# Alpha Shape Analysis (ASA) Framework for Post Clustering Property Determination in Atom Probe Tomographic Data

Evan Still,[1] Daniel K Schreiber,[2], Jing Wang,[2] and Peter Hosemann[1]

[1]*Department of Nuclear Engineering, University of California, Berkeley, CA, 94720, USA*

[2]*Energy and Environment Directorate, Pacific Northwest National Laboratory, Richland, WA, 99354, USA*

*Corresponding Author: Evan Still <evanstill@berkeley.edu>*

**Abstract**

While application of clustering algorithms to atom probe tomography data have enabled quantification of solute clusters in terms of number density, size, and subcomposition there exist other properties (e.g. volume, surface area, and composition) that are better determined by defining an interface between the cluster and the surrounding matrix. Namely volume, surface area, and total composition. The limitation in composition results from a pre-filtering step where the expected matrix ion types are omitted from the cluster search to enhance the contrast between the matrix and cluster and to reduce the complexity of the search. Previously composition determination within solute clusters has utilized a secondary envelopment and erosion step on top of conventional methods such as maximum separation. In this work we present a novel and automated method that combines the particle identification fidelity of a conventional clustering algorithm with the analytical flexibility of mesh-based approaches through the generation of alpha shapes for each identified cluster. The corresponding mesh accounts for concave components of the clusters and determines the volume and surface area of the clusters, additionally the mesh boundary is utilized to update the total composition

according to the internal ions.

**Key Words:** atom probe tomography, alpha shapes, clustering (Received XX Y 20ZZ; revised XX Y 20ZZ; accepted XX Y 20ZZ)

## Introduction

Atom probe tomography (APT) provides the unique capability to probe 3-dimensional spatial and chemical identity information simultaneously at the nanoscale (Geiser et al., 2007, 2009; Bas et al., 1995; Gault et al., 2012). Due to this joint capability APT has become a standard method for visualizing solute-rich microstructural features in a wide variety of materials (Marquis & Hyde, 2010). Despite APT's sub-nm, 3D spatial resolution, accurate quantification of nanoscale features is still a source of ongoing research with many groups developing new frameworks and algorithms to improve the detection of atomic ordering and clustering within the generated point cloud (Wang et al., 2019; Ghamarian & Marquis, 2019). Much of this work is focused on the characterization of precipitates during thermal aging or irradiation due to the impact such precipitates have on the mechanical properties of the alloys in question (Bailey et al., 2015; Bachhav et al., 2014; Miller & Kenik, 2004). Yet, more recently attempts have been made to extend past approaches to classification of solute-enriched dislocations, which prove difficult for conventional concentration based methods, such as isoconcentration surfaces, and density based methods to detect (Ghamarian et al., 2020). Whether analyzing precipitates or dislocations the properties of interest are generally the elemental composition and feature size, the latter of which is often measured in terms of the ions contained within the features or radial estimations assuming a spherical feature. In this work, we propose a novel method to enhance composition estimates for clusters regardless of the clustering algorithm used and extend the computation of feature size to volume and surface area through the use of alpha shapes.

For both concentration and density-based clustering approaches, a fundamental pre-

processing step is the filtration of the relevant solute species for clustering. In the case of isoconcentration surfaces, the filtration is performed with respect to user-selected ions that are used for the concentration threshold. For spatial search algorithms the filtering is performed externally and produces a subset of the spatial data that corresponds to only the ions of interest. In the latter, the filtration directly limits composition analysis to the subcomposition of the selected ions, i.e. for a search of yttrium-titanium-oxygen (YTiO) precipitates using only yttrium (Y) and titanium (Ti) as the ions of interest, the presence of oxygen (O) and possible inclusion of other solute ions can not be determined. While, for isosurfaces the generation of a mesh via the marching cubes algorithm enables recovery of the full local composition as well as volume and surface area.

This constraint on composition analysis following clustering has long been acknowledged as a dilemma, and algorithms such as the envelopment method were designed specifically to account for such discrepancies (Vaumousse et al., 2003). The envelopment and erosion approach is a modification of maximum separation where two additional distance parameters are supplied. The first of which determines the inclusion of "non-clustered" ions of all species relative to the distance to the nearest clustered ion. However, as noted by the authors this results in an enlargement of the clusters, so they then erode the clusters by removing all ions within the erosion distance from the cluster/matrix interface (Vaumousse et al., 2003; Guo et al., 2003). In Felfer et al. (2015) and Ghamarian & Marquis (2019) the convex hull of a cluster was identified as one possible way to better reflect the structure of the individual clusters while determining the composition. The convex hull however cannot account for concave features such as the internal cavity expected when identifying the shell of a core-shell structure or in identifying dislocations which possess a degree of curvature. Instead alpha shapes are identified as one possible approach for determining the concave hull. Alpha shapes have already been used to identify a mesh for the overall atom probe specimen in order to improve reconstructions as in Felfer & Cairney (2016) or to identify clusters existing near the boundary of the data as in Jenkins et al. (2020). We propose the alpha shape analysis (ASA)

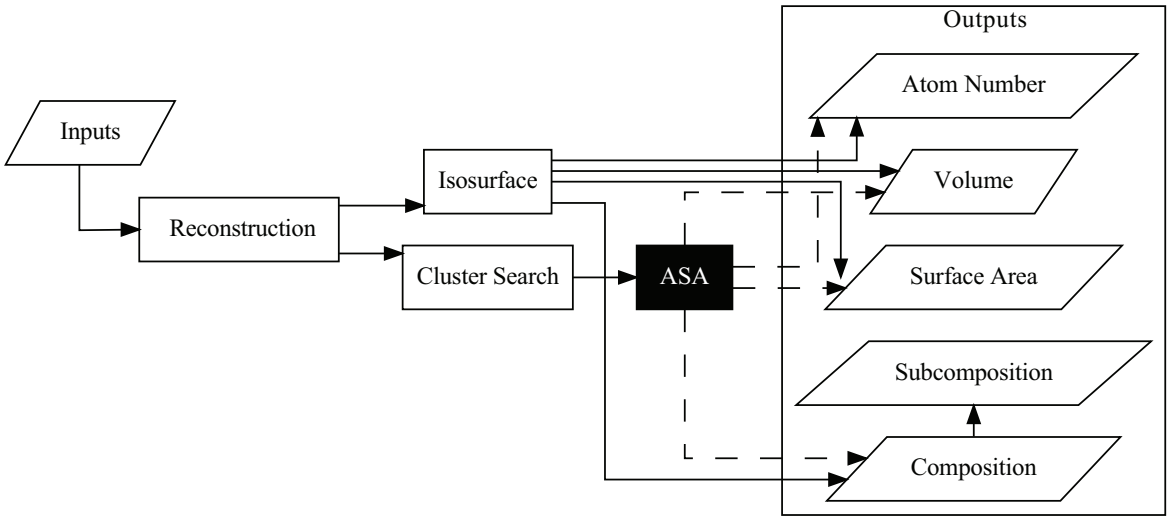



**Fig. 1.** Modified Atom Probe Analysis procedure for feature property extraction with the addendum of alpha shape analysis (ASA). ASA enables conventional clustering algorithms to compete with the Isosurface method through the determination of composition, surface area, and volume in addition to the expected subcomposition and atom number.

framework which enables irregular meshing of complex point clouds through a cluster-specific automated heuristic for selection of its singular input parameter, $\alpha$, which describes the radius of a carving sphere. Within the atom probe data analysis pipeline displayed in figure 1 ASA is performed as a post-clustering step that enables computation of composition, volume, and surface area for the identified clusters.

## Materials and Methods

### *Computational Background and Methodology*

We first briefly review the clustering algorithm, OPTICS, applied to our experimental test case. Following this the basis of the ASA procedure's subroutines, as presented in figure 2, are discussed. The potential limitations will later be assessed through a set of five synthetic datasets designed to explore three possible concave and two possible convex point clouds.

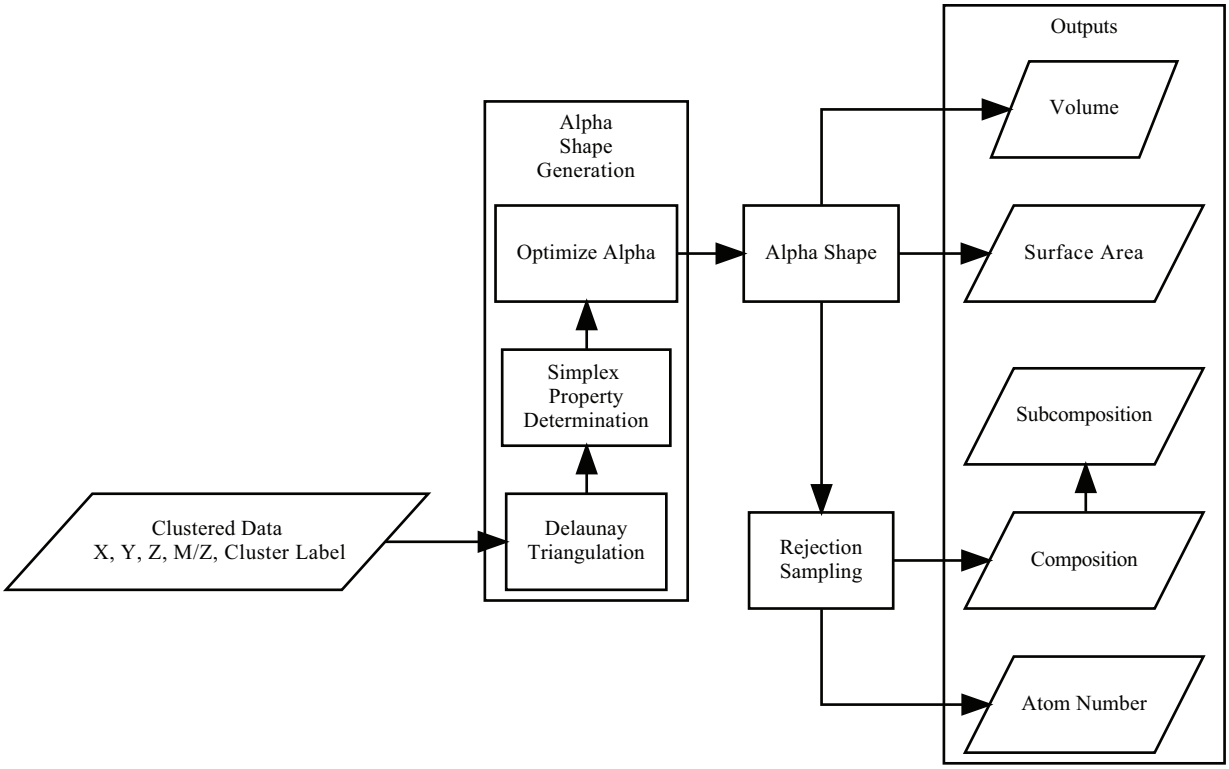



**Fig. 2.** Flowchart of the ASA frameworks subroutines. The output of the cluster algorithm alongside non-clustered data is supplied to the alpha shape generation subroutine. The alpha shape generation subroutine is furthered compartmentalized into three components which produce the Delaunay triangulation, calculates relevant simplex properties such as volume, and determines the $\alpha$ parameter. Following this the alpha shape determines the volume and surface Area while rejection sampling is performed to update the composition and atom number of the clusters.

### *OPTICS*

Proposed by Ankerst et al. (1999), in order to overcome inherent limitations of density-based spatial clustering on applications with noise (DBSCAN) when applied to clusters of varying atomic densities and noise the ordering points to identify the clustering structures (OPTICS) algorithm makes use of the notion of reachability distance to define a visualization of the cluster structure Ester et al. (1996). OPTICS requires two input parameters, $\epsilon$ and MinPts, which define the search distance from a point O, and the number of additional points within $\epsilon$ for O to be considered a core point. Starting at a random initial point, o, OPTICS identifies the $\epsilon$-neighbors and places them within a queue for further processing. If O is found to be a core point the core-distance (CD) is defined as the distance from O to its MinPts neighbor. Then for a point, P, in the queue the reachability-distance (RD) is defined as either the CD of O or the distance between P and O. The RD is then used to reorder the queue and the point with minimal RD is processed next. The OPTICS procedure and its requirements are further described in Ankerst et al. (1999).

The RD is plotted with respect to the index ordering to generate a reachability plot which provides a visualization of the cluster hierarchy as shown in figure 3 for a system with four clusters over top uniform noise. This plot is then processed to determine the final cluster structure with the most common analysis being based on identification of local minima alongside a significance test to determine whether a cluster should be segmented into sub-clusters. For further information on hierarchical extraction reference Sander et al. (2003) or Wang et al. (2019), for an atom probe specific algorithm.

### *Alpha Shape Generation*

Originally conceived by Edelsbrunner et al. (1983), alpha shapes and the related alpha hulls and alpha complexes were designed as a generalization of the 2-dimensional convex hull in which a point cloud's representative shape is generated through the eraser intuition and expanded to higher dimensions by considering the relationship between the alpha complexes and the

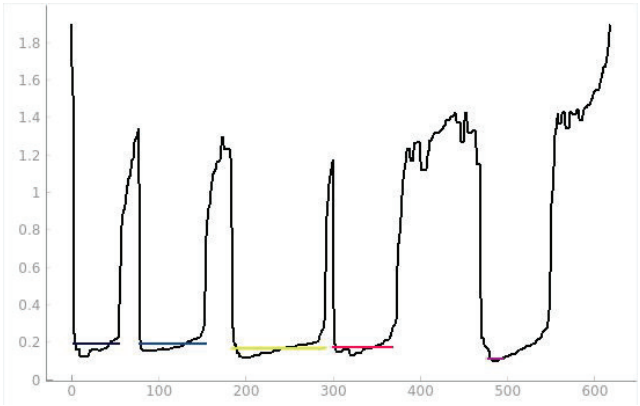



**Fig. 3.** Example reachability plot for a system of four clusters, each represented in a different color with black indicating non-clustered points. Cluster boundaries are determined according to a slope threshold. The reachability distance is presented on the y-axis, while the x-axis indicates the indices of the ordered list.

Delaunay triangulation (Edelsbrunner et al., 1983; Edelsbrunner & Mücke, 1994). The eraser intuition is based upon the movement of an eraser or carving ball, of radius $\alpha$, about the point cloud's convex hull. The carving ball is defined to be open and thus the points of the cloud may be circumscribed but not contained. As the carving ball moves throughout space the content of the convex hull contained within is deleted producing an alpha hull where circular arcs of curvature $\frac{1}{\alpha}$ bound the surface. The alpha shape is then defined as the linearized alpha hull as presented in figure 4. Intuitively, when $\alpha = \inf$, the alpha hull is the convex hull as the carving ball cannot enter the point cloud, while sufficiently small $\alpha$ can result in the formation of cavities and eventually a disconnected set.

The alpha hull and thus shape can be more formally defined and generalized for $\mathbb{R}^d$ by considering a filtration of the Delaunay triangulation referred to as the alpha complex. The Delaunay triangulation of a point set $S$, $DT(S)$, can be defined as the set of d-simplices, $\Delta_T$, such that the circum-hypersphere of any $\Delta_T$ does not contains a point in $S$. Given that $S$ is in general position, i.e. within a hyperplane no three points are collinear and no four points are cocircular, $DT(S)$ is unique and it is guarantied that each $\Delta_T$ possesses an open ball with a radius, $\sigma_T$, and centroid, $\mu_T$. Additionally, all lower dimensional simplices are bound by

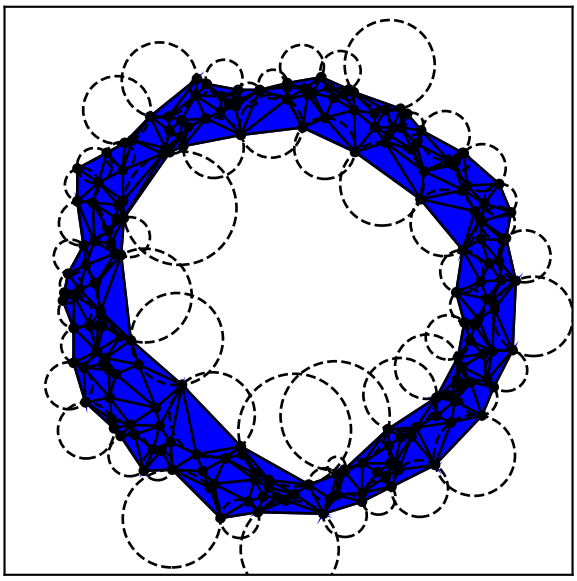


**Fig. 4.** Example alpha shape of a loop restricted to 2-simplices. The required carving balls to remove the simplices on the exterior and interior boundaries are dashed, while the solid lines represent the triangulation.

the higher dimensional simplices. Thus, the alpha complex can be defined as subset of the Delaunay triangulation, for which each simplex with $\sigma_T$ greater than $\alpha$ is removed. Further refinement of the alpha complex by identifying the exterior and interior boundaries results in the identification of the alpha shape.

### *Simplex Property Determination*

For each $\Delta_T$ within the $DT(S)$ the circumscribing radius, $\sigma_T$, must be determined for the generation of the alpha complex, and the content, volume in $\mathbb{R}^3$ and area in $\mathbb{R}^2$, must be calculated for the computation of the alpha complex's volume and surface area. All properties can be related to the Cayley Menger matrix of a $\Delta_T$, $CM(\Delta_T)$ presented in eq 1 (Sommerville,

1958).

$$CM(\Delta_T) = \begin{bmatrix} 0 & 1 & 1 & 1 & 1 & \cdots & 1 \\ 1 & 0 & l_{01}^2 & l_{02}^2 & l_{03}^2 & \cdots & l_{0d}^2 \\ 1 & l_{10}^2 & 0 & l_{12}^2 & l_{13}^2 & \cdots & l_{1d}^2 \\ 1 & l_{20}^2 & l_{21}^2 & 0 & l_{23}^2 & \cdots & l_{1d}^2 \\ 1 & l_{30}^2 & l_{31}^2 & l_{32}^2 & 0 & \cdots & l_{1d}^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & l_{d0}^2 & l_{d1}^2 & l_{d2}^2 & l_{d3}^2 & \cdots & 0 \end{bmatrix} \tag{1}$$

Here $l_{ij}$ denotes the distance between the $ith$ and $jth$ vertex of $\Delta_T$ for an arbitrary number of dimensions, note that $l_{i=j} = 0$. The relationship of $|CM(\Delta_T)|$ to the simplex content, $v_d$ is given in eq 2 (Sommerville, 1958). For a 3-simplex the content, $v_3$, describes the volume and for a 2-simplex the content, $v_2$, describes the area. All simplices are bounded by a set of lower dimensional simplices, and for a 3-simplex there are four 2-simplices composing its bounds.The surface area of a 3-simplex is then a summation over the content of the 2-simplices. Thus, eq 1 enables computation of each simplex's volume and surface area prior to selection of alpha and the subsequent generation of the alpha complex.

$$v_d^2 = \frac{(-1)^{d+1}}{(d!)^2 2^d} |CM(\Delta_T)| \tag{2}$$

To derive the relationship between the circumradius, $\sigma_T$, and $CM(\Delta_T)$ consider a $\Delta_T$, with sidelengths $l_{ij}$ and a center of mass referred to as a barycenter, $[x_0, x_1, x_2, x_3, ...x_d]$. Coxeter (1930), provides eq 3 to relate $\sigma_T$ to all $l_{ij}$ given the barycenter of $\Delta_T$ is known. Coxeter (1930) also establishes that a point, $P$ with barycentric coordinates $[y_0, y_1, y_2, \cdots, y_d]$, exists within $\Delta_T$ only if condition 4 is met.

$$\sum_{j=0}^{d} x_j l_{ij}^2 = 2\sigma_T^2 \quad for \quad i = (0, 1, 2, 3, \ldots, d) \tag{3}$$

$$\sum_{j=0}^{d} y_j \leqslant 1 \tag{4}$$

Consider the following constraints derived from eq 3.

$$\sum_{j=0}^{d} x_j l_{ij}^2 - 2\sigma_T^2 = 0 \quad for \quad i = (0, 1, 2, 3, \ldots, d) \tag{5}$$

Eq 4 and eq 5 provides a set of d + 2 linear equations from which it is evident that the coefficient matrix is identical to the $CM(\Delta_T)$ as shown in eq 6 and eq 7.

$$\begin{bmatrix} 0 & 1 & 1 & 1 & 1 & \cdots & 1 \\ 1 & 0 & l_{01}^2 & l_{02}^2 & l_{03}^2 & \cdots & l_{0d}^2 \\ 1 & l_{10}^2 & 0 & l_{12}^2 & l_{13}^2 & \cdots & l_{1d}^2 \\ 1 & l_{20}^2 & l_{21}^2 & 0 & l_{23}^2 & \cdots & l_{1d}^2 \\ 1 & l_{30}^2 & l_{31}^2 & l_{32}^2 & 0 & \cdots & l_{1d}^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & l_{d0}^2 & l_{d1}^2 & l_{d2}^2 & l_{d3}^2 & \cdots & 0 \end{bmatrix} \begin{bmatrix} -2\sigma_T^2 \\ x_0 \\ x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_d \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{6}$$

$$CM(\Delta_T) \begin{bmatrix} -2\sigma_T^2 \\ x_0 \\ x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_d \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{7}$$

Left multiplying by $CM^{-1}(\Delta_T)$ results in the final relationship between the Cayley-Menger

matrix and the circumradius established in eq 8.

$$CM^{-1}(\Delta_T)\begin{bmatrix}1\\0\\0\\0\\0\\\vdots\\0\end{bmatrix} = \begin{bmatrix}-2\sigma_T^2\\x_0\\x_1\\x_2\\x_3\\\vdots\\x_d\end{bmatrix} \tag{8}$$

Thus, utilizing only the Cayley-Menger matrix enables computation of the content, circumradius, and barycenter of each simplex within the Delaunay triangulation.

### *Alpha Selection Heuristic*

As clusters are not necessarily globally uniform in size and shape, selection of $\alpha$ must be done on a per cluster basis. A set of N clusters would require manual selection of N $\alpha$ values which is prohibitive for all but the smallest cluster sets. Overcoming this drawback requires a robust and sensitive heuristic for determining each optimal $\alpha$. The simplest and most commonly used heuristic for identifying $\alpha$ is to select the smallest $\alpha$ for which all clustered points belong on or interior to the alpha shape and leaves no disjoint members (Wilson et al., 2009). However, such an approach can lead to erroneous cavities for irregularly sampled volumetric objects. Additionally, an ideal heuristic would provide meaningful information about the potential shapes, such as the volume.

Gardiner et al. (2018), provides an alternate heuristic based instead upon the volume of the corresponding alpha shape for analysis of bacula tomographs. Figure 5 referred to as an alpha volume curve displays the volume with respect to $\alpha$. Here morphologically concave features such as voids appear as sharp decreases in volume as $\alpha$ is decreased, hereby referred to as a volume drop. Additionally sufficiently small alphas result in a final volume drop corresponding

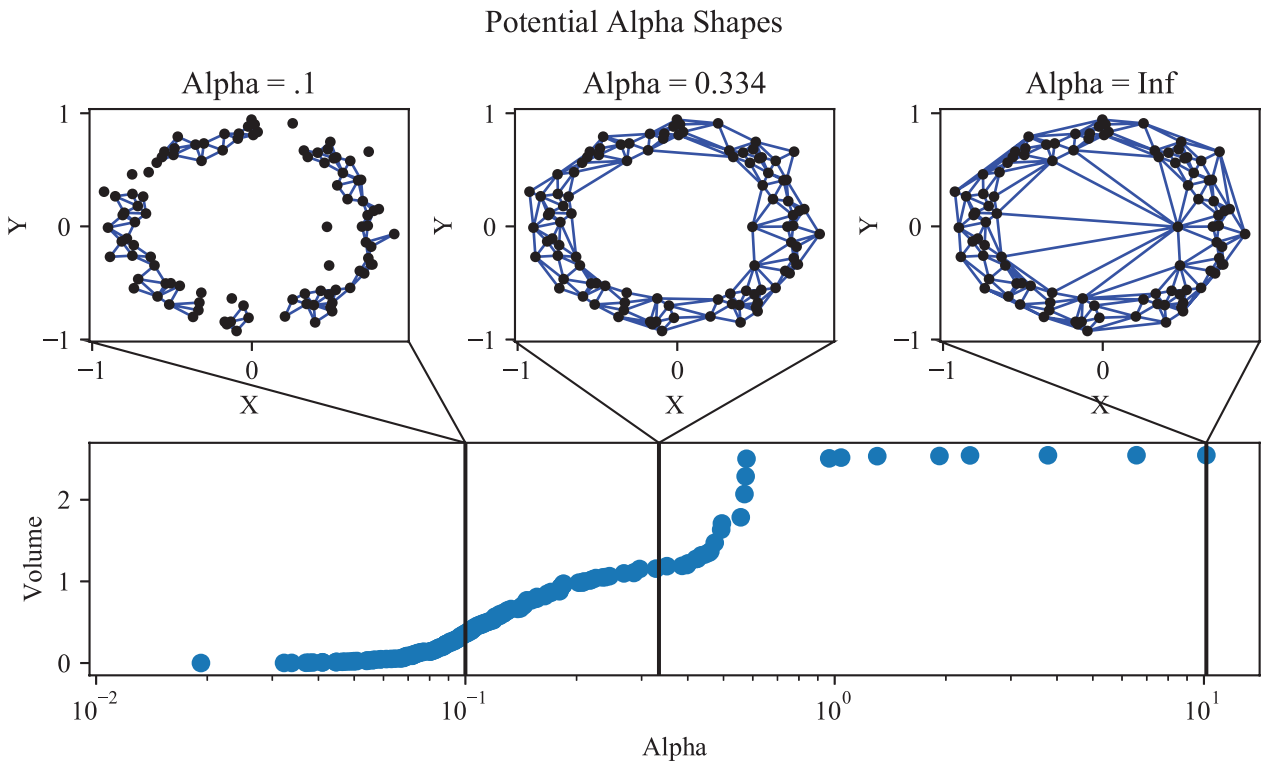



**Fig. 5.** Example plot of complex volume with respect to the selected $\alpha$ for a 2-d loop. Three example $\alpha$ are selected to show a fragmented complex, an ideal complex, and the convex hull.

to the fragmentation of the complex (Gardiner et al., 2018). In the case of Gardiner et al. (2018) the optimal $\alpha$ corresponds with a known volume of the object pre-tomography, which is not applicable to features within APT data. Building upon their work, a heuristic was designed in order to automatically identify the regions of minimal and maximal volume change by first fitting the volume drops.

From figure 5 and Gardiner et al. (2018) it is evident that the volume drops possess an approximately sigmoidal shape with varying levels of asymmetry. An ideal function for the model of such shapes is the generalized logistic function (GLF), alternatively known as the Richard's curve, for which one possible paramaterization is presented in eq 9 and was designed to model growth processes (RICHARDS, 1959).

$$Y(t) = A + \frac{K - A}{(1 + Qe^{-Bt})^{\frac{1}{v}}} \tag{9}$$

Here $Y(t)$ generally denotes the growth property of interest, volume, as a function of the input, $t$. $A$ and $K$ represent the lower and upper asymptotes respectively, which can be bound in (0, 1) if the alpha complex's volume is measured as a fraction of the convex volume. $B$ denotes the unbounded growth rate and $Q$ is a scaling parameter. The parameter $v$ controls the position of the inflection point with respect to the two asymptotes and must be greater than zero. In order to account for multiple volume drops we propose utilizing a summation of multiple GLFs acting over different ranges of $\alpha$. Due to the complexity of identifying the optimal number of GLFs and fitting a complex curve, we restrict our current heuristic to the summation of two GLFs, eq 10. In doing so we limit ASA to accurately characterize features that possess either a singular concave region or a singular $\alpha$ which corresponds to that concave regions.

$$V(\alpha) = A_1 + \frac{A_2 - A_1}{(1 + Q_1 e^{-B_1 \alpha})^{\frac{1}{v_1}}} + \frac{A_3 - A_2}{(1 + Q_2 e^{-B_2 \alpha})^{\frac{1}{v_2}}} \tag{10}$$

Parameters; $A_1$, $A_2$, $A_3$, define the possible asymptotes, where $A_1$ denotes the lower asymptote. $A_2 - A_1$ and $A_3 - A_2$ defines the upper asymptotes of the first and second GLF. Alternatively, $A_3$ is the upper asymptote of the summation and $A_3 - A_2$ denotes the expected value at the region of minimal growth. The remaining parameters remain identical to eq 9, but are subscripted to denote which GLF is parameterized. Furthermore, by constraining our heuristic to eq 10, numeric approximation is unneeded as the first and second derivatives can be analytically derived in eq 11 and eq 12.

$$\begin{aligned} \frac{\delta V(\alpha)}{\delta \alpha} = & \frac{B_1 Q_1 (A_2 - A_1) e^{-B_1 \alpha} (1 + Q_1 e^{-B_1 \alpha})^{\frac{-1}{v_1} - 1}}{v_1} \\ & + \frac{B_2 Q_2 (A_3 - A_2) e^{-B_2 \alpha} (1 + Q_2 e^{-B_2 \alpha})^{\frac{-1}{v_2} - 1}}{v_2} \end{aligned} \tag{11}$$

$$\begin{aligned} \frac{\delta^2 V(\alpha)}{\delta \alpha^2} = & \frac{(A_2 - A_1)(-B_1^2 Q_1^2 (\frac{-1}{v_1} - 1) e^{-2B_1 \alpha} (1 + Q_1 e^{-B_1 \alpha})^{\frac{-1}{v_1} - 2}}{v_1} \\ & + \frac{(A_3 - A_2)(-B_2^2 Q_2^2 (\frac{-1}{v_2} - 1) e^{-2B_2 \alpha} (1 + Q_2 e^{-B_2 \alpha})^{\frac{-1}{v_2} - 2}}{v_2} \\ & - \frac{B_1^2 e^{-B_1 \alpha} (1 + Q_1 e^{-B_1 \alpha})^{\frac{-1}{v_1} - 1}}{v_1} - \frac{B_2^2 e^{-B_2 \alpha} (1 + Q_2 e^{-B_2 \alpha})^{\frac{-1}{v_2} - 1}}{v_2} \end{aligned} \tag{12}$$

From eq 11 and eq 12, the inflection points of the alpha volume curve can be determined

and used as the potential optimal values of $\alpha$. The general heuristic for selection is then as follows: if only a singular inflection point is identified, then the cluster is labeled as convex and the complex for $\alpha = \infty$ is returned, else successive inflection points are checked and the first alpha complex which contains all points on the boundary or interior and has only one connected component is returned. In the event that all inflection points fail this check the simplices composing the convex hull are returned. In the case of a 2-compoment mixture of GLF's, there are at most three inflection points, two of which correspond to regions of maximal growth and a singular inflection point which corresponds to minimal growth.

### *Alpha Complex Composition*

Given the accepted simplices of the alpha complex, the remaining data points that were not used for clustering are reincorporated via rejection sampling to determine a full composition. A point is accepted if it is contained within a simplex of the convex hull and that simplex exists within the alpha complex. If an ion is not accepted for any alpha complex it is labeled as a matrix ion, and can be used to establish the matrix composition. One possible option for the containment check is the condition that the barycentric coordinates must sum to 1 for the containing simplex, presented in eq 4. Alternate approaches such as random or directed walks through the Delaunay triangulation are also applicable. The final composition is calculated according to the relative abundances of ions within the alpha complex.

## ***Experimental Data Acquisition and Analysis Procedure***

An irradiated 304 stainless steel with known radiation induced precipitation at the nanoscale was used as a test case for the proposed algorithm. Conventional manufacturing of APT specimens was performed using a Ga focused ion beam (FIB) for both rough cuts and annular sharpening (Thompson et al., 2007). APT data was collected using a CAMECA LEAP 4000X HR equipped with a UV 355 nm laser operated in laser mode using a pulse energy of 50 pJ/pulse, 200 kHz pulse rate, and a detection rate and temperature of .5% and 50K respectively.

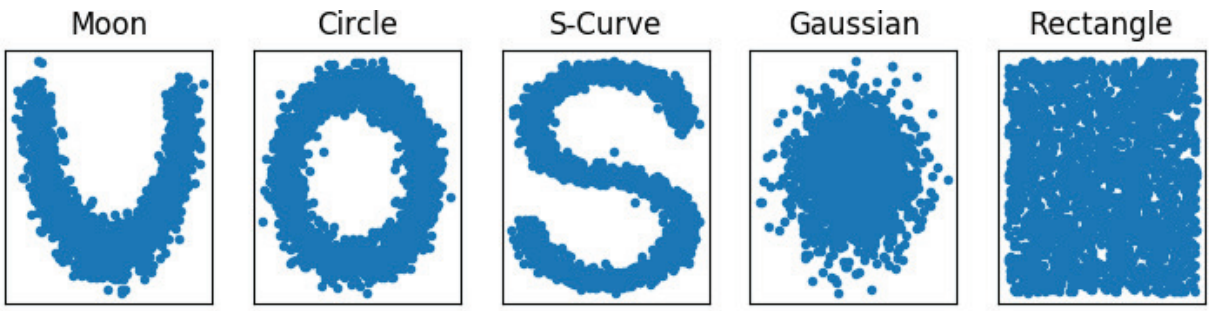



**Fig. 6.** Synthetic datasets used to evaluate the accuracy of the ASA procedure. The datasets consist of three concave and two convex samples.

APT reconstruction was performed using the CAMECA IVAS software package.

### *Synthetic Data Generation and Analysis Procedure*

The ASA framework was applied to a set of five synthetic datasets, presented in figure 6, designed to be possess representative shapes for concave and convex features evident in atom probe tomography data as projected to 2-dimensions for simplicity. The synthetic data was not generated to possess a target density that is reflective of APT data due to the alpha shape procedure having dependencies on the number of points and the uniformity of the density and not the actual point density.

The three concave features types were selected to be representative of dislocation lines, dislocation loops, and shell structures. Dislocation lines are represented through the use of a moon and s-curve, which possess one and two sources of concavity, while dislocation loops and shell structures are represented through a circular dataset. The circle is a direct projection of a loop or cylindrical shell into lower dimensions, while the rotation about the centroid would give a typical spherical shell. The three real equivalents would each have one concave feature. Testing of convex features was accomplished using a rectangular distribution of points to represent planar features such as grain boundaries, as well as a Gaussian blobs to represent spherical precipitates.

The synthetic datasets were treated as post-clustered data, so uncertainties due to clus-

tering and prefiltering would be omitted from the analysis. The goal of this analysis is to determine the variability in results for ASA when applied to similar features and the overall accuracy when applied to convex and concave features with respect to the number of reference points and point density. Finally, the data from the concave and convex test cases are aggregated and used to determine the overall accuracy of the approach with respect to number of reference points and point density.

All concave features; s-curve, moon, circle, were generated through SciPy's dataset utility with Gaussian noise overlaying the true structure (Virtanen et al., 2020). Convex features were generated using numpy's random sampling functionality to produce a uniformly random planar features and a Gaussian feature (van der Walt et al., 2011). Each dataset was generated with a maximum of 2,000 points with which to generate a representative alpha shape. Following generation, the datasets were downsampled without replacement to a lower number of points 100 times, to generate a set of distributions with similar shape to the parent which ASA was then applied to. The number of reference points for these tests ranged from 50 to 1,000 points in 50 point increments. For each trial the inflection point of the concave solution and the percent of the convex volume was recorded. The inflection points of the concave solution were used to determine the true positive rates (TPR) and true negative rates (TNR) and thus the overall accuracy of the approach. The percent volume was used to compare the expected volume distributions for each test case. Surface area/perimeter estimations and composition estimates are left out of this analysis.

### *ASA Implementation*

We combined the above steps into a single framework for robust analysis of composition, surface area, and volume for a set of clusters from a conventional clustering algorithm. The framework is developed purely in Python 3.8.2 with open-source projects and user-written programs. All analyses were performed on a 32 GB Asus Vivobook with a 1.8 GHz Intel Corei7 processor. All required Python packages and codes are documented and will be available on Github and

on PyPa as a package.

Given the input data consisting of spatial coordinates X, Y and Z and a mass-to-charge state ratio (M/Z), and the cluster label of the initial dataset, ions are first grouped according to the cluster label as determined by a clustering algorithm of choice. Each cluster is then processed sequentially and the scipy.spatial package is used to generate a nearest site Delaunay triangulation of the cluster points according to the quickhull algorithm from Qhull (Virtanen et al., 2020; Barber et al., 1996). From the Delaunay triangulation, the 3-simplices and vertices are collected and the Cayley-Menger matrix is calculated to determine the circumradius and volume. Unbounded lower dimensional simplices are omitted from the subsequent alpha shape generation. Simplices are then sorted according to circumradius and the cumulative volume is calculated to generate the alpha volume plot.

In order to fit eq 10, a stochastic optimization is performed using the adaptive memory programming for constrained global optimization (AMPGO) algorithm as provided by the non-linear least-square minimization and curve-fitting for Python (LMFIT) package (Lasdon et al., 2010; Newville et al., 2014). It should be noted that any optimization that adequately explores parameter space and ideally supports unbounded parameters could be used instead. Constrained optimizations such as AMPGO are applicable if the variables bounds are set broad enough to allow sufficient exploration of the parameter space. Despite being a constrained optimization AMPGO was selected due to its competitive convergence rate and ability to avoid becoming trapped in local minima. The basis of AMPGO is the concept of tabu tunneling, where after finding a local minima in the objective function the algorithm attempts to find a point that has a similar evaluation to the current minima and is sufficiently far from all previously identified minima. Previously identified minima are labeled tabu, and are tunneled away from in this construction. An additional local search is performed at each minima in order to more fully explore parameter space. AMPGO has a set of hyperparameters, which are defined in table 1, and the values set for the ASA procedure are provided. Given the fit parameters, the number of local maxima is then calculated using a conventional peak-finding

**Table 1.** User selected hyperparameters for AMPGO stochastic optimization.

| Variable Name | Variable Interpretation | Value |
|---|---|---|
| maxfunevals | Maximum number of function evaluations | 25 |
| totaliter | Number of global iterations | 25 |
| glbtotal | Tolerance for solution acceptance | 1e-5 |
| eps1 | Weighting parameter for objection function during tunneling | .5 |
| eps2 | Perturbation factor to move from current minima prior to tunneling | 1e-4 |

algorithm, data with a singular maxima uses the convex hull for further analysis. In the case of multimodal data, generated from the 2-component GLF mixture, the three inflection points corresponding to maximal, minimal, and maximal growth regions are checked to determine if the cluster points are fully contained within the corresponding alpha complex and that the alpha complex is composed of one connected component. The lowest inflection point is then selected as $\alpha$, and Cayley-Menger analysis is used to determine the area of each remaining 2-simplex bounding an exterior 3-simplex.

Finally, rejection sampling is applied to all unlabeled ions through a scipy.spatial application of a directed search algorithm of the containing simplex which starts at the closest facet of the lifting paraboloid (Virtanen et al., 2020). The search algorithm is based upon Qhull's find best facet function which maps the Delaunay triangulation to a higher dimension generating a lifting paraboloid and computes the distance of a test point to the convex hull of the paraboloid (Barber et al., 1996). This process is repeated for each cluster, to provide two output files. The first of which contains the initial parameters as well as the ASA cluster label, while the latter contains the cluster properties such as volume, surface area, and composition. In the event of overlapping alpha shapes for clusters sharing similar space, the ions existing in the shared content are included in the first cluster analyzed.

## Synthetic Data Results

### *Concave Feature Results*

The overall TPR for the concave datasets is defined according to eq 13 which is graphically presented in figure 7 as a function of point cloud properties. Figure 7a depicts the TPR with respect to the number of downsampled datapoints (N_points) used for the triangulation, while figure 7b is with respect to the effective density given by the number of points divided by the overall content of the alpha shape. These plots are used to assess the conditional accuracy of the classification heuristics performed during ASA. In the event that an individual cluster of 50 ions is observed in APT and is concave then the cluster is correctly labeled as concave approximately 77% of the time when the ASA heuristic is limited to the 2nd inflection point.

$$TPR = \frac{N_{Concave\ Labels|Concave}}{N_{Concave\ Simulations}} \tag{13}$$

To analyze the volume distributions of true concave features, shown in figure 8, as a function of the inflection points the parent distributions were downsampled to fifty points prior to application of ASA. 50 was selected for the amount of downsampling as it provided the maximum difference in TPR according to figure 7. Figure 7 is used to probe the stability and modality of the volume distribution in the worst case scenario with respect to accuracy. For the three concave datasets, where only one true volume is correct the volume distribution should ideally be unimodal with a peak centered on the true value and possessing minimal spread.

### *Convex Feature Results*

The overall TNR for the convex datasets defined in eq 14 and is graphically presented in figure 9 as a function of point cloud properties. Figure 9a depicts the TNR with respect to

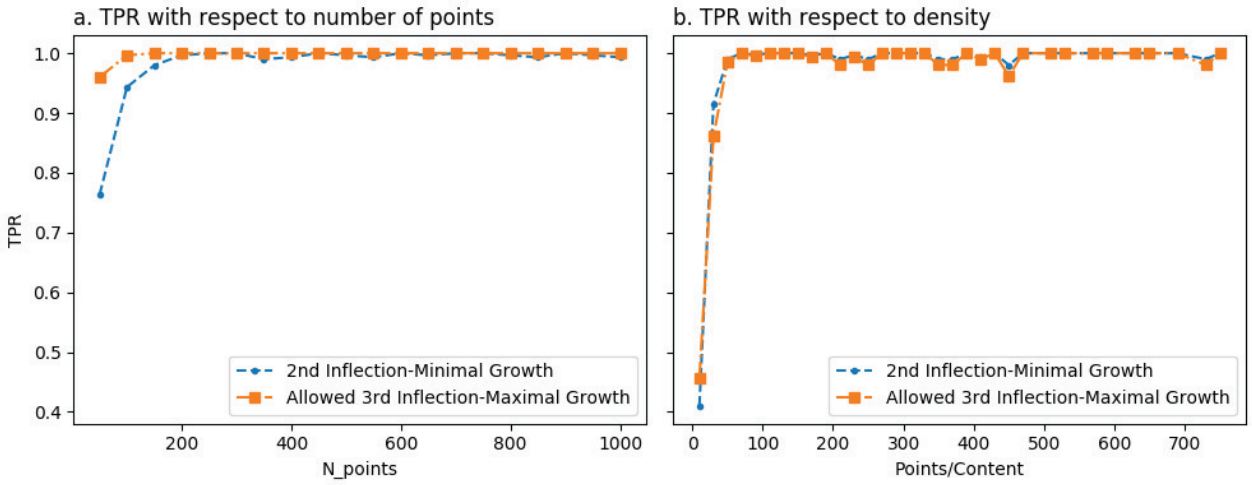



**Fig. 7.** a. True Positive Rate (TPR) of the three concave datasets with respect to the number of reference points (N_points) used in triangulation. b. True Positive Rate (TPR) of the three concave datasets with respect to the point cloud density (Points/Content) as determined by the alpha shape prior to downsampling. Both heuristic options are presented.

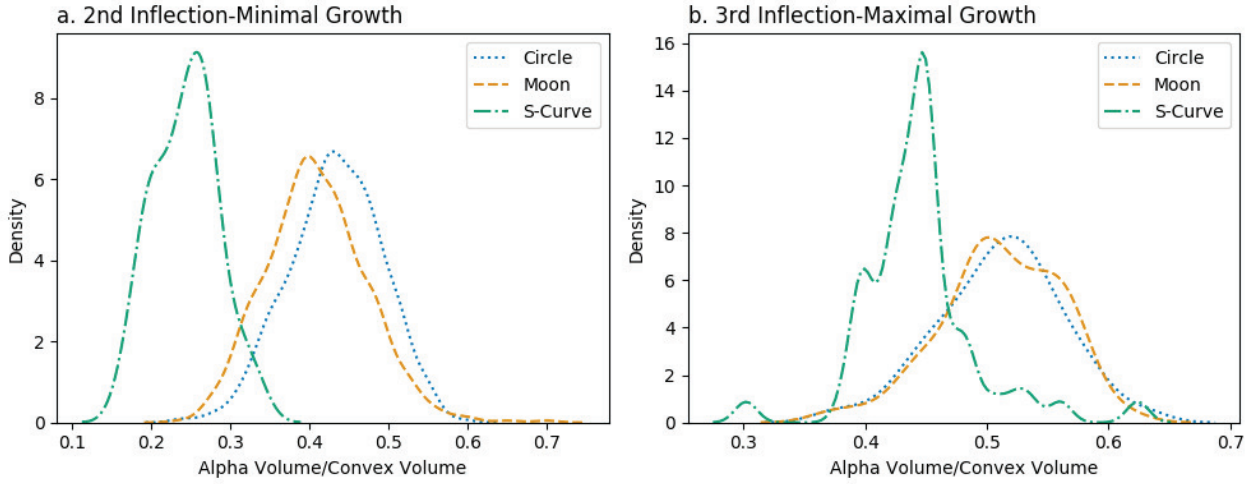



**Fig. 8.** a. Volume distributions of correctly labeled concave datasets given 50 reference points for samples where the 2nd inflection point was optimal. b. Volume distributions of correctly labeled concave datasets given 50 reference points for samples where the 3rd inflection point was optimal. The Volume distributions were generated via a gaussian kernel density estimate (KDE) using Silverman's rule for bandwidth selection.

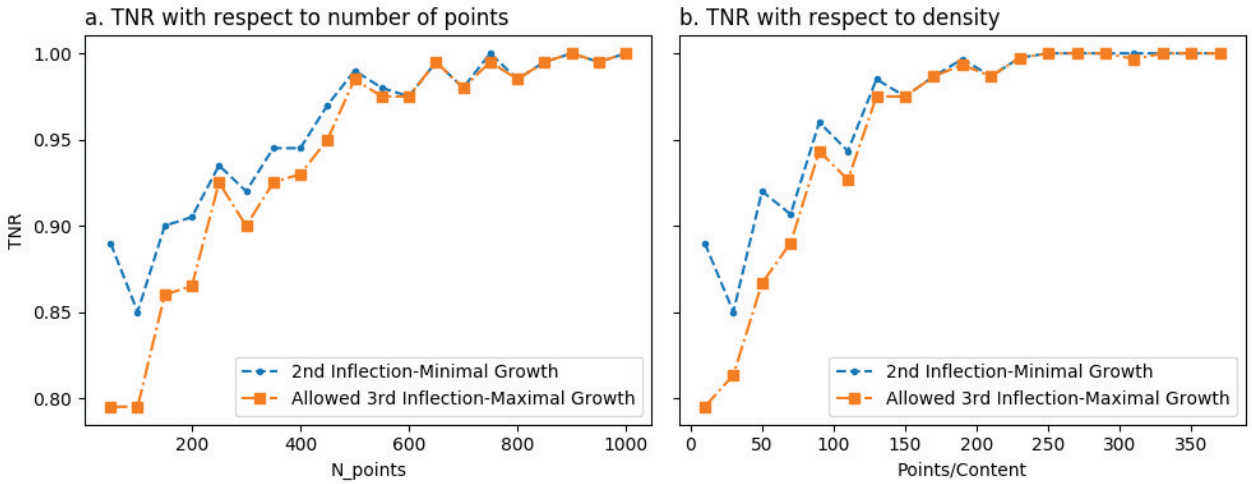



**Fig. 9.** a. True Negative Rate (TNR) of the two convex datasets with respect to the number of reference points (N_points) used in triangulation. b. True Negative Rate (TNR) of the two convex datasets with respect to the point cloud density (points/content) as determined by the alpha shape prior to downsampling. Both heuristic options are presented.

the number of downsampled datapoints (N_points) used for the triangulation, while figure 9b is with respect to the effective density given by the number of points divided by the overall content of the alpha shape. Whereas figure 7 is used to assess a conditional accuracy for concave data, figure 9 provides an effective assessment of classification for the convex datasets given either the number of reference points or the effective point density.

$$TNR = \frac{N_{Convex\ Labels|Convex}}{N_{Convex\ Simulations}} \tag{14}$$

To analyze the volume distributions of false concave features, shown in figure 10, as a function of the inflection points the parent convex distributions were downsampled to fifty points prior to application of ASA. 50 was maintained for the amount of downsampling because it provided the maximum difference in TNR according to figure 9 and was the downsampling amount selected for comparing the concave volume distributions. Volume distributions are once more used to assess stability in the resulting alpha shapes for the same feature. Similar

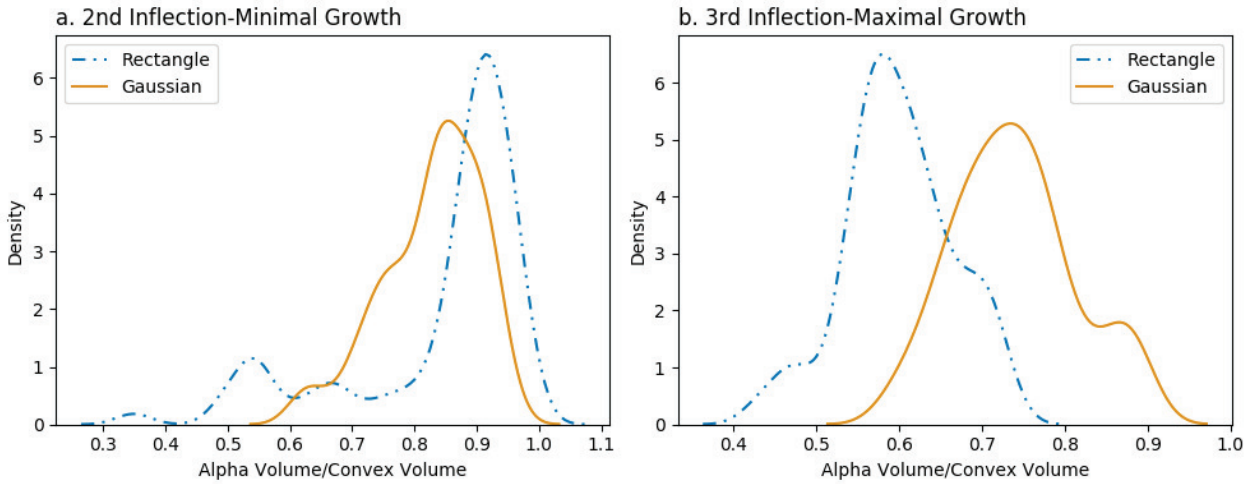



**Fig. 10.** a. Volume distributions of incorrectly labeled convex datasets given 50 reference points for samples where the 2nd inflection point was optimal. b. Volume distributions of incorrectly labeled convex datasets given 50 reference points for samples where the 3rd inflection point was optimal. The volume distributions were generated via a gaussian kernel density estimate (KDE) using Silverman's rule for bandwidth selection.

to figure 8 an ideal heuristic would provide a sharply peaked unimodal distribution, however for the convex datasets the selection of $\alpha$ and not just the labeling step is expected to contribute to the variance in the computed volume.

### *Aggregated Results*

The overall accuracy for all datasets is defined according to eq 15 which is graphically presented in figure 11 as a function of point cloud properties. Figure 11a displays the accuracy with respect to the number of downsampled data points (N_points) used for the triangulation, while figure 11b is with respect to the effective density given by the number of points divided by the overall content of the alpha shape. This aggregate accuracy does not condition on a know truth for a feature being concave or convex, and thus is more reflective of the clusters extracted from APT.

$$Accuracy = \frac{N_{Concave\ Labels|Concave} + N_{Convex\ Labels|Convex}}{N_{Concave\ Simulations} + N_{Convex\ Simulations}} \tag{15}$$

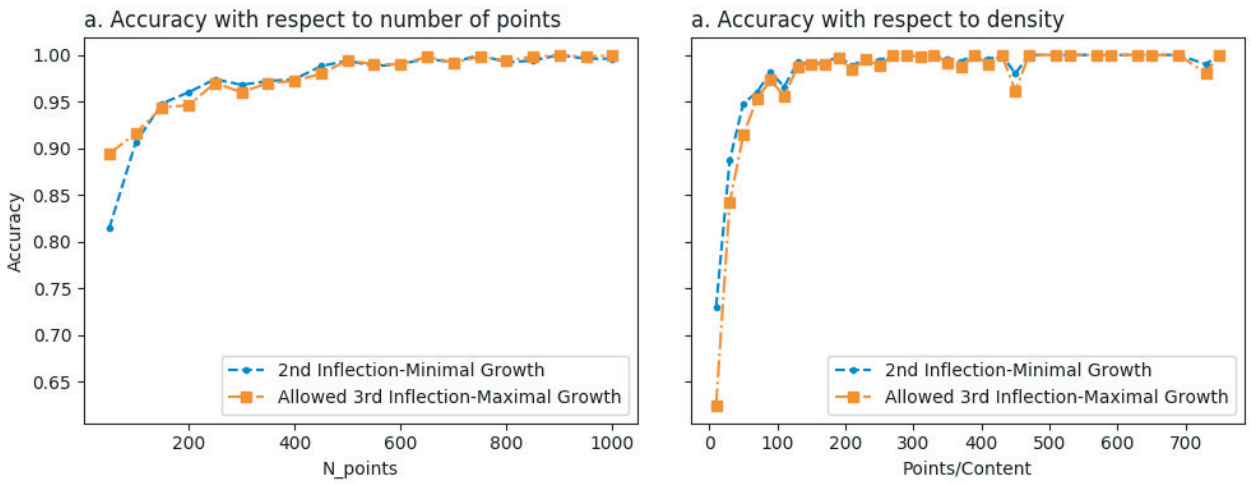



**Fig. 11.** a. Aggregated accuracy over all synthetic datasets with respect to the number of reference points (N_points) used in triangulation. b. Aggregated accuracy over all synthetic datasets with respect to the point cloud density (points/content) as determined by the alpha shape prior to downsampling. Both heuristic options are presented.

The results from the volume distribution analysis for the correctly labeled concave data are shown in figure 8 and the incorrectly labeled convex data from figure 10 are aggregated and displayed in figure 12.

## Experimental Results

### *Ionic Distributions*

304SS is a 18/8 Cr/Ni austenitic steel alloy by weight percent. Si, Mn, P, S, Mo, and N are all expected contaminants in 304SS. The presence of Si in 304SS is often manifested through grain boundary and dislocation enrichment as well as precipitation. As the goal of using 304SS was to test the application of ASA to a variety of microstructural features investigation of the ionic distributions were focused upon Fe, Mn, Ni, and Si observed as elemental peaks. Analysis of Si was restricted to the $Si^{2+}$ avoid overlap between the $Si^{1+}$, $Fe^{2+}$, and $Ni^{2+}$ peaks. Mn and Ni

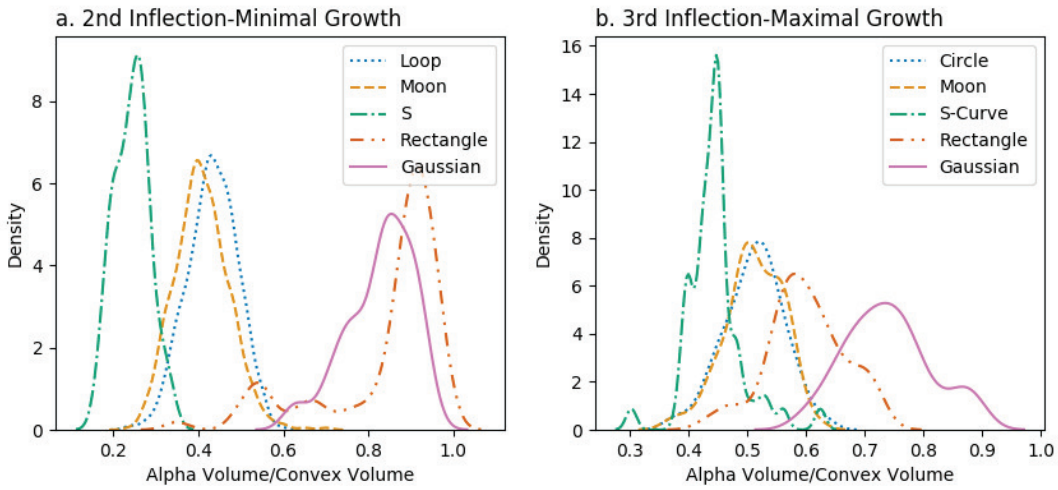


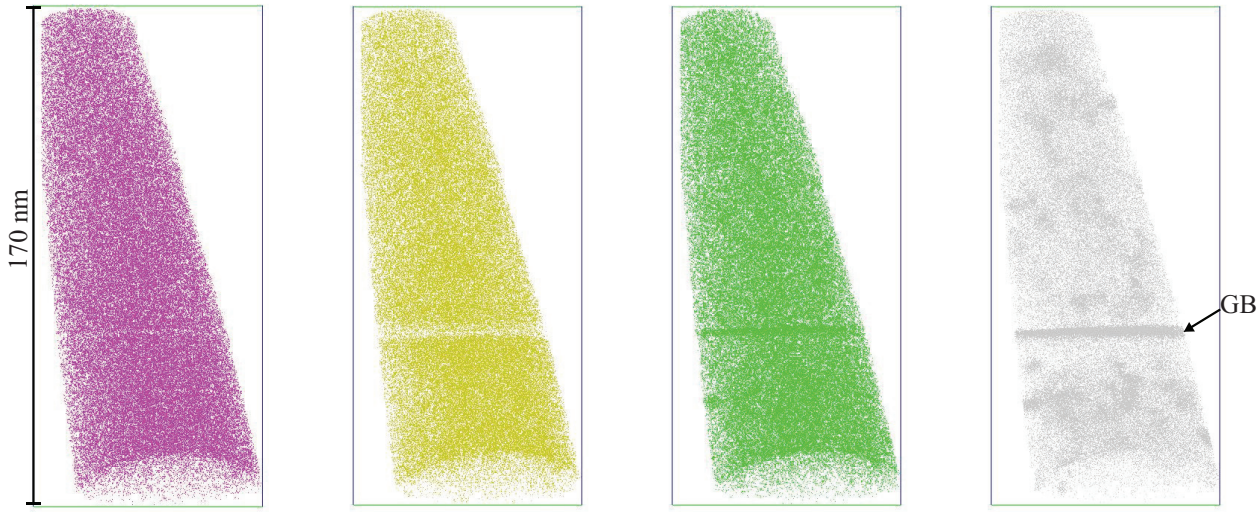

**Fig. 12.** a. Volume distributions of all synthetic datasets given 50 reference points for samples where the 2nd inflection point was optimal. b. Volume distributions of all synthetic datasets given 50 reference points for samples where the 3rd Inflection Point was optimal. The volume distributions were generated via a gaussian kernel density estimate (KDE) using silverman's rule for bandwidth selection.

commonly accompany Si enrichment in 304SS while the Fe concentration should be depleted within the vicinity of such features, as shown in figure 13.

## *Application of OPTICS and ASA*

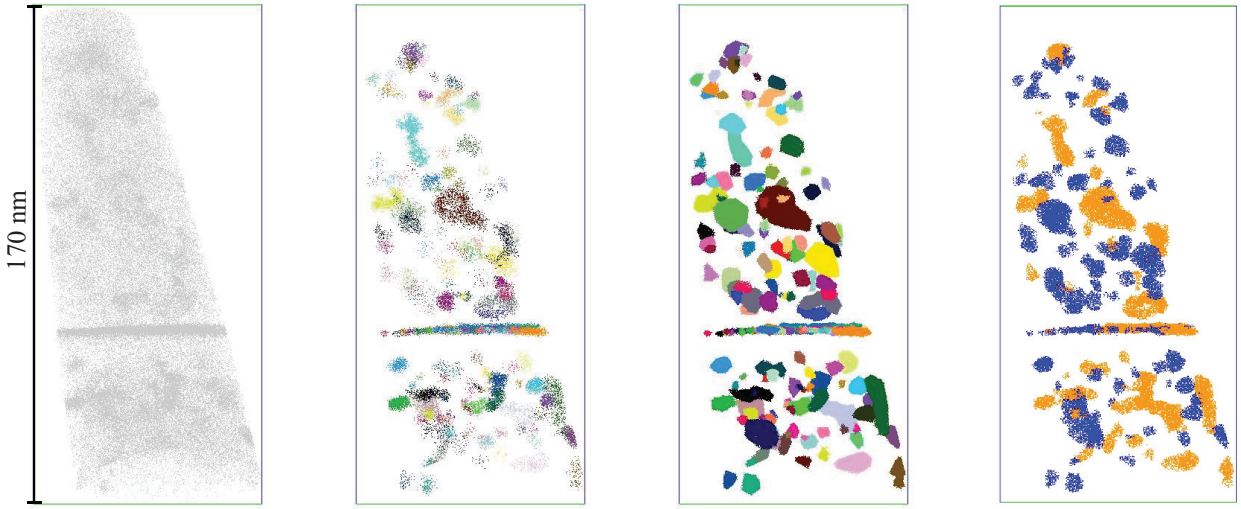
Selection of Si ions acted as the prefiltering step for both OPTICS and subsequent applications of isoconcentration surfaces. The selection of Si ions only resulted in a reduction in the number of ions from 30 million down to approximately 130,000. OPTICS analysis utilized $\epsilon$ of 25 and MinPts of 20, producing the set of 203 clusters shown in figure 14b.

Commonly, post clustering analysis of precipitates is focused on the ratios of their ionic constituents as presented in figure 15. However, unlike the conventional analysis the application of ASA enables comparison of all ionic species as opposed to just those used for clustering.




**(a)** Fe Ion Map **(b)** Mn Ion Map **(c)** Ni Ion Map **(d)** Si Ion Map

**Fig. 13.** 304SS ion distribution maps for Fe, Mn, Ni, and Si Ions. Ni and Si show enrichment along the grain boundary and within both grains via precipitates. Mn shows a relative depletion at the grain boundary.




**(a)** Si Filtering **(b)** OPTICS Clusters **(c)** ASA Clusters **(d)** ASA Classification

**Fig. 14.** APT Maps depicting the initial Si distribution, the clustering output of OPTICS for $\epsilon$ of 25 and MinPts of 20, as well as the final cluster output and classification output corresponding to the ASA framework. In figure 14d orange points represent ions belonging to concave features while blue points represent those belonging to convex features.

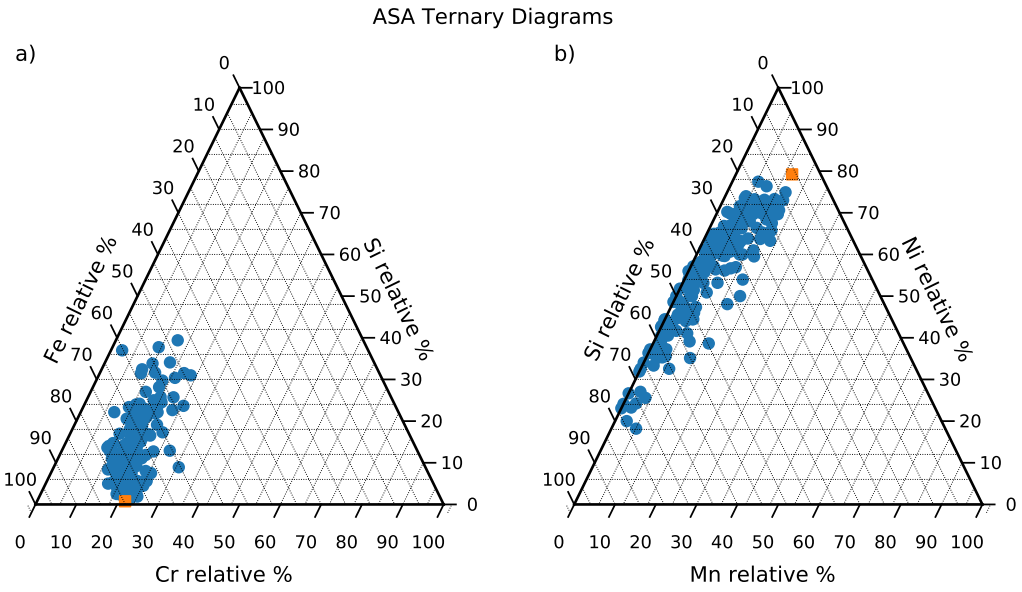
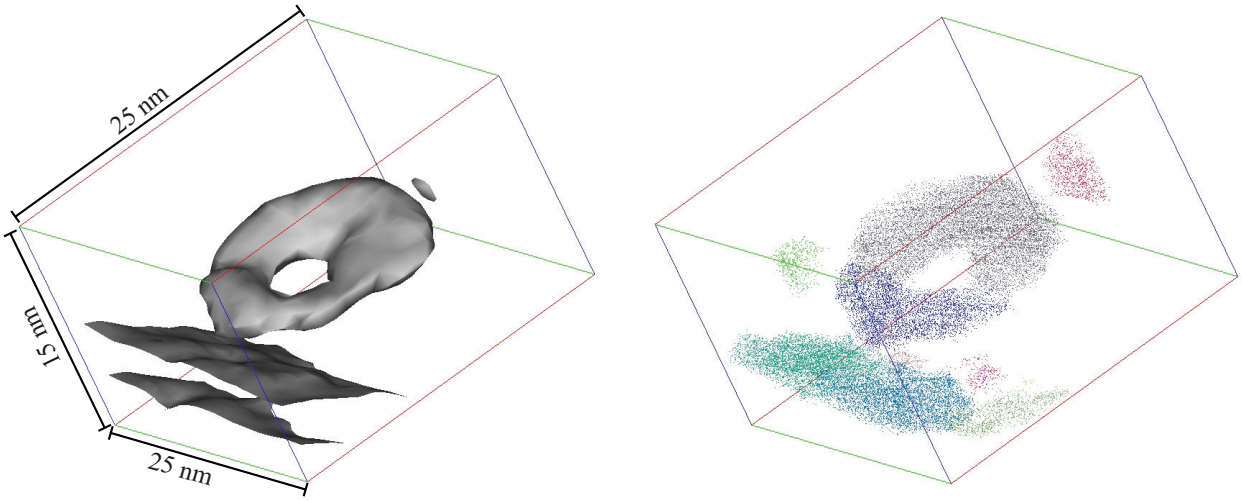



**Fig. 15.** a. Ternary diagram consisting of Fe, Cr, and Si ions. b. Ternary diagram consisting of Si, Mn, and Ni ions. Blue circular points denote the relative abundances of selected ions for clusters, while the orange square indicates species abundance within the matrix.

### *Analysis of a Dislocation Loop*

In order to assess the similarity of ASA and Isosurfaces for analysis of dislocation loops, the previously identified dislocation and grain boundary were extracted with a cubic region of interest. As seen in figure 16a a 2.8 at% Si isoconcentration surface extracts two features of interest: the dislocation loop and a portion of the grain boundary. In figure 16b, the post-ASA clusters representing the dislocation loop and grain boundary are shown to be fragmented due to the inability of OPTICS to preserve these features as whole extracted clusters. The application of OPTICS only found 2,240 clustered Si ions in this region of interest, while ASA reintroduced an additional 29,367 ions from the entire mass spectrum. The dislocation loop is composed of two fragments which were forcibly merged into one cluster prior to reapplication of ASA in order to ensure the extracted properties are representative of the complete loop and comparable to those reported by the isoconcentration surfaces.

The ionic concentrations of different species and size descriptors such as volume, surface




**(a)** 2.8 at% Si Isosurface

**(b)** Fragmented Loop-ASA

**Fig. 16.** Si enriched dislocation loop and grain boundary as observed by Isosurfaces and the ASA framework.

**Table 2.** Comparison of dislocation loop concentrations determined by isococentration surface method and ASA

| | Fe | Cr | Cu | Ni | Si | Mn |
|---|---|---|---|---|---|---|
| 2.8 at% Si Iso | 66.59 | 14.65 | .1431 | 12.86 | 4.95 | .818 |
| 2.4 at% Si Iso | 67.56 | 14.88 | .1439 | 12.13 | 4.32 | .862 |
| ASA | 67.22 | 14.87 | .1723 | 11.84 | 5.01 | .88 |

area, and ion count for ASA as well as two isosurfaces with 2.4% and 2.8% Si are presented in tables 2 and 3 respectively.

## Discussion on Synthetic Data

### *Heuristic Accuracy*

The TPR of the alpha selection heuristic when applied to the concave datasets was presented in figures 7 as a function of the number of datapoints as well as the point density. In both scenarios the TPR approaches 1 as the number of points for downsampling and thus the density

[]

**Table 3.** Comparison of Dislocation Loop Size measured in number of ions, volume, and surface area determined according to Isosurfaces and ASA

| | Number of Ions | Volume ($nm^3$) | Surface Area ($nm^2$) |
|---|---|---|---|
| 2.8% Si Iso | 12678 | 370.6 | 416.6 |
| 2.4% Si Iso | 16820 | 477.3 | 487.9 |
| ASA | 16964 | 470.6 | 524.8 |

increases. The differences between using solely using the 2nd inflection point and allowing the use of the 3rd inflection point of the alpha volume curve is only evident when using N_points as the metric, where allowing the 3rd inflection point results in a higher TPR for N_points $< 250$.

Figure 9 instead displays the TNR of the heuristic for the evaluated convex datasets. Regardless of the constraints on inflection points applied to the heuristic the TNR approaches 1 as the number of points for downsampling and thus the density increases. The differences between using solely using the 2nd inflection point and allowing the use of the 3rd inflection point of the alpha volume curve is evident for both metrics where allowing the 2nd inflection point results in a higher or equal TNR for all N_points and densities.

In order to assess the overall efficacy of the heuristic the accuracy was computed considering both the concave and convex datasets as shown in 11. As expected the accuracy displays similar trends to the TNR and TPR for it approaches 1 for an increasing number of downsampled points and thus density. The behavior of the accuracy with respect to density is consistent regardless of which heuristic is used, but limiting the inflection point selection to the 2nd point maximizes the accuracy for most densities. The only noticeable deviation in accuracy with respect to N_points occurs for 50 points where restricting selection to the 2nd inflection point results in an accuracy drop of approximately 0.15.

### *Percent Volume Distributions*

From the concave data presented in figure 8 depicting the percent volume distributions for both heuristic options it is clear that using the third inflection point introduces a right translation of 0.1 to a more convex percent volume. Additionally, the moon and circle datasets experienced a left skew while the s-curve experienced a right skew.

In opposition to this behavior is the convex data's percent volume distributions visible in figure 10. When restricted to the 2nd inflection point of the alpha volume curve, the rectangle dataset was peaked at a highly convex percent ratio of 0.9, with a long left skew extending down to 0.4, while the gaussian dataset was peaked near 0.85 and possessed a lesser left skew. The data with an optimal 3rd inflection point exhibit a left translation of approximately 0.3 and a diminishing amount of left skew in the case of the rectangle dataset. The Gaussian dataset showcases a lesser translation of 0.1 and a possible right skew.

The combined percent volume distributions are depicted in figure 12. In figure 12a the distributions are shown when the selection heuristic is restricted to the 2nd inflection point of the alpha volume curve. In this scenario given the tested synthetic datasets there is clear separation between the false concave data with a convex parent distribution and true concave data occurring at 0.6. When the heuristic is allowed to select the 3rd inflection point as an optimal $\alpha$, the right and left shifts in the concave and convex data respectively result in a smearing of volume distributions such that there is no longer a clear separation in the false concave and true concave data.

### *Final Heuristic Selection*

In determining the constraints on the final selection of $\alpha$ it is imperative to consider which processes are influenced by the number of datapoints and which are influenced by a point density. From the perspective of atom probe tomography it would be ideal if the ASA frameworks accuracy is primarily impacted by point density as that has historically been the method for assessing clustering accuracy given a background density. However, assuming that the point

cloud possesses a uniform and isotropic density the efficacy of the delaunay triangulation and the corresponding alpha shape is governed purely by the number of datapoints. Additionally, the fitting process of the alpha volume curve is dependent on the number of simplices in the delaunay triangulation and the proportion of the convex volume which is contained within concave regions. Neither of which can be related to the initial point density. Thus, looking at the accuracy with respect to the number of datapoints is most representative of the ASA framework.

From our synthetic datasets this would suggest that allowing an unrestricted selection of inflection point is ideal. However, as noted in the discussion of figure 12, there is significant smearing of the volume distributions that is not present when we restrict our selection to the 2nd inflection point. Allowing the enhanced overlap in volume distributions would make it exceedingly difficult to apply a secondary categorizing step based on the percent volume as suggested by Ghamarian et al. (2020) for their approach to quantifying solute topology. As such we propose that the $\alpha$ selection process should be limited to the 2nd inflection point of the alpha volume curve despite a slight loss in accuracy for datasets with less than 200 datapoints in order to maintain tighter volume distributions.

## Experimental Discussion

### *OPTICS and ASA*

The high visibility of Si enrichment compared to Mn and Ni only Si ions, as evident in figure 13d was the justification for clustering only the Si ions via the OPTICS algorithm. OPTICS performed as expected within each grain identifying visibly observable clusters as well as lower relative density features. Of particular note is the grain boundary which was segmented into many sub-clusters due to the reachability distance corresponding to points at opposite ends of the boundary growing prohibitively large. Additionally, directly above the grain boundary on the right-hand side exists a loop-like structure that was fragmented by OPTICS.

For the majority of clusters, as observed in figure 14, the ASA and OPTICS results are comparable, with the most salient observation being that post-ASA the features are higher density. This is due to the rejection sampling component of ASA which attempts to reintroduce the ions absent from OPTICS back into the clusters. Many of the features observed after ASA maintain concavities visible in the cluster results of OPTICS. One exception to this is the brown cluster front and center in the topmost grain. In the case of OPTICS there is a clear internal cavity as well as an outward concave region, however ASA only identified the outer region. It is believed that OPTICS returned a combined feature consisting of a dislocation loop and precipitate in close proximity. If the two concave regions had varying representative $\alpha$ values one of the assumptions for ASA was violated explaining the inaccuracy.

In figure 15a, due to the singular cluster within composition space there appears to be no significant relationship between Mn, Fe, and Si. Investigating the relationships between Si, Mn, and Ni, in figure 15b, shows an elongated distribution of variable Si/Ni concentration possessing low Mn amounts as well as a potential secondary distribution with a tighter Ni and higher Mn spread. This observation may correspond with clusters that originate in the grain boundary due to the observed Ni, Si enrichment and Mn depletion. Finally, due to the nature of rejection sampling, all ions that were rejected from the set of clusters are used to estimate the matrix composition denoted by the orange squares.

### *Analysis of a Dislocation Loop*

The Si isconcentration surface presented in figure 16a shows the beginnings of a pinch point which may explain the fragmentation of the dislocation loop into two clusters that OPTICS reported. This hypothesis is also supported by the presence of the adjacent grain boundary. During irradiation both grain boundaries and dislocations act as competing defect sinks for Si, which could result in a lower density of Si ions in the bordering portion of the loop.

Analysis of the compositions from table 2 shows that ASA reports lower concentrations of Ni1 and higher amounts of Mn, Cr, Cu, and Fe relative to the 2.8 at% Si isoconcentration

surface. While ASA reports higher Si compared to this isoconcentration surface the difference is likely to be within experimental error. Applying a lower threshold of 2.4 at% Si results in the Ni and Si content dropping and the amounts of Cr, Mn, and Fe to rise while Cu remains similar. This change results in the Fe and Ni content exceeding that of ASA, while Mn and Cr approach that reported by ASA. The difference in reported Si content is amplified with an approximate difference of 0.7 at%. It is believed that the 2.4 at% Si isoconcentration surface results in a higher proportion of the matrix being contained within the surface, explaining the overall decrease in Si and increase in expected matrix components such as Fe and Cr/

The size metrics provided in table 3 show that ASA reports a total of 16,964 ions within the dislocation loop which exceeds that of the 2.8 at% Si isoconcentration surface by 3,000 ions, while the 2.4 at% Si isoconcentration surface contains a comparable 16,820 ions. As expected from observations with contained ions ASA also reports a higher volume and surface area of 470.6 $nm^3$ and 524.8 $nm^2$ compared to the 2.8 at% isoconcentration surface. Interestingly, while the 2.4 at% Si isoconcentration surface reports a comparable volume of 477.3 $nm^3$ the surface area is only 487.9 $nm^2$. This may be due to the inherent smoothing of the data that isoconcentration surface introduce during their voxelization and kernel smoothing steps, while the ASA framework does not inherently guarantee smooth features nor does the framework currently include a smoothing process.

## Conclusions

In this work a novel and automated post-clustering framework based on alpha shapes was developed to enhance quantification of volume, surface area, and composition for solute clusters identified by arbitrary clustering algorithms using simulated and experimental APT data. The Alpha Shape Analysis (ASA) framework provided accuracy rates exceeding 80% regardless of the number of points provided and minimized the overlap of resulting percent volume distributions for a set of three concave and two convex datasets. Additionally, ASA was applied to an experimental 304SS data set containing silicon (Si) enrichment enabling measurement of

volume, surface area, and composition for a set of 203 clusters as identified by the OPTICS algorithm with only Si ions as the ions of interest. The ASA clusters maintained visual equivalence with the clusters as reported by OPTICS in the majority of cases and maintains any deficiencies that the cluster algorithm of choice possesses, i.e. fragmentation of clusters with variable Si content such as the grain boundary as observed by OPTICS.

In the current state ASA assumes that each provided cluster possesses uniform and isotropic density. One avenue to address this current limitation is the implementation of the anisotropic density-scaled alpha shape reconstruction procedure which would minimize connections over sparse regions and enable a scaling $\alpha$ parameter as a function of local density. Additionally, the alpha selection heuristic can only account for concave features with a singular representative $\alpha$ value due to the reliance on a 2-component GLF mixture. Ideally, a Dirichlet mixture model composed of GLFs could be designed to enable higher dimensional fits without overfitting. The final and most accessible direction is to include topological characterization of the alpha shapes via Betti numbers.

**Acknowledgements**

This work was supported as part of FUTURE (Fundamental Understanding of Transport Under Reactor Extremes), an Energy Frontier Research Center funded by the U.S. Department of Energy, Office of Science, Basic Energy Sciences. E.K.S acknowledges a graduate fellowship through the Nuclear Regulatory Committee.

## References

**Ankerst, M., Breunig, M.M., Kriegel, H.P. & Sander, J.** (1999). Optics: Ordering points to identify the clustering structure, *SIGMOD Rec* **28**, 49–60, URL `https://doi.org/10.1145/304181.304187`.

**Bachhav, M., Odette, G. & Marquis, E.** (2014). $\alpha$ precipitation in neutron-irradiated fe–cr alloys, *Scripta Materialia* **74**, 48–51.

**Bailey, N.A., Stergar, E., Toloczko, M. & Hosemann, P.** (2015). Atom probe tomography

analysis of high dose ma957 at selected irradiation temperatures, *Journal of Nuclear Materials* **459**, 225–234.

**Barber, C.B., Dobkin, D.P. & Huhdanpaa, H.** (1996). The quickhull algorithm for convex hulls, *ACM Trans Math Softw* **22**, 469–483, URL https://doi.org/10.1145/235815.235821.

**Bas, P., Bostel, A., Deconihout, B. & Blavette, D.** (1995). A general protocol for the reconstruction of 3d atom probe data, *Applied Surface Science* **87-88**, 298 – 304, URL http://www.sciencedirect.com/science/article/pii/0169433294005613, proceedings of the 41st International Field Emission Symposium.

**Coxeter, H.S.M.** (1930). The circumradius of the general simplex, *The Mathematical Gazette* **15**, 229–231.

**Edelsbrunner, H., Kirkpatrick, D. & Seidel, R.** (1983). On the shape of a set of points in the plane, *IEEE Transactions on Information Theory* **29**, 551–559.

**Edelsbrunner, H. & Mücke, E.P.** (1994). Three-dimensional alpha shapes, *ACM Trans Graph* **13**, 43–72, URL https://doi.org/10.1145/174462.156635.

**Ester, M., Kriegel, H.P., Sander, J. & Xu, X.** (1996). A density-based algorithm for discovering clusters in large spatial databases with noise, *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, KDD'96, 226–231, AAAI Press.

**Felfer, P. & Cairney, J.** (2016). A computational geometry framework for the optimisation of atom probe reconstructions, *Ultramicroscopy* **169**, 62 – 68, URL http://www.sciencedirect.com/science/article/pii/S0304399116300961.

**Felfer, P., Ceguerra, A., Ringer, S. & Cairney, J.** (2015). Detecting and extracting clusters in atom probe data: A simple, automated method using voronoi cells, *Ultramicroscopy* **150**, 30 – 36, URL http://www.sciencedirect.com/science/article/pii/S0304399114002307.

**Gardiner, J.D., Behnsen, J. & Brassey, C.A.** (2018). Alpha shapes: determining 3d shape complexity across morphologically diverse structures, *BMC Evolutionary Biology* **18**, 184, URL https://doi.org/10.1186/s12862-018-1305-z.

**Gault, B., Moody, M.P., Cairney, J.M. & Ringer, S.P.** (2012). New York, NY: Springer New York, URL https://doi.org/10.1007/978-1-4614-3436-8_1.

**Geiser, B., Larson, D., Oltman, E., Gerstl, S., Reinhard, D., Kelly, T. & Prosa, T.** (2009). Wide-field-of-view atom probe reconstruction, *Microscopy and Microanalysis* **15**, 292 – 293.

**Geiser, B.P., Kelly, T.F., Larson, D.J., Schneir, J. & Roberts, J.P.** (2007). Spatial distribution maps for atom probe tomography, *Microscopy and Microanalysis* **13**, 437–447.

**Ghamarian, I. & Marquis, E.** (2019). Hierarchical density-based cluster analysis framework for atom probe tomography data, *Ultramicroscopy* **200**, 28 – 38, URL http://www.sciencedirect.com/science/article/pii/S0304399118303267.

**Ghamarian, I., Yu, L.J. & Marquis, E.A.** (2020). Quantification of solute topology in atom probe tomography data: Application to the microstructure of a proton-irradiated alloy 625, *Metallurgical and Materials Transactions A* **51**, 42–50, URL https://doi.org/10.1007/s11661-019-05520-6.

**Guo, Z., Sha, W. & Vaumousse, D.** (2003). Microstructural evolution in a ph13-8 stainless steel after ageing, *Acta Materialia* **51**, 101 – 116, URL http://www.sciencedirect.com/science/article/pii/S1359645402003531.

**Jenkins, B.M., London, A.J., Riddle, N., Hyde, J.M., Bagot, P.A. & Moody, M.P.** (2020). Using alpha hulls to automatically and reproducibly detect edge clusters in atom probe tomography datasets, *Materials Characterization* **160**, 110078, URL http://www.sciencedirect.com/science/article/pii/S1044580319329511.

**Lasdon, L., Duarte, A., Glover, F., Laguna, M. & Martí, R.** (2010). Adaptive memory programming for constrained global optimization, *Computers & Operations Research* **37**, 1500 – 1509, URL http://www.sciencedirect.com/science/article/pii/S0305054809002937, operations Research and Data Mining in Biological Systems.

**Marquis, E.A. & Hyde, J.M.** (2010). Applications of atom-probe tomography to the characterisation of solute behaviours, *Materials Science and Engineering: R: Reports* **69**, 37 – 62, URL http://www.sciencedirect.com/science/article/pii/S0927796X10000525.

**Miller, M. & Kenik, E.** (2004). Atom probe tomography: A technique for nanoscale characterization, *Microscopy and Microanalysis* **10**, 336–341.

**Newville, M., Stensitzki, T., Allen, D.B. & Ingargiola, A.** (2014). LMFIT: Non-Linear Least-Square Minimization and Curve-Fitting for Python, URL `https://doi.org/10.5281/zenodo.11813`.

**RICHARDS, F.J.** (1959). A Flexible Growth Function for Empirical Use, *Journal of Experimental Botany* **10**, 290–301, URL `https://doi.org/10.1093/jxb/10.2.290`.

**Sander, J., Qin, X., lu, Z., Niu, N. & Kovarsky, A.** (2003). Automatic extraction of clusters from hierarchical clustering representations, 75–87.

**Sommerville, D.M.Y.** (1958). *An introduction to the geometry of N dimensions*, Dover New York.

**Thompson, K., Lawrence, D., Larson, D., Olson, J., Kelly, T. & Gorman, B.** (2007). In situ site-specific specimen preparation for atom probe tomography, *Ultramicroscopy* **107**, 131 – 139, URL `http://www.sciencedirect.com/science/article/pii/S0304399106001203`.

**van der Walt, S., Colbert, S.C. & Varoquaux, G.** (2011). The numpy array: A structure for efficient numerical computation, *Computing in Science Engineering* **13**, 22–30.

**Vaumousse, D., Cerezo, A. & Warren, P.** (2003). A procedure for quantification of precipitate microstructures from three-dimensional atom probe data, *Ultramicroscopy* **95**, 215 – 221, URL `http://www.sciencedirect.com/science/article/pii/S0304399102003194`, iFES 2001.

**Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S.J., Brett, M., Wilson, J., Jarrod Millman, K., Mayorov, N., Nelson, A.R.J., Jones, E., Kern, R., Larson, E., Carey, C., Polat, İ., Feng, Y., Moore, E.W., Vand erPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E.A., Harris, C.R., Archibald, A.M., Ribeiro, A.H., Pedregosa, F., van Mulbregt, P. & Contributors, S...** (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python, *Nature Methods* **17**, 261–272.

**Wang, J., Schreiber, D.K., Bailey, N., Hosemann, P. & Toloczko, M.B.** (2019). The application of the optics algorithm to cluster analysis in atom probe tomography data, *Microscopy and Microanalysis* **25**, 338–348.

**Wilson, J.A., Bender, A., Kaya, T. & Clemons, P.A.** (2009). Alpha shapes applied to molecular shape characterization exhibit novel properties compared to established shape descriptors, *Journal of Chemical Information and Modeling* **49**, 2231–2241, URL `https://doi.org/10.1021/ci900190z`, pMID: 19775113.