

**Development of Structure property linkage with a machine
learning approach using experimental data**

Thesis submitted to

Indian Institute of Science, Bangalore

in partial fulfilment for the award of the degree of

Master of Science(Research)

in

Materials Science
by

Mohammed Abdul Wadood

(11-01-04-20-92-19-1-17679)

Under the supervision of

Dr. Abhik Choudhury



UG Programme

Indian Institute of Science, Bangalore

Academic Year, 2019-20

June, 2020

DECLARATION

I certify that

- (a) The work contained in this report has been done by me under the guidance of my supervisor.
- (b) The work has not been submitted to any other Institute for any degree or diploma.
- (c) I have conformed to the norms and guidelines given in the Ethical Code of Conduct of the Institute.
- (d) Whenever I have used materials (data, theoretical analysis, figures, and text) from other sources, I have given due credit to them by citing them in the text of the thesis and giving their details in the references.

Date: June 25, 2020

Place: Bangalore

(Mohammed Abdul Wadood)

(11-01-04-20-92-19-1-17679)

UG PROGRAMME**INDIAN INSTITUTE OF SCIENCE, BANGALORE****BANGALORE - 560012, INDIA****Certificate**

This is to certify that the project report entitled “Development of Structure property linkage with a machine learning approach using experimental data” submitted by Mohammed abdul Wadood (Sr No. 11-01-04- 20-92-19-1-17679) to Indian Institute of Science, Bangalore towards partial fulfilment of requirements for the award of degree of Master of Science(Research) in Materials Science is a record of bona fide work carried out by him under my supervision and guidance during Academic Year, 2019-20.

Date: June 25, 2020

Place: Bangalore

Dr. Abhik Choudhury
Materials Engineering
Indian Institute of Science, Bangalore
Bangalore - 560012, India

Abstract

Microstructure of a material is a result of the composition and processing parameters it had gone through. This microstructure, which represents the spatial arrangement of phases, predominantly dictates the macroscopic properties of the material. In eutectic solidification, the microstructures consist of various patterns like lamella, rods and labyrinths in binary alloys. However, as the number of phases and components increases, the number of possible patterns that might be obtained during bulk solidification also become larger. This incentivizes the need for an optimized method to develop a relationship between the microstructure and processing.

In this study, we develop a method to obtain a linkage between the processing and microstructure of a material. This is done using the various microstructures obtained at different growth rates of directionally solidified AgCuSb.

Contents

1 Introduction	1
2. Literature Review	2
2.1 Ternary eutectic solidification	2
2.2 Quantification and reconstruction of microstructures	3
2.2.1 Quantification using 2-point statistics.	3
2.2.2 Reconstructing 2 point data.	8
2.3 Microstructure variables and processing parameters linkage	9
2.3.1 Support Vector Regression	9
3. Methods	12
3.1 Experimental Details	12
3.1.1 Alloy preparation	12
3.1.2 Directional solidification	13
3.1.3 Sample preparation for imaging	14
3.2 Image Preprocessing	15
4. Results	17
5.Discussion	19
6.Work Planned	20
7.Problems faced	20
8.Conclusions and future work	20
8.1 Reconstruction of microstructure from 2 pt statistics	21

8.2 Alternative method: Quantification, dimensionality reduction and reconstruction as a single step	21
9. References	22

List of figures

Figure 1: The soft margin loss setting for linear SVM [17]	11
Figure 2: Solidified alloy after induction melting(left). Crushing of the alloy (right)	13
Figure 3. Modified Bridgman apparatus [20]	14
Figure 4. Transverse sectioning of directionally solidified sample	14
Figure 5. Sample preparation for imaging(optical for preliminary investigation)	15
Figure 6. Image of eutectic composition at 1micron/s obtained from SEM, with the magnification details cropped out	15
Figure 7. The image shown in figure 6 after thresholding(left) and bucket filling(right).	16
Figure 8. (a)microstructure obtained at growth rate of $16\mu\text{s}$. (b) cross correlation of the sb phase with the Ag ₃ Sb phase. (c) Autocorrelation of the Sb phase	17
Figure 9. Best fit line using support vector regression.	18
Figure 10. Figure 8.1 Proposed workflow of microstructure encoding and reconstruction according to Li, Xiaolin,2018[21].	20

List of Tables

Table 1 The PC coefficients for different growth rates	18
--	----

Acknowledgements

Undertaking this project has been an immense learning experience for me and it would not have been possible without the support and guidance that I received from many people. I take this opportunity to extend my sincere gratitude and appreciation to all those who made this possible.

Foremost, I would like to acknowledge my indebtedness to my supervisor - Dr. Abhik Choudhury, who made this work possible. His friendly guidance and expert advice has been invaluable throughout all stages of the work. I would also like to thank A.S. Kiran for guiding and helping me throughout the course of my project.

I also thank the UG Programme and the Dept. of Materials Engineering for the facilities provided.

The members of our lab group have contributed immensely to my time during the project. The group has been a source of friendships as well as good advice and discussions. Thanks to Sumit, Fiyanshu, Bikram.

I have been fortunate to come across many good friends, without whom life would be bleak, and Shirin deserves a special mention here. Thanks to Sarthak, Pratitee, Teja, Jkub.

I feel a deep sense of gratitude for my parents and sibling without whose unconditional love, emotional and psychological support, this would not have been possible.

1 Introduction

Obtaining microstructure images experimentally is an expensive affair, it's costs arising from the labour put in, the equipment used and the resources consumed. Same goes for simulation methods, whose costliest aspects are the time and computational resources it takes. In this study, we aim to establish a linkage between the processing parameters (composition and solidification speed) and resultant microstructure, such that, the microstructure for a random combination of processing parameters can be predicted. Development of such techniques will help us obtain a fundamental understanding of principle mechanisms of pattern formation in alloys which are of general importance for the development of multicomponent alloys for advanced technical applications.

The fundamental understanding of pattern formation for binary eutectics was established by Jackson and Hunt ^[1], who proposed when lamellar and rod eutectics are formed, the correlation between eutectic spacing and growth rate is expressed as $\lambda^2 v = k$ (K is a constant). Ternary eutectic alloys have received increasing attention in recent years owing to their more complex structures and thus different physical properties as compared with those of binary alloys ^[2]. The simultaneous growth of three distinct phases from the liquid allows for a broad variety of different phase arrangements far beyond the typical microstructures of rods, lamellae and labyrinths in binary eutectics ^[3]. It has been summarized that as solidification condition and/or alloy system vary, such eutectic morphologies as three lamellar-phase structures, one fibrous- and two lamellar-phase structures, two fibrous- and one lamellar-phase structures, two fibrous phase in continuous matrix structure, three fibrous-phase structures and so on can be formed in ternary eutectic alloys ^[4,5].

In the present work, ternary $Ag_{42.4}Cu_{21.6}Sb_{36}$ eutectic alloy – whose solidification process involves the cooperative growth of two intermetallic compounds $\theta(Cu_2Sb)$ and $\epsilon(Ag_3Sb)$ and one solid solution (Sb) phase, is studied for its variation in microstructure as a function of solidification speed/growth rate. Additionally, two off eutectic compositions are studied for the same variable too.

The following sections cover the literature used to proceed with work, the methods used to obtain the results, the results, discussion and conclusion respectively.

2. Literature Review

The problem at hand, where we attempt to predict the 2 point statistics of a microstructure undergoing a given set of processing conditions, involves the following steps. Firstly, since we are using supervised learning algorithms, the procurement of microstructure data whose processing conditions are known. Then, a method to obtain the 2 point statistics of a given microstructure. After which , we can make use of a regression algorithm to predict the 2 point statistics for any given processing conditions .

In the following subsections, the theory and present literature used to solve the problem at hand are presented.

2.1 Ternary eutectic solidification

Basic theories of solidification and segregation suggest that the quality of grown crystals is primarily controlled by the crystal growth rate, the growth interface morphology, and the solute redistribution ^[6].In directional solidification, the advancement of the growth interface into the melt(the growth rate) is governed by the balance of the heat fluxes at the interface.

In this present work, the microstructure patterns of directionally solidified AgCuSb is studied at growth rates 0.5, 1, 2, 4, 8, 16, 32, 64 $\mu\text{m/s}$. The eutectic composition of this alloy is $\text{Ag}_{42.4}\text{Cu}_{21.6}\text{Sb}_{36}$. ^[7]

The variation in microstructures obtained at different growth rates was shown by Wei Zhai et al, 2015 ^[7]. This suggests that the chosen alloy would be appropriate for the problem at hand.

2.2 Quantification and reconstruction of microstructures

To form a linkage between the processing parameters and the resultant microstructure, the first step in the analysis would be to represent both the parameters and the microstructures in a suitable format. Only after this is achieved, can the process of linking the variables be done.

2.2.1 Quantification using 2-point statistics.

The goal here is to utilize the periodic properties of microstructures. Microstructure when viewed as an image has variables equivalent to the number of pixels in the image. However, this isn't the most efficient method to represent a microstructure and the dimensionality of this data can be reduced drastically. Two-point statistics (i.e., 2-point spatial correlations) are the most basic of the n-point spatial correlations that can be used to capture the details of the microstructure ^[8].

Microstructure Discretization[9]

The first step in all of MKS workflows is to discretize microstructures. To do this, we create a probabilistic description of a microstructure by introducing the continuous local state variable h , the local state space H and the microstructure function $m(h,x)$. The local state space H can be thought of as all of the thermodynamic state variables that

are needed to uniquely define the material structure at a given location. The local state variable h is one instance of the local state space, or one configuration of state variables. The microstructure function $m(h,x)$ is a probability density function of finding a local state h at location x . For instance, let $\mu(x)$ be a microstructure that we plan to discretize, then μ is the expectation of the microstructure function.

$$\mu(x) = \int_H h m(h, x) dh$$

Now, we will discretize the microstructure in space by averaging over small cubic domains in the microstructure function. The local state can be discretized, using two methods: one is to bin the microstructure, using the primitive basis Λ

$$\frac{1}{\Delta x} \int_H \int_s \Lambda(h-l) m(h, x) dx dh = m[l, s]$$

the other is to use a spectral representation using some orthogonal basis function ξ

$$\frac{1}{\Delta x} \int_s m(h, x) dx dt = \sum_{l=0}^{L-1} m[l, s] \xi_l(h)$$

In the notation above, all of the round brackets variables are continuous variables and the square brackets variables are the discrete variables. The variables s and S represent a discrete position and the total volume, while l and L represent the discrete versions of h and H . In PyMKS the Legendre polynomials are currently the only orthogonal basis functions available. Either of these two discretization methods are used to discretize the microstructure. In the current work, legendre basis was used.

n-Point Spatial Correlations [9]

1-Point Spatial Correlations (or 1-point statistics)

N-point spatial correlations provide a way to rigorously quantify material structure, using statistics. As an introduction to n-point spatial correlations, let's first discuss 1-point statistics. 1-point statistics are the probability that a specified local state will be found in any randomly selected spatial bin in a microstructure. 1-point statistics compute the

volume fractions of the local states in the microstructure. 1-point statistics are computed as

$$f[l] = \frac{1}{S} \sum_s m[s, l]$$

In this equation, $f[l]$ is the probability of finding the local state l in any randomly selected spatial bin in the microstructure, $m[s, l]$ is the microstructure function (the digital representation of the microstructure), S is the total number of spatial bins in the microstructure and s refers to a specific spatial bin.

While 1-point statistics provide information on the relative amounts of the different local states, they do not provide any information about how those local states are spatially arranged in the microstructure. Therefore, 1-point statistics are a limited set of metrics to describe the structure of materials.

2-Point Spatial Correlations

2-point spatial correlations (also known as 2-point statistics) contain information about the fractions of local states as well as the first order information on how the different local states are distributed in the microstructure.

2-point statistics can be thought of as the probability of having a vector placed randomly in the microstructure and having one end of the vector be on one specified local state and the other end on another specified local state. This vector could have any length or orientation that the discrete microstructure allows. The equation for 2-point statistics can be found below.

$$f[r|l, l'] = \frac{1}{S} m[s, l] m[s + r, l']$$

In this equation, $f[r|l, l']$ is the conditional probability of finding the local states l and l' at a distance and orientation away from each other defined by the vector r . All other variables are the same as those in the 1-point statistics equation. In the case that we have an eigen microstructure function (it only contains values of 0 or 1) and we are using an indicator basis, the $r = 0$ vector will recover the 1-point statistics.

When the 2 local states are the same, $l = l'$, it is referred to as an autocorrelation. If the 2 local states are not the same, it is referred to as a cross-correlation.

In 2-pt statistics, we find the conditional probabilities for each possible vector. However, this produces a very large set of descriptors for the material structure. Indeed, the number of descriptors is equal to the number of distinct discretized vectors included in the analyses. For example, if one were to define a 21×21 neighborhood (which corresponds to including only ten voxels on each side of the voxel of interest) in a 2-D microstructure (or a cross-section of a 3D microstructure), one ends up with 441 descriptors of the material structure. Clearly, this number of material structure descriptors will be unwieldy for any subsequent analyses (e.g., extraction of process-structure-property linkages).

Therefore, a dimensionality reduction technique should be used which preserves the most important/significant descriptors. In prior work ^[10], principal component analysis (PCA) has been demonstrated to provide objective low dimensional representation of the 2-point statistics of the material microstructure.

The eutectic alloys explored in the present study exhibit three distinct phases in their evolution histories. The total number of 2-point spatial correlation functions that can be defined in a three-phase composite is nine. This is the number of different choices for the ordered pair (h, h') , since each index can take three different values. These nine correlation functions consist of three autocorrelations (i.e., $h=h'$) and six cross correlations (i.e., $h \neq h'$). Niezgoda et al. ^[11] have shown that at most, only two of these correlation functions will be independent. ^[12]

In the present study, we have selected an autocorrelation of the Sb phase and its cross correlation with Ag₃Sb to represent the set of independent 2-point statistics. This choice was made because the Sb phase shows the highest variance and Ag₃Sb exhibits the highest values of the volume fraction in the entire ensemble of the simulation results produced for this study.

PCA [13]

PCA is a dimensionality reduction technique which basically maps/transforms data linearly such that the covariance matrix of the transformed data is a diagonal matrix, i.e, it has zero covariance among dimensions and non zero variance. Post this step, the dimensions with the lowest variance can be discarded. To minimize the error, k such dimensions can be selected. The steps taken to achieve this are described below.

Let $x_1, x_2, \dots, x_m \in R^n$ be m data points and let X be a matrix such that

x_1, x_2, \dots, x_m are the rows of this matrix. Assuming standardized data, i.e, data with zero mean and unit variance.

Let p_1, p_2, \dots, p_n be a set of such n linearly independent orthonormal vectors. Let P be a $n \times n$ matrix such that p_1, p_2, \dots, p_n are the columns of P .

Then,

$$\hat{X} = XP$$

Where \hat{X} is the matrix of transformed points.

And the original data points, x_i , can be represented as

$$x_i = \alpha_{i1}p_1 + \alpha_{i2}p_2 + \alpha_{i3}p_3 + \dots + \alpha_{in}p_n$$

Where, for an orthonormal basis, α_{ij} is

$$\alpha_{ij} = x_i^T \cdot p_j$$

Now, we want the covariance matrix of the transformed data to have specific properties, i.e,

$$C_{ij} = 0 \text{ when } i \neq j \text{ (covariance = 0)}$$

$$C_{ii} \neq 0 \text{ when } i = j \text{ (variance } \neq 0)$$

From our previous assumption of X being a standardized matrix, the covariance matrix of our transformed data, \hat{X} , is given by

$$\frac{1}{m} \hat{X}^T \hat{X}$$

This follows since \hat{X} inherits the standardized property of X and

$$\begin{aligned} C_{ij} &= \frac{1}{m} \sum_{k=1}^m (X_{ki} - \mu_i)(X_{kj} - \mu_j) \\ &= \frac{1}{m} \sum_{k=1}^m X_{ki} X_{kj} \\ &= \frac{1}{m} X_i^T X_j = \frac{1}{m} (X^T X)_{ij} \end{aligned}$$

We can write the covariance matrix of the transformed data as follows,

$$\frac{1}{m} \hat{X}^T \hat{X} = \frac{1}{m} (XP)^T XP = \frac{1}{m} P^T X^T XP = P^T (\frac{1}{m} X^T X) P = P^T \Sigma P$$

From the desired properties mentioned above it follows that,

$$\frac{1}{m} \hat{X}^T \hat{X} = P^T \Sigma P = D \quad [\text{where } D \text{ is a diagonal matrix}]$$

This is achieved when the column of matrix p are eigenvectors of

$$\Sigma = X^T X \quad (\text{By singular value decomposition})$$

Thus, the new basis P used to transform X is the basis consisting of the Eigen Vectors of $X^T X$

2.2.2 Reconstructing 2 point data.

We make use of machine learning algorithms to find a relation between the transformed data points and the processing parameters. The next step would then be to reconstruct the original data from the predicted transformed data. However, there will be an error during reconstruction. It arises from the discarding of the non dominant eigenvectors.

The error will be equivalent to the variance captured by the selected eigen vectors subtracted from unity.

To transform the data back to its original space, we can multiply it with the inverse of the initial transformation matrix. However, since the initial transformation matrix is an orthonormal matrix, the inverse is equal to transpose.

Therefore, the retransformed data, X' , when the matrix of the chosen PC vectors is V is,

$$X' = \hat{X} V^T$$

Where, \hat{X} is the transformed data. Therefore,

$$X' = X V V^T$$

2.3 Microstructure variables and processing parameters linkage

2.3.1 Support Vector Regression

To predict the PCs from the processing parameter variables, various regression algorithms can be used. In this study, Support vector regression was used. The Support Vector Regression (SVR) uses the same principles as the SVM for classification, with only a few minor differences. Like classification, in the case of regression, a margin of tolerance (epsilon) is set in approximation to the SVM which would have already requested from the problem. The tacit assumption here is that such a function f actually exists that approximates all pairs (x, y) with ϵ precision, or in other words, that the convex optimization problem is feasible. Sometimes, however, this may not be the case, or we also may want to allow for some errors and therefore the algorithm could be more complicated than SVM. However, the main idea is always the same: to minimize error, individualizing the hyperplane which maximizes the margin, keeping in mind that part of the error is tolerated.

Sometimes, however, this may not be the case, or we also may want to allow for some errors. Analogously to the “soft margin” loss function ^[14] which was adapted to SV machines by Cortes and Vapnik ^[14], one can introduce slack variables ξ_i, ξ_i^* to cope with otherwise in-feasible constraints of the optimization problem . Hence we arrive at the formulation stated in ^[16].

$$\text{minimize} \quad \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*)$$

$$\begin{aligned} \text{Subject to, } & y_i - \langle w, x_i \rangle - b \leq \varepsilon + \xi_i \\ & \langle w, x_i \rangle + b - y_i \leq \varepsilon + \xi_i^* \\ & \xi_i, \xi_i^* \geq 0 \end{aligned}$$

The constant $C > 0$ determines the trade-off between the flat-ness of f and the amount up to which deviations larger than ε are tolerated. This corresponds to dealing with a so called ε -insensitive loss function $|\xi| \varepsilon$ described by

$$\begin{aligned} |\xi|_\varepsilon &:= 0 \quad \text{if } |\xi| \leq \varepsilon \\ &:= |\xi| - \varepsilon \quad \text{otherwise} \end{aligned}$$

Fig.1 depicts the situation graphically. Only the points out-side the shaded region contribute to the cost insofar, as the deviations are penalized in a linear fashion.

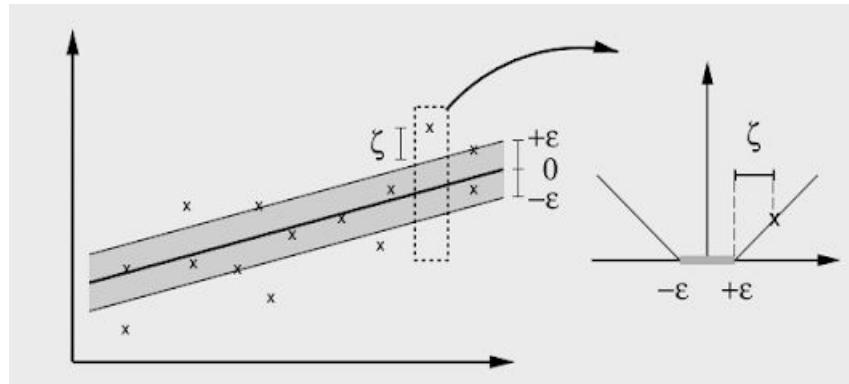


Figure 1 The soft margin loss setting for linear SVM ^[17]

At its core, this is similar to linear regression algorithm. The next step is to make the SV algorithm nonlinear. This, for instance, could be achieved by simply preprocessing the training patterns x_i by a map $\Phi : X \rightarrow F$ into some feature space F , as described in [18] [19] and then applying the standard SV regression algorithm.

In this work, the kernel used to transform the data into higher dimension space to make the linear separation possible was the gaussian radial basis function.

$$k(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right)$$

3. Methods

3.1 Experimental Details

3.1.1 Alloy preparation

Ag42.4Cu21.6Sb36 is the eutectic composition for the Ag-Cu-Sb alloy. This composition was based on the eutectic composition mentioned in [7]. Preliminary experiments were conducted to verify this. For the preparation of the alloy, high purity, i.e., 99.999% pure

Ag, Cu and Sb were used. Once these were weighed and taken in the right proportion, they were filled in quartz tubes of 4mm inner diameter(6mm outer diameter).Next, they were sealed at 10-5 mbar pressure using a rotary and then a diffusion pump subsequently.

Induction melting was used to melt all the components of the alloy thoroughly and then the melt was left to solidify and cool at room temperature. Once the sample reached ambient temperature, it was taken out of the quartz tube to be cut and crushed into chunks of size that fit into the quartz tube for the directional solidification step. Once the quartz tube was filled, the steps to vacuum seal were repeated.



Figure 2. Solidified alloy after induction melting(left). Crushing of the alloy (right).

3.1.2 Directional solidification

A modified bridgman solidification apparatus was used for these set of experiments. Here, the furnace had three zones in order to have good control over the thermal gradient. These were separated by thin ceramic based insulation zones. This established a gradient between the hot and the cold zones. Kanthal-A was selected as

a heating element for the isothermal heating zone, while the chill zone has a water circulating cooling system. The arrangement of the zones are as shown in the figure below.

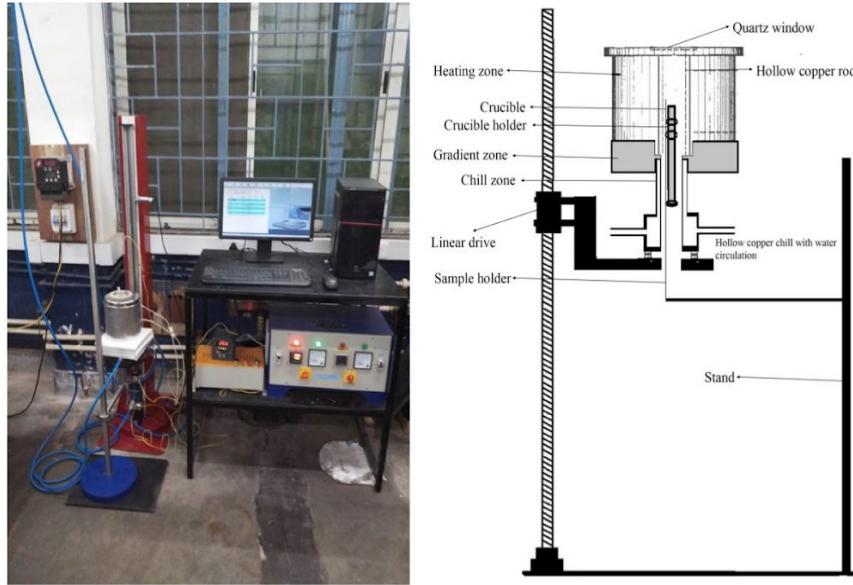


Figure 3. Modified Bridgman apparatus [20]

3.1.3 Sample preparation for imaging

The directionally solidified sample was broken out of the quartz tube and then sectioned transversely into pieces of equal lengths. The cutting was done using a low speed precision cutting saw. Post this step, the cut pieces were then polished thorough polishing papers 400 to 3000 and then polished till 0.1micron alumina. The end result was a mirror finish surface , perpendicular to the direction of solidification. These samples were then mounted for SEM imaging.



Figure 4. Transverse sectioning of directionally solidified sample.



Figure 5. Sample preparation for imaging(optical for preliminary investigation).

3.2 Image Preprocessing

The images obtained were 2560x2018 ,256 bit, of .tif format.

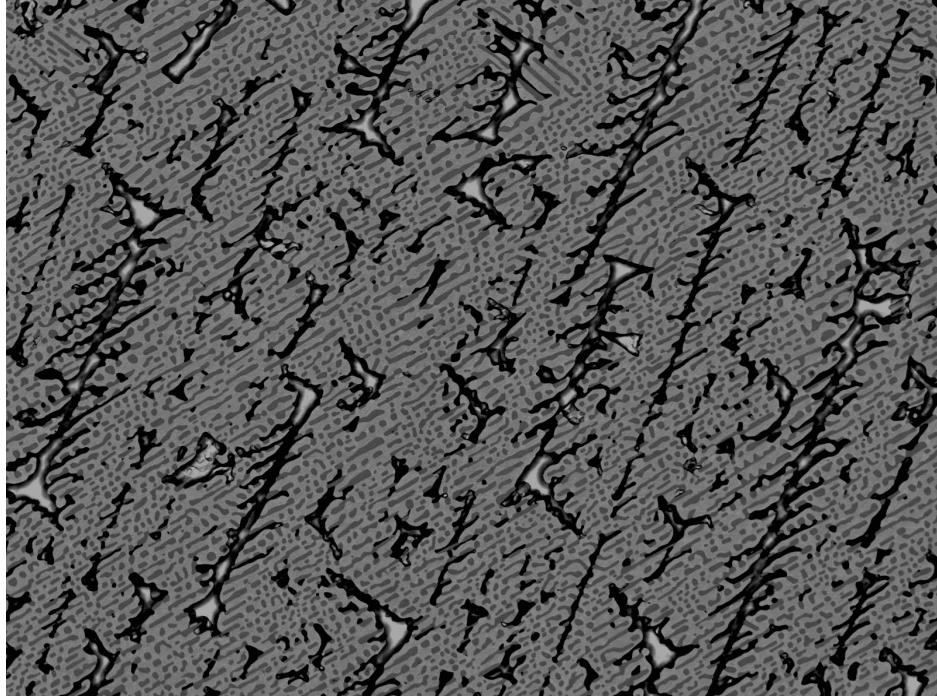


Figure 6. Image of eutectic composition at 1micron/s obtained from SEM, with the magnification details cropped out.

Here, we can see that microstructure doesn't have a specific pixel value assigned with a phase. Also the magnification scale should be cropped as it serves no value in our analysis. To proceed with the quantification of these microstructures, we first need to preprocess the images such that each phase is assigned a single pixel value and convert it to an 8 bit image.

This was achieved by first equalizing the image histogram and then thresholding the pixel values with two thresholds. This would ideally result in the three phases taking three distinct pixel values. Thresholding is when a range of possible pixel values are converted to a single pixel value of choice. Since the microstructure images had sections where a single phase had pixel values from the lightest to the darkest, thresholding alone didn't suffice and additional steps had to be taken. Namely, it was first ensured that the phase boundaries were closed and then each phase was bucket filled with its corresponding pixel value.

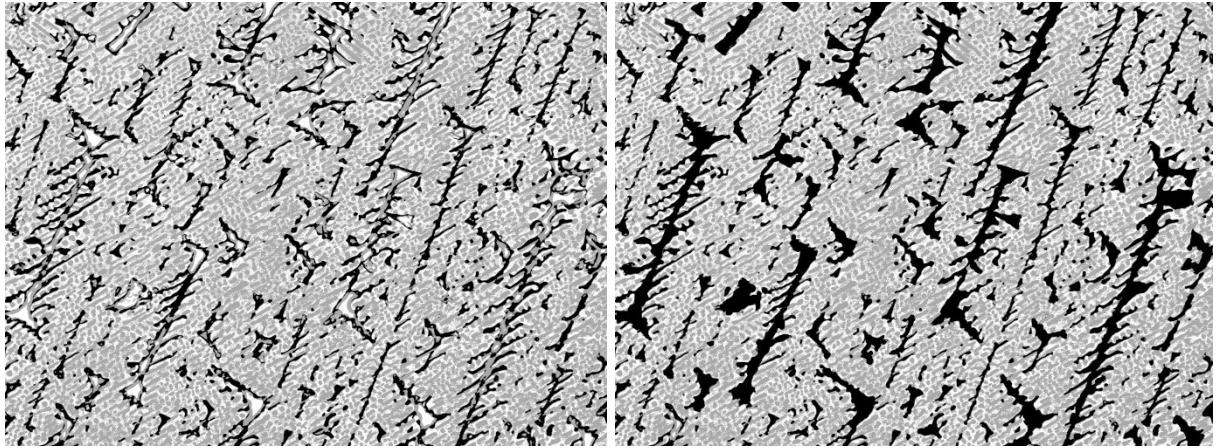


Figure 7. The image shown in figure 3.5 after thresholding(left) and bucket filling(right).

At this point, the images have to be cropped such that each image can be flattened and placed in an array for computation of 2 point statistics. The cropping has to be done such that we get a square image, i.e, height and width of the image are the same. The code used to achieve this is presented in the appendix along with other codes for the preprocessing steps.

4. Results

The microstructures obtained had varying patterns with variation in growth rate. The following is the microstructure obtained for a growth rate of $16\mu\text{s}$

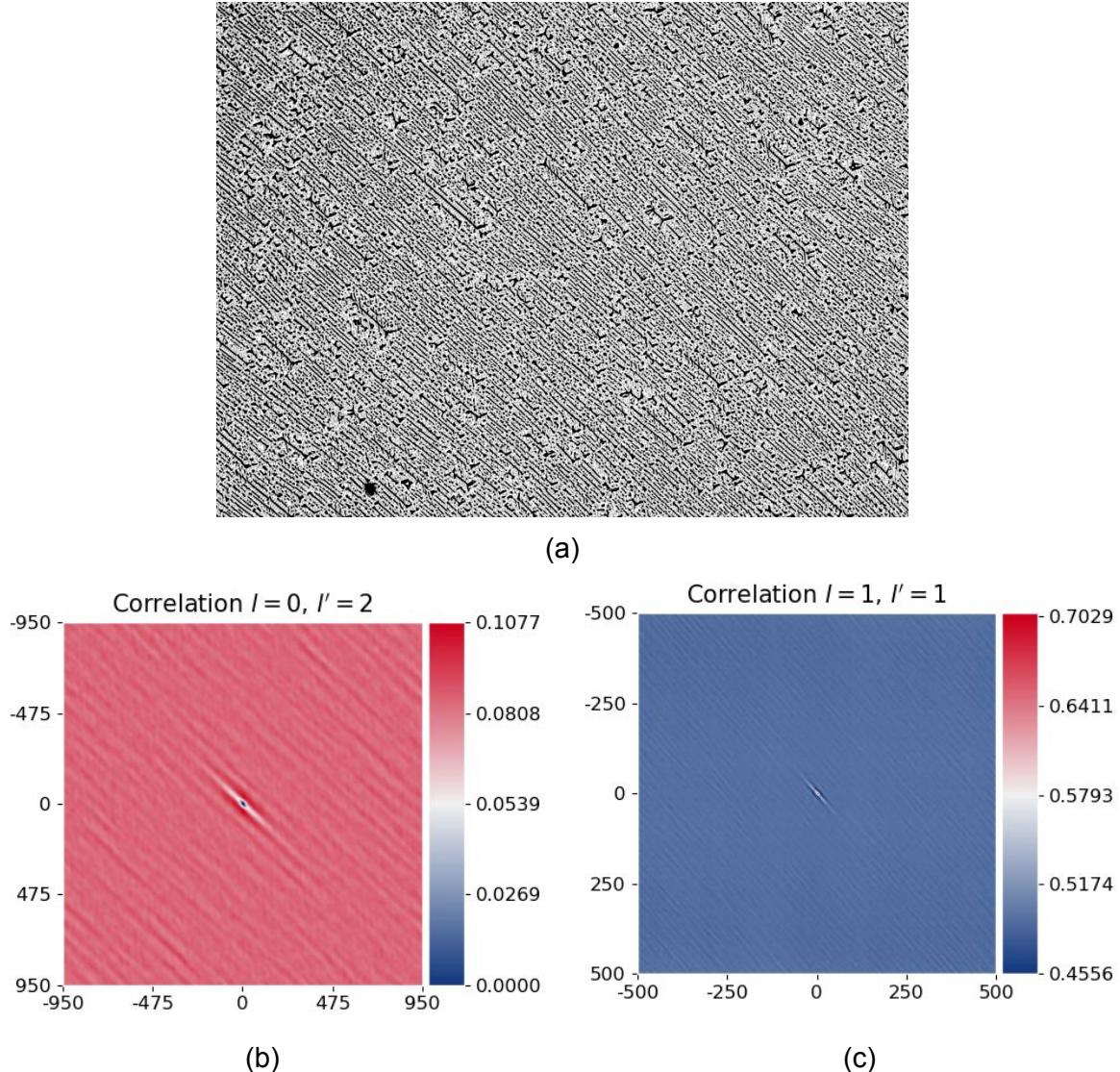


Figure 8. (a)microstructure obtained at growth rate of $16\mu\text{s}/\text{s}$. (b) cross correlation of the sb phase with the Ag₃Sb phase. (c) Autocorrelation of the Sb phase

Applying PCA on these 2-point data yielded an eigenvector which captured 95.19% variance of the initial data.

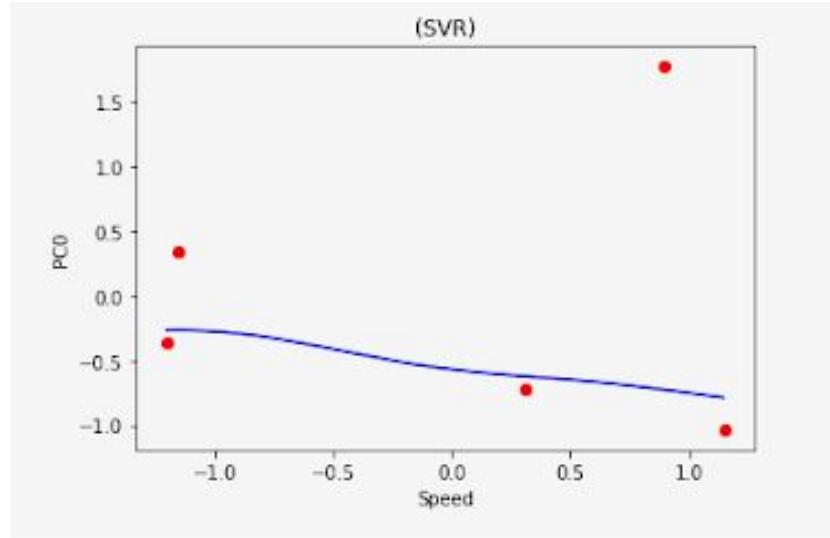


Figure 9. Best fit line using support vector regression.

Applying Here, the X-axis is the growth rate and the y-axis is the coefficient of the principal vector. Both these values have been centered around zero with a deviation of 1.

PC coeff	Growth rate
553.17	1
235.16	8
-339.87	16
-320.5	32
455.613	64

Table 1 The PC coefficients for different growth rates

5. Discussion

It can be seen in figure 4.1 (b) and (c) that for small vector sizes, the probability of auto correlation is high while the probability for cross correlation is low this makes sense, since each phase would have a certain volume.

To get a good fit between the PC coefficients and the growth rate, various images at each growth rate should be used. This would ensure a reliable relationship.

Since only single microstructure images of different growth rates were used, the fit is not reliable, as can be seen from the figure 4.2.

One of the observations was that the number of correlations required changes with the number of phases in the material as, $\# \text{phases} - 1$. This affects the scalability of this method as the input size is dependent on the number of phases. An alternative approach to tackle this is discussed in section 8.2.

6. Work Planned

Post the experiments conducted to obtain the 24 different microstructures of 8 solidification speeds and 3 compositions, it was planned to preprocess all the microstructures and then obtain the PCs of their 2 point statistics.

Then the accuracy of the linkage formed could be tested by dividing the data between a test and a training set and then measuring the distance between the test and predicted values.

Post this, the 2 point data of an unknown could have been predicted.

7. Problems faced

- Although the experiments were conducted at off eutectic compositions, the microstructures couldn't be imaged due to the unavailability of SEM facility.
- Unavailability of SEM facilities led to lack of ample data which is much needed for data based methods.
- Procurement of EBSD data would have made the preprocessing step practically error free. Since this was not possible, it led to an additional source of error in the process.

8. Conclusions and future work

- Dimensionality reduction on the limited dataset showed that most of the variance is captured by few PC vectors
- The validity of the regression method remains ambiguous as conclusive results can only be established with more data.

The following subsections contain threads that can be followed for future work.

8.1 Reconstruction of microstructure from 2 pt statistics

Loosely based on the gerchberg-saxton algorithm, an algorithm was suggested by Fullwood,2007_[21]. Using this method, the microstructure can be reconstructed from the 2- point data and therefore completing the linkage process.

8.2 Alternative method: Quantification, dimensionality reduction and reconstruction as a single step

The methods shown above make use of the features which humans perceive as important.An alternative approach would be to make use of numerous images of microstructures and to obtain the most essential features as an optimization problem. A type of artificial neural network called the convolutional autoencoder best fits our task of learning these encodings.Autoencoding is a data compression algorithm where the compression and decompression functions are data-specific, lossy, and learned automatically from examples rather than engineered by a human. In contrast to PCA, an autoencoder can map the data non-linearly and is bound to be at least as good as the former method(provided that the network was trained on ample relevant data). Unlike 2 point statistics, where the input data size increases linearly with number of phases,this method is not affected by the number of phases in the microstructure

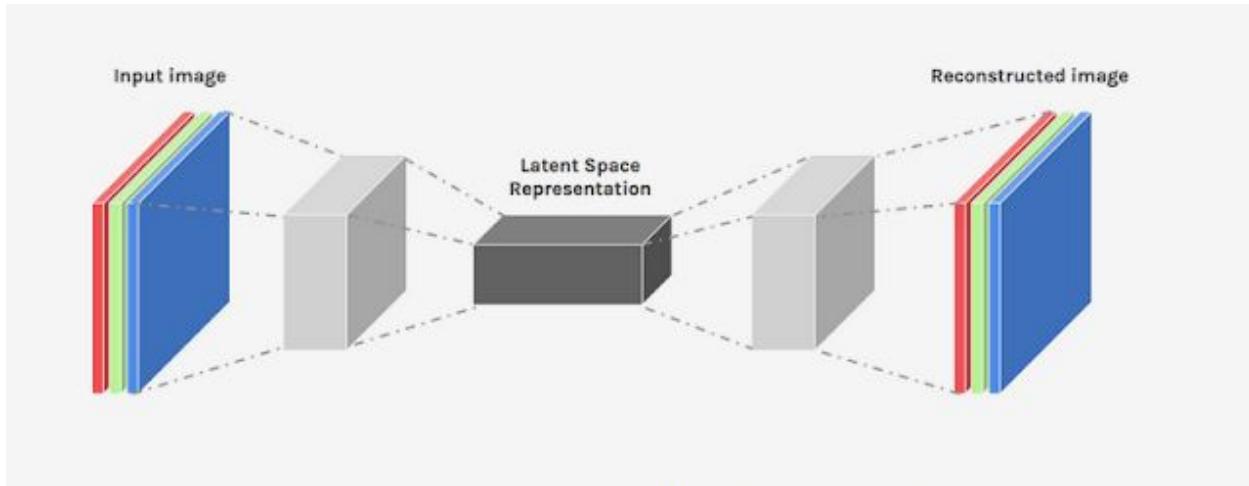


Figure 10. Figure 8.1 Proposed workflow of microstructure encoding and reconstruction according to Li, Xiaolin, 2018[21]

The problem of huge requirements of data can be tackled with the use of machine learning. Since the filters required to identify basic features of an image would remain the same at low level, the initial training can be done on random images, then the network can be trained again on pictures containing patterns while only updating the mid to terminal layers and finally the last layers could be trained on microstructure data.

9. References

1. K.A. Jackson and J.D. Hunt, Trans. Metall. Soc. AIME 236 (1966) p.1129.
2. A. Choudhury, M. Plapp and B. Nestler, Phys. Rev. E 83 (2011) p.051608.
3. Choudhury, A. Pattern-Formation During Self-organization in Three-Phase Eutectic Solidification. *Trans Indian Inst Met* 68, 1137–1143 (2015).
4. M.A. Ruggiero and J. Rutter, Mater. Sci. Technol. 13 (1997) p.5.
5. F.P. Dai, W.J. Xie and B. Wei, Phil. Mag. Lett. 89 (2009) p.170.
6. Crystal growth and segregation in vertical bridgman configuration.
7. Wei Zhai, Baojian Wang, Liang Hu & Bingbo Wei (2015) Ternary eutectic growth during directional solidification of Ag–Cu–Sb alloy, Philosophical Magazine Letters, 95:4, 187-193,
8. Microstructure reconstructions from 2-point statistics using phase-recovery algorithms
9. “Technical Overview”, as accessed on 20th june 2020
http://pymks.org/en/latest/rst/notebooks/tech_overview.html#2-Point-Statistics
10. Fullwood, David T., et al. “Microstructure Reconstructions from 2-Point Statistics Using Phase-Recovery Algorithms.” *Acta Materialia*, vol. 56, no. 5, 2008, pp. 942–948
11. S. Niezgoda, D. Fullwood, S. KalidindiDelineation of the space of 2-point correlations in a composite material system
Acta Mater, 56 (2008), pp. 5285-5292
12. Analytics for microstructure datasets produced by phase-field simulations
13. Lecture notes of CS 7015, by Mitesh Khapra.
14. K. P. Bennett and O. L. Mangasarian. Robust linear programming discrimination of two linearly inseparable sets. *Optimization Methods and Software*, 1:23–34, 1992.
15. C. Cortes and V. Vapnik. Support vector networks. *Machine Learning*, 20:273–297, 1995.
16. V. Vapnik., *The Nature of Statistical Learning Theory*. Springer, New York, 1995.
17. Alex J. Smola and Bernhard Schölkopf. *A Tutorial on Support Vector Regression*, 2003
18. M. A. Aizerman, É. M. Braverman, and L. I. Rozonoér. Theoretical foundations of the potential function method in pattern recognition learning. *Automation and Remote Control*, 25:821–837, 1964.
19. N. J. Nilsson. *Learning machines: Foundations of Trainable Pattern Classifying Systems*. McGraw-Hill, 1965.
20. Modified bridgman setup, A.S.Kiran

21. Li, Xiaolin, et al. "A Transfer Learning Approach for Microstructure Reconstruction and Structure-Property Predictions." *Scientific Reports*, vol. 8, no. 1, 2018,

Appendix

Preprocessing code

```

import cv2
from PIL import Image
import skimage.color
from PIL import Image, ImageOps
import glob

for name in glob.glob('/home/wadood/Desktop/wadood/python/preprocessed/*'):
    im = Image.open(name)
    # Setting the points for cropped image
    left = 0
    top = 0
    right = 1000
    bottom = 1000

    im1 = im.crop((left, top, right, bottom))
    im3 = im1.save(name)

for name in glob.glob('/home/wadood/Desktop/wadood/python/preprocessed/*'):
    im = Image.open(name)
    #im = Image.open("expt/10_.png")
    #print(im.format, im.size, im.mode)
    im2 = ImageOps.equalize(im, mask = None)
    #im2.show()
    im3 = ImageOps.grayscale(im2)
    #im3.show()
    px = im3.load()
    for i in range (2560):
        for j in range (1910):
            if px[i,j] > 170:
                px[i,j] = 2
            elif px[i,j] > 25 and px[i,j] <= 170:
                px[i,j] = 1
            elif px[i,j] <= 25:
                px[i,j] = 0
    im3 = im3.save(name)

```

2pt correlation and PCA

```

from __future__ import print_function
import os

```

```

import numpy as np

from pymks.datasets import make_delta_microstructures
from pymks import PrimitiveBasis
from pymks.bases import LegendreBasis
from pymks.stats import correlate
from sklearn.decomposition import PCA
import pickle as pk
import meshio
import cv2
from numpy import array
from PIL import Image
import glob

n_components = 8 # for PCA
n_frac = 0.95
n_states = 3
n_samples = 70
domain = [0, 2]

sample_count = 0
m_X_sampled = np.zeros([12,1000000])
for name in glob.glob('/home/wadood/Desktop/wadood/python/preprocessed/*'):
    im = array(Image.open(name))
    a = im.flatten()
    m_X_sampled[sample_count] = np.copy(a)
    sample_count += 1

# p_basis = PrimitiveBasis(n_states=n_states, domain=domain)
p_basis = LegendreBasis(n_states=n_states, domain=domain)
# i = 1
# for j in range(3):
i, j = 'all', 'all'
#all_correlations = [(i, j)]
X_stats = correlate(m_X_sampled, p_basis, correlations=[(2,2),(1,2)])

X_reshaped = X_stats.reshape((X_stats.shape[0], -1))
# pca = PCA(n_components)
pca = PCA(n_frac)
# X_mean = np.mean(X_reshaped, axis=1)[:, None]
pca_out = pca.fit_transform(X_reshaped)
f_write = "pc_scores_corr_" + str(i) + "_" + str(j) + ".csv"
f_model_write = "pca_fit_" + str(i) + "_" + str(j) + ".csv"
hdr = []
hdr = hdr + ["PC_" + str(i) for i in range(pca.n_components_)]
np.savetxt(f_write, np.column_stack([pca_out]), delimiter=',', header=str(hdr) )

```

```
# saving PCA fit data
np.savetxt(f_model_write, np.row_stack([pca.singular_values_, pca.explained_variance_ratio_]),
delimiter=',')
pk.dump(pca, open("pca_" + str(i) + "_" + str(j) + ".pkl", "wb"))

pc_vector = pca.components_
im = array(pc_vector)
```

Support Vector Regression

```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
from sklearn.preprocessing import StandardScaler
from sklearn.svm import SVR

dataset = pd.read_csv('linkagedata.csv')
X = dataset.iloc[:, 0:1].values
y = dataset.iloc[:, 1:2].values

sc_X = StandardScaler()
sc_y = StandardScaler()
X = sc_X.fit_transform(X)
y = sc_y.fit_transform(y)

regressor = SVR(kernel = 'rbf')
regressor.fit(X, y)

X_grid = np.arange(min(X), max(X), 0.01) #this step required because data is feature scaled.
X_grid = X_grid.reshape((len(X_grid), 1))
plt.scatter(X, y, color = 'red')
plt.plot(X_grid, regressor.predict(X_grid), color = 'blue')
plt.title('(SVR)')
plt.xlabel('Speed')
plt.ylabel('PC0')
plt.show()

y_pred = regressor.predict([[2]])
y_pred = sc_y.inverse_transform(y_pred)
print(y_pred)
```


Introduction

Obtaining microstructure images experimentally is an expensive affair, its costs arising from the labour put in, the equipment used and the resources consumed. Same goes for simulation methods, whose costliest aspects are the time and computational resources it takes. In this study, we aim to establish a linkage between the processing parameters (composition and solidification speed) and resultant microstructure, such that the microstructure for a random combination of processing parameters can be predicted. Development of such techniques will help us obtain a fundamental understanding of principle mechanisms of pattern formation in alloys which are of general importance for the development of multicomponent alloys for advanced technical applications.

The fundamental understanding of pattern formation for binary eutectics was established by Jackson and Hunt [1], who proposed when lamellar and rod eutectics are formed, the correlation between eutectic spacing and growth rate is expressed as $\lambda^2 v = k$ (K is a constant). Ternary eutectic alloys have received increasing attention in recent years owing to their more complex structures and thus different physical properties as compared with those of binary alloys [2]. The simultaneous growth of three distinct phases from the liquid allows for a broad variety of different phase arrangements far beyond the typical microstructures of rods, lamellae and labyrinths in binary eutectics[3]. It has been summarized that as solidification condition and/or alloy system vary, such eutectic morphologies as three lamellar-phase structures, one fibrous- and two lamellar-phase structures, two fibrous- and one lamellar-phase structures, two fibrous phase in continuous matrix structure, three fibrous-phase structures and so on can be formed in ternary eutectic alloys [4,5].

In the present work, ternary $Ag_{42.4}Cu_{21.6}Sb_{36}$ eutectic alloy – whose solidification process involves the cooperative growth of two intermetallic compounds $\theta(Cu_2Sb)$ and $\epsilon(Ag_3Sb)$ and one solid solution (Sb) phase, is studied for its variation in microstructure as a function of solidification speed/growth rate. Additionally, two off eutectic compositions are studied for the same variable too.

The various microstructures obtained are dependent on a lot of variables. By taking into consideration a specific alloy, we can treat much of these variables as constant. This leaves us with the two main variables of focus here, namely, growth rate and composition(since off eutectic alloys are tested too).

Capturing this variation as a result of these variables using an ML science approach involves preprocessing & processing of the data which is mentioned in the next section. The following sections cover the literature used to proceed with work, the methods used to obtain the results, the results, discussion and conclusion respectively.

2. Literature Review

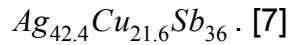
The problem at hand, where we attempt to predict the 2 point statistics of a microstructure undergoing a given set of processing conditions, involves the following steps. Firstly, since we are using supervised learning algorithms, the procurement of microstructure data whose processing conditions are known. Then, a method to obtain the 2 point statistics of a given microstructure. After which , we can make use of a regression algorithm to predict the 2 point statistics for any given processing conditions .

In the following subsections, the theory and present literature used to solve the problem at hand are presented.

2.1 Ternary eutectic solidification

Basic theories of solidification and segregation suggest that the quality of grown crystals is primarily controlled by the crystal growth rate, the growth interface morphology, and the solute redistribution[6].In directional solidification, the advancement of the growth interface into the melt(the growth rate) is governed by the balance of the heat fluxes at the interface.

In this work, the microstructure patterns of directionally solidified AgCuSb is studied at growth rates 0.5, 1, 2, 4, 8, 16, 32, 64 $\mu\text{m/s}$. The eutectic composition of this alloy is



The variation in microstructural pattern obtained at different growth rates was shown by Wei Zhai et al, 2015 [7]. This suggests that the chosen alloy would be appropriate for the problem at hand.

2.2 Quantification and reconstruction of microstructures

To form a linkage between the processing parameters and the resultant microstructure, the first step in the analysis would be to represent both the parameters and the microstructures in a suitable format. Only after this is achieved, can the process of linking the variables be done.

2.2.1 Quantification using 2-point statistics.

The goal here is to utilize the periodic properties of microstructures. Microstructure when viewed as an image has variables equivalent to the number of pixels in the image. However, this isn't the most efficient method to represent a microstructure and the dimensionality of this data can be reduced drastically. Two-point statistics (i.e., 2-point spatial correlations) are the most basic of the n-point spatial correlations that can be used to capture the details of the microstructure[8].

Microstructure Discretization[9]

The first step in all of MKS workflows is to discretize microstructures. To do this, we create a probabilistic description of a microstructure by introducing the continuous local state variable h , the local state space H and the microstructure function $m(h,x)$. The local state space H can be thought of as all of the thermodynamic state variables that are needed to uniquely define the material structure at a given location. The local state variable h is one instance of the local state space, or one configuration of state variables. The microstructure function $m(h,x)$ is a probability density function of finding a local state h at location x . For instance, let $\mu(x)$ be a microstructure that we plan to discretize, then μ is the expectation of the microstructure function.

$$\mu(x) = \int_H hm(h, x)dh$$

Now, we will discretize the microstructure in space by averaging over small cubic domains in the microstructure function. The local state can be discretized, using two methods: one is to bin the microstructure, using the primitive basis Λ

$$\frac{1}{\Delta x} \int_H \int_s \Lambda(h-l)m(h, x)dx dh = m[l, s]$$

the other is to use a spectral representation using some orthogonal basis function ξ

$$\frac{1}{\Delta x} \int_S m(h, x)dx dt = \sum_{l=0}^{L-1} m[l, s] \xi_l(h)$$

In the notation above, all of the round brackets variables are continuous variables and the square brackets variables are the discrete variables. The variables s and S represent a discrete position and the total volume, while l and L represent the discrete versions of h and H . In PyMKS the Legendre polynomials are currently the only orthogonal basis functions available. Either of these two discretization methods are used to discretize the microstructure. In the current work, legendre basis was used.

n-Point Spatial Correlations[9]

1-Point Spatial Correlations (or 1-point statistics)

N-point spatial correlations provide a way to rigorously quantify material structure, using statistics. As an introduction to n-point spatial correlations, let's first discuss 1-point statistics. 1-point statistics are the probability that a specified local state will be found in any randomly selected spatial bin in a microstructure. 1-point statistics compute the volume fractions of the local states in the microstructure. 1-point statistics are computed as

$$f[l] = \frac{1}{S} \sum_s m[s, l]$$

In this equation, $f[l]$ is the probability of finding the local state l in any randomly selected spatial bin in the microstructure, $m[s, l]$ is the microstructure function (the digital representation of the microstructure), S is the total number of spatial bins in the microstructure and s refers to a specific spatial bin.

While 1-point statistics provide information on the relative amounts of the different local states, they do not provide any information about how those local states are spatially arranged in the microstructure. Therefore, 1-point statistics are a limited set of metrics to describe the structure of materials.

2-Point Spatial Correlations

2-point spatial correlations (also known as 2-point statistics) contain information about the fractions of local states as well as the first order information on how the different local states are distributed in the microstructure.

2-point statistics can be thought of as the probability of having a vector placed randomly in the microstructure and having one end of the vector be on one specified local state and the other end on another specified local state. This vector could have any length or orientation that the discrete microstructure allows. The equation for 2-point statistics can be found below.

$$f[r|l, l'] = \frac{1}{S} m[s, l] m[s + r, l']$$

In this equation, $f[r|l, l']$ is the conditional probability of finding the local states l and l' at a distance and orientation away from each other defined by the vector r . All other

variables are the same as those in the 1-point statistics equation. In the case that we have an eigen microstructure function (it only contains values of 0 or 1) and we are using an indicator basis, the $r = 0$ vector will recover the 1-point statistics.

When the 2 local states are the same, $l = l'$, it is referred to as an autocorrelation. If the 2 local states are not the same, it is referred to as a cross-correlation.

In 2-pt statistics, we find the conditional probabilities for each possible vector. However, this produces a very large set of descriptors for the material structure. Indeed, the number of descriptors is equal to the number of distinct discretized vectors included in the analyses. For example, if one were to define a 21×21 neighborhood (which corresponds to including only ten voxels on each side of the voxel of interest) in a 2-D microstructure (or a cross-section of a 3D microstructure), one ends up with 441 descriptors of the material structure. Clearly, this number of material structure descriptors will be unwieldy for any subsequent analyses (e.g., extraction of process-structure-property linkages).

Therefore, a dimensionality reduction technique should be used which preserves the most important/significant descriptors. In prior work[10], principal component analysis (PCA) has been demonstrated to provide objective low dimensional representation of the 2-point statistics of the material microstructure.

The eutectic alloys explored in the present study exhibit three distinct phases in their evolution histories. The total number of 2-point spatial correlation functions that can be defined in a three-phase composite is nine. This is the number of different choices for the ordered pair (h, h') , since each index can take three different values. These nine correlation functions consist of three autocorrelations (i.e., $h=h'$) and six cross correlations (i.e., $h \neq h'$). Niezgoda et al. [11] have shown that at most, only two of these correlation functions will be independent.[12]

In the present study, we have selected an autocorrelation of the Sb phase and its cross correlation with Ag₃Sb to represent the set of independent 2-point statistics. This choice

was made because the Sb phase shows the highest variance and Ag₃Sb exhibits the highest values of the volume fraction in the entire ensemble of the simulation results produced for this study.

PCA [13]

PCA is a dimensionality reduction technique which basically maps/transforms data linearly such that the covariance matrix of the transformed data is a diagonal matrix, i.e., it has zero covariance among dimensions and non zero variance. Post this step, the dimensions with the lowest variance can be discarded. To minimize the error, k such dimensions can be selected. The steps taken to achieve this are described below.

Let $x_1, x_2, \dots, x_m \in R^n$ be m data points and let X be a matrix such that

x_1, x_2, \dots, x_m are the rows of this matrix. Assuming standardized data, i.e., data with zero mean and unit variance.

Let p_1, p_2, \dots, p_n be a set of such n linearly independent orthonormal vectors. Let P be a $n \times n$ matrix such that p_1, p_2, \dots, p_n are the columns of P .

Then,

$$\hat{X} = XP$$

Where \hat{X} is the matrix of transformed points.

And the original data points, x_i , can be represented as

$$x_i = \alpha_{i1}p_1 + \alpha_{i2}p_2 + \alpha_{i3}p_3 + \dots + \alpha_{in}p_n$$

Where, for an orthonormal basis, α_{ij} is

$$\alpha_{ij} = x_i^T \cdot p_j$$

Now, we want the covariance matrix of the transformed data to have specific properties, i.e.,

$$C_{ij} = 0 \text{ when } i \neq j \text{ (covariance = 0)}$$

$$C_{ii} \neq 0 \text{ when } i = j \text{ (variance } \neq 0)$$

From our previous assumption of X being a standardized matrix, the covariance matrix of our transformed data, \hat{X} , is given by

$$\frac{1}{m} \hat{X}^T \hat{X}$$

This follows since \hat{X} inherits the standardized property of X and

$$\begin{aligned} C_{ij} &= \frac{1}{m} \sum_{k=1}^m (X_{ki} - \mu_i)(X_{kj} - \mu_j) \\ &= \frac{1}{m} \sum_{k=1}^m X_{ki} X_{kj} \\ &= \frac{1}{m} X_i^T X_j = \frac{1}{m} (X^T X)_{ij} \end{aligned}$$

We can write the covariance matrix of the transformed data as follows,

$$\frac{1}{m} \hat{X}^T \hat{X} = \frac{1}{m} (XP)^T XP = \frac{1}{m} P^T X^T XP = P^T \left(\frac{1}{m} X^T X \right) P = P^T \Sigma P$$

From the desired properties mentioned above it follows that,

$$\frac{1}{m} \hat{X}^T \hat{X} = P^T \Sigma P = D \quad [\text{where } D \text{ is a diagonal matrix}]$$

This is achieved when the column of matrix P are eigenvectors of

$$\Sigma = X^T X \quad (\text{By singular value decomposition})$$

Thus, the new basis P used to transform X is the basis consisting of the Eigen Vectors of $X^T X$

2.2.2 Reconstructing 2 point data.

We make use of machine learning algorithms to find a relation between the transformed data points and the processing parameters. The next step would then be to reconstruct the original data from the predicted transformed data. However, there will be an error

during reconstruction. It arises from the discarding of the non dominant eigenvectors. The error will be equivalent to the variance captured by the selected eigen vectors subtracted from unity.

To transform the data back to its original space, we can multiply it with the inverse of the initial transformation matrix. However, since the initial transformation matrix is an orthonormal matrix, the inverse is equal to transpose.

Therefore, the retransformed data, X' , when the matrix of the chosen PC vectors is V is,

$$X' = \hat{X} V^T$$

Where, \hat{X} is the transformed data. Therefore,

$$X' = X V V^T$$

2.3 Microstructure variables and processing parameters linkage

2.3.1 Support Vector Regression

To predict the PCs from the processing parameter variables, various regression algorithms can be used. In this study, Support vector regression was used. The Support Vector Regression (SVR) uses the same principles as the SVM for classification, with only a few minor differences. Like classification, in the case of regression, a margin of tolerance (ϵ) is set in approximation to the SVM which would have already requested from the problem. The tacit assumption here is that such a function f actually exists that approximates all pairs (x, y) with ϵ precision, or in other words, that the convex optimization problem is feasible. Sometimes, however, this may not be the case, or we also may want to allow for some errors and therefore the algorithm could be more complicated than SVM. However, the main idea is always the same: to minimize

error, individualizing the hyperplane which maximizes the margin, keeping in mind that part of the error is tolerated.

Sometimes, however, this may not be the case, or we also may want to allow for some errors. Analogously to the “soft margin” loss function [14] which was adapted to SV machines by Cortes and Vapnik[15] , one can introduce slack variables ξ_i, ξ_i^* to cope with otherwise in-feasible constraints of the optimization problem . Hence we arrive at the formulation stated in [16].

$$\text{minimize} \quad \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*)$$

$$\begin{aligned} \text{Subject to, } & y_i - \langle w, x_i \rangle - b \leq \varepsilon + \xi_i \\ & \langle w, x_i \rangle + b - y_i \leq \varepsilon + \xi_i^* \\ & \xi_i, \xi_i^* \geq 0 \end{aligned}$$

The constant $C > 0$ determines the trade-off between the flat-ness of f and the amount up to which deviations larger than ε are tolerated. This corresponds to dealing with a so called ε -insensitive loss function $|\xi|_\varepsilon$ described by

$$\begin{aligned} |\xi|_\varepsilon &:= 0 \quad \text{if } |\xi| \leq \varepsilon \\ &:= |\xi| - \varepsilon \quad \text{otherwise} \end{aligned}$$

Fig.1 depicts the situation graphically. Only the points out-side the shaded region contribute to the cost insofar, as the deviations are penalized in a linear fashion.

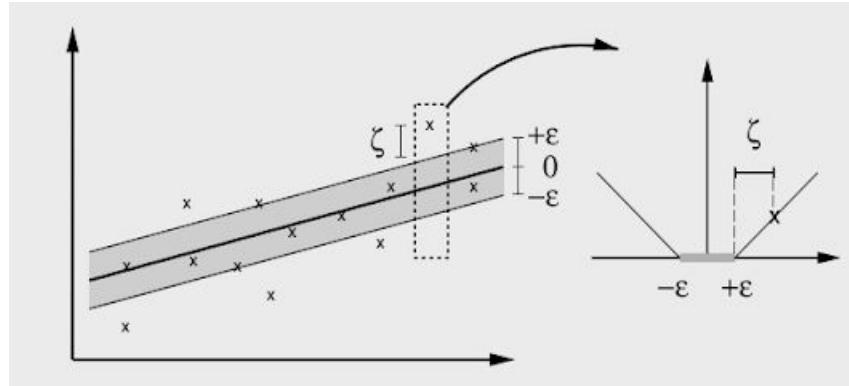


Figure 1 The soft margin loss setting for linear SVM [17]

At its core, this is similar to linear regression algorithm. The next step is to make the SV algorithm nonlinear. This, for instance, could be achieved by simply preprocessing the training patterns x_i by a map $\Phi : X \rightarrow F$ into some feature space F , as described in [18] [19] and then applying the standard SV regression algorithm.

In this work, the kernel used to transform the data into higher dimension space to make the linear separation possible was the gaussian radial basis function.

$$k(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right)$$

3. Methods

3.1 Experimental Details

3.1.1 Alloy preparation

$Ag_{42.4}Cu_{21.6}Sb_{36}$ is the eutectic composition for the Ag-Cu-Sb alloy. This composition was based on the eutectic composition mentioned in [7]. Preliminary experiments were conducted to verify this. For the preparation of the alloy, high purity, i.e., 99.999% pure Ag, Cu and Sb were used. Once these were weighed and taken in the right proportion, they were filled in quartz tubes of 4mm inner diameter(6mm outer diameter). Next, they were sealed at 10-5 mbar pressure using a rotary and then a diffusion pump subsequently.

Induction melting was used to melt all the components of the alloy thoroughly and then the melt was left to solidify and cool at room temperature. Once the sample reached ambient temperature, it was taken out of the quartz tube to be cut and crushed into chunks of size that fit into the quartz tube for the directional solidification step. Once the quartz tube was filled, the steps to vacuum seal were repeated.



Figure 2. Solidified alloy after induction melting(left). Crushing of the alloy (right).

3.1.2 Directional solidification

A modified bridgman solidification apparatus was used for these set of experiments. Here, the furnace had three zones in order to have good control over the thermal gradient. These were separated by thin ceramic based insulation zones. This established a gradient between the hot and the cold zones. Kanthal-A was selected as a heating element for the isothermal heating zone, while the chill zone has a water circulating cooling system. The arrangement of the zones are as shown in the figure below.

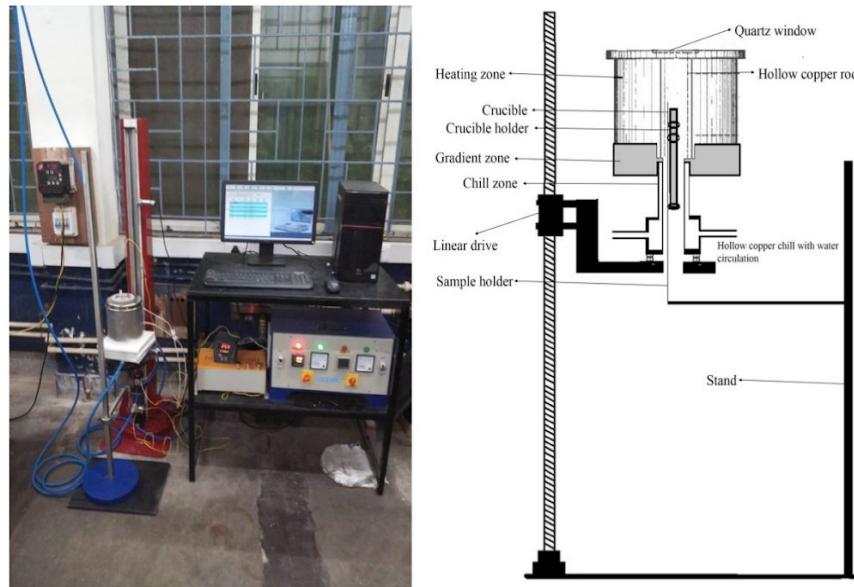


Figure 3. Modified Bridgman apparatus[20]

3.1.3 Sample preparation for imaging

The directionally solidified sample was broken out of the quartz tube and then sectioned transversely into pieces of equal lengths. The cutting was done using a low speed precision cutting saw. Post this step, the cut pieces were then polished thorough polishing papers 400 to 3000 and then polished till 0.1micron alumina. The end result was a mirror finish surface , perpendicular to the direction of solidification. These samples were then mounted for SEM imaging.



Figure 4. Transverse sectioning of directionally solidified sample.



Figure 5. Sample preparation for imaging(optical for preliminary investigation).

3.2 Image Preprocessing

The images obtained were 2560x2018 ,256 bit, of .tif format.

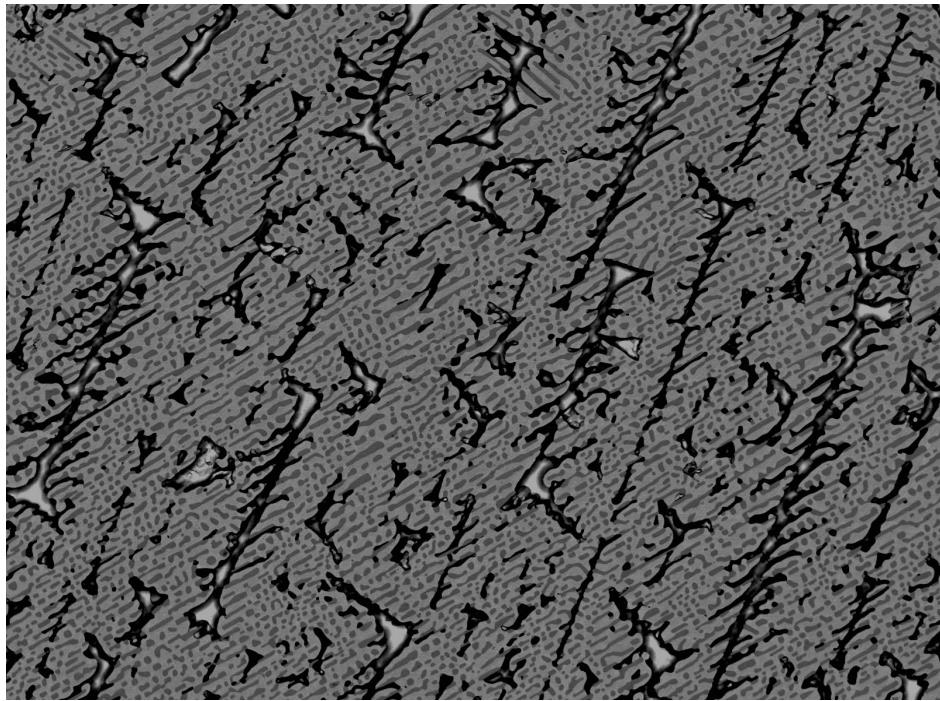


Figure 6. Image of eutectic composition at 1micron/s obtained from SEM, with the magnification details cropped out.

Here, we can see that microstructure doesn't have a specific pixel value assigned with a phase. Also the magnification scale should be cropped as it serves no value in our

analysis. To proceed with the quantification of these microstructures, we first need to preprocess the images such that each phase is assigned a single pixel value and convert it to an 8 bit image.

This was achieved by first equalizing the image histogram and then thresholding the pixel values with two thresholds. This would ideally result in the three phases taking three distinct pixel values. Thresholding is when a range of possible pixel values are converted to a single pixel value of choice. Since the microstructure images had sections where a single phase had pixel values from the lightest to the darkest, thresholding alone didn't suffice and additional steps had to be taken. Namely, it was first ensured that the phase boundaries were closed and then each phase was bucket filled with its corresponding pixel value.

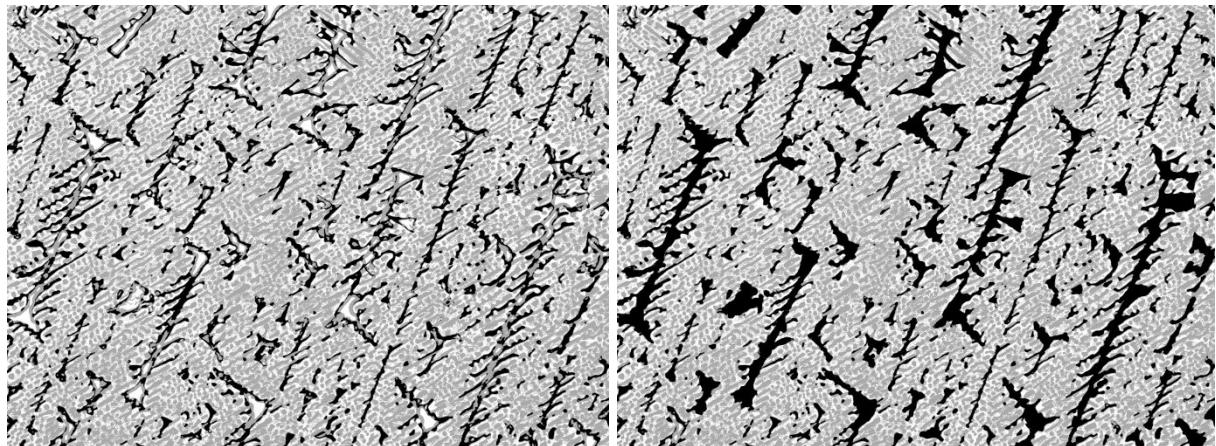


Figure 7. The image shown in figure 6 after thresholding(left) and bucket filling(right).

At this point, the images have to be cropped such that each image can be flattened and placed in an array for computation of 2 point statistics. The cropping has to be done such that we get a square image, i.e, height and width of the image are the same. The code used to achieve this is presented in the appendix along with other codes for the preprocessing steps.

4. Results

The microstructures obtained had varying patterns with variation in growth rate. The following is the microstructure obtained for a growth rate of $16\mu\text{s}$

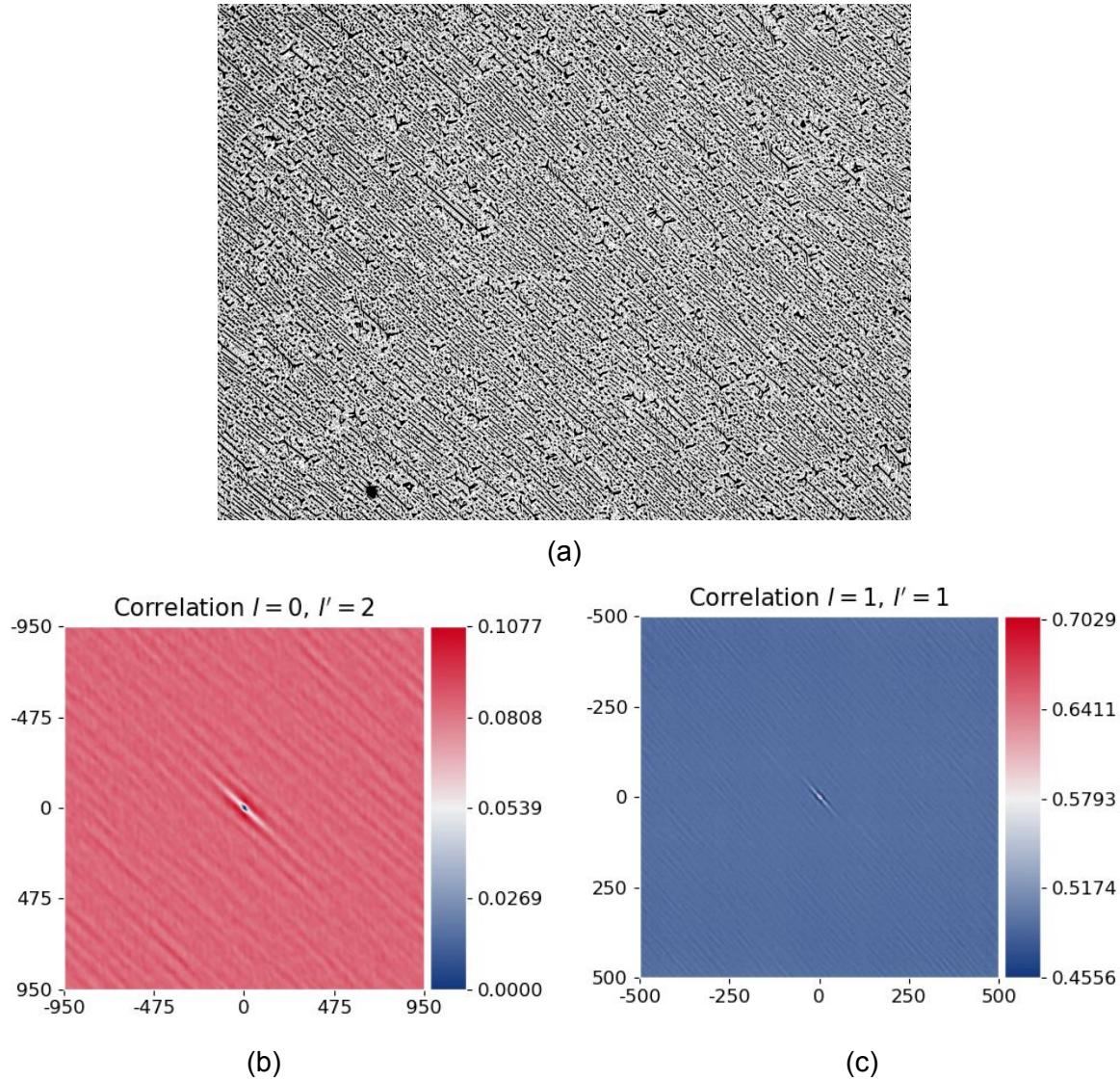


Figure 8. (a)microstructure obtained at growth rate of $16\mu\text{s}/\text{s}$. (b) cross correlation of the sb phase with the Ag₃Sb phase. (c) Autocorrelation of the Sb phase

Applying PCA on these 2-point data yielded an eigenvector which captured 98.9% variance of the initial data.

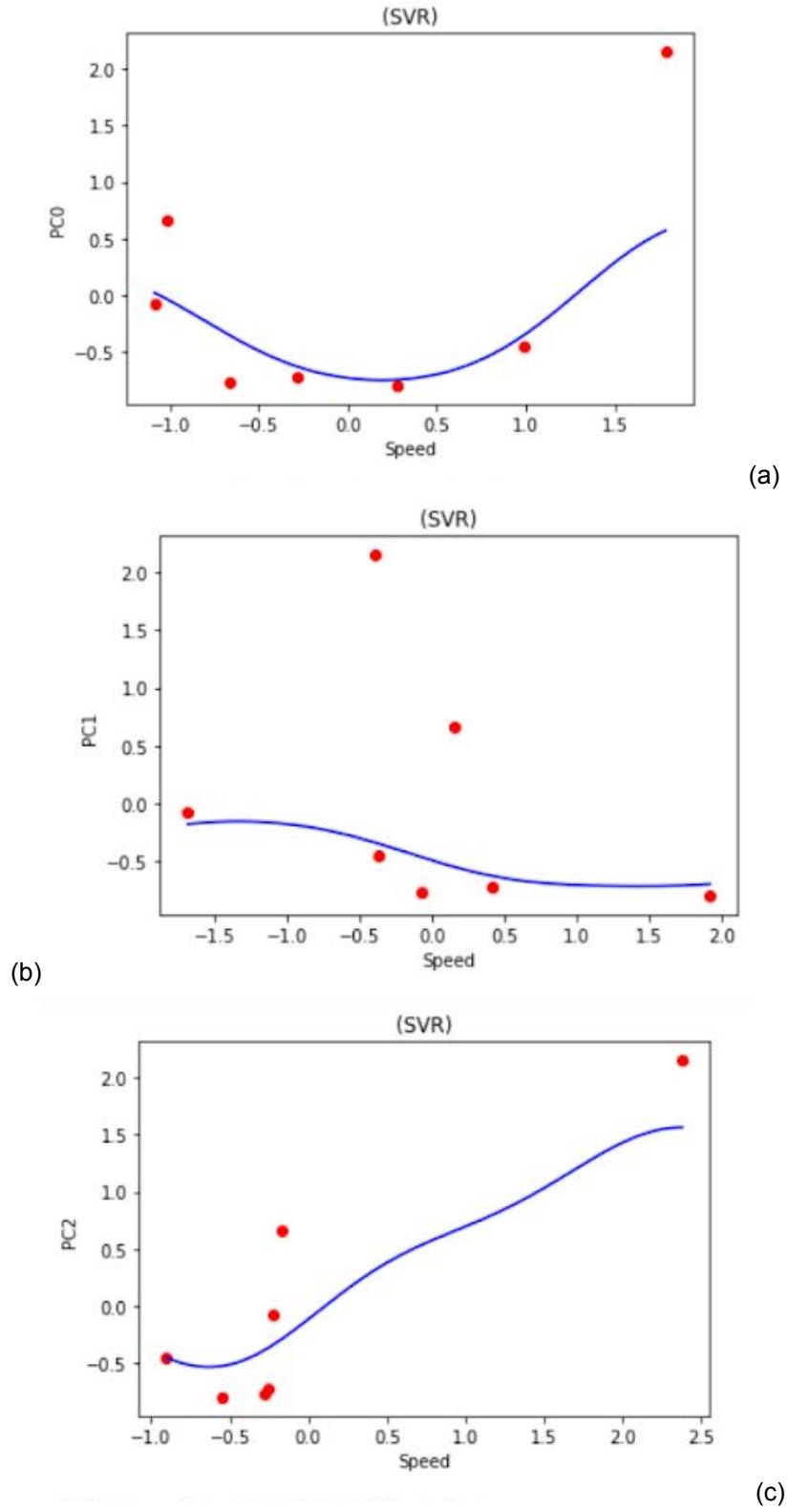


Figure 9. Best fit line of scaled values using support vector regression.

Applying Here, the X-axis is the growth rate and the y-axis is the coefficient of the principal vector. Both these values have been centered around zero with a deviation of 1.

PC_0	PC_1	PC_2	Growth rate
36.52	87.44	-13.22	0.5
-224.43	-0.97	-0.80	1
-119.08	20.89	-0.22	2
235.15	-13.94	-28.82	8
-339.87	-72.59	1.58	16
-320.50	9.03	3.74	32
455.61	-15.07	117.44	64

Table 1 The PC coefficients for different growth rates

5. Discussion

It can be seen in figure 4.1 (b) and (c) that for small vector sizes, the probability of auto correlation is high while the probability for cross correlation is low this makes sense, since each phase would have a certain volume.

To get a good fit between the PC coefficients and the growth rate, various images at each growth rate should be used. This would ensure a reliable relationship.

Since only single microstructure images of different growth rates were used, the fit is not reliable, as can be seen from the figure 4.2.

One of the observations was that the number of correlations required changes with the number of phases in the material as, #phases-1. This affects the scalability of this method as the input size is dependent on the number of phases. An alternative approach to tackle this is discussed in section 8.2.

6. Work Planned

Post the experiments conducted to obtain the 24 different microstructures of 8 solidification speeds and 3 compositions, it was planned to preprocess all the microstructures and then obtain the PCs of their 2 point statistics.

Then the accuracy of the linkage formed could be tested by dividing the data between a test and a training set and then measuring the distance between the test and predicted values.

Post this, the 2 point data of an unknown could have been predicted.

7. Problems faced

- Although the experiments were conducted at off eutectic compositions, the microstructures couldn't be imaged due to the unavailability of SEM facility.
- Unavailability of SEM facilities led to lack of ample data which is much needed for data based methods.
- Procurement of EBSD data would have made the preprocessing step practically error free. Since this was not possible, it led to an additional source of error in the process.

8. Conclusions and future work

- Dimensionality reduction on the limited dataset showed that most of the variance is captured by few PC vectors
- The validity of the regression method remains ambiguous as conclusive results can only be established with more data.

The following subsections contain threads that can be followed for future work.

8.1 Reconstruction of microstructure from 2 pt statistics

Loosely based on the gerchberg-saxton algorithm, an algorithm was suggested by Fullwood,2007[20]. Using this method, the microstructure can be reconstructed from the 2- point data and therefore completing the linkage process.

8.2 Alternative method: Quantification, dimensionality reduction and reconstruction as a single step

The methods shown above make use of the features which humans perceive as important.An alternative approach would be to make use of numerous images of microstructures and to obtain the most essential features as an optimization problem. A type of artificial neural network called the convolutional autoencoder best fits our task of learning these encodings.Autoencoding is a data compression algorithm where the compression and decompression functions are data-specific, lossy, and learned automatically from examples rather than engineered by a human. In contrast to PCA, an autoencoder can map the data non-linearly and is bound to be at least as good as the former method(provided that the network was trained on ample relevant data). Unlike 2 point statistics, where the input data size increases linearly with number of phases,this method is not affected by the number of phases in the microstructure

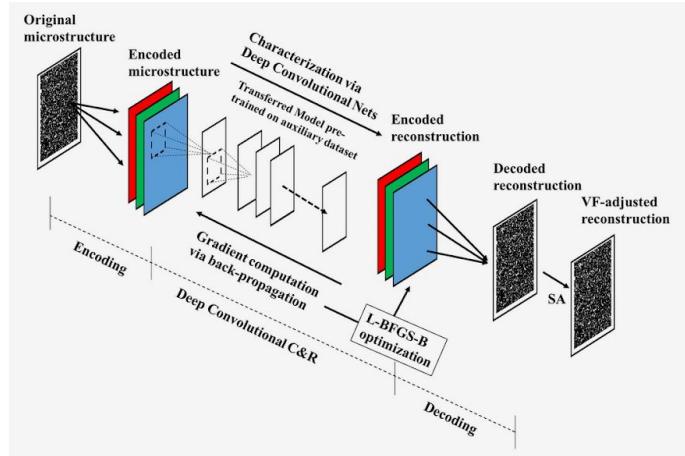


Figure 10. Figure 8.1 Proposed workflow of microstructure encoding and reconstruction according to Li, Xiaolin, 2018[21]

The problem of huge requirements of data can be tackled with the use of machine learning. Since the filters required to identify basic features of an image would remain the same at low level, the initial training can be done on random images, then the network can be trained again on pictures containing patterns while only updating the mid to terminal layers and finally the the last layers could be trained on microstructure data.

9. References

1. K.A. Jackson and J.D. Hunt, Trans. Metall. Soc. AIME 236 (1966) p.1129.
2. A. Choudhury, M. Plapp and B. Nestler, Phys. Rev. E 83 (2011) p.051608.
3. Choudhury, A. Pattern-Formation During Self-organization in Three-Phase Eutectic Solidification. *Trans Indian Inst Met* 68, 1137–1143 (2015).
4. M.A. Ruggiero and J. Rutter, Mater. Sci. Technol. 13 (1997) p.5.
5. F.P. Dai, W.J. Xie and B. Wei, Phil. Mag. Lett. 89 (2009) p.170.
6. Crystal growth and segregation in vertical bridgman configuration.
7. Wei Zhai, Baojian Wang, Liang Hu & Bingbo Wei (2015) Ternary eutectic growth during directional solidification of Ag–Cu–Sb alloy, Philosophical Magazine Letters, 95:4, 187-193,
8. Microstructure reconstructions from 2-point statistics using phase-recovery algorithms
9. “Technical Overview”, as accessed on 20th june 2020
http://pymks.org/en/latest/rst/notebooks/tech_overview.html#2-Point-Statistics
10. Fullwood, David T., et al. “Microstructure Reconstructions from 2-Point Statistics Using Phase-Recovery Algorithms.” *Acta Materialia*, vol. 56, no. 5, 2008, pp. 942–948
11. S. Niezgoda, D. Fullwood, S. KalidindiDelineation of the space of 2-point correlations in a composite material system
Acta Mater, 56 (2008), pp. 5285-5292
12. Analytics for microstructure datasets produced by phase-field simulations
13. Lecture notes of CS 7015, by Mitesh Khapra.
14. K. P. Bennett and O. L. Mangasarian. Robust linear programming discrimination of two linearly inseparable sets. *Optimization Methods and Software*, 1:23–34, 1992.
15. C. Cortes and V. Vapnik. Support vector networks. *Machine Learning*, 20:273–297, 1995.
16. V. Vapnik., *The Nature of Statistical Learning Theory*. Springer, New York, 1995.
17. Alex J. Smola and Bernhard Schölkopf. A Tutorial on Support Vector Regression, 2003
18. M. A. Aizerman, É. M. Braverman, and L. I. Rozonoér. Theoretical foundations of the potential function method in pattern recognition learning. *Automation and Remote Control*, 25:821–837, 1964.
19. N. J. Nilsson. *Learning machines: Foundations of Trainable Pattern Classifying Systems*. McGraw-Hill, 1965.
20. Modified bridgman setup, A.S.Kiran
21. Li, Xiaolin, et al. “A Transfer Learning Approach for Microstructure Reconstruction and Structure-Property Predictions.” *Scientific Reports*, vol. 8, no. 1, 2018,

Appendix

Preprocessing code

```

import cv2
from PIL import Image
import skimage.color
from PIL import Image, ImageOps
import glob

for name in glob.glob('/home/wadood/Desktop/wadood/python/preprocessed/*'):
    im = Image.open(name)
    # Setting the points for cropped image
    left = 0
    top = 0
    right = 1000
    bottom = 1000

    im1 = im.crop((left, top, right, bottom))
    im3 = im1.save(name)

for name in glob.glob('/home/wadood/Desktop/wadood/python/preprocessed/*'):
    im = Image.open(name)
    #im = Image.open("expt/10_.png")
    #print(im.format, im.size, im.mode)
    im2 = ImageOps.equalize(im, mask = None)
    #im2.show()
    im3 = ImageOps.grayscale(im2)
    #im3.show()
    px = im3.load()
    for i in range (2560):
        for j in range (1910):
            if px[i,j] > 170:
                px[i,j] = 2
            elif px[i,j] > 25 and px[i,j] <= 170:
                px[i,j] = 1
            elif px[i,j] <= 25:
                px[i,j] = 0
    im3 = im3.save(name)

```

2pt correlation and PCA

```

from __future__ import print_function
import os
import numpy as np

from pymks.datasets import make_delta_microstructures
from pymks import PrimitiveBasis
from pymks.bases import LegendreBasis
from pymks.stats import correlate

```

```

from sklearn.decomposition import PCA
import pickle as pk
import meshio
import cv2
from numpy import array
from PIL import Image
import glob

n_components = 8 # for PCA
n_frac = 0.95
n_states = 3
n_samples = 70
domain = [0, 2]

sample_count = 0
m_X_sampled = np.zeros([12,1000000])
for name in glob.glob('/home/wadood/Desktop/wadood/python/preprocessed/*'):
    im = array(Image.open(name))
    a = im.flatten()
    m_X_sampled[sample_count] = np.copy(a)
    sample_count += 1

# p_basis = PrimitiveBasis(n_states=n_states, domain=domain)
p_basis = LegendreBasis(n_states=n_states, domain=domain)
# i = 1
# for j in range(3):
i, j = 'all', 'all'
#all_correlations = [(i, j)]
X_stats = correlate(m_X_sampled, p_basis, correlations=[(2,2),(1,2)])

X_reshaped = X_stats.reshape((X_stats.shape[0], -1))
# pca = PCA(n_components)
pca = PCA(n_frac)
# X_mean = np.mean(X_reshaped, axis=1)[:, None]
pca_out = pca.fit_transform(X_reshaped)
f_write = "pc_scores_corr_" + str(i) + "_" + str(j) + ".csv"
f_model_write = "pca_fit_" + str(i) + "_" + str(j) + ".csv"
hdr = []
hdr = hdr + ["PC_" + str(i) for i in range(pca.n_components_)]
np.savetxt(f_write, np.column_stack([pca_out]), delimiter=',', header=str(hdr) )
# saving PCA fit data
np.savetxt(f_model_write, np.row_stack([pca.singular_values_, pca.explained_variance_ratio_]),
delimiter=',')
pk.dump(pca, open("pca_" + str(i) + "_" + str(j) + ".pkl", "wb"))

```

```
pc_vector = pca.components_
im = array(pc_vector)
```

Support Vector Regression

```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
from sklearn.preprocessing import StandardScaler
from sklearn.svm import SVR

dataset = pd.read_csv('linkagedata.csv')
X = dataset.iloc[:, 0:1].values
y = dataset.iloc[:, 1:2].values

sc_X = StandardScaler()
sc_y = StandardScaler()
X = sc_X.fit_transform(X)
y = sc_y.fit_transform(y)

regressor = SVR(kernel = 'rbf')
regressor.fit(X, y)

X_grid = np.arange(min(X), max(X), 0.01) #this step required because data is feature scaled.
X_grid = X_grid.reshape((len(X_grid), 1))
plt.scatter(X, y, color = 'red')
plt.plot(X_grid, regressor.predict(X_grid), color = 'blue')
plt.title('SVR')
plt.xlabel('Speed')
plt.ylabel('PC0')
plt.show()

y_pred = regressor.predict([[2]])
y_pred = sc_y.inverse_transform(y_pred)
print(y_pred)
```

