Quantitative chemical imaging in atom probe tomography

Kaustubh Kaluskar

Iowa State University

Follow this and additional works at: https://lib.dr.iastate.edu/etd

Part of the Materials Science and Engineering Commons, and the Mechanics of Materials Commons

Recommended Citation
Kaluskar, Kaustubh, "Quantitative chemical imaging in atom probe tomography" (2013). Graduate Theses and Dissertations. 17713.
https://lib.dr.iastate.edu/etd/17713

This Dissertation is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
Quantitative chemical imaging in atom probe tomography

by

Kaustubh Kaluskar

A dissertation submitted to the graduate faculty
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Materials Science and Engineering

Program of Study Committee:

Krishna Rajan, Major Professor
Kaitlin Bratlie
Kristen Constant
Baskar Ganapathysubramanian
Richard LeSar

Iowa State University

Ames, Iowa

2013

Copyright © Kaustubh Kaluskar, 2013. All rights reserved.
# TABLE OF CONTENTS

ACKNOWLEDGEMENTS ........................................................................................................ v

ABSTRACT .......................................................................................................................... vii

CHAPTER 1: INTRODUCTION .............................................................................................. 1

1.1 Objective of the Thesis ................................................................................................. 2
  1.1.1 Imaging ..................................................................................................................... 2
  1.1.2 Novelty of the Work ............................................................................................... 7
  1.1.3 Voxel Size ............................................................................................................... 8
  1.1.4 Concentration Threshold ....................................................................................... 9
  1.1.5 Uncertainty Quantification and Visualization ....................................................... 9
  1.2 Thesis Outline ........................................................................................................... 10

CHAPTER 2: IMAGING IN ATOM PROBE TOMOGRAPHY .............................................. 12

2.1 APT Basics .................................................................................................................. 12

2.2 Imaging Basics and Challenges .................................................................................. 15

2.3 Statistical Methods to Image Chemistry .................................................................... 18
  2.3.1 Interatomic Distance .............................................................................................. 19
  2.3.2 Concentration Threshold ....................................................................................... 21

2.4 Role of Input Parameters ............................................................................................ 23

2.5 Chapter Summary ....................................................................................................... 24
CHAPTER 3: VOXEL SIZE SELECTION IN ATOM PROBE TOMOGRAPHY.... 26

3.1 Introduction ........................................................................................................... 26

3.2 Role of Voxel in Density Estimation...................................................................... 28

3.3 Proposed Approach............................................................................................. 34

3.4 Results.................................................................................................................. 37

3.5 Chapter Summary............................................................................................... 39

CHAPTER 4: TOPOLOGY FOR APT IMAGING.............................................................. 41

4.1 Introduction ........................................................................................................... 41

4.2 Previous Work....................................................................................................... 44

4.3 Proposed Approach............................................................................................. 45

4.3.1 Structure from Point Data................................................................................ 46

4.3.2 Quantification of Structure Using Betti Numbers............................................ 47

4.4 Material................................................................................................................ 49

4.5 Results and Discussion....................................................................................... 50

4.5.1 Morphology Analysis...................................................................................... 50

4.5.2 Comparison...................................................................................................... 61

4.5.3 Feature Identification ..................................................................................... 63

4.6 Chapter Summary............................................................................................... 65

CHAPTER 5: SPATIAL UNCERTAINTY IN ISOSURFACE............................................ 66
5.1 Introduction ........................................................................................................................................ 66

5.2 Isosurface as Data ................................................................................................................................ 68

5.3 Error and Uncertainty ............................................................................................................................ 71

5.4 Proposed Approach .................................................................................................................................. 72

5.4.1 Parameter Calculation for Gaussian Function ................................................................................... 75

5.5 Results ..................................................................................................................................................... 76

5.6 Chapter Summary .................................................................................................................................... 83

CHAPTER 6: SUMMARY AND FUTURE WORK ....................................................................................... 85

6.1 Summary ................................................................................................................................................ 85

APPENDIX: CODES ....................................................................................................................................... 88

A1: Voxel Size Selection ................................................................................................................................. 88

A2: Uncertainty Code .................................................................................................................................. 94

REFERENCES ................................................................................................................................................ 98
ACKNOWLEDGEMENTS

This work would not have been possible without constant support and help of many individuals, both in and outside the lab. First I would like to acknowledge all the support provided by my advisor Dr. Rajan. This work began with his suggestions and evolved into its present form with his continuous guidance at each stage of this work. In particular, I would like to acknowledge the freedom and resources available to any group member for pursuing new ideas.

I would also like to acknowledge the support from group members at various stages of my work. They helped me in and out of lab. I learnt many of skills during my work and this learning process would not have been possible without their involvement. I would like to acknowledge Joaquin and Claudia for their help in understanding physics from computational view point, Scott and Prasanna for help in data analysis techniques, and Santosh and Mike for Atom Probe Tomography.

My experience during the PhD improved a lot due to excellent company I had in lab and outside. I would like to acknowledge all the group members for supportive and interesting environment in the lab. Special thanks to Dr. Chitra Rajan for all the cakes, beside all the support she provided. My friends Varun and Vidya were very helpful when I was trying to settle down and acted as guide and mentor throughout my stay.
Along with this thesis, I would like to believe, I have learnt a lot. This thesis is result of the work and support of many people. I would like to end this with a sincere thank you to all those people and hope you find this work a good output.
ABSTRACT

Atom Probe Tomography (APT) resolves atoms in real space and detects their chemical identity. In the realm of microscopic techniques, it has its unique place with sub nanometer spatial resolution and high chemical sensitivity. In this thesis, data driven techniques have been developed to identify nano scale chemical features from reconstructed atomic data obtained from APT experiments. Main drawbacks in present day methods, to detect nano scale features, are use of input parameters involving heuristics on part of user. Techniques developed in this work use the APT data to determine the input parameters, thereby making the process more quantitative. In particular, techniques have been developed to select optimal voxel size to calculate the concentration profile, and to select the concentration threshold using ideas from computational topology. Voxel size is selected using an error minimization technique. A framework to quantify and visualize spatial uncertainty in isosurface has also been developed. Approaches developed in this work are generic in nature and can be applied on any APT data. In this work, results have been shown for Ni based superalloy data.
CHAPTER 1: INTRODUCTION

In the realm of microscopic techniques used to study material systems, Atom Probe Tomography (APT) has a unique place. Its spatial resolution is less than a nanometer allowing it to image single atoms. Its chemical sensitivity is same for all the elements. This allows it to recreate the 3 Dimensional image of material system under investigation where atomic positions and their chemical identities are resolved [1], [2], [3]. In this sense, it can image smallest engineering building blocks of engineered material systems. Introduction of laser assisted atom probe has made it possible to study material systems having low electrical conductivity such as polymers and biological systems [4], [5], [6]. Wide field of view atom probe has enabled study of larger specimen volume [7], [8]. In computational materials science, atomic scale techniques such as density functional theory or molecular dynamics provide an idea about relative distribution of atoms as per some theory. Atom Probe Tomography provides an experimental evidence of atomic distribution relative to other atoms [9]. Materials science finds its usefulness from the fact that response of a material to external stimuli is a function of its atomic chemistry as well as how the atoms are arranged relative to each other. Atom Probe Tomography, therefore, is a unique and very useful tool to study materials.
From an engineering perspective, atoms are most fundamental building blocks. However, many of the theories in materials science have been developed over the ages for nanostructures or microstructures. It is because in a material collection of atoms, rather than a single atom, play important role. Grain boundary or precipitates are such example where collection of atoms sharing some common features, such as chemistry or structure, affects the final property. Due to high resolution and chemical sensitivity, APT has been used for many studies of nano scale features [10]. These features are not imaged in APT but are defined by the user by fixing some parameter such as inter atomic distance threshold or concentration threshold. Appropriate selection of these parameters is necessary for meaningful feature identification. Some methods have been proposed in the past to guide the selection of these parameters. Selection of these parameters involves heuristic decisions on part of user. In this work, I have developed an approach where selection is data driven removing the ambiguity. This results in images of features which are more meaningful. In addition, I have developed the procedure to quantify and visualize the spatial uncertainty involved in imaging features.

1.1 Objective of the Thesis

1.1.1 Imaging

At the outset, we state our objective: To image chemical features using APT data. An example of APT output is shown in Fig.1.1. Also shown in the figure are outputs
Figure 1.1. Final reconstruction result using Atom Probe Tomography. An example electron micrograph using TEM. While atoms of different chemistries are resolved in APT, boundaries between different features become diffused. In the TEM, imaging mechanism produces the required contrast which differentiates between phases. To identify feature boundaries, it is required to define a contrast function and limit the contrast threshold [11], [12].

obtained from TEM. It can be seen that APT is able to resolve individual atoms along with their chemistry. However, when it comes to resolving boundaries between distinct features, APT falls short of it. On the other hand, in TEM while individual atoms are not resolved, yet the boundaries are clearly visible. If we can identify the boundary region
in Atom Probe then it will be very helpful because it will allow us to analyze the region with respect to distribution of surrounding atoms in 3 Dimension.

![Diagram](image)

**Figure 1.2.** In Atom Probe Tomography, we get spatial location and chemical identity of feature. To identify features, we are looking to develop some kind of functional rule which take as an input some spatial/chemical data and then classify the atom/region as belonging to Matrix or Precipitate.

Problem of finding a functional definition of precipitate for an APT output is a difficult problem. In other microscopy techniques, such as Transmission Electron Microscopy, electron beam interacts with the energy field resulting in a contrast in output thereby resolving the feature. In APT, user has to define the property for contrast and limit the extent or threshold of the contrast. Metallurgical precipitates are well defined features based on concentration and crystal structure. In Atom Probe Tomography, since almost half of the atoms are not detected and even for those which are detected it is not always possible to extract structural information [13], [14], it is not possible to calculate these defining properties. As a consequence of these difficulties, it is not possible to exactly identify the precipitate. Rather clusters of atoms are identified based on some similarity measure such as nearest neighbor distance, radial distribution...
function or concentration in a region [15], [16]. These are based on the fact that concentration of solute atoms in the precipitate region is greater than the concentration in matrix region. Since concentration is not directly obtained in APT, it is measured indirectly. In first case, concentration argument translates into the fact that interatomic distance in solute atoms will be less than in precipitate region because solute atoms are closer to each other in a precipitate than in the matrix. It requires user to define a threshold for the maximum distance between the atoms, $d_{\text{max}}$, and minimum number of atoms, $N_{\text{min}}$, to fix cluster. Second method requires calculation of concentration from atomic data. For this, reconstructed volume is first sub divided in smaller volumes, called voxels. For each voxel, concentration is calculated. Isoconcentration surfaces are drawn for different concentration threshold and appropriate concentration threshold is chosen for which concentration gradient across the isoconcentration surface is optimum. It requires user to decide on voxel size and concentration threshold. There are other methods but they are variant of these two methods. Table 1.1 presents a list of these methods. In present work, we follow the similar approach as isoconcentration surface, but develop a methodology to determine every input parameter. Determination is based on the data thereby eliminating the heuristic on the part of user. Objective of the thesis is, thus:

i) To determine optimum voxel size based on data distribution
ii) To determine concentration threshold by mapping topology of data to expected morphology of precipitate

iii) To develop a framework to quantify and visualize uncertainty associated with above features

<table>
<thead>
<tr>
<th>Table 1.1. Previous Work</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Author</strong></td>
</tr>
<tr>
<td>Hyde[17], Vaumousse and Cerezo[18]</td>
</tr>
<tr>
<td>Hellman et al.[19]</td>
</tr>
<tr>
<td>Stephenson et al.[20]</td>
</tr>
<tr>
<td>Lefebvre et al.[21]</td>
</tr>
</tbody>
</table>
Figure 1.3. Steps and inputs required in feature visualization in Atom Probe Tomography. In first step, atomic data is gridded in smaller volume called voxels. User has to decide voxel size. From the voxelized data, concentration is calculated and isoconcentration surfaces are drawn. A concentration threshold is to be chosen based on concentration gradient across the isoconcentration surface. Once the feature is detected, associated uncertainty should be quantified and included in the final visualization. There is no way to accurately decide on these parameters. But optimum input and computation at each step will result in an image which is close the data characteristics.

1.1.2 Novelty of the Work

Present work is a unified attempt dealing with each step involved in visualization pipeline of metallurgical features in Atom Probe Tomography. Each of these steps requires user input and final results are therefore user dependent. In this thesis, I have laid down steps which provide a data driven approach to decide these parameters thereby removing heuristics. Fig.1.3 shows steps involved in realizing an image of a feature from the Atom Probe Tomography data.
1.1.3 Voxel Size

Voxel Size determination is an important step for subsequent analysis [22], [23], [24]. It determines the level of details available for calculation. Too big a size results in averaging of data and may result in loss of feature information. Too little a voxel size may result in final output affected by statistical noise [25]. Optimum choice would be the voxel size for which data averaging takes place in such a way that features are preserved. For this, we use an error minimization approach. For different voxel sizes, we calculate the error associated with result and choose the voxel size which minimizes the error. Error measure is chosen so as to indicate departure of final result from underlying data [26]. This makes the voxel size determination an automated task.

Table 1.2. Previous works on voxel size determination

<table>
<thead>
<tr>
<th>Author</th>
<th>Contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hetherington and Miller[22]</td>
<td>Suggested voxel size optimization based on minimization of variance of concentration in neighboring voxels</td>
</tr>
<tr>
<td>Torres et al. [24]</td>
<td>Implemented the above method for LEAP data</td>
</tr>
</tbody>
</table>
1.1.4 Concentration Threshold

Interfaces used to define precipitate regions in Atom Probe Tomography are drawn using isoconcentration surfaces at concentration threshold. Therefore it is important to choose correct concentration threshold so that appropriate regions are selected. One approach to select concentration threshold is to draw proximity histogram [19] which represent average concentration gradient across the interface. However, choice of concentration gradient value is user dependent. In this thesis, I propose an alternate approach where concentration threshold is determined by mapping the topology of the dataset to expected topology. Precipitates in material systems have well defined topology. We use techniques derived from computational homology [27] to analyze APT dataset for concentration thresholds at which topology of dataset matches with expected topology.

1.1.5 Uncertainty Quantification and Visualization

Uncertainty is inherent in all kind of data. Source of uncertainty can be instrumental – limited resolution, or it may enter in final result via computational steps. Final output of Atom Probe Tomography is essentially the outcome of computational operations on instrument parameters. While it is common practice to display error bars in 2D plots, idea of uncertainty is seldom used in 3D visualization. In this work, we have incorporated recent developments in the field of statistical analysis and computer
Table 1.3. Previous work on uncertainty of isosurface

<table>
<thead>
<tr>
<th>Author</th>
<th>Contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hellman et al. [25]</td>
<td>Introduced the idea of sampling error and counting statistics error. Did not include it in final visualization scheme.</td>
</tr>
</tbody>
</table>

graphics with APT data to quantify and display uncertainty due to sampling process [28]. This approach lays down the steps involved in uncertainty quantification by identifying sources of uncertainty, ways to rectify and quantify them, and display them on computer screen so that information is available to the user in an intuitive way.

1.2 Thesis Outline

Main underlying theme of the thesis is to develop methods which lead to automated imaging of features of interest and involve as less human involvement as possible to limit the heuristics involved. This chapter presented an overview of the thesis. Main contributions of this thesis are presented in

- Chapter 3: On Voxel Size Determination
- Chapter 4: On Topological Analysis of Atom Probe Tomography Data
- Chapter 5: On Uncertainty Quantification and Visualization
Basics of APT and state of the art in chemical imaging in APT are explained in Chapter 2: Atom Probe Tomography basics. In the final chapter, Chapter 6, we summarize our findings and suggest future directions this research can take.
CHAPTER 2: IMAGING IN ATOM PROBE TOMOGRAPHY

This chapter explains the basics of Atom Probe Tomography (APT) and imaging of features obtained from it. It is divided into four parts. In first part, operational principles of APT are mentioned. In second part, basic idea of imaging data and particular challenges in imaging APT data are discussed. In third part, review of available technique is done. In fourth part, influence of various input parameters of final image is discussed.

2.1 APT Basics

From material to final analysis, APT can broadly be divided into 3 steps: Specimen preparation, experimentation, and analysis. APT works on the principle of field evaporation of surface atoms of specimen under the influence of electric field [29], [30], [31], [32]. For an applied voltage \( V \), electric field \( F \) experienced by an atom at the end is proportional to radius as shown in equation (2.1).

\[
F \propto \frac{V}{R}
\]  

(2.1)

For a very low end radius, high enough electric field can be achieved at application of low voltage which is sufficient for field evaporation of atoms as positive ions [33], [34], [35], [36]. End radius of specimen is of the order of nm. Electropolishing
Focused Ion Beam milling [38], and thin film deposition [39] are some of the techniques used to produce the specimen. To this specimen, positive voltage is applied in the APT analysis chamber. To control the sequence of evaporation, voltage is slightly less than the required voltage. To this base voltage, a voltage pulse of high frequency is applied for a short duration. This is done to fix the time of evaporation. For material systems having low electrical conductivity, laser pulse is used to assist in evaporation [40]. Ion evaporated from the sample flows through the electric field line and hits the detector at the end. Time from start of evaporation to detector hit, $\Delta t$, is used to determine the mass to charge ratio of the ion using Time of Flight Mass Spectrometry [41] [42], [43], given by equation (2.2).

$$\frac{\text{Mass (m)}}{\text{Charge (ne)}} = \frac{2V_{tot}\Delta t^2}{L^2}$$

Here $V_{tot}$ is the net applied voltage, $L$ is the length of flight path, and $\Delta t$ is the flight time. Figure 2.1 shows a schematic of APT process as taking place in analysis chamber.
Spatial location of atoms in the specimen is reconstructed based on the location where ion hits the detector and its sequence of evaporation. Hit location is used to fix the X and Y coordinates and sequence of evaporation is used to fix the Z coordinate.

![Figure 2.1](image)

**Figure 2.1.** Schematic of Atom Probe Tomography. R is the radius of the tip of the sample which is normally in the range of 20 – 100 nm. L is the flight length. $V_{DC}$ is the voltage applied to a specimen; $V_{pulse}$ is the pulse voltage to control the evaporation process. $X_D$, $Y_D$ is the location of ion hit on the detector.

Reconstruction is done using some reconstruction algorithm which is based on the idea that ions in the electric field follow the field lines and work on some assumptions about tip shape. More details can be found here [7], [44], [45], [46]. These
reconstruction procedures are usually implemented in proprietary software accompanying the APT. After the reconstruction process is complete, output data contains information about spatial location of detected atoms and mass to charge ratio. Ranges of mass to charge ratio is then assigned to different elements by user.

2.2 Imaging Basics and Challenges

Data in science and engineering is always present in the form of numbers. When this data is useful for some particular application, it becomes information. It is common for data point in datasets to run into thousands and millions of data points. While useful inference can be drawn by reading the data if number of data points are less, it becomes increasingly difficult, time consuming, prone to error, and ultimately impossible to do so with increasing size of dataset. Visualization of dataset is a useful way to convey information contained in the dataset. When the data correspond to some physical feature then by visualizing the data, we are visualizing the feature. In this case, visualization is called imaging.

Reconstructed atom probe data contains millions of data points where each data point maps to an atom in the real space. This location in real space is fixed by first three columns and chemical identity of atom is fixed by fourth column. If this reconstructed data is visualized, then each data point will be an image of an atom in specimen. APT datasets runs into millions of data points. Size of dataset presents two problems. First, it
stretches to limit the capability of present day computer. Second, image of the dataset is very dense which makes it difficult to fully understand the useful features. Moreover, this dataset is in 3 Dimension. Therefore, at a time only one layer of data is visible [47].

While atoms represent smallest unit of information contained in APT data, it is seldom the point of interest. Rather, it is the group of atoms sharing some similar property as opposed to other atoms. For instance, precipitates are a feature in microstructure playing important role in mechanical and electrical property of material. They act as pinning sites for dislocation movement [48]. At atomic level, they are characterized by different local chemistry and/or crystal structure. In APT, dataset maps individual atoms to image in real space. In order to image features such as precipitate, it is first necessary to identify atoms which map to the feature.

In microscopic techniques such as optical microscopy or transmission electron microscopy, imaging mechanism interacts with the feature properties and contrast between the features is natural outcome of this interaction. For example, in TEM, electron beam interacts with the material resulting in electron micrograph where intensity is indicative of feature [49]. In APT atoms are imaged and from relative distribution of atoms, different features are detected. This is done by mapping the expected distribution to the distribution of imaged atoms. However, due to
experimental limitations and limitations of reconstruction algorithm, this mapping is not straightforward.

Figure 2.2. Schematic of a) atoms in the specimen b) atoms after the reconstruction with only 50% - 60% atoms detected.

During the experiment, around 45 % of the atoms are not detected. Due to limitations of reconstruction algorithm, reconstructed atoms are displaced from their ideal position. Thus, some chemical and structural information is lost in the reconstruction. Due to this loss of information, mapping of features to reconstructed data is difficult.

Fig.2.2. shows a schematic of an ideal lattice and lattice after detection. Due to such detection rate, and distorted geometry, direct application of geometry based rule is
not possible. For instance, a circular distribution of atoms in the original material may not retain its shape in the reconstructed image. Therefore statistical techniques are used which are based on averaging of the detected data to derive useful conclusion [50], [51], [52].

2.3 Statistical Methods to Image Chemistry

In Atom Probe Tomography, chemical features such as precipitates are not identified automatically rather they are defined by the user. This definition, depending on the method used, is based on what is expected in the feature. In case of chemically distinct feature, difference in chemistry is the obvious choice to identify the features. Local chemistry in a region will be different from the neighborhood chemistry. This logic is implemented in different methods, of which two are most common. One method is to identify the clusters based on interatomic distance for a given chemistry and decide the threshold for interatomic distance or maximum separation to decide the cluster definition. Other method is to first draw interfaces and calculate the concentration gradient across these interfaces. Appropriate interfaces are chosen based on concentration gradient across them. Some other techniques used are pair correlation function [53], local chemistry analysis [54], contingency analysis[55], Fourier transform [56], auto correlation methods etc.
Consider a material system of elements $A$ and $B$. Let $A$ be the major element and $B$ be the minor element. Let $A_M$ and $B_M$ be the concentration of respective elements in matrix and $A_P$ and $B_P$ be their concentration in precipitates.

### 2.3.1 Interatomic Distance

Clusters are defined in terms of minor element. Let $d$ be the interatomic distance between any two $B$ atoms. $B$ atoms are closer to each other in precipitate than they are in matrix. Problem of defining cluster, then, reduces to defining the appropriate cutoff distance $d$ and this distance is chosen as the maximum distance, $d_{\text{max}}$. If solute atoms falls within or less than distance $d_{\text{max}}$ then they are included in the cluster. This, however, results in inclusion of even 2 atoms if their interatomic distance is less than $d_{\text{max}}$. To control this, another parameter, $N_{\text{min}}$ is defined which limits the selection of only those clusters where number of atoms is equal to or more than $N_{\text{min}}$. Using this method, therefore, requires defining two parameters: $d_{\text{max}}$ and $N_{\text{min}}$. This method was proposed by Hyde in 1993 [17]. Some approaches to select these parameters were proposed [18], [57]. Variations of this method have been proposed. For instance, core-linkage algorithm developed by Stephenson [20] is a combination of maximum separation based method and density based analysis. It focuses on higher nearest neighbor distances to decide the distance threshold. Lefebvre [21] introduced a Delaunay tessellation based approach to decide the distance threshold. Functionality of maximum
separation based method is increased by envelop method [58]. Envelop method overlays a three dimensional grid on the APT data. If the solute atom falls in a grid element, then that grid element is assigned to be part of precipitate. This allows for calculation of concentration and to image the feature [58]. It is explained schematically in Fig. 2.3.
**Figure 2.3.** Schematic of maximum separation method and envelop method. Red atoms form the solute atoms. As long as a red atom finds another red atom within distance \( d_{\text{max}} \), it is included in the fine scale cluster. Once all such atoms are detected, all atoms within the outer envelope of red atoms are included in the solute definition, envelop region shown in green squares [47].

### 2.3.2 Concentration Threshold

In concentration threshold method, precipitates are defined by fixing the concentration of isosurface [19], [57], [59]. Isodensity and Isoconcentration surfaces are used to draw the surfaces and concentration variation across these surfaces is calculated. Isosurfaces are 3D analogue of contours. They are drawn for a given property such that value of that property is constant at each point of the surface. To
draw these surfaces, reconstructed volume is first divided into smaller sub volumes and for each volume, density and concentration is calculated. Then surfaces are drawn for different threshold values [60]. For each threshold value, proximity histogram is calculated. Proximity histogram is a histogram of concentration variation of an element normal to the surface averaged across the three dimensions and plotted as a function of distance across the interface. Appropriate concentration threshold for an isosurface is chosen based on the concentration variation of elements across the surface. Concentration gradient of solute atom should be positive in going from matrix to precipitate.

Figure 2.4. Circles represent atoms and different colors signify different chemistry. Red surface is isoconcentration surface drawn for concentration of red atoms. To calculate
proximity histogram, distance of atoms from the surface along surface normal is calculated and number of atoms are binned as per their distance [47].

**Figure 2.5.** Proximity histogram for the isoconcentration surface in Fig. 2.5. 0 on the x axis represent the surface. Negative X axis is for region outside the interface and positive X axis is for region contained in the interface. It can be seen that concentration of Al increases across the interface from matrix to cluster region and concentration decreases for rest of the elements [47].

### 2.4 Role of Input Parameters

Clusters defined using these methods are dependent on the input variables. In case of maximum separation method, distance and number of atoms are required variable. Further, if envelop method is applied, then grid spacing also needs to be defined. If too large distance is chosen, then atoms outside the precipitate will be part of the cluster and too small distance will leave out the precipitate atoms. Similarly, for number of atoms, choice will decide the size of final cluster. These choices are to be
made by the user and can be made based on material property, earlier known values, and/or simulations. Still, there is no foolproof way to decide on these values.

For concentration threshold method, input parameter include voxel size and concentration threshold. Voxel size decides the extent of averaging of discrete data in density and concentration calculation. If too large a size is chosen, than variations in data will be averaged out and if too small size is chosen than result will be affected by noise in the data. In case of concentration threshold, decision is mostly based on heuristic. There is no accepted method to decide the appropriate concentration gradient. It is an exploratory process where concentration gradients for different concentration threshold are calculated.

Proper selection of input parameters is necessary for correct identification of chemical features. Unfortunately, there is no way to check for its correctness. In case of materials system already studied by some other technique, cross validation can be done but in case of new material systems, it is difficult to validate the results and therefore improved statistical techniques needs to be developed to select appropriate input parameters.

### 2.5 Chapter Summary

This chapter covered basics of Atom Probe Tomography. Imaging of chemical features involves application of statistical technique on reconstructed data. Therefore
nature of reconstructed data, difficulty in imaging chemical features, different techniques developed for imaging, their advantages and limitations were covered. Experimental details of Atom Probe Tomography were briefly covered.
CHAPTER 3: VOXEL SIZE SELECTION IN ATOM PROBE TOMOGRAPHY

Data driven approach to select voxel size in Atom Probe Tomography is proposed. Voxelization is used in Atom Probe Tomography to calculate density/concentration profile of the sample from the atomic data. For the decisions based on threshold of density/concentration value, such as drawing isosurfaces, it is necessary that resultant profile represents the underlying data. Due to destructive nature of technique, it is not possible to validate the result and inaccurate selection of voxel size may result in averaging out of features or may be affected by noise. To overcome this, an error minimization based approach is developed in this work.

3.1 Introduction

Atom Probe Tomography (APT) detects spatial location and chemical identity of millions of atoms per successful run. This information, combined with available graphics capability of desktop computers, creates very detailed image of internal atomic level chemical variation [61]. APT works on the principle of field evaporation and time of flight mass spectrometry. Atoms from the surface of sample are field evaporated on application of positive voltage. Ions travel through the field lines and hit the detector. Their position in the sample is reconstructed based on their location and sequence of hit on the detector. Time taken from the application of pulse to hit registration is used to
determine mass to charge ratio of the ion. This information reconstructs the atomic level
image of sample [43].

Reconstructed APT data is a data set where each row represents an atom and
three columns fix atom’s spatial location and fourth column fix its chemical identity. For
many analyses it is the average atomic distribution which is required. For instance,
density and concentration values across the sample are used to identify clusters. Since
these values are not naturally obtained from the experiment, they are to be computed
using atomic data and involve averaging over a region, known as voxel in APT [24].
This poses the question what should be the voxel size for averaging. In a homogeneous
material system, density or concentration will not depend on the size of averaging but
for a heterogeneous system, extent of averaging will affect the final result. Small voxel
size may result in statistical noise whereas large voxel size may lead to smoothing of
important data variations. For density/concentration values to represent underlying
data accurately, it is therefore necessary to select a voxel size which is derived from the
underlying data and takes into account the variation of data.

With the introduction of 3 Dimensional Atom Probe, it was suggested to divide
the analyzed data into regions of small volume and to measure the composition in that
small volume[22], [23]. Method based on selecting the voxel size by minimizing the
concentration difference between neighboring voxels was proposed. Recently this
method was applied to data obtained from LEAP instrument. In present work, we approach this problem as density estimation problem of scattered data [53], [26], [62][63]. Kernel density estimation or KDE is a non-parametric way to estimate data intensity in a given region. Kernels refer to the mathematical function used to sample the data. When solving the problem as kernel density estimation problem, voxel size represents the smoothing parameter which is indicative of the extent of smoothing applied in final representation.

Rest of the chapter is organized as follows. In next part, we explain the steps involved in calculating the density or calculation value from the reconstructed atomic data. With this information, we propose our approach. We then apply this approach on an experimental dataset and discuss the result.

3.2 Role of Voxel in Density Estimation

Atom Probe Tomography reconstructed dataset is an n by 4 dataset. Each of the n rows represents an atom. First three columns fix the geometric position and fourth column fixes the mass to charge ratio of atom. Mass to charge ratios are ranged to a list of elements and thus based on where it falls, atom’s elemental identity is also fixed.

Atomic density of element \( i \) is defined as number of atoms of element \( i \) per unit volume. Atomic concentration of element \( i \) is defined as number of atoms of element \( i \) in a given volume divided by total number of atoms in the volume. Thus, final outcome is
dependent on the definition of unit volume. If too large a unit volume is chosen, then important variation in data will be averaged out. If too small a unit volume is chosen, then final outcome will be affected by the noise [25]. For rest of the chapter, we use the term density. Once the unit volume is selected for density, concentration calculations can easily be done.

First step in density calculation is dividing the reconstructed volume in smaller cubical region. Each atom contributes some value at the center. This contribution is a function of its location and sampling formula used. Contributions from all atoms are added up. This gives the net value at the center. This is repeated for all the voxels to get the density variation across the sample. Fig.3.1. shows the voxel with atoms. h is the voxel edge length. \(x_i\) are the atoms distributed in the voxel. Contribution from each atom is calculated at the center \(X\).

Each of the atoms contributes some value to the center of the voxel. This value depends on the function used to calculate the contribution, known as sampling function. At the location of each atom, a kernel function is superimposed. At the voxel center, all the net contributions due to that kernel function are added up to give resultant density. This is shown in Fig. 3.2.
Figure 3.1. h is the edge length of the voxel. $x_i$ denotes the atom locations. $X$ denotes the center of the voxel at which the density is calculated.

Figure 3.2. Density Calculation explained in 1D. Green circles on the x axis represent the data points. At each data point, a kernel function is imposed as shown in red curves. Contribution of each data point in its neighborhood is function of its distance as determined by the kernel function. Net density across the region is determined by adding up respective contributions from different data points, as shown by the blue curve.
Figure 3.3. Kernel Functions in 1D. On left is shown a box function in which every atom contributes 1 to the center irrespective of its position. On right is shown the Gaussian function where contribution is based on its distance and Gaussian kernel.

Fig. 3.3 displays examples of two sampling function or kernel function. Shown on left is a box function. In this, an atom contributes a value of one when it falls in a voxel, irrespective of its position in the voxel. While this is the simplest way to average the data, it does not take into account the difference in spatial location of atoms. Thus, an atom at the edge of the voxel will contribute same amount as the atom which is near the center of the atom.

On the right is shown an example of a Gaussian kernel. When Gaussian kernel is used, contribution is calculated based on the Gaussian function and distance of the atom from the center of the voxel. If X is the atom position and Y is the location where contribution is measured, then contribution will be:
And for Gaussian Kernel

$$Box\ Kernel\ Contribution = \begin{cases} 1 \text{ if } |Y - X| < h/2 \\ 0 \text{ if } |Y - X| > h/2 \end{cases}$$ (3.1)

$$Gaussian\ Kernel\ Contribution = \frac{1}{\sqrt{2\pi}\sigma^2} e^{-\frac{(Y-X)^2}{2\sigma^2}}$$ (3.2)

Here, h is the voxel edge length. Since contribution is measured at the center of the voxel, any point which lies further than h/2 will contribute to 0. In case of Gaussian kernel, there is no spatial limit to which an atom can contribute. To limit this, it is artificially cut to some value. This value, known as the width of the kernel, is dependent on voxel size and the quantity on which final density variation depends. Fig. 3.4 shows the effect of kernel window width on final density estimation. Fig. 3.5 shows the isoconcentration surface drawn for same concentration threshold but at different voxel sizes. Resultant surface is different in shape, morphology, and connectivity. Voxel size decides the density and concentration value in a particular region and geometry of surface is dependent on density and concentration value. Correct voxel size, therefore, is necessary for accurate representation of isosurfaces.
Figure 3.4. Different width for Gaussian kernel. Width of the kernel is important as it determines whether the variation of underlying density is captured in final output.
Figure 3.5. An example isoconcentration surface at same concentration threshold but at different voxel size. It can be seen that with varying voxel size, resultant morphology changes and final output is different.

3.3 Proposed Approach

Proposed approach is based on the idea that final density estimation should follow the same trend as the underlying data. It should have a maxima where density of data point is high, minima where density is low. We cannot have exact density value but we aim for trend in the density estimation to follow same as that of underlying data.
For this, we define an error function to compute the difference between real data and estimated data. Estimated data is defined at all the points. Real data is available only at discrete data points. At these discrete data points, we calculate the error between these data. As mentioned earlier, we want the estimated density profile to have same trend as the data. For this, we need some estimate which varies based on how much estimate differs from the actual data.

We define an error measure to quantify the variation between estimate and actual data. Let $E$ be the estimated density and $A$ be the actual data. Error at the actual data point is defined as the difference between the actual and error data:

$$\text{Error } (X_i) = E(X_i) - A(X_i)$$ (3.3)

Estimated value can be more or less than the actual value. We are not concerned with the sign of the error but the magnitude. To account for this, we square the error term. Also, since we want to minimize the error for the whole dataset rather than at some selected point, we integrate the error for whole dataset. This is known as Mean Integrated Square Error or MISE:

$$MISE = \int_{-\infty}^{+\infty} [E(x) - A(x)]^2 \, dx$$ (3.4)
Since actual data is available only at few points, integral becomes a summation over those points:

\[ MISE = \sum_{i=1}^{N} [E(x_i) - A(x_i)]^2 \]  

(3.5)

This is shown graphically in Fig. 3.6. Here difference between estimated density and actual density is shown. This is done for each voxel separately and errors are added up. Procedure is then repeated for different voxel sizes. Voxel size which gives lowest net error is selected for density estimation.
**Figure 3.6.** Difference between estimated density and actual density. Shaded red lines are actual density and continuous blue line is estimated density. Orange colored line is the difference between estimated and actual density. Proposed approach of voxel selection is based on the logic that estimated density should follow underlying data as much as possible and therefore the difference between estimated data and actual data should be as low as possible.

### 3.4 Results

We applied this approach on an experimental dataset. Dataset was in the form of atomic coordinate data points along with chemical labels. This dataset was divided into
smaller voxel dataset where atoms were assigned based on their location. For each atom in a voxel dataset, a Gaussian was fit at the atom coordinate and its amplitude was set at 1 with parameter such that Full width at half maxima was half of voxel size. Based on this Gaussian distribution centered at atomic coordinate, value of distribution was calculated at the voxel center and all such values were added up. This gave the amplitude of density at the voxel center. Error was calculated between the actual density and thus estimated density. Fig. 3.7 shows the result of the calculations.

**Figure 3.7.** Error between estimated density and actual density at atomic data points. Voxel size was varied from 0.7 nm to 2.4 nm. Minimum error was measured at 1.3 nm.
Voxel size was varied from 0.7 nm to 2.4 nm in the step of 0.1 nm. Calculations were done for each voxel size. Minimum error was obtained for 1.3 nm voxel size.

Computations were done using in-house developed MATLAB code. Code was run on a personal computer with 4 GB RAM and 2.46 GHz processor. Dataset consisted of 8 million atomic data points. It took around 8 minutes to process the dataset obtained for 1 nm voxel size.

In the result, it can be seen at higher voxel size that there is not much difference for subsequent voxel size. It may be due to the fact that at higher voxel size, data becomes averaged out. At lower voxel size, difference between subsequent voxel sizes is greater as results are more prone to noise. We expect minima at voxel size which is a tradeoff between the noise and average and it occurs at 1.3 nm.

3.5 Chapter Summary

Voxel size is an important parameter for APT data analysis. A data driven approach for voxel size selection is presented. This approach is based on the idea that estimated density should follow same trend as the data. To this end, an error measure is defined to quantify the deviation of estimated density from actual data and it is proposed that voxel size which minimizes this deviation should be selected for further analyses. We have used Gaussian kernel and Mean Integrated Square Error in this work but the approach is general and can be applied with other functions and error measures.
We have applied our method to an existing dataset and shown how one particular voxel size is better at following the data trend compared to others.
CHAPTER 4: TOPOLOGY FOR APT IMAGING

Atom Probe Tomography experiments produce millions of data points per run. This makes analysis of APT data a challenge. In this work, we present a new analysis technique which detects useful features using topology of APT point cloud data. In particular, we use tools from computational homology to quantify Isosurfaces in APT using Betti numbers, a topological invariant of an object. Utility of this approach is shown by demonstrating how it is able to: (1) capture and display spatial distribution of scalar fields (density/concentration) which is not apparent on visual inspection due to high dimensionality of dataset (2) quantify effect of input parameters on scalar fields (3) guide in feature selection.

4.1 Introduction

Atom Probe Tomography is a useful tool to analyze atomic level chemical distribution in materials. Its sub nanometer spatial resolution images atoms in direct lattice space[64], [65]. Development in instrumentation, such as laser assisted and wide field of view APT has enabled analysis of different materials system and increased data collection[40], [66], [67]. While this has led to more information, it also poses new challenges for data analysis.

APT experiments regularly produce millions of data points per run. Each valid hit on the detector can be reconstructed to locate spatial location and to detect chemical
Atom maps is most common way to represent APT data where each atom is represented as a dot and color of dot is used to differentiate between different chemistries. While this completely represents the data, it is difficult to extract useful information due to two reasons. First, atoms at the front make it difficult to get the whole picture. Second, boundaries of certain features such as grain boundary or interface between precipitate and matrix are not apparent. It is not possible to represent the 3D information on 2D surfaces completely [47]. Still, techniques such as proximity histogram represent useful chemical information in 3D on a 2D surface [19]. For visualization of concentration or density, isosurfaces are drawn at certain threshold. For imaging features of interest, it is first necessary to segment the region in reconstructed dataset. This makes it a data analysis problem.

APT gives spatial location and chemical identity of atoms in the sample. Development of data analysis tools for APT data plays an important role in extracting useful information from reconstructed data. APT is a destructive technique – once the experiment is over, there is very little which can be done to validate the result if any prior information is not available. APT dataset is a point set data distributed in space and lend itself very nicely to analysis by various statistical techniques. These are probably the reason why statistical techniques find such application in APT data analysis [68], [69]. For instance, crystal structure information is not readily available but
technique such as Fourier transform [70], [71], spatial distribution maps [14], [72], [73], [74] allows us to extract these information. While most of the analysis techniques in use mostly use geometrical approach, recently some data dimensionality reduction techniques and graph theory based approaches have also been proposed to analyze the complex data [75], [76].

In present work, we motivate the use of topological data analysis techniques to mine information APT dataset. Lately, topological data analysis has been applied to many data analysis problems where spatial distribution of data is important and resulting relationship between the data is important [77], [78], [79], [80], [81]. From this point of view, we can expect topological analysis of PAT data to provide useful information. We are interested in how atoms are distributed in the space, what the relationship between them is, and when different atoms are connected, how this connectivity influences the final shape. In fact it is the last point, how the connectivity influences the final shape, which is the focus point.

Rest of the chapter is arranged as follows. We explain some of the previous work done in the field. Then explain our proposed approach. We make use of tools from computational topology which has its own set of vocabulary. In this section, we also explain the terms and their relationship with APT data. Finally we explain three areas where topological data analysis can be helpful.
4.2 Previous Work

Topology based approach has been used in studying material microstructure in cases where finding geometrical relationships is difficult [82], [83], [84], [85]. Two topological concepts in particular, Betti number and Euler number has been used to quantify connectivity or network information in microstructure. Euler number gives information about the handle density in microstructure. Betti number gives information about the number of connected surface in zero, one, and two dimensions [86]. These are explained in detail later. Connectivity plays an important role in deciding mechanical response of material [87], [88], [89], [90]. With recent developments in computational topology, it has been applied in the field of microstructure characterization [91], [92].

In the field of Atom Probe, there have been few studies related to topology of dataset. All of these studies were related to spinodal decomposition [93], [94]. Recently, there was a comparison of spinodal decomposition simulation data with experimental atom probe data using Betti numbers parameters [95], [96]. Lack of topology based studies can be attributed to two main reasons. First is that topology, till very recently, was more of a theoretical subject with very little application on applied side. In recent years, developments in the field of computational homology and computational topology have made it easier to apply its technique to real world problems. Secondly, computations are very expensive. This barrier has reduced in recent years.
4.3 Proposed Approach

Atomic information obtained from APT is in the form of a 3D point cloud where each point represents an atom. Points themselves do not have any shape or size. However, there exists mathematical techniques which when applied on collection of data points provides useful information about its spatial structure. For the purpose of extracting useful structural information from this data it is necessary to first identify set of points which when taken together form useful structures. Once we have identified these points, we can then proceed to calculate useful parameters. First step is essentially a classification problem wherein our input consists of whole dataset and based on how we define useful features, output is classified dataset.

Atom Probe Tomography, due to limitations imposed by instrument, does not have a well-defined classification criteria for the atoms. In this work, we come up with classification criteria by examining structures formed by different combinations of atoms and comparing it with expected structures. We use ideas from topology to guide our process and tools from computational topology to do necessary calculations.

Our approach consists of three steps. First we generate structures using APT data. For density/concentration data, we use cubical complexes. These structures are generated at different length scales. Next, we quantify topology of these structures. For
this we compute first 3 Betti numbers of each structure. Finally, we come up with a classification criteria based on topology of structures at different scales.

4.3.1 Structure from Point Data

While using APT data, two kinds of point cloud data structures are encountered. Atomic point cloud data is present in the form of unorganized point cloud. Each atom is restricted in spatial space by x, y, and z coordinate and in chemical space by its mass to charge ratio. Density and concentration data is present in the form of structured data where its value is defined at points on a cubical grid.

For the purpose of extracting topology of data, we do not need to know exact shape of the data. We are interested in knowing if the points are connected and not how they are connected. Given set of points can be connected in many possible combinations. We need some connectivity criteria so that resulting structure has same topology as the structure we are interested in. Simplicial complexes are used to approximate the structure of unstructured point cloud which has same topology as original structure[97], [98], [99].

For the density data, marching cube is a common way to generate shapes corresponding to isovalue. However, for the purpose of our calculations, we only need connectivity information. For this, we make use of cubical complexes. Cubical complexes are complexes where each successive object can be specified by increasing
the index by 1. Vertex is a zero dimensional cubical complex. Edge is 1 dimensional, face is 2 dimensional and so on.

4.3.2 Quantification of Structure Using Betti Numbers

![Figure 4.1](image)

**Figure 4.1.** Distribution of point clouds, after introducing connectivity, can assume different shapes have well defined topology. When points in (a) are connected, they result in a single simply connected component. Whereas points in (b) after connectivity, in addition to a single component, also result in a loop. Points in 3D can potentially result in a cavity. To characterize topology of structures, it suffices to specify number of simply connected components and number of loops present in final structure.

Topological features are those which depend only on the connectivity and not on how they are connected [100]. Thus a circle and an ellipse have same topology although they are geometrically different. While a structure can have infinite shapes, many of which cannot be quantified, it can have only limited topological features depending on
its dimension. For instance, in 1D, structure can be simply connected or not connected. In 2D, a structure can be simply connected, or it can be connected to itself so that a loop is formed, or it cannot be connected. In 3D, a structure can be simply connected, or it can be connected such that a tunnel passes through it, or it can be connected to itself such that it encloses a cavity, or it cannot be connected. This is explained graphically in Fig. 4.1.

Thus, we can characterize the topology of a structure by counting number of simply connected components, number of tunnels and number of cavities. We have explained the concept of topological features in terms of geometrical figures but it should be understood that this is a mathematical concept, operating on the numbers and in case of real data, we never get continuous data. Topology is property of the dataset, and it depends on how close data points are to each other. Mathematically, topology resulting from this closeness can be quantified using the concept of Betti numbers. Based on some notion of closeness, if numbers are so arranged that when plotted in a 3D space they form m simply connected components, n tunnels, and p cavities, then first three Betti number of the dataset \( (\beta_0, \beta_1, \beta_2) \) will be m, n, and p. It should be understood that Betti numbers are property of the dataset. Since the data lies in a 3D space, we are able to associate physical shape to specific Betti number but this can be extended to higher dimensional data also [101].
4.4 Material

Method was applied on a Ni based superalloy to which Platinum group metals were added.

Figure 4.2. Mass spectrum of APT specimen used for the analysis. Specimen was Ni based superalloy with platinum group metals.

Figure 4.3. Atom map of Ni based super alloy used for analyses. Only major elements are shown. Color of the sphere corresponds to color in the mass spectra in Fig. 4.2.
4.5 Results and Discussion

For the purpose of calculation, reconstructed volume was sub divided into voxels of equal shapes. To study the effect of voxel size on final morphology, voxel size was varied from 0.7 nm to 1.3 nm in step size of 0.1 nm. For each voxel size, concentration was calculated using standard procedure as applied in IVAS.

4.5.1 Morphology Analysis

For a given dataset, concentration profile can assume different shapes based on concentration threshold. In this section, we examine change in concentration profile for Ni–Al dataset binned at 1 nm voxel size. Isosurfaces are 3D analogue of contours. When these are drawn for concentration value, they are called isoconcentration surfaces. Isoconcentration surfaces pass through the voxels having concentration value equal to the threshold concentration value. Thus we can get an idea about the nature of isoconcentration surfaces by understanding the connectivity of voxel data. Arrangement of voxels can give rise to three different kinds of morphology. These morphologies are characterized by three different Betti numbers. It is necessary to define connectivity in this case. Two voxels are considered connected if they share a face, an edge, or a vertex together. This is shown in Fig. 4.4.
Figure 4.4. Definition of connectivity is an important parameter as it decides the final shape of the surface. Voxels are connected if they share (a) Face (b) An edge or (c) A Vertex.

Figure 4.5. (a) Simply Connected (b) Simply connected with tunnel. (a) and (b) when considered together will be two simply connected plus one tunnel.

A 100 nm * 50 nm * 50 nm dataset will have $2.5 \times 10^4$ voxels. Depending on their concentration value, there arrangement can take variety of shapes. However, in terms of connectivity, there are only three possible ways, two of which are shown in Fig. 4.5. They can be arranged such that they form a simply connected morphology. They can also be arranged such that there is a tunnel passing through them. They can be connected such that they enclose a cavity in them. For any arrangement, simply
connected surface will always be present. Tunnels and Cavities may be present in addition to simply connected surfaces. Thus, for each concentration threshold, we first isolated those voxels which had concentration value within the limits. Next, coordinates of these were subjected to calculation. Output of calculation gave Zeroth Betti number, First Betti number, and Second Betti number which are equal to number of simply connected components, number of tunnels, and number of enclosed cavities. Fig. 4.6 shows the Betti number for different concentration thresholds.

**Figure 4.6.** Topology of Aluminium concentration profile at different concentration threshold. Different Betti numbers quantify different features in the concentration profile. Betti 0 quantifies number of simply connected components. Betti 1 quantifies number of loops through the surfaces and Betti 2 quantifies number of cavities enclosed by surfaces. This graph is a compact representation in 2D space of the morphology of concentration field over entire range of concentration in the sample.
On the X axis is concentration threshold. Maximum concentration was calculated to be 29%. For each concentration threshold, three values were calculated: Betti 0, Betti 1, and Betti 2. Betti 0 is the number of simply connected components. Betti 1 is the number of tunnels passing through the simply connected components and Betti 2 is the number of cavities present in simply connected components.

This graph represents in a compact 2D representation of full topology of concentration profile which is hard to extract just by visual inspection of reconstructed data. Consider 20% concentration threshold. It has Betti 0 = 110, Betti 1 = 365, and Betti 2 = 192 which means voxels having concentration value 20% or more are arranged such that they form 110 simply connected surfaces, 365 tunnels in total pass through these surfaces and they enclose 192 cavities in them. Another feature is trend of individual graph. For instance, focus on variation of Betti 0, the blue line. At high tail of the graph, we see a very low number of simply connected components. This is due to the fact that very few voxels have concentration value equal to or more than this threshold. As concentration threshold is decreased, more and more voxel qualify to be included in the group leading to an increase in Betti 0, which reaches maxima at 23%. For further decrease of concentration threshold, we see a decrease in Betti 0 which is indicative of the fact that new voxels which are now included in the calculation are leading to connectivity of earlier unconnected components. Changes in isoconcentration at
different concentration threshold and their relationship with Betti numbers is explained in Fig 4.7. and explained schematically in Fig. 4.8.

However, we see another maximum at concentration value of 13%, implying increase in number of unconnected components. It is indicative of two phases in the sample, $\gamma$ and $\gamma'$. Thus, Betti number signatures of concentration field provide a compact low dimensional representation of its morphological information.
Figure 4.7. Isoconcentration Surfaces from IVAS at different concentration threshold. On left side are shown surfaces for full dataset, arranged from maximum concentration at the top, where no voxels qualify for the threshold and hence blank box, to minimum concentration at the bottom, where all the voxels are included. Empty regions in last isoconcentration surface are due to edge effects. It can be seen that as concentration threshold is reduced, leading to higher number of voxel inclusion, more interconnections are introduced and visual interpretation and differentiation of features becomes difficult. On right is shown a small region in the dataset at decreasing concentration threshold to display how different regions appear at different concentration threshold. At first, three simply connected Isoconcentration surfaces appear, with Betti Zeroth equal to 0. As concentration threshold is decreased, neighboring voxels are also included. This results in appearance of a tunnel through the surface. Betti zero still remains 3 but Betti 1 increases to 1. Finally, neighboring unconnected components combine to form one surface resulting in Betti 0 = 0 and Betti 1 = 1. Isoconcentration surface may also contain cavities which are not visible due to limitation of visualization medium.
Figure 4.8. With change in concentration threshold, new morphologies appear in concentration profile. In (a), concentration threshold is 23%. This results in two simply connected components as voxel between them has concentration value of 21%. In (b), concentration threshold is reduced to 21% leading to inclusion of central voxel. This connects the whole structure and results in only one simply connected component. From (a) to (b), Betti 0 decreases from 2 to 1.

Another feature of this graph is relative trend of different Betti number graphs. We observe that as concentration threshold is decreased, trend of Betti 2 follows that of Betti 1 which in turn follows the trend of Betti 0. This trailing behavior is expected and explained in Fig. 4.9. Fig. 4.10 to Fig. 4.14 show how the visual representation changes with changing concentration threshold and corresponding change in Betti numbers.

Figure 4.9. As concentration threshold is decreased from (a) to (c), more voxels are added leading to first an increase in simply connected components, Betti 0, as in (b). Further decrease in concentration threshold leads to connectivity of earlier unconnected component but in such a way that they result in a tunnel, as in (c). Intensity of color is indicative of concentration.
Figure 4.10. At high concentration threshold set by user, very few voxels qualify to be included in the calculation. Here, isoconcentration surfaces are shown for the concentration threshold indicated by (1). In the image, very few simply connected surfaces are present. Our calculation shows that for this value, number of unconnected components is less than 50.
Figure 4.11. As concentration threshold is decreased, more region in the reconstructed specimen falls within the specified concentration range. These regions, in this case, are distributed throughout the sample. Therefore the increasing number of regions lead to increase in number of simply connected components. Shown here is an image for isoconcentration surface drawn at concentration threshold of 23%. At this concentration threshold, our calculation shows that highest numbers of unconnected components are present in specimen, more than 450.
Figure 4.12. For a given concentration threshold, isoconcentration surfaces drawn enclose the region where concentration is at least the concentration threshold. (2) was the point where maximum number of unconnected components were found. As concentration threshold is further decreased, more area is included and it leads to joining of two components which were earlier separated. Therefore, we see a decrease in number of components when decreasing concentration threshold from 23% to 16%.
We observe increase in number of component when concentration is decreased to 13% as compared to 15%. We know that as concentration is decreased, more voxels are included in the isoconcentration surface. Increase in number of unconnected components indicates that these new voxels are not in the neighborhood of already drawn isoconcentration surface. These new voxels, which are included, form some sort of island in the space. This is not clear due to limitation of visualization but our calculations can capture these developments.

**Figure 4.13.** We observe increase in number of component when concentration is decreased to 13% as compared to 15%. We know that as concentration is decreased, more voxels are included in the isoconcentration surface. Increase in number of unconnected components indicates that these new voxels are not in the neighborhood of already drawn isoconcentration surface. These new voxels, which are included, form some sort of island in the space. This is not clear due to limitation of visualization but our calculations can capture these developments.
Figure 4.14. Decrease in Zeroth Betti number on further decrease in concentration threshold as we go from (4) to (5) is for the similar reason as we go from (2) to (3). As more voxels are being included, not only the island voxels are being joined with each other but they are also joined with rest of the isoconcentration surface.

4.5.2 Comparison

Betti number of a dataset, in present case concentration dataset at particular threshold, allows us to study its topology. Topology of a dataset is one of its characteristics and thus can be used to compare different datasets. We use this property to compare and quantify effect of voxel size on concentration profile. It is known that concentration is not the natural outcome of APT experiment but is dependent on user
inputs. Table 4.1 lists the maximum and minimum concentration value obtained for Al at different voxel sizes.

**Table 4.1. Change in estimated concentrations with voxel size**

<table>
<thead>
<tr>
<th>Voxel Size</th>
<th>Minimum Concentration</th>
<th>Maximum Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.7</td>
<td>0.258</td>
<td>36.8</td>
</tr>
<tr>
<td>0.8</td>
<td>1.44</td>
<td>32.8</td>
</tr>
<tr>
<td>0.9</td>
<td>1.84</td>
<td>30</td>
</tr>
<tr>
<td>1.0</td>
<td>3.75</td>
<td>28.3</td>
</tr>
<tr>
<td>1.1</td>
<td>5.19</td>
<td>28</td>
</tr>
<tr>
<td>1.2</td>
<td>5.51</td>
<td>27</td>
</tr>
</tbody>
</table>

We can see that as voxel size increase, range of concentration value which voxels assume decreases. This is expected as higher voxel size leads to averaging of data points.

Fig. 4.15. shows the variation in topology of composition profile with change in voxel size. γ'.

Figure 4.15. Effect of voxel size on morphology of concentration profile. We do not see any change in the pattern of Betti 0 but see changes in number of Betti numbers. This is due to the fact that at lower voxel size, region is divided into higher number of voxel which results in higher number of simply connected components.

For the purpose of comparison only Zeroth Betti number, which gives number of unconnected components, is compared. There are two things to note: (1) Trend in topology is similar at different voxel sizes (2) Number of unconnected components is higher for lower voxel size for non-extreme concentration values.

4.5.3 Feature Identification

Simply connected components are collection of voxels having concentration value greater than or equal to concentration threshold. Thus, of all the possible voxels,
simply connected components define the clusters based on concentration value. This information can be used to limit the concentration threshold value for analyses.

**Figure 4.16.** In maximum separation method, \( d_{\text{max}} \) is chosen as the distance where number of clusters is stable with respect to the distance. In the above example, from [57], 0.6 nm was chosen as the the distance to limit the definition of cluster.

Maximum separation distance method is commonly used to identify clusters of solute atoms in the APT. In that method, as shown in Fig. 4.16, clusters are identified as a function of distance between the atoms and distance threshold is chosen so that resulting cluster numbers is unperturbed by small changes in distance. In this work, where Betti 0 is equal to number of clusters, number of clusters is a function of
concentration. Applying the same logic, we find that concentration threshold for Al in $\gamma'$ phase should be at 17%. Sample we used for our analysis was studied independently using electron microprobe analysis and using Atom Probe Tomography by other group[102]. For their analysis, they reached the value of 19% for Al which is very close to the value obtained from topology based approach presented here.

4.6 Chapter Summary

Application of Topological techniques for APT data analyses and visualization were explained. Value of the technique was shown by testing on a Ni-Al super alloy dataset. Topological data analysis provides a way to explore the relationships in spatial data at different length scales. Relationships thus obtained were then correlated with the properties of precipitates and a quantitative value for concentration threshold was determined for it. This was compared with the values obtained with value obtained from proximity histogram and from cluster analysis method. Utility of the method as a tool to represent the higher dimensional information at multiple length scale in a compact way was established with the use of Betti number graphs for different levels of connectivity at different input values.
CHAPTER 5: SPATIAL UNCERTAINTY IN ISOSURFACE

Atom Probe Tomography is a powerful visualization tool to image internal chemical structure of a material at atomic level resolution. Like all the scientific measurements and calculations, there is some level of uncertainty in APT reconstructed data. In this work, we propose a formalism to quantify and then visualize spatial uncertainty associated with isosurfaces. Sources of uncertainty are identified and mathematical approach to quantify them is developed. High dimensionality of data presents unique challenges in visualization of uncertainty. Graphical capabilities of state of the art software tools are used to represent this uncertainty in an easy to understand manner.

5.1 Introduction

Atom Probe Tomography resolves individual atoms of a bulk specimen in spatial and chemical domain in real space [103]. It works on the principle of field ionization and time of flight mass spectrometry [104]. Specimen is in the form of a very sharp needle [38]. To this specimen, positive voltage is applied and atoms are field evaporated as ions under the influence of electric field. Time elapsed between the evaporation from the specimen to its hit at the detector is used to determine mass to charge ratio. Location where ion hits the detector is used to determine its original location. For the specimen which is not a good conductor of electricity, laser pulse can be used. At the end of the
experiment, instrumental data is subjected to reconstruction algorithms which results in the reconstructed image of specimen. Individual atoms are resolved in this image making it an excellent tool to visualize internal chemical structure of material at an unprecedented resolution [105].

Typical dataset, reconstructed using Atom Probe Tomography data, runs into millions of data points. Each data point corresponds to an atom in the real space. Atoms map is most common method to visualize this information. In atom maps, each atom is represented as a dot on the screen. While this represents the full spatial and chemical information available, it is not very useful. As all the atoms are shown, features of interest such as precipitates or grain boundaries are hidden. Another method to explore internal chemistry is to image the density or concentration fields using isosurfaces. As name suggests, density field tracks the density variation across the sample and concentration field tracks the concentration variation across the sample. Isosurfaces are surface drawn in 3D such that each point on these surfaces has same value. They are 3D analogue of contours. Imaging chemical features using Atom Probe Tomography involves drawing isosurface at appropriate threshold. In this work, we assume that appropriate threshold has been chosen by the methods described in previous chapters. Focus of this chapter is to calculate the computational uncertainty in spatial location of isosurface and to use visualization techniques leading to incorporation of uncertainty information in final image.
5.2 Isosurface as Data

Isosurfaces are visualization of calculated data which fulfills certain criteria. Steps involved in these calculations add some uncertainty to final result. To understand and quantify the uncertainty, it is necessary to first understand how the information evolves during different stages of Atom Probe experiments and subsequent analysis stages.

Atoms in the specimen are evaporated as ions. These ions hit the detector producing electrical signals. These electrical signals are fed as an input to reconstruction algorithm which recreates the specimen image. Density and concentration are resulted of operations on atomic data. In other microscopy techniques, such as optical microscope or transmission electron microscope, imaging mechanism interacts with the density and concentration fields of the material to produce an image which are a reflection of these properties. It is easy to see grain boundaries due to natural contrast produced in the optical micrograph. This is not the case in Atom Probe Tomography. In Atom Probe Tomography, hits detected at the detector are due to a single ion. Reconstructed image is of a single atom. Therefore, density and concentration are not the natural outcome in APT but are calculated quantity. Unlike atomic positions, which can be mapped to a single point in space, density and concentration are calculated for a region in space. When calculated from atomic data, it involves averaging of data. During averaging of data, a point in space represents properties of its neighborhood.
Due to scattered nature of data, there is always some uncertainty in final output. It is this uncertainty which is quantified and visualized in this work.

Isosurface visualization involves following steps: First reconstructed volume is subdivided into smaller regions known as voxels. For a voxel, density is assumed to be constant and calculated at its center. All the atoms within the neighborhood of voxel centers contribute to the net value at the center. Contribution is calculated according to some sampling function which takes into account the distance of atom from the voxel center and extent of contribution. Thus an atom near to voxel center will contribute more than an atom at the edge of the voxel. Density of an element is calculated by dividing number of atoms by the voxel volume. Concentration of an element is calculated by dividing the number of atoms of that element by total number of atoms contained in the volume. Isosurface is drawn by joining voxels which have same value of density or concentration.
Figure 5.1. Schematic of the atoms distributed in the voxel. h is the edge length of the voxel making the voxel volume as $h^3$. $x_i$ are the atomic positions. x is the center of the voxel at which density is calculated.

Figure 5.2. Schematic of density calculation in 1D. Green dots represent the atoms distributed across the space. At each atomic position, a sampling function is fitted, in this case Gaussian. Contribution due to each atom is added up and net value is calculated at the center.
5.3 Error and Uncertainty

Let $D$ be the true data value we are interested in and $Y$ be the measured or calculated value. Then:

$$Y = D + \epsilon$$

(5.1)

Where $\epsilon$ is the difference by which measured value deviates from the true value. Difference can be due to the error or uncertainty or combination of both. It is necessary to distinguish between the error and uncertainty. An uncertainty is a measure of confidence in the value. A highly uncertain value may still be very close to the true value.

There can be two kinds of error: Systematic error and random error. Let us consider a data acquisition process where repeated measurements are made. In case of systematic error, every recorded measurement will deviate from true value in same fashion. For instance, deviation from true value can be constant in subsequent measurement or additive or multiplicative. Common thing in systematic error is mode of deviation from true value will be same in all the measurements. Random error, on the other hand, as name suggests random in nature. Deviation of measured value from true value does not follow any pattern. These can be quantified using statistical
methods. In case of systematic errors, quantification is difficult unless some model is known or if data can be validated from some other source.

5.4 Proposed Approach

Present work is motivated by recent work of Pothkow and Hege [28] on uncertain isocontours. Following the approach proposed in this work, we model the uncertainty in density data calculated from atomic positions. Methodology explained in the work is of general nature for any kind of data. In our case, since we will be working only with Atom Probe Data, we can relax some of the assumptions and also use some properties of the APT data to help in the calculations.

APT data is in 3 dimensions. We assume that density is realizable at all the points of $I$ where $I$ is the domain of specimen volume and $I \subseteq \mathbb{R}^3$. The true values at all positions $x \in I$ is given by the continuous function $h : I \rightarrow \mathbb{R}$. $x$ here represents the position of voxel centers and is available at finite sets of points $x_i \in I$, $i \in \{1,2,...,n\}$. If $g(x_i)$ is the calculated density at some point $x_i$, then

$$g(x_i) = h(x_i) + \epsilon$$ (5.4)
Here $\epsilon$ is the difference between true density and calculated density. This difference can be modeled only if we have some prior information or assume some distribution of it. We model it as a normal distribution centered at the voxel. The normal distribution appears frequently in many natural phenomena, in particular the distribution of measured values of deterministic quantities. Moreover, the calculations involved in density estimation at voxel center make use of Gaussian function to sample the data.

Normal distribution, also called as Gaussian distribution is given by:

$$f(x) = \frac{1}{\sigma \sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{x - \mu}{\sigma}\right)^2\right)$$

(5.5)

Here $\mu$ is the mean and $\sigma$ is the standard deviation. $t$ is the calculated density at the voxel center. On calculating this, we will not only have the density at the voxel center but also the spread of the density around it which will be based on the spatial distribution of atom data points.
Figure 5.3. Schematic of distribution of atoms in space.

Figure 5.4. A rectangular grid is superimposed on the reconstructed space. Net density is calculated at discrete points and depends on the spatial distribution of atoms in the neighborhood. These calculated values are assumed to be constant for given rectangular region. Thus, discrete atomic data, which is available only at selected data points, is converted into continuous density data across the analysis domain.
Figure 5.5. Schematic of a small region with a grid point. Density of the distributed atoms is calculated at the intersection of grid lines. In present method, as shown in (a), density is calculated as the single value defined at a single point. In our proposed approach, we calculate density and model the distribution around the grid point as a normal distribution. Resulting normal distribution is superimposed on the calculated density. Width of Gaussian distribution is influenced by the distribution of atoms around it.

5.4.1 Parameter Calculation for Gaussian Function

Let us consider a single voxel \( v \) and its voxel center \( x \). Let \( a_i \) be the atoms distributed in the voxel with their respective coordinates as: \( (a_i^x, a_i^y, a_i^z) \). Then mean along different axes will be defined as

\[
\mu_x = \frac{\sum_{i=1}^{N} a_i^x}{N}
\]  

(5.6)
And standard deviation will be

\[
\sigma_x = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (a_i^x - \mu_x)^2}
\]  \hspace{1cm} (5.7)

With this information, we can get the full definition of normal distribution at the voxel center and accordingly interpolate across them to get the uncertain isosurfaces.

Let us first take the example of density based isosurfaces or isodensity surfaces. To generate the uncertainty visualization as shown in Fig.5.1, we need three parameters per voxel: Density value, Mean, and Standard Deviation. Mean position is used to decide where the distribution will be centered. In case of Atom Probe Data, it is assumed that distribution is centered at the voxel center.

5.5 Results

We present our approach by applying it on a real dataset. Sample was Ni based superalloy with platinum group metal added. Reconstructed data was divided into voxels of 1 nm volume. For each voxel, net contribution from involved atoms was calculated at the center. Based on spatial position of the atoms, mean and standard deviation was calculated. Isosurfaces were drawn for Al.
In the first step, atoms which contribute to net density for a given voxel were identified based on their distance from voxel center. Next, their contribution at voxel center was calculated by using a Gaussian kernel. Contribution from all the atoms was added to get the net contribution. Other parameters required for normal distribution were also calculated as explained in previous section. Thus a new dataset where required value for each voxel was calculated was obtained as in Table 5.2.

Table 5.1. Reconstructed Dataset

<table>
<thead>
<tr>
<th>X coordinate</th>
<th>Y coordinate</th>
<th>Z coordinate</th>
<th>Chemical label</th>
</tr>
</thead>
<tbody>
<tr>
<td>X₁</td>
<td>Y₁</td>
<td>Z₁</td>
<td>L₁</td>
</tr>
<tr>
<td>X₂</td>
<td>Y₂</td>
<td>Z₂</td>
<td>L₂</td>
</tr>
<tr>
<td>..</td>
<td>..</td>
<td>..</td>
<td>..</td>
</tr>
<tr>
<td>Xₙ</td>
<td>Yₙ</td>
<td>Zₙ</td>
<td>Lₙ</td>
</tr>
</tbody>
</table>
### Table 5.2. Voxel Dataset

<table>
<thead>
<tr>
<th>Voxel Coordinate</th>
<th>Voxel Coordinate</th>
<th>Voxel Coordinate</th>
<th>Al density</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>Y</td>
<td>Z</td>
<td>D</td>
<td></td>
<td></td>
</tr>
<tr>
<td>X_1</td>
<td>Y_1</td>
<td>Z_1</td>
<td>D_1</td>
<td>μ_1</td>
<td>σ_x, σ_y, σ_z</td>
</tr>
<tr>
<td>X_2</td>
<td>Y_2</td>
<td>Z_2</td>
<td>D_2</td>
<td>μ_2</td>
<td>σ_x, σ_y, σ_z</td>
</tr>
<tr>
<td>..</td>
<td>..</td>
<td>..</td>
<td>..</td>
<td>..</td>
<td></td>
</tr>
<tr>
<td>X_n</td>
<td>Y_n</td>
<td>Z_n</td>
<td>D_n</td>
<td>μ_n</td>
<td>σ_x, σ_y, σ_z</td>
</tr>
</tbody>
</table>

Visualization software such as PARAVIEW and VISIt were used. Voxel data was fed to PARAVIEW to convert it into structured grid format which can be read by VISIt. VISIt was then used to visualize the data. In first step, crisp isosurfaces were drawn. In second step uncertainty information was added by assigning a shaded region around it. Intensity of shade was dependent on uncertainty value. Fig. 2. shows the steps involved in generating uncertainty data and its visualization.
Figure 5.6. Sequence of operations to generate uncertainty visualization in APT data.

Uncertainty is a measure of confidence in the result. It is user and application dependent as for some applications, a result with large uncertainty can be acceptable. Once we have information about the result and its spread, the question of how much of that uncertainty to include in the final visualization is to be taken by the end user. For instance, here we have used the value of +/- 1%. Other measures can be Full width at half maximum of the Gaussian which is used to model uncertainty. This will be given by $2\sqrt{2\ln2} \sigma$. If more uncertainty is to be included, then Full width at tenth of maximum can be used. Since in our calculation we already have calculated these values, including these in visualization is straightforward.
Figure 5.7. Isosurfaces drawn using conventional calculations. To each voxel, a value is assigned which is assumed to be constant throughout the voxel. To draw isosurfaces, all the voxels which have value equal to the threshold value are joined. This results in the ‘crisp isosurface’. These give an indication that the surfaces are well defined in space with no uncertainty in their position.
Figure 5.8. Isosurfaces drawn with uncertainty information. Red color is the value at 12%. Blue color indicates the two sigma probability of value of 11%. It can be seen that different isosurfaces have different levels and spatial extent of uncertainty associated with them.
Figure 5.9. One isosurface is isolated and shown for explanation. Threshold at the surface is 12%. All those voxels where value was equal to 12% were joined as per the marching cube algorithm. Algorithm assumes a constant value across the voxel and therefore resultant isosurface is crisp giving an indication that its shape is well defined in the space without any uncertainty.

Figure 5.10. Isoconcentration surface as in previous figure. In this representation, uncertainty information is added. Intensity of the color maps to the heat map at the left. Isosurface threshold is 12%. Uncertainty is function of spatial distribution of atoms. Distribution of atoms is less at far off and therefore no affect is seen. Near the surface, uncertainty in the definition of isosurface decreases. In the image, it is seen as increased level of intensity as one approaches the surface. Values are calculated assuming 2 times of variance in the normal distribution of data.
Figure 5.11. Isoconcentration surface at 12%. In this image, representation of regions where uncertainty leads to higher value than the threshold value is shown. Isosurfaces are drawn such that they enclose higher value regions inside them. For instance, interior of 12% isosurface will include region where concentration value is more than 12%. Therefore there is some region outside the isosurface where value, within some uncertainty, can reach higher than the threshold value is new information revealed through this visualization.

5.6 Chapter Summary

Uncertainty in final results is always present in scientific data. Reducing uncertainty requires extra information which may not be always available. Quantification and subsequent representation of uncertainty is therefore important to convey complete information. Isosurface visualization of Atom Probe Tomography data is a very common and useful tool to image chemical features. In this work, we have developed a formalism to quantify and visualize uncertainty associated with its calculation. Its calculation involves averaging of discrete spatial data to generate density data in continuous domain. Uncertainty arising due to this is quantified.
Visualization of uncertainty in 3D poses its own challenges as available degrees of freedom are reduced due to high dimensionality of data. Conventional features such as error bars cannot be used. We have used shading to convey uncertain information. For ease of use, it is not only necessary to render the data, but user should also be able to navigate through the data at a reasonable rate. Huge size of APT data makes this difficult for conventional engineering software. Therefore we have made use of graphical software such as PARAVIEW and VISit.

Future work in this direction will involve inclusion of uncertainty due to experiments. More work is required to be done to understand how the equipment interacts with the field evaporated ions. For now, we have quantified the uncertainty due to mathematical operations involved in generating density and/or concentration information from atomic data and have used state of the art graphical software to visualize the same.
CHAPTER 6: SUMMARY AND FUTURE WORK

6.1 Summary

Primary objective of this thesis was to develop data driven techniques to obtain images of chemical features which are quantitatively meaningful. Two parameters were found to affect the final image: Voxel size and concentration threshold. In this thesis, we have identified the limitations of present approaches and laid down detailed steps to achieve the objective. To select voxel size, we have laid down the procedure to minimize the error between the estimated density and detected data points. To select the threshold for isosurfaces, we have made use of expected topology of the features and mapped it to topology of the dataset at appropriate connectivity. Uncertainty is an inherent part of engineering data and it is necessary to represent it so that user has complete picture. We have identified the sources of uncertainty in isosurfaces and developed a framework to quantify and visualize it.

While this thesis has focused on imaging data, it can generalize to any form of data. When exploring extremely large and diverse data spaces in materials modeling and simulation, a fundamental question that is often overlooked is whether we have the necessary data beyond our original assumptions and if so, where in the data landscape does it exist? Hence, we can based on this thesis develop computational
methods to search for “holes” in data sets that cannot be detected readily when we are dealing with an n-dimensional data space. In low dimensions, it is sometimes possible to visually infer necessary information. As dimensionality of data is increased, this approach becomes increasingly difficult.

Regions where there are holes, cavities, and other features of missing data must be defined for two reasons: first, to address the data regions where calculations should be performed to fill in the data and second, to determine the regions that should be used to build the informatics models. To determine where these regions are present, we will apply computational homology to quantitatively characterize the connectivity of data points without explicitly defining the Euclidean distance between points. Computational homology provides a framework to assess the discrete data in n-dimensions. We shall be taking advantage of recent developments in the field which makes the computation of regions where there is gap in data very efficient in n-dimensions. One can for example compute Betti Numbers of data points at different resolutions to extract information about gaps in the data. Other regions of interest include regions where functions assume maximal or minimal value. Based on these computational homology calculations, informatics modeling can be performed in both
the high density and the low density region of data to demonstrate the importance to
define data connectivity in high dimensional space.
APPENDIX: CODES

A1: Voxel Size Selection

% Author: Kaustubh Kaluskar

% dtst = n X 3 matrix of data points

%%%%%%%%%% Voxel Size

%%%%%%%%%%

vx = 0.3;
vy = 0.3;
vz = 0.3;

sigma_g = vx*vx/(8*log(2)); % global sigma for individual data point Gaussian

% Voxel numbers determination

a = min(dtst); % a1 contains 3 elements which are minimum value along X,Y, Z
z = max(dtst); % z1 contains 3 elements which are maximum value along X,Y, Z
dx = z(1) - a(1); % Length of dataset along X axis
dy = z(2) - a(2); % Length of dataset along Y axis
\( \text{dz} = z(3) - a(3); \) % Length of dataset along Z axis

\( \text{v\_x} = \text{ceil(dx/vx)}; \) % number of voxels along x axis

\( \text{v\_y} = \text{ceil(dy/vy)}; \) % number of voxels along y axis

\( \text{v\_z} = \text{ceil(dz/vz)}; \) % number of voxels along z axis

\( \text{vL} = \text{zeros(v\_x*v\_y*v\_z,12);} \) % Matrix to store values per voxels

% Voxel number determination ends

\( [m,n] = \text{size(dtst);} \) % m == number of data points. n == 3

% Calculating Mean of data points in each voxel

\( \text{for i} = 1:1:m \)

\( \text{x} = \text{dtst}(i,1) - a(1); \)

\( \text{y} = \text{dtst}(i,2) - a(2); \)

\( \text{z} = \text{dtst}(i,3) - a(3); \)

\( \text{v\_i} = \text{floor(z/vz)*v\_x*v\_y + floor(y/vy)*v\_x + floor(x/vx) + 1;} \) % voxel index in which atom falls

\( \text{vL(v\_i,1)} = \text{vL(v\_i,1)} + 1; \) % Number of atom counts in the voxel

\( \text{vL(v\_i,2)} = \text{vL(v\_i,2)} + x; \) % Adding x coordinate for mean value calculation

\( \text{vL(v\_i,3)} = \text{vL(v\_i,3)} + y; \) % Adding y coordinate for mean value calculation

\( \text{vL(v\_i,4)} = \text{vL(v\_i,4)} + z; \) % Adding z coordinate for mean value calculation
for i = 1:1:v_x*v_y*v_z
    vL(i,5) = vL(i,2)/vL(i,1); % mean along x axis
    vL(i,6) = vL(i,3)/vL(i,1); % mean along y axis
    vL(i,7) = vL(i,4)/vL(i,1); % mean along z axis
end

% Mean Calculation ends

% Variance Calculation
for i = 1:1:m
    x = dtst(i,1) - a(1);
    y = dtst(i,2) - a(2);
    z = dtst(i,3) - a(3);
    v_i = floor(z/vz)*v_x*v_y + floor(y/vy)*v_x + floor(x/vx) + 1;
    vL(v_i,8) = vL(v_i,8) + (x - vL(v_i,5))^2;
    vL(v_i,9) = vL(v_i,9) + (y - vL(v_i,6))^2;
    vL(v_i,10) = vL(v_i,10) + (z - vL(v_i,7))^2;
end
for i = 1:1:v_x*v_y*v_z

vL(i,8) = sqrt(vL(i,8)/(vL(i,1)-1)); % sigma along x axis
vL(i,9) = sqrt(vL(i,9)/(vL(i,1)-1)); % sigma along y axis
vL(i,10) = sqrt(vL(i,10)/(vL(i,1)-1)); % sigma along z axis
end

% Variance Calculation end

% Density Calculation

for i = 1:1:m

x = dtst(i,1) - a(1);

y = dtst(i,2) - a(2);

z = dtst(i,3) - a(3);

v_i = floor(z/vz)*v_x*v_y + floor(y/vy)*v_x + floor(x/vx) + 1;

vxx = floor(x/vx) + vx/2; % Voxel center x coordinate
vyy = floor(y/vy) + vy/2; % Voxel center y coordinate
vzz = floor(z/vz) + vz/2; % Voxel center z coordinate

distance = (vxx - x)^2 + (vyy - y)^2 + (vzz - z)^2;
\[ v_{L(v_i,11)} = v_{L(v_i,11)} + \exp(-\text{distance}/(2*\sigma_g)); \]

end

% Density Calculation ends

% Error Calculation

for i = 1:1:m

\[ x = \text{dtst}(i,1) - a(1); \]
\[ y = \text{dtst}(i,2) - a(2); \]
\[ z = \text{dtst}(i,3) - a(3); \]

\[ v_i = \text{floor}(z/v_z)*v_x*v_y + \text{floor}(y/v_y)*v_x + \text{floor}(x/v_x) + 1; \]

\[ v_{xx} = \text{floor}(x/v_x) + v_x/2; \]
\[ v_{yy} = \text{floor}(y/v_y) + v_y/2; \]
\[ v_{zz} = \text{floor}(z/v_z) + v_z/2; \]
amplitude_c = vL(v_i,11); % Density amplitude at the voxel center

%amplitude_m = vL(v_i,15); % Density amplitude at the Mean position of data

sigma_x = vL(v_i,8); % Variance of the total voxel data along X axis

sigma_y = vL(v_i,9); % Variance of the total voxel data along y axis

sigma_z = vL(v_i,10); % Variance of the total voxel data along z axis

tr1_c = ((vxx-x)/(sigma_x))^2;

tr2_c = ((vyy-y)/(sigma_y))^2;

tr3_c = ((vzz-z)/(sigma_z))^2;

den_c = amplitude_c*exp(-0.5*(tr1_c+tr2_c+tr3_c));

vL(v_i,12) = vL(v_i,12) + (den_c - 1)^2; % Error calculation

end

vL(isnan(vL)) = 0;
A2: Uncertainty Code

[m,n] = size(uc);

limit_max = max(uc);
limit_min = min(uc);

x_range = limit_max(1) - limit_min(1);
y_range = limit_max(2) - limit_min(2);
z_range = limit_max(3) - limit_min(3);
c_range = limit_max(4) - limit_min(4) + 1;

vxl_size = 1;

n_vxl_x = ceil(x_range/vxl_size); % number of voxels along x direction
n_vxl_y = ceil(y_range/vxl_size); % number of voxels along y direction
n_vxl_z = ceil(z_range/vxl_size); % number of voxels along z direction
n_vxl = n_vxl_x*n_vxl_y*n_vxl_z; % total number of voxels
vxl_data = zeros(n_vxl,c_range+3); % matrix containing voxel center coordinate in first three columns and net contribution of each element in respective columns

dtst_num = zeros(n_vxl, c_range); % integer number of atoms in each voxel

dtst_var = zeros(n_vxl, c_range); % Variance of each element in each voxel

% This part of code assigns voxel coordinates to voxles

c = 1;

for i = 0:1:n_vxl_x-1
    for j = 0:1:n_vxl_y-1
        for k = 0:1:n_vxl_z-1
            vxl_data(c,1) = limit_min(1) + i*vxl_size+0.5;
            vxl_data(c,2) = limit_min(2) + j*vxl_size+0.5;
            vxl_data(c,3) = limit_min(3) + k*vxl_size+0.5;
            c = c+1;
        end
    end
end

% Ends
% Contribution of atoms to voxel centers

for i = 1:1:m

    x = uc(i,1) - limit_min(1);
    y = uc(i,2) - limit_min(2);
    z = uc(i,3) - limit_min(3);
    chem_id = uc(i,4) + 3;

    vxl_index = floor(z/vxl_size)*n_vxl_x*n_vxl_y + floor(y/vxl_size)*n_vxl_x + floor(x/vxl_size) + 1; % voxel index in which atom falls

    x_vxl = vxl_data(vxl_index,1); % X coordinate of Center of the voxel
    y_vxl = vxl_data(vxl_index,2); % Y coordinate of Center of the voxel
    z_vxl = vxl_data(vxl_index,3); % Z coordinate of Center of the voxel

    contri = gauss_contri(uc(i,1),uc(i,2),uc(i,3),x_vxl,y_vxl,z_vxl); % How much an atom contributes at the voxel center
vxl_data(vxl_index,chem_id) = vxl_data(vxl_index,chem_id) + contri;%adding that contribution to get the net contribution

chem_id = chem_id - 3;

var = var_calc(uc(i,1),uc(i,2),uc(i,3),x_vxl,y_vxl,z_vxl);

dtst_var(vxl_index,chem_id) = dtst_var(vxl_index, chem_id)+var;

dtst_num(vxl_index,chem_id) = dtst_num(vxl_index, chem_id)+1;

end

dtst_var_c = dtst_var./dtst_num;

dtst_sd = sqrt(dtst_var./dtst_num);

dtst_sd(isnan(dtst_sd)) = 0 ;

dlmwrite('dtst_sd.dat',dtst_sd);

dlmwrite('vxl_data.dat',vxl_data);
REFERENCES


