Using alpha Hulls to Automatically and Reproducibly Detect Edge Clusters in Atom Probe Tomography Datasets

Benjamin M. Jenkins[a]*, Andrew J. London[b]*, Nick Riddle[c], Jonathan M. Hyde[a,d], Paul A. J. Bagot[a], Michael P. Moody[a]

[a] - Dept. of Materials, Parks Road, Oxford, OX1 3PH, UK

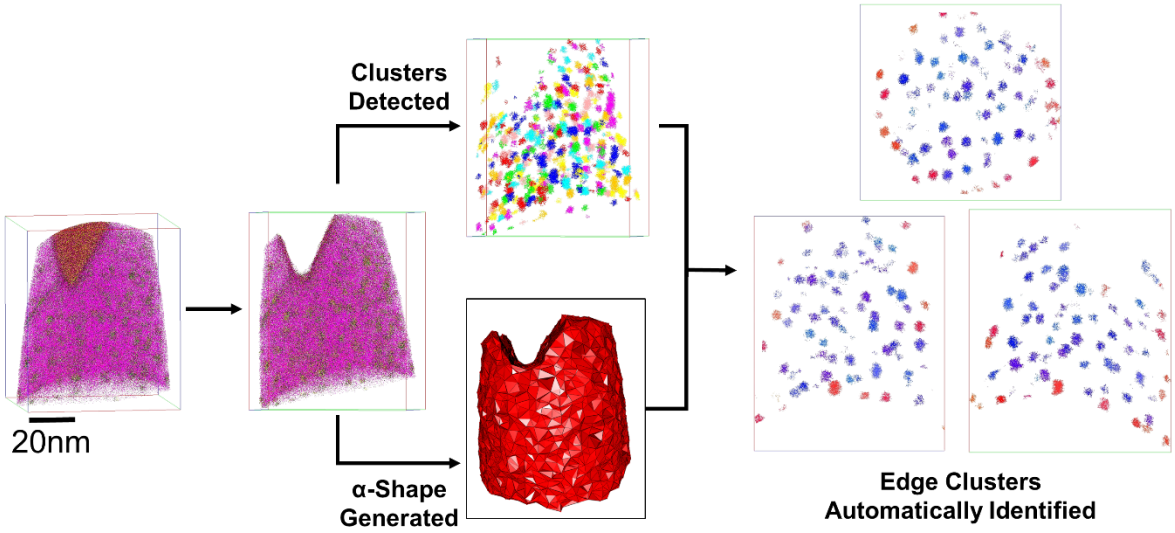
[b] - UKAEA, Culham Science Centre, Abingdon, Oxfordshire, OX14 3DB, United Kingdom

[c] – Rolls Royce, PO Box 2000, Raynesway, Derby DE21 7XX, UK

[d] – National Nuclear Laboratory, Building D5, First Floor, Culham Science Centre, Abingdon, Oxfordshire, OX14 3DB

* Corresponding authors: benjamin.jenkins@materials.ox.ac.uk (B. M. Jenkins), andy.london@ukaea.uk (A. J. London)

Graphical Abstract



Highlights

- Novel method developed for automatically detecting edge clusters in atom probe datasets
- Application to simulated data and its effect on measurement accuracy is assessed
- Successfully applied to real data – including datasets with complex shapes

Abstract

An automated way to accurately and reproducibly identify edge clusters within atom probe tomography datasets has been developed. The alpha-hull algorithm is used to generate a concave alpha-shape around an atom probe dataset. Information from core-linkage cluster searches is used, in combination with the calculated alpha-shape, to determine which clusters are on the edge of the dataset.

The potential effects that not removing edge clusters may have on calculated cluster sizes, number densities and compositions is discussed. The viability of the methodology is demonstrated via application to real datasets, one of which was a non-standard shape.

The sensitivity of the method to user parameter selection is explored. Sampling fractions > 0.1 % and an alpha value, used to make the alpha shape, greater than twice the maximum measured nearest neighbour distance were found to be suitable.

Keywords

Atom probe tomography, cluster analysis, alpha hull, precipitates

1 Introduction

The presence of nanometre-scale solute clusters and precipitates can greatly affect macroscopic mechanical properties in a range of material systems. For example, solute clustering in aluminium alloys has been observed to lead to significant changes in alloy hardness[1] and precipitate formation during operation is responsible for embrittlement of reactor pressure vessel steels[2]. The magnitude to which the presence of these features affects the material's properties depends upon their size distribution, chemical composition, and number density. Accurate determination of these parameters is essential for the reliable prediction of a material's properties.

Atom probe tomography (APT) is a three-dimensional characterisation technique that, due to its high chemical and spatial resolution, is routinely used to characterise material systems containing solute clusters[3]. Solute clusters can be identified within APT data by the application of different cluster search algorithms, the most commonly used are based on the maximum separation method [4] [5]. Once solute clusters have been located, their size distribution, composition and number density can be determined.

However, some of the features in the original material are likely to not be wholly contained within the volume sampled during the APT experiment. Only a fraction of the original feature will therefore be detected in cluster searches of the dataset. These clusters are likely to influence and bias the measured cluster number density and size distributions, they may also affect compositional measurements if the features display a core-shell structure. It is therefore important not to include these partial "edge" clusters when calculating the aforementioned values.

Many publications that characterise solute clusters via APT do not state if edge clusters were identified and removed before further analysis. Authors that do state that edge clusters were accounted for[6–12] fail to state how edge clusters were identified. This makes it difficult for other researchers to reproduce experimental results. There is, therefore, the need for a protocol that can be consistently applied by researchers to identify edge clusters within APT datasets.

It is desirable that any protocol to identify edge clusters can be performed in a reasonable time. The large number of clusters that can be present in a single APT dataset means that it may take a prohibitively long time to manually identify edge clusters. As a result of this, and to remove subjectivity in the identification of edge clusters, it is preferred to develop an automated approach.

A novel methodology has been developed to reproducibly and automatically identify solute clusters that are on the edge of the volume sampled during an APT experiment. The method was applied to a series of simulated datasets, and the calculated cluster size distribution, number density and composition were compared to their known values before and after edge cluster removal. The approach was then used to identify edge clusters in two real APT datasets, one of which was irregularly shaped after a region of interest containing a carbide had been removed from it.

2 Materials and Methods

2.1 Simulated Data

To investigate the validity of the method proposed in Section 2.5 Method For Detecting Edge Clusters, two sets of simulated data were generated and analysed. The simulated data took the form of a spherical volume with radius *R* = 20 nm and atoms were randomly located such that the overall ionic density was 20 ions/nm^3. Randomly placing the ions in the simulated volume mimics the limited spatial resolution commonly observed in experimental APT datasets. An atomic density of 20 ions/nm^3 is similar to the atomic density observed in APT datasets, after accounting for the limited detection efficiency. Initially, solute clusters were generated in this volume with a uniform concentration, however, in a second subsequent simulation, clusters with a core-shell structure were produced. In both cases, 40 clusters with a radius *r* = 2 nm were placed in a uniform random distribution with centres within 22 nm of the centre of the spherical volume. The clusters had a composition of 50 at.% A and 50 at.% B (where A and B were both solute atoms). The matrix solute concentration of 0.2 at.% was sufficiently low so as to minimise the chances of neighbouring clusters being connected by the cluster search algorithm.

The posgen program [13] was used to generate the ion position (pos) files from XML scripts written by a custom script written in R [14]; these scripts are available in the supplementary material. Clusters were placed with a separation margin of at least 1 nm to prevent their volumes overlapping. Any cluster placed at a distance from the centre of the dataset between *R-r* and *R+r* will intersect the edge of the spherical volume and therefore will be an edge cluster (Figure 1). Based on a random placement of clusters, this would result in approximately 50 % of the clusters being classified as edge clusters. Posgen was also used to detect the clusters in the volume using the core-linkage [15] variant of the maximum separation method. This algorithm was implemented using a kth nearest neighbour (kNN) order of k = 1, d_{max} = 0.5 and an inclusion envelope (L) equal to an erosion (E) distance of 0.2 nm. N_{min}, the minimum number of solute ions that had to be contained in a cluster for it to be counted, was set equal to 20; this value is much lower than the number of ions (670) expected within a cluster of radius 2 nm. A modified version of the posgen code was used which reported the bounding box of each cluster as well as the composition and centre of mass; a patch for the modified code is included in the supplementary material.

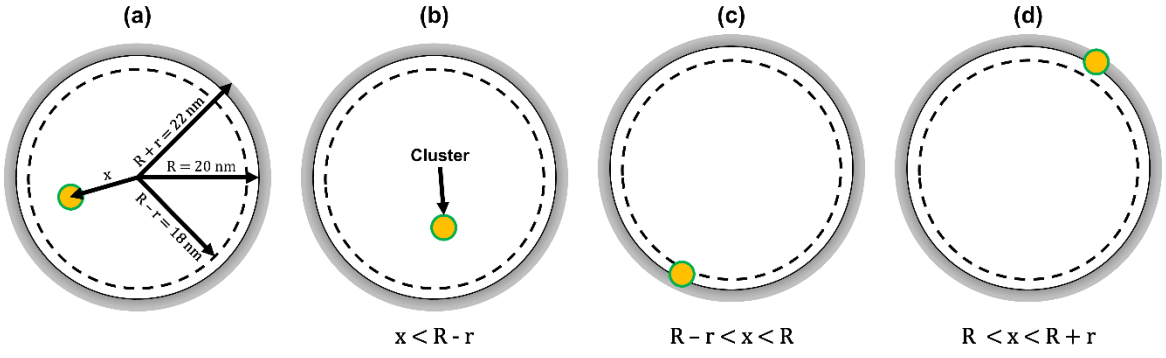


Figure 1: Schematic diagram (2D) of (a) simulated volume and dimensions, (b) Cluster, with centre (x) < R – r, that does not intersect the edge of the dataset, (c) Cluster, with R – r < x < R, that intersects the edge of the dataset and (d) Cluster, which intersects the edge of the datasets, with R < x < R + r. x is the radial distance from the centre of the sphere.

For the core-shell simulation, each cluster was composed of an inner sphere (1.59 nm radius) containing only A ions and an outer shell (radius 2 nm) containing B ions. This results in an average cluster composition of 50 at.% A and 50 at.% B. The same cluster selection parameters were used as above.

The classification of edge clusters was performed using the alpha-hull methodology presented below. For calculating the alpha hull, a sampling fraction of 0.01 and alpha value of 12 were used. In both scenarios above, 2000 random datasets were created and the edge cluster classification was performed.

2.2 Materials

A model steel was exposed to neutron irradiation in the Advanced Test Reactor 2 experiment at the University of California, Santa Barbara. The irradiation was conducted at a temperature of 290°C and the specimen experienced a fluence of $1.4x10^{20}n/cm^2$ at a flux of $3.6x10^{12}n/cm^2s$. The damage was calculated to be equivalent to 0.17 displacements per atom (dpa). The nominal composition of the alloy is shown in Table 1.

Cu	Si	Ni	Mn	P	Cr	Mo	C	Fe
0.04	0.37	3.11	1.90	0.01	0.10	0.27	0.87	Bal.

Table 1 Composition of model steel alloy (at.%)

2.3 Sample Preparation and APT Acquisition

Specimens were prepared for APT analysis using an FEI Quanta 3D FEG. The APT tips were fabricated following standard focussed ion beam lift-out methods[16,17]. A final polishing stage was conducted at 5 kV to limit Ga implantation into the volume analysed during the APT experiment[18].

APT data was acquired using voltage mode on a LEAP 4000-XR with an operating temperature of 50 K, a pulse fraction of 20 %, a pulse frequency of 200 kHz, and a detection rate of 0.003 ions per pulse.

2.4 APT Data Analysis

Reconstructions of the atom probe data were generated using Integrated Visualization and Analysis Software (IVAS) version 3.8.4. Current reconstruction protocol utilised in this work are summarised in detail elsewhere [19,20]. Cluster searches were conducted via the IVAS cluster search tool, which implements the core-linkage method [15]. Cu, Ni, Mn, Si, P were selected as solute ions; the $^{58}Fe^{2+}/^{58}Ni^{2+}$overlap peak at 29 was included in the cluster search [21]. The following cluster search parameters were used: kNN Order = 5-6, d_{max} = 0.55nm, L = E = 0.275nm, N_{min} = 25-30. Cluster search parameters were chosen such that the ratio of real:random clusters was maximised.

2.5 Method For Detecting Edge Clusters

We define edge clusters as any cluster from the original material whose true extent is not completely contained within the reconstructed APT volume. Clusters may be assigned as edge clusters if any of the ions within the detected cluster are on the edge of the analysed volume. However, this can make calculations computationally complex as clusters may contain many thousands of ions. To reduce computational complexity, a convex hull was generated around each cluster. Clusters were then assigned as edge if any of the vertices of the convex hull generated around the cluster were not contained within the analysed volume. This is shown schematically in Figure 2.

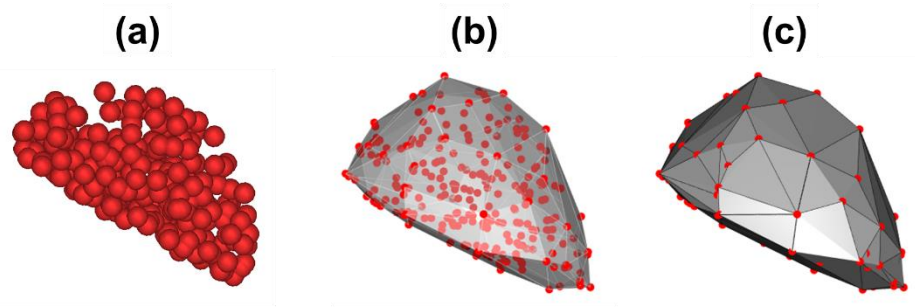


Figure 2 Schematic diagram demonstrating that for (a) each cluster (b) a convex hull is fitted and (c) the vertices that define the hull are identified.

Before one can determine if the vertices of the convex shape lie within the reconstructed volume, the shape of the reconstructed volume itself must be defined. To define the surface of the analysed volume, an alpha shape[22] which contained all points in a downsampled subset of the original .pos file was generated using the alphashape3d (v1.3) package[23] in R (v3.4.2). The alpha-hull allows a concave hull to be generated around a series of points and has been previously used to generate an outline around APT datasets[24].

A key input parameter of the alpha-hull algorithm is the value of alpha (α). This determines the radius of the spheres that are used to generate the alpha-shape around the point cloud. Low values of alpha will lead to many volumes each containing few points. Meanwhile, high values of alpha will generate one volume containing all of the points, however, the shape may be “low resolution” and not accurately reflect the shape of the original point cloud. Figure 3 demonstrates how, when applied to an APT dataset, too small values of alpha (a, e, i) will lead to multiple volumes being created, whilst too large values of alpha (c, d, h, l) lead to a low resolution shape, which may not capture the concave curvature of the original data, being generated. It is important that a reasonable alpha value, one which generates a single volume but with high enough resolution, is used to generate the alpha shape. Note that some sampling of the data is required to enable the alpha shape to be generated within the RAM capacity of a desktop computer. Using an alpha value equal to the largest nearest neighbour distance (NND) between points in the sampled data should ensure that all points within the sampled data are contained within the alpha shape. Using the maximum NND

means the alpha value scales with the selected sampling rate. The effect of sampling, coupled with choice of alpha, value is discussed in Section 4.2.1.

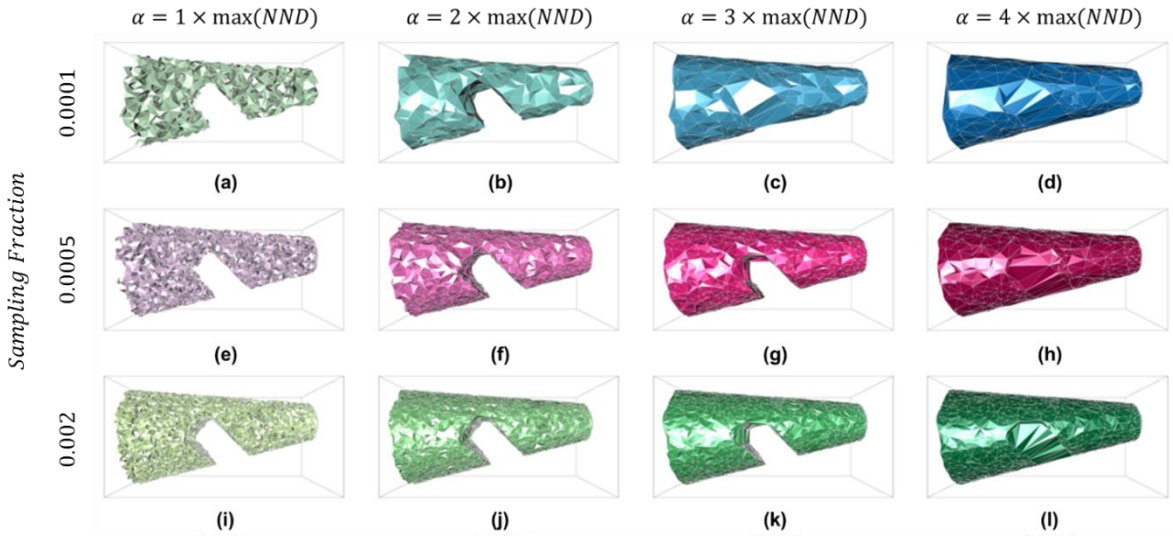


Figure 3 Effect of varying sampling fraction and the alpha value used to generate the alpha-Hull around a dataset with a section removed. (a)-(d): Sampling = 0.0001, (e)-(h): Sampling = 0.0005, (i)-(l) Sampling = 0.002.

Once the alpha surface had been generated, a test was performed to determine if all the vertices of each cluster were contained within the alpha surface. If any were found to lie outside of the surface, the cluster was defined as a potential edge cluster.

A laptop computer with 16GB RAM and an Intel Core i7 CPU 2.60GHz duo core processor was used to generate alpha shapes around the datasets in this study.

3 Theory and Calculation

In an experiment, clusters may, or may not, intersect the edge of the analysed volume. Before calculating the cluster number density, it must first be decided whether calculations will included all clusters detected within the volume, or only those clusters whose centres are within the volume. Since it is not possible to know whether the true centre of a cluster detected in an APT dataset lies within the reconstructed volume, all detected clusters should be utilised in number density calculations.

If all of the detected clusters are used for the calculations and counted with equal weight when estimating the number density, an overestimate will be measured (assuming that all clusters are detected by the search algorithm). If edge clusters can be discriminated from those not on the edge, a more accurate estimate of cluster number, and therefore number density, can be made by multiplying by a correction factor:

$$N_{True} = N_{nonedge} + \kappa N_{edge}$$

where the number of clusters whose true centre is contained within the sampled volume (N_{True}) is given by the number of non-edge clusters ($N_{Non-Edge}$) plus some fraction, κ, of the number of edge clusters (N_{Edge}), where κ is the fraction of edge clusters which have their true centre within the analysed volume. For a simple volume and a known number of clusters, the correction factor (κ) can be calculated by rearranging the previous equation:

$$\kappa = \frac{N_{True} - N_{Non-Edge}}{N_{Edge}}$$

For a sphere radius *R* and clusters radius *r*, if *N* clusters are placed with a centre between *0 < x < R + r*. Clusters with centre between *R – r < x < R + r* will be on the edge of the volume (Figure 1).

The true number density (n_d) should be the number of clusters (N) whose true centre is contained within the whole volume divided by the volume they are placed in:

$$n_d = \frac{3.N}{4\pi(R+r)^3}$$

Therefore, the number of clusters (N_{True}) with their true centre within the sampled sphere radius R is given by the true number density times the sampled volume:

$$N_{True} = \frac{3.N}{4\pi(R+r)^3}.\frac{4}{3}\pi R^3 = \frac{N.R^3}{(R+r)^3}$$

Note that the geometric volume ($4\pi/3$) factor cancels out, meaning this result is true for cubic and other volumes as well. The number of non-edge clusters is simply those clusters with a centre less than the outer radius minus the cluster radius. This is given by the true number density times the inner volume:

$$N_{Non-Edge} = n_d.\frac{4}{3}\pi(R-r)^3 = \frac{3.N}{4\pi(R+r)^3}.\frac{4}{3}\pi(R-r)^3 = \frac{N.(R-r)^3}{(R+r)^3}$$

The number of edge clusters is simply the true number of clusters minus the non-edge clusters:

$$N_{Edge} = N - \frac{3.N}{4\pi(R+r)^3}.\frac{4}{3}\pi(R-r)^3 = N.\left(1 - \frac{(R-r)^3}{(R+r)^3}\right)$$

Then the correction factor above is given by:

$$\kappa = \frac{\frac{N.R^3}{(R+r)^3} - \frac{N.(R-r)^3}{(R+r)^3}}{N.\left(1 - \frac{(R-r)^3}{(R+r)^3}\right)}$$

Which simplifies to:

$$\kappa = \frac{1}{2} - \frac{3rR}{2(r^2+3R^2)}$$

Equation 1

This factor is always less than 0.5 and tends to 0.5 with large R or small r. This agrees with the accepted procedure to count the non-edge clusters and add half of the count of edge clusters when performing number density calculations [25].

4 Results

4.1 Simulated Data

As described in section 2.1, two sets of simulations were made to validate the edge detection method and investigate the influence of edge clusters on the measured values. For both cases the number density, size and volume fraction are the same, only the measured composition is affected by the distribution of the solute within the clusters.

4.1.1 Number density

Table 2: Summary of the mean number of different types of cluster in the simulated dataset. These values were calculated based on the results from 2,000 simulations. Uncertainty quoted is the standard error in the mean (SEM).

Classification	Truth table	Cluster count
Non-edge/Correct	FF	20.6 ± 0.06
Edge but classified non-edge	TF	0.003 ± 0.001
Non-edge but classified edge	FT	0.60 ± 0.02
Edge/Correct	TT	16.0 ± 0.06
Not detected	ND	2.80 ± 0.04

If there was perfect discrimination of whether a cluster's true centre was inside the sampled volume or outside of it, then the number density could be measured exactly, however in practice this is not possible. As detailed in Section 3, clusters on the edge can be counted with a weight to give a more accurate estimate of the cluster count and therefore number density. In the simulated data, the ground truth of whether or not clusters are of edge type is known; this can be compared to the classification made by the alpha-hull method.

A summary of the average number of clusters of each type in each simulation is given in Table 2. On average, the mean number of clusters that intersected the edge of the sampled volume per simulation was 18.8 ± 0.05 (SEM) and for non-edge clusters was 21.2 ± 0.05 (SEM), where SEM is the Standard Error of the Mean. In 2000 simulations, 74,439 clusters were detected by the cluster search algorithm. Of these, 41,271 non-edge clusters and 31,976 edge clusters were correctly identified. 1,192 clusters were falsely identified as being edge while not actually being on the edge of the volume. This represents that the algorithm has a 98.4 ± 0.1 % accuracy with 95 % CI for determining whether detected clusters are edge or not; accuracy given by (TT+FF/total number of detected clusters). Figure 4 shows the classification of the simulated clusters based on their true centre position. Clusters with a radius position less than 17.5 or greater than 18 nm are classified correctly and those with a centre beyond 21.5 nm are rarely detected by the cluster search algorithm. Clusters with a centre between 17.5 and 18 nm are often classified as being edge type. These are false positives as only clusters with a centre position beyond 18 nm can be edge type. Possible reason that these false positives arise is discussed in Section 5.2. Additionally, 5 clusters were not correctly identified as being on the edge by the alpha-edge method when they actually were edge-type clusters. These TF classified clusters all had a true radius position between 18 and 18.2 nm.

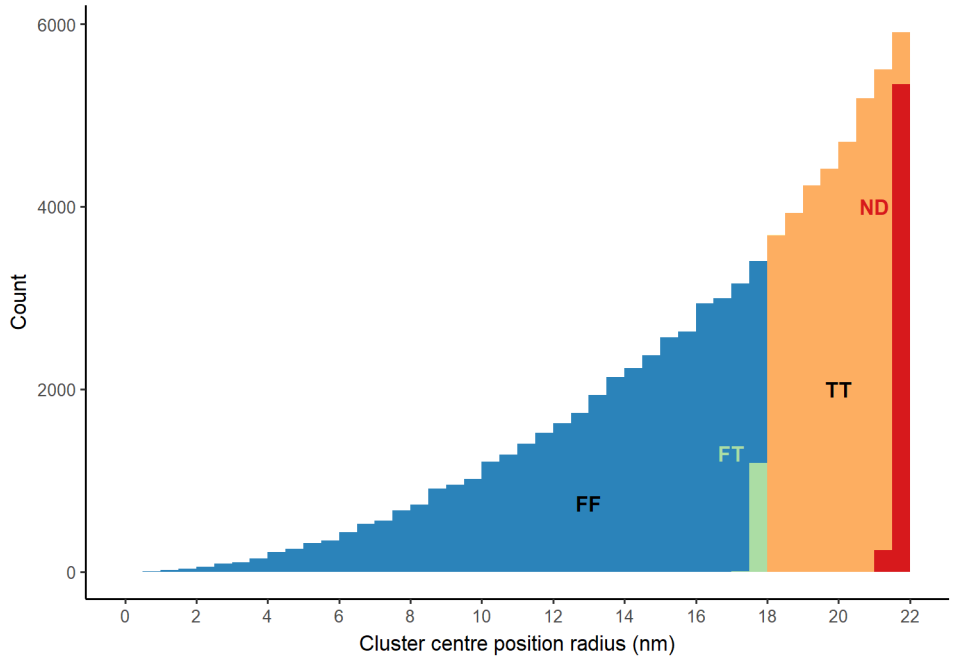


Figure 4: Classification of clusters as a function of their simulated centre position as a radius from the centre of the volume. FF are correctly identified non-edge clusters, FT are non-edge clusters classified wrongly as edge, TT are correctly identified edge and ND were not detected by the cluster search algorithm.

The number density estimated using all detected clusters is 11.1×10^{23} clusters/m^3 which is a 24 % over-estimate compared to the true number density of 8.97×10^{23} clusters/m^3. However, if a correction factor (κ) of 0.5 is applied to the edge clusters, then a better estimate of 8.63×10^{23} clusters/m^3 is given, this is a 4 % underestimate. The underestimate of number density comes from two issues: imperfect detection of some clusters that are present in the original data and classification as edge type. These are discussed in more detail later.

4.1.2 Cluster size

Figure 5 shows a histogram of the measured cluster size of both edge and non-edge clusters that were detected. Clusters which intersect the edge of the volume have a reduced size. The exact distribution of sizes depends on the cluster and volume shape. The non-edge clusters have a binominal distribution of sizes, as measured by number of solute ions, because the clusters were generated from cubes containing a fixed number of randomly positioned atoms, from which spheres of radius 2 nm were cropped. Similarly, Figure 5 shows the false-positive (FT) clusters have a size distribution which matches the true non-edge clusters (FF). The size of the edge clusters can be described exactly by the product of the sphere-sphere intersection volume and the centre placing probability, which increases with radius squared. The detected edge clusters contained, on average, 47 % fewer ions than the non-edge clusters and had mean radii 17 % smaller (as measured by radius of gyration). Evidently, to correctly estimate the true cluster size distribution, the edge clusters should be discounted.

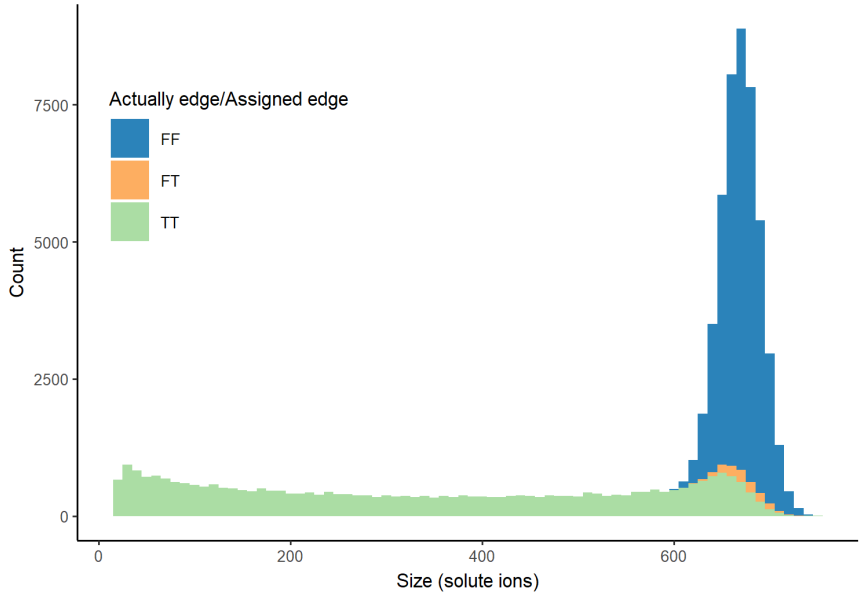


Figure 5: Cluster size histogram (stacked) for all detected clusters. Bar fill colour indicates if the cluster was an edge cluster and whether it was assigned as an edge cluster. TT and FF equals correct classification, FT are false-positives; non-edge clusters falsely identified as edge clusters.

4.1.3 Cluster Composition

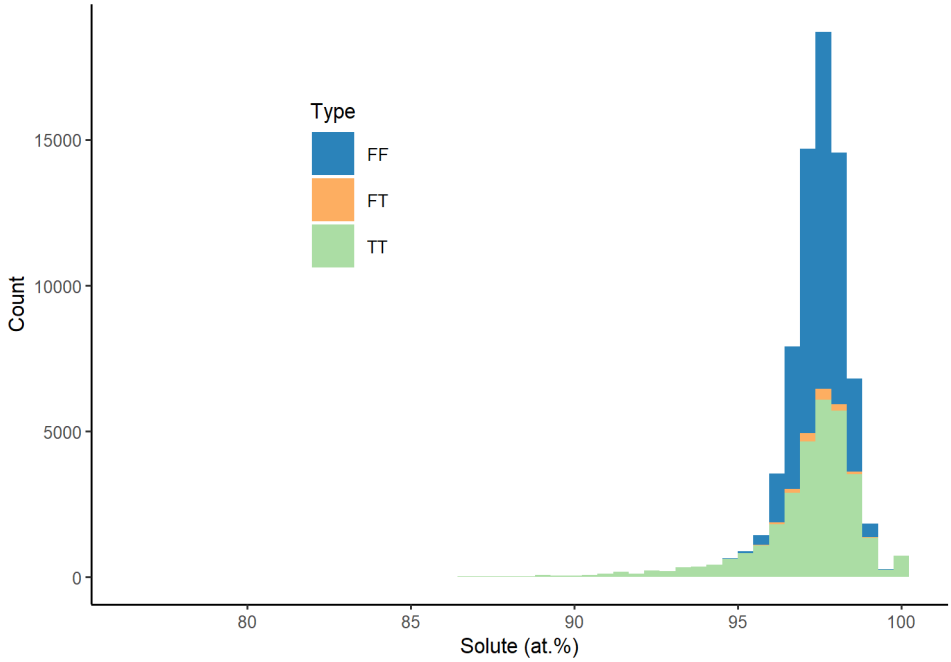


Figure 6: Stacked histogram of cluster composition of core ions.

Figure 6 shows the composition of the clusters after the envelope and erosion steps of the cluster selection. Whilst the true composition of the clusters is 100 at.% solute, the measured composition is slightly lower because the implemented cluster search algorithm will include some matrix ions within the clusters. The distribution of composition of the non-edge clusters results from the stochastic selection of atoms during the data generation. However, the edge clusters have a much broader distribution of compositions, many with a lower solute content and some with a higher solute content. Being on the edge of the data

volume may slightly increase the overall solute fraction or it may greatly decrease the solute fraction. The increase occurs when the cluster closely approaches the edge and some matrix is removed. The solute fraction decreases when the cluster has more matrix-containing surface area compared to the solute-rich volume. This effect is exacerbated for the core-shell simulation data, shown in Figure 7, where the measured composition of the cluster can greatly vary. The fraction of core ions (A) within the whole cluster should be 50 at.%, but this can be increased or decreased depending on the location of the cluster with respect to the edge of the simulation volume.

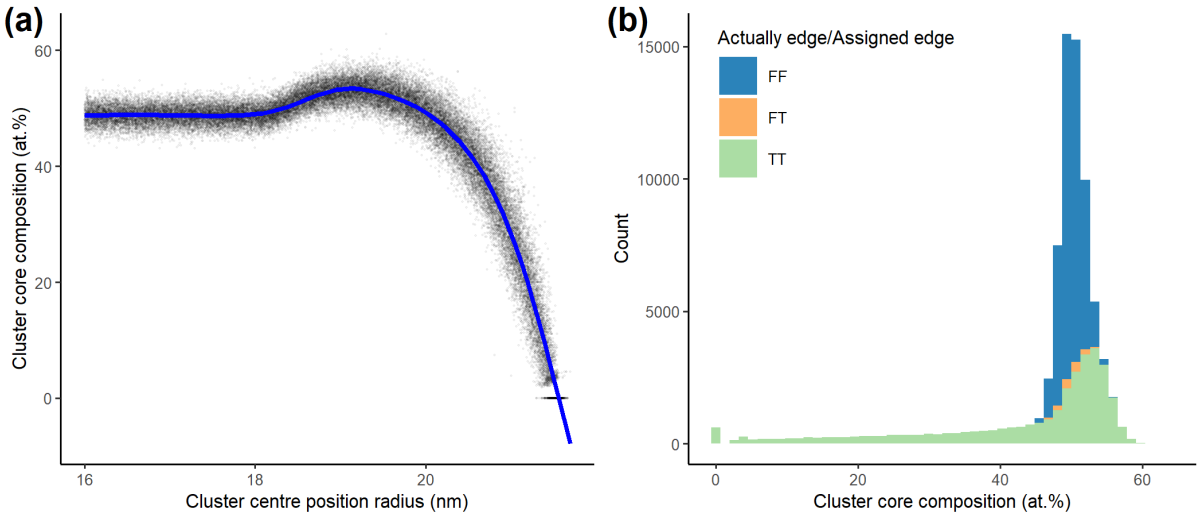


Figure 7: a) Composition of the cluster's core ions as a function of the simulated cluster centre for core-shell clusters. >18 nm centre position means increasing intersection with the edge of the volume. b) Stacked histogram of the core ion composition depending on the edge-type classification. The core composition is 50 at.% when not intersecting the edge.

4.2 Experimental Data

To demonstrate the suitability of the aforementioned protocol to experimental data, the method was applied to a dataset that contained 6.7 million detected ions. 505 clusters were identified in this specimen after cluster searches were performed. Whilst the atomic density of the reconstructed volume (31.5 atoms nm^3) is different to the atomic density of the simulated data, the authors do not believe that this will directly affect the results.

4.2.1 Effect of Sampling Data

As it took a long time (> 2 hours) for the alpha-hull algorithm to generate an alpha-shape around the atom probe reconstruction in the original .pos file, and due to RAM constraints, it was necessary to downsample the data. The effect of data downsampling on the alpha-shape computation time and the number of edge-classified clusters was investigated.

To investigate the relationship between downsampling and computational time for this dataset, the original data was randomly sampled 20 times at five different sampling fractions (0.02, 0.01, 0.005, 0.001, 0.0001). The time taken to generate an alpha-shape, using an alpha value of $\alpha = 2 \times max(NND)$, around the remaining points and determine which clusters were wholly contained within this volume was recorded. Figure 8 (a) shows the effect that sampling fraction had on the time required to generate the alpha-shape and determine which clusters were edge clusters. There was a clear trend for increased time required to perform the algorithm as the level of downsampling became less aggressive.

The number of edge clusters determined for each sampling fraction (0.02, 0.01, 0.005, 0.001, 0.0001) was also recorded (Figure 8 (b)).

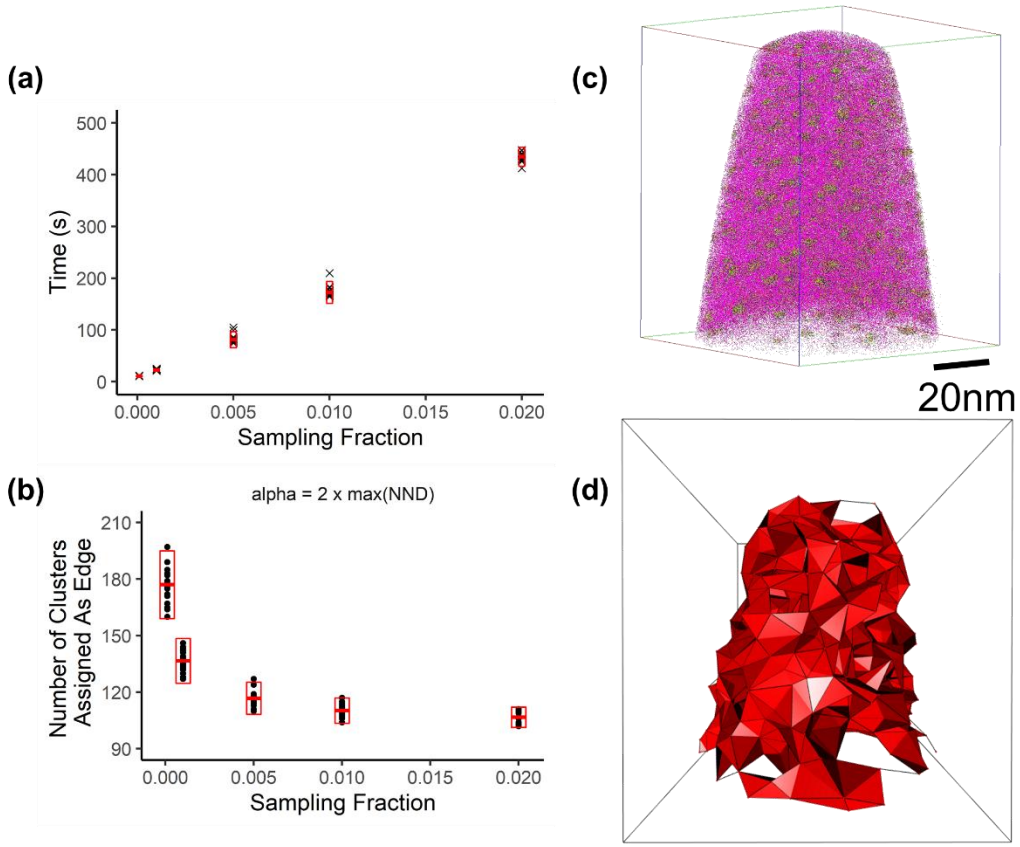


Figure 8: Effect of varying sampling fraction on (a) time taken to perform algorithm and (b) the number clusters assigned as edge clusters for $\alpha = 2 \times max(NND)$. Mean values and 95 % CI are displayed. (c) shows the original dataset and (d) shows the alpha-shape generated when using a sampling fraction of 0.0001 and setting $\alpha = max(NND)$

The effect of the alpha value was then investigated by randomly sampling the dataset twenty times for each sampling fraction between 0.0001 and 0.01. For each sampled dataset, alpha shapes were then generated for alpha values equal to $\alpha = x \times max(NND)$, with $x = 1, 2, 3, or\ 4$. The number of clusters that were assigned as edge in each case was determined. Figure 9 shows the effect that different alpha values and levels of downsampling had on the number of edge clusters detected. When $\alpha > max(NND)$, there was much lower variance in the number of clusters that were assigned as edge clusters for lower sampling fractions. The number of clusters assigned as edge clusters decreased as sampling fraction increased but converges after a sampling fraction of 0.005.

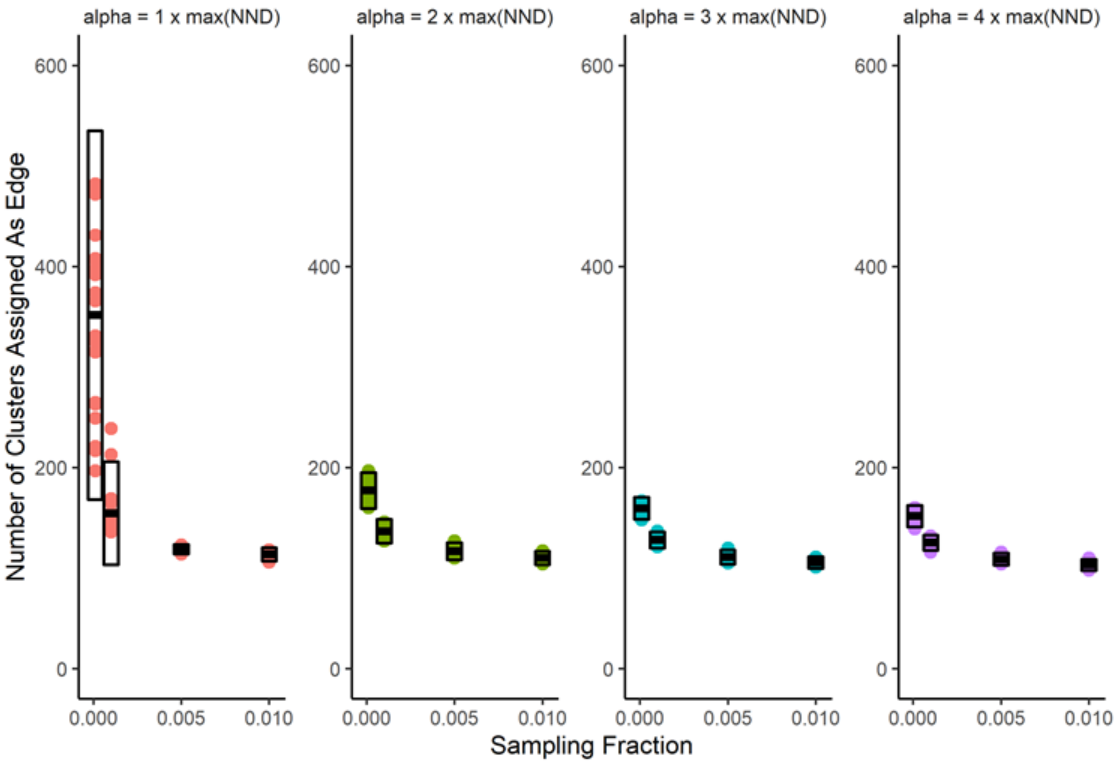


Figure 9: The effect that varying sampling fraction and alpha value has on the number clusters assigned as edge clusters by the algorithm. Mean values and 95 % confidence intervals are displayed.

4.2.2 Application to Experimental Data

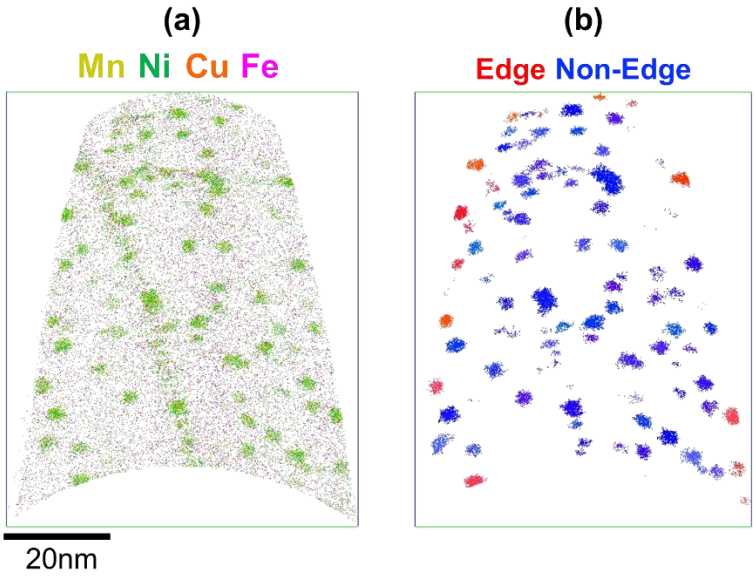


Figure 10: 5nm thick sections through an atom probe dataset showing (a) the distribution of ions and (b) clusters which are designated as edge (red) and non-edge (blue)

When using a sampling fraction of 0.005 and $\alpha = 2 \times max(NND)$ the algorithm took, on average, 81.3 seconds to identify 117 edge clusters within the dataset. Figure 10 (b) shows

the clusters that were assigned as edge and non-edge clusters after the algorithm was applied.

4.2.3 Application to Complex Shapes

The method was also successfully used to identify edge clusters in a reconstruction that was complex in shape. The original dataset contained a carbide that, as it was enriched in Mn, interfered with the initial cluster search (Figure 11). Therefore, the carbide was removed from the APT dataset using an 18 at.% C isosurface. After the carbide was removed, cluster searches detected 356 clusters in the matrix (Figure 11 (c)). After downsampling to a fraction of 0.005 and setting $\alpha = 2 \times max(NND)$, an alpha-shape was fitted to the matrix data and 89 clusters were found to be on the edge of the dataset (Figure 11 (d)-(g)).

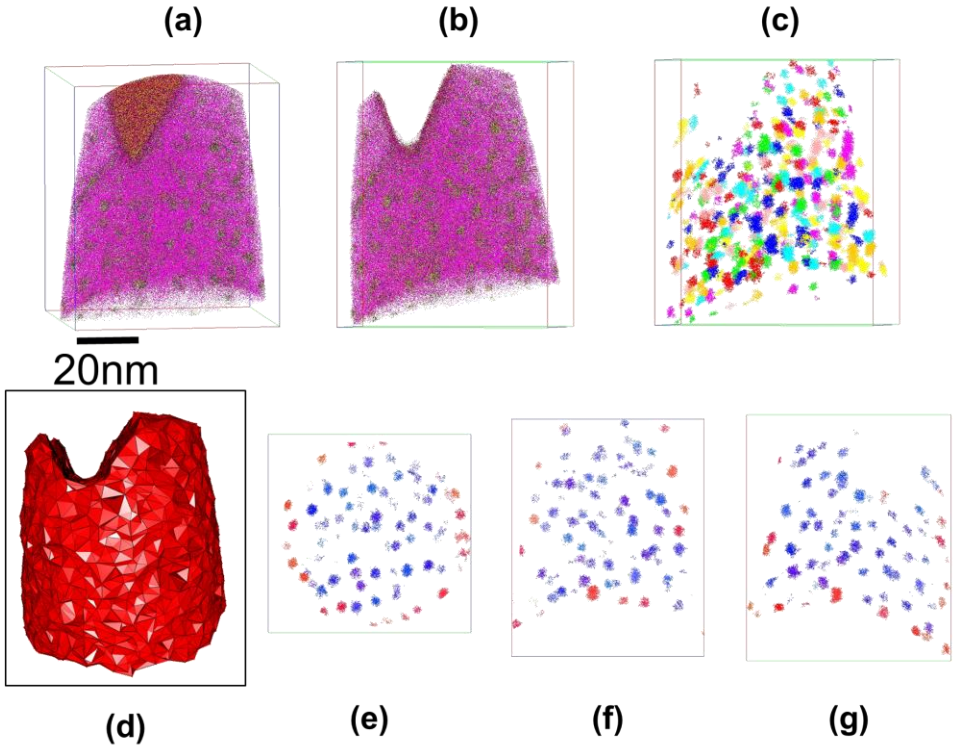


Figure 11: (a) original data, (b) data after removal of isosurface region containing more than 18 at.% C, (c) the clusters detected in the matrix of the dataset, (d) alpha-shape generated around dataset after carbide removal. (e), (f), and (g) show the clusters that were defined as edge (red) and non-edge (blue) in 10nm sections of the y-z, x-z, and x-y planes respectively.

4.2.4 Effect on Calculated Cluster Statistics

After cluster searches had been performed on the dataset shown in Figure 8 (c), a cluster volume fraction of 2.45 % was measured. The measured number density of clusters was 2.4 x 10^{24} m^{-3}. However, after edge clusters were identified and accounted for, a number density of 2.1 x 10^{24} m^{-3} was determined. Calculated cluster sizes were also affected by the inclusion/exclusion of edge clusters in the calculations. The edge clusters contained, on average, 286 ± 23 solute ions. Meanwhile, non-edge clusters were larger and contained 349 ± 13 ions. This is shown graphically for the clusters detected in eight datasets of this alloy in Figure 12. The solute content fraction of the clusters was not significantly affected by the inclusion or exclusion of edge clusters (Figure 13).

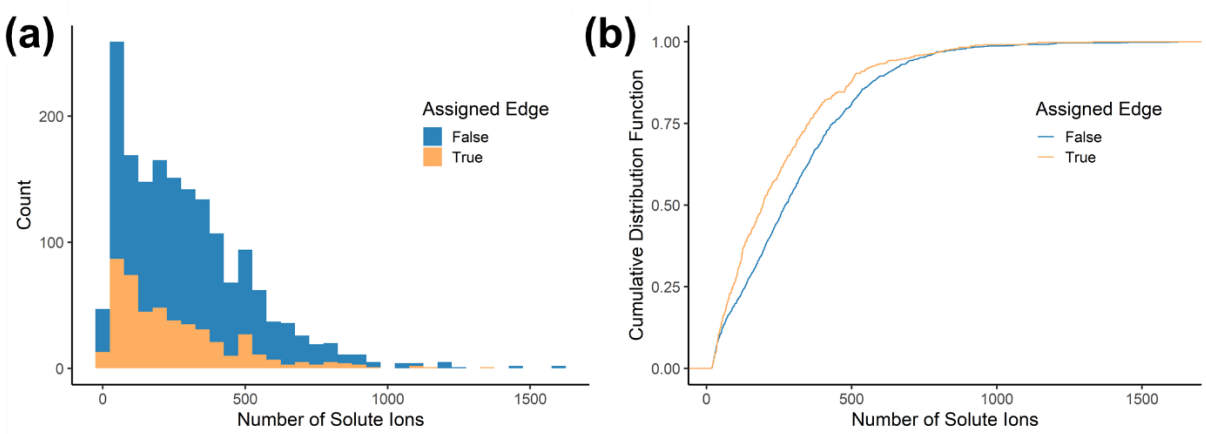


Figure 12: (a) Cluster size histogram (stacked) for all detected clusters ($N_{min} = 25$). Bar fill colour indicates if the cluster was assigned as an edge cluster. (b) Cumulative frequency plot for cluster sizes of edge and non-edge clusters.

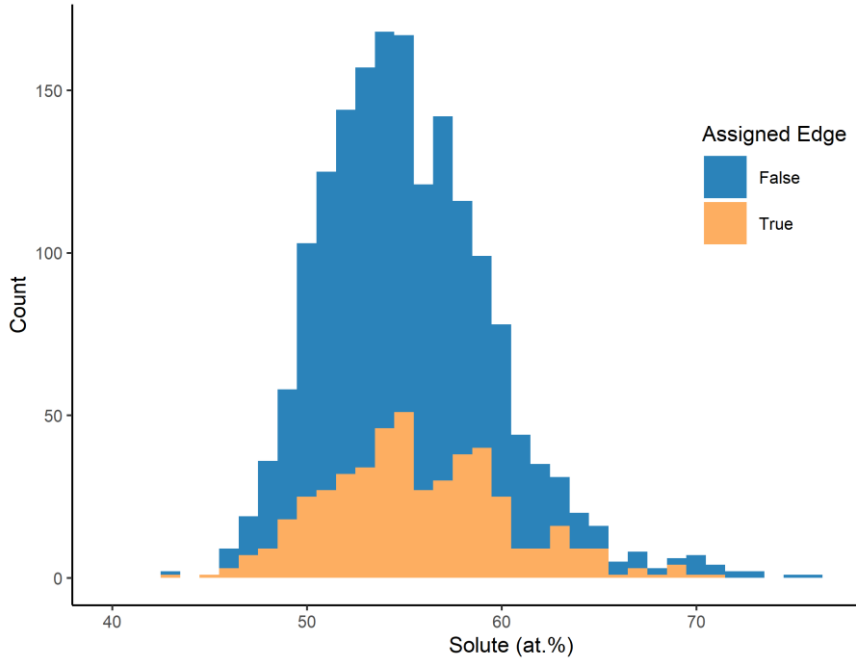


Figure 13: Stacked histogram of cluster composition of non-Fe ions (N.B. the majority of Fe ions detected in the clusters are likely the result of trajectory aberrations).

5 Discussion

5.1 Methodology

A reproducible method for identifying edge clusters in simulated and real datasets has been presented. Downsampling real datasets is necessary for the algorithm to be performed in a reasonable time and an alpha value which generates one volume containing all of the sample points must be chosen. These user-defined parameters must be carefully selected.

5.1.1 *Effect of Downsampling*

More aggressive downsampling of the original data has been shown to decrease the time required for the alpha shape to be generated and edge clusters to be identified (Figure 8 (a)). As the number of clusters that must be tested to determine if they are edge or not remained constant, the main difference in time taken to identify edge clusters is the result of the time required to generate the alpha-shape around the data. In the future it may be possible to perform the inside/outside point test in a more computationally efficient way than is currently done by alphashape3d package that is utilised in this study, thus reducing the time required for the algorithm to be performed.

As the downsampling becomes more aggressive there will be fewer points in the sampled point cloud. This is likely to cause the maximum nearest neighbour distance between ions to increase, hence increasing the alpha values used to generate the alpha-shape. To investigate algorithm performance, alpha-surfaces were generated for a variety of sampling levels at constant alpha-value of alpha = 100 nm, before a variety of alpha values were used at a fixed sampling fraction of 0.005. It was found that reducing sampling fraction, i.e. reducing the number of points in the sampled point cloud, was responsible for reducing the time taken to generate the alpha-shape.

Figure 9 indicates that more edge clusters were detected as the original data was downsampled more aggressively. This was probably due to two reasons. Firstly, sampling artificially reduced the volume of the alpha-shape that was generated around the dataset. This reduction in volume may lead to clusters, which were not on the edge of the original dataset, being outside of the computed alpha-shape. In the most extreme case of downsampling, one ion would remain in the sampled dataset. The alpha-shape generated around this would encapsulate only one ion and all clusters would lie outside of this shape and be assigned as edge. Secondly, more aggressive downsampling can be expected to reduce the likelihood that the alpha-shape generated around the remaining points in the filtered .pos file is representative of the original reconstructed specimen shape.

Mistakenly assigning several non-edge clusters as edge should not greatly affect size and composition calculations, since most datasets will contain many other clusters that can be used to determine these characteristics. However, due to the small analysed volume in APT, incorrectly assigning non-edge clusters as edge clusters may lead to underestimations in calculated cluster number densities. Meanwhile, not removing edge clusters prior to carrying out number density calculations will lead to an over-calculation of cluster number density.

Since it is important to minimise the number of non-edge clusters that may be defined as edge, the level of applied downsampling should not be too aggressive. However, a suitable trade-off must be reached between reducing the probability of non-edge clusters being assigned as edge clusters and minimising the time taken to perform the algorithm. Figure 9 indicates that the number of assigned edge clusters decreased significantly as the sampling fraction increased from 0.0001 to 0.001. Above 0.001, the number of assigned edge clusters was observed to decrease slightly, but the computational time significantly increased.

Consequently, it was decided that a sampling fraction of 0.005 provides a reasonable balance of these considerations for this dataset. The optimal sampling fraction may vary for different datasets and is like to depend on the number of detected ions in the original dataset, as well as the complexity of the shape of the original dataset. It is recommended that a sampling fraction should be used such that there is not a great degree of variance in the number of clusters that are detected by the algorithm.

5.1.2 *Effect of alpha value*

Setting $\alpha = max(NND)$ should generate one volume that contains all of the points in the sampled dataset. The alpha-shape will represent the shape of the sampled dataset with high resolution. However, due to the downsampling process, the remaining points and the generated alpha-shape will not necessarily be representative of the unsampled data. For example, using an alpha value of $\alpha = max(NND)$ led to a large degree of variance in the number of clusters assigned as edge and many non-edge clusters being assigned as edge. Several of the generated alpha-shapes were not representative of the shape of the original dataset. The generated alpha-shape in Figure 8 (d) had voids passing through it and was clearly not representative of the original data. As this will lead to clusters being incorrectly assigned as edge, a larger value should be selected for this dataset. Figure 8 shows that, for sampling fractions above 0.001, alpha values of $\alpha \geq 2 \times max(NND)$ did not significantly affect the number of clusters assigned as edge by the algorithm. Therefore, an alpha value of $\alpha = 2 \times max(NND)$ was used for this dataset. We propose an alpha value of $2 \times max(NND)$ is a reasonable value of general applicability which scales well with the level of sampling used.

5.2 Edge cluster effect: Simulated data

The algorithm was shown to improve the accuracy of both size (Figure 5) and composition measurements of the clusters in the simulated datasets. The measured composition of core-shell clusters was shown to be particularly susceptible to error if edge clusters were included in the analyses (Figure 7).

The measured number density of clusters was shown to be an overestimate of the true value when edge clusters were not accounted for, whilst the number density was underestimated when edge clusters were identified and accounted for using Equation 1. However, when edge clusters were accounted for, the accuracy of the number density measurement was increased. The underestimation in the calculated number density of the simulated datasets is due to the false-positives that are detected by the algorithm and due to the non-identification of clusters by the cluster search algorithm. The false-positive rate (FT) is where non-edge clusters are mistakenly identified as edge clusters.

The false-positives are caused by alpha-shape underestimating the volume boundary because the input data is sampled. In the simulated data, these false-positives accounted for about 7 % of all the detected clusters, thus giving roughly a 4 % underestimate of the number density (when edge clusters are counted as a half). On average, the alpha-shape underestimated the true volume by 3.1 % in the simulated test cases shown here. However, if the shape of data is concave then volume may be under or overestimated by the alpha-shape generated from the sampled points. Note that the volume of the dataset can be quickly and simply estimated using the total ionic volume, measured without sampling. This could be used to check the alpha-shape generated has a volume representative of unsampled data.

The second issue relates to the detection of the clusters by the cluster selection algorithm. To limit the inclusion of clusters arising purely from statistical fluctuations in the bulk a minimum number of core ions is specified, N_{min} [26]. In the simulated data a realistic value was chosen, resulting in about 3 non-detected clusters on the edge of the volume for each simulation. The effect of using a more sensitive cluster detection can be shown by including the non-detected clusters in the edge-cluster count. Then the number density estimate would be of 9.05×10^{23} clusters/m^3, just 1 % over the true value. This highlights the issue of not selecting small clusters leading to an underestimate of the number density. However, these

clusters should not be included in further measurements because they will bias the cluster size and composition measurements.

5.3 Edge cluster effect: Irradiated steel data

When applied to real data in Figure 8 (c), the mean measured cluster size was shown to increase if edge clusters were identified and removed from the analysis. This is due to the clusters detected on the edge of the dataset only containing a fraction of the ions that were present in the cluster prior to undergoing APT analysis. The measured number density of clusters was found to decrease by around 13 % when edge clusters were accounted for.

The composition of the clusters in the real dataset was not observed to be significantly affected by the inclusion/exclusion of edge clusters. The clusters did not display a core shell structure and this may be one of the reasons for this observation. However, the edge clusters did display a slightly higher variance in their compositions than the non-edge clusters.

5.4 Applicability of method

The ability to apply the proposed methodology to automatically and reproducibly detect edge clusters in a regularly shaped APT specimen, within a reasonable time frame, has been demonstrated. Accounting for edge clusters has been shown to improve the accuracy of size and compositional measurements in simulated datasets. Meanwhile, failure to account for edge clusters has been shown to lead to an overestimate in cluster number density measurements.

It has also been shown that edge clusters in real datasets can be identified and accounted for, including in irregularly shaped and concave specimens. Irregularly shaped datasets are frequently generated when isosurfaces are used to extract chemically enriched regions, such as carbides, from the original .pos file. As these chemically enriched volumes interfere with maximum separation cluster searches, it is often necessary to remove them prior to performing cluster analyses. It is desirable to determine which clusters are on the edge of the cropped volume not only for accurate cluster calculations, but also because clusters that have nucleated at the carbide-matrix interface may have different compositions and sizes to those that nucleated homogeneously in the matrix.

To generate the alpha-shape, which defines the surface of the data volume, some degree of downsampling was necessary. The number of identified edge clusters was shown to converge once the sampling rate exceeded 0.001 (0.1 % of the data). The measured maximum nearest neighbour distance of the sampled data can be used to give a sampling-fraction independent alpha value to use to generate the alpha shape. With $\alpha \geq 2 \times max(NND)$ found as a generally applicable value. These values should be appropriate for any experimental data. However, where high-accuracy of sharp concave features is required, smaller alpha values are required and, therefore, the data should be sampled less aggressively.

5.5 Considerations

Whilst the applicability of the proposed method has been demonstrated via application to simulated and real data, there are several necessary considerations.

Estimating the sample boundary using an alpha shape of sampled data will always be an approximation, and the exact shape will be stochastic due to the random sampling. As shown in Figure 4, some clusters will be falsely identified as being edge and the sampling fraction and alpha value will affect the number of clusters assigned as edge. This conservatism should not affect mean size and composition calculations, but may lead to an

under-calculation in the measured number density. However, not removing edge clusters will lead to more erroneous over-calculated number densities.

Reasonable values for the alpha value and sampling fraction can be determined using the approaches shown in this paper. For increased confidence a convergence test of alpha and sample fraction should be performed. More complex computational approaches may eliminate the need for sampling the data [27], however the sampling method presented here is quick and gives reasonable results in a few seconds.

6 Conclusions

In this study, the effect that edge clusters can have on atom probe tomography data analyses has been investigated. It has been found that:

- Failure to identify and account for edge clusters can reduce the accuracy of the number density, size, and compositional calculations.
- A novel method has been developed for automatically and reproducibly identifying clusters on the edge of simulated atom probe datasets. The algorithm can be applied in a timeframe of the order of two minutes on a desktop computer.
- These clusters were then accounted for in number density, size, and compositional calculations.
- The same algorithm has been successfully applied to real datasets, including datasets of complex shape. The effect of accounting for these edge clusters on number density, size and compositional measurements has been investigated in one alloy.

In conclusion, identifying and accounting for edge clusters is an imperative step if one is to accurately determine the characteristics of clusters present in APT datasets. This present work provides a much-needed development that will contribute towards improving the accuracy of quantitative results from APT data.

7 Appendices

8 Acknowledgements

We acknowledge funding from Rolls-Royce Plc., Derby, UK. The authors would also like to thank Prof. G. R. Odette for providing the materials used in this study and for helpful discussions during the course of this work. This work was supported by the U.S. Department of Energy, Office of Nuclear Energy under DOE Idaho Operations Office Contract DE-AC07-051D14517 as part of a Nuclear Science User Facilities experiment.

Funding Sources

This work was supported by Rolls-Royce Plc., Derby, UK. © Rolls-Royce plc 2019. The atom probe facilities at the University of Oxford are funded by the EPSRC (EP/M022803/1).

This work has been carried out within the framework of the EUROfusion Consortium and has received funding from the Euratom research and training programme 2014-2018 and 2019-2020 under grant agreement No 633053. The views and opinions expressed herein do not necessarily reflect those of the European Commission.

Data availability

The raw/processed data required to reproduce these findings cannot be shared at this time due to legal or ethical reasons.

References

[1] R.K.W. Marceau, G. Sha, R. Ferragut, A. Dupasquier, S.P. Ringer, Solute clustering in Al-Cu-Mg alloys during the early stages of elevated temperature ageing, Acta Mater. 58 (2010) 4923–4939. doi:10.1016/j.actamat.2010.05.020.

[2] P.B. Wells, T. Yamamoto, B. Miller, T. Milot, J. Cole, Y. Wu, G.R. Odette, Evolution of manganese-nickel-silicon-dominated phases in highly irradiated reactor pressure vessel steels, Acta Mater. 80 (2014) 205–219. doi:10.1016/j.actamat.2014.07.040.

[3] B. Gault, M.P. Moody, J.M. Cairney, S.P. Ringer, Atom Probe Microscopy, 1st ed., Springer-Verlag New York, 2012. doi:10.1007/978-1-4614-3436-8.

[4] J.M. Hyde, C.A. English, An Analysis of the Structure of Irradiation induced Cu-enriched Clusters in Low and High Nickel Welds, in: Symp. R 'Microstructural Process. Irradiat. Mater. Fall MRS, 2000. doi:https://doi.org/10.1557/PROC-650-R6.6.

[5] L.T. Stephenson, M.P. Moody, S.P. Ringer, Theory of Solute Clustering in Materials for Atom Probe, Philos. Mag. 91 (2011) 2200–2215. doi:10.1080/14786435.2011.554909.

[6] M.K. Miller, B.D. Wirth, G.R. Odette, Precipitation in neutron-irradiated Fe–Cu and Fe–Cu–Mn model alloys: a comparison of APT and SANS data, Mater. Sci. Eng. A. 353 (2003) 133–139. doi:10.1016/S0921-5093(02)00679-2.

[7] P.D. Styman, J.M. Hyde, K. Wilford, A. Morley, G.D.W. Smith, Precipitation in long term thermally aged high copper, high nickel model RPV steel welds, Prog. Nucl. Energy. 57 (2012) 86–92. doi:10.1016/j.pnucene.2011.10.010.

[8] I. Arslan, E.A. Marquis, M. Homer, M.A. Hekmaty, N.C. Bartelt, Towards better 3-D reconstructions by combining electron tomography and atom-probe tomography, Ultramicroscopy. 108 (2008) 1579–1585. doi:10.1016/j.ultramic.2008.05.008.

[9] E. Meslin, M. Lambrecht, M. Hernández-Mayoral, F. Bergner, L. Malerba, P. Pareige, B. Radiguet, A. Barbu, D. Gómez-Briceño, A. Ulbricht, A. Almazouzi, Characterization of neutron-irradiated ferritic model alloys and a RPV steel from combined APT, SANS, TEM and PAS analyses, J. Nucl. Mater. 406 (2010) 73–83. doi:10.1016/j.jnucmat.2009.12.021.

[10] S. Shu, B.D. Wirth, P.B. Wells, D.D. Morgan, G.R. Odette, Multi-technique characterization of the precipitates in thermally aged and neutron irradiated Fe-Cu and Fe-Cu-Mn model alloys: Atom probe tomography reconstruction implications, Acta Mater. 146 (2017) 237–252. doi:10.1016/j.actamat.2017.12.006.

[11] S. Shu, N. Almirall, P.B. Wells, T. Yamamoto, G.R. Odette, D.D. Morgan, Precipitation in Fe-Cu and Fe-Cu-Mn model alloys under irradiation: Dose rate effects, Acta Mater. 157 (2018) 72–82. doi:10.1016/j.actamat.2018.07.017.

[12] J.W. Valley, A.J. Cavosie, T. Ushikubo, D.A. Reinhard, D.F. Lawrence, D.J. Larson, P.H. Clifton, T.F. Kelly, S.A. Wilde, D.E. Moser, M.J. Spicuzza, Hadean age for a post-magma-ocean zircon confirmed by atom-probe tomography, Nat. Geosci. 7 (2014) 219–223. doi:10.1038/ngeo2075.

[13] D.J. Haley, posgen, (2017). https://sourceforge.net/projects/apttools/files/posgen/.

[14] R.C. Team, R: A language and environment for statistical computing, (2013). http://www.r-project.org/.

[15] L.T. Stephenson, M.P. Moody, P.V. Liddicoat, S.P. Ringer, New techniques for the analysis of fine-scaled clustering phenomena within atom probe tomography (APT)

data, Microsc. Microanal. 13 (2007) 448–463. doi:10.1017/S1431927607070900.

[16] M.K. Miller, K.F. Russell, Atom probe specimen preparation with a dual beam SEM/FIB miller, Ultramicroscopy. 107 (2007) 761–766. doi:10.1016/j.ultramic.2007.02.023.

[17] M.K. Miller, K.F. Russell, G.B. Thompson, Strategies for fabricating atom probe specimens with a dual beam FIB, Ultramicroscopy. 102 (2005) 287–298. doi:10.1016/j.ultramic.2004.10.011.

[18] K. Thompson, D. Lawrence, D.J. Larson, J.D. Olson, T.F. Kelly, B. Gorman, In situ site-specific specimen preparation for atom probe tomography, Ultramicroscopy. 107 (2007) 131–139. doi:10.1016/j.ultramic.2006.06.008.

[19] B.P. Geiser, D.J. Larson, E. Oltman, S. Gerstl, D. Reinhard, T.F. Kelly, T.J. Prose, Wide-Field-Of-View Atom Probe Reconstruction, Microsc. Microanal. 15 (2009) 292–293. doi:10.1017/S1431927609098249.

[20] B. Gault, D. Haley, F. de Geuser, M.P. Moody, E.A. Marquis, D.J. Larson, B.P. Geiser, Advances in the reconstruction of atom probe tomography data, Ultramicroscopy. 111 (2011) 448–457. doi:10.1016/j.ultramic.2010.11.016.

[21] P.D. Styman, J.M. Hyde, D. Parfitt, K. Wilford, M.G. Burke, C.A. English, P. Efsing, Post-irradiation annealing of Ni-Mn-Si-enriched clusters in a neutron-irradiated RPV steel weld using Atom Probe Tomography, J. Nucl. Mater. 459 (2015) 127–134. doi:10.1016/j.jnucmat.2015.01.027.

[22] H. Edelsbrunner, E.P. Mücke, Three-dimensional alpha shapes, ACM Trans. Graph. 13 (1994) 43–72. doi:10.1145/174462.156635.

[23] B. Pateiro-López, A. Rodríguez-Casal, Generalizing the convex hull of a sample: The R package alphahull, J. Stat. Softw. 34 (2010) 1–28. doi:http://dx.doi.org/10.18637/jss.v034.i05.

[24] P. Felfer, A. Ceguerra, S. Ringer, J. Cairney, Applying computational geometry techniques for advanced feature analysis in atom probe data, Ultramicroscopy. 132 (2013) 100–106. doi:10.1016/j.ultramic.2013.03.004.

[25] E.A. Marquis, V. Araullo-Peters, Y. Dong, A. Etienne, S. Fedotova, K. Fujii, K. Fukuya, E. Kuleshova, A. Lopez, A. London, S. Lozano-Perez, Y. Nagai, K. Nishida, B. Radiguet, D. Schreiber, N. Soneda, M. Thuvander, T. Toyama, F. Sefta, P. Chou, On the Use of Density-Based Algorithms for the Analysis of Solute Clustering in Atom Probe Tomography Data, in: Miner. Met. {&} Mater. Ser., Springer International Publishing, 2017: pp. 2097–2113. doi:10.1007/978-3-030-04639-2_141.

[26] J.M. Hyde, E.A. Marquis, K.B. Wilford, T.J. Williams, A sensitivity analysis of the maximum separation method for the characterisation of solute clusters, Ultramicroscopy. 111 (2011) 440–447. doi:10.1016/j.ultramic.2010.12.015.

[27] M. Kühbach, P. Bajaj, A. Breen, E.A. Jägle, B. Gault, On Strong Scaling Open Source Tools for Mining Atom Probe Tomography Data, Microsc. Microanal. 25 (2019) 298–299. doi:10.1017/s1431927619002228.