CALCULATION, UTILIZATION, AND INFERENCE OF SPATIAL STATISTICS IN PRACTICAL SPATIO-TEMPORAL DATA

A Dissertation
Presented to
The Academic Faculty

by

Ahmet Cecen

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Computational Science and Engineering

Georgia Institute of Technology
August 2017

CALCULATION, UTILIZATION, AND INFERENCE OF SPATIAL STATISTICS IN PRACTICAL SPATIO-TEMPORAL DATA

Approved by:

Dr. Surya Kalidindi, Advisor
The George W. Woodruff School of Mechanical Engineering
Georgia Institute of Technology

Dr. Hamid Garmestani
School of Materials Science and Engineering
Georgia Institute of Technology

Dr. Le Song
School of Computational Science and Engineering
Georgia Institute of Technology

Dr. Duen Horng Chau
School of Computational Science and Engineering
Georgia Institute of Technology

Dr. Sung Ha Kang
School of Mathematics
Georgia Institute of Technology

Date Approved: July 20, 2017

ACKNOWLEDGEMENTS

I would like to acknowledge the invaluable contributions of my advisor, Surya R. Kalidindi to both my academic and personal growth over the 8 years we have worked together, through my undergraduate and graduate studies.

I acknowledge the financial support of Air Force Office of Scientific Research (AFOSR) through the Multidisciplinary University Research Initiative (MURI) award FA9550-12-1-0458. I would also like to recognize my collaborators through this initiative for their personal and academic contributions, especially Yue Sun whose ideas are directly utilized in parts of this dissertation.

I would like to thank my committee members Hamid Garmestani, Le Song, Sung Ha Kang and Duen Horng Chau for their input in this work. I would also like to thank David M.Turner for the fruitful discussions we had over the years on some of the key concepts that form a basis for this work. Finally, I would like to thank all members of our research group for their conversation and support.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

S	Total number of discrete spatial bins in a microstructure
m_s^h	Microstructure function
c_s	Mask function
$\odot$	Element-wise matrix product
$\widetilde{m}_s^h$	Masked microstructure function
$f_t^{hh'}$	Two-point statistic corresponding to vector *t*
$f^{hh'}$	Matrix of two-point statistics
FFT	Fast Fourier Transform
$\Im$	Forward Fourier Transform
$\Im^{-1}$	Backward Fourier Transform
$F^{hh'}$	Fourier Transform of matrix of two-point statistics
τ	Time step
κ_1	First principal curvature
κ_2	Second principal curvature
H	Mean curvature
V	Interfacial velocity
OCF	Overlap Cost Function
DCF	Distance Cost Function
SCF	Statistical Cost Function

SUMMARY

The direct influence of spatial and structural arrangement in various length scales to the performance characteristics of materials is a core premise of materials science. Spatial correlations in the form of n-point statistics have been shown to be very effective in robustly describing the structural features of a plethora of materials systems, with a high number of cases where the obtained futures were successfully used to establish highly accurate and precise relationships to performance measures and manufacturing parameters.

This work addresses issues in calculation, representation, inference and utilization of spatial statistics under practical considerations to the materials researcher. Modifications are presented to the theory and algorithms of the existing convolution based computation framework in order to accommodate deformed, irregular, rotated, missing or degenerate data, with complex or non-probabilistic state definitions. Memory efficient personal computer oriented implementations are discussed for the extended framework. A universal microstructure generation framework with the ability to efficiently address a vast variety of geometric or statistical constraints including those imposed by spatial statistics is assembled while maintaining scalability, and compatibility with structure generators in literature.

CHAPTER 1. INTRODUCTION

Most materials of interest to modern technology display internal architectures of various complexity at multiple length scales, ranging from atomic/molecular arrangements to imposed structural patterns at the visible scales. Certain, not yet fully identified characteristics of these architectures directly control the properties exhibited by a given material. Understanding the complex hierarchical internal structure of materials (referred simply as the material microstructure) and the related effects on the physical phenomenon of interest is central to virtually all considerations in the field of materials science and engineering. As such, one of the core products of materials science is the expression of such relationships in the form of process-structure-property linkages, which requires robust identification and quantification of the important microstructure metrics and the associated properties.

While the quantification of process and property metrics have a rich history arguably dating back millennia, the quantification of structure is a fairly recent field with accumulated legacy knowledge based mostly on anecdotes and empirical metrics. Until the advent of modern imaging techniques such as electron microscopy or x-ray computed tomography, it was impossible to obtain any meaningful representation of structure in lower length scales for most material systems. Therefore, most early work on description of structures relied on hypothetical constructs of simple geometric shapes, such as tightly packed spheres. With the rapid development of sophisticated high-resolution experimental techniques and imaging tools, there is now a critical need for efficient strategies to distill useful insights and descriptive metrics from massive amounts of 3D and 4D materials

datasets. Although the end goals are usually different, many parallels are ripe to be drawn from signal/image processing and machine learning paradigms as enablers for efficient processing of such datasets.

A microstructure quantification framework heavily relying on signal processing and feature selection techniques already exist in the literature as spatial correlations, practical computation of which requires the utilization of moving windows. Spatial correlations of various forms have been widely used in both analytical [1, 2] and computational [3-10] linkages tying material structure to effective properties. However, significant limitations exist in the computation and application of spatial statistics in real data obtained through imaging. Furthermore, much work is needed on the proper interpretation of spatial statistics in order to guide physical models, as well the quantification of the effect of modifications to microstructure on its spatial statistics.

This dissertation is focused on improvements and extensions to the existing framework in both forward mapping from microstructure to spatial statistics descriptors and backwards mapping from spatial statistics to microstructural information. Specifically:

1) Efficient accommodation of conditions and problems commonly encountered real world materials data in computation of spatial statistics. Modifications are derived and described to the existing theory and algorithms of the convolution based computation framework in order to accommodate deformed, irregular, rotated, missing or degenerate data, with complex or non-probabilistic state definitions. Strategies are described for memory efficient personal computer oriented implementations of the extended framework.

2) Assembly of a universal microstructure generation framework with the ability to efficiently address a wide variety of geometric/physical or statistical constraint including those imposed by spatial statistics with a focus on scalability; while maintaining compatibility with any and all specialized structure generators in literature. This framework can be used to solve the reverse mapping problem to an extent, as well being useful as a dataset design or enrichment tool.

CHAPTER 2. BACKGROUND

This dissertation deals predominantly with tools and methodologies related to the understanding of material microstructure. What follows is a brief survey of existing metrics and methodologies for structural description.

2.1 Microstructure Function

A microstructure function expresses spatially resolved material structure information gathered from any source, either experiments or simulations. Conceptually, one can think of the microstructure function as $l(x)$, where l denotes the local state occupying the spatial position x. In this notation, the local state refers to any combination of attributes used to define the material locally (e.g., a combination of elemental composition, phase identifier, crystal lattice orientation, and dislocation density may be used to define the local state in multiphase polycrystalline materials at the mesoscale). Brief reflection will expose the unwieldy nature of such a description, especially when one tries to include a diverse set of local state attributes over multiple hierarchical length scales. In an effort to overcome this challenge, the concept of a stochastic microstructure function was introduced [11]. In this novel concept, the microstructure function is defined as $m(l,x)$, where m denotes the probability density associated with finding the local state l at the spatial position x. Consequently, $m(l,x)dldx$ captures the corresponding probability measure.

Although it is theoretically possible to extract a digital representation of the microstructure function using a multitude of choices in the selection of the basis functions for both the spatial and local state variables [12, 13], for clarity, we focus our attention in

the background section on the simplest of these bases corresponding to the primitive binning of the spatial domain as well as the local state space. With this choice, $m(l,x)$ admits a simple digital description as:

$$m(l,x)dldx \approx \sum_{h}\sum_{s} m_s^h \chi_s(x)\chi^h(l) \tag{1}$$

where $\chi_i()$ denotes a set of indicator basis functions, and m_s^h denotes a digital microstructure signal. For example, $\chi_s(x)$ allows partitioning of the spatial domain into non-overlapping volumes (typically employed as uniform binning of the space so that DFT methods can be applied later), with the function taking the value one for all points inside the sub-volume enumerated by s and the value zero for all other points. Note that $\chi^h(l)$ can be defined in a similar manner for any local state space of interest. Figure 1 presents a simple illustration of these concepts. It is also important to recognize that m_s^h can be physically interpreted as the probability of finding any of the local states corresponding to local state bin enumerated by h in the spatial bin enumerated by s. Consequently, it should be noted that m_s^h reflects a spatially resolved description of the material structure in a broadly applicable form. Note that $0 \leq m_s^h \leq 1$. It is also emphasized that the digital microstructure signal is inherently tied to a specific length scale (defined by size of spatial bins) and a specific resolution of the local state (defined by size of local state bins).

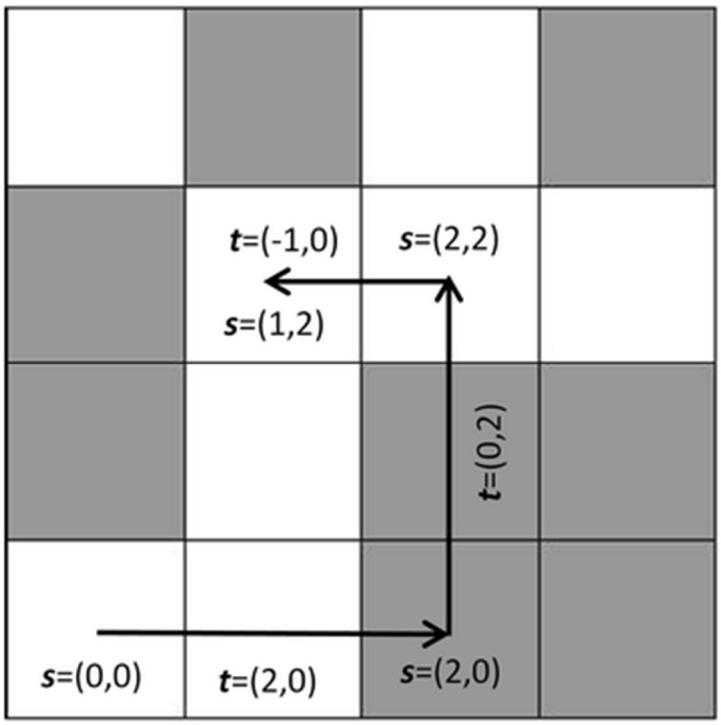

Figure 1 – Illustration of the discretized microstructure, m_s^n. In this highly simplified microstructure, there are only two local states that are conveniently indexed by h, with $h = 1$ denoting the phase represented by white and $h = 2$ denoting the phase represented by gray. Example values of the microstructure signal are $m_{(1,2)}^1 = 1$, $m_{(1,2)}^2 = 0$, $m_{(2,0)}^1 = 0$, and $m_{(2,0)}^2 = 1$. The interpretation for the index *t* used to label the discretized vector space is also illustrated. Note that both *s* and *t* are used as vector indices in this figure.

2.2 Empirical Metrics of Structure

Conventional practices for the quantification of the material microstructure have largely relied on accumulated legacy knowledge by domain experts and intuition. These measures usually include simple scalar values representing some average quantity of the entire structure (or image). The most common and most intuitive of these metrics is the volume fraction [14-21], representing the percentage of the volume under consideration belonging to a particular local state (equivalent to a color or intensity histogram for regular images). Local states represent any discrete mapping of material qualities to

numbers/intensities in an image. For the purposes of this paper a local state can be defined as any attribute relevant to a problem of consideration, such as whether a particular voxel contains a solid or a liquid, a hard material or a soft material, aluminum or copper etc. Other traditional metrics include but are not limited to particle size [15-17], grain size [22], particle spacing [14, 15, 17, 23], aspect ratio [14, 17, 24], roundness [24], coordination number [21], surface area [16, 19], connectivity [18] and tortuosity [25-27]. It can easily be shown that metrics composed of a small set of scalars are unlikely to be able to fully describe a structure, because it is easy to imagine multiple shapes/patterns that would exhibit the same average qualities but have different structures, resulting in vastly different performance characteristics.

2.3 2-Point Statistics

Given the vast separation between the length scales of interest in most materials development activities, any specific microstructure dataset collected for the material of interest is best interpreted as an experimental outcome of a stochastic process. Under this consideration, attempts to quantify the material microstructure should be able accommodate statistical interpretations, be insensitive enough to account for issues such as limited sampling size or experimental errors but sensitive enough to reveal common or distinctive structural features.

Robust statistical descriptions of structure are found in literature in the form of spatial correlations [1, 11, 28-33]. Although a number of different measures of the spatial correlations in the microstructure are possible, the n-point spatial correlations (or n-point statistics) provide the most complete set of measures that are naturally organized by

increasing amounts of structure information. The most basic of the n-point statistics are the 1-point statistics, and they reflect the probability of finding a specific local state of interest at any randomly selected voxel in the material structure. In other words, they essentially capture the information on volume fractions of the various distinct local states present in the material system.

The next higher level of structure information is contained in the 2-point statistics, which capture the probability of finding specified local states h and h' at the tail and head, respectively, of a prescribed vector r randomly placed into the material structure. This can be expressed using the concept of the microstructure function, as:

$$f(l, l'|r) = \frac{1}{|\Omega(r)|} \int_{\Omega(r)} m(l, x) m(l', x + r) dx \tag{2}$$

Note that the vector here r carries both the magnitude and direction in this definition. $\Omega(r)$ here denotes the volumetric domain of the material internal structure analyzed, with $|\Omega(r)|$ denoting the measure of the corresponding volume. It is important to note the dependence of the volumetric domain on the vector itself. This is because material structures studied often have finite domains (except when periodicity is invoked) and the domain available for evaluating the 2-point spatial correlation depends on the vector r. This is because only those points where it is possible to evaluate both $m(l, x)$ and $m(l', x + r)$ can be included in the evaluation of Eq. 2. As one might imagine there are certain regions near the boundaries of a given microstructure image where this condition is not met (i.e., either x or $x + r$ fall outside the given image) and therefore the region available for use should be expected to show a strong dependence on r.

Analogous to the treatment of the microstructure function earlier, we can express the probability measure as $f(l,l'|r)dldl'$ and establish a simple digital representation of this function as:

$$f(l,l'|r)dldl' \approx \sum_{h'}\sum_{h}\sum_{t} f_t^{ll'}\chi_t(r)\chi^h(l)\chi^{h'}(l') \tag{3}$$

where the index t effectively bins the vector space associated with r as illustrated in Figure 1. Starting with the above notions, one can establish the desired relationship between the digital representations of microstructure and the (directionally resolved) 2-point spatial correlations as [29, 30]:

$$f_t^{hh'} = \frac{1}{S_t}\sum_{s=1}^{S_t} m_s^h m_{s+t}^{h'} \tag{4}$$

where S_t captures the r-dependence of $\Omega(r)$. The denominator S_t is essentially the total number of trials conducted (where each trial denotes checking what local states exist in spatial bins marked s and $s+t$) and the numerator $\sum_{s=1}^{S_t} m_s^h m_{s+t}^{h'}$ denotes an expected measure of total success in these trials (i.e., actually finding the selected local states h and h' at the two bins, respectively). Recognizing this feature allows one to make any needed corrections for different situations.

The computation of $f_t^{hh'}$ for a specified combination of h and h', essentially requires $\mathrm{O}(S^2)$ (i.e., of the order of S^2) computations ($\mathrm{O}(S)$ for each value of t and there are $\mathrm{O}(S)$ different values of t). Such calculations are generally very expensive and are not easily scalable for datasets with high values of S. In recent years, it has been demonstrated

that these computations can be accomplished at $O(Slog\ S)$ using Fast Fourier Transforms (FFTs) [29, 30]. However, FFTs implicitly assume that the microstructure being studied is periodic in all directions (i.e., it can be extended by simply repeating the entire domain as many times as needed). With the assumption of periodicity, S_t can be taken to be the same as S (the total number of spatial bins in the microstructure). This is because every spatial bin in the microstructure can be used to place the tail (or equivalently the head) of the vector in evaluating the 2-point statistics. Furthermore, one can simply use the properties of DFTs to compute $f_t^{hh'}$. This is because Eq. 4, with the assumption of periodicity, translates to the following in the DFT space via the convolution theorem:

$$F_k^{hh'} = \frac{1}{S} M_k^{h*} \odot M_k^{h'}, \qquad F_k^{hh'} = \Im\left(f_t^{hh'}\right), \qquad M_k^h = \Im(m_s^h) \tag{5}$$

where $\odot$ is the element-wise product operator (also known as Hadamard or Schur product). Superscript * denotes the complex conjugate and $\Im()$ denotes the DFT transformation of the data to the frequency space enumerated by k (in the context of this paper, this is the spatial frequency space). As a result the computation of the 2-point statistics is reduced to computing the DFT of m_s^h, performing requisite products in the frequency space (where they are fully uncoupled), and performing an inverse DFT. For plotting the 2-point statistics, the most intuitive visualizations of 2-pt. statistics would result if $t = 0$ lies in the center of plot. This shift is accomplished trivially by making use of the periodicity implied in the DFT based computations.

Figure 2 illustrates the above concepts through a simple "honeycomb" microstructure, where each pixel or voxel is colored either white or black. Since there are

two local states, we can potentially compute a total of four different 2-point spatial correlations functions: f_t^{11}, f_t^{12}, f_t^{21}, and f_t^{22}, where $h = 1$ refers to the white-colored phase and $h = 2$ refers to the black-colored phase in Figure 2. Exploiting the known properties of DFTs, Niezgoda et al. [29] have demonstrated that the number of independent 2-point spatial correlations is only $H - 1$, where H is the total number of distinct local states present in the material system of interest. Consequently, for two-phase microstructures in most of the examples in this dissertation, we generally need to compute only one of the autocorrelations. Figure 2 shows a plot of white-white autocorrelation.

The autocorrelations presented in Figure 2 capture a number of salient features of the microstructure. The hexagonal symmetry, the feature shape, and the feature spacing are readily apparent. Furthermore, the periodicity implied in the use of DFTs resulted in the autocorrelations also exhibiting the same periodicity. Note also that the autocorrelation for the zero vector (at the center of the plot) provides the phase volume fraction. An important consequence of invoking periodicity assumptions is that the number of trials for all vectors is exactly the same and is equal to the number of pixels or voxels in the microstructure studied. In other words, all vectors of interest have been sampled fairly.

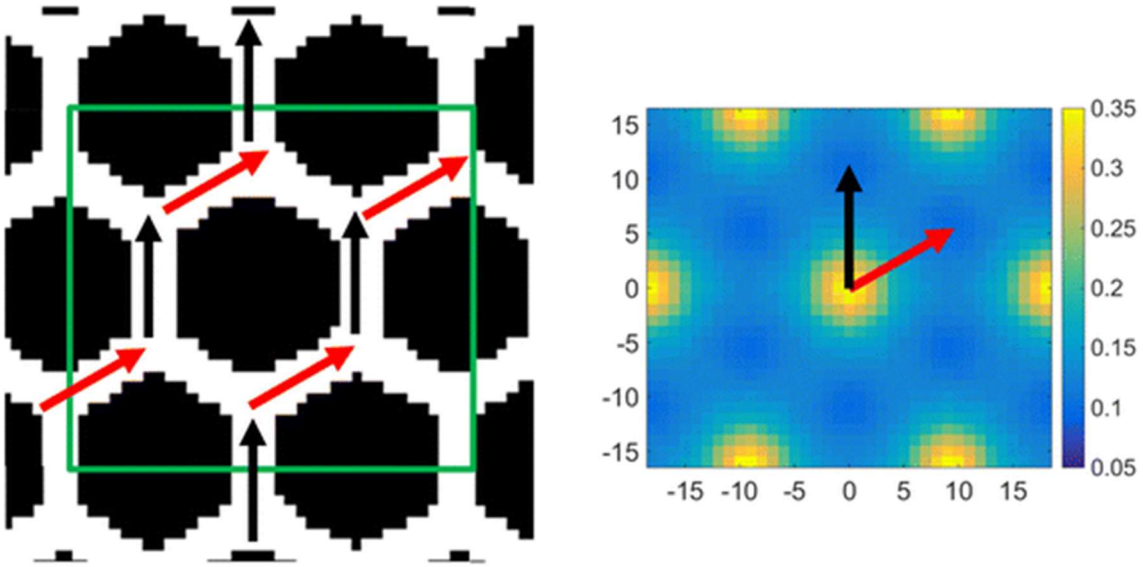

Figure 2 – Illustration of the computation and visualization of 2-point statistics while invoking the periodicity assumption. Left: the microstructure used in the computation. The actual microstructure, shown in the green box in the center, is extended by invoking the periodicity assumption. This extension is only for visualization purposes, and allows us to see the use of the exact same sampling size for all vectors of interest in the microstructure domain. Right: the corresponding white-white autocorrelation map.

It is also possible to use DFTs to compute 2 point statistics with non-periodic boundary conditions, As a specific example, we will revisit the same structure illustrated earlier, but without invoking the assumption of periodicity. In other words, our interest is to compute the autocorrelations as defined in Eq. (4), while accounting for the fact that $S_t \neq S$. However, as stated earlier, a direct implementation would incur $O(S^2)$ computations. A much better computational strategy would result if one "pads" the microstructure such that only long vectors (larger than the vectors of interest in computing the 2-point statistics) can wrap around from one edge of the original image to the opposite edge when the periodic assumption is implicitly invoked to take advantage of the computational expediency of the DFTs.

The padding strategy described above is illustrated in Figure 3. Let $S = (S_1, S_2)$ denote the number of spatial bins in the original two-dimensional microstructure (shown in the inner green box). The padding now extends the microstructure function to $\tilde{S} = (S_1 + T_1, S_2 + T_2)$, where $T = (T_1, T_2)$ identifies the range of the vectors for which the 2-point statistics are to be computed. The reader is cautioned that use of very high values of T can produce meaningless answers. As an example, if one chooses $T = (S_1, S_2)$, then one can see that the number of trials conducted for the largest vector in computing the 2-point statistics is just one. Based on prior experience, $T < (S_1/2\,, S_2/2)$ is recommended. Let the padded microstructure be denoted as $\tilde{m}_s^n$. The spatial bins in the padded region of the microstructure may be assigned any of the local states that are not involved in the computation of the desired 2-point statistics. For example, if we are interested in computing f_t^{11} only, then the spatial bins in the padded region can be assigned a local state enumerated by 2 or a completely new local state enumerated by 3 (making the padded microstructure a 3-phase microstructure).

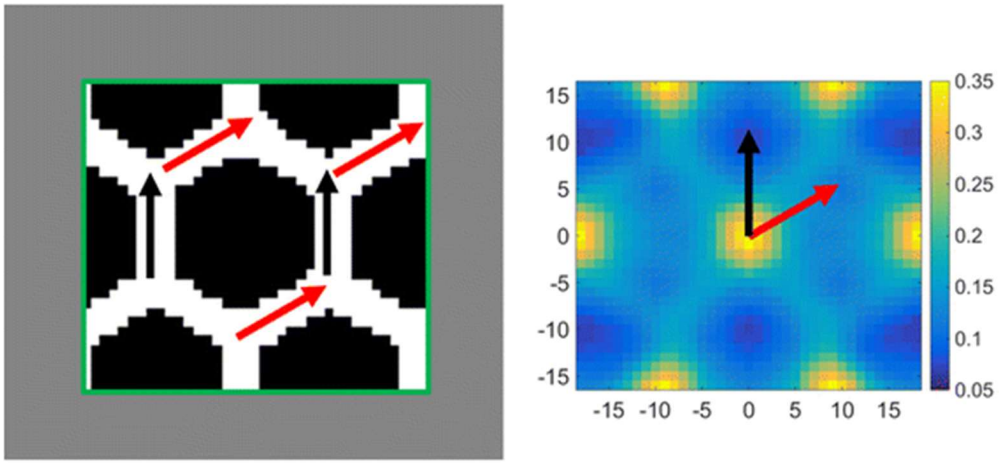

Figure 3 – Illustration of the padding strategy to compute the 2-point statistics using DFT representations while avoiding the errors associated with the implicit periodic boundary assumptions. The green box around the original microstructure is only for visualization.

With the padded microstructures, we are now in a position to take advantage of DFTs. We can first compute $\widetilde{M}_k^h = \Im(\widetilde{m}_s^h)$, and then $\Im^{-1}\left(\widetilde{M}_k^{h*} \circ \widetilde{M}_k^h\right)$, which produces an accurate count of the number of successes in finding local states n and p separated by all vectors $t \leq T$. In fact, the computation described above produces results even for vectors $t > T$, but these results are corrupted by vectors wrapping around the padded region because of the periodicity assumption implicit in the DFTs. However, since our interest here is exclusively in $t \leq T$, we will only take these results from the DFT computation described above. In order to compute the 2-point statistics of interest, we simply need to divide these numbers with a suitable denominator denoting the total number of trials involved, which is expressed simply as $(S_1 - |t_1|)(S_2 - |t_2|)$. The padding in Figure 3 is shown such that it equally envelopes all sides of the original microstructure. This is just for easy visualization and interpretation. In reality, any placement of the original microstructure inside the overall padded region (i.e., any unequal distribution of the padding as long as the extended microstructure has the same overall size) will produce identical results for the computed 2-point statistics (this is, once again, a consequence of using DFTs).

Figure 3 also depicts a plot of the f_t^{11} (white-white) autocorrelations that are not tainted by the periodicity assumptions implied in the use of DFTs. A comparison of the autocorrelations in Figures 2 and 3 reveals important consequences of the assumption of periodicity. For example, the hexagonal symmetry is no longer evident in the autocorrelations (see the values corresponding to the black and red vectors shown in these figures). This is mainly because the different vectors are no longer sampled the exact same number of times. Although this may not be as important when one deals with a very large

image, it clearly has an effect for the relatively small image shown in Figure 3. In this simple example, one can easily reconcile the different values of the autocorrelations for the red and black vectors depicted in Figure 3, by noting that we can indeed place many more red vectors with both endpoints in a white pixel, when compared to the similar placement of the black vectors. It is therefore important to recognize that the assumption of periodicity can indeed influence significantly the computed 2-point statistics, especially when one has a limited number of features in the image. Note that the strategy described above can be applied selectively on any of the bounding planes of the image. In other words, one can decide to invoke periodicity assumption on certain bounding planes and employ the padding strategy described above selectively on the other bounding planes.

2.4 Other Statistical Measures of Structure

There are a few other statistical descriptions/functions of structure in literature. Some of the most prominent are:

- *Surface Correlation Function (SFC)*: The probability of finding an interface between local states in the microstructure at the tail and head, respectively, of a prescribed vector *r* randomly placed into the material structure. [1, 9]
- *Radial Distribution Function (RDF)*: The probability of finding specified local states *h* and *h'* at the tail and head, respectively, of a prescribed vector of length $|r|$ randomly placed into the material structure. RDF discards any angular information on the prescribed vectors.[1, 2, 8, 34, 35]

- *Lineal Path Function (LPF)*: The probability of finding the specified local state h wholly within a prescribed line segment r randomly placed into the material structure.[7, 9, 10, 36]
- *Chord Length Distribution (CLD)*: The probability of finding the specified local state h wholly within a prescribed chord r randomly placed into the material structure. The distinction between LPF and CLD is that CLD only considers chords (vectors that are drawn from interface to interface) while LPF includes all line segments.[1, 3, 37-39]

Despite claims to the contrary in literature [1, 10], brief inspection will immediately reveal that in general all 4 of these functions are specific subsets of n-point statistics. The relationship between these functions are shown in Figure 4. The figure is only a reference for subset relationships, any depiction of the relative size of these sets are highly inaccurate. We can redefine these functions as:

- *Surface Correlation Function (SFC)*: The probability of finding specified local states h, which is an interface indicator function, at the tail and head, respectively, of a prescribed vector of length r. Related to 2-point statistics by a narrower definition of local states.
- *Radial Distribution Function (RDF)*: The probability of finding specified local states h and h' at the tail and head, respectively, of a prescribed vector of length $|r|$ randomly placed into the material structure. This is an angular integral of 2-point statistics.
- *Lineal Path Function (LPF)*: For a line segment with the length of r spatial bins, the probability of finding the specified local state h at every spatial bin on over the

r points. As such, LPF consists of a collection of very particular subsets of *r* point statistics.

- *Chord Length Distribution (CLD)*: For a line segment with the length of *r* spatial bins, the probability of finding the specified local state *h* at every spatial bin on over the *r* points, as well finding the local state $h' \neq h$ at a bin before and a bin after the line segment. As such, *CLD* consists of a collection of very particular subsets of *r+2* point statistics.

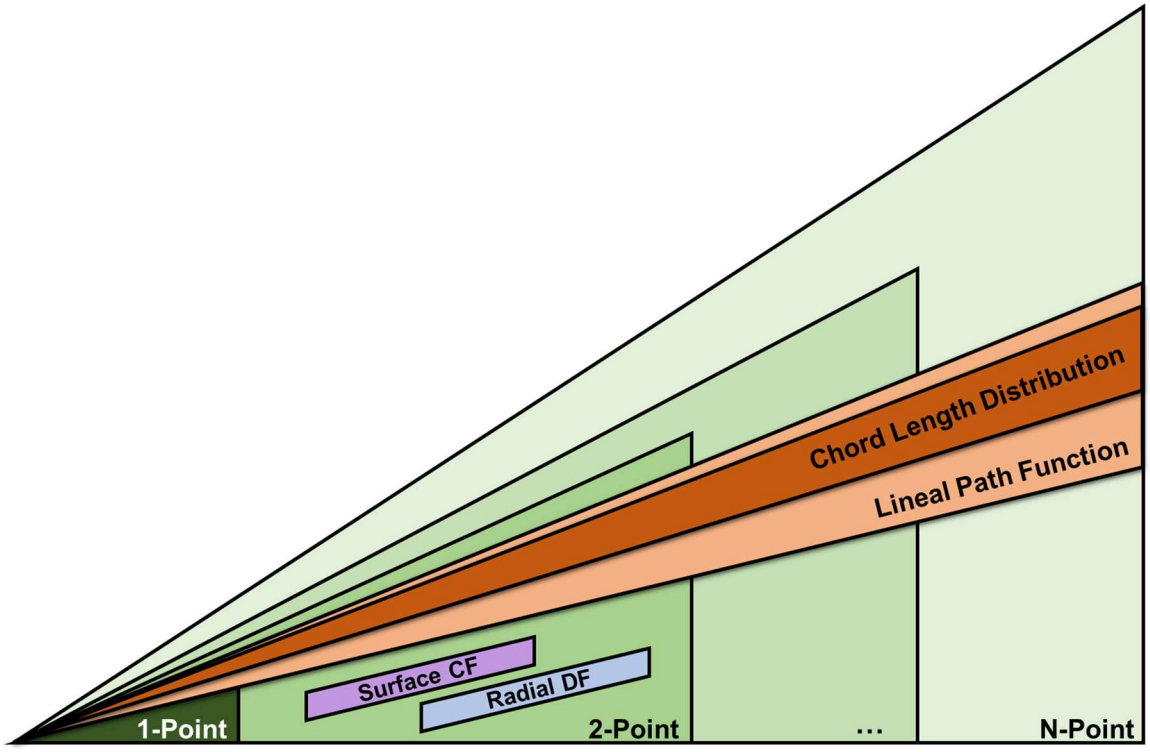

Figure 4 – Relationship between various subsets of n-point statistics. Each stage of the pyramid represents information content in addition to the previous stages. The graphic is intended only to show subset relationships and does not reflect an accurate size of information content.

2.5 Supporting Arguments for 2-point Statistics as Structure Descriptors

2-point statistics is the structural measure of choice for the work in this dissertation. There are several reasons for this:

- *Natural Origin Problem:*

  Because of the absence of a natural origin from where one might start indexing the spatial bins, only the relative placement of local states in the material structure contains meaningful information. In other words, only the spatial correlations in the material structure contain high value information.

- *Anisotropy of Real Data:*

  Most complex material systems show considerable anisotropy, as such average measures or measures that discard directional information such as the RDF cannot be used effectively.

- *Convolution Based Formulations:*

  2-point statistics can be computed in a linear moving window using convolutions in O(SlogS) time via FFTs. This enables seamless interplay with many image processing concepts.

- *Analytical Work Showing Direct Relationships to Material Properties:*

  The strongest support for the choice of n-point spatial correlations as the most appropriate measures of material structure comes from the pioneering work of Kroner [40], who has shown that the effective properties of composite material systems can be conveniently expressed as a series sum with the structure details entering this series explicitly in the form of n-point spatial correlations. These composite theories have been generalized to a broad range of materials phenomena,

and have been summarized in several books [1, 41, 42]. There are also several reports in literature, where they have been successfully applied to estimate effective properties (both linear and nonlinear) of a broad range of materials with complex structures [43-48]. Physically, the n-point spatial correlations are very effective in rigorously quantifying the local neighbourhoods in the complex internal structure of most advanced materials. Since the local neighbourhoods control the local response, it is only logical that the n-point spatial correlations are the ideal measures of the material structure in formulating process-structure-property (PSP) linkages of interest in designing high performance engineering components.

- *Past Success in Establishing Direct Relationships to Material Properties and Process Parameters:*

  In prior work, 2-point statistics have been successfully utilized to establish structure relations to property or process in multiphase composite systems [6, 49-53], atomistic datasets [5, 54], and polycrystalline microstructures [55, 56].

2.6 Dimensionality Reduction

It should be noted that there is a tremendous leap in the amount of structure information contained in the 2-point statistics compared to the 1-point statistics. In a microstructure image with a 1000x1000 pixels (or spatial bins) resolution, there will be millions to tens of millions of statistics depending on the local state definitions. In order to make sense of these metrics when compared with different microstructures, it is essential to utilize a feature extraction method. The method of choice here is Principal Component Analysis, which projects a dataset of 2-point statistics of multiple microstructures from millions (the statistic corresponding to each vector is treated as a dimension) to a handful

of orthogonal dimensions in order of decreasing explained variance. Singular Value Decomposition is the method of choice for computing the principal components in examples shown in this dissertation. For a centered (column means are subtracted) data matrix A with each row representing a microstructure and each column representing a statistic corresponding to a vector, the SVD is defined as:

$$A = U\Sigma V^T \tag{6}$$

where U is the orthonormal matrix representing the left singular vectors or the axis vectors corresponding to PCA embedding, Σ is the diagonal matrix of singular values and V is the orthonormal matrix representing the right singular vectors or the projection matrix from the statistics space to the principal space.

Utility of PCA in the context of microstructure datasets can be shown with an example. A toy dataset consisting of 10 microstructures per 4 different "classes" were generated resulting in a total of 40 structures. Figure 5 shows a member from each class. The classes used are: randomly distributed randomly oriented non-overlapping rectangles, randomly distributed randomly oriented non-overlapping circles, randomly distributed horizontal non-overlapping rectangles and randomly distributed vertical non-overlapping rectangles. The number of shapes in these 40 microstructures are also random. These classes were chosen due to ease of interpretability of the principal component embedding of their statistics. Figure 6 shows the principal component embedding of this dataset from PC1 vs PC2 and PC2 vs PC3 views. Note that even though PCA is uninformed of the

specific classes of each microstructure, the dataset is naturally organized into distinct regions in the principal space corresponding to each class.

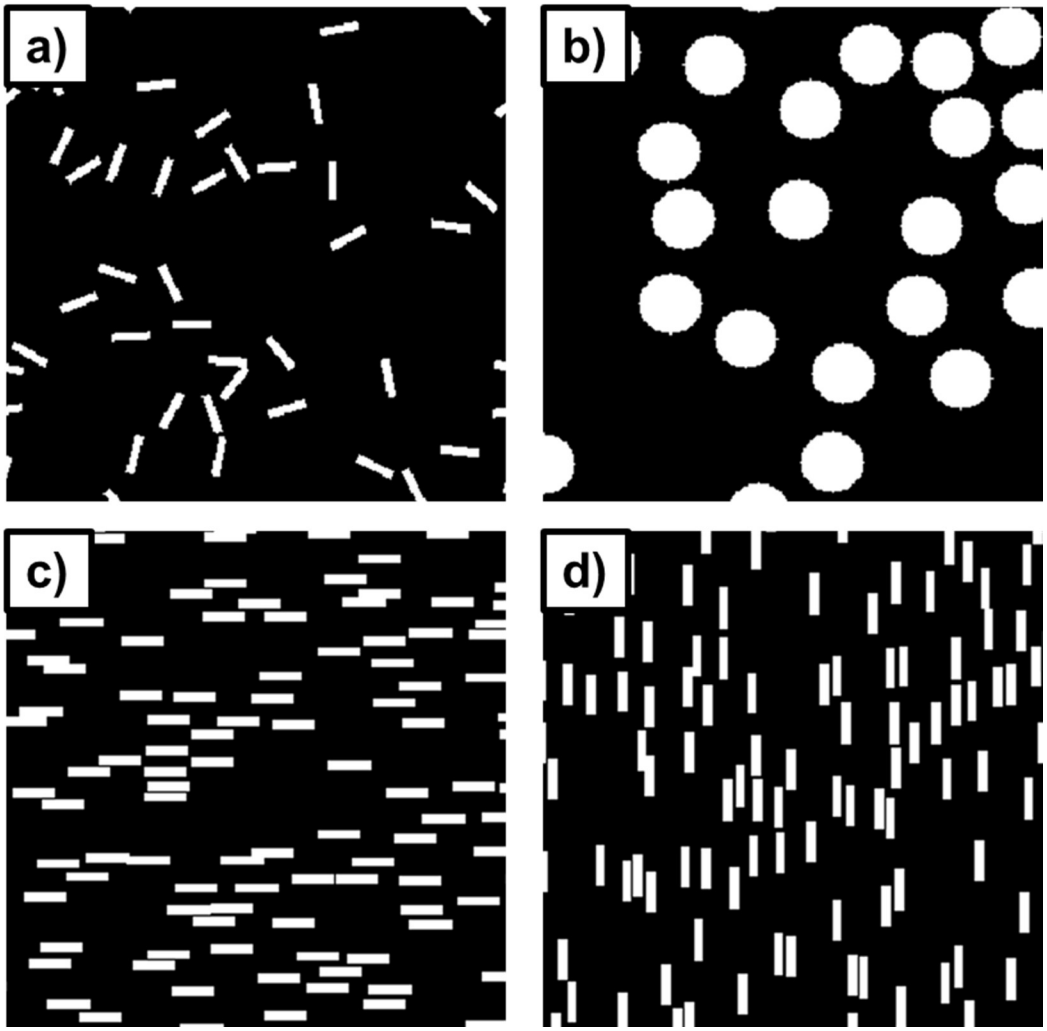

Figure 5 – A member form each class for the example data. a) randomly distributed randomly oriented non-overlapping rectangles b) randomly distributed randomly oriented non-overlapping circles c) randomly distributed horizontal non-overlapping rectangles d) randomly distributed vertical non-overlapping rectangles. The number of shapes in each of the 40 microstructures is also determined randomly.

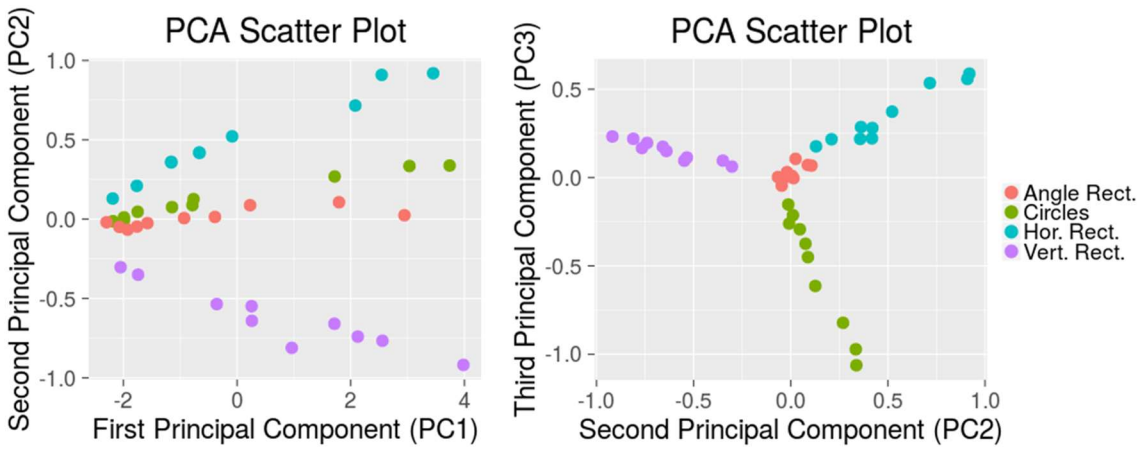

Figure 6 – Principal component embedding of the example dataset. Notice that between the first 3 principal components, all 4 classes are clearly separated.

Further insight is possible when the basis vectors for this embedding are inspected. Recall that the PCA of a centered data matrix measures the deviation from the average trend. Figure 7.a shows the average statistics for these 40 microstructures. The average trend shows preference to horizontal and vertical alignment as expected, since half of the dataset is enforced to such an arrangement. Figure 7.b shows the basis vector for the first principal component, which represents the highest variance. The first basis shows clear sensitivity to $t = 0$ vector which corresponds to volume fraction, which is not surprising as the number of objects in each structure was uncontrolled. The spread of different volume fractions can easily be seen on along the PC1 axis of Figure 6. Figure 7.c shows the basis vector for the second principal component, which appears to discriminate strict alignment in exactly horizontal and exactly vertical directions, while remaining indifferent to any other trend. This can visually be verified from Figure 6 as well, since the class that consistent only of horizontal rectangles is strictly contained to the positive side of the axis (in line with the positive weights in the basis vector in the same direction), while the class

of vertical rectangles is strictly contained in the negative side. While these two classes are very well separated, PC2 does a poor job in distinguishing the other two classes that did not impose a strict alignment. Finally Figure 7.d shows the basis vector for the third principal component, which discriminates positively to both strictly horizontal and strictly vertical alignment, and discriminates negatively to alignment in any other direction. As such, in Figure 6 both horizontal and vertical rectangle classes remain on the positive side of the axis, while the randomly oriented rectangle class is collected on the negative axis. Since circles have no particular alignment overall (or equal "alignment" in all directions), they result in an average (or 0) score on the third axis.

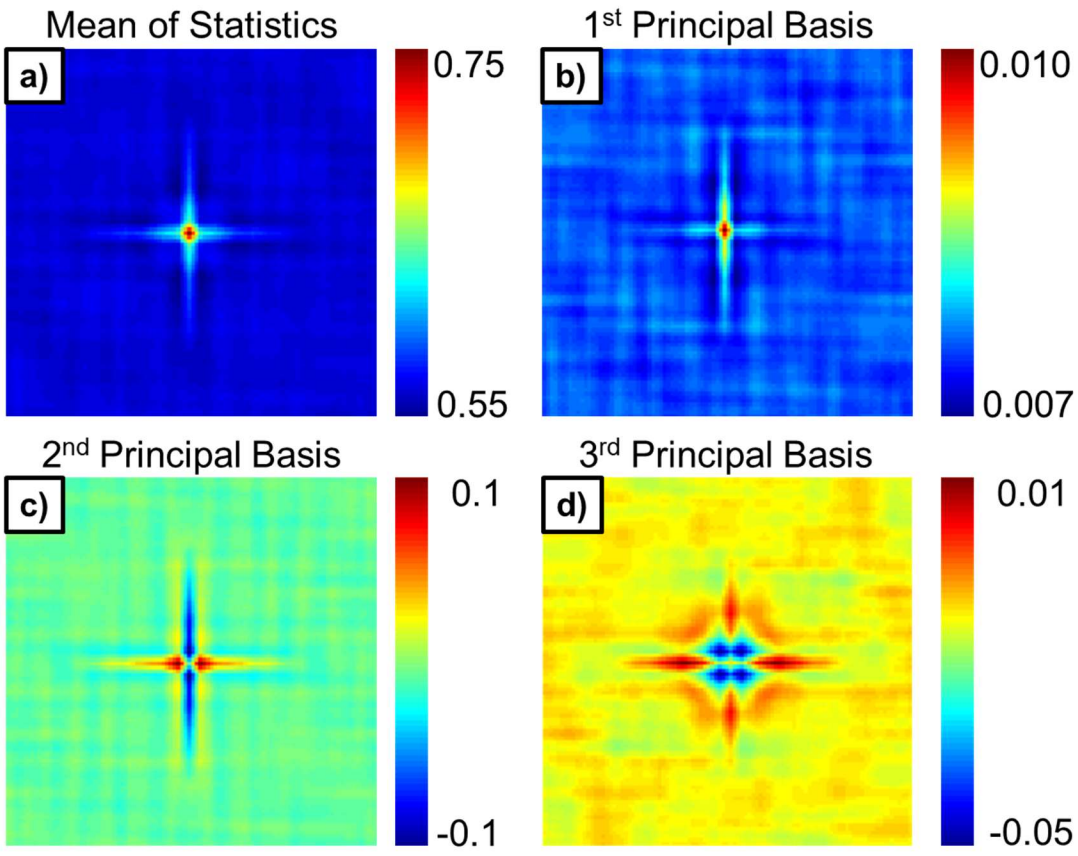

Figure 7 – a) The mean of statistics for all 40 datasets. b,c,d) Principal component basis vectors.

2.7 Reconstruction of Microstructure Function from Statistical Descriptions

So far the discussion has been entirely focused on obtaining statistical representations of microstructures. The reverse problem is unsurprisingly also of popular interest, where given a set of statistical measures it is desired to obtain the corresponding structure representation in real space (or the original image or microstructure function). While there have been studies using many statistical functions in conjunction such as the cluster functions [1], surface correlation functions [9] and lineal path functions [7, 36], these have been mostly successful in limited to niche cases and structures of specific form and shape. Reobtaining a physical representation from a complete set of 2-point statistics is analogous to the extensively studied signal processing problem of phase recovery. Given a set of two point statistics for instance, the probability of occurrence (amplitude) of each vector is known while the information necessary to pinpoint the locations (phase) of those vectors is lost. Previously, Fullwood et al. [30] adapted the well-established phase recovery method Gerchberg–Saxton algorithm to reconstruct physical representations of binary (two local states) microstructures described by a complete set of two-point statistics. Hasanabadi et al. [57] proposed an approach that estimates the information in the missing 3D orientations by exploiting the theoretical bounds and trends to the two and three point statistics based on the values in the known planes. They have then followed the Gerchberg–Saxton method to reconstruct a 3D microstructure using the complete set of statistics they have estimated.

2.8 Microstructure Generation

The fundamental assumption in a reconstruction problem is that the statistical description at hand contains a single underlying truth, and this truth needs to be reached through optimization. However, in most realistic cases where reconstructions are desirable, the statistical information at hand is deficient in some manner. These cases can include rank deficiency for statistics estimated by an established process-structure or structure-property linkage (from only knowing a few PCs and the mean), tainted/missing statistics for a particular length scale (such as those obtained experimentally as diffraction data), or perhaps most notoriously cases where statistical information is only available in a limited number of oblique 2D planes of orientation. Recognizing the extreme lack of information in the cases described above it can clearly be seen that the only reasonable approach to this problem is the generation of microstructures, instead of reconstruction. From this point of view, the statistical information is regarded as incomplete with many microstructure variations satisfying the statistical criteria with minimal error. The objective now is to create a viable microstructure instance that comes within a margin of having the same incomplete statistical description as the reference. The viability measure can come from physical constraints like manufacturability or theoretical constraints.

For instance, the last case listed above (3D image from 2D exemplars) is of significant interest in many fields and material systems where obtaining 3D structural information via current imaging technologies is either very expensive or impossible, however it is relatively straightforward to obtain 2D images from certain angles. Thus, there is a potential for high impact for any method that can successfully generate a statistically representative 3D microstructure from a given set of 2D cross sections. For

this problem Turner et al. [58] recently adapted a solid texture synthesis approach from Kopf et al. used in computer graphics to construct an estimate of a 3D structure by using the 2D images of the microstructures in real space. While the method is not guided by 2-point statistics, it was still able to produce microstructures that had very minimal error in 2-point statistics with respect to the reference in certain cases, while requiring significant improvement in others. To the best of the author's knowledge, no microstructure generation method exists in literature that can function under a spatial statistics (including RDF, SCF, LPF, CLD or cluster functions) constraint.

Computer vision literature contains numerous works in texture synthesis which can at first seem directly applicable, however a distinction needs to be drawn between the goals of microstructure generation and the goals of texture synthesis. Texture synthesis is only concerned with the generation of a smooth and visually appealing image from a given reference. For all intents and purposes, texture synthesis methods mostly *exclusively utilize the information from a sample or samples* and create variations. However, the goal of microstructure generation is to create *instances that could have come from the population that a given reference was sampled*. The information available is also very different in the two problems: a texture is a *product of imagination*, while a microstructure is a *product of a measurement of a material that was manufactured*, and as such a synthesized microstructure should also be realizable at least in theory.

In consideration of the above objectives, microstructure generators are very rare in literature, with most successful cases exploiting the constraints on a very narrow niche or a special geometric case. An example of these is the generation of images consisting of non-overlapping geometric shapes like the ones used for the discussion in the

dimensionality reduction section. Such a microstructure representation is usually a gross over simplification and only useful for theoretical arguments, however generation of randomly packed shapes is perhaps the most common family of random generation problems across all engineering and scientific domains. Yet, only a very few methods exists in literature that can solve this problem to minimal satisfaction, and none that can exert any kind of control over the nature of generation such as exact distances between shapes or allowing some range of overlap. The most promising implementation in literature is by Tschopp et al. [59], which is essentially a highly optimized brute force solution, where a random point is picked and checked for overlap until a point without overlap is found. Another example exploits the nature of polycrystalline microstructures in order to generate more instances from a reference [60, 61]. However these methods prioritize the matching of the relative spatial distribution of clusters of individual local states, without paying attention to the shapes of these clusters (i.e. a blue blob and a red blob has to be neighbors to each other but the shapes of these blobs are not important). As such, these microstructure generators are in essence solving an entirely different class of problems that are only valid in their specific application area.

CHAPTER 3. OVERVIEW OF CONTRIBUTIONS

This dissertation is organized into the following tasks:

TASK 1: Efficient Tools to Map Real Microstructures to Low Dimensional Descriptors

The main objective of this task is to expand the existing framework for structure descriptions using spatial correlations beyond the underlying trivial/simplistic assumptions in their original derivation. Microstructures of real material systems acquired through imaging techniques such as various methods of microscopy or tomography, can contain many irregularities, have elaborate local state definitions, be susceptible to observer rotations and have very high resolutions resulting in gigabytes or terabytes of data. This task deals with methods and strategies that accommodate:

1. Missing Data and Irregular Structural Domains
2. Complex/Unbounded Local States
3. Invariance to Observer Rotations
4. Very Large Data Sizes

TASK 2: Universal Microstructure Generation Framework

The main objective of this task is to formally describe a heavily extensible and scalable microstructure generation framework that can accommodate a wide array of structural manipulation tasks, which include microstructure building, modulation and evolution.

CHAPTER 4. EFFICIENT TOOLS TO MAP REAL MICROSTRUCTURES TO LOW DIMENSIONAL DESCRIPTORS

4.1 Missing Data and Irregular Domains

A key requirement to fast calculation of spatial statistics is that the data is on a uniform grid, so that convolution theorem can be exploited with the use of DFTs. Synthetically generated structural data is generally reliable and in most cases can be made to conform to a perfect rectangular domain. However experimental datasets can easily have non-rectangular boundaries, and frequently contain regions of high uncertainty or gaps within the uniform mesh. These concerns require the framework to be able accommodate an indicator function mapping the valid points within a dataset. A Boolean representation of such a function on the same grid as the dataset under consideration is generally referred to as a "mask" in the image processing community. This work incorporates the concept of masks to calculation of spatial correlations.

Consider the simplified case in Figure 8. An otherwise periodic hexagonal mesh was only accurately observed (due to perhaps, experimental difficulties) in a very irregularly bounded region. Naïve computation of spatial correlations would result in the imposition of the irregular boundary as an inherent property of the microstructure, which is a fallacy. In this situation, it is convenient to define two microstructure functions: (i) an extended microstructure function denoted as $\tilde{m}^h = m^h \odot c$, where we have introduced an additional fictitious local state (i.e., the third phase colored gray) in the masked region as well as the boundary padded regions, and (ii) a mask function denoted as c such that it

takes a value of zero for spatial bins (shown as black) in the masked regions and one (shown as white) everywhere else.

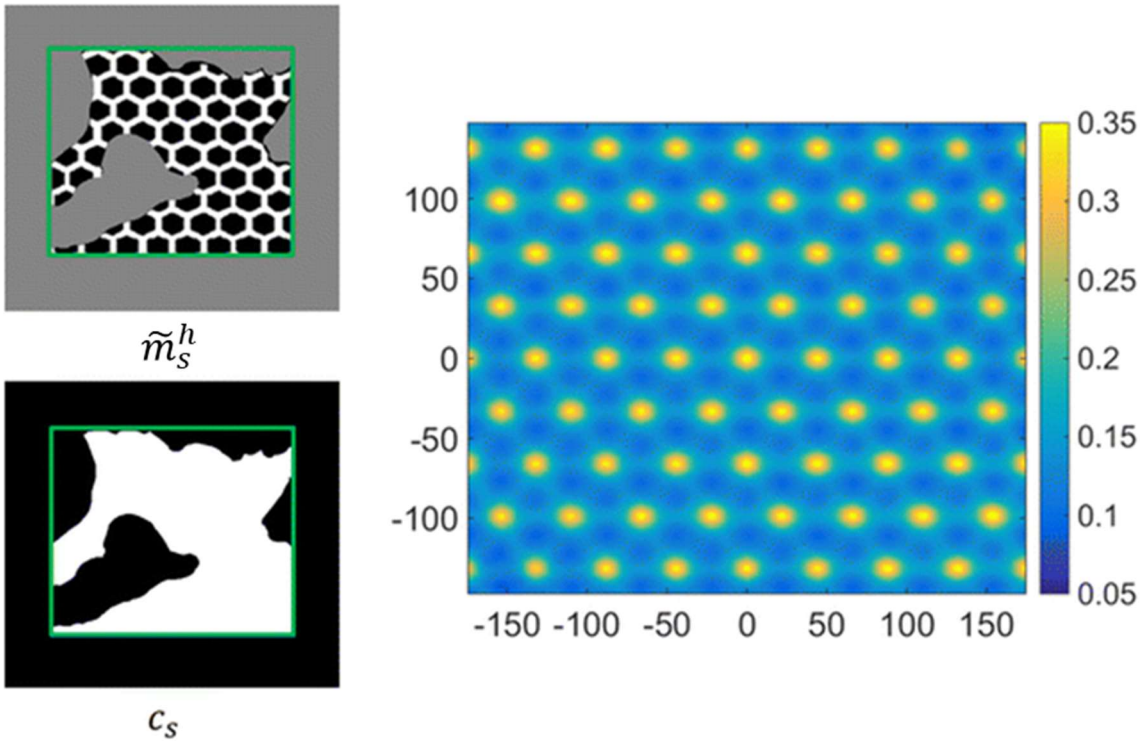

Figure 8 – Illustration of the masking strategy to compute 2-point statistics on irregular domains. The green boxes around the original microstructure are only for visualization.

Following the methodology described in the earlier sections, we compute $\widetilde{M}^h = \mathfrak{I}(\widetilde{m}^h)$ and then $\mathfrak{I}^{-1}\left(\widetilde{M}^{h*} \odot \widetilde{M}^{h'}\right)$ to accurately count of the number of successes in finding local states n and p separated by all vectors of interest (as mentioned earlier it is important to include padding if we wish to avoid the default assumption of periodicity implicit in the use of DFTs). In order to compute the 2-point statistics of interest, we simply need to divide these numbers with a suitable denominator denoting the total number of trials involved. For the masked microstructures described here, the denominator can be computed easily as $\mathfrak{I}^{-1}(C^* \odot C)$, where $C = \mathfrak{I}(c)$. This leads to the generalized form:

$$f(h,h'|t) = f_t^{\,hh'} = \frac{\sum_s \widetilde{m}_s^h \widetilde{m}_{s+t}^{h'}}{\sum_s c_s c_{s+t}} \tag{7}$$

$$f^{hh'} = \frac{\mathfrak{I}^{-1}\left(\widetilde{M}^{h*} \odot \widetilde{M}^{h'}\right)}{\mathfrak{I}^{-1}(C^* \odot C\,)} = \frac{\mathfrak{I}^{-1}\left(\mathfrak{I}(\widetilde{m}^h)^* \odot \mathfrak{I}\left(\widetilde{m}^{h'}\right)\right)}{\mathfrak{I}^{-1}(\mathfrak{I}(c\,)^* \odot \mathfrak{I}(c\,)\,)} \tag{8}$$

Figure 8 also depicts a plot of the f^{11} (white-white) autocorrelations where the computations were limited to the unmasked regions (the white region of the mask). Furthermore, there was no assumption of periodicity in this computation. However, it is seen that these autocorrelations are indeed very similar to the ones shown in Figure 2 (performed assuming periodicity and limited to a much smaller range of vectors).

It should be noted that the computed two-point statistics from the above equations, if computed assuming a periodic domain *S* and taking full advantage of the FFT algorithm, are valid only for $|t| \leq t_{\max}$, with $t_{\max}$ determined by the extent of the padding employed. Although the padding by itself increases the computational cost of the DFTs by a small amount, the cost savings in the overall computation of the two-point statistics using this strategy for non-periodic microstructures, compared to direct computation is quite significant.

The application of the concept of masking/padding for computing spatial correlations is further illustrated in Figure 9 using an example of simple two-dimensional irregular, non-periodic, domain with excluded patches (white squares). In this microstructure, let us say that we are specifically interested in the spatial correlations between the two local states colored red (h = 1) and blue (h = 2), respectively. The rest of

the spatial bins in Figure 9.a are occupied by other local states and are colored gray. In order to apply the protocols described above, we first pad the microstructure domain as shown in Figure 9.b and Figure 9.c and introduce the mask function shown in Figure 9.d. The microstructure functions m^1 and m^2 are defined such that they take values of one in red and blue voxels, respectively, and zeroes for all other voxels. The mask function, c , is created by assigning values of one to the voxels colored black, and zeroes to all other voxels. The mask function essentially marks all of the voxels in the original microstructure identified in Figure 9.a.

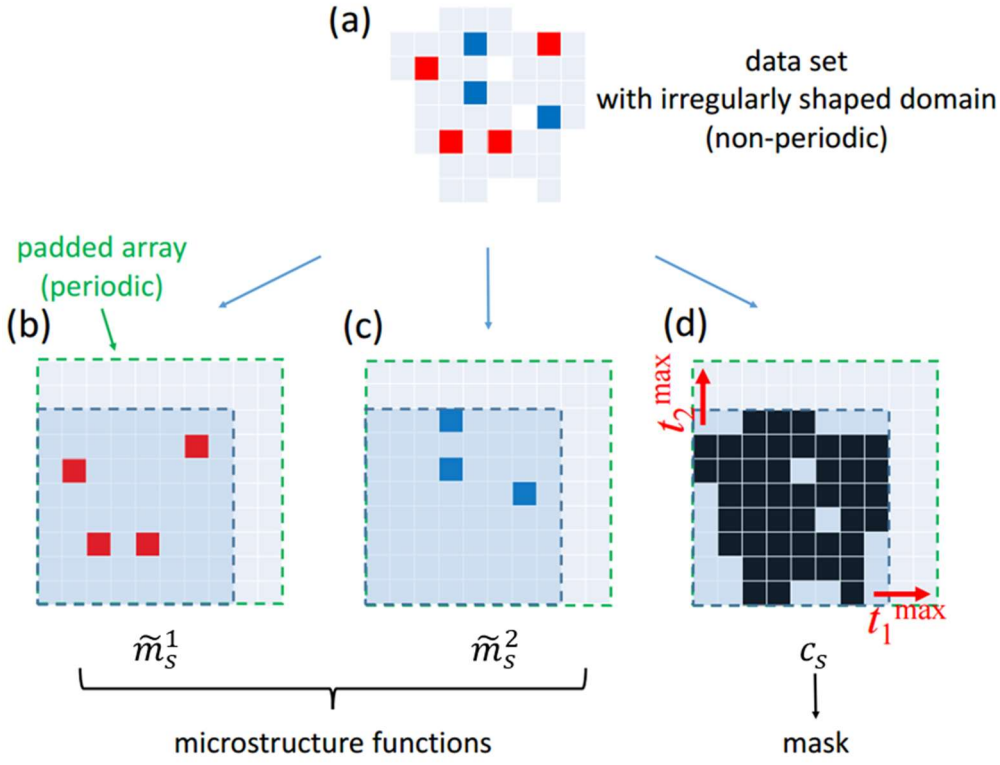

Figure 9 – Illustration of padding and masking for irregularly shaped, non-periodic, microstructure domains.

4.2 Conditional Two-Point Probability

Although two-point correlations capture the probability of simultaneously finding two selected local states in space for a specified vector of translation r, it is sometimes desirable to calculate the conditional probabilities between the two local states, namely the probability of finding one local state h', at a location r from another given local state h. This conditional probability is defined as:

$$f_t^{h'|h} = f(h'\,|h\,,t) = \frac{f(h,h'|t)}{f(h\,|t)} = \frac{\sum_s \widetilde{m}_s^h \widetilde{m}_{s+t}^{h'}}{\sum_s c_s c_{s+t}} \tag{9}$$

$$f^{h'|h} = \frac{\Im^{-1}\left(\Im(\widetilde{m}^h)^* \odot \Im\left(\widetilde{m}^{h'}\right)\right)}{\Im^{-1}(\Im(\widetilde{m}^h)^* \odot \Im(\,c\,))} \tag{10}$$

Note that the only difference in expression with respect to the joint probability is in the denominator (in both definitions this normalization term computes the number of valid trials). As illustrated graphically in Figure 9c, the conditional probability uses the number of vectors t with tails on 1 (red) and heads in the original microstructure domain as the normalization factor, whereas the joint probability uses the total number of vectors t regardless of the local states present at their heads or tails (Figure 9b). It is easy to see that the conditional two-point statistics will have values between 0 and 1. When two local states *n* and *p* are perfectly correlated in a microstructure, the conditional probability is equal to 1; when there is no correlation between the two local states, the conditional probability is equal to the expected value of m^p, which is equal to the volume fraction of p in the microstructure domain.

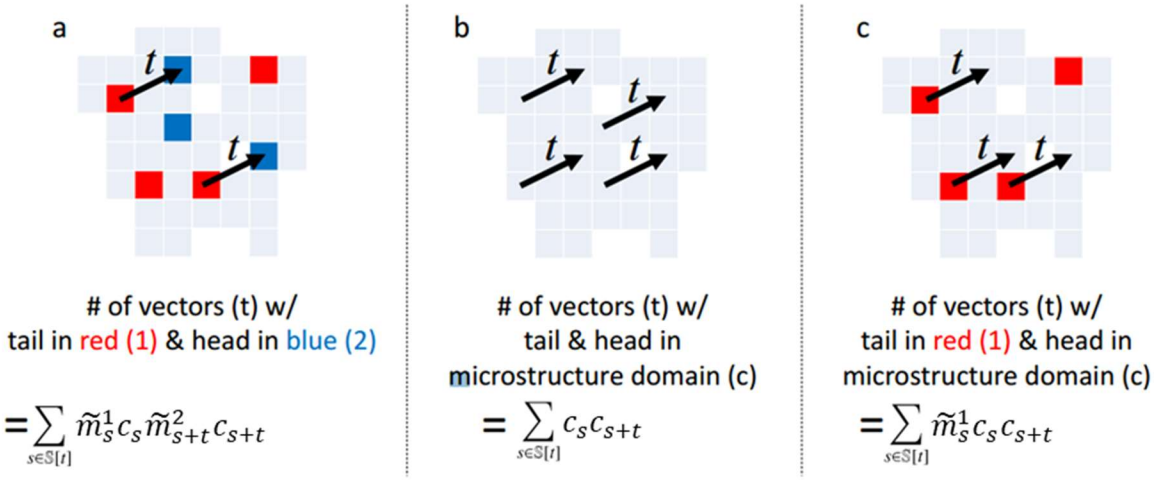

Figure 10 – (a) The joint probability finds the number of vectors with tails and heads on red and blue voxels, respectively, and is (b) normalized by the number of valid trials possible in the given microstructure domain. The conditional probability uses the same numerator (a), but the normalization factor on the denominator is replaced by the number of vectors with tails on a red voxel and heads anywhere in the original microstructure domain (c).

4.3 Complex/Unbounded Local States

The current framework for spatial statistics can accommodate multiple local states with unique labels/colors, as well as fractional local states that indicate a mixture of multiple local states within a single voxel. However, there exists cases where a single voxel can be occupied by multiple independent local states, similar for example to the RGB channels of a color image. Furthermore, some local states are also defined across a range of values but do not represent fractional mixture information, including local states with negative values.

One promising avenue of analysis for such datasets is to quantify how correlated two spatial fields are with respect to each other in ways that allow positive and negative correlations (note that probability-based concepts discussed thus far do not allow such an

interpretation in a direct manner). Pearson product-moment correlation coefficient is a commonly used value in statistics to quantify such correlations. It is first useful observe the following relation:

$$f_t^{\,hh'} = \mathrm{E}\left[m_s^h m_{s+t}^{h'}\right] \tag{11}$$

We can then define a two-point Pearson correlation function as:

$$\rho_t^{hh'} = \frac{Cov\left(\widetilde{m}_s^h, \widetilde{m}_{s+t}^{h'}\right)}{\sigma_t^h \sigma_t^{h'}} = \frac{\mathrm{E}\left[\widetilde{m}_s^h \widetilde{m}_{s+t}^{h'}\right] - \mathrm{E}[\widetilde{m}_s^h]\mathrm{E}\left[\widetilde{m}_{s+t}^{h'}\right]}{\sigma_t^h \sigma_t^{h'}} = \frac{f_t^{hh'} - \mu_t^h \mu_t^{h'}}{\sigma_t^h \sigma_t^{h'}} \tag{12}$$

$$\rho^{hh'} = \frac{f^{hh'} - \mu^h \odot \mu^{h'}}{\sigma^h \odot \sigma^{h'}} \tag{13}$$

where, for any given t:

$$\mu_t^h = f_t^h = \frac{\sum_s \widetilde{m}_s^h c_{s+t}}{\sum_s c_s c_{s+t}} \tag{14}$$

$$\mu_t^{h'} = f_t^{h'} = \frac{\sum_s c_s \widetilde{m}_{s+t}^{h'}}{\sum_s c_s c_{s+t}} \tag{15}$$

$$\sigma_t^h = \sqrt{\frac{\sum_s \widetilde{m}_s^h \widetilde{m}_s^h c_{s+t}}{\sum_s c_s c_{s+t}} - \left(\frac{\sum_s \widetilde{m}_s^h c_{s+t}}{\sum_s c_s c_{s+t}}\right)^2} \tag{16}$$

$$\sigma_t^{h'} = \sqrt{\frac{\sum_s c_s \widetilde{m}_{s+t}^{h'} \widetilde{m}_{s+t}^{h'}}{\sum_s c_s c_{s+t}} - \left(\frac{\sum_s c_s \widetilde{m}_{s+t}^{h}}{\sum_s c_s c_{s+t}}\right)^2} \tag{17}$$

and in their matrix form:

$$\mu^h = \frac{\Im^{-1}(\Im(\widetilde{m}^h)^* \odot \Im(\mathrm{c}))}{\Im^{-1}(\Im(\mathrm{c})^* \odot \Im(c))} \tag{18}$$

$$\mu^{h'} = \frac{\Im^{-1}\left(\Im(c)^* \odot \Im\left(\widetilde{m}^{h'}\right)\right)}{\Im^{-1}(\Im(\mathrm{c})^* \odot \Im(c))} \tag{19}$$

$$\sigma^h = \sqrt{\frac{\Im^{-1}(\Im(\widetilde{m}^h \odot \widetilde{m}^h)^* \odot \Im(\mathrm{c}))}{\Im^{-1}(\Im(\mathrm{c})^* \odot \Im(c))} - \left(\frac{\Im^{-1}(\Im(\widetilde{m}^h)^* \odot \Im(\mathrm{c}))}{\Im^{-1}(\Im(\mathrm{c})^* \odot \Im(c))}\right)^2} \tag{20}$$

$$\sigma^{h'} = \sqrt{\frac{\Im^{-1}\left(\Im(c)^* \odot \Im\left(\widetilde{m}^{h'} \odot \widetilde{m}^{h'}\right)\right)}{\Im^{-1}(\Im(\mathrm{c})^* \odot \Im(c))} - \left(\frac{\Im^{-1}\left(\Im(c)^* \odot \Im\left(\widetilde{m}^{h'}\right)\right)}{\Im^{-1}(\Im(\mathrm{c})^* \odot \Im(c))}\right)^2} \tag{21}$$

Assumption of a global mean for each phase h and h', without accounting for shifts in the image corresponding to head and tail of placed vectors, will simplify this calculation quite a bit, while mostly preserving the desired effect of eliminating mean contributions to calculated correlations. Under these conditions we can define:

$$\mu^h = \frac{\sum_s \widetilde{m}_s^h}{\sum_s c_s} \tag{22}$$

$$\mu^{h'} = \frac{\sum_s \tilde{m}_s^{h'}}{\sum_s c_s} \tag{23}$$

$$\sigma^{h} = \sqrt{\frac{\mathfrak{F}^{-1}(\mathfrak{F}((\tilde{m}^{h} - \mu^{h}) \odot (\tilde{m}^{h} - \mu^{h}))^{*} \odot \mathfrak{F}(\mathrm{c}))}{\mathfrak{F}^{-1}(\mathfrak{F}(\mathrm{c})^{*} \odot \mathfrak{F}(c))}} \tag{24}$$

$$\sigma^{h'} = \sqrt{\frac{\mathfrak{F}^{-1}\left(\mathfrak{F}(\mathrm{c})^{*} \odot \mathfrak{F}\left((\tilde{m}^{h'} - \mu^{h'}) \odot (\tilde{m}^{h'} - \mu^{h'})\right)\right)}{\mathfrak{F}^{-1}(\mathfrak{F}(\mathrm{c})^{*} \odot \mathfrak{F}(c))}} \tag{25}$$

This simplified definition is used in any examples and discussions hereafter.

In comparison with the two-point spatial correlations described earlier, the two-point Pearson correlation defined above incorporates two main changes: (i) the global means are subtracted from the spatial fields of interest before performing the correlation, and (ii) the correlations are normalized by the standard deviations. The central advantage of introducing these modifications is that the two-point Pearson correlations are valued between 1 and −1, with $\rho = 1$ meaning perfection correlation, $\rho = -1$ meaning perfect anti-correlation, and $\rho = 0$ indicating that the two functions are uncorrelated in space.

4.4 Rotationally Invariant Spatial Statistics

Yet another potential concern in application of the concept of two point statistics is the sensitivity to observer rotations. Consider for example the micrographs presented in Figure 11. Two samples of similar microstructures were imaged by a microscope. However, the view angle of the images are different, perhaps simply due to the way the

samples were mounted. The current framework requires explicit correction for such cases, and otherwise will assume the orientation difference in the microstructures to be an inherent feature, resulting in the identification of the microstructures in Figure 11.a and Figure 11.b to be drastically different from each other.

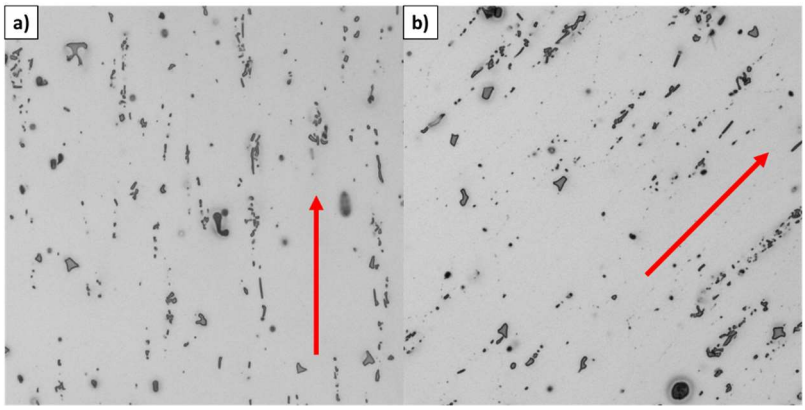

Figure 11 – An example demonstrating the potential shortcomings of two-point statistics in datasets containing observer rotations.

Another complication of the same kind can be observed when microstructure data from two different sources are compared for similarities. Figure 12 shows an example where an experimentally obtained dataset in Figure 12.a is compared to the result of a simulation of the same system in Figure 12.b. Without a manual correction, these two structures will be identified to contain drastically different patterns, when in fact they have reasonably similar patterns. It is possible to imagine many other cases where this shortcoming can adversely affect analysis.

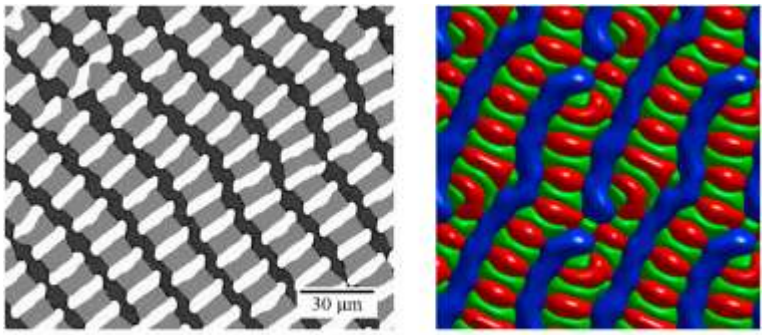

Figure 12 – Example of a case when two-point statistics may classify similar looking microstructures as being very different. Reproduced with permission from [62].

Typical work arounds to these cases either involve manual corrections via rotations, which may not be determined with reliable accuracy in most real cases, or resorting to the use of radial distribution functions. While RDFs are invariant to observer rotations, they ignore the relative spatial arrangements within a structure as well as observer rotations. Therefore, an alternative is desirable which retains the angular resolution of patterns contained in two point statistics, while discarding the effects of a rotation of the observer frame of reference.

Recall the expression for Discrete Fourier Transform:

$$X_k = \sum_{n=0}^{N-1} x_n e^{-i2\pi kn/N} \tag{26}$$

where x_n is the data in the real space and X_k is the data in Fourier space. According to the Shift Theorem, when a periodic shift Δ is introduced to the data in real space we can trace its affect in the Fourier space as:

$$\bar{X}_k = \sum_{n=0}^{N-1} x_{n-\Delta} e^{-i2\pi kn/N} \tag{27}$$

$$\bar{X}_k = \sum_{m=-\Delta}^{N-1-\Delta} x_m e^{-i2\pi k(m+\Delta)/N} \qquad (m = n - \Delta) \tag{28}$$

$$\bar{X}_k = e^{-i2\pi k\Delta/N} \sum_{m=0}^{N-1} x_m e^{-i2\pi km/N} \tag{29}$$

Notice that a shift in real space only affects the phase of the Fourier coefficients. It follows:

$$|X_k| = |\bar{X}_k| \tag{30}$$

Utilizing this property, we can render two point statistics invariant to observer rotations by calculating DFTs along the angular dimensions at discrete values of radii, eliminating the phase information, then returning to real space via IDFTs. For a 2D case this amounts to:

$$f_{t_x,t_y}^{hh'} \rightarrow f_{r,\theta}^{hh'} \tag{31}$$

$$F_{r,k}^{hh'} = \sum_{\theta=0}^{\mathbf{N}_\theta - \mathbf{1}} f_{r,\theta}^{hh'} e^{-i2\pi k\theta/\mathbf{N}_\theta} \tag{32}$$

$$f_{r,\psi}^{hh'} = \frac{1}{\mathbf{N}_\theta} \sum_{k=0}^{\mathbf{N}_\theta - \mathbf{1}} \left| F_{r,k}^{hh'} \right| e^{i2\pi k\psi/\mathbf{N}_\theta} \tag{33}$$

which can be trivially extended to 3D cases.

We can demonstrate this concept with an example. Consider the example microstructure shown in Figure 13.a. A representation of the two point autocorrelation of the white phase calculated using the existing FFT framework is shown in Figure 13.b. In order to operate on a subset of these statistics defined by a constant radius using FFTs, it is more convenient to cast them in to polar coordinates. This can easily be achieved with reasonable accuracy by interpolation. For the current discussion, linear interpolation will be sufficient. Figure 13.c shows the same statistics in a polar grid, discretized by 1 pixel steps in radius and 1 degree steps in angle. We can now take Fourier transforms along the constant radius grids and eliminate the phase information. Figure 13.d shows the resulting rotationally invariant representation of two point statistics in real space.

An immediate instinct is to assume that Figure 13.d is simply a rigid rotation of Figure 13.b. However, close inspection will reveal that this is definitely not the case. It is important to recognize that Figure 13b is in the Cartesian coordinates while Figure 6 is in polar coordinates. Also, the statistics corresponding to the angle of zero in Figure 6d will be subsequently interpreted as the radial distribution function (RDF). This function is not easily observed in Figure 13b. The non-equivalence between these plots can be easily understood by noting that the shifts obtained through the DFT at each constant value of the radius are independent of each other. Consequently, barring trivial structures, it is impossible to achieve this effect by a rigid rotation. It is also important to note that while the current example allows easy interpretation in the form of an overall rotational trend in the clockwise direction, other examples may include vastly varying angles of effective rotation in both clockwise and counter-clockwise directions between different radius values.

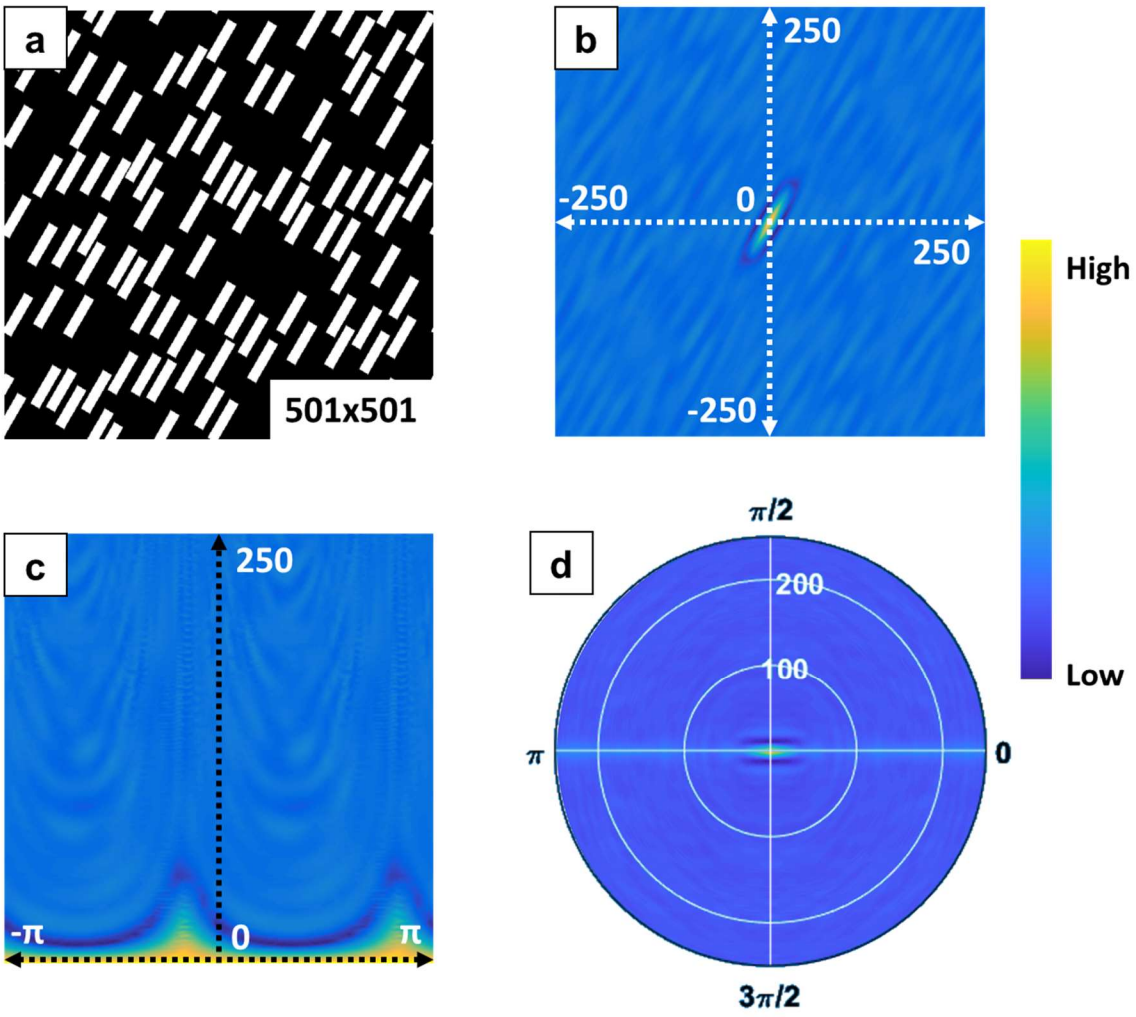

Figure 13 – a) An example microstructure b) its 2-point statistics in Cartesian coordinates c) its 2-point statistics in polar coordinates. d) The rotationally invariant 2-point statistics of the example structure in Cartesian coordinates.

We can also determine how the newly defined rotationally invariant 2-point statistics compare to RDFs:

$$RDF_r^{hh'} = \frac{1}{\mathbf{N}_\theta} \sum_{\theta=0}^{\mathbf{N}_\theta - \mathbf{1}} f_{r,\theta}^{hh'} = \frac{1}{\mathbf{N}_\theta} F_{r,0}^{hh'} \tag{34}$$

Figure 14 shows a plot of $\frac{1}{\mathbf{N}_\theta} F_{r,k}^{hh'}$ for the example structure in Figure 13.a, centered at the 0 frequency for convenience. The central line, also shown from a profile view in Figure 15 corresponds to the RDF within machine precision, whereas values at any other frequency in Figure 14 reflect information not captured by the RDF.

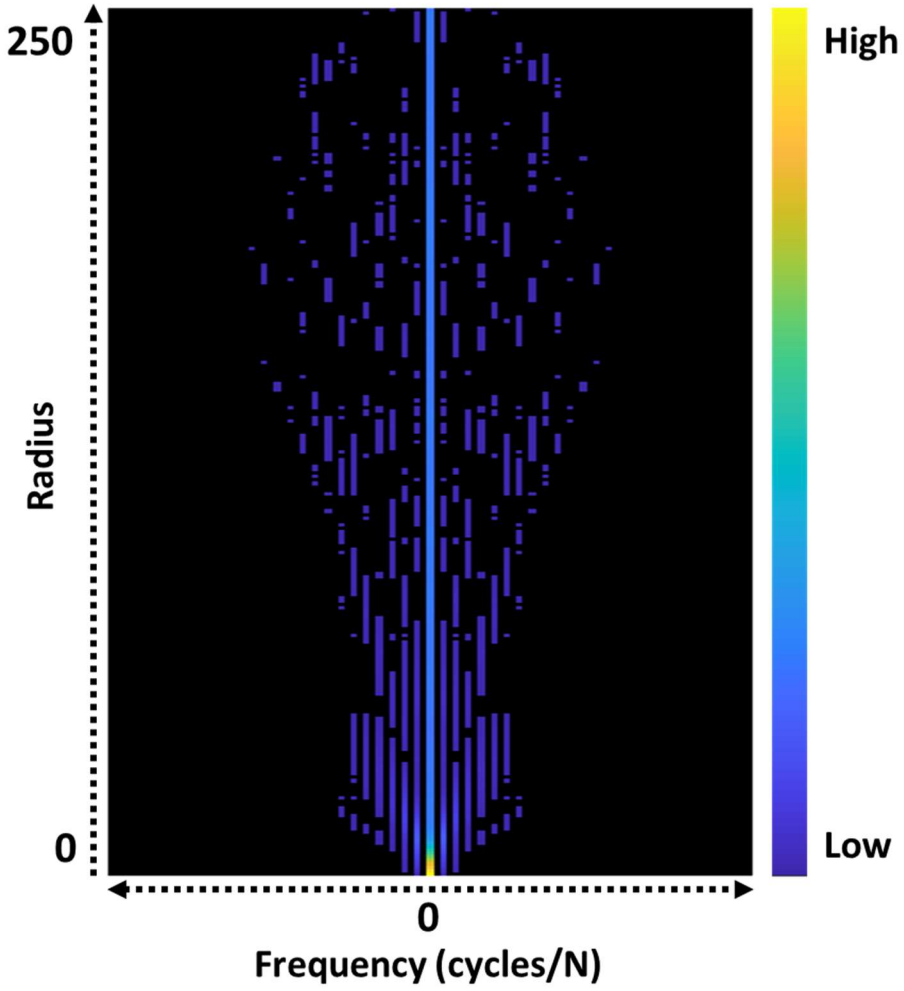

Figure 14 – Radius vs frequency space representation of 2-point statistics in polar coordinates.

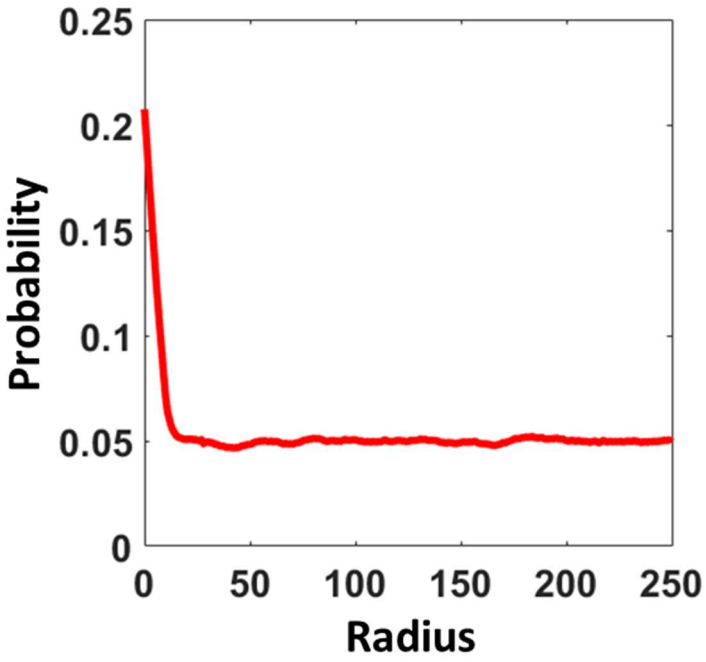

Figure 15 – The line profile of the 0 frequency across all radii from Figure 14.

Another example is in order, to show that the rotationally invariant statistics achieve the intended results in distinguishing between different patterns, while identifying the rotated variations of a pattern as similar. For this experiment a total of 400 artificial patterns are generated with the following composition: 20 instances of randomly placed non-overlapping rectangles all oriented at the same angle, which is changed by increments of 10 degrees in the range of 0 to 170 degrees, resulting in a total of 360 patterns; 20 instances of randomly placed non-overlapping rectangles of random orientations within a pattern; 20 instances of randomly placed chamfered circles. The dataset contains 20 classes in total, 18 of which is the same pattern with different rotations and 2 of which are unique. All 400 patterns contain the same volume fraction of the white phase. A visualization of one member from each class is presented in Figure 16. Figure 17 shows visualizations of the

corresponding traditional two-point auto correlations of the white phases for each structure in Figure 16, while Figure 18 shows the visualizations of the corresponding rotationally invariant two-point statistics. The slight differences in the statistics of the similar patterns can be attributed to the fact that they are generated from scratch for each instance and are not simply rotations of the same image. Figure 19 contains a visualization of the PCA embedding of the 400 instances using the traditional statistics. The embedding is well organized that one can intuitively make the associations with the different classes of microstructures employed in this case study. Notice that the 18 classes of rotated rectangle patterns are identified individually. Figure 20 shows the PCA embedding using rotationally invariant statistics, which now shows all 18 rotated rectangle classes lumped together as a single class, while still retaining the ability to distinguish the other 2 unique classes. The formalism defined here for rotational invariance in two point statistics clearly achieve the desired result.

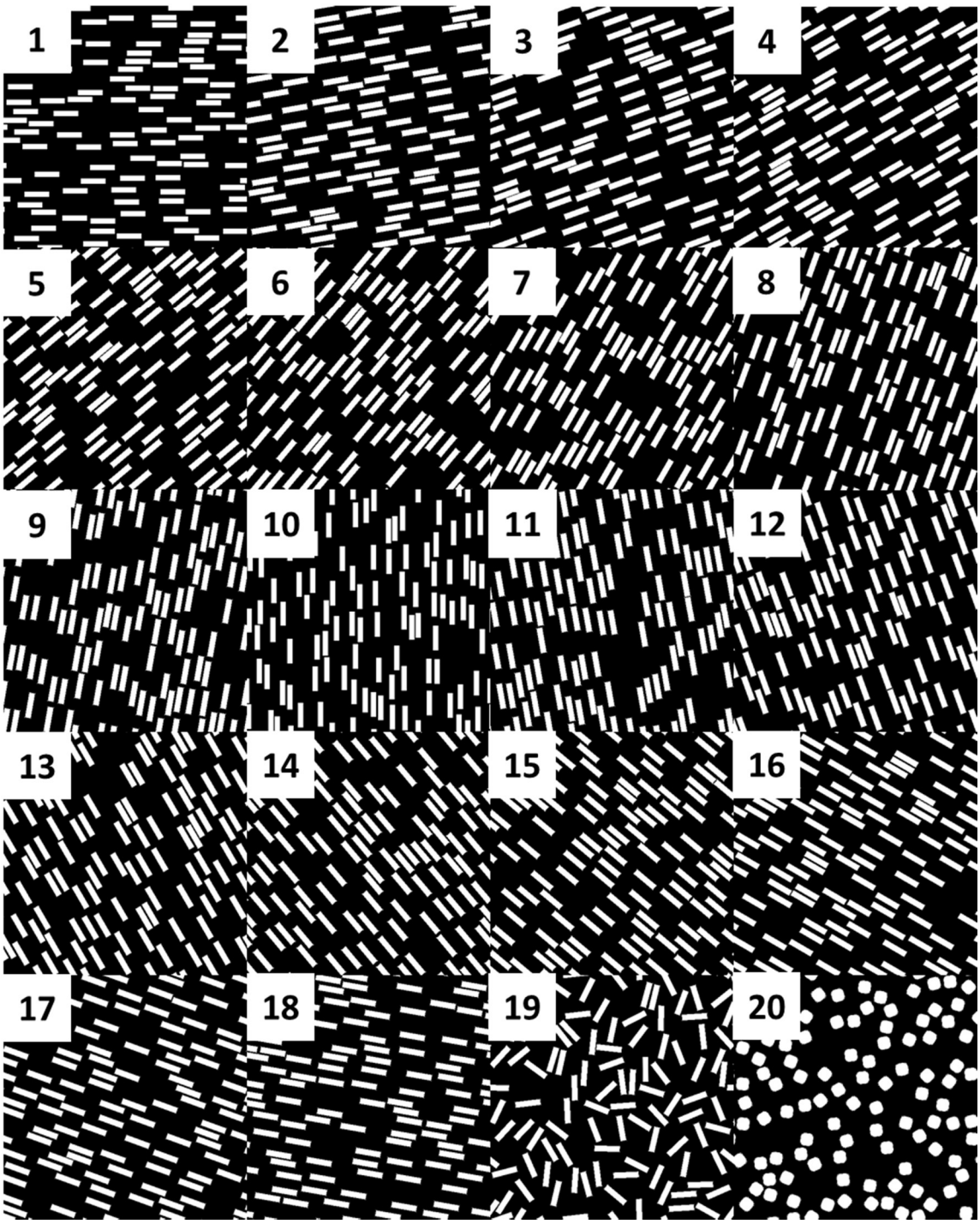

Figure 16 – Visualizations of microstructures from each of the 20 classes.

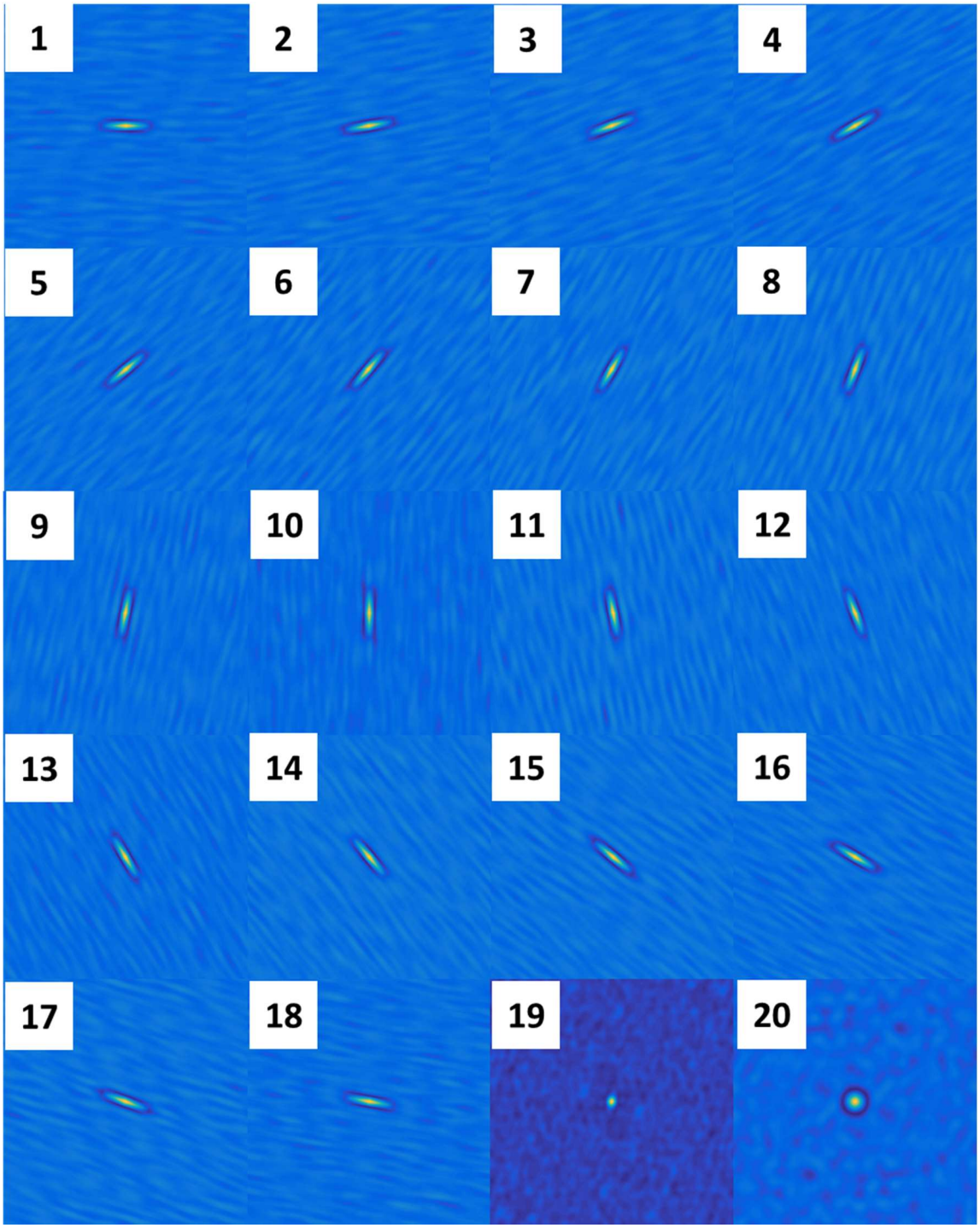

Figure 17 – Visualizations of the 2-point statistic of microstructures from each of the 20 classes.

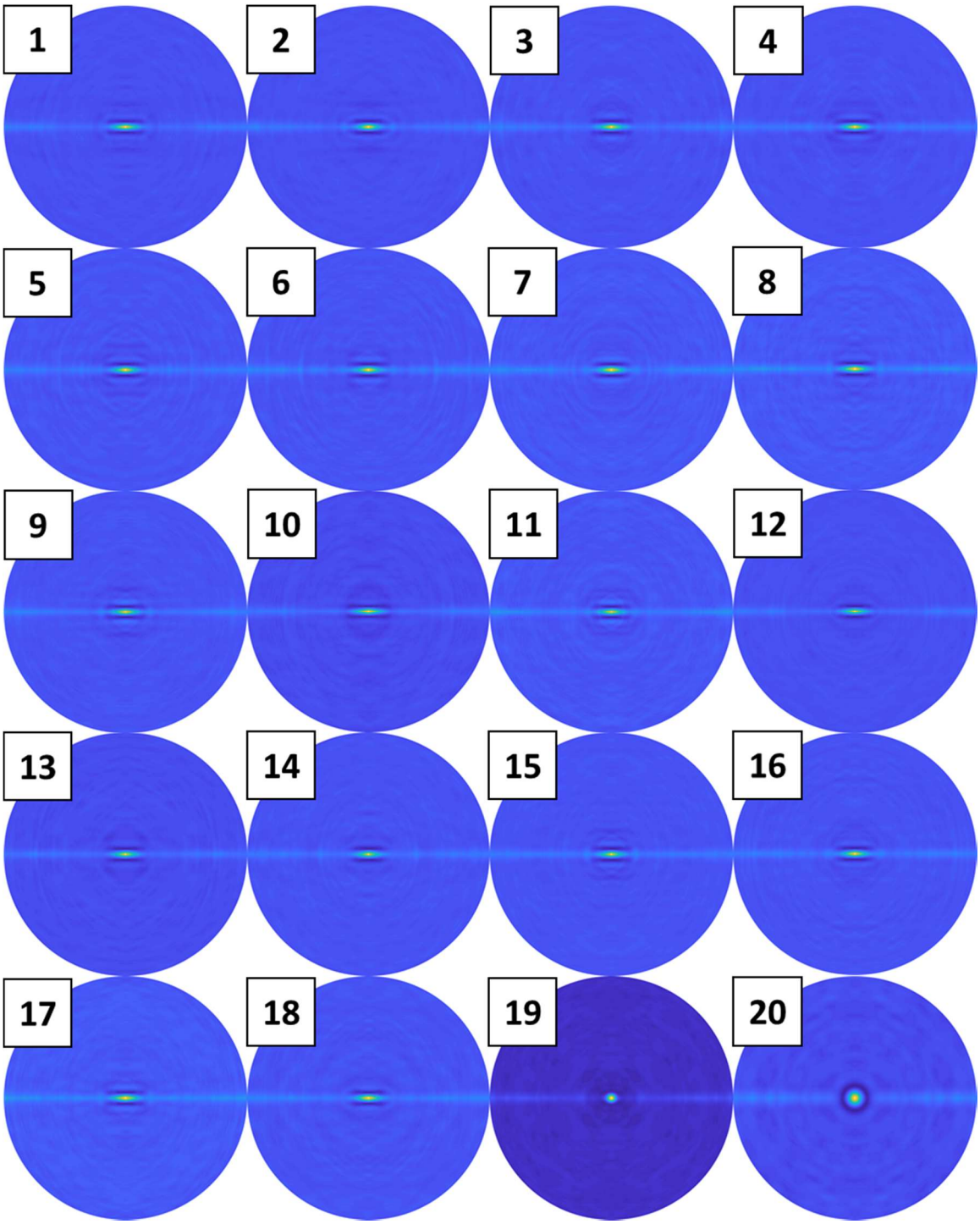

Figure 18 – Visualizations of the rotationally invariant 2-point statistic of microstructures from each of the 20 classes.

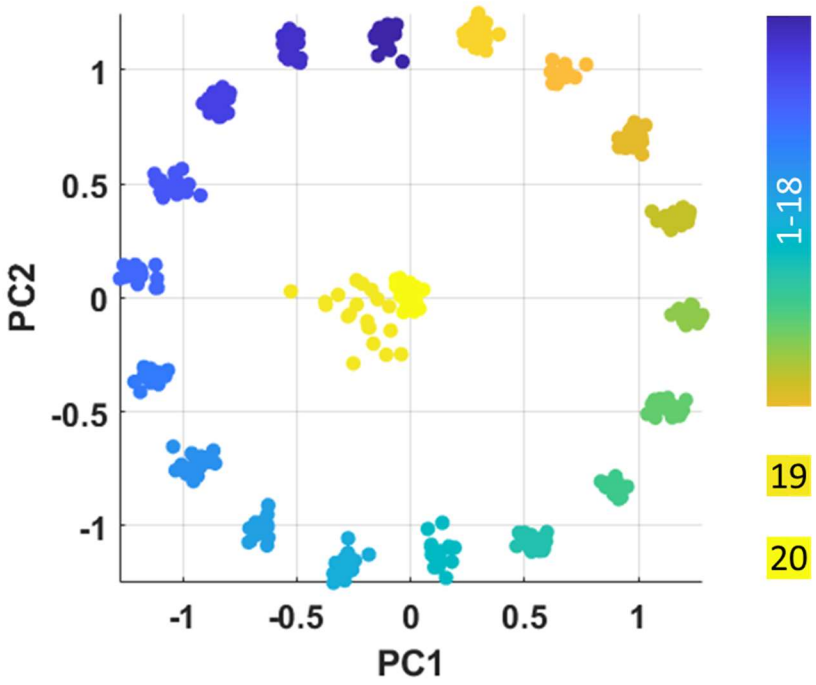

Figure 19 – PCA embedding of the 400 microstructures using traditional 2-point statistics.

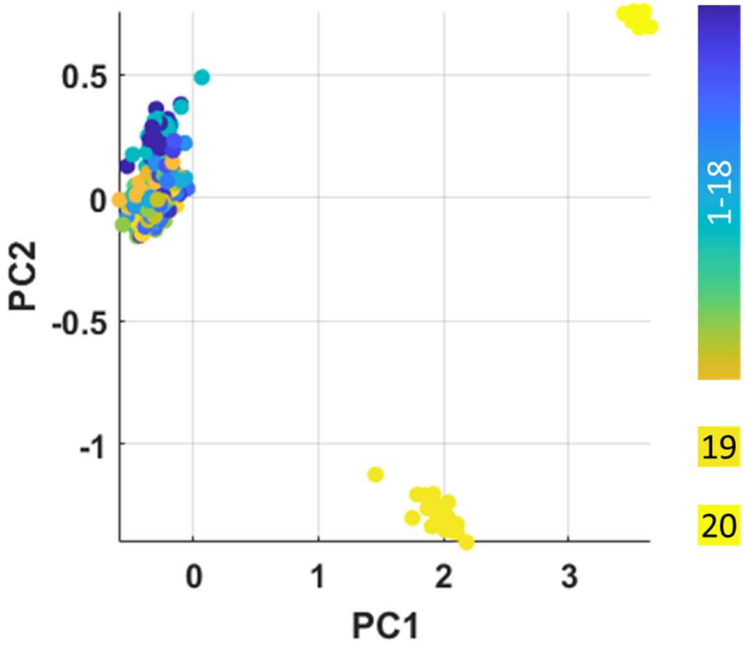

Figure 20 – PCA embedding of the 400 microstructures using rotationally invariant statistics.

4.5 Strategies to Accommodate Large Data Sizes

A final consideration here is the accommodation of datasets that are too large to fit in the memory of a personal computer during the calculation of two point statistics or PCA. This is of importance due to the fact that many engineering researchers neither have access to HPC resources nor possess the knowhow to properly utilize said resources. Furthermore, many experimental datasets require extensive preparation and down time for acquisition, meaning collected datasets are usually instanced and are not streamed.

The analysis methods described in this thesis do not require parallelism to be effective under these considerations and will execute in mere seconds for the overwhelming majority of cases, since they either have O(SlogS) runtime complexities or have to only be executed once (PCA). Therefore, memory mapping is an ideal solution to enable the everyday engineer to perform these analyses on the nearest resource while being able to accommodate large datasets. The design specifics of memory mapping implementations are beyond the scope of this thesis. Instead, strategies will be highlighted here that will enable a divide and conquer approach.

Calculation of two point statistics utilizes convolutions by FFTs, which lend themselves to a divide and conquer approach fairly trivially. For a 2D dataset with coordinates indexed by $x_1$and x_2, the Fourier transform will be taken sequentially as:

$$M_{k_1,x_2} = \sum_{x_1=0}^{N_1-1} m^h_{x_1,x_2} e^{-i2\pi k_1 x_1/N_1}$$
$$M_{k_1,k_2} = \sum_{x_2=0}^{N_2-1} M^h_{k_1,x_2} e^{-i2\pi k_2 x_2/N_2} \tag{35}$$

only requiring access to 1 row or column at a time. The convolution operation itself is an element-wise product in the Fourier domain, only requiring access to 2 elements at a time. Both operations can be batched adaptively for available memory for an efficient memory mapped implementation.

Depending on the number of microstructures and choice of local states, it is easily possible to hit the memory limit of most personal computers or even small clusters for realistic datasets during the calculation of PCA. Typical problems can have thousands of microstructures, with millions to billions of dimensions each corresponding to a choice of *h*, *h'* and *t*. To address this issue, the common formulation of Kernel PCA can be exploited with a linear kernel to allow partial access to the data. Given $A = USV^T$, the linear kernel PCA follows:

$$AA^T = USV^T VS^T U^T = US^2 U^T \tag{36}$$

which requires the explicit formation of the covariance matrix AA^T. This matrix can be formed while accessing only 2 rows of the matrix A at any given time. In most realistic cases AA^T is small enough that no further memory mapping is required during the calculation of PCA. The basis vectors can be obtained by:

$$V^T = SU^T A \tag{37}$$

which can be computed as a vector-matrix product accessing only 1 column of U and 1 column of A at a time. Both operations allow batched execution for efficiency.

4.6 Case Study: Coarsening of Dendritic Structures

Dendrites are one of the most frequently seen microstructures during solidification of metals and alloys. The morphologies of these dendritic microstructures can strongly affect the physical, chemical and mechanical properties of the subsequent products of many metallurgical technologies, such as casting [63]. This makes the study of the evolution of dendritic microstructures a crucial part in predicting and controlling the properties of metallic materials. However, like many other microstructures present in nature, dendrites are complex structures. They have irregular shapes, and are often highly interconnected. Moreover, as most real problems demand three-dimensional or even four-dimensional (including time) considerations, it makes it practically difficult to visualize the data, and even more so to analyze it. As a result, two-dimensional sections of the data are usually made in order to obtain salient attributes of the three-dimensional dataset, such as the secondary and tertiary arm spacings (λ_2, λ_3) of the dendrite [64]. As the system size grows larger, the dataset may become unwieldy to handle as a whole, and hence data processing and interpretation must be done in a piece by piece manner. This adds significantly to the challenges in arriving at meaningful global (bulk) descriptors of the microstructure that are also of high value in correlating to its performance characteristics.

A number of statistical quantities have been utilized in the past to serve as microstructure descriptors. For example, the interface shape distribution (ISD) [65-67] and the interface normal distribution (IND) [68-70] have been used to capture the statistical distribution of the shapes and orientations of the interface areal elements. However, these descriptors capture the desired information in a grossly simplified manner by treating each measurement at a randomly selected point in the microstructure as a completely independent sample (i.e. as one-point statistics) and do not capture information on the spatial distribution of the selected features in a representative domain (volume) of the microstructure. Radial distribution functions have also found some use in previous studies, however shortcomings of the RDF in comparison to angularly resolved two point statistics has been extensively described in earlier sections.

This case study will demonstrate the application of the new concepts presented in Task 1 on experimental datasets of a dendritic microstructures in Al-Cu alloys undergoing coarsening. The datasets consists of time snapshots of 4 experiments on samples of directionally solidified Al-Cu alloys. The raw material was cut into cylinders that are 3 to 5 cm in length by 1 cm diameter, with the axial direction approximately parallel to the direction of solidification. During the experiments, samples were heated to 558°C (5°C above the eutectic temperature), forming a liquid-solid mixture with a constant amount of the two phases with 28/35/55/80% solid volume fractions. The samples were then held at that temperature, in which state the interfaces between liquid and solid can rapidly coarsen to decrease the total interfacial area and thus total energy of the system. The isothermal condition was held for 2 to 15 hours while the sample was being scanned by X-ray

computed tomography (XCT) every 50 to 250 seconds. The XCT scans were performed at the TOMCAT beamline at the Swiss Light Source.

This case study was chosen due to the fact that it is an experimental dataset (i.e. the underlying problems are real and to be expected in many other cases) which requires the use of every method described/derived in this chapter to be properly analyzed:

- The regions of interest are the solid-liquid interfaces, which requires the utilization of masks to select **highly irregular sub-regions** of rectangularly bounded datasets.

- The quantities of interests (i.e. phases) are interface curvature and velocity, both of which have **continuous values** that can be negative.

- The 4 experiments are each in contained in cylindrical volumes that are **susceptible to observer rotations** around the cylindrical axis.

- The dataset contains 500 time steps, with each time frame consisting of on average an 1100x700x700x3 voxel double precision array, resulting in statistics computations normally requiring upwards of **50GB memory**, and PCA computations normally requiring around **1TB memory**.

4.6.1 Data Preparation and Pre-Processing

In order to determine the interface locations between the liquid and solid phases with sufficient accuracy and hence be able to calculate interfacial curvature and velocity, the grayscale data reconstructed from XCT scans was segmented using a level-set based method [71]. In this method, a signed distance function (SDF) is used to implicitly define the interface locations. An SDF is an array of data in which the value at each voxel gives the distance to the nearest interface, with opposite signs on the two sides of the interface. In this study, positive values are used to represent voxels inside the solid phase, and negative values the liquid phase. Therefore, the interface is defined as where the SDF crosses zero. Two-dimensional slices of the SDF for a selected time snapshot are shown in Figure 21, with zero value contour lines marked by white lines. In 3D, the SDF = 0 isosurfaces for the same time snapshot are shown as a mesh in Figure 22.

For the efficient computation of two-point statistics utilizing the methods described in this thesis, voxelized representations of the interface is needed. To achieve this, the SDF arrays were first segmented by thresholding at 0 into a binary array representing the liquid and solid phases. The segmented microstructures were shifted by one voxel and subtracted from the original microstructure to identify all of the interface voxels (only these voxels have non-zero values after this operation). This operation was performed in both directions in each of the three reference directions to make sure all of the interface voxels were captured. Additionally, we selected only the interface voxels in the solid phase at the solid-liquid boundary. This results in an interface with thickness of one voxel length.

Since the samples were directionally solidified along the axial direction of the cylinder, the primary dendrite arms are all aligned approximately in the axial direction of the cylindrical samples. Perpendicular to the primary arm are the secondary arms, and perpendicular to the secondary arms are the tertiary arms. This is shown in Figure 21, where the primary, secondary and tertiary arm directions are labeled as a_1, a_2 and a_3, respectively on the signed distance function plots. Note that the dendrite arm axes (a_1, a_2, a_3) are slightly different from the data axes (x, y, z). In Figure 21b and c, quaternary dendrite arms can be seen along directions roughly perpendicular to tertiary arms. However, due to the small spacing between the tertiary arms, the quaternary arms are short and tilted.

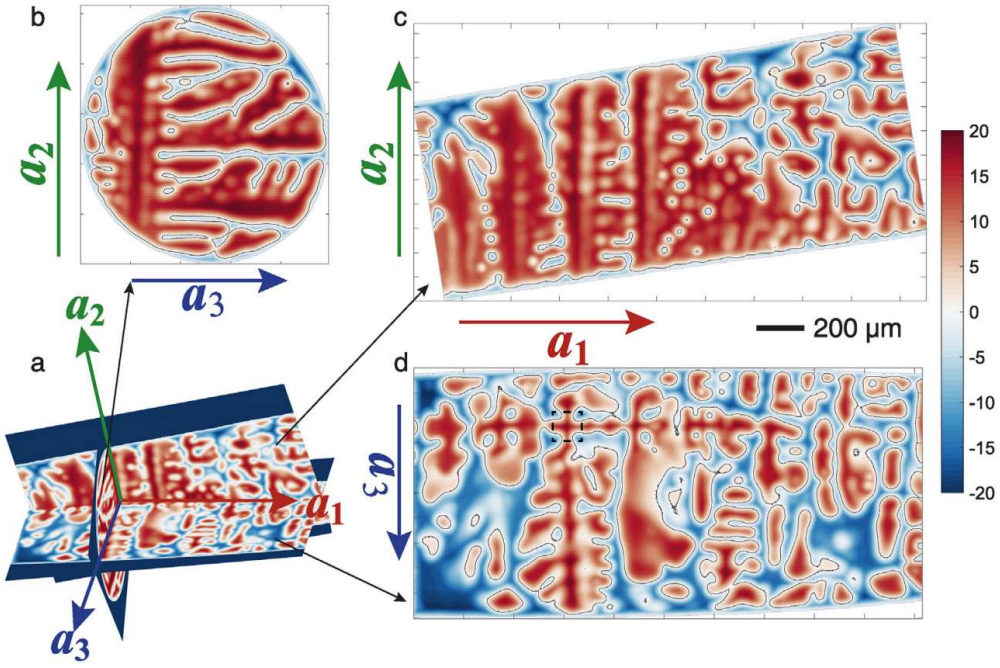

Figure 21 – 2D slices of the signed distance function (SDF). The value of the SDF indicates the distance in pixels to the nearest interface, with positive values being the solid phase and negative being liquid. (a) 3D view of the 2D SDF slices. The three slicing planes are orthogonal to each other and chosen to be aligned with the primary dendrite arm (a_1), secondary dendrite arm (a_2), and tertiary dendrite arm (a_3) directions. The *x, y, z* are the original data axes. (b–d) Enlarged 2D views of the SDF slices in (a). The white lines indicate the location of the interface.

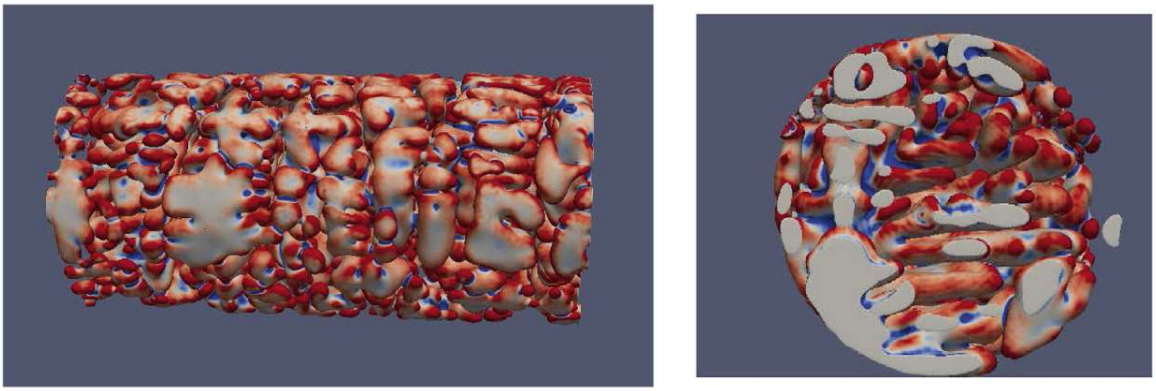

Figure 22 – Solid-liquid interface of Al-Cu alloy during coarsening. The locations of the interface and the interfacial curvatures and velocity are calculated from X-ray tomography experimental data using level-set method. The colors of the interfaces shown in the figures indicates the local mean curvature H of the interface. Red colors indicate high (positive) curvature, and blue colors indicate low (negative) curvature.

4.6.2 Local State Selection

The microstructure variables of interest for this study are the interfacial curvatures and interfacial velocity (V, in the direction normal to the interface, where the normal is directed from the solid to the liquid. Thus positive V denotes local solidification). The interfacial curvature can be represented by the two principle curvatures κ_1 and κ_2 or the combination of the mean curvature H = $(\kappa_1 + \kappa_2)/2$ and the Gaussian curvature K = $\kappa_1\kappa_2$. These interfacial properties can be easily extracted from X-ray tomography data using a level-set method based on signed distance function (SDF) [71].

In Figure 23c and d, the histograms of H and V are shown. While the histogram of the velocity behaves like a Gaussian distribution centered on V = 0, the histogram of H is skewed towards the H > 0 direction. Moreover, the difference in the shape of the distribution profiles of H and V indicates that there is not a one-to-one correspondence between H and V. This is also illustrated by the H − V joint probability distribution (Figure 23a). For each value of H, there is a broad range of V values present, and vice versa.

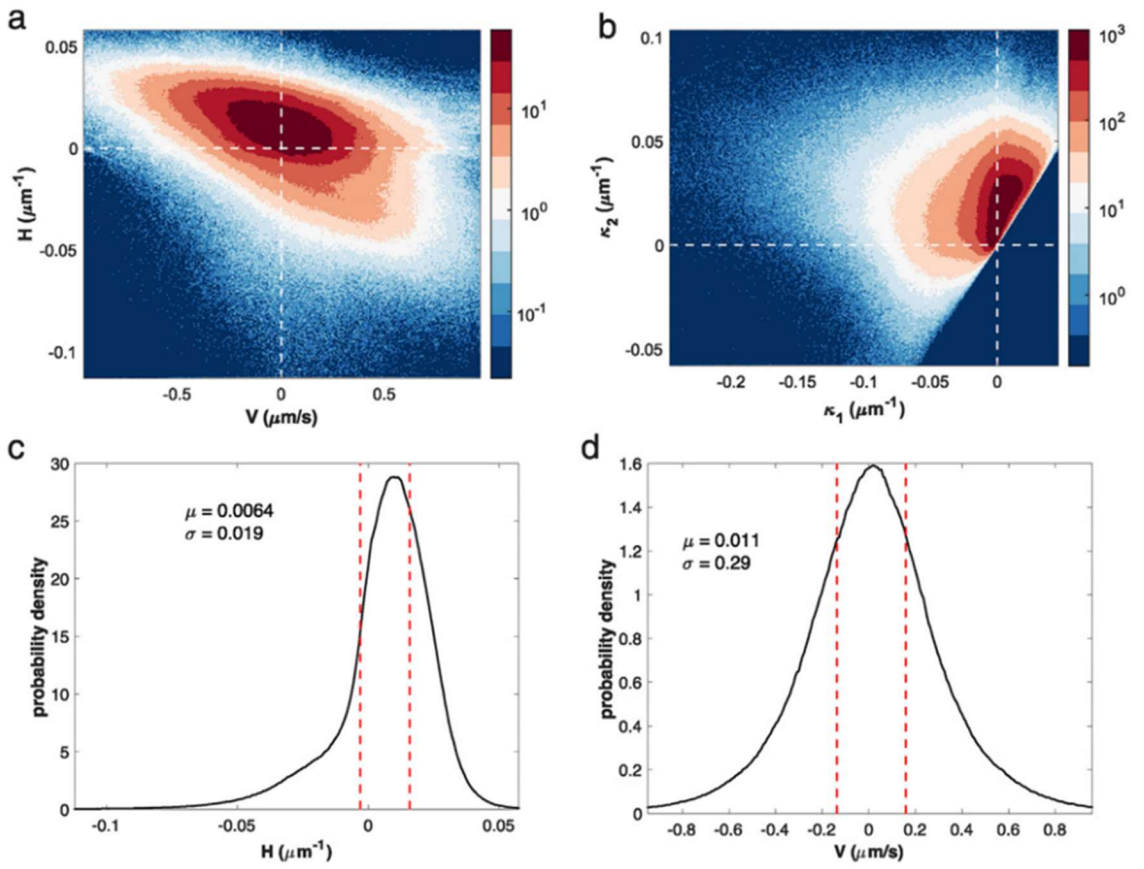

Figure 23 – Two-dimensional probability density distribution of interfacial mean curvature H vs velocity V (a) and principal curvatures κ_1 vs κ_2 (b). Probability density distribution of interfacial mean curvature H (c) and velocity V (d). The high and low 0.1% of the data are clipped. The histograms are taken with 256 evenly spaced bins. The mean (m) and standard deviation (s) are also shown. Red dashed lines indicate 1s deviation from mean.

Although the classic Allen-Cahn theory for interfacial dynamics suggests a negative proportionality between the interfacial mean curvature H and interfacial velocity V in the ideal case [72], the lack of one-to-one correspondence between H and V may come from two reasons. Firstly, given a certain value of H, the actually shape of the interface may not be set since the two principle curvatures κ_1 and κ_2 can still vary. Secondly, even if we pick a combination of κ_1 and κ_2, the velocity of the interface can be influenced by its neighboring interfaces, since unlike in the Allen-Cahn case, coarsening proceeds by the

long range diffusion of solute, thus leading to non-local interactions and a dependence of the velocity on the surrounding interfaces. Therefore, it is important to account for these local effects by including the spatial variance of H and V in our analysis, justifying the case for the use of two-point spatial correlations.

Since H and V are continuous local state variables, we can directly use their values as state definitions and use Eq. 13 for the computation of two-point Pearson correlation, ρ_t^{HV}. However, for computing the probabilistic two-point correlations, it is necessary to discretize the local state space. In this study, the local state spaces for H and V were each discretized into ten bins in such a way that the number of interfacial voxels (these are the only voxels for which H and V are defined) corresponding to each bin were equal (see Figure 24). In other words, the first bin H_1 represents the complete range of local states (i.e., mean curvatures) encountered in the lowest decile of the interfacial voxels, when sorted by the values of H. H_2 represents the next decile of interfacial voxels in the same sorting (see also Figure 24a). The local state space for V is also binned in an analogous manner, as shown in Figure 24b. Note that a finer binning (i.e. using more bins) will give a finer resolution in the local state space, whereas, at the same time, reduce the number of data points in each bin hence impairing the statistical significance of the result. With the use of the discretized local state spaces described above, the discretized microstructure functions $m_s^{H_i}$ and $m_s^{V_i}$ for $i = 1..10$ essentially take values of zero or one. In prior work [30], microstructures of this type were termed eigenmicrostructures.

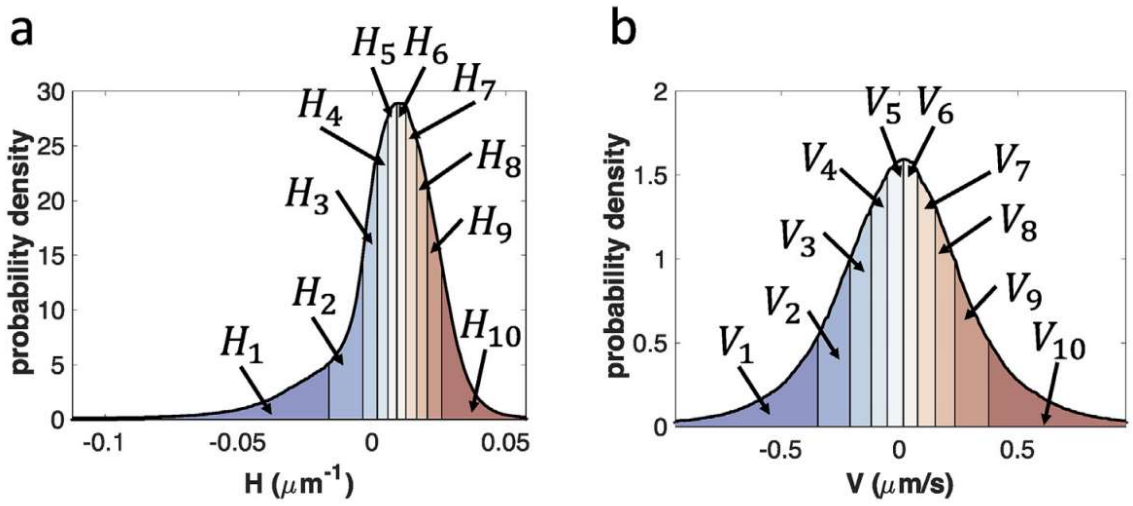

Figure 24 – Discretization of the local states H and V. The local state spaces for H and V are each discretized into ten bins with equal number of interfacial voxels in each bin. The first bin H_1 represents the complete range of H encountered in the lowest decile of the interfacial voxels, when sorted by the values of H. H_2 represents the next decile of interfacial voxels in the same sorting. The discretization of V is done analogously.

It is also possible to bin the local states with respect to interfacial shape. The two principal curvatures κ_1 and κ_2 can be used to classify interface patches based on their specific shapes from the $\kappa_1 - \kappa_2$ distribution map (Figure 23b), or the so called interface shape distribution (ISD). As shown in Figure 25, we can select regions from the ISD corresponding to positive (solid) spheres ($\kappa_1 > 0$, $\kappa_2 > 0$), negative (liquid) spheres ($\kappa_1 < 0$, $\kappa_2 < 0$), positive (solid) cylinders ($\kappa_1 \approx 0$, $\kappa_2 > 0$), negative (liquid) cylinders ($\kappa_1 < 0$, $\kappa_2 \approx 0$), and planar interfaces ($\kappa_1 \approx 0$, $\kappa_2 \approx 0$), etc. This classification is tantamount to discretization (i.e., binning) in the $\kappa_1 - \kappa_2$ space.

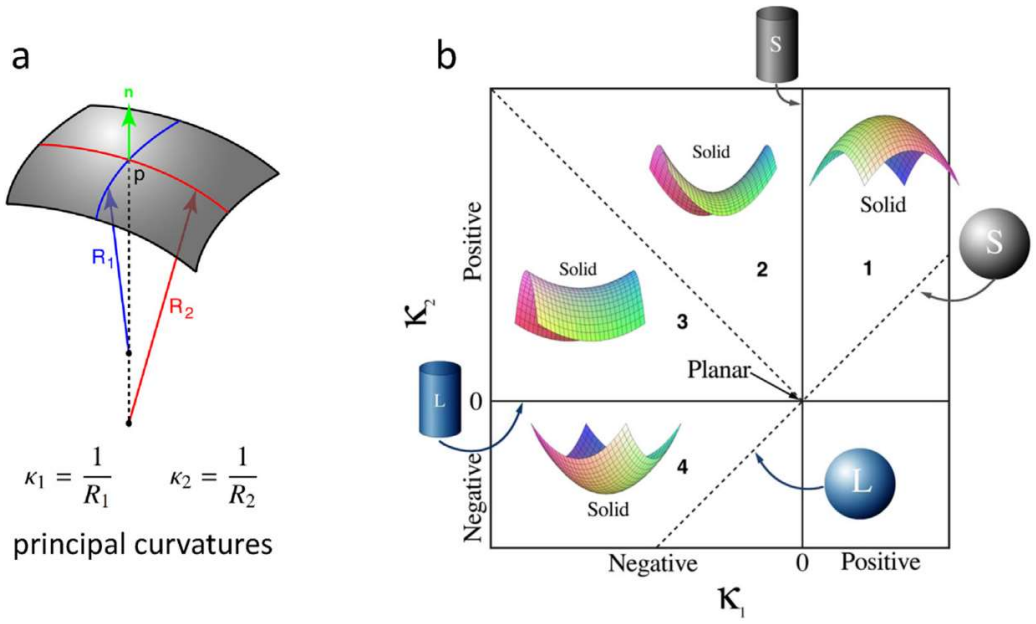

Figure 25 – Illustration of the principal curvatures κ_1, κ_2 (a) and the interfacial shape distribution (ISD) (b) Interfaces with specific shapes can be selected from the ISD by putting constraints on κ_1 and κ_2. Near the origin are the flat interfaces; on the κ_1 = 0 and κ_2 = 0 axes are the solid (positive, solid on the concave side) and liquid (negative, liquid on the concave side) cylinders respectively; and along the line of κ_1 = κ_2 are the solid and liquid spheres. In regions between these extremes are interfaces of intermediate shapes.

As noted earlier, the values of the local state variables H and V are available only in the cylindrical sample volume (an irregularly shaped volume) and only for the solid-liquid interface voxels. Consequently, it is necessary to use masks to compute all the different two-point correlations defined earlier in a computationally efficient manner. For the present work, two mask functions are defined as:

$$c_s = \begin{cases} 1, & s \text{ is on the interface} \\ 0, & \text{otherwise} \end{cases} \tag{38}$$

$$\bar{c}_s = \begin{cases} 1, & s \text{ is inside the cylinder} \\ 0, & \text{otherwise} \end{cases} \tag{39}$$

Additionally, in order to implement a non-periodic treatment of the microstructure, the microstructure domain was padded with a minimum of 200 voxels in each direction. Padding was accomplished in a manner that converted the given cylindrical volume into a regular rectangular parallelepiped volume with uniformly sized voxels. Because the extent of padding was at least 200 voxels in each reference direction (*x, y, z*), the correlations computed using the equations presented in earlier sections are valid at least for all vectors with all three components smaller than 200 voxels.

4.6.3 Analysis within the Scope of a Time Step

We will first treat H and V data as continuous local states and compute the two-point Pearson correlations, ρ_t^{HH}, ρ_t^{VV}, and ρ_t^{HV}. Of these, the first two are called auto-correlations, while the third is referred to as a cross- correlation. All three correlation maps are shown in Figure 26 in forms of orthogonal two-dimensional slices, up to vectors whose components are smaller than 200 pixels (358 μm). The 2D slices shown in this figure are selected perpendicular to secondary (a_2) and tertiary (a_3) arm directions; compare with the sections of the microstructure volume shown in Figure 21.

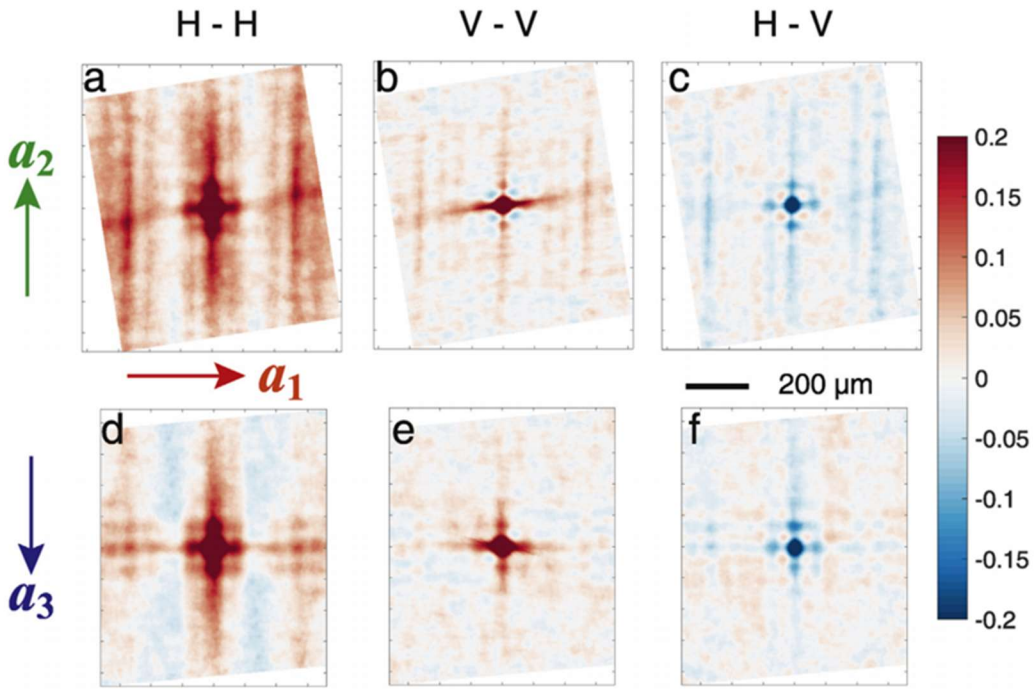

Figure 26 – Two-dimensional slices of the two-point Pearson auto-correlation of interfacial mean curvature H (a,d) and velocity V (b,e), and two-point Pearson cross-correlation of H and V (c,f). The slices are selected perpendicular to the tertiary dendrite arms a_3 (a-c), and perpendicular to the secondary dendrite arms a_2 (d-f). These slices correspond to the 2D slices of the SDF in Figure 21.

In all of the three correlation maps (H − H, V − V, H − V), we see four-fold symmetric patterns, reflecting the four-fold symmetry of the dendrite arms. In the H − H auto-correlation (Figure 26a,d), we see long range correlations along the secondary and tertiary dendrite arm axes (a_2 and a_3). More specifically, the two-point Pearson correlation function $\rho_t^{HH} \geq 0.2$ up to a distance of $|t| \approx 250$ μm along a_3. This correlation distance is even larger for the a_2 direction, where $\rho_t^{HH} \geq 0.2$ for distances as far as $|t| \approx 350$ μm. These long range correlations of the interfacial curvatures along the tertiary and secondary dendrite arms indicate that the secondary and tertiary dendrite arms are aligned parallel to each other (along the a_2 and a_3 axes, respectively), and that their interfacial shapes are similar along their respective axial directions. This can also be verified from the slices of

the SDF in Figure 21. The continuous correlations come from the stems of the secondary and tertiary dendrite arms, which have extended structures along directions matching the strong correlation stripes. Along the a_1 axis, the Pearson auto-correlation coefficient of H shows periodic patterns with a periodicity of $\sim$ 270 μm (Figure 26a,d). This indicates periodically repeated microstructures along the primary dendrite arm direction in the sample, which naturally suggests that this periodicity comes from the secondary branches on the primary arm. Therefore, this distance of $\sim$ 270 μm should correspond to the secondary dendrite arm spacing λ_2, the value of which is in good agreement with the SDF plots (Figure 21). The correlation along a_1 does not show continuous patterns in the long range as seen along a_2 and a_3. This indicates an absence of extended microstructure along a_1. In other words, there is no primary dendrite arm in the microstructure domain studied. From Figure 21c, we can confirm that it is indeed the case (the primary dendrite arm is cut out of the sample cylinder during sample preparation).

While the interfacial curvature H shows long range auto-correlations with $\rho_t^{HH} \geq 0.2$ at distances $|t| \geq 250$ μm along all three dendrite arm axes, we see a much weaker correlation in the interfacial velocity V auto- correlation maps (Figure 26b,e). In fact, we only see strong ($\rho_t^{VV} \geq 0.2$) correlation within $|t| \leq 100$ μm along a_1 and a_3, and $|t| \leq 65$ μm along a_2. The first correlation range (100 μm) is smaller than the inter-dendritic distance λ_2 of the secondary arms along the primary arm axis, as obtained from the H auto-correlation; and, as we shall see later, the second correlation range (65 μm) is about the same as the tertiary dendrite arm spacing λ_3 of the tertiary arms on the same secondary arm. This implies that the interfacial velocity V only strongly correlates with each other within the same secondary or tertiary dendrite arm. Since the motion of the interface is

determined by the diffusion of solute in the alloy, the distance of velocity correlation (i.e., 100 μm) is a good measurement of the solute diffusion length. Moreover, in contrast to the H correlation which shows only positive correlation coefficient in the short range ($|t| \leq 100$ μm), the interfacial velocity V shows negative correlation near the strong center peak in the correlation map. Since the velocity has near-zero mean value, the negative correlation indicates a sign reversion of the velocity across a distance of $\sim$ 30 μm on the interface. This in turn reflects the wavelength of perturbation of the diffusion field by the fluctuations in the solute concentration.

We also present the computed H − V cross-correlation in Figure 26c,f. H and V are negatively correlated at r = 0 with $\rho_t^{HV} = -0.52$. The patterns in the H − V correlation map show mixed contributions from the auto- correlations of H and V. More specifically, in the short range, H is positively auto-correlated, but V has both positive and negative auto-correlations, therefore, the correlation pattern resembles that of the V auto-correlation in this range. In the long range, auto-correlation of H is strong yet the auto-correlation of V is weak, so the combined effect is that the cross-correlation of H and V is also weak.

In order to better understand the spatial distribution of H and V, we can now utilize the bins we defined in their respective local state spaces. These discretized local states were defined earlier and labeled as $H_1, \ldots, H_{10}, V_1, \ldots, V_{10}$. Using these discrete local states, we can compute two-point Pearson auto-correlation functions of the corresponding microstructure functions $m_s^{H_1}, \ldots, m_s^{H_{10}}, m_s^{V_1}, \ldots, m_s^{V_{10}}$. These are presented in Figure 27 and Figure 28. It is seen that only first three and the last bins of H (H_1, H_2, H_3, H_{10}) show strong long range auto-correlations. In other words, only the 30% most negative and

10% most positive values of H have strong long range auto-correlations. The rest of the H values do not have strong long range correlations in space. Among the 10 bins of H, the first bin (H_1) is the most correlated in the long range, and it contributes to most of the features in the continuous H auto-correlation pattern in Figure 26. The high negative H values represented by H_1 correspond to interfaces with concave curvatures on the liquid side, which can be either cylindrical or spherical.

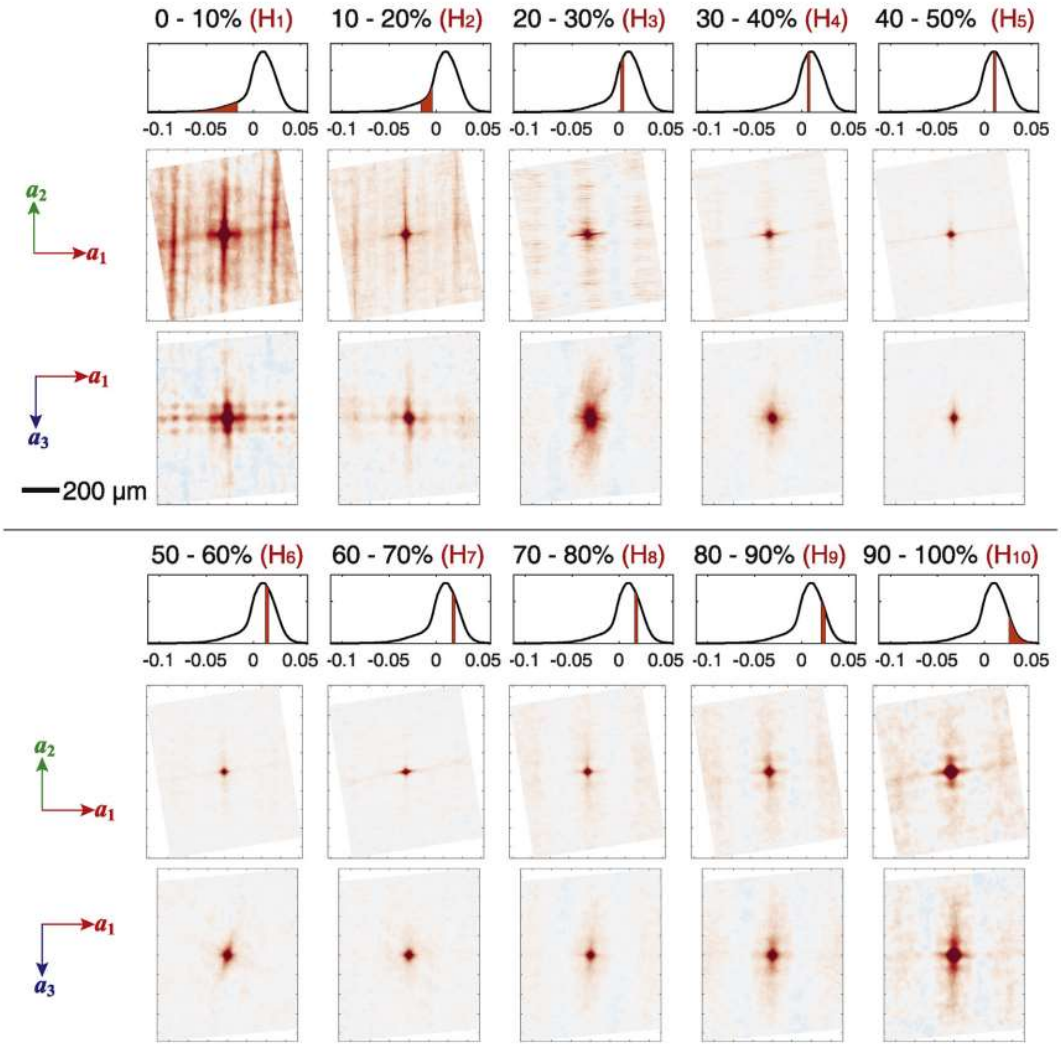

Figure 27 – Two-dimensional slices of the two-point Pearson auto-correlation of interfacial mean curvature H in 10% bins. The Pearson correlations are computed with the microstructure functions built from the discretized local states H_1 .. H_{10}.

The correlations of V in 10% bins (Figure 28) show the same trend as those of H in that the most positive and most negative extremes are the most correlated in the longer range, while the center values are only correlated at around t = 0. In other words, the spatial distribution of the non-extreme values of V are essentially equivalent. In all of the cases, V is only spatially correlated in the short range, which agrees with the result from the full V correlation.

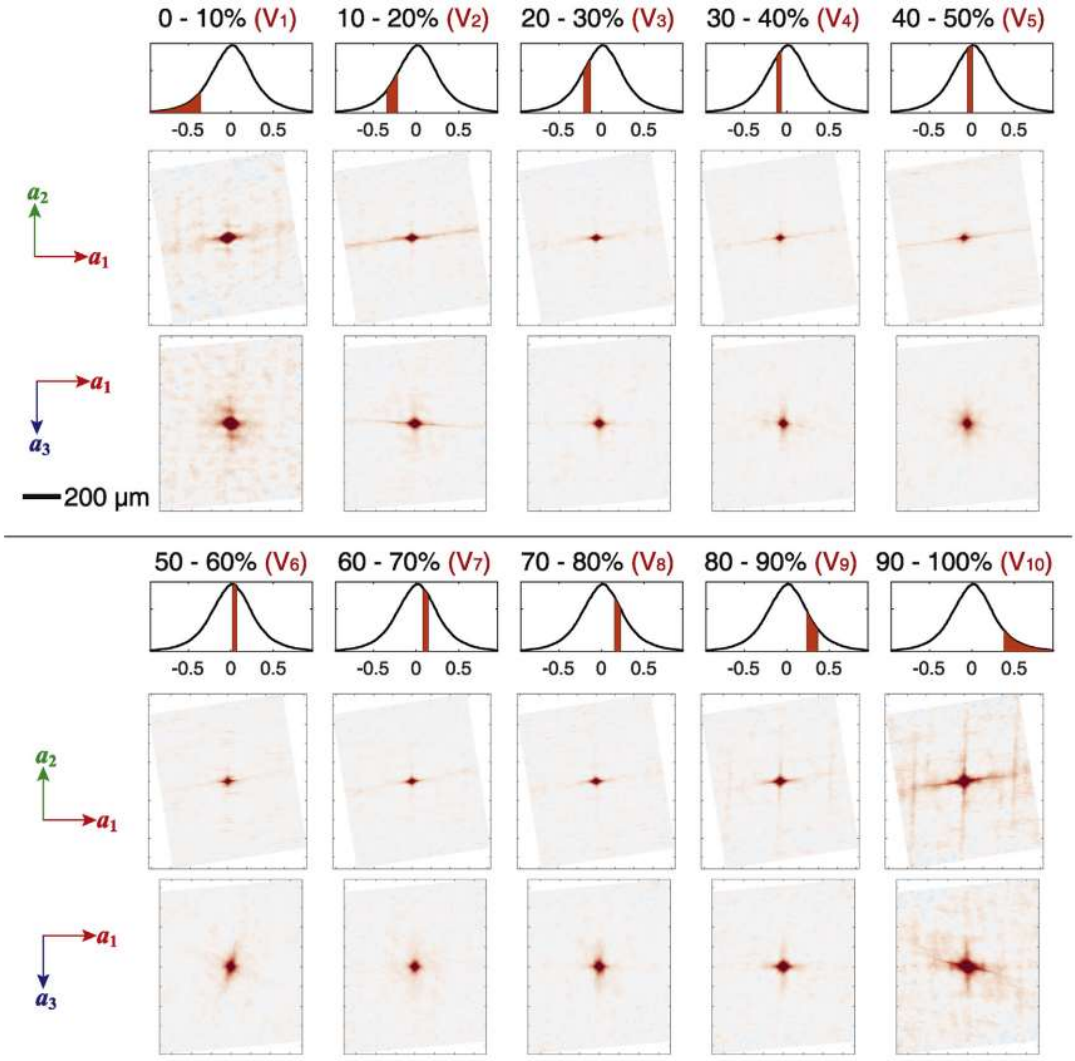

Figure 28 – Two-dimensional slices of the two-point Pearson auto-correlation of interfacial velocity V in 10% bins. The Pearson correlations are computed with the microstructure functions built from the discretized local states V_1 .. V_{10}.

In order to determine the specific microstructures resulting in the long-range correlation pattern seen in the auto-correlation of the most negative values of H, we select on the ISD according to Figure 25, curvature regions corresponding to negative (liquid) cylinders ($\kappa_1 < -0.02, |\kappa_2| < 0.01$; region A) and negative (liquid) spheroids ($\kappa_1 < -0.02, \kappa_2 < -0.02$; region B) as shown in Figure 29h. We use these regions as discrete local states in the $\kappa_1 - \kappa_2$ space and calculated the two-point Pearson correlation functions of their microstructure functions. The results are shown in Figure 29a-g. Comparing Figure 29a,b,d,e with Figure 26, we find that the four-fold "lattice"-like correlation pattern comes from the auto-correlation of cylindrical interfaces with negative H. These interfaces are the troughs on the stems of the secondary (or tertiary) dendrite arms, between tertiary (or quaternary) dendrite branches. These interfaces are well-aligned along the a_2 and a_3 directions and with even spacings along a_1. These contribute to the long stripes in the auto-correlation map of H. In the $a_1 - a_3$ plane (i.e. perpendicular to the secondary dendrite arms), the cross-sections of the stems of the secondary dendrite arms show "cross" shapes with negative curvature troughs between the roots of tertiary dendrite arms, as highlighted by the black dashed box in Figure 21d. An enlarged view of the interface around this area is shown in Figure 30a, with the 10% most negative H highlighted in red. These "troughs on cross" structures will give a correlation pattern that resembles a "cubic lattice". In Figure 30b and c we demonstrate this correlation through a 2D model interfacial structure, in which the four-fold lattice-like correlation pattern is reproduced. Following this approach, the distance between the center and diagonal peaks in the "lattice" pattern should be the diameter of the secondary dendrite arms (d_2), as illustrated by the arrows in Figure 30. From Figure 29d we know that d_2 = 90 μm. From Figure 26c,f we see that the H and V

show anti-correlations pertaining the same four-fold symmetric pattern as seen in the H - H auto-correlation. In fact, these H - V anti-correlations happen at the same locations where H is autocorrelated in the short range. Combining with the argument made just above, it suggests that the trough structures on the secondary dendrite arms are moving away from each other (i.e. toward the liquid), which will make the curvature at these locations smaller.

As opposed to H_1, H_3 shows strong correlations on planes perpendicular to the a_2 axis. This segment of H corresponds to the near-zero curvature interfaces. Analysis on the Gaussian curvature K shows that these are primarily flat interfaces which constitute the surface of the tertiary dendrite arms that undergo columnar growth perpendicular to the secondary dendrite arms (Figure 21b). To verify this, we utilize the ISD again and compute the two-point Pearson auto-correlation with the near-zero curvature bin $\kappa_1, \kappa_2 \in [-0.005, 0.005]$, represented by region D in Figure 29h. The resulting correlation pattern greatly resembles that of H_3, suggesting that the latter is indeed from the correlation of flat interfaces on the columnar tertiary dendrite arms. Therefore, the spacings between the "correlation planes" should represent the columnar width or equivalently the inter-dendritic spacing of the tertiary dendrite arms. Since the first strong correlation appears at |t| = 65 μm, the average columnar width should be on the same scale.

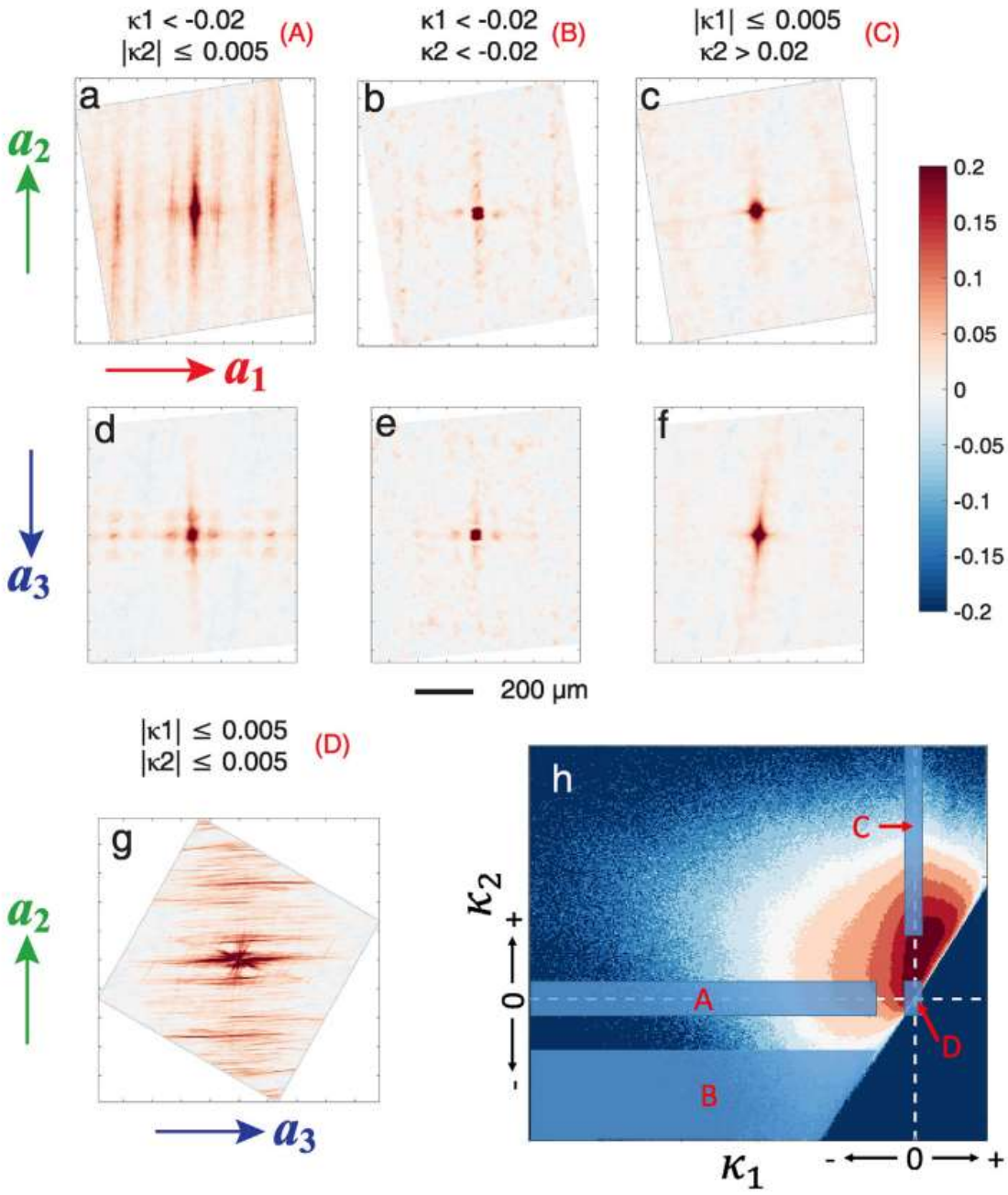

Figure 29 – Two-dimensional slices of two-point Pearson auto-correlation with discrete local states selected from the ISD (Figure 25b) according to Figure 23b. (a,d) $\boldsymbol{\kappa_1 < -0.02, |\kappa_2| < 0.01}$; region A in (h), representing negative cylindrical shapes, and constituting ~ 3.18% of the whole interface. (b,e) $\boldsymbol{\kappa_1 < -0.02, \kappa_2 < -0.02}$; region B in (h), representing negative spherical and spheroidal shapes, and constituting ~ 0.87% of the whole interface. (c,f) $\boldsymbol{|\kappa_1| \leq 0.005, \kappa_2 > 0.02}$; region C in (h), representing positive cylindrical shapes, and constituting ~ 13.17% of the whole interface. (g) $\boldsymbol{\kappa_1, \kappa_2 \in [-0.005, 0.005]}$; region D in (h), representing flat interfaces, and constituting ~ 2.94% of the whole interface.

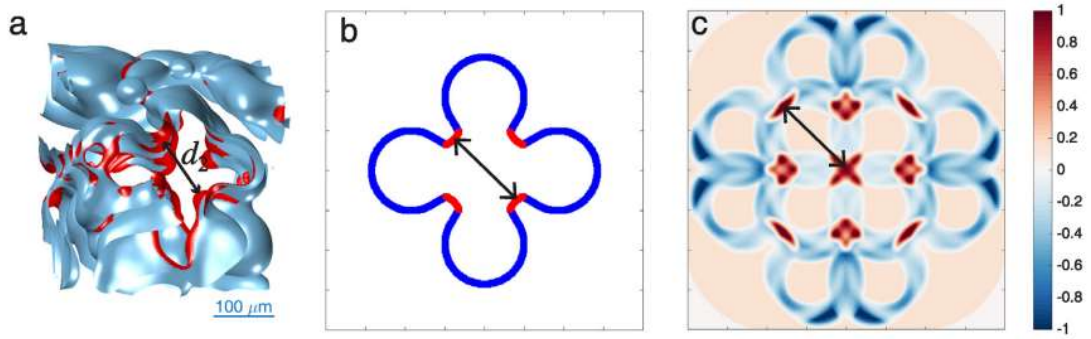

Figure 30 – (a) A subset of the sample interface showing the stem of a secondary dendrite arm (as highlighted with the black dashed box in Figure 21d), with the 10% most negative H interfaces highlighted in red. (b) A 2D model structure representing a 2D cross-section of the structure in (a), with high negative curvature interfaces highlighted in red and the rest in blue. (c) The two-point Pearson auto-correlation of the red local state (microstructure function) in the model structure. The whole "interface" (blue and red in b) is used as mask in the calculation. The double-headed arrows indicate distances of equal value in the three plots.

Finally, to obtain a quantitative measure of the probability distribution of H and V in space, we calculate the conditional two-point probability function of discretized local states H_1 and V_{10}. These two should be correlated according to Allen-Cahn theory. In Figure 31a and c, we show the auto-correlation conditional probabilities $f_t^{H_1|H_1}$ and $f_t^{V_{10}|V_{10}}$, respectively. The same long stripes and lattice-like patterns appear in the two-point probability function of H_1 as expected. At these "hotspots", the probability of finding an interface with local state H_1 (i.e. 10% most negative H) is about 30 %, which is 3 times the value in the case where H_1 is randomly distributed $f_{t\to\infty}^{H_1|H_1} = 10\%$. While the probability of finding an interface with H_1 is greater than 20 % even at distances |t| = 250 μm from another interface with H_1, the $f_t^{V_{10}|V_{10}}$ tapers off quickly to the random distribution value (10 %), as shown in Figure 31b and d. Figure 31e shows line profiles of both probability distributions along the white dashed lines in Figure 31a-d. It is clear that the auto-correlation of H_1 shows peaks with a height of 0.3 at ~ ±270 μm, which indicates

strong correlation between neighboring secondary arms. On the other hand, the same line for V_1 shows no discernible peaks at ~ ±270 µm. This again supports the argument that the diffusion of solute is only limited within one dendrite arm.

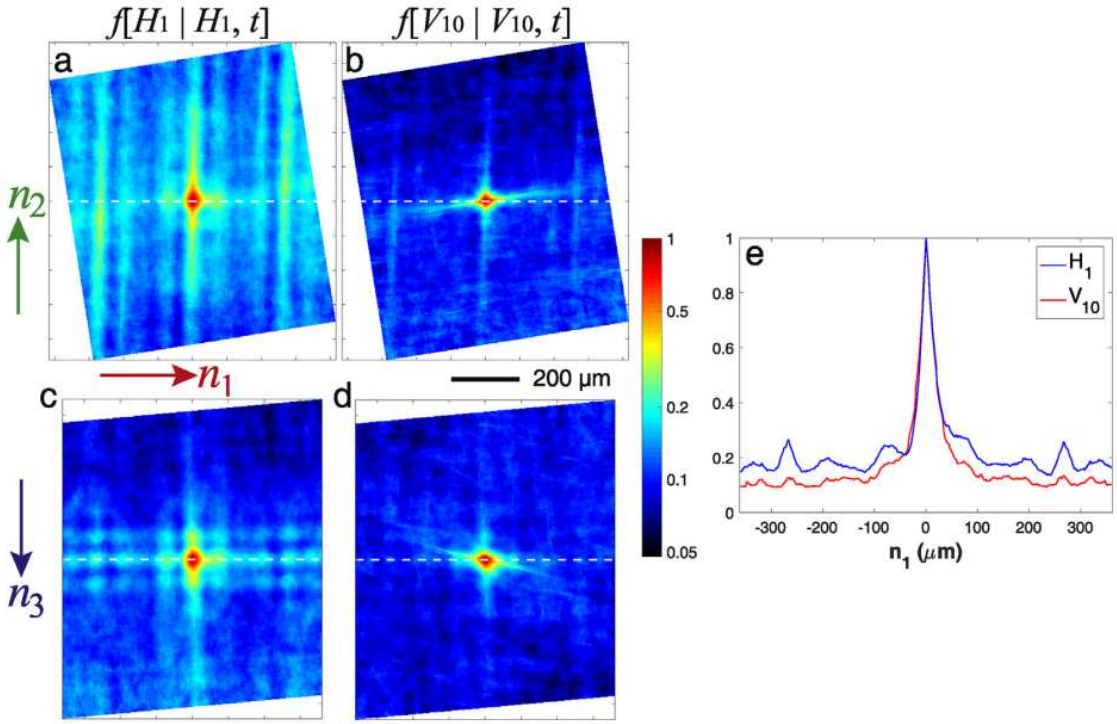

Figure 31 – Two-dimensional slices of the conditional two-point probability functions (auto-correlation) of discretized local states H_1 (a,c) and (V_{10}, b,d). (e) Line profiles of $f_t^{H_1|H_1}$ and $f_t^{V_{10}|V_{10}}$ shown in (a-d), along the white dashed lines.

4.6.4 Analysis within the Scope of an Experiment

This section will deal with the analysis of an experiment as a whole with all of its time steps. As the scope of the data is broader, the local states utilized will be limited to the interface indicator c_s (i.e the interface mask function) and the extreme negative 10% bin of the mean curvature H_1, as the methods can easily be extended to other states further analysis. The reader is reminded that ultimately, the purpose of this case study is to demonstrate the utility of the various tools developed in this thesis and not the exhaustive

analysis of this dataset. The experiment with 55% solid volume fraction is chosen as the example for this section. This experiment contains 94 time steps which will be labeled τ = 0..93. Figure 32 shows a visualization of the solid liquid interface (zoomed into a cubic sub-volume for clarity), and a visualization of the corresponding two point autocorrelation of the interface mask function c_s (represented by f_t^{II}) for the entire cylinder for τ = 0 (a), τ = 20 (b), τ = 93 (c). The 3D visualizations of two point statistics contain 2 isosurfaces at f_t^{II} = 0.0024 and f_t^{II} = 0.0038 on all 3 cases.

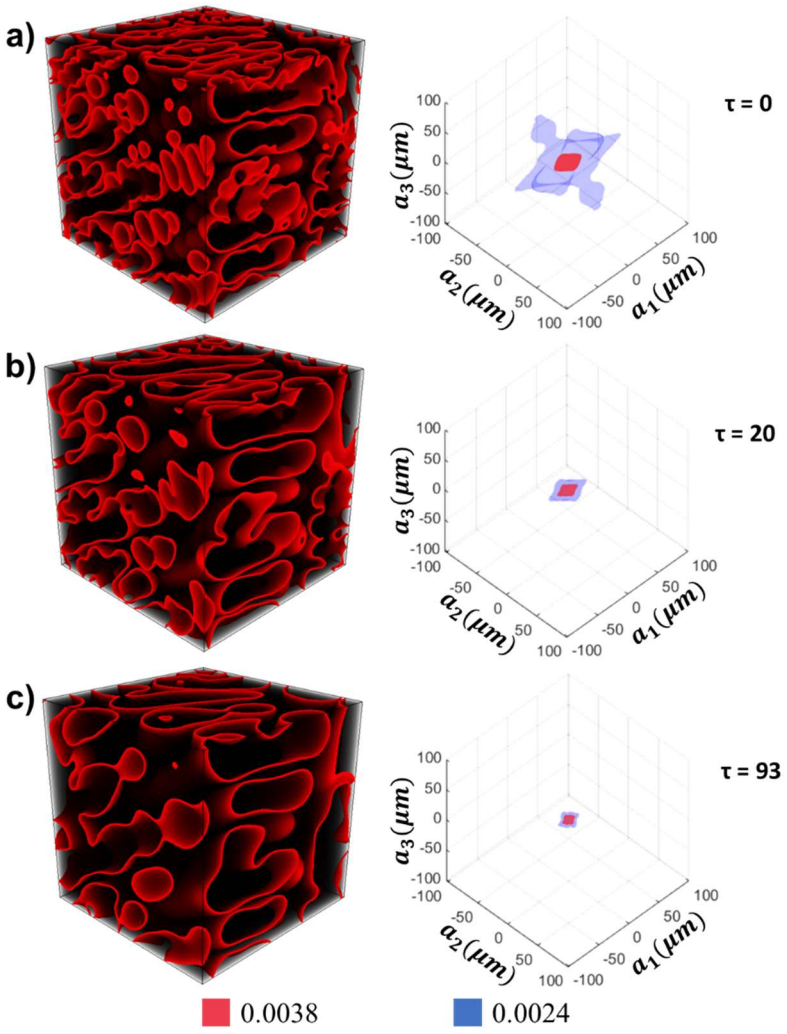

Figure 32 – Visualizations of the solid liquid interface and two point autocorrelation of the interface mask function c_s are shown for τ = 0 (a), τ = 20 (b), τ = 93 (c).

It can clearly be seen from Figure 32 that the dendritic structure is becoming more and more dominant and organized as time progresses, which results in a monotonic decrease in interface surface area. This notion is supported in the principal component 2D projection of the 94 data points as shown in Figure 33. In this plot, each point represents the two point autocorrelation of c_s for a given time step, and the points corresponding to the time steps shown in Figure 32 are shown explicitly. Since PC scores denote deviations from the mean, the mean of two point statistics corresponding to all 94 time steps (a,b,c) and the basis vector for the first principal component (d,e,f) are shown in three orthogonal slices at $z = 0$, $y = 0$ and $x = 0$ planes respectively in Figure 34. The red arrow indicates the approximate direction of the cylindrical axis. In Figure 34a,b we can see that there is a dominant dendritic pattern almost exactly along and perpendicular to the cylinder axis throughout all 94 time steps. Figure 34c shows that the dendritic structure perpendicular to the cylinder favors a particular alignment with a roughly 45 degree slant. The existence and angle of this slant is a direct result of the particular placement of the experiment cylinder with respect to the imaging device and hence an artifact of observer rotations, which will be addressed later in the multi-experiment scope analysis section. From Figure 34d,e we can see that the interface is losing vectors at the tertiary arm spacing range (notice the negative sign on the color bar) , as well as along the cylinder axis at a much faster rate than other directions. This is a result of independent blobs or small tertiary arms getting absorbed into continuous well-formed dendritic arms with homogenous thickness as a result of the coarsening process. Figure 34f shows that this absorption has a directional preference within the circular plane, resulting in the preservation (within some tolerance) of the dominant orientation we see in Figure 34c. In all three slices of the first principal

bases, a very sharp peak at the center (which corresponds to volume fraction) is visible. Since PC1 scores are increasing over time in the positive direction, this corresponds to a sharp rate of decrease in the center statistics, once again showing that with increasing time the volume fraction and hence the surface area of the interface is monotonically decreasing.

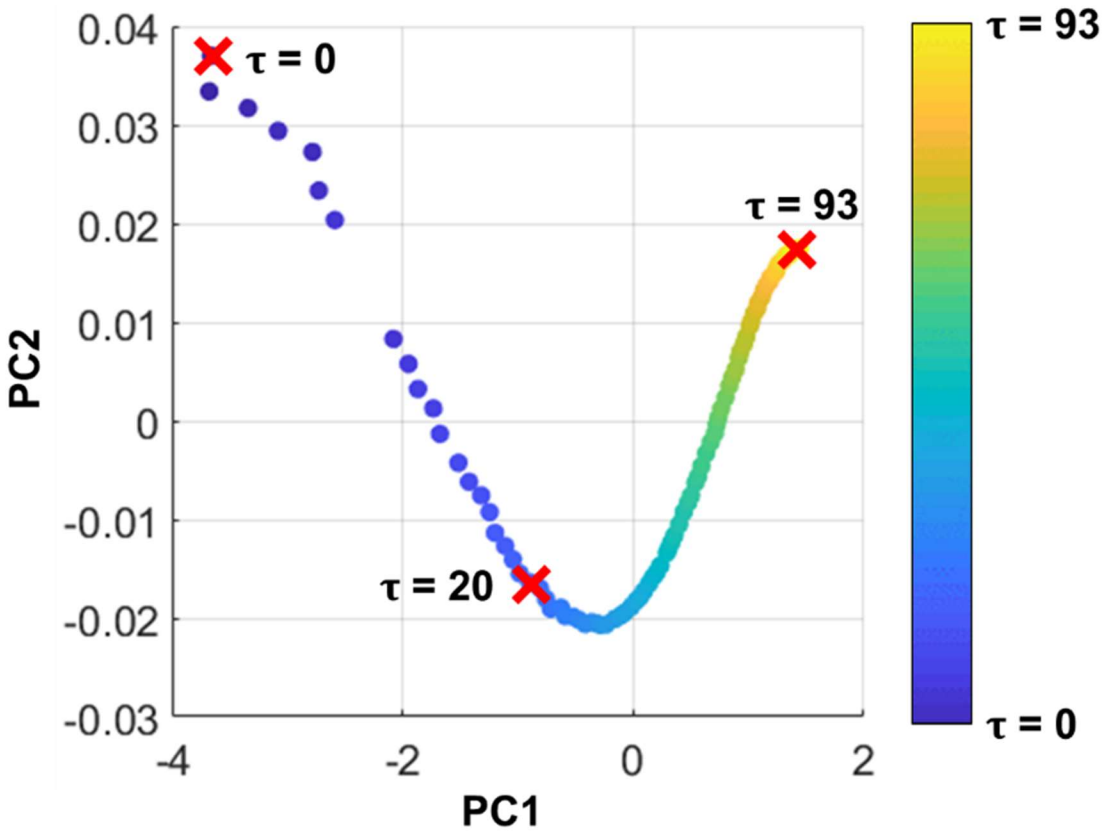

Figure 33 – Principal component scores for the interface autocorrelation across all 94 time steps.

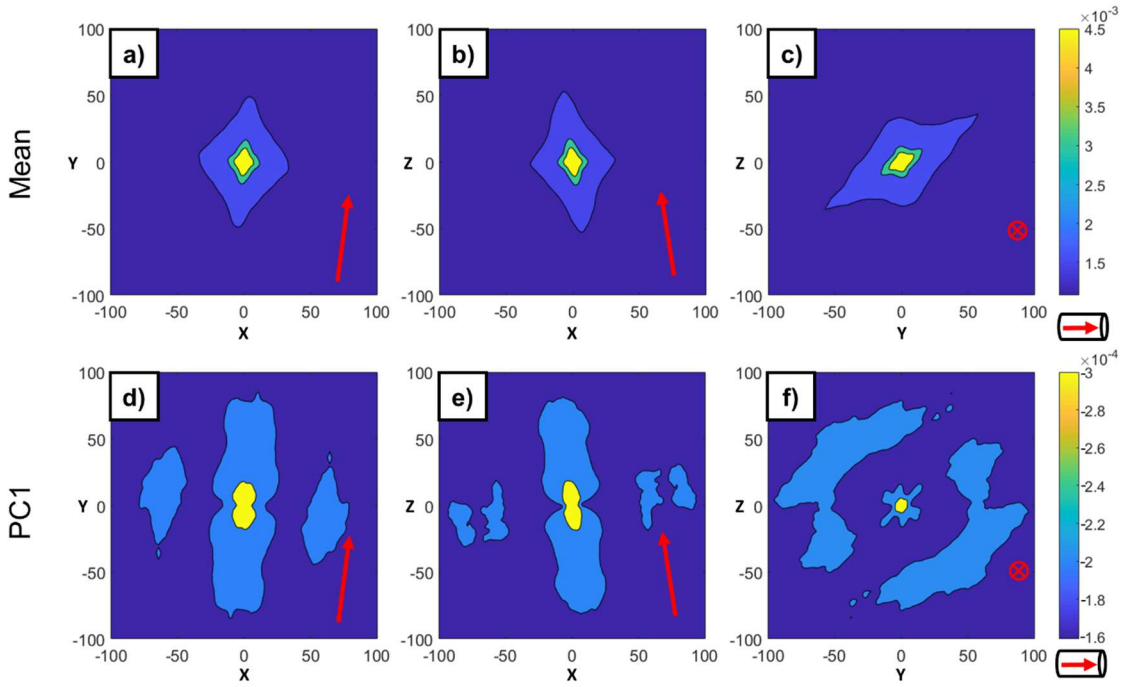

Figure 34 – Orthogonal slices for the mean (a,b,c) and first principal component basis (d,e,f) of interface autocorrelations across all 94 time steps. The red arrow indicates the direction of the cylindrical axis.

We can now inspect whether features and trends observed in the single time step scoped analysis are applicable to the entire experiment. The strongest features and meaningful results for the single time step analysis were observed in the autocorrelation of the extreme negative curvature autocorrelation (H_1), hence we can take this case as a suitable example. Firstly, Figure 35 shows a 3D visualization of the mean of the autocorrelation of H_1 across all time steps. We can once again observe that the dendritic pattern is dominant throughout the experiment. Furthermore, there are clear peaks corresponding to long range order which were previously identified to be associated with λ_2 and d_2, which implies an experiment wide order to the dendritic arrangement. We can further observe the evolution of these features by tracking the line profile previously shown in Figure 31e where both λ_2 and d_2 were identifiable. Figure 36 shows the evolution of

this line profile over time. While the defining peaks for both λ_2 and d_2 exist throughout the experiment, they become sharper as time passes, resulting in a more strictly imposed and well defined dendritic structure. This effect can be somewhat observed in Figure 32, where the dendrite thicknesses and spacings appear to be more uniform in Figure 32c in comparison to Figure 32a.

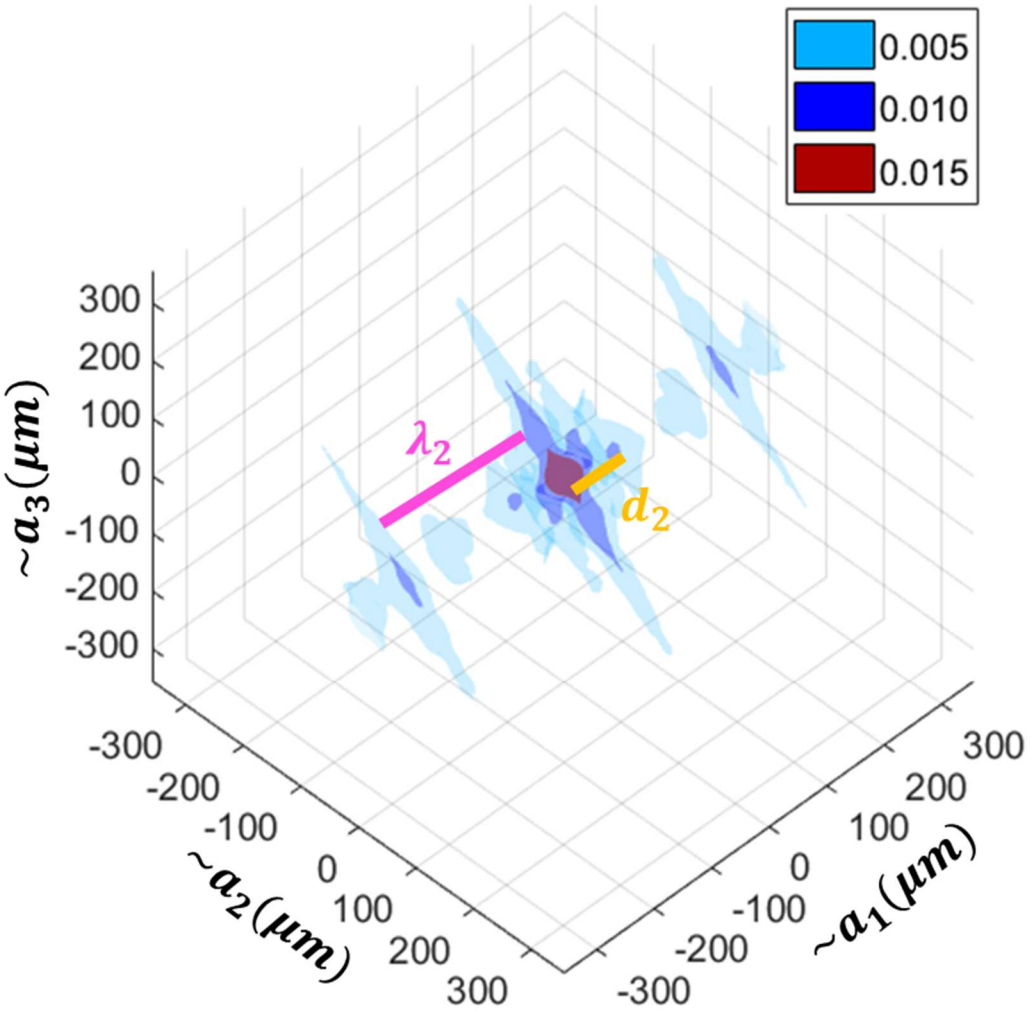

Figure 35 – 3D visualization of the mean of the autocorrelation of H_1 across all time steps. Features λ_2 and d_2 are highlighted.

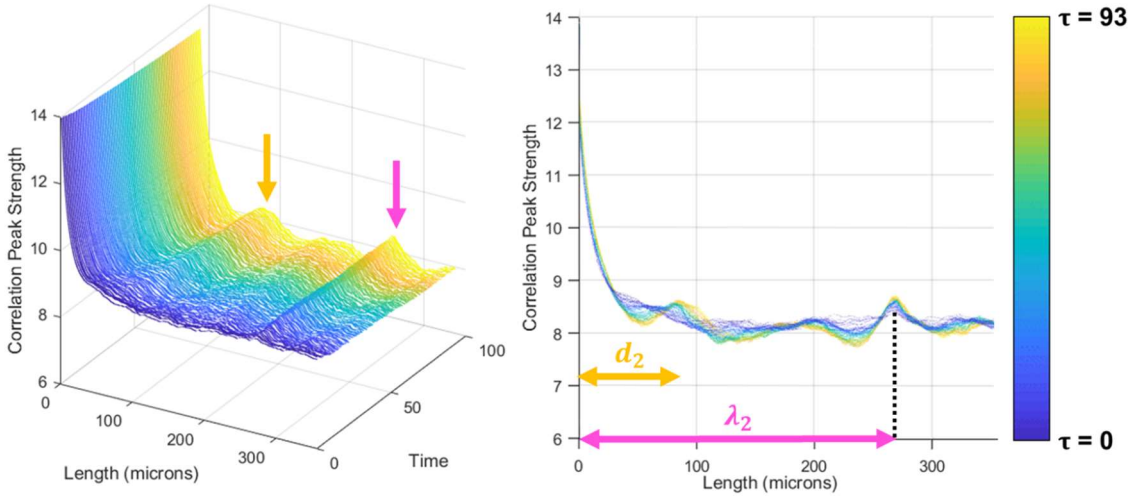

Figure 36 – The evolution of the features λ_2 and d_2 over time. Notice that while the defining peaks exist throughout the experiment, they become sharper as time passes, resulting in a more strictly imposed dendritic structure (compare the blue profile to yellow).

4.6.5 Analysis within the Scope of All 4 Experiments

Finally, we will look at analysis techniques at the scope covering all 500 time steps from 4 different experiments. While the utilization of selective local state definitions and correlation combinations will surely yield additional insights, within the scope of this thesis, the analysis will be limited to the interface indicator c_s as the local state description. As usual however, the methodologies applied here are applicable to any state definition and any mode of spatial correlation.

We start by naively combining the interface autocorrelations of all 500 time steps and projecting them to their principal component space, which is shown in Figure 37. The first principal component dominates by 3 orders of magnitude in the explained variance. Hence the most important observation here is the fact that PC1 is monotonically decreasing, which is again a result of the monotonic decrease in interface surface area. The orthogonal cross-sections for the mean and the first principal component basis is shown in

Figure 38. While Figure 38 is informative in many ways, Figure 38c and f specifically reveal a potential problem with this avenue of analysis. Recall that for the single experiment scoped evolution analysis, the mean and first principal component showed preference to dendritic patterns of specific slants (Figure 34c and f). However, especially Figure 38c is utterly devoid of any anisotropy in mid-range order, showing no signs of a dendritic pattern. As mentioned before, the slant in the patterns are a direct result of the particular placement of the experiment cylinder with respect to the imaging device, and the fact that the combined average is isotropic proves that there is mismatch caused by observer rotations.

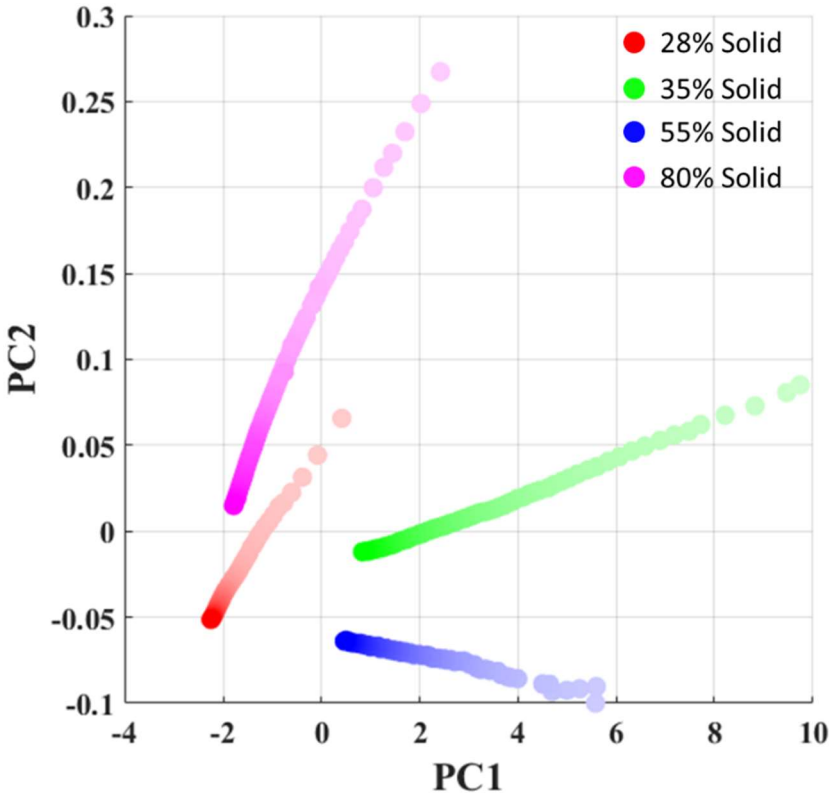

Figure 37 – Principal component scores of the 500 time steps. The gradient within each color implies time evolution, lightest shade being the earliest and the darkest shade being the latest time step within a given experiment.

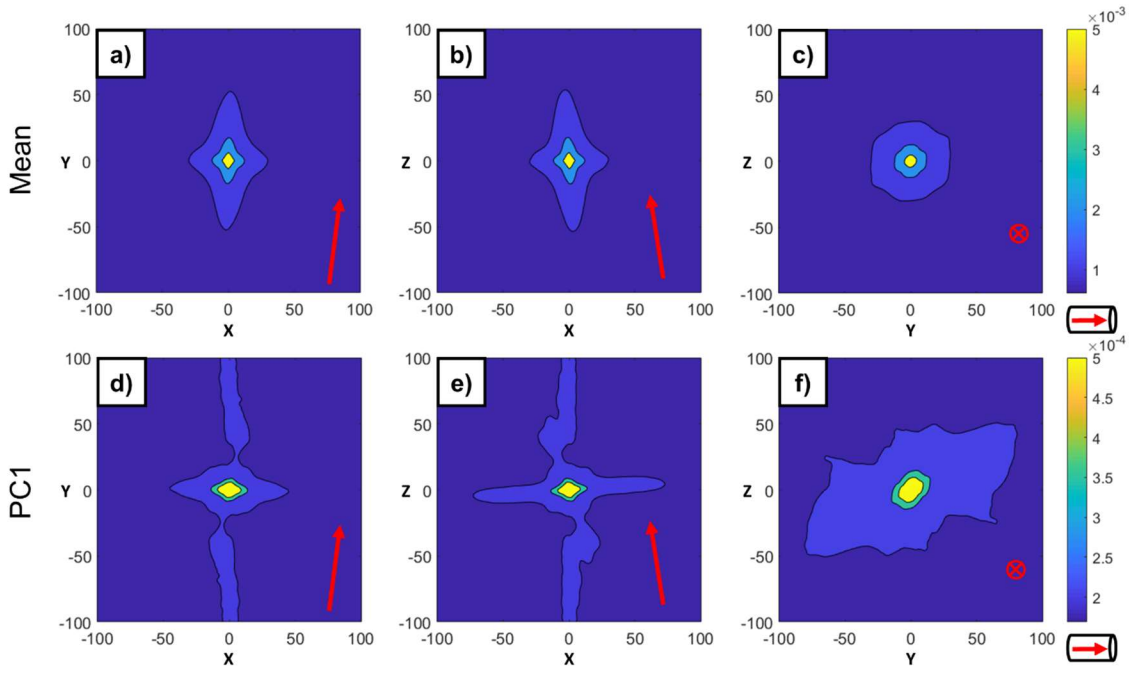

Figure 38 – Orthogonal slices for the mean (a,b,c) and first principal component basis (d,e,f) of interface autocorrelations across all 500 time steps. The red arrow indicates the approximate direction of the cylindrical axis.

These rotations need to be accounted for if an accurate analysis of the dendritic structure is desired. However going back to Figure 21 and Figure 22, a dominant orientation that can be used to align these structures by a rotation of the cylinder around its axis is not immediately apparent, if at all exists. This is where the rotationally invariant statistics derived in this thesis will shine by enabling an accurate analysis of the dendritic structure not possible with any other analysis method in literature. The use of pair correlation functions here will result in the assumption of isotropy and dissolution of the dendritic structure in a similar way. Since the orientation of the cylindrical axis itself is not prone to observer rotations, we require a solution that will only impose rotational invariance in planes perpendicular to the cylindrical axis. Hence we will employ the cylindrical coordinate system for this calculation with a trivial extension to Eq. 33. It is now possible to look at the principal component analysis of the rotationally invariant

interface autocorrelations of all 500 time steps. The projection to principal space is shown in Figure 39. Notice that the projection changed significantly compared to Figure 37. Figure 40 shows comparisons between cross-sectional slices of the mean and first principal component basis without(a,c,e) and with(b,d,f) imposed rotational invariance. Figure 40(a,b) and (c,d) show slices perpendicular to the cylinder axis where the invariance was imposed for the mean and first principal basis respectively. Notice that the dendritic structure is successfully preserved in the mean and the basis across all 500 time steps. (e,f) shows slices along the cylindrical axis for the mean, which shows that the trends for the basis observed in regular statistics are preserved in rotational invariant statistics (this is also true for slices not shown here). This analysis method can now be applied to local states of interest in further extraction of physics and features from all 4 experiments, which is beyond the scope of this thesis.

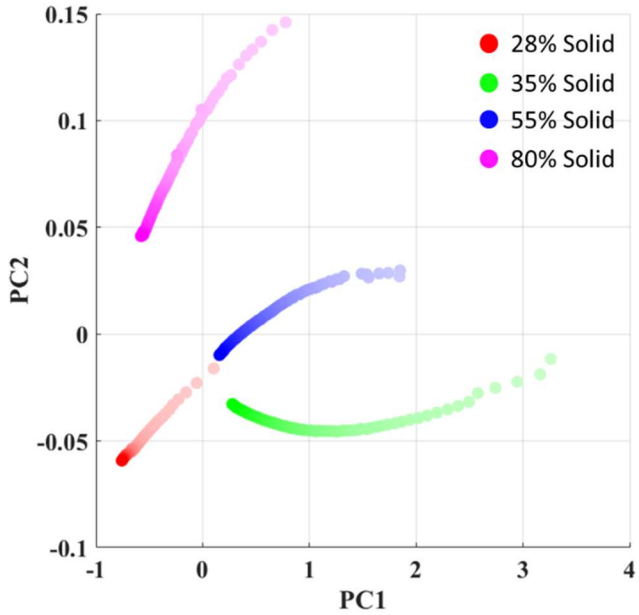

Figure 39 – Principal component scores of the 500 time steps using rotationally invariant two point statistics. The gradient within each color implies time evolution, lightest shade being the earliest and the darkest shade being the latest time step within a given experiment.

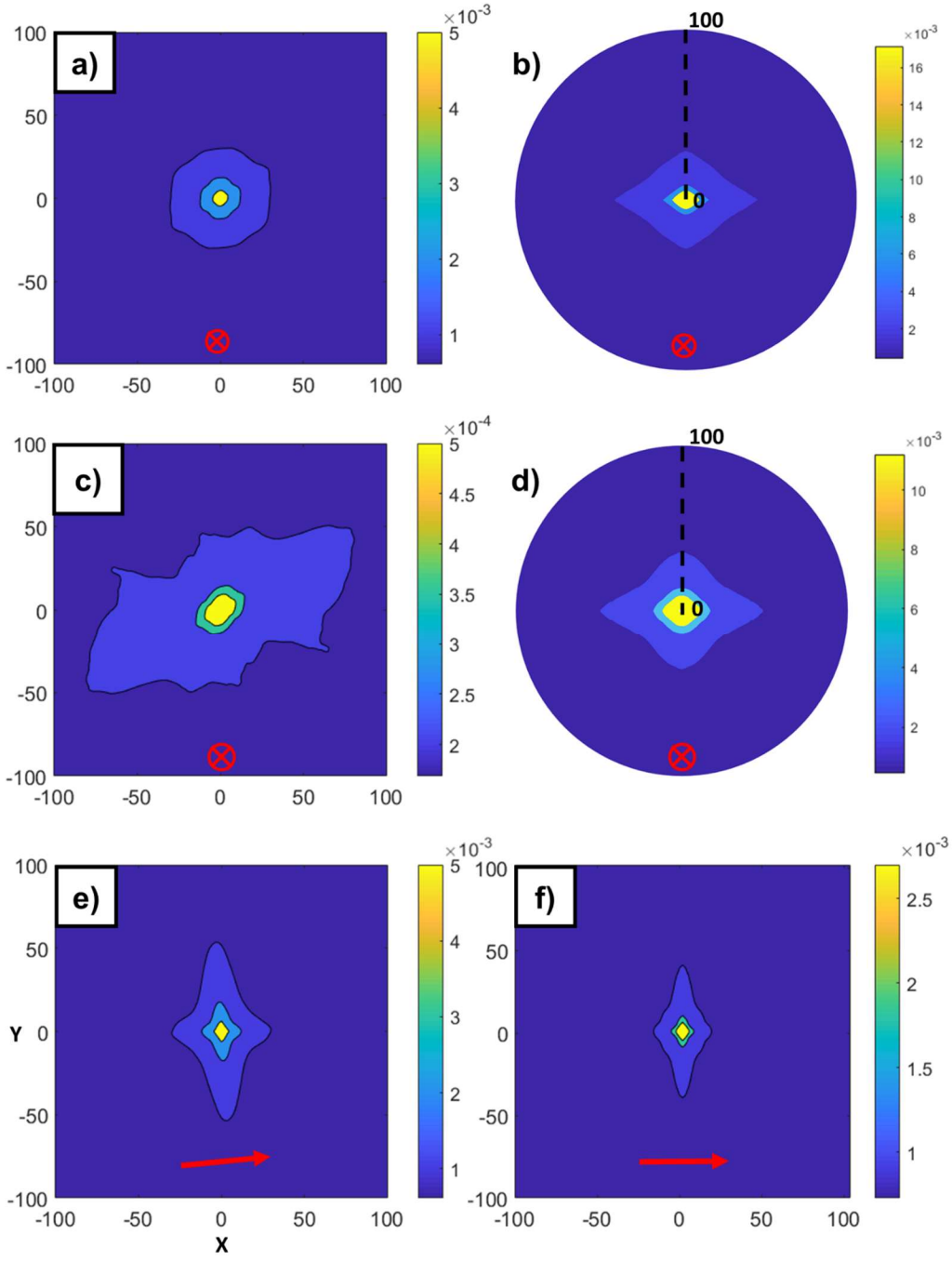

Figure 40 – Comparison of cross-sectional slices of the mean and first principal component basis without(a,c,e) and with(b,d,f) imposed rotational invariance. (a,b) and (c,d) show slices perpendicular to the cylinder axis where the invariance was imposed for the mean and first principal basis respectively. (e,f) shows slices along the cylindrical axis for the mean.

CHAPTER 5. UNIVERSAL MICROSTRUCTURE GENERATION FRAMEWORK

This section derives a foundation for a universal framework of microstructure generation and showcases a few simple examples. Significant research is needed in exploration, extension and utilization of this framework before it can be adopted for widespread use. This task of maturing this framework is left as a future work to next generation of researchers. A single iteration of the framework consists of the following steps: (1) select object to be placed, (2) obtain geometric and statistical cost functions, (3) place object to an optimum location according to some heuristic, (4) check for convergence (see Figure 41). Details of each step are presented in the next sections.

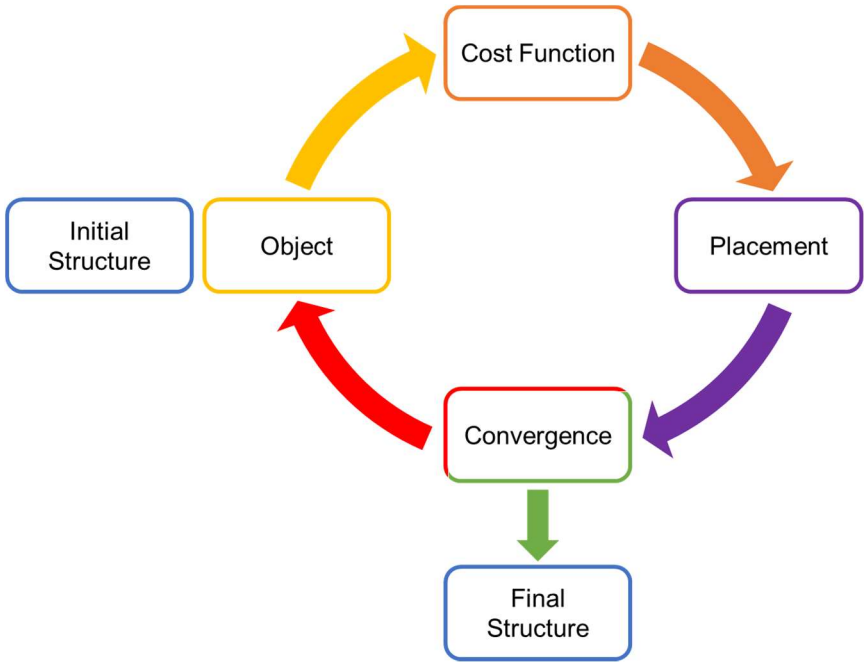

Figure 41 – The main workflow of the generation framework. A single iteration of the framework consists of the following steps: (1) select object to be placed, (2) obtain geometric and statistical cost functions, (3) place object to an optimum location according to some heuristic, (4) check for convergence.

5.1 Initial Structure

In order for the microstructure generation framework to be truly universal, it must have the ability to accept any initial structure definition adaptively. The proposed framework can be utilized in a multitude of modes to ensure any and all such definitions can be accommodated:

1) *Microstructure Building*, in which a structure is generated on a blank canvas. In this mode, the generator can rapidly build many near-optimal structural instances under a variety of constraints.

2) *Microstructure Modulation*, in which a structure is generated from another structure. The input can be an experimentally obtained image, the output of a simulation or outputs from any other pixel/voxel based structure generator in literature.

3) *Microstructure Evolution*, in which an existing structure is evolved under the statistical description of a process, such as the one obtained in the previous chapter for coarsening of an Al-Cu alloy in the form of a principal component basis vector.

5.2 Object

The object selection and design is one of the primary means of injecting prior domain knowledge into the structure generation process and plays a crucial role in obtaining visually realistic microstructures. While it is possible to generate virtually infinitely different patterns or textures with different object selections, for any real materials problem, the set of patterns that can be actualized as a material microstructure through real world natural/chemical or manufacturing constraints is very tightly and

nonlinearly bound (albeit still containing practically infinite variations). This requirement of practical manufacturability or replicability is the primary differentiator of a microstructure generator from common texture synthesis methods found abundantly in literature. Regardless of the guidelines described here, proper object selection requires extensive study and is a major future work area in further development of this framework.

The most basic choice of a single pixel as an object in conjunction with the practicality requirement often leads to the hardest optimization task due to a combinatorial explosion on the order of the number of pixels in an image. This choice should be appropriate either when the structure can be manufactured at the same resolution via additive manufacturing, or when the structure is a result of a complex dynamic process such as diffusion or solidification. Objects on the order of a few pixels are also appropriate when the objective is the evolution of an existing structure according to a statistical process description.

Common geometric shapes are another class of objects that are of interest to materials researches due to the fact that most models and theory developed before the availability of advanced imaging and computing resources heavily rely on them in their derivation. In fact, randomly generated patterns of non-overlapping circles is perhaps the most popular problem of this kind in almost all scientific fields.

A significant differentiator of this framework is the ability accommodate object libraries mined from existing images of similar material structures. Availability of such a library significantly reduces the degrees of freedom in the optimization process, and will naturally result in the most realistic generation of randomized microstructures theoretically

possible. The object library can be enriched by variations of an object under various linear transforms such as rotation and scaling where applicable. Mechanisms of automated extraction of regions of interest from image libraries is a heavily studied field in computer vision, and not in the scope of this thesis.

Finally, objects can also be handcrafted in order to accommodate very elaborate modes of microstructure building such as directional overlaps, Lego-like connections, dendrite-axon type connections, perfect alignments etc. This requires the definition of intermediate local states that "paint" a specific region on the object for tighter constraints. This is further explained in the next section.

5.3 Cost Functions

Microstructure instances are generated by a combination of physical/geometric constraints to accommodate manufacturing/domain expertise and statistical constraints to measure similarity. The central premise of the generation framework is that an incredible variety of microstructures with very tight and specific constraints can be generated entirely with $O(S \log S)$ computational complexity using filters, where N is the number of pixels in the image. Note that under this regime, the computational cost is independent of the state of the image (i.e. how many objects are placed) or the type or size of the object. Within the scope of this thesis, 3 types of cost functions will be discussed.

5.3.1 Overlap Based Cost Functions

The first and most notorious of the cost functions is the overlap between objects. The overwhelming majority of pattern or microstructure generators in literature follow a

brute force approach. A random location is chosen in the image and placement of the object is attempted at that location. If there is any overlap, a new random location is chosen and the attempt is repeated until there is no overlap, then the process continues until specified number of objects are placed. Numerous disadvantages of such an approach can immediately be spotted:

- The cost of placing a new object increases as more objects are placed, due to more frequent failed placement attempts. This is bounded roughly at $O(S^2)$.
- Only 1 object can be placed at a single iteration.
- The placement becomes even more costly and unstable as the object size increases.
- The complexity increases with a combinatorial explosion, if partial overlaps at specific percentages, or overlaps at specific locations are desired.
- The complexity increases with a combinatorial explosion if objects are to be placed at a specific distance from each other.

Conversely, in a filtering based approach:

- The cost of placing a new object is always $O(SlogS)$.
- Multiple objects can be placed simultaneously in a single iteration, especially in earlier stages.
- The placement cost is independent of object topology or size.
- Shapes can be placed at partial overlaps such as a percentage, and can be made to overlap at a particular location or direction while remaining at $O(SlogS)$ complexity.
- Shapes can be placed with specified distances while remaining at $O(SlogS)$ complexity.

The filtering approach is described pictorially in 2D in Figure 42. The object is convolved with the initial structure and normalized by its size in pixel/voxels, which results in an overlap cost map. Each pixel in this map describes the percent overlap of the new object with existing objects, if the new object were to be placed centered on that pixel. To provide context, we can overlay the existing objects over this map and normalize, resulting in the Overlap Cost Function (OCF). The OCF can be thresholded and run through a distance transform in linear time to obtain a signed distance function. In this image each pixel now contains the distance from the nearest object already in the image (away from the nearest object if positive, overlapping the nearest object if negative), if the new object were to be placed centered on that pixel. Overlaying the existing objects over this image and normalizing will yield the Distance Cost Function (DCF). This approach can accommodate microstructures with multiple or continuous local state definitions without an increase in complexity. Specifically, the OCF is defined as:

$$OCF = \frac{\Im^{-1}\left(\Im(Image^{h})^{*} \odot \Im\left(Object^{h'}\right)\right)}{ObjectSize} \tag{40}$$

The DCF is obtained as follows:

$$DCF_{s}^{+} = \min_{s'} \operatorname{dist}(s, s'), \qquad s \in OCF = 0\%, \qquad s' \in OCF > 0\% \tag{41}$$

$$DCF_{s}^{-} = \min_{s'} \operatorname{dist}(s, s'), \qquad s \in OCF > 0\%, \qquad s' \in OCF = 0\% \tag{42}$$

$$DCF = DCF_s^+ + DCF_s^- \tag{43}$$

where for the purposes of this thesis, the function dist() is the Euclidean distance function $\text{dist}(s, s') = \|s - s'\|$. However other measures of distance can easily be accommodated.

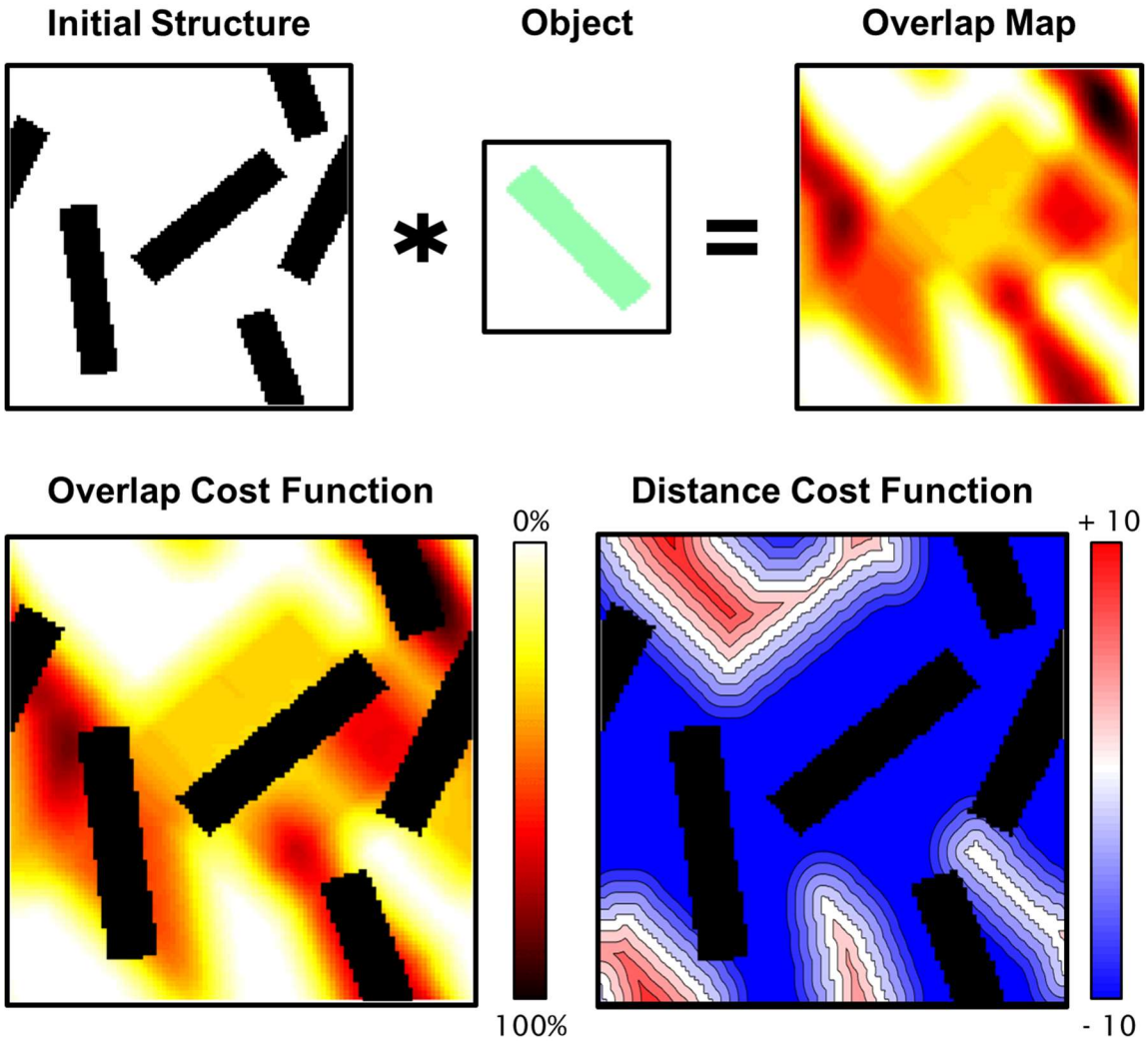

Figure 42 – Graphic depiction of the operations to obtain the overlap based cost functions OCF and DCF.

An incredibly wide variety of microstructures or patterns can be generated simply by joining various overlap and distance constraints. Figure 43 shows an example of how overlap and distance constraints can be used in conjunction to create different interactions

between objects. In this case, the blue object is already placed and the green object is to be placed. Only providing a no-overlap constraint can result in very unpredictable relative placement, which can be alleviated by imposing a distance range to the placement. Only providing an overlap percentage can result in unintended behavior as well, such as two objects joining slightly along a single edge. In these cases, imposing a negative distance constraint will ensure the new object is intruding at least a certain amount into the existing object.

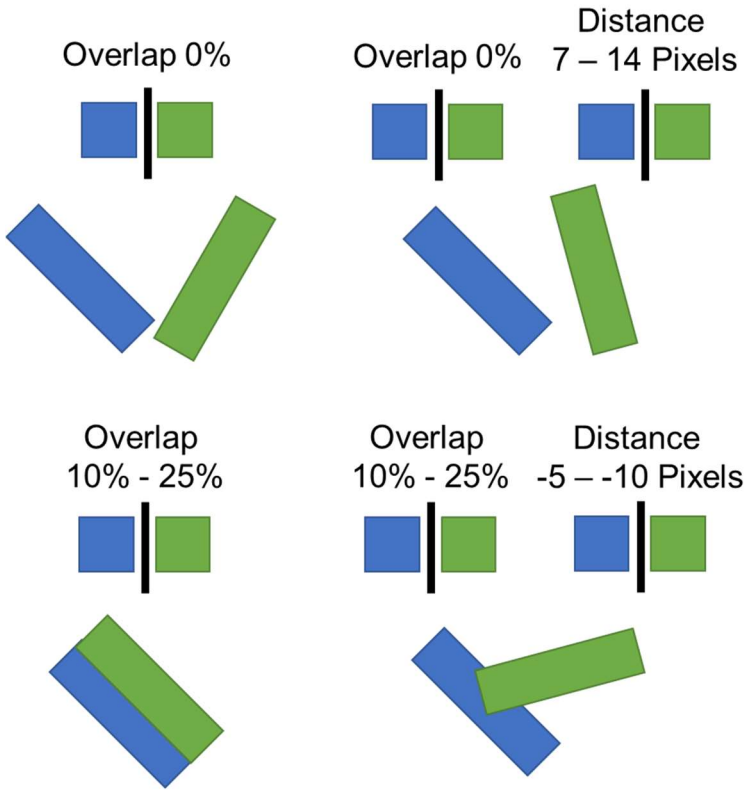

Figure 43 – Examples showing various placement modes that can be achieved using an OCF and a DCF in conjunction. For each example, the blue rectangle is already in the microstructure, and the green rectangle is to be placed.

A more elaborate example is shown in Figure 44 to demonstrate the intricacy, complexity and specificity of the structures that can be generated simply by chaining overlap and distance based constraints. In this case, we have more elaborately designed

objects with multiple meta local states that color different regions of interest. The objective is to make sure objects connect only on the yellow phases without overlap in any other phase, while also maintaining a wide angle between connected objects. This is enforced first by a yellow-yellow overlap constraint, followed by a yellow to non-yellow and non-yellow to non-yellow no overlap constraint. The angle of incidence is bound using a minimum separation distance constraint to the green/blue body of the objects.

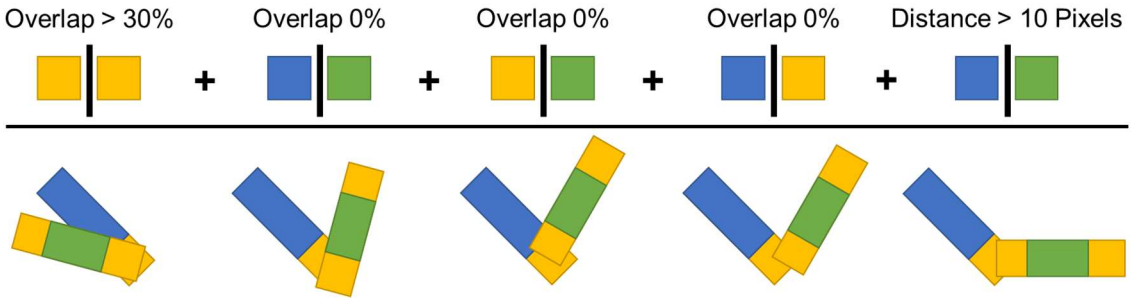

Figure 44 – An examples showing how the placement mode evolves as 4 OCFs and a DCF is chained to achieve a very specific criteria. For each example, the blue rectangle is already in the microstructure, and the green rectangle is to be placed. Yellow areas are meta-states to identify the tips of the rectangles. The final goal is to achieve a placement where rectangles overlap exclusively in the yellow regions, while having a wide angle between them.

5.3.2 2-Point Statistics Based Cost Functions

Before the derivation of the Statistical Cost Function (SCF), it is important to quantify the effect of an individual pixel/voxel to the overall set of 2-point statistics. Recall that the set of all two point statistics is given by:

$$f_t^{hh'} = \frac{\sum_s \tilde{m}_s^h \tilde{m}_{s+t}^{h'}}{\sum_s c_s c_{s+t}}, \qquad \forall t \in T \tag{44}$$

where T is the set of all vectors considered. We can limit our attention to only $\widetilde{m}_s^h$ and quantify the effect a particular spatial bin has on the overall set of statistics as:

$$\widetilde{m}_x^h \widetilde{m}_{x+t}^{h'}, \qquad \forall t \in T \tag{45}$$

This expression is essentially defining a point x in the microstructure with respect to every point (including itself) in the microstructure, bounded by a neighborhood T. Therefore an analogous definition of 2-point statistics would be:

> *The frequency that a prescribed vector **t** appears with a local state **n** at the tail and **p** at the head, if each point in the microstructure were to be defined with respect to every point in the microstructure (via a vector), bounded by a neighborhood **T**.*

This definition allows the explicit identification the exact contribution of a single pixel to the overall set of 2-point statistics. Suppose we would like to observe the effect of changing one particular spatial bin in the structure shown in Figure 45 from the blue state to yellow state. Figure 45.a shows a red pixel selected in the image to be painted yellow. The change in spatial statistics can simply be obtained by centering the image on this pixel (the green frame in Figure 45.a) in consideration of the boundary conditions. The statistics of the new image is now simply the summation of the statistics of the old image with the image contained by the green frame normalized by the appropriate S_t for each t, where the yellow state is represented by 1s and blue state is represented by 0s. (Figure 45.b) . Note that the normalization of vectors are unaffected by changes in the structure and are defined by the boundary conditions and c_s.

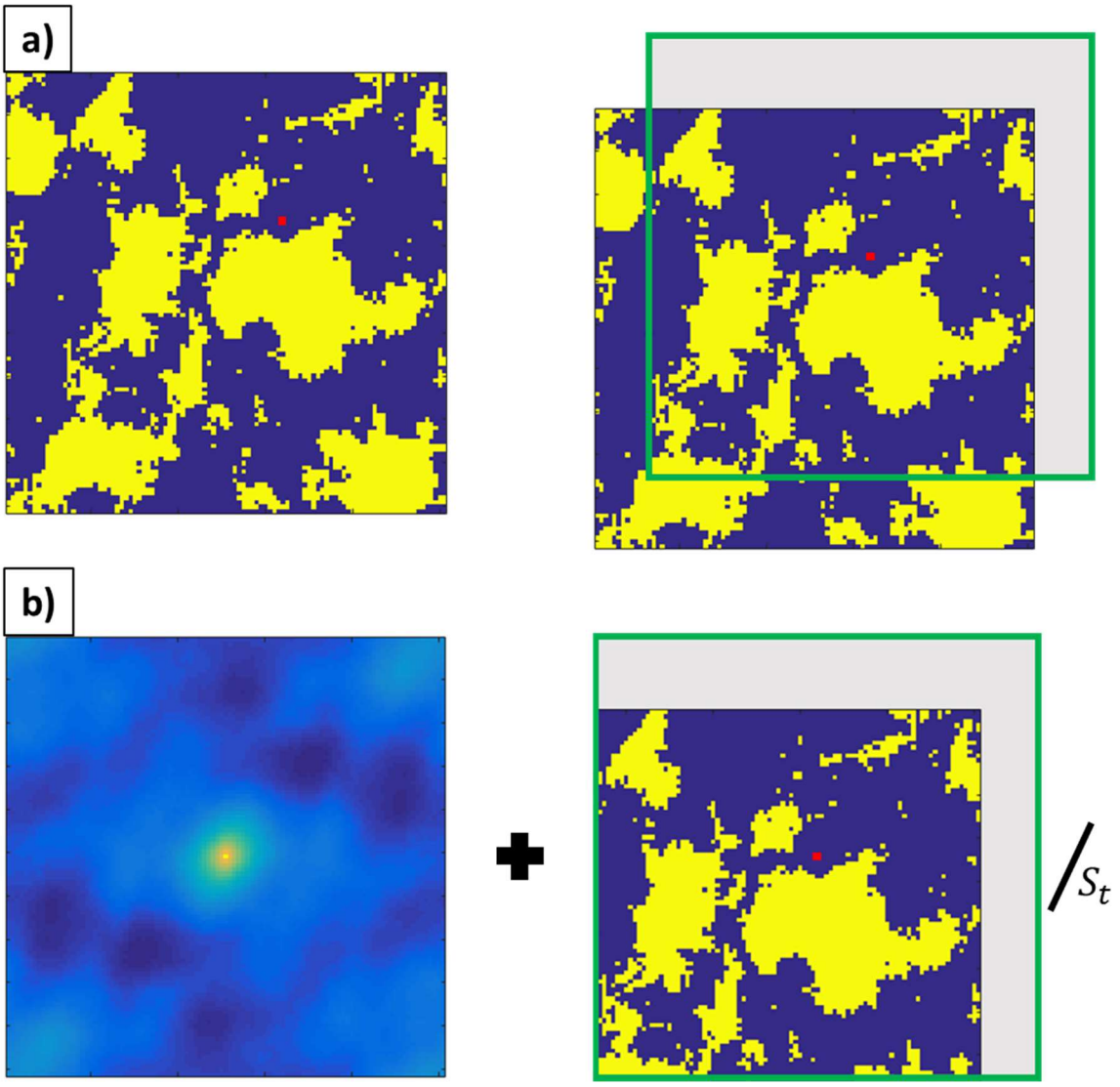

Figure 45 – An example of the Rank 1 Update process for spatial statistics.

In order to make the generation process conform to a particular set of spatial statistics, it is necessary to quantify how much closer the spatial statistics of the current microstructure will approach to the target statistics, if a particular pixel is filled with a particular phase. Let us assume that we have a microstructure that only needs a single pixel modification to reach the target 2-point statistics. In this case, by applying the concept from Figure 45 we can test the effect of modifying each pixel until we find the right one (Figure 46.a). However, another way achieving the same effect is to instead find the difference

between the target and current statistics to obtain the ideal neighborhood directly, and searching the microstructure for this neighborhood (Figure 46.b). While straightforward application of both concepts wield yield a search of $O(S^2)$ complexity, the workflow described in Figure 46.b can be actualized by convolving the ideal neighborhood with the current microstructure, resulting in an $O(SlogS)$ complexity.

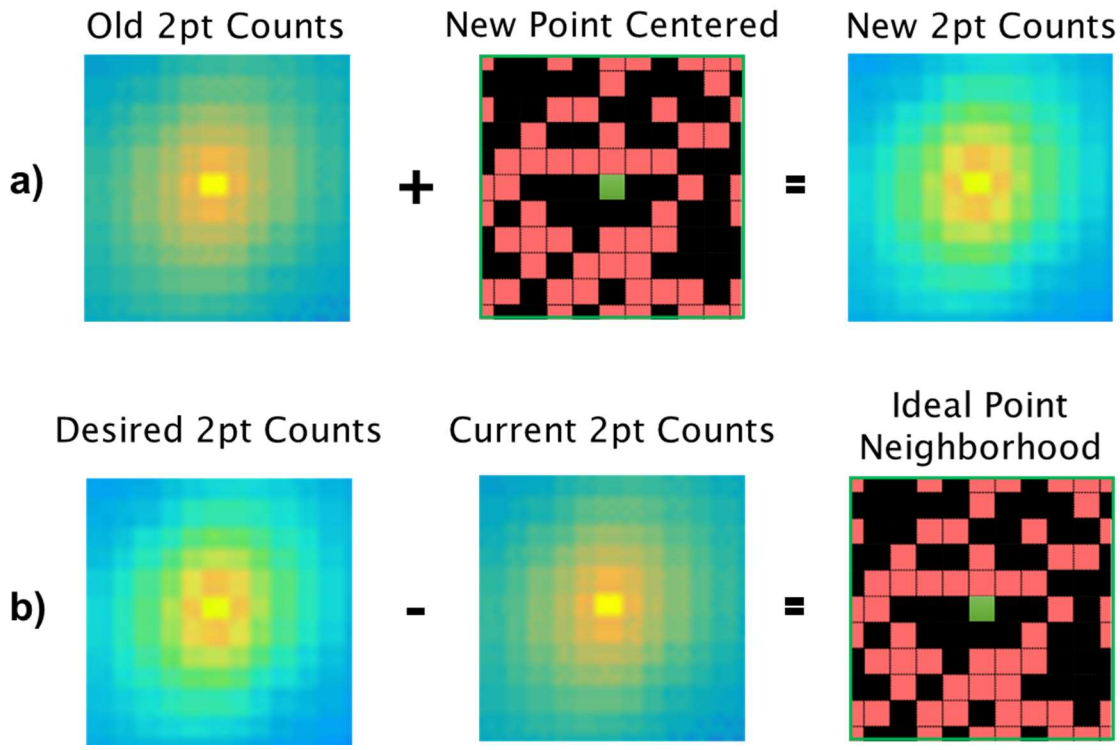

Figure 46 – a) The rank-1 update concept for 2-point statistics is shown graphically. b) The graphical equation is re-arranged to yield the ideal neighborhood a pixel to be placed should possess.

This concept can be generalized to microstructures that require heavy modification to reach desired 2-point statistics by treating the difference in statistics as a weighted filter describing the potential gain towards the target statistics. In this filter, larger weights will signify high priority neighborhoods. As such modifications to the structure should

prioritize pixels that contain a large quantity of high priority neighborhoods. Such a filter will be referred henceforth as the Placement Gain Filter (PGF).

It is now possible to demonstrate the calculation and utilization of the PGF for the placement of an object to an existing microstructure. Figure 47 shows an initial structure and an object to placed, as well as both of their 2-point autocorrelation plots. Figure 48 pictorially outlines the necessary operations to obtain a proper PCF for this object and initial structure pair, which when chained together can be expressed as:

$$PGF = \left(\bar{f}^{\,hh'} - \Im^{-1}\left(\Im(Image^{h})^{*} \odot \Im\left(Image^{h'}\right)\right) - \Im^{-1}\left(\Im(Object^{h})^{*} \odot \Im\left(Object^{h'}\right)\right)\right) \odot W \tag{46}$$

where $\bar{f}^{hh'}$ is the target 2-point statistics and $W$is the matrix of neighborhood weights. Wenables the generation process to prioritize a specific subset of the statistics, such as vectors of a particular length or orientation, based on physics/manufacturing requirements. The subtraction of the object 2-point statistics from the target assumes that the object will be placed without overlap, however the error introduced is minimal in most overlapping cases. Calculation of PGF only contains element wise products, subtractions and Fourier transforms for an $O(SlogS)$ complexity.

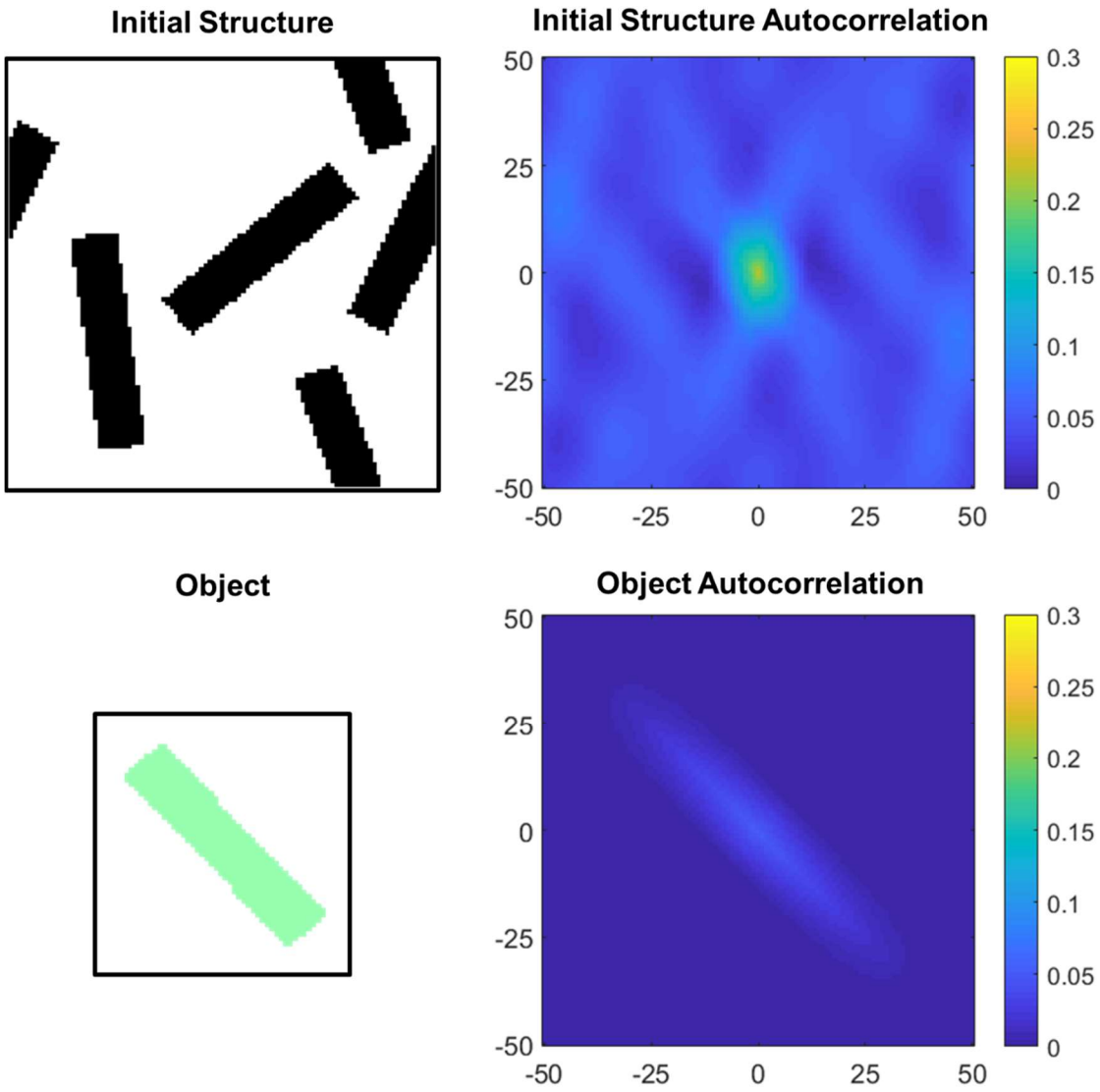

Figure 47 – The example computation of the statistical cost function uses the Initial Structure, the Object, and their autocorrelations. A formulation that uses cross-correlations can trivially be derived.

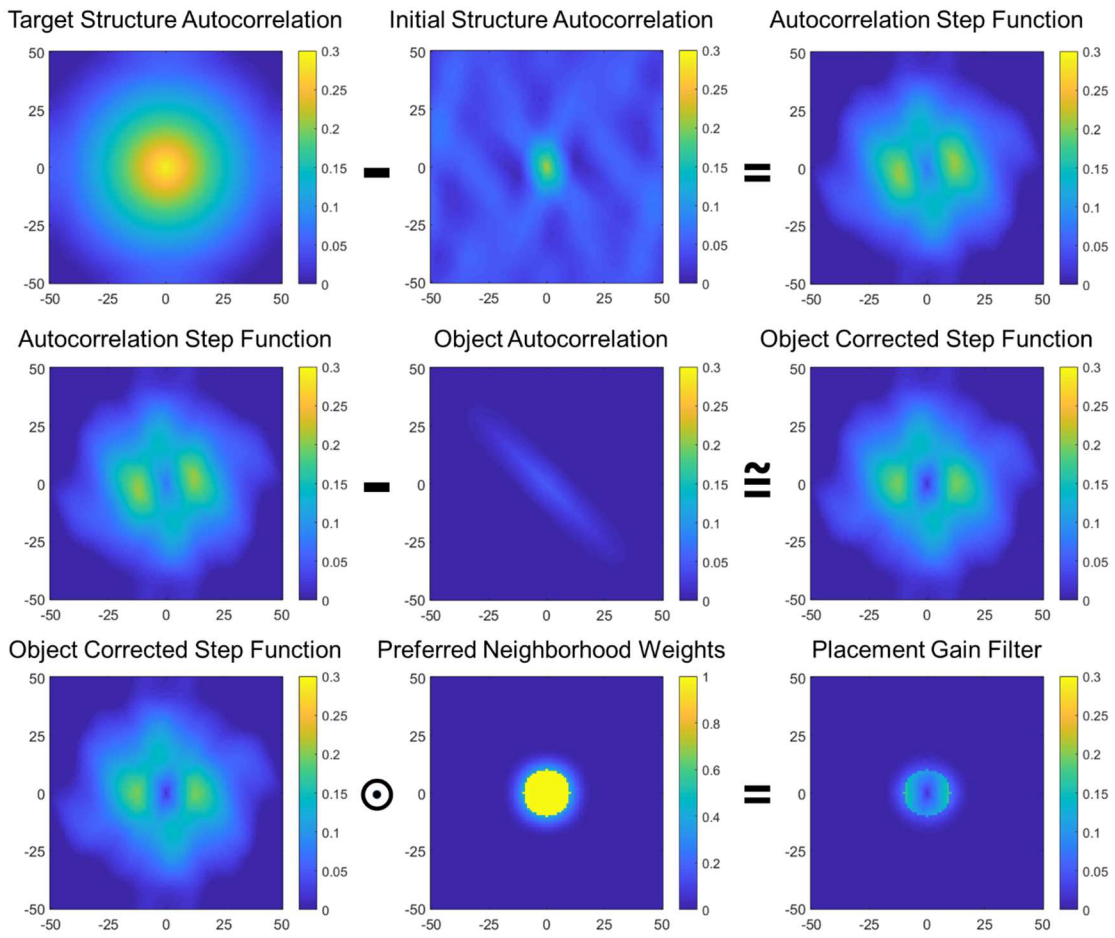

Figure 48 – Graphical depiction of the operations involved in the calculation of the PFG. (See Eq. 46).

In order to utilize the obtained PGF to calculate the Statistical Cost Function, two more convolutions and an element-wise multiplication are necessary, thus the Fourier transform remains the operation with the highest complexity. SCF can be found as:

$$SCF = \Im^{-1}(\Im(\Im^{-1}(\Im(Image^h)^* \odot \Im(\mathrm{PGF})) \odot Image^h)^* \odot \Im(Object^h)) \quad (47)$$

This operation is pictorially shown in Figure 49 for clarity.

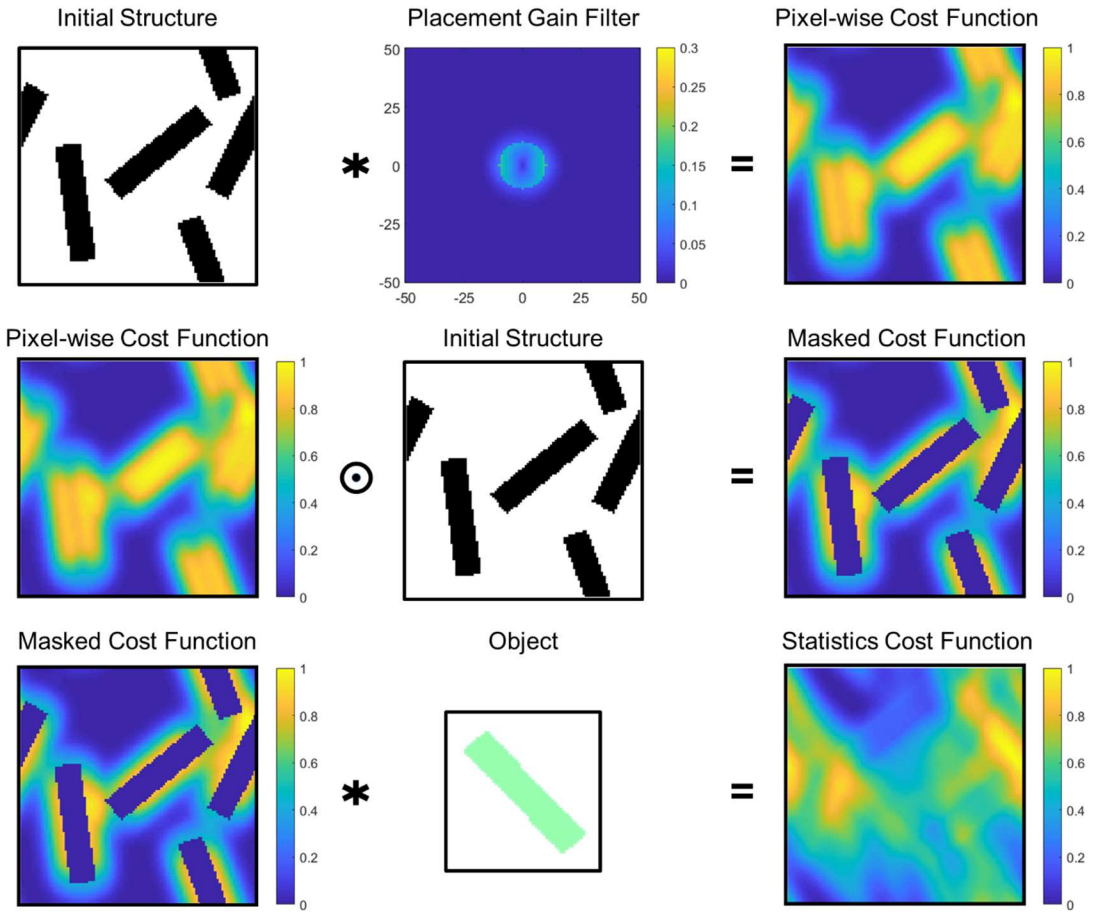

Figure 49 – Graphical depiction of the operations involved in the calculation of the SCF. (See Eq. 47).

5.4 Object Placement

The next step in a given generation iteration after cost function calculation is the actual placement of the object. There are two key stages to this decision both of which have centuries of history in the field of optimization and will not be explored in depth. The first one involves the merging of information from the OCF, DCF and SCF (Figure 50). This can be achieved in various ways such as addition (soft constraints), multiplication (hard constraints) or weighted independent voting. The examples in this thesis are generated using multiplicative joining so that:

$$Joint\ Cost\ Function = OCF \odot \text{DCF} \odot \text{SCF} \tag{48}$$

Note that the joint cost function can involve multiple OCFs, DCFs and SCFs corresponding to different local states. The second stage involves the selection of one or more points the actually place the object according to the joint cost function. To the very ill-posed nature of most microstructure generation problems, it is usually desirable to allow for random variation in the selection process so that the generation doesn't directly converge to the nearest local minima. The examples here treat the normalized joint cost function as a probability distribution and draw a sample for placement.

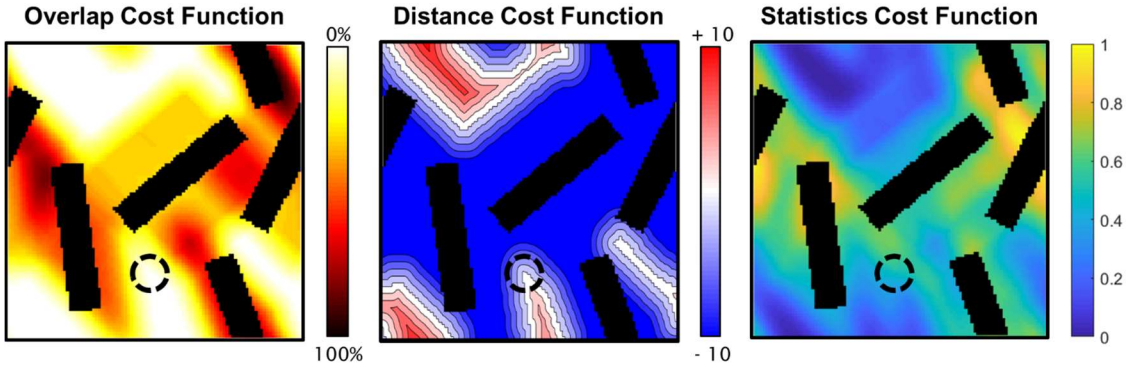

Figure 50 – Joint consideration of all 3 example cost functions computed. In a real case, there can be multiple OCFs, DCFs and SCFs.

5.5 Convergence

The final consideration in the microstructure generation framework is the stopping criteria. Either 1-point (equivalent to volume fraction or number of objects) or 2-point statistics convergence can be used to terminate the generation process. As with placement heuristics, error metrics and convergence methods are extremely abundant in optimization

literature and will not be explored here. For the examples here the following error metric is used when statistical criteria are imposed:

$$error = \frac{\left\| (\overline{f}^{\,hh'} - f^{hh'}) \odot \mathrm{W} \right\|}{\left\| \overline{f}^{\,hh'} \odot \mathrm{W} \right\|} \tag{49}$$

where S is the number of spatial bins in the image and W is the neighborhood weight matrix from the PGF calculation.

5.6 Examples

This section presents some brief examples to demonstrate the utility and performance of the microstructure generation framework and the underlying concepts.

5.6.1 Versatility of Overlap Based Constraints

First, let us focus our attention to microstructures generated only under physical constraints without statistical guidance. The framework is set up in the following manner:

1) Start with a blank canvas.
2) While number of objects placed is smaller than number of objects desired:
    a. Randomly rotate the object.
    b. Compute Overlap Cost Functions with desired overlap range.
    c. Compute Distance Cost Functions with desired distance range.
    d. Multiplicatively join the OCFs and DCFs.
    e. Sample a point to place the object assuming the resulting cost function represents a probability distribution (with proper normalization).

Figure 51 shows 6 outputs under generated the current setup. Figure 51.a is generated by using non-overlapping circles. The is the only microstructure out of the 6 that can be generated by the brute force method in $O(S^2)$, while the others are bounded roughly at $O(S^3)$ under brute force regime. Figure 51.b shows a microstructure generated using non-overlapping circles with a prescribed minimum separation, while Figure 51.c shows an example using circles that overlap by 5% to 25% of their area. Figure 51.d,e,f are generated using shapes with meta-states like the ones shown in Figure 44. Figure 51.d is made by ensuring wider angled overlap between tips of placed rectangles, while Figure 51.e enforces overlap between the centers of already placed rectangles and tips of rectangles to be placed. Figure 51.f is first generated using random non-overlapping circles, then modified by placing rectangles with tip regions that overlap the existing circles.

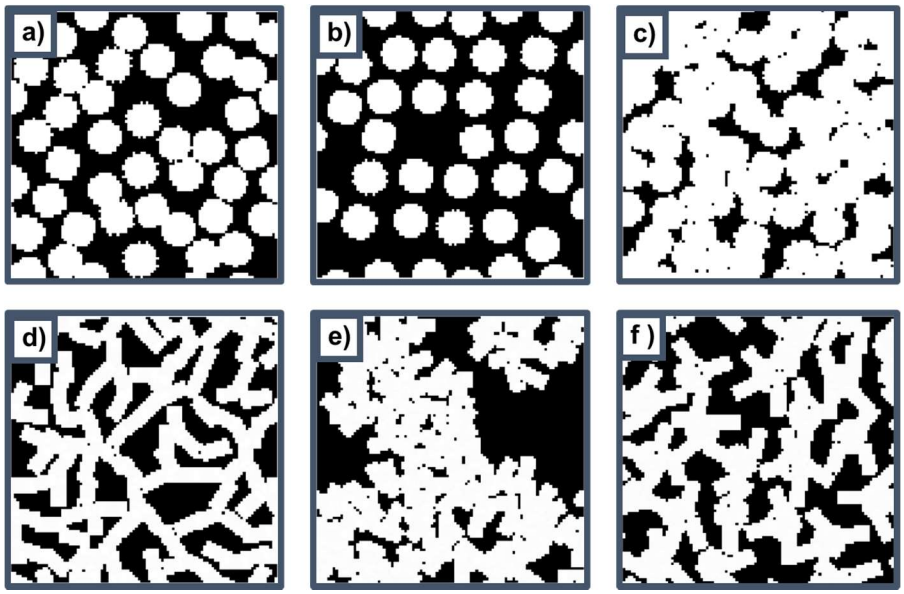

Figure 51 – Various examples of microstructures generated using only overlap and distance constraints.

The variety in these structures demonstrate the versatility that can be achieved by the chaining of overlap and distance constraints, even when only basic geometric shapes

are used. With proper object selection or the utilization of object libraries, versatility will be even further improved.

5.6.2 Generation under Statistical Constraints

The next example demonstrates microstructure generation under 2-point statistical guidance. Figure 52.a shows an experimental image obtained by Scanning Electron Microscopy (SEM). We will assume now that this image is extremely costly to acquire, and we only have this one sample to work with. The task is to generate additional microstructure instances that could have come from the same population this sample was taken (a very tall order with only one image). We first segment the image in to "flakes" (white) and background (black) states (see. Figure 52.b), then we follow the microstructure generation procedure in the following way:

1) Start with a blank canvas.
2) While the volume fraction of the current image is less than the target image:
    a. Sample an object randomly from the object library
    b. Randomly rotate the object
    c. Compute Overlap Cost Functions with 0% to 50% overlap range, as suggested by a visual inspection of the SEM image.
    d. Compute Statistical Cost Functions, set target as the statistics of the segmented SEM image.
    e. Multiplicatively join the OCFs and SCFs.
    f. Sample a point to place the object assuming the resulting cost function represents a probability distribution (with proper normalization).

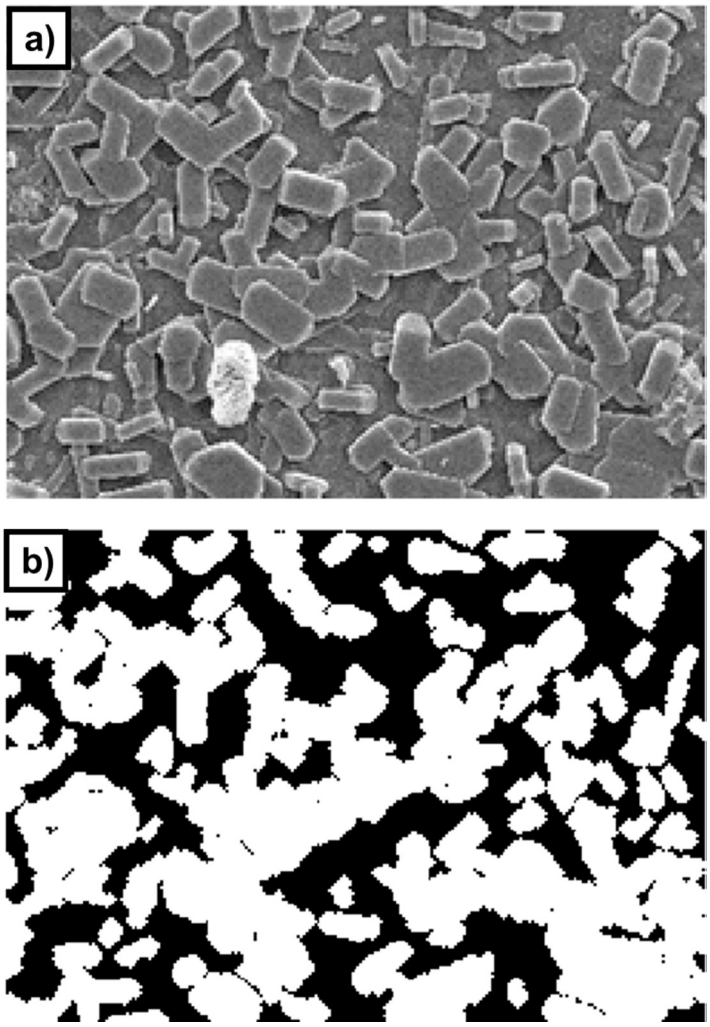

Figure 52 – An experimental SEM image and its segmentation used as the statistical target in the generation process. (Reused with permission from [73])

The information from the original sample is used in two ways: (i) the 2-point statistics of the image is used as the target statistics, (ii) an object library is mined from the image. A library consisting of 36 objects was manually mined for this example, and is shown in Figure 53. Notice that while there are objects of various size and shapes in the library, all objects predominantly contain rounded features.

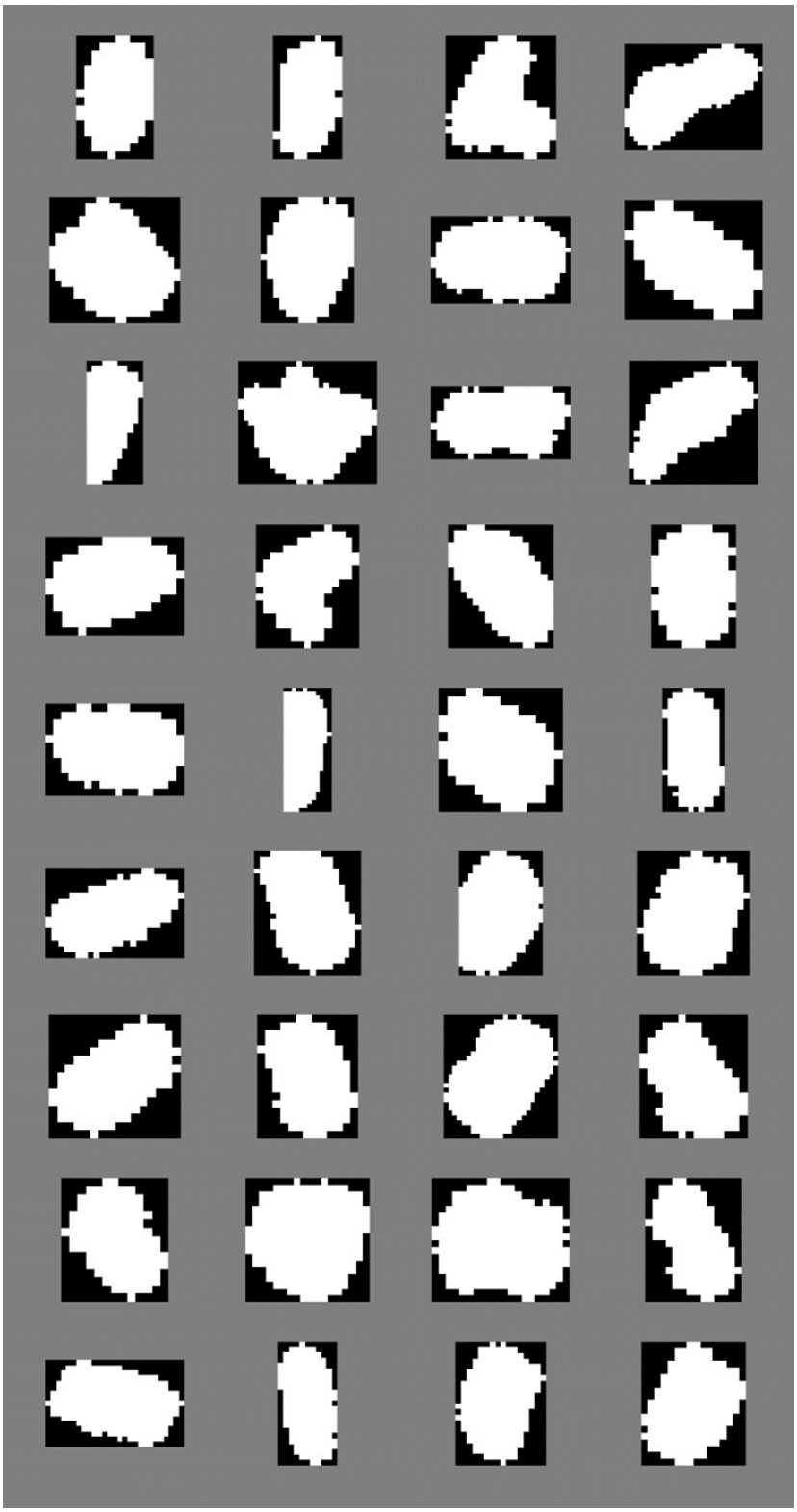

Figure 53 – The object library mined from Figure 52 for the statistically guided microstructure generation example.

Figure 54 shows 3 examples that were generated using the procedure described earlier. Notice that while looking significantly different overall, all 3 examples share many common features with the original sample including but not limited to: (i) the presence and frequency of small island-like formations, (ii) presence of major directional chains composed of many overlapping objects, (iii) occasional large gaps between features. While the visual judgment of whether is it believable that these 3 examples could have come from the same population as the SEM sample is left to the reader, we can investigate further to see if such a relationship is statistically believable.

Figure 55.a shows the error in 2-point statistics compared to the original sample, averaged over 100 generated microstructures. It can be seen that the error in statistics has some very particular trends. To inspect whether this error comes from the object library or overlap conditions, we can generate another 100 microstructures without any statistical guidance and obtain their average error in statistics, which is shown in Figure 55.b. Comparing the two average error plots, Figure 55.c quantifies the effect of the inclusion of the statistical constraints in the generation process. Note that the negative and positive error peaks outside of the small range vectors were not affected by the inclusion of the SCF.

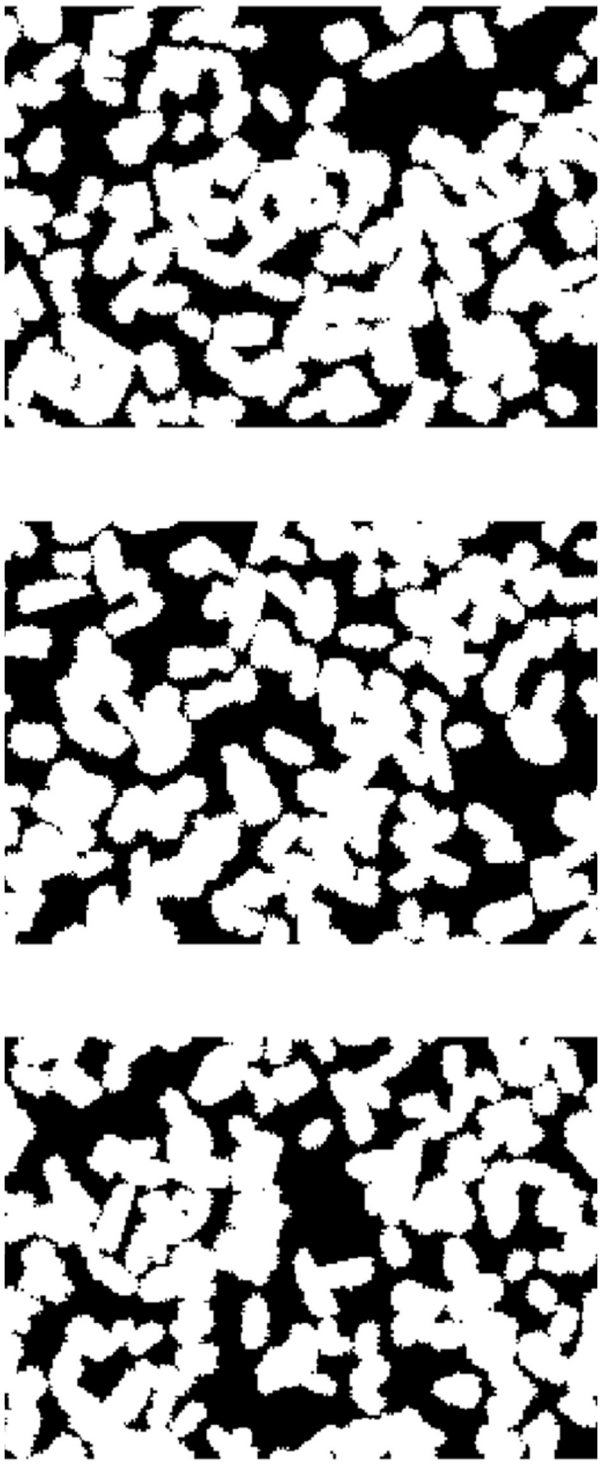

Figure 54 – 3 examples of microstructures generated using the mined object library under statistical guidance.

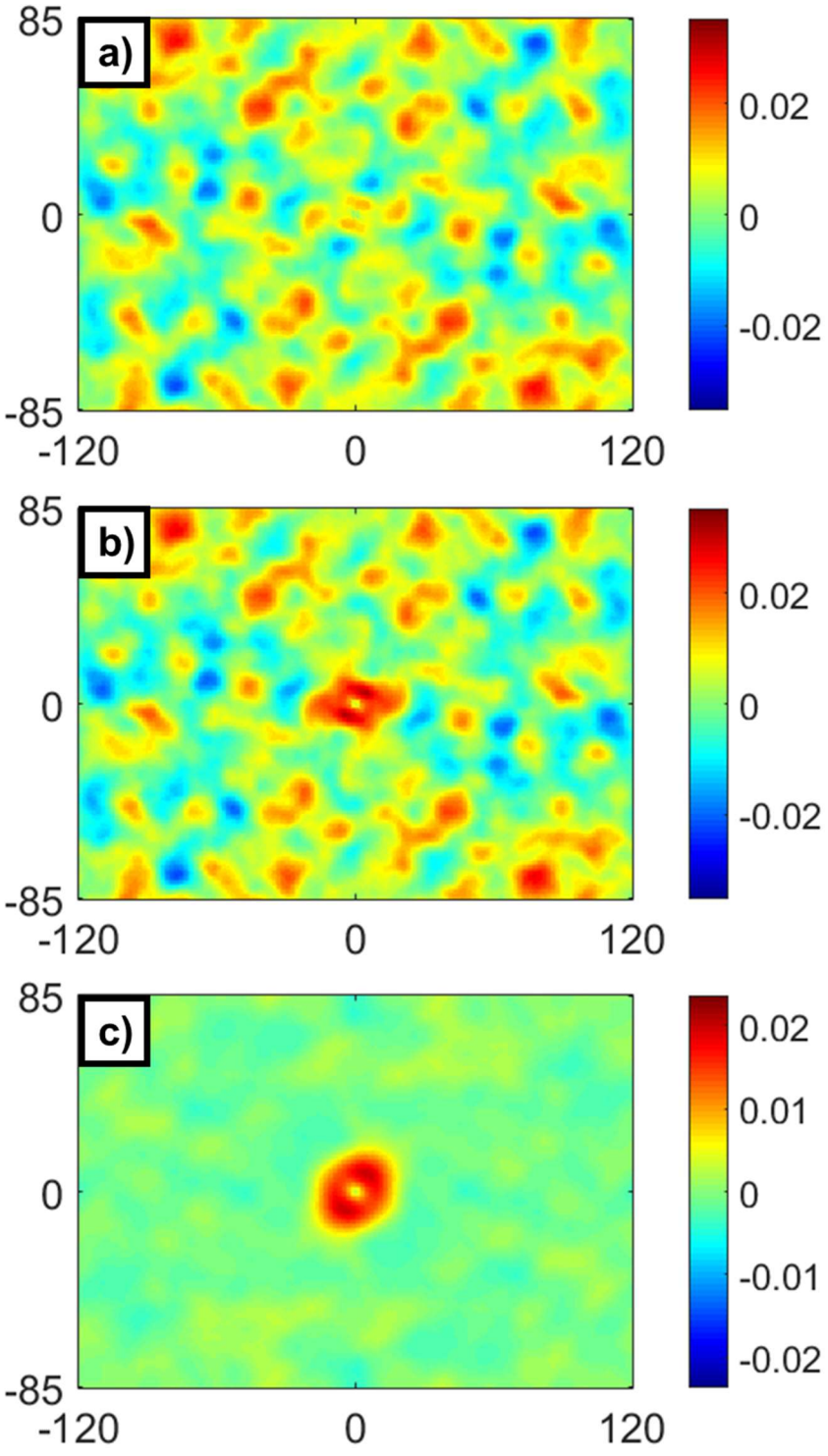

Figure 55 – a) The error in 2-point statistics with respect to the reference, when SCF is utilized. b) The error in 2-point statistics with respect to the reference, when SCF is not utilized. c) The difference in error between a) and b).

To reveal the reason why these peaks exist in their specific locations, we can inspect the 2-point statistics of the SEM sample (Figure 56.a). Figure 56.b shows a modified version of Figure 56.a where values greater than the average of all 2-point statistics plus the standard deviation of the 2-point statistics are colored dark red, and values smaller than the average of all 2-point statistics minus the standard deviation of the 2-point statistics are colored dark blue. Figure 56.c shows these peaks overlaid on top of the error plot in Figure 55.a. Notice that they align perfectly. We can thus conclude that these very specific trends in error peaks are a result of the placement heuristic of the objects into the structure. The error in reaching the target of a 2-point statistic corresponding to a particular vector, is proportional to its deviation from the mean of all 2-point statistics. This means that over enough placements, the current placement heuristic is governed by the average trend. In order to achieve a more homogenous error profile across all vectors, a placement heuristic that shows some consistent bias towards high error regions is necessary.

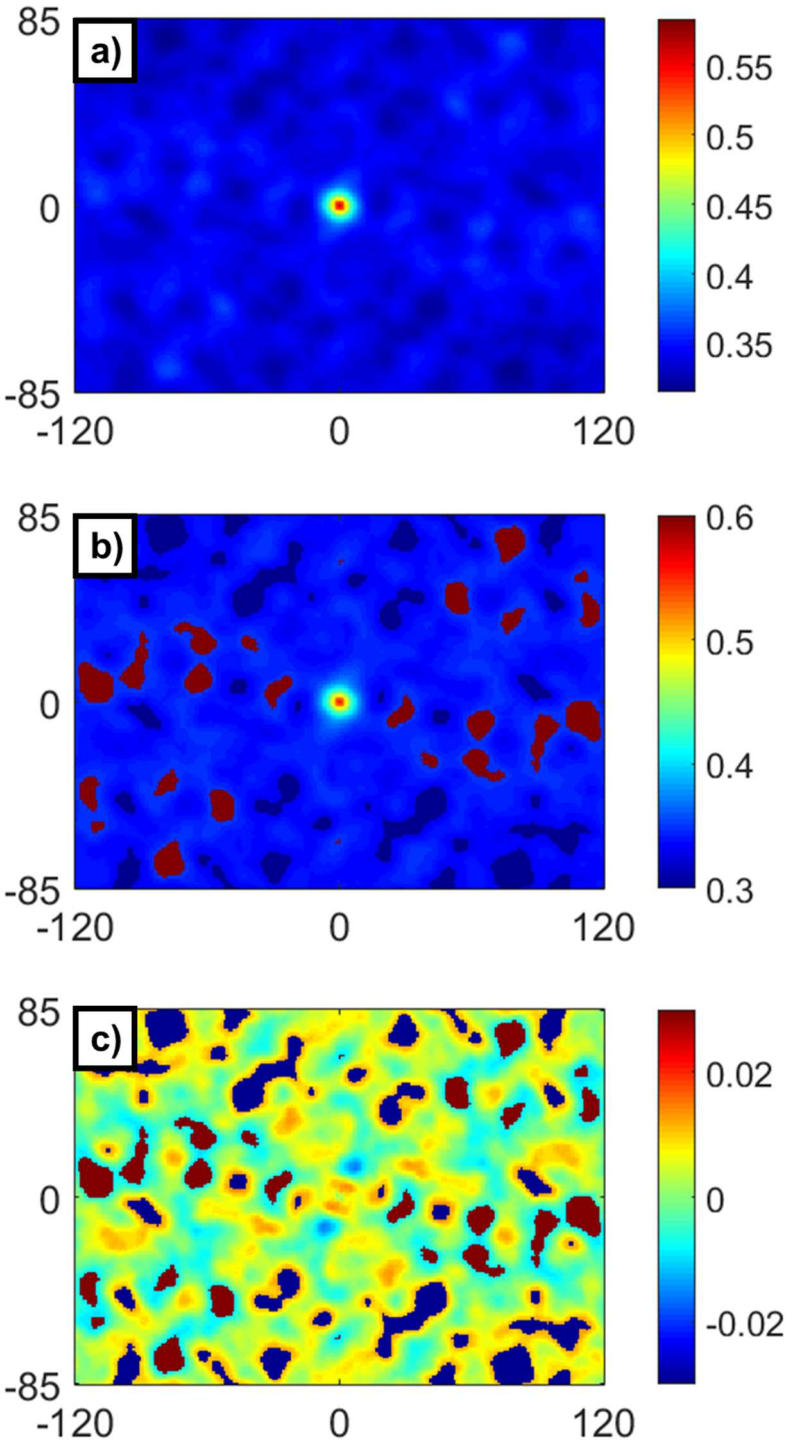

Figure 56 – a) The 2-point statistics of the reference. b) The 2-point statistics of the reference with the regions above and below a standard deviation away from the mean are colored dark red and dark blue respectively. c) The same highlighted regions overlaid on the error in 2-point statistics of generation. Notice that the peaks in error and the highlighted region align perfectly.

CHAPTER 6. CONCLUSIONS AND FUTURE WORK

6.1 Efficient Tools to Map Real Microstructures to Low Dimensional Descriptors

Extensions to the spatial statistics computational framework to accommodate datasets with irregularities and deficiencies such as non-rectangular boundaries, missing data, continuous local states and susceptibility to observer rotations have been described. Strategies to scale the analysis methods to very large datasets were presented. These attributes are very common in datasets obtained via experiments and imaging equipment. The improvements in this work should enable the rapid and seamless analysis of real datasets in a plethora of material systems across a wide range of scientific fields. The issues addressed here were the most frequently encountered in real datasets that the author have been exposed to. Future work will be required as additional complications from real data sources arise, such as a unique consideration for a particular materials system or unique representation issues for a new experimental data acquisition method.

6.2 Universal Microstructure Generation Framework

A framework enabling the rapid and scalable generation of microstructures under geometrical and statistical constraints have been assembled with examples to showcase its versatility. This framework requires heavy future work, perhaps in the form of a self-contained dissertation, before it can be utilized by the common materials researcher. There are numerous hyper-parameters and optimization alternatives to be explored in-depth through case studies on specific material systems. Upon completion, the framework would shave countless hours off from the development process of new materials.

REFERENCES

1. Torquato, S., *Random heterogeneous materials: microstructure and macroscopic properties*. Vol. 16. 2013: Springer Science & Business Media.

2. Jr., W.F.B., *Solid Mixture Permittivities.* The Journal of Chemical Physics, 1955. **23**(8): p. 1514-1517.

3. Popova, E., et al., *Process-Structure Linkages Using a Data Science Approach: Application to Simulated Additive Manufacturing Data.* Integrating Materials and Manufacturing Innovation, 2017. **6**(1): p. 54-68.

4. Gupta, A., et al., *Structure–property linkages using a data science approach: Application to a non-metallic inclusion/steel composite system.* Acta Materialia, 2015. **91**: p. 239-254.

5. Kalidindi, S.R., et al., *Application of Data Science Tools to Quantify and Distinguish between Structures and Models in Molecular Dynamics Datasets.* Nanotechnology, 2015. **submitted**.

6. CeCen, A., et al., *A Data-driven Approach to Establishing Microstructure-Property Relationships in Porous Transport Layers of Polymer Electrolyte Fuel Cells.* Journal of Power Sources, 2014. **245**: p. 144-153.

7. Schröder, J., D. Balzani, and D. Brands, *Approximation of random microstructures by periodic statistically similar representative volume elements based on lineal-path functions.* Archive of Applied Mechanics, 2011. **81**(7): p. 975-997.

8. Rollett, A., et al., *Stress hot spots in viscoplastic deformation of polycrystals.* Modelling and Simulation in Materials Science and Engineering, 2010. **18**(7): p. 074005.

9. Jiao, Y., F.H. Stillinger, and S. Torquato, *A superior descriptor of random textures and its predictive capacity.* Proceedings of the National Academy of Sciences, 2009. **106**(42): p. 17634-17639.

10. Singh, H., et al., *Image based computations of lineal path probability distributions for microstructure representation.* Materials Science and Engineering: A, 2008. **474**(1): p. 104-111.

11. Adams, B.L., X. Gao, and S.R. Kalidindi, *Finite approximations to the second-order properties closure in single phase polycrystals.* Acta Materialia, 2005. **53**(13): p. 3563-3577.

12. Yabansu, Y.C., D.K. Patel, and S.R. Kalidindi, *Calibrated localization relationships for elastic response of polycrystalline aggregates.* Acta Materialia, 2014. **81**: p. 151-160.

13. Yabansu, Y.C. and S.R. Kalidindi, *Representation and calibration of elastic localization kernels for a broad class of cubic polycrystals.* Acta Materialia, 2015. **94**: p. 26-35.

14. Baker, T.J., K.B. Gave, and J.A. Charles, *Inclusion deformation and toughness anisotropy in hot-rolled steels.* Metals Technology, 1976. **3**(1): p. 183-193.

15. Garrison, W.M. and A.L. Wojcieszynski, *A discussion of the effect of inclusion volume fraction on the toughness of steel.* Materials Science and Engineering: A, 2007. **464**(1): p. 321-329.

16. Holzer, L., et al., *Quantitative relationships between composition, particle size, triple phase boundary length and surface area in nickel-cermet anodes for Solid Oxide Fuel Cells.* Journal of Power Sources, 2011. **196**(17): p. 7076-7089.

17. Murakami, Y., *Chapter 6 - Effects of Nonmetallic Inclusions on Fatigue Strength*, in *Metal Fatigue*. 2002, Elsevier Science Ltd: Oxford. p. 75-127.

18. Promentilla, M.A.B., et al., *Quantification of tortuosity in hardened cement pastes using synchrotron-based X-ray computed microtomography.* Cement and Concrete Research, 2009. **39**(6): p. 548-557.

19. Shearing, P.R., et al., *Characterization of the 3-dimensional microstructure of a graphite negative electrode from a Li-ion battery.* Electrochemistry Communications, 2010. **12**(3): p. 374-377.

20. Su, H., et al., *Processing, microstructure and tensile properties of nano-sized Al2O3 particle reinforced aluminum matrix composites.* Materials & Design (1980-2015), 2012. **36**: p. 590-596.

21. van de Lagemaat, J., K.D. Benkstein, and A.J. Frank, *Relation between Particle Coordination Number and Porosity in Nanoparticle Films: Implications to Dye-Sensitized Solar Cells.* The Journal of Physical Chemistry B, 2001. **105**(50): p. 12433-12436.

22. Dimiduk, D.M., et al., *The role of grain size and selected microstructural parameters in strengthening fully lamellar TiAl alloys.* Metallurgical and Materials Transactions A, 1998. **29**(1): p. 37-47.

23. Bei, H. and E.P. George, *Microstructures and mechanical properties of a directionally solidified NiAl–Mo eutectic alloy.* Acta Materialia, 2005. **53**(1): p. 69-77.

24. Li, M., D. Wilkinson, and K. Patchigolla, *Comparison of Particle Size Distributions Measured Using Different Techniques.* Particulate Science and Technology, 2005. **23**(3): p. 265-284.

25. Cecen, A., et al., *3-D microstructure analysis of fuel cell materials: spatial distributions of tortuosity, void size and diffusivity.* Journal of The Electrochemical Society, 2012. **159**(3): p. B299-B307.

26. Gommes, C.J., et al., *Practical methods for measuring the tortuosity of porous materials from binary or gray-tone tomographic reconstructions.* AIChE Journal, 2009. **55**(8): p. 2000-2012.

27. Pharoah, J.G., K. Karan, and W. Sun, *On effective transport coefficients in PEM fuel cell electrodes: Anisotropy of the porous transport layers.* Journal of Power Sources, 2006. **161**(1): p. 214-224.

28. Fullwood, D.T., et al., *Microstructure sensitive design for performance optimization.* Progress in Materials Science, 2010. **55**(6): p. 477-562.

29. Niezgoda, S.R., D.T. Fullwood, and S.R. Kalidindi, *Delineation of the space of 2-point correlations in a composite material system.* Acta Materialia, 2008. **56**(18): p. 5285-5292.

30. Fullwood, D.T., S.R. Niezgoda, and S.R. Kalidindi, *Microstructure reconstructions from 2-point statistics using phase-recovery algorithms.* Acta Materialia, 2008. **56**(5): p. 942-948.

31. Kalidindi, S.R., S.R. Niezgoda, and A.A. Salem, *Microstructure informatics using higher-order statistics and efficient data-mining protocols.* JOM, 2011. **63**(4): p. 34-41.

32. Niezgoda, S.R., et al., *Optimized structure based representative volume element sets reflecting the ensemble-averaged 2-point statistics.* Acta Materialia, 2010. **58**(13): p. 4432-4445.

33. Kalidindi, S.R., *Microstructure Informatics.* Informatics for Materials Science and Engineering: Data-driven Discovery for Accelerated Experimentation and Application, 2013: p. 443.

34. Li, G., et al., *Microstructural analysis of the radial distribution function for liquid and amorphous Al.* Journal of Physics: Condensed Matter, 2003. **15**(14): p. 2259.

35. Park, S.H. and D.N. Lee, *A study on the microstructure and phase transformation of electroless nickel deposits.* Journal of Materials Science, 1988. **23**(5): p. 1643-1654.

36. Yeong, C.L.Y. and S. Torquato, *Reconstructing random media.* Physical Review E, 1998. **57**(1): p. 495-506.

37. Staśko, R., H. Adrian, and A. Adrian, *Effect of nitrogen and vanadium on austenite grain growth kinetics of a low alloy steel.* Materials Characterization, 2006. **56**(4): p. 340-347.

38. Turner, D.M., S.R. Niezgoda, and S.R. Kalidindi, *Efficient computation of the angularly resolved chord length distributions and lineal path functions in large microstructure datasets.* Modelling and Simulation in Materials Science and Engineering, 2016. **24**(7): p. 075002.

39. Yu, W. and K. Erickson, *Chord length characterization using focused beam reflectance measurement probe - methodologies and pitfalls.* Powder Technology, 2008. **185**(1): p. 24-30.

40. Kröner, E., *Statistical modelling*, in *Modelling small deformations of polycrystals*. 1986, Springer. p. 229-291.

41. Milton, G.W., *The theory of composites (Cambridge monographs on applied and computational mathematics).* 2002.

42. Adams, B.L., S.R. Kalidindi, and D. Fullwood, *Microstructure Sensitive Design for Performance Optimization*. 2012: Butterworth-Heinemann.

43. Fullwood, D.T., B.L. Adams, and S.R. Kalidindi, *A strong contrast homogenization formulation for multi-phase anisotropic materials.* Journal of the Mechanics and Physics of Solids, 2008. **56**(6): p. 2287-2297.

44. Adams, B.L., H. Garmestani, and G. Saheli, *Microstructure design of a two phase composite using two-point correlation functions.* Journal of Computer-Aided Materials Design, 2004. **11**: p. 103–115.

45. Garmestani, H., et al., *Statistical Continuum Theory for Large Plastic Deformation of Polycrystalline Materials.* Journal of the Mechanics and Physics of Solids, 2001. **49**(3): p. 589-607.

46. Adams, B.L., *Use of microstructural statistics in predicting polycrystalline material properties.* Metallurgical and materials Transactions, 1999. **30A**: p. 969.

47. Adams, B.L. and T. Olson, *The mesostructure--properties linkage in polycrystals.* Progress in Materials Science, 1998. **43**(1): p. 1-87.

48. Beran, M.J., et al., *Bounding elastic constants of an orthotropic polycrystal using measurements of the microstructure.* Journal of the Mechanics and Physics of Solids, 1996. **44**(9): p. 1543-1563.

49. Gupta, A., et al., *Structure-Property Linkages for Non-Metallic Inclusions/Steel Composite System using a Data Science Approach.* Acta Materialia, 2015. **91**: p. 239–254.

50. Kalidindi, S.R., *Data science and cyberinfrastructure: critical enablers for accelerated development of hierarchical materials.* International Materials Reviews, 2015. **60**(3): p. 150-168.

51. Niezgoda, S.R., A.K. Kanjarla, and S.R. Kalidindi, *Novel microstructure quantification framework for databasing, visualization, and analysis of microstructure data.* Integrating Materials and Manufacturing Innovation, 2013. **2:3**.

52. Wargo, E.A., et al., *Selection of Representative Volume Elements for Pore-Scale Analysis of Transport in Fuel Cell Materials.* Journal of Power Sources, 2012. **197**: p. 168-179.

53. Niezgoda, S.R., Y.C. Yabansu, and S.R. Kalidindi, *Understanding and Visualizing Microstructure and Microstructure Variance as a Stochastic Process.* Acta Materialia, 2011. **59**: p. 6387-6400.

54. Dong, X., et al., *Dependence of mechanical properties on crystal orientation of semi-crystalline polyethylene structures.* Polymer, 2014. **55**(16): p. 4248-4257.

55. Salem, A.A., et al., *Workflow for integrating mesoscale heterogeneities in materials structure with process simulation of titanium alloys.* Integrating Materials and Manufacturing Innovation, 2014. **3**(1): p. 1-22.

56. Qidwai, S.M., et al., *Estimating response of polycrystalline materials using sets of weighted statistical volume elements (WSVEs).* Acta Materialia, 2012. **60**: p. 5284–5299.

57. Hasanabadi, A., et al., *Efficient three-phase reconstruction of heterogeneous material from 2D cross-sections via phase-recovery algorithm.* Journal of microscopy, 2016. **264**(3): p. 384-393.

58. Turner, D.M. and S.R. Kalidindi, *Statistical construction of 3-D microstructures from 2-D exemplars collected on oblique sections.* Acta Materialia, 2016. **102**: p. 136-148.

59. Tschopp, M., G. Wilks, and J. Spowart, *Multi-scale characterization of orthotropic microstructures.* Modelling and Simulation in Materials Science and Engineering, 2008. **16**(6): p. 065009.

60. Hou, Y., et al., *A novel artificial dual-phase microstructure generator based on topology optimization.* Computational Materials Science, 2016. **123**: p. 188-200.

61. Groeber, M., et al., *A framework for automated analysis and simulation of 3D polycrystalline microstructures.* Acta Materialia, 2008. **56**(6): p. 1257-1273.

62. Choudhury, A., et al., *Quantification and classification of microstructures in ternary eutectic alloys using 2-point spatial correlations and principal component analyses.* Acta Materialia, 2016. **110**: p. 131-141.

63. Glicksman, M.E., *Mechanism of dendritic branching.* Metallurgical and Materials Transactions A, 2012. **43**(2): p. 391-404.

64. Grugel, R., *Secondary and tertiary dendrite arm spacing relationships in directionally solidified Al-Si alloys.* Journal of materials science, 1993. **28**(3): p. 677-683.

65. Kwon, Y., K. Thornton, and P.W. Voorhees, *Coarsening of bicontinuous structures via nonconserved and conserved dynamics.* Physical Review E, 2007. **75**(2): p. 021120.

66. Cai, B., et al., *4D synchrotron X-ray tomographic quantification of the transition from cellular to dendrite growth during directional solidification.* Acta Materialia, 2016. **117**: p. 160-169.

67. Gibbs, J., et al., *The three-dimensional morphology of growing dendrites.* Scientific reports, 2015. **5**: p. 11824.

68. Rowenhorst, D.J., et al., *3D Crystallographic and morphological analysis of coarse martensite: Combining EBSD and serial sectioning.* Scripta Materialia, 2006. **55**(1): p. 11-16.

69. Kammer, D. and P.W. Voorhees, *The morphological evolution of dendritic microstructures during coarsening.* Acta Materialia, 2006. **54**(6): p. 1549-1558.

70. Ziehmer, M., et al., *A principle curvatures analysis of the isothermal evolution of nanoporous gold: Quantifying the characteristic length-scales.* Acta Materialia, 2016. **120**: p. 24-31.

71. Gibbs, J.W. and P.W. Voorhees, *Segmentation of four-dimensional, X-ray computed tomography data.* Integrating Materials and Manufacturing Innovation, 2014. **3**(1): p. 6.

72. Allen, S.M. and J.W. Cahn, *A microscopic theory for antiphase boundary motion and its application to antiphase domain coarsening.* Acta Metallurgica, 1979. **27**(6): p. 1085-1095.

73. Agrawal, K.V., et al., *Oriented MFI Membranes by Gel-Less Secondary Growth of Sub-100 nm MFI-Nanosheet Seed Layers.* Advanced Materials, 2015. **27**(21): p. 3243-3249.