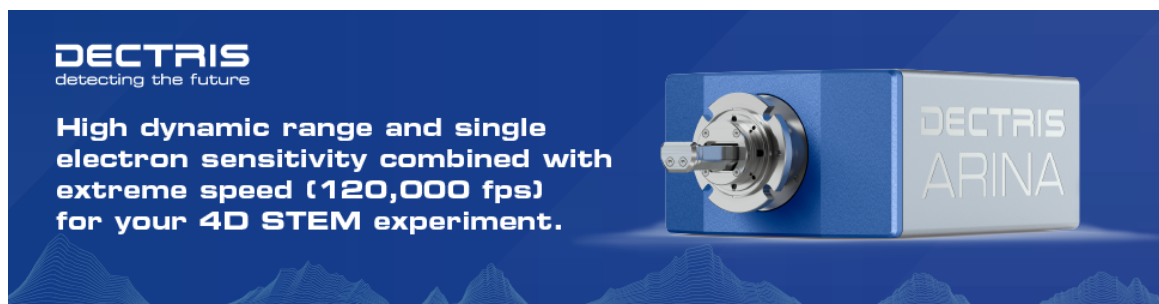





# A Machine Learning Framework for Quantifying Chemical Segregation and Microstructural Features in Atom Probe Tomography Data

Alaukik Saxena, Nikita Polin, Navyanth Kusampudi, Shyam Katnagallu, Leopoldo Molina-Luna, Oliver Gutfleisch, Benjamin Berkels, Baptiste Gault, Jörg Neugebauer, Christoph Freysoldt



# A Machine Learning Framework for Quantifying Chemical Segregation and Microstructural Features in Atom Probe Tomography Data

Alaukik Saxena<sup>1</sup> , Nikita Polin<sup>1</sup>, Navyanth Kusampudi<sup>1</sup> , Shyam Katnagallu<sup>1</sup>, Leopoldo Molina-Luna<sup>2</sup>, Oliver Gutfleisch<sup>3</sup>, Benjamin Berkels<sup>4</sup>, Baptiste Gault<sup>1,5</sup> , Jörg Neugebauer<sup>1</sup>, and Christoph Freysoldt<sup>1,\*</sup>

<sup>1</sup>Max-Planck-Institut für Eisenforschung GmbH, Max-Planck-Straße 1, 40237 Düsseldorf, Germany

<sup>2</sup>Department of Materials and Earth Sciences, Technische Universität Darmstadt, Peter-Grünberg-Straße 2, 64287 Darmstadt, Germany

<sup>3</sup>Functional Materials, Institute of Materials Science, Technical University of Darmstadt, Alarich-Weiss-Straße 16, 64287 Darmstadt, Germany

<sup>4</sup>Aachen Institute for Advanced Study in Computational Engineering Science (AICES), RWTH Aachen University, Schinkelstr. 2, 52062 Aachen, Germany

<sup>5</sup>Department of Materials, Royal School of Mines, Imperial College London, SW7 2AZ London, UK

\*Corresponding author: Christoph Freysoldt, E-mail: [freysoldt@mpie.de](mailto:freysoldt@mpie.de)

## Abstract

Atom probe tomography (APT) is ideally suited to characterize and understand the interplay of segregation and microstructure in modern multi-component materials. Yet, the quantitative analysis typically relies on human expertise to define regions of interest. We introduce a computationally efficient, multi-stage machine learning strategy to identify compositionally distinct domains in a semi-automated way, and subsequently quantify their geometric and compositional characteristics. In our algorithmic pipeline, we first coarse-grain the APT data into voxels, collect the composition statistics, and decompose it via clustering in composition space. The composition classification then enables the real-space segmentation via a density-based clustering algorithm, thus revealing the microstructure at voxel resolution. Our approach is demonstrated for a Sm–(Co,Fe)–Zr–Cu alloy. The alloy exhibits two precipitate phases with a plate-like, but intertwined morphology. The primary segmentation is further refined to disentangle these geometrically complex precipitates into individual plate-like parts by an unsupervised approach based on principle component analysis, or a U-Net-based semantic segmentation trained on the former. Following the composition and geometric analysis, detailed composition distribution and segregation effects relative to the predominant plate-like geometry can be readily mapped from the point cloud, without resorting to the voxel compositions.

**Key words:** atom probe tomography, Fe-doped Sm–Co alloys, image segmentation, junction detection, machine learning

## Introduction

Atom probe tomography (APT) is a unique technique that provides the three-dimensional (3D) distribution of atoms in a material at sub-nanometer resolution along with their chemical identities (Gault et al., 2012a, 2021; Lefebvre et al., 2016; Miller & Forbes, 2014). APT provides insights into the composition of a variety of microstructural features discernible through their chemical (composition) fingerprint, including grain boundaries, dislocations, and secondary phase precipitates (Medrano et al., 2018; Zhou et al., 2021; Prithiv et al., 2022). The interplay between these microstructural features along with their chemistry, on the other hand, determines the material's macroscopic properties. For exploring structure-property relations and, ultimately, for tailoring new materials, the individual objects forming the microstructure must be quantified in abundance, size, composition, geometrical shape, etc.

Obtaining such a comprehensive quantitative description from APT datasets in a reliable, controlled, and reproducible way remains challenging. The size of the data sets has steadily

increased with a wider field-of-view (Kelly et al., 2004) and the implementation of laser pulsing capabilities (Gault et al., 2006; Cerezo et al., 2007). APT datasets now routinely contain 10–100 s of millions of ions. This limits the choice of the applicable algorithms. Quantifying the microstructural features including their composition, volume fraction, size, geometry, or spatial distribution of secondary phases, requires first some form of segmentation of the APT data (Hellman & Seidman, 2003). Indeed, segregation means that atoms of certain species are on average closer together, i.e., the local density of atoms of a specific species can be used to distinguish microstructural features (Vaumousse et al., 2003). Certainly, the density may be affected by aberrations in reconstruction due to the differences in phases' evaporation or chromatic aberrations (Marquis & Vurpillot, 2008). Therefore, one should be cautious in interpreting the results.

The most common approach in this direction is the use of iso-concentration surfaces, based on a marching cubes algorithm (Lorenson & Cline, 1987), that delineate microstructural features above a threshold composition compared to their

Received: March 27, 2023. Revised: June 6, 2023. Accepted: August 2, 2023

© The Author(s) 2023. Published by Oxford University Press on behalf of the Microscopy Society of America.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

surrounding, facilitating visualization. Iso-concentration surfaces, often called isosurfaces, provide quantitative information about the composition, and qualitatively define the geometry and distribution of the microstructural features. The caveat is that, as with most aspects of APT data processing, the information from isosurfaces may be sensitive to user-defined parameters. These parameters are often chosen ad hoc, e.g., to highlight specific characteristics of the data. This may lead to inconsistencies when a single dataset is analyzed by different scientists (Exertier et al., 2018; Barton et al., 2019; Dong et al., 2019).

More recently, various clustering algorithms, as a complementary approach to isosurfaces, have been proposed as a more controlled alternative, for instance, GMM (Gaussian mixture model) (Zelenty et al., 2017), DBSCAN (Density-Based Spatial Clustering of Applications with Noise) (Stephenson et al., 2007), HDBSCAN (Hierarchical DBSCAN) (Ghamarian & Marquis, 2019), and OPTICS (Wang et al., 2019). Clustering algorithms applied directly to the atomic coordinates of APT datasets tend to be computationally expensive due to the huge number of atoms. Therefore, the impact of the clustering algorithm's hyper-parameters (e.g., number of components or clusters in GMM) on the outcome of this process is rarely evaluated systematically.

Additionally, the morphology of microstructural regions of interest can be geometrically complex, another hurdle for human-guided quantification methods. In recent years, machine learning algorithms have been used more often in the analysis of APT data, with the aim of enhancing automation, reliability, and efficiency of the analysis process. Madireddy et al. presented a deep learning-based edge detection to identify the interface between a matrix and a precipitate phase (Madireddy et al., 2019). Although this technique offers a scalable (high-throughput) alternative to iso-concentration surfaces, it does not provide any information about the shape and geometry of the precipitate phase. Peng et al. quantified segregation at grain boundaries through a clustering algorithm applied on spatial coordinates of atoms (Peng et al., 2019), followed by calculating composition and thickness fluctuations. Zhou et al. introduced a more precise approach to identify junctions and straight segments in connected interface networks, notably grain boundaries, via deep neural-network image recognition in 2D projections (Zhou et al., 2022). The approach works best for columnar grains projected into their basal plane. An extension to interface networks with random orientation, which lack a common projection plane, is unfortunately not straightforward.

Here, we propose an approach based on unsupervised machine learning to semi-automatically extract “chemical domains,” i.e., regions with a characteristic composition that corresponds to different phases and segregation zones in the APT data. Contrary to performing clustering on the spatial coordinates of the atoms, we perform clustering in a space containing local composition information of the APT data, with the possibility to also provide the atom clusters in the point cloud within a few minutes on a single CPU core even for datasets containing 500 million atoms. We showcase this approach in the analysis of Fe-doped Sm–Co hard magnets, whose magnetic properties depend on their microstructure. The application of the developed workflows helps to identify and segment different phases into separate precipitates and further disentangle their geometry into plate-like structures enabling meaningful compositional and morphology quantification.

## Materials and Methods

### Experimental Data

The production-grade Fe-doped Sm–Co alloy that has been analyzed here is prepared by first milling and crushing the book mold ingots and subsequently mixing the prepared powders to achieve the desired composition (wt%)  $\text{Sm}_{2.5}\text{Zr}_3\text{Co}_{49}\text{Fe}_{19}\text{Cu}_5$ . This is followed by isostatic pressing and sintering of the powders. For further details see Sample synthesis in Duerrschabel et al. (2017).

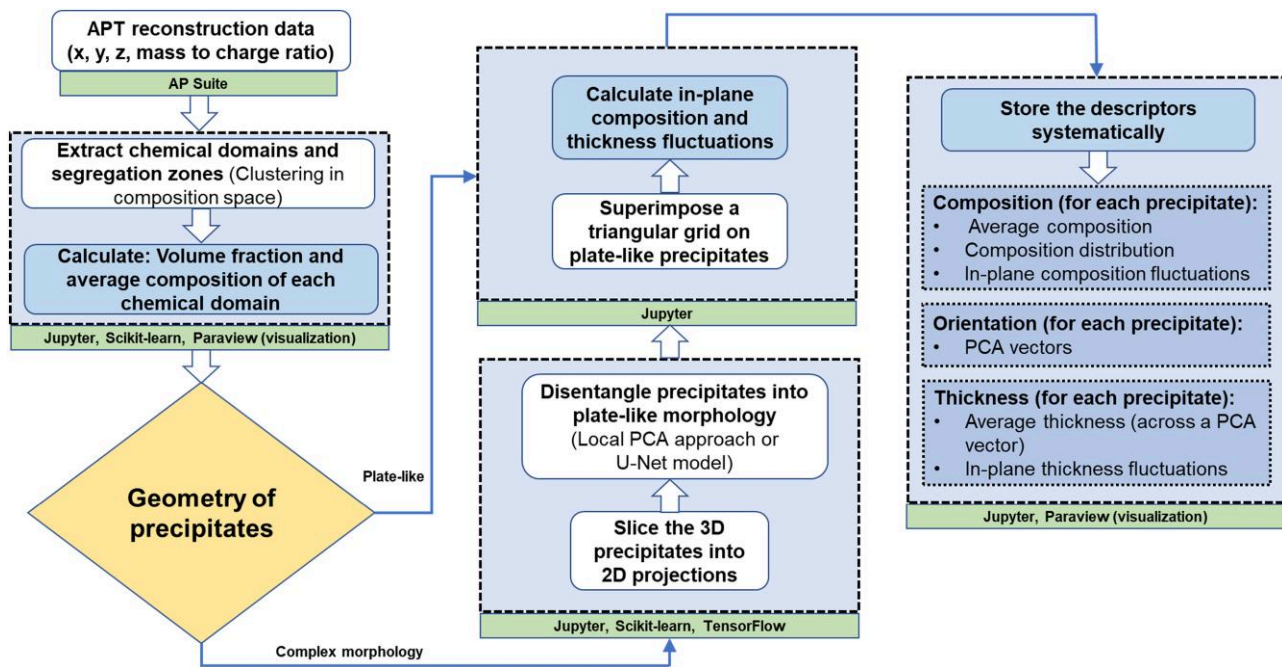
Specimens for APT were prepared by focused ion beam on a Dual-Beam Helios Nanolab 600i System using the approach outlined in Thompson et al. (2007). APT measurements were performed on a CAMECA local electrode atom probe (LEAP 5000 XS) in laser pulsing mode, with 10 ps laser pulses at a wavelength of 355 nm (UV), 45 pJ pulse energy at a repetition rate of 200 kHz, in ultra-high vacuum ( $1 \times 10^{-10}$  mbar). Further, the specimen base temperature was kept at 60 K with 6–10 ions detected per 100 pulses on average.

### Workflow

Figure 1 shows the current overarching workflow to classify different chemical domains in any APT data and to reduce morphologically-complex precipitates into simpler geometries. The spatial coordinates ( $x, y, z$ ), mass-to-charge state, and ranges defining the elemental or molecular nature of each ion are extracted from AP Suite, a proprietary software used to reconstruct and analyze APT data (POS and RRNG files) Day et al. (2019). Further, different chemical domains are extracted from the reconstructed data using clustering in composition space, which is discussed in detail in Section “Classification of chemical domains: clustering in composition space.” After a DBSCAN-based post-processing step, the morphology of the individual chemical domains is considered. If the APT reconstruction contains precipitates with a complex morphology they are disentangled into plate-like simpler structures using a local PCA approach or a U-Net model, which are described in Section “Decomposing geometrically complex precipitates.” The plate-like precipitates or structures can be further analyzed by calculating the in-plane composition and thickness fluctuations. Finally, descriptors corresponding to composition, orientation, and thickness for each phase's precipitates are stored systematically using the hdf5 file format (The HDF Group, 2022). All codes for reading and analyzing APT data are developed in Python, in a Jupyter environment, an open-source web-based tool used for code development and data visualization (Kluyver et al., 2016). ParaView, an open-source scientific visualization tool, is used for additional analysis and visualization (Ahrens et al., 2005). The code for the demonstrated workflow presented in this paper is accessible at [https://github.com/Alaukiksaxena/APT\\_Machine\\_learning](https://github.com/Alaukiksaxena/APT_Machine_learning).

### Classification of Chemical Domains: Clustering in Composition Space

First, to identify the chemical domains present in the given APT dataset, we introduce a composition-informed segmentation algorithm to visualize quickly and statistically understand clusters or precipitates and segregation zones in APT data. The algorithm is summarized in Figure 2. The APT data set is divided into equally-sized voxels (here:  $2 \times 2 \times 2 \text{ nm}^3$ ), without the application of any delocalization scheme (Hellman &



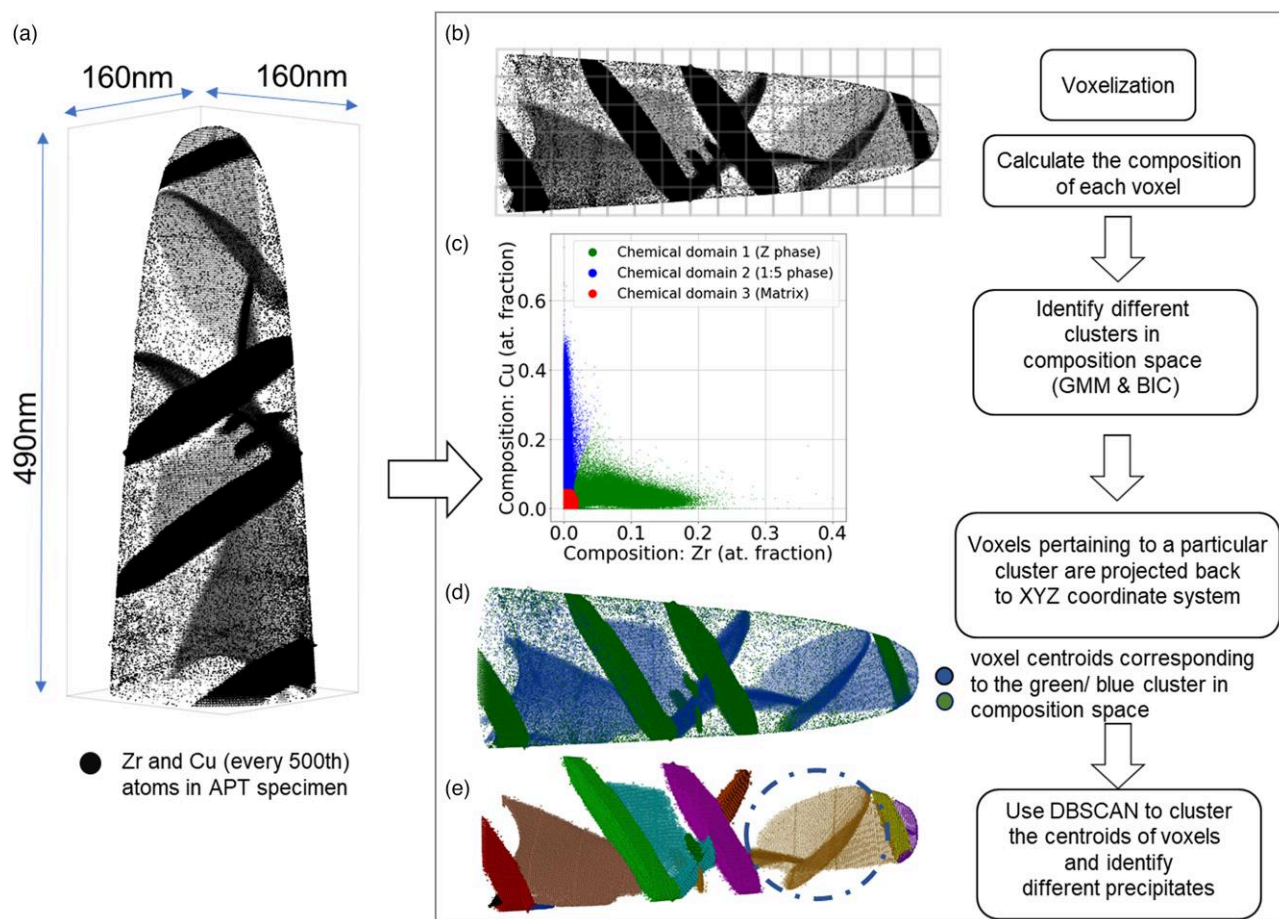
**Fig. 1.** The steps to quantify a 3D microstructure obtained from an APT reconstruction are summarized in the flow chart. The tools and libraries employed in each step are mentioned in the green boxes at the bottom.

Seidman, 2003). For each voxel, the composition, i.e., the relative fraction of all chemical species, is calculated as a vector, cf. Figure 2b. These vectors describe points in a multi-dimensional unit simplex which we call composition space whose dimensions depend on the number of relevant chemical species in the APT dataset. In this showcase, the APT dataset consists of five elements (Sm, Co, Fe, Zr, Cu), so the composition space is five-dimensional. The rationale behind the proposed method is that the voxels that are present inside a precipitate or cluster, of a particular phase or segregation zone, will have approximately the same composition and therefore appear as clusters in composition space. These “clusters in composition space” implies that many voxels exhibit similar composition; they must not be confused with spatial clusters of atoms or ions that form e.g., in segregation zones. The size of voxels is a compromise between spatial resolution (smaller voxels can map smaller features) and reducing statistical noise (larger voxels have less scatter in the compositions and better distinguish chemical domains of close compositions) (Torres et al., 2011). While we use cubes of side 2 nm in our case study, this should be adapted for other datasets, depending on the complexity of the underlying phases and length scale of phase separation. For optimization of the voxel size for a given APT dataset, it is recommended to follow the guidelines given in Mason & London (2020).

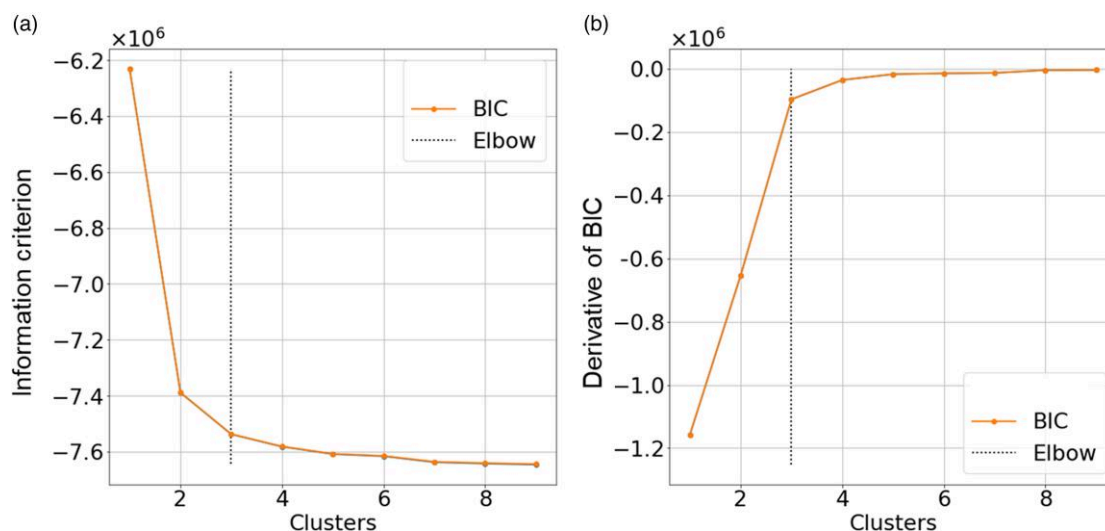
The second step is to identify different clusters in the composition space. For this, we use the GMM clustering algorithm (Hastie et al., 2009), Figure 2c. The rationale behind choosing the GMM algorithm is that deviation in composition arises from stochastic effects (e.g., due to intrinsic disorder in the material or the limited detection efficiency) and in the limit of large numbers the underlying distribution becomes normal. The GMM algorithm cannot automatically guess the number of clusters in a given dataset. This information has to be fed to the algorithm as a hyperparameter. The approximate number of clusters or chemical domains in the composition space can

be determined using the Bayesian information criterion (BIC) or can be inferred from, e.g., thermodynamic databases. In the current work, we have used minimization of the BIC to get the number of clusters as shown in Figure 3a, Geron (2019). For the given APT datasets, we get a visible kink or elbow (Zelenty et al., 2017), i.e., the number of clusters after which the decrease in the BIC value is not significant (threshold derivative of  $BIC = -0.1 \times 10^6$ ), at three clusters, which physically correspond to the 2:17 phase (matrix), Z-phase (rich in Zr) and 1:5 phase (rich in Cu). The number of clusters is also in agreement with the known phases of this material (Duerrschabel et al., 2017). The kink or change in slope corresponding to the derivative threshold  $= -0.1 \times 10^6$  can be seen in the central difference numerical derivative of the BIC as shown in Figure 3b. Figure 4b, shows a 3-D (Cu, Fe, Sm) slice of the 5-D composition space where each point corresponds to a voxel in the spatial coordinate system (refer to Supplementary Video 1). The three clusters or chemical domains found by the GMM algorithm are marked in Figure 4b. Further, normalized frequencies of voxel composition for each chemical species shown in Figures 4a, 4c, 4d, 4e, 4f, obtained using kernel density estimate (kernel-bandwidth = 0.1), highlight that no single element is sufficient to assign a voxel to a certain phase, but definitely their combination is required. The voxels pertaining to a particular cluster in composition space are translated back to the spatial coordinate system so that voxels present in the matrix, precipitates, and segregation zones can be separated, Figure 2d.

After 3D phase segmentation using clustering in composition space, DBSCAN is applied separately to the centroids of the voxels belonging to the same composition space cluster (Ester et al., 1996). Here, a voxel centroid is the arithmetic mean of the positions of the atoms present in the given voxel. In this process, the noisy voxels, which often lie at the boundary of the dataset, are removed and various precipitates pertaining to each phase are identified as shown in Figures 2d



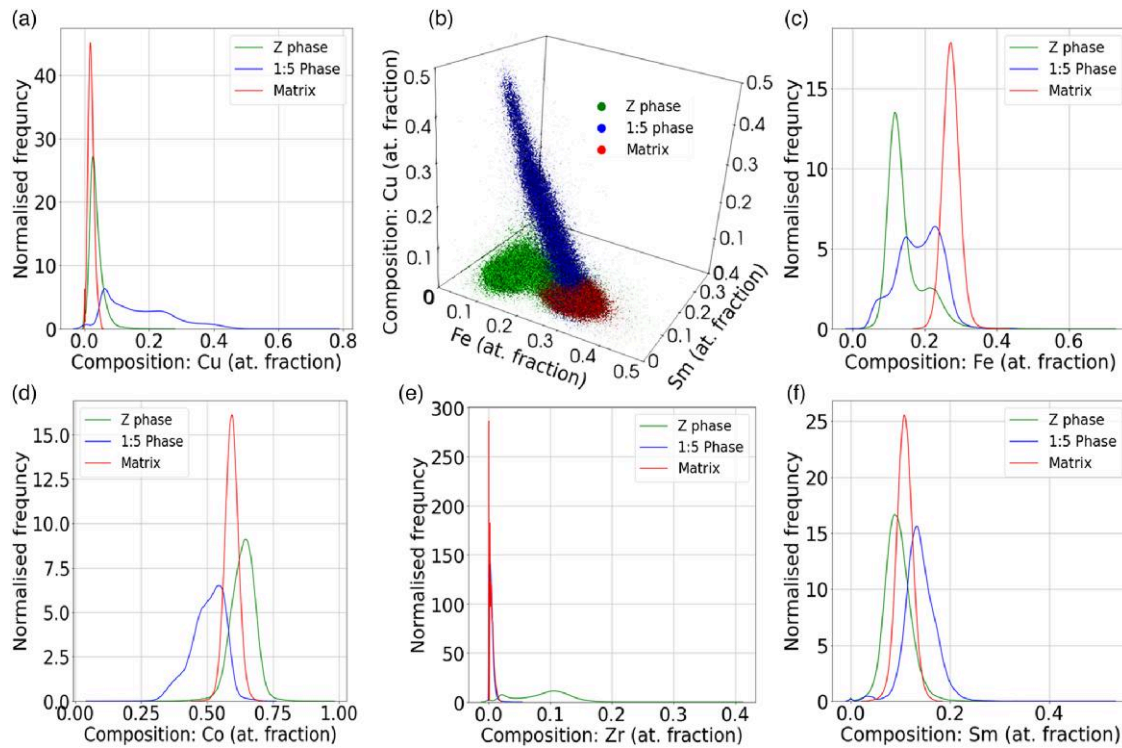
**Fig. 2.** (a) APT reconstruction for a Fe-doped Sm–Co alloy. For clarity, only minority elements (Cu, Zr) are shown. (b) Voxelization and per-voxel composition analysis, the gray grid schematically illustrates the voxels. (c) 2D projection of the 5D composition space. Each point corresponds to one voxel. Color coding according to 3 composition clusters identified. (d) Phases in real space at voxel resolution according to composition classification. (e) DBSCAN is used to cluster voxel centroids belonging to each phase to recognize separate individual precipitates. The circled precipitate has a complex morphology.



**Fig. 3.** (a) An elbow, shown with a dotted line, at clusters =3 is observed in the BIC curve. The number of clusters is in agreement with the known phases of the given material. (b) The central difference numerical derivative of BIC is invariant from clusters =4.

and 2e (refer to [Supplementary Video 2](#)). Since both the atoms residing in each voxel and the voxels present in each precipitate are registered, the last step enables us to study each

precipitate separately. Moving from an atom-by-atom representation of the data set to voxel centers is particularly advantageous for DBSCAN since the atomic density is subjected



**Fig. 4.** Normalized frequencies of voxel composition (at. fraction) of (a) Cu, (c) Fe, (d) Co, (e) Zr, and (f) Sm, respectively, for each phase present in one of the Sm–Co alloy APT dataset. (b) 3D projection (Cu–Fe–Sm composition) of the 5D composition space with clusters pertaining to respective phases.

to fluctuations due to the APT experiment [e.g., pole formation (Gault et al., 2012b) or in-homogeneous evaporation due to evaporation–field differences (Vurpillot et al., 2000)], while the voxel density is uniform. The voxel centers, by construction, are equally spaced and provide a practically constant density of points within each precipitate. Since now the number of voxels that are present in each precipitate is known, statistical descriptors like mean volume, number density, and nearest neighbor distribution of the precipitates can be readily obtained. For our analysis, we have used scikit-learn’s implementation of DBSCAN (parameters  $\text{eps} = 9$  and  $\text{min\_samples} = 80$  are used for all the identified phases) (Pedregosa et al., 2011).

### Decomposing Geometrically Complex Precipitates

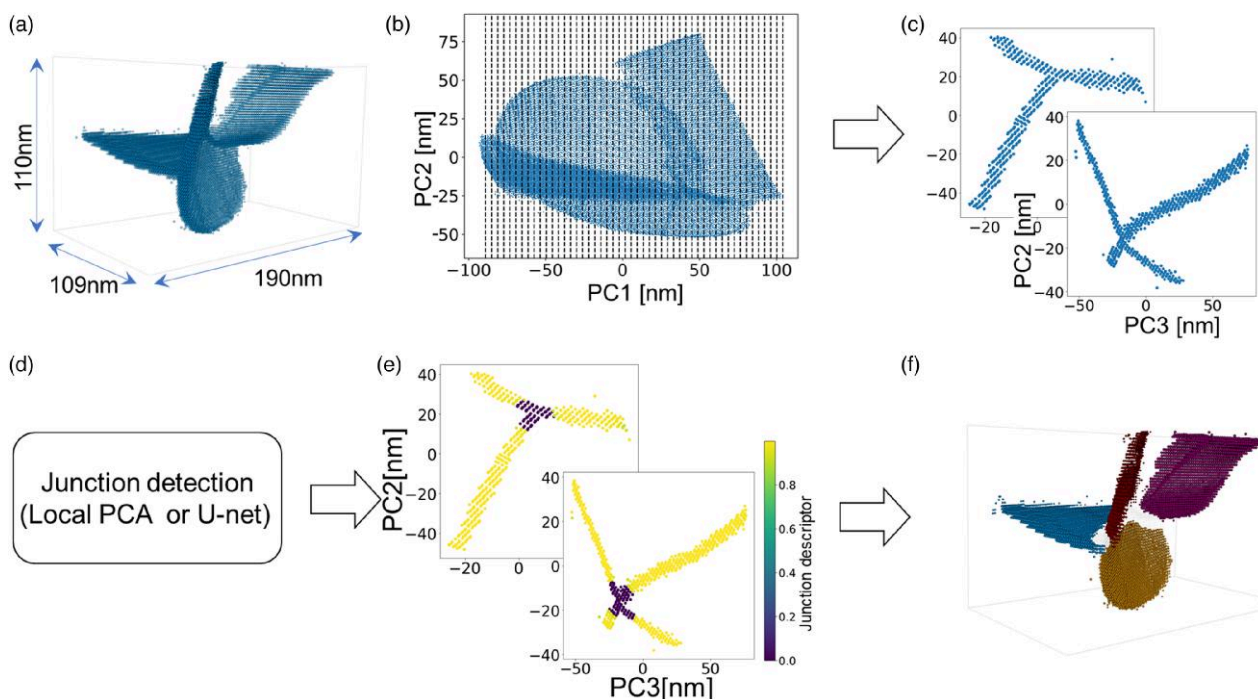
#### Local principal component analysis (PCA)-based method

In APT datasets, especially while studying nano-grain structures, we often encounter quasi-planar interfaces or inter-phase boundaries that cross each other forming triple or quadruple junctions. This geometric complexity makes the quantification of such microstructures cumbersome and challenging. In the Sm–Co alloy under our investigation, the Cu-rich 1:5 phase has precipitates that do not cross the entire field of view, making the approach of Zhou et al. (2022) discussed in Section “Introduction” impossible to use. To separate complex precipitates into plate-like substructures visible to the human eye, we propose a local PCA-based method (Jolliffe, 2002). This method is applied to all identified precipitates to find plate-like structures automatically.

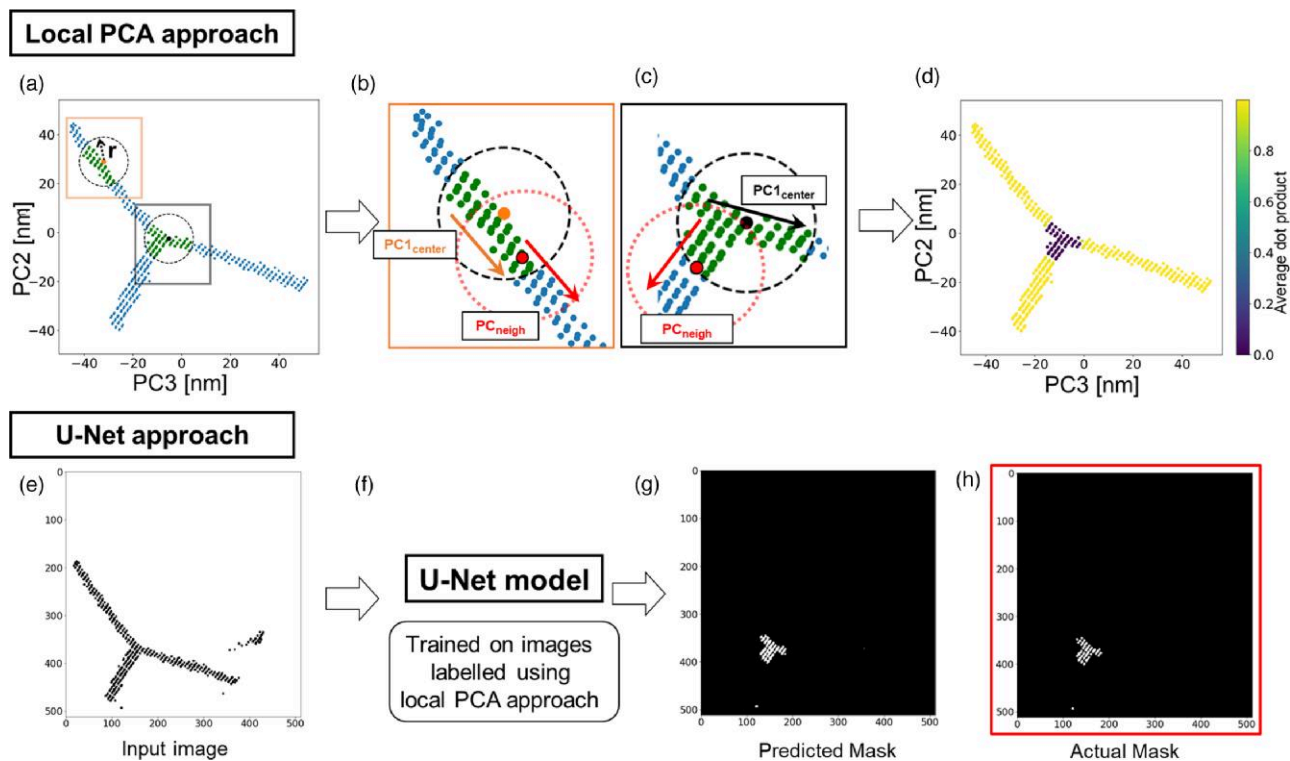
This method is demonstrated on a 1:5 phase precipitate, as shown in Figure 5a, with an entangled morphology. First, PCA is performed on the spatial coordinates of voxel centroids belonging to the chosen DBSCAN cluster (or precipitate), and all the voxel centroids are transformed to the PCA coordinate

system (PC1, PC2, PC3), Figure 5b. For visualization a 2D projection of the 3D precipitate in the PC3 direction is shown. Next, the precipitate is divided into  $N_{\text{cut}}$  slices, such that each slice is 4 nm thick, shown by the dotted lines in the figure, along the PC1 direction. Each slice is projected in the PC1 direction, i.e., to the PC2–PC3 plane, Figure 5b. The slicing process is repeated in the PC2 and PC3 directions to generate 2D projections in the respective perpendicular planes. The created 2D projections are shown in Figure 5c and will be used to identify the junctions connecting the ribbon-like features in the projection plane.

For this, the local PCA approach is applied to each of the 2D projections or slices. This method helps in the classification of voxel centroids (data points in the 2D projection) to be in junction-like or non-junction-like environments, Figures 5d to 5f. The local PCA algorithm is summarized in Figures 6a to 6d. As shown in Figure 6a, for each data point in a 2D projection (e.g., the orange point), we find all neighborhood points lying within a radius  $r$  (green). Subsequently, from the PCA of the distance vectors, the direction of maximum variance is identified,  $\text{PC1}_{\text{center}}$ , Figure 6b. The radius  $r$  is provided as a hyperparameter for each precipitate. The  $\text{PC1}_{\text{center}}$  direction is then compared to the corresponding one,  $\text{PC1}_{\text{neigh}_i}$ , for all the data points in the neighborhood (green) by computing the dot product ( $\text{PC1}_{\text{center}} \cdot \text{PC1}_{\text{neigh}_i}$ , where  $i$  is a given neighborhood data point). If the selected data point (orange) lies in a quasi-plate-like structure then most of the  $\text{PC1}_{\text{neigh}}$  directions will point in the same direction as  $\text{PC1}_{\text{center}}$  and the average of their dot products will be close to one (threshold for not junction: average dot product  $> 0.8$ ). However, if the orange data point lies near the junction then most of the  $\text{PC1}_{\text{neigh}}$  directions will not point along  $\text{PC1}_{\text{center}}$  and the average of their dot product is lower (threshold for junction: average dot product  $\leq 0.8$ ), Figure 6c. In this way, the local PCA approach,



**Fig. 5.** Workflow for precipitate disentanglement via junction detection and removal. (a) Voxel centroids corresponding to a 1:5 phase precipitate having four plate-like structures meeting at a junction, also shown in Figure 2e. (b) The voxel centroids are transformed to the PCA coordinate system and divided into  $N_{cut}$  slices along PC1 (direction of maximum data variance). For visualization a 2D projection of the 3D precipitate in the PC3 direction is shown. (c) 2D projections for slices of voxel centroids in the plane perpendicular to the cutting direction. (d, e) Local PCA approach or U-Net model trained on similar data is used to detect junctions in the 2D projections. (f) All the voxel centroids that are labeled as plate-like back-projected to 3D space where they are clustered using DBSCAN.



**Fig. 6.** Junction detection. (a) A 2D projection of the voxel centroids in a slice extracted from a 3D precipitate. The orange and the black box highlight a plate-like and junction region, respectively. The neighborhood (green) within a radius  $r$  around a chosen point (orange) is selected within the orange box. (b) Magnified region away from the junction. The direction of maximum variance,  $PC1_{center}$  and  $PC1_{neigh}$  corresponding to the neighborhood of the orange and red points within radius  $r$ , respectively. The red point is chosen from the neighborhood of the orange point. (c) Magnified region near the junction. (d) Average dot product mapped on the 2D projection delineating the junction region. (e-h) A U-Net model trained on local PCA approach labeled images is applied on a previously unseen input image to predict the junction region or mask. The actual or true mask is also shown.

which is applied to all the data points in a 2-D slice, is able to classify the data points that lie at the junction of the precipitates, **Figure 6d**. All the junction data points from each 2-D slice are stored in a list and tracked back to the original points (or voxel centroids) in the 3-D spatial coordinate system. Finally, the junction centroids in 3D space are removed from the set of all the voxel centroids pertaining to a particular precipitate, separating the quasi-plate-like structures. These structures can now be identified as separate clusters by applying DBSCAN since the connecting junction between them has been removed, **Figure 5f**.

### U-Net based approach

In the local PCA-based approach, the accuracy of the classification of data points as junction depends on the hyperparameter  $r$ , which has to be fed for each precipitate, shown in **Figure 6a**. For example, if the radius  $r$  is larger than the precipitate the local PCA approach will not be able to find the data points at the junction. To circumvent this problem, we have developed a more robust approach based on supervised machine learning (U-Net) and image augmentation. A U-Net is a convolutional neural network (CNN) based architecture, which is used in computer vision problems, especially biomedical image data, for semantic segmentation (Ronneberger et al., 2015). Semantic segmentation means that the deep learning algorithm predicts the class for each pixel in a given image.

First, the local PCA approach is applied to 1:5 phase precipitates extracted from a given APT dataset. This process created 500 2D projections or slices. The data points in the projections are already labeled as junction or not junction using the local PCA approach. All the projections are converted to gray-scale images of shape  $512 \times 512 \times 1$  and based on the label information a mask depicting the junction in each image is also

extracted with shape  $512 \times 512 \times 2$ , where the last 2 channels correspond to the classes: junction and not junction.

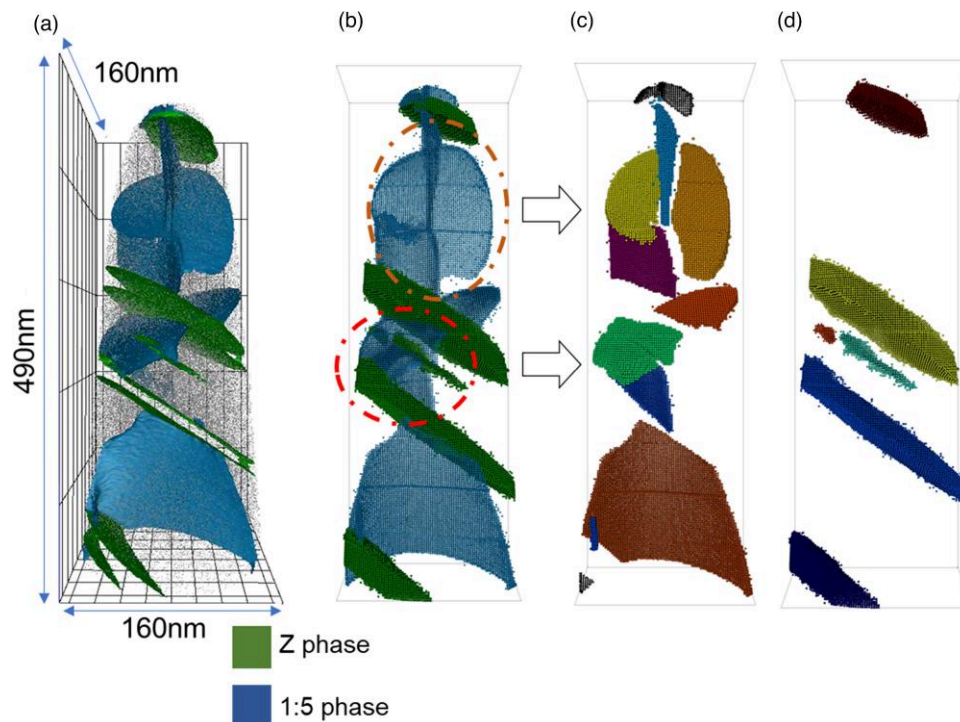
To increase the diversity of the training data, image augmentation steps like rotation, translation, flipping and resizing are applied to the original images to create a pool of 3000 training images and their corresponding masks. These training images ( $512 \times 512 \times 1$ ) and the masks ( $512 \times 512 \times 2$ ) are used to train a U-Net. The trained model, as shown in **Figures 6e to 6g**, takes an image corresponding to a projection as input ( $512 \times 512 \times 1$ ) and predicts a mask of shape  $512 \times 512 \times 2$ , where the last channel contains the probability of a pixel to be a junction. Based on the predicted mask, the data points in the original projection/slices are classified as junction or not.

The U-Net based model is more robust and is directly applied to the other samples of the same material without the need to tune the parameter  $r$  for each precipitate. The local PCA-based approach is semi-automatic and can be tuned according to the precipitate size. However, the U-Net-based approach is fully automatic irrespective of the precipitate size. Nevertheless, retraining of the U-Net model might be required when applying the same model to other material datasets, in which the geometry of precipitates is different. Further details of the U-Net architecture used in this study are given in **Supplementary Figure 14**.

## Results and Discussion

### Visual Validation with Isosurface Results

The proposed approach of clustering in composition space has been applied to all the given Fe-doped Sm–Co alloy APT datasets. **Figure 7** provides a visual comparison between the isosurfaces obtained using AP suite and the results of our new



**Fig. 7.** (a) 6 at% Zr (green: Z-phase) and 12 at% Cu (blue: 1:5 phase) isosurfaces in a Fe-doped Sm–Co alloy APT dataset extracted using AP suite. (b) The voxel centroids lying in the Z-phase (green) and 1:5 phase (blue) precipitates extracted using clustering in composition space. The 1:5 phase precipitates with complex morphology are highlighted. (c) Segmented 1:5 precipitates. Complex morphology precipitates are separated into plate-like structures. (d) Segmented precipitates of the Z-phase.



algorithm. **Figure 7a**, shows the 6 at% Zr (Z-phase) and 12 at% Cu (1:5 phase) isosurfaces, and **Figure 7b** shows the voxel centroids that lie in Z-phase and 1:5 phase. These phases have been identified by the composition space clustering, while the final assignment of the voxels as part of a spatial cluster additionally relies on the DBSCAN algorithm on the voxel centroids of each phase to eliminate noise. The phases obtained in **Figure 7b** are visually identical to the isosurfaces for the respective phases shown in **Figure 7a**.

For clearer visualization of the precipitates, we separated them in subfigures (c) and (d): **Figure 7c** depicts only the 1:5 phase precipitates, and **Figure 7d** the ones for the Z-phase. The precipitates of 1:5 phase with complex morphologies containing bi and quad junctions are highlighted in **Figure 7b**. The junction voxel centroids have been identified using the U-Net approach and removed (Section “Decomposing geometrically complex precipitates”). The remaining voxel centroids are clustered by employing DBSCAN to obtain plate-like precipitates as shown in **Figure 7c** (refer to **Supplementary Videos 2 and 3**). A quantitative comparison between isosurfaces from Paraprobe (Kühbach et al., 2022) and the developed composition space approach is discussed in the **Supplementary Section “Quantitative comparison with isosurfaces”**.

### Calculation of Volume Fraction and Composition of Each Phase

The segmentation of each phase or chemical domain allows for the calculation of the approximate volume fraction, as the number of voxels included in a phase is known. The volume fraction of each phase in one of the APT datasets is summarized in **Table 1**. The average composition of a phase is the mean of the compositions of the precipitates or domains it comprises. Average compositions and standard error (confidence interval = 95%) for each phase in the APT sample are summarized in, **Table 1**. They are calculated by averaging over the subsets, encompassing 15,000 adjacent voxels each. The dependence of the standard error on subset size is shown in **Supplementary Figure 18**.

### Composition fluctuations in precipitates

The comparison between the voxel composition distribution of each chemical species in a particular precipitate and the corresponding random distribution are plotted to understand if any clusters or segregation zones are present inside the chosen precipitate. Such a comparison for randomly selected precipitates of the 1:5 and Z-phase and a cell of 2:17 matrix phase for Cu, Fe, and Zr is shown in **Figure 8**. As seen in the figure, the experimental distribution deviates from the random distribution for the Z and 1:5 phases as compared to the 2:17 phase. This behavior is seen across all APT data sets of the current material. In order to understand and quantify these deviations, all the precipitates are reduced to simpler plate-like structures enabling us to study in-plane composition fluctuations. The

computation of volume fraction, the composition of each phase, and composition fluctuations in precipitates can be executed in just a few seconds using a single CPU core for the given APT dataset.

### In-plane Composition and Thickness Fluctuations

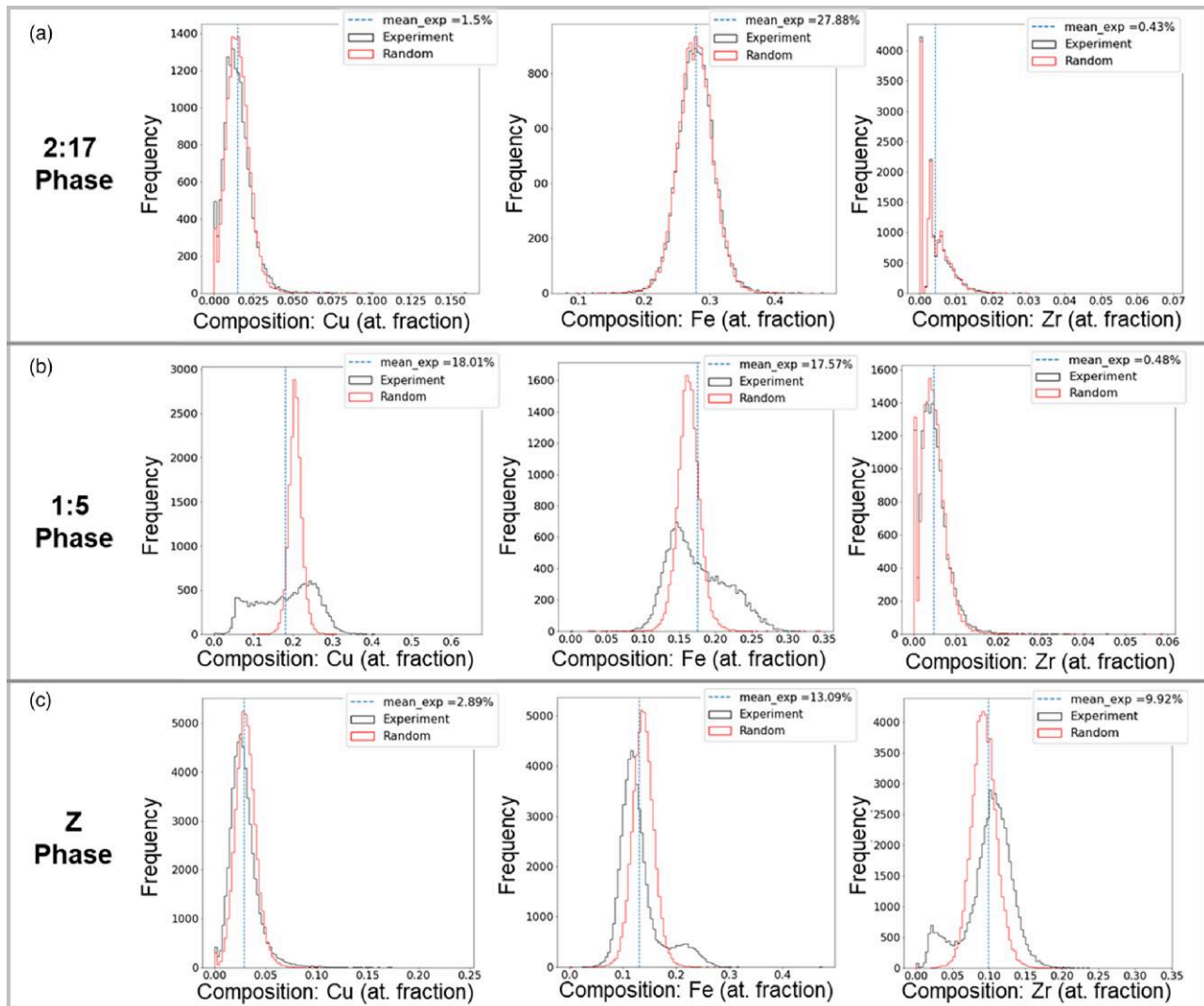
We modified the algorithm of Peng et al. to find the in-plane composition and thickness fluctuations in the plate-like precipitates (Peng et al., 2019). The first step is to apply a PCA transformation on the voxel centroids of a given quasi-planar precipitate. Next, a regular 2-D triangular grid of size 4 nm is superimposed on the precipitate in the plane formed by the two maximum variance PCA directions. The grid triangles not containing voxel centroids are removed. This is followed by moving each node of the trimmed grid to the center of mass of the voxel centroids located in the cuboidal region overlaid at that node, shown in **Supplementary Figure 15a**. We use voxel centroids of the precipitates for analysis while the original algorithm directly uses atoms at the interface to be analyzed. Since some of the precipitates in our case studies have a million atoms or more, the choice of voxel centroids makes it faster. As a result of moving the nodes, the trimmed grid takes the shape of the precipitate, **Supplementary Figures 15b and 15c**. The geometry of the precipitate is still preserved if we use voxel centroids for our analysis.

After obtaining the grid that conforms with the shape of the precipitate, the local surface orientation (normal vector) at each node is defined from the average of the adjacent triangles. This is followed by selecting atoms in a cylindrical region of interest (ROI) at each node with the axis of the cylinder along the node’s normal. To speed up the atom search, we exploit the space partitioning provided by the voxels. First, relevant precipitate voxels are identified, whose centroids fall inside a cylindrical ROI with a radius,  $r_{out}$  (= 4 nm), and length  $L_{out}$  (= 34 nm). From the selected voxels, the atoms within a cylindrical ROI of dimensions,  $r_{in}$  (= 2 nm), and length  $L_{in}$  (= 30 nm), are chosen for the calculation of in-plane thickness and composition fluctuations.

Each cylindrical ROI is divided into bins along the cylinder axis and the composition of each bin is calculated. We arbitrarily selected a bin width of  $L_{bin}$  (= 0.6 nm) as it provided a good compromise of limited statistical fluctuation and details of the feature of interest. The 1D composition profile for all the elements, obtained at a certain node in a precipitate of the Z-phase along the normal shown in **Figure 9a**, is plotted in **Figure 9b**. To analyze the clustering or segregation of the chemical species across the quasi-planar precipitates, we study the 1D composition profiles in more detail. To exemplify our approach, **Figure 9c** presents a 1D composition profile of Cu across the aforementioned node in the Z-phase precipitate. A six-degree polynomial is fitted on the 1D composition profile. Subsequently, the number and location of local maxima in the fitted polynomial are obtained using `find_peaks`, a tool

**Table 1.** Volume Fraction, Average Thickness and Average Composition of 2:17, Z, 1:5 Phase in One of the Given APT Datasets.

Phase	Volume Fraction	Thickness (nm)	Composition at% (mean $\pm$ 1.96 standard error)				
			Sm	Co	Fe	Cu	Zr
2:17	0.8		10.9 $\pm$ 0.4	59.4 $\pm$ 1.0	27.2 $\pm$ 0.8	2.0 $\pm$ 0.8	0.5 $\pm$ 0.2
1:5	0.06	17.6 $\pm$ 4.0	14.4 $\pm$ 1.6	48.0 $\pm$ 5.3	16.1 $\pm$ 5.1	21.1 $\pm$ 8.8	0.4 $\pm$ 0.2
Z	0.14	10.9 $\pm$ 1.4	9.7 $\pm$ 1.4	63.4 $\pm$ 2.5	14.5 $\pm$ 1.4	3.8 $\pm$ 1.4	8.5 $\pm$ 1.74



**Fig. 8.** Voxel composition distributions of Cu, Fe and Zr present in (a) a 2:17 matrix cell as well as in 2 precipitates chosen for (b) 1:5 phase and (c) Z-phase. The experimental composition distribution (black) deviates significantly from the random distribution (red) for 1:5 and Z precipitates.

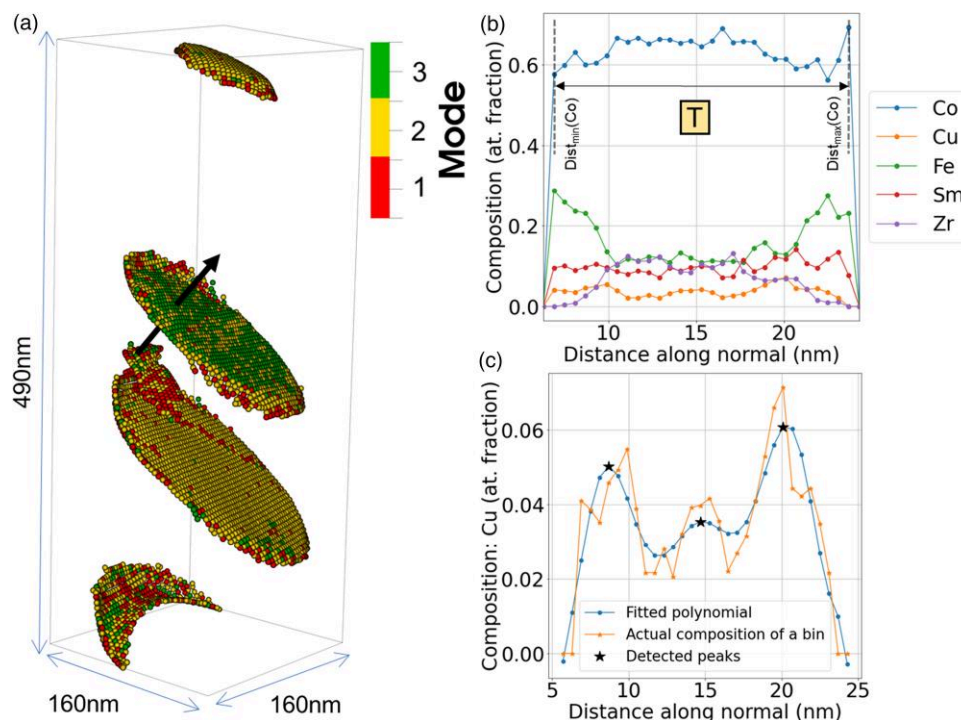
implemented in open-source scipy library (Virtanen et al., 2020). Figure 9a shows the fluctuations in the number of modes (or peaks) in 1D Cu composition profiles of the Z-phase precipitates in one of the APT data sets. The precipitates exhibit 1, 2, or 3 modes in their composition profiles, suggesting that Cu is either uniformly distributed, segregated at the edges, or additionally segregated at the middle of the precipitates, respectively.

Based on the 1D composition profiles at each node, it is possible to reveal possible in-plane composition fluctuations, as shown in Figure 10 for Zr (A and B) and Cu (C and D) for precipitates of Z (A and C) and 1:5 phase (B and D). Since the composition of Co is significantly higher than any other element in both Z and 1:5 phases (refer Table 1) and also its composition is higher in the Z-phase and lower in the 1:5 phase compared to the matrix (2:17 phase), it is the best choice to mark the start and end of a precipitate. The in-plane fluctuations are calculated by taking the average of an element's composition over the length of the ROI where Co composition is not zero. As shown in Figure 9b, the average is calculated over the bins lying between  $\text{Dist}_{\max}(\text{Co})$  and  $\text{Dist}_{\min}(\text{Co})$ . This line of reasoning is also used to calculate the thickness

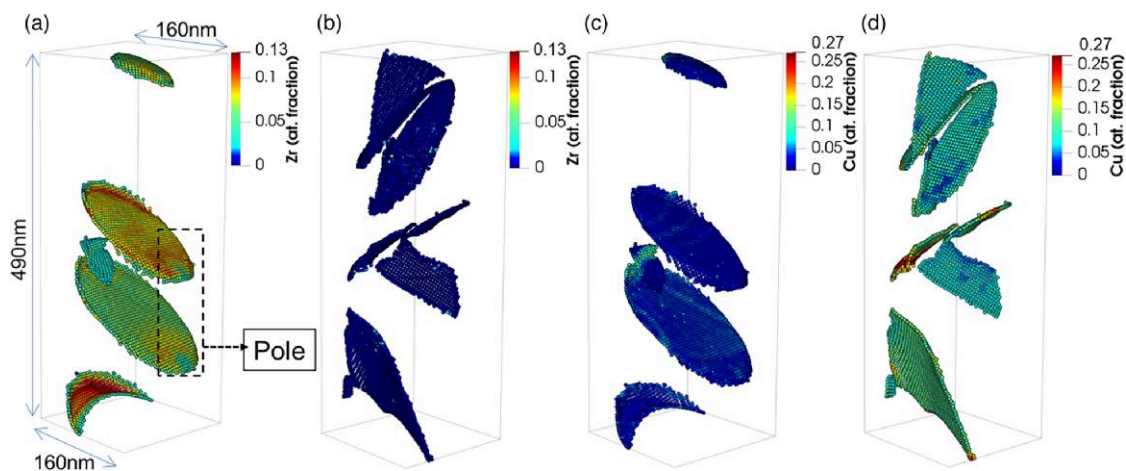
of each precipitate at every node of the grid. Thickness,  $T = \text{Dist}_{\max}(\text{Co}) - \text{Dist}_{\min}(\text{Co})$ , of the Z precipitate at the node, is shown in Figure 9b. Thickness fluctuations for precipitates of Z and 1:5 phase are summarized in Supplementary Figures 16b and 16c. The computation of in-plane composition and thickness fluctuations in the precipitates can take up to 30 minutes on a single CPU core, for a fine grid size (4nm), for the given APT dataset.

### Quantification of the Microstructure

To quantify the Sm–Co alloy microstructure, in addition to the thickness, orientation, and spacing between the extracted precipitates have to be calculated. The parameters like the distance between parallel precipitates and the angle between the intersecting ones are used to define a 2D model microstructure of a Sm–Co alloy (Katter et al., 1996). Similarly, we have extended this concept to quantify 3D microstructures. To define the orientation of each plate-like precipitate, PCA is carried out for the spatial coordinates of voxel centroids present in the given precipitate. The direction of minimum variance in the spatial coordinates is then taken as an



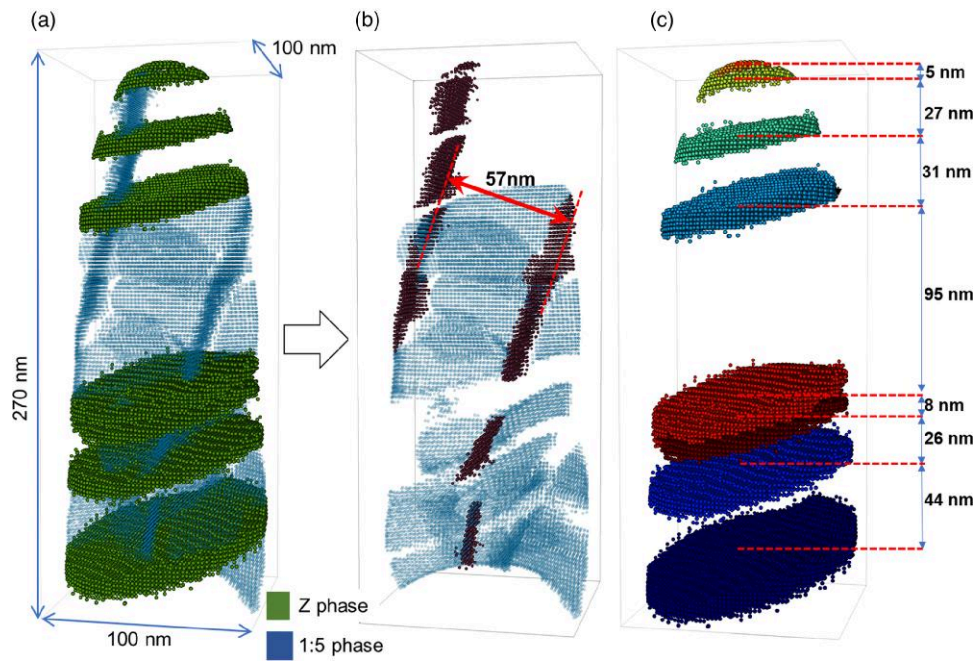
**Fig. 9.** (a) The nodes in a triangular grid superimposed on the Z-phase precipitates. The color of each node pertains to the number of modes or peaks in the 1D composition profile of Cu at that particular node. (b) 1D composition profile of Co, Cu, Fe, Sm, and Zr along a cylindrical ROI around the black arrow (normal at a node), crossing through the precipitate shown in panel (a). Here, only the atoms within the precipitate are considered to plot the 1D composition profiles. Also, the composition profiles while considering all atoms in the vicinity of the precipitate as well are shown in [Supplementary Figure 17](#). The thickness (T) of the precipitate along the ROI is estimated by calculating the maximum distance between the bins that have non-zero Co composition. (c) 1D composition (at. fraction) profile of Cu along the same ROI. A six-degree polynomial is fitted on the 1D composition profile. Local maxima or peaks are estimated in the fitted polynomial to approximate the local segregation zones of Cu across the precipitate.



**Fig. 10.** In-plane composition fluctuations: The nodes associated with 2D grids fitted on plate-like precipitates in one of the APT samples. The color coding corresponds to the average composition (at. fraction) of Zr (a,b) and Cu (c,d) at each node in Z phase (a,c) and 1:5 phase (b,d). Further, in panel (a), Zr is depleted in a circular region in two precipitates, marked with a rectangular box. This can be a pole in the dataset, which is an APT artifact.

approximate orientation normal to the precipitate plane. [Figure 11a](#) shows voxel centroids corresponding to Z and 1:5 phase in one of the Sm–Co alloy APT datasets. The 1:5 phase precipitates with complex morphology in the given microstructure have been reduced to plate-like structures. The angle between each pair of planar precipitates is calculated and a set of parallel precipitates (the angle between the normal vectors is in range  $0^\circ$  to  $5^\circ$ ) forming the magnetic domain walls is shown in [Figure 11b](#). Similarly, a set of parallel

Z-phase precipitates in the same APT sample is shown in [Figure 11c](#). The distance between a pair of parallel precipitates is the distance of the centroid of one of the precipitates from the plane of the other precipitate. The distances between the consecutive Z-phase plates, and a set of parallel 1:5 phase precipitates are summarized in [Figures 11b](#) and [11c](#) (refer to [Supplementary Video 4](#)). This computational analysis can be executed in just a few seconds using a single CPU core for the given APT dataset.



**Fig. 11.** (a) Voxel centroids of Z and 1:5 phase, green and blue respectively, present in one of the APT samples with complex microstructure. (b,c) Sets of parallel 1:5 and Z-phase precipitates, respectively, with distances between them.

Such quantification of the microstructure from APT datasets is directly dependent on the way APT data is reconstructed. In the present study, the APT data has been reconstructed in AP Suite Software 6.1 using voltage reconstruction. The image compression factor (ICF) was set to 1.65 and the radius was adjusted such that the Z-phase precipitates have the form of straight platelets and that these platelets are aligned parallel to each other in the reconstruction as observed in transmission electron microscope (TEM) measurements. The distances between the platelets observed in APT are compatible with those in TEM measurements on the same sample (Giron et al., 2023).

Our quantitative analysis of composition and geometric aspects of the chemical domains identified by the 3D segmentation algorithm illustrates opportunities that come with automatic identifications of microstructural features. It is very easy to apply new APT data analysis methods as shown in this section, or adapt already existing ones like the construction of 1D composition profiles or composition fluctuation maps. Thanks to the controlled workflow, this can be done not only for selected regions of interest but automatically for all features identified. This allows for a more comprehensive description of the APT data than what was possible so far.

## Summary and Conclusion

In this paper, we developed a python based ML workflow to compositionally and geometrically quantify the 3D microstructure from APT data. The first part of the workflow extracts existing chemical domains or segregation zones. The key steps are (1) voxelisation of the APT data set, (2) calculating the composition of each voxel, (3) clustering in the composition space to obtain the chemical domains in the dataset and (4) separation of voxels of the extended regions of each chemical domain by DBSCAN clustering algorithm. The relevant data of all steps is stored in hdf5 files to aid subsequent

postprocessing. This enables a robust and systematic analysis of the microstructure from its chemical fingerprint, yielding for instance the average volume fraction, domain size distribution, and average composition of the identified chemical domains (which could be stable phases from the thermochemical phase diagram, or segregation zones with a unique composition near grain boundaries or other crystallographic defects).

Geometrically quantifying the microstructure requires to reduce morphologically entangled phases to individual structures. This was achieved here for a particular type, namely agglomerates exhibiting plate-like substructures. Our novel local PCA and/or U-Net approaches allow us to distinguish junction voxels from those lying in the planar regions, followed by a DBSCAN clustering to get the separated planar structures. This procedure can efficiently and rigorously segment out predominantly flat subdomains with arbitrary orientations in 3D. Indeed, planar or plate-like morphologies are common microstructural features, for example, segregation at the interface (grain boundary) and inter-phase boundaries or phases formed in some Al (Khushaim et al., 2015; Zandbergen et al., 2015) and Ni alloys (Vogel et al., 2015). Then descriptors such as PCA vectors, their explained variance, average composition, centroid, or the angle of the inclination with the axis of the APT tip, are calculated for each planar substructure. Using the geometrical information, the compositional analysis can be further refined to yield in-plane composition and thickness fluctuations. A method to map the segregation of solute atoms across interfaces is also developed.

Using the developed workflow, we identified three phases, matrix (2:17 phase), Z-phase and 1:5 phase, for Fe-doped Sm-Co alloy APT datasets. Notably, this computational analysis can be executed in just a few minutes using a single CPU core, even for APT datasets with 500 million atoms. We used the workflow to disentangle each precipitate, allowing a fast and robust analysis. We were able to identify the deviation

of the experimental distribution of Zr, Fe and Cu in precipitates of Z and 1:5 phase from their corresponding random distributions. The across-plane composition fluctuation of Cu was very distinct with enrichment at the edges or also at the center of the Z-phase precipitates. Such compositional fluctuations together with geometrical quantification of the microstructure are very relevant for the magnetic performance of this and other alloys.

The developed workflow can readily be applied to a wide variety of microstructures analyzed by APT. This enables a consistent quantification of compositional and structural features in an APT dataset.

## Code Availability

The workflow is available at [https://github.com/Alaukiksaxena/APT\\_Machine\\_learning](https://github.com/Alaukiksaxena/APT_Machine_learning).

## Supplementary Material

To view supplementary material for this article, please visit <https://doi.org/10.1093/micmic/ozad086>.

## Acknowledgments

A.S. appreciates funding by Helmholtz School for Data Science in Life, Earth and Energy (HDS-LEE). NP is grateful for funding from the IMPRS SURMAT. S.K., N.P. and B.G. acknowledge support from the Deutsche Forschungsgemeinschaft (DFG) for funding from the Leibniz Prize 2020 (GA 2450/2-1). O.G., L.M.L. and B.G. are grateful to DFG for funding of the TRR 270 HoMMage (INST 163/578-1). N.K., S.K., B.G., J.N. and C.F. are grateful for financial support from BiGmax, the Max Planck Society's Research Network on Big-Data-Driven Materials Science.

## Financial Support

The current study hasn't received any fund from any organizations or institutions.

## Conflict of Interest

The authors declare that they have no competing interest.

## References

Ahrens JP, Geveci B & Law CC (2005). 36 – Paraview: An end-user tool for large-data visualization. In *Visualization Handbook*, Hansen CD & Johnson CR (Eds.), pp. 717–731. Burlington: Butterworth-Heinemann. Available at <https://www.sciencedirect.com/science/article/pii/B9780123875822500381>.

Barton DJ, Hornbuckle BC, Darling KA & Thompson GB (2019). The influence of isoconcentration surface selection in quantitative outputs from proximity histograms. *Microsc Microanal* 25, 401–409. <https://doi.org/10.1017/S143192761900014X>

Cerezo A, Clifton PH, Gumberg A & Smith GDW (2007). Aspects of the performance of a femtosecond laser-pulsed 3-dimensional atom probe. *Ultramicroscopy* 107, 720–725. <https://doi.org/10.1016/j.ultramic.2007.02.025>

Day A, Francois-Saint-Cyr H, Geiser B, Payne T, Oltman E, Ringer S & Reinhard D (2019). Recent developments in apt analysis automation and support for user-defined custom analysis procedures in IVAS 4. *Microsc Microanal* 25, 338–339. <https://doi.org/10.1017/S1431927619002423>

Dong Y, Etienne A, Frolov A, Fedotova S, Fujii K, Fukuya K, Hatzoglou C, Kuleshova E, Lindgren K, London A, Lopez A, Lozano-Perez S, Miyahara Y, Nagai Y, Nishida K, Radiguet B, Schreiber DK, Soneda N, Thuvander M, Toyama T, Wang J, Sefta F, Chou P & Marquis EA (2019). Atom probe tomography interlaboratory study on clustering analysis in experimental data using the maximum separation distance approach. *Microsc Microanal* 25, 356–366. <https://doi.org/10.1017/S1431927618015581>

Duerrschabel M, Yi M, Uestuener K, Liesegang M, Katter M, Kleebe HJ, Xu B, Gutfleisch O & Molina-Luna L (2017). Atomic structure and domain wall pinning in samarium-cobalt-based permanent magnets. *Nat Commun* 8, 1–7. <https://doi.org/10.1038/s41467-017-00059-9>

Ester M, Kriegel HP, Sander J & Xu X (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In *Knowledge Discovery and Data Mining*, KDD'96, pp. 226–231. AAAI Press. Available at <https://www.osti.gov/biblio/421283>.

Exertier F, La Fontaine A, Corcoran C, Piazzolo S, Belousova E, Peng Z, Gault B, Saxey D, Fougerouse D, Reddy S, Pedrazzini S, Bagot P, Moody M, Langelier B, Moser D, Botton G, Vogel F, Thompson G, Blanchard P, Chiaromonte A, Reinhard D, Rice K, Schreiber D, Kruska K, Wang J & Cairney J (2018). Atom probe tomography analysis of the reference zircon gi-1: An interlaboratory study. *Chem Geol* 495, 27–35. <https://www.sciencedirect.com/science/article/pii/S0009254118303723>

Gault B, Chiaromonte A, Cojocaru-Mirédin O, Stender P, Dubosq R, Freysoldt C, Makineni SK, Li T, Moody M & Cairney JM (2021). Atom probe tomography. *Nat Rev Methods Primers* 1, 51. <https://doi.org/10.1038/s43586-021-00047-w>

Gault B, Moody MP, Cairney JM & Ringer SP (2012a). Atom probe crystallography. *Mater Today* 15, 378–386. [https://doi.org/10.1016/S1369-7021\(12\)70164-5](https://doi.org/10.1016/S1369-7021(12)70164-5)

Gault B, Moody MP, Cairney JM & Ringer SP (2012b). *Atom Probe Microscopy*. New York, NY: Springer. Available at <https://doi.org/10.1007/978-1-4614-3436-8>.

Gault B, Vurpillot F, Vella A, Gilbert M, Menand A, Blavette D & Deconihout B (2006). Design of a femtosecond laser assisted tomographic atom probe. *Rev Sci Instrum* 77, 43705. <https://doi.org/10.1063/1.2194089>

Geron A (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*, 2nd ed. Sebastopol, CA: O'Reilly Media, Inc. Available at <https://www.oreilly.com/library/view/hands-on-machine-learning/9781492032632/>.

Ghamarian I & Marquis E (2019). Hierarchical density-based cluster analysis framework for atom probe tomography data. *Ultramicroscopy* 200, 28–38. <https://doi.org/10.1016/j.ultramic.2019.01.011>

Giron S, Polin N, Adabifiroozjaei E, Yang Y, Kovács A, Almeida TP, Ohmer D, Üstüner K, Katter M, Radulov IA, Dunin-Borkowski RE, Farle M, Durst K, Zhang H, Alff L, Ollefs K, Xu BX, Gutfleisch O, Molina-Luna L, Gault B & Skokov KP (2023). Towards engineering the perfect defect in high-performing permanent magnets. Available at <https://arxiv.org/abs/2304.14958>.

Hastie T, Tibshirani R & Friedman J (2009). *The Elements of Statistical Learning: Data Mining, Inference and Prediction*, 2nd ed. New York, NY: Springer. Available at <https://www-stat.stanford.edu/tibs/ElemStatLearn/>.

The HDF Group (1997–2022). Hierarchical Data Format, version 5. Available at <https://www.hdfgroup.org/HDF5/>.

Hellman OC & Seidman DN (2003). Efficient sampling for three-dimensional atom probe microscopy data. *Ultramicroscopy* 95, 199–205. [https://doi.org/10.1016/S0304-3991\(02\)00317-0](https://doi.org/10.1016/S0304-3991(02)00317-0)

Jolliffe I (2002). *Principal Component Analysis*. New York: Springer Verlag.

Katter M, Weber J, Assmus W, Schrey P & Rodewald W (1996). A new model for the coercivity mechanism of Sm<sub>2</sub>(Co,Fe,Cu,Zr)<sub>17</sub> magnets. *IEEE Trans Magn* 32, 4815–4817. <https://doi.org/10.1109/20.539161>

- Kelly TF, Gribb TT, Olson JD, Martens RL, Shepard JD, Wiener SA, Kunicki TC, Ulfing RM, Lenz DR, Strennen EM, Oltman E, Bunton JH & Strait DR (2004). First data from a commercial local electrode atom probe (LEAP). *Microsc Microanal* 10, 373–383. <https://doi.org/10.1017/S1431927604040565>
- Khushaim M, Boll T, Seibert J, Haider F & Al-Kassab T (2015). Characterization of precipitation in Al–Li alloy AA2195 by means of atom probe tomography and transmission electron microscopy. *Adv Condens Matter Phys* 2015, 647468. <https://doi.org/10.1155/2015/647468>
- Kluyver T, Ragan-Kelley B, Pérez F, Granger B, Bussonnier M, Frederic J, Kelley K, Hamrick J, Grout J, Corlay S, Ivanov P, Avila D, Abdalla S & Willing C (2016). Jupyter notebooks—A publishing format for reproducible computational workflows. In *Positioning and Power in Academic Publishing: Players, Agents and Agendas*, Loizides F & Schmidt B (Eds.), pp. 87–90. IOS Press.
- Kühbach M, Rielli VV, Primig S, Saxena A, Mayweg D, Jenkins B, Antonov S, Reichmann A, Kardos S, Romaner L & Brockhauser S (2022). On strong-scaling and open-source tools for high-throughput quantification of material point cloud data: Composition gradients, microstructural object reconstruction, and spatial correlations. Available at <https://arxiv.org/abs/2205.13510>.
- Lefebvre W, Vurpillot F & Sauvage X (Ed.) (2016). *Atom Probe Tomography: Put Theory into Practice*. San Diego, CA: Academic Press. Available at <https://www.sciencedirect.com/book/9780128046470/atom-probe-tomography>.
- Lorensen WE & Cline HE (1987). Marching cubes: A high resolution 3D surface construction algorithm. *SIGGRAPH Comput Graph* 21, 163–169. <https://doi.org/10.1145/37402.37422>
- Madireddy S, Chung DW, Loeffler T, Sankaranarayanan SKRS, Seidman DN, Balaprakash P & Heinonen O (2019). Phase segmentation in atom-probe tomography using deep learning-based edge detection. *Sci Rep* 9, 20140. <https://doi.org/10.1038/s41598-019-56649-8>
- Marquis EA & Vurpillot F (2008). Chromatic aberrations in the field evaporation behavior of small precipitates. *Microsc Microanal* 14, 561–570. <https://doi.org/10.1017/S1431927608080793>
- Mason DR & London AJ (2020). Morphological analysis of 3d atom probe data using Minkowski functionals. *Ultramicroscopy* 211, 112940. <https://doi.org/10.1016/j.ultramic.2020.112940>
- Medrano S, Zhao H, De Geuser F, Gault B, Stephenson L, Deschamps A, Ponge D, Raabe D & Sinclair C (2018). Cluster hardening in Al–3Mg triggered by small Cu additions. *Acta Mater* 161, 12–20. <https://doi.org/10.1016/j.actamat.2018.08.050>
- Miller MK & Forbes RG (2014). *Atom-Probe Tomography*. New York, NY: Springer. Available at <https://doi.org/10.1007/978-1-4899-7430-3>.
- Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M & Duchesnay E (2011). Scikit-learn: Machine learning in Python. *J Mach Learn Res* 12, 2825–2830.
- Peng Z, Lu Y, Hatzoglou C, Kwiatkowski Da Silva A, Vurpillot F, Ponge D, Raabe D & Gault B (2019). An automated computational approach for complete in-plane compositional interface analysis by atom probe tomography. *Microsc Microanal* 25, 389–400. <https://doi.org/10.1017/S1431927618016112>
- Prithiv TS, Kloenne Z, Li D, Shi R, Zheng Y, Fraser HL, Gault B & Antonov S (2022). Grain boundary segregation and its implications regarding the formation of the grain boundary  $\alpha$  phase in the metastable  $\beta$ -Titanium Ti–5Al–5Mo–5V–3Cr alloy. *Scr Mater* 207, 114320. <https://doi.org/10.1016/j.scriptamat.2021.114320>
- Ronneberger O, Fischer P & Brox T (2015). U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, Navab N, Hornegger J, Wells WM & Frangi AF (Eds.), pp. 234–241. Cham: Springer International Publishing. Available at [https://link.springer.com/chapter/10.1007/978-3-319-24574-4\\_28](https://link.springer.com/chapter/10.1007/978-3-319-24574-4_28).
- Stephenson LT, Moody MP, Liddicoat PV & Ringer SP (2007). New techniques for the analysis of fine-scaled clustering phenomena within Atom Probe Tomography (APT) data. *Microsc Microanal* 13, 448–463.
- 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. <https://doi.org/10.1016/j.ultramic.2006.06.008>
- Torres K, Daniil M, Willard M & Thompson G (2011). The influence of voxel size on atom probe tomography data. *Ultramicroscopy* 111, 464–468, special Issue: 52nd International Field Emission Symposium. <https://doi.org/10.1016/j.ultramic.2011.01.001>
- Vaumousse D, Cerezo A & Warren P (2003). A procedure for quantification of precipitate microstructures from three-dimensional atom probe data. *Ultramicroscopy* 95, 215–221. [https://doi.org/10.1016/S0304-3991\(02\)00319-4](https://doi.org/10.1016/S0304-3991(02)00319-4)
- Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski E, Peterson P, Weckesser W, Bright J, van der Walt SJ, Brett M, Wilson J, Millman KJ, Mayorov N, Nelson ARJ, Jones E, Kern R, Larson E, Carey CJ, Polat İ, Feng Y, Moore EW, VanderPlas J, Laxalde D, Perktold J, Cimrman R, Henriksen I, Quintero EA, Harris CR, Archibald AM, Ribeiro AH, Pedregosa F & van Mulbregt P, SciPy 1.0 Contributors (2020). SciPy 1.0: Fundamental algorithms for scientific computing in python. *Nat Methods* 17, 261–272. <https://doi.org/10.1038/s41592-019-0686-2>
- Vogel F, Wanderka N, Balogh Z, Ibrahim M, Stender P, Schmitz G, Fedorova T & Banhart J (2015). Evolution of nanoscale clusters in  $\gamma'$  precipitates of a Ni–Al–Ti model alloy. *Ultramicroscopy* 159, 278–284. <https://doi.org/10.1016/j.ultramic.2015.04.021>
- Vurpillot F, Bostel A & Blavette D (2000). Trajectory overlaps and local magnification in three-dimensional atom probe. *Appl Phys Lett* 76, 3127–3129. <https://doi.org/10.1063/1.126545>
- Wang J, Schreiber DK, Bailey N, Hosemann P & Toloczko MB (2019). The application of the optics algorithm to cluster analysis in atom probe tomography data. *Microsc Microanal* 25, 338–348. <https://doi.org/10.1017/S1431927618015386>
- Zandbergen MW, Xu Q, Cerezo A & Smith GD (2015). Study of precipitation in Al–Mg–Si alloys by atom probe tomography I. Microstructural changes as a function of ageing temperature. *Acta Mater* 101, 136–148. <https://doi.org/10.1016/j.actamat.2015.08.017>
- Zelenty J, Dahl A, Hyde J, Smith GD & Moody MP (2017). Detecting clusters in atom probe data with gaussian mixture models. *Microsc Microanal* 23, 269–278. <https://doi.org/10.1017/S1431927617000320>
- Zhou X, Mianroodi JR, da Silva AK, Koenig T, Thompson GB, Shanthraj P, Ponge D, Gault B, Svendsen B & Raabe D (2021). The hidden structure dependence of the chemical life of dislocations. *Sci Adv* 7, 1–9. <https://doi.org/10.1126/sciadv.abf0563>
- Zhou X, Wei Y, Kühbach M, Zhao H, Vogel F, Kamachali RD, Thompson GB, Raabe D & Gault B (2022). Revealing in-plane grain boundary composition features through machine learning from atom probe tomography data. *Acta Mater* 226, 117633. <https://doi.org/10.1016/j.actamat.2022.117633>