).
- [11] R. Voskoboinikov, *Nucl. Instr. Meth. Phys. R. B* **479**, 18 (2020).
- [12] A. Jay, M. Raine, N. Richard, N. Mousseau, V. Goiffon, A. Hémercyck, and P. Magnan, *IEEE Trans. Nucl. Sc.* **64**, 141 (2017).
- [13] T. Jarrin, A. Jay, M. Raine, N. Mousseau, A. Hémercyck, and N. Richard, *IEEE Trans. Nucl. Sc.* **67**, 1273 (2020).
- [14] T. Jarrin, A. Jay, A. Hémercyck, and N. Richard, accepted for publication in *Nucl. Instr. Meth. Phys. R. B* (2020).
- [15] S. Plimton, *J. Comput. Phys.* **117**, 1 (1995).
- [16] F. H. Stillinger and T. A. Weber, *Phys. Rev. B* **31**, 5262 (1985).
- [17] J. Ziegler, J. Biersack, and U. Littmark, *The Stopping Range of Ions in Solids*, 1st ed. (Pergamon Press, N.Y, 1983).
- [18] N. Chen, E. Rasch, D. Huang, E. Heller, and F. Gao, *IEEE Trans. Nucl. Sc.* **65**, 1108 (2018).
- [19] D. Duffy and A. Rutherford, *J. Phys.: Cond. Matter.* **19**, 1 (2007).
- [20] H. Hensel and H. M. Urbassek, *Phys. Rev. B* **57**, 4756 (1998).
- [21] C. Kittel, *Introduction to Solid State Physics*, 8th ed. (John Wiley and Sons, 2004).