Microstructure Taxonomy based on Spatial Correlations: Application to Microstructure Coarsening

Tony Fast^{a,*}, Olga Wodo^b, Baskar Ganapathysubramanian^c, Surya R. Kalidindi^a

^aGeorge W. Woodruff School of Mechanical Engineering, Georgia Institute of Technology, 801 Ferst Drive, 30332, Atlanta, GA, USA

^bDepartment of Materials Design and Innovation, Department of Mechanical and Aerospace Engineering, SUNY at Buffalo, 318 Jarvis Hall, 14260, Buffalo, NY, USA

^cDepartment of Mechanical Engineering, Iowa State University, 2025 Black Engineering, 50011, Ames, IA, USA

Abstract

To build materials knowledge, rigorous description of the material structure and associated tools to explore and exploit information encoded in the structure are needed. These enable recognition, categorization and identification of different classes of microstructure and ultimately enable to link structure with properties of materials. Particular interest lies in the protocols capable of mining the essential information in large microstructure datasets and building robust knowledge systems that can be easily accessed, searched, and shared by the broader materials community.

In this paper, we develop a protocol based on automated tools to classify microstructure taxonomies in the context of coarsening behavior which is important for long term stability of materials. Our new concepts for enhanced description of the local microstructure state provide flexibility of description. The mathematical description of microstructure that capture crucial attributes of the material, although central to building materials knowledge, is still elusive. The new description captures important higher order spatial information, but at the same time, allows down sampling if less information is needed. We showcase the classification protocol by studying coarsening of binary polymer blends and classifying steady state structures. We study several microstructure descriptions by changing the microstructure local state order and discretization and critically evaluate their efficacy. Our analysis revealed the superior properties of microstructure representation is based on the first order-gradient of the atomic fraction.

Keywords: microstructure representation, data analytics, dimensionality reduction, microstructure, spinodal decomposition, taxonomy

1. Introduction

The complex hierarchical internal spatial organization of the material, often referred simply as the material microstructure, is known to play a dominant role in controlling the performance of engineering components. Various processing options offer a means of altering the material microstructure and thereby modulate the performance characteristics of the material. The microstructure space can be viewed as the common base where the structure-property-processing linkages can be expressed, curated and communicated effectively [1]. Consequently, a rigorous description of the material structure is essential to the ongoing efforts aimed at building the materials knowledge and at facilitating accelerated design of new advanced materials and more efficient manufacturing. However, the mathematical description of microstructure that capture crucial attributes of the material and facilitate linkage between process and properties is still elusive.

*Corresponding author

Email addresses: anthony.fast@me.gatech.edu (Tony Fast), olgawodo@buffalo.edu (Olga Wodo), baskarg@iastate.edu (Baskar Ganapathysubramanian), surya.kalidindi@me.gatech.edu (Surya R. Kalidindi)

The most commonly used tools to describe microstructure are based on the concept of n -point spatial correlations [1, 2, 3, 4, 5, 6, 7]. The n -point spatial correlations provide a detailed set of measures that are naturally organized by increasing amounts of microstructure information, with increasing n . At the same time, however, for most material systems of interest, the set of n -point statistics is an extremely large, even for $n = 2$. In such cases, rigorous analyses towards establishing structure-processing- property linkages can be enabled with the aid of data-mining analytic tools. For example, techniques such as principal component analysis (PCA) can be used to obtain objective low-dimensional representations of the 2-point statistics [1, 2, 8, 9, 10, 11, 12, 13, 14]. With reduced-order representation of the microstructure, it is possible to approach the classification of an ensemble of microstructures and the data-driven formulation of microstructure-property linkages [1], that otherwise would be very challenging, if at all feasible. In such cases, providing the microstructure description is only a preliminary step for building and mining large collections of structure-property-processing relationships. Therefore, morphology quantification along with the corresponding data mining utilities should be seen as a synergistic ensemble of tools centered around microstructure. Several such tools already exist and emerge as a nucleus of the new field – microstructure informatics [15].

Microstructure informatics focuses on development of novel algorithms and computationally efficient protocols that are capable of mining the information in large microstructure datasets and building robust knowledge systems that can be easily accessed, searched, and shared by the broader materials community. The central concept of microstructure informatics is realized by the sequential execution of three systems: (1) Segmentation Systems that convert data generated from physical models into suitable formats, (2) Analytic Systems that provide a host of effective, spatial statistics, and (3) Data-mining Systems that produce reduced dimensional forms of large datasets. Each system is independent and perform very specific tasks.

In this paper, we develop new modules and enhance preexisting modules in basic systems of microstructure informatics. In particular, we develop a workflow to build an automated tool to classify microstructure taxonomies in the context of coarsening behavior that is important for long term stability of materials. We report the following significant advances: (i) We describe novel concepts for enhanced description of the local microstructure state amenable to both discrete and continuous representation. The new descriptions captures important higher-order spatial information of the microstructure (Segmentation System). This extension segues to the first application of n -pt statistics (Statistical System) with $n > 2$ in the microstructure classification. (ii) We introduce novel modular workflows and visualization tools that facilitate critical evaluation of the efficacy of various descriptions of the local state. In particular, the visualization tools allow to evaluate the ability of descriptors to succinctly capture the salient features of 4-D (three dimensions in space and one dimension in time) microstructure datasets. (iii) We demonstrate the tremendous flexibility and advantages of these novel data science concepts by applying them to a case study involving coarsening of binary polymer blends.

This work is motivated by a technologically relevant and physically important microstructure evolution process for a variety of alloy systems as well as polymer blends spinodal decomposition. We are interested in understanding the dynamics of long-term behavior of binary systems. This has particular implications in the stability of coarsening microstructures, especially for photovoltaic applications [16]. To understand this, we performed high throughput computational studies using the Cahn Hilliard equation that describes phase separation and coarsening of various systems [17]. These equations are computationally intensive to solve as well as produce large, complex datasets [18]. The primary motivation of this work is to comprehensively analyze the datasets by data dimensionality reduction for efficient visual representation, and utilize the reduced representation to enable faster prediction of long-term behavior.

2. Segmentation Systems

Segmentation Systems convert input data – generated from physical models – into an alternative format that is easier to handle, while still retaining sufficient information to perform quantitative data analysis. This focuses on selected aspects of the microstructure but often neglects certain level of details. Designing a converter that offers flexibility to choose the granularity of details accounted for is a crucial feature for a good

Segmentation System. In particular, ease of extending to higher orders of detail is of critical importance. In this work, we define the discretized representation of microstructure that accounts for higher order aspects of the microstructure description. This representation allows down sampling if less information is needed.

A discretized representation of microstructure offers many advantages in fast computation of microstructure metrics [3, 19]. It has been exploited in automated identification of salient microstructure features in large datasets [4], extraction of representative volume elements from an ensemble of datasets [5], reconstructions of microstructures from measured statistics [6, 7], building of real-time searchable microstructure databases [15], and mining of high fidelity multi-scale structure-performance evolution linkages from physics-based models [20, 21, 22, 23, 24, 25, 26]. In this section, we first detail the microstructure function to subsequently motivate novel higher order description and provide details of the enhanced microstructure function.

2.1. Microstructure Function

Any microstructure dataset collected from a series of experiments can be interpreted as an outcome of a stochastic process. This notion is foundational to the stochastic description of microstructure introduced in recent work [1, 2, 8]. In this description, the microstructure function is denoted as $m(\mathbf{x}, \mathbf{h})$, and represents the probability density of finding local state $\mathbf{h}$ in the immediate neighborhood of the spatial location $\mathbf{x}$. The local state is often described by a set of parameters that may be organized as a k -dimensional vector, $\mathbf{h} = \{h_1, h_2, \dots, h_k\}$. For example, in an annealed multi-phase metallic alloy, the complete description of the local state may include the thermodynamic phase identifier, the elemental composition of the phase, and the crystal lattice orientation, measures of various defect densities, among other descriptors. In the notation presented above, the space of all possible local states would be a k -dimensional space called the local state space, and is denoted as $\mathbf{H}$.

A discretized version of the microstructure function can be extracted from $m(\mathbf{x}, \mathbf{h})$ by binning (typically a uniform tessellation) the spatial domain as well as the local state space [1]. Let $s = 1, 2, \dots, S$ and $h = 1, 2, \dots, H$ enumerate the individual bins in spatial domain and the local state space, respectively. With these conventions introduced, the discretized microstructure function is denoted as m_s^h and represents the total volume fraction of all local states from bin h in the spatial bin s . With this definition, the microstructure function is subject to the following constraints:

$$\sum_{h=1}^H m_s^h = 1, \quad 0 \leq m_s^h \leq 1 \quad (1)$$

Note that the framework for the description of the microstructure presented above is fairly general as it is not limited to any specific length or time scales. In particular, it can accommodate any combination of material features by simply increasing the dimensionality of the local state space. However, it introduces an inherent approximation in the discretization process. The fidelity of the approximation depends on the details of the discretization. In general, higher values of S and H improve the accuracy of the discretization. For large spatial domains and local state spaces, however, the storage demands on the discretized microstructure function can increase rapidly. Therefore, one has to be extremely judicious in selecting the list of the parameters of interest in identifying the local state, especially when seeking a viable, but sufficiently accurate, description of the microstructure function.

2.2. Higher-order Microstructure Description

In most prior work, the material system was treated as a multi-phase composite with a small number of distinct phases. In other words, the local state space was inherently discrete and one-dimensional. However, in recent work [24] when dealing with extraction of reliable structure-performance linkages in high-contrast composite systems, we found that local state space based solely on data about local phase is not sufficient. Therefore, we have introduced the notion of a higher-order microstructure function. In addition to the local phase, the function also captures the details of the local conformations (i.e. the local topology). As the local state in this function is based on the conformations, we call this function conformation-based microstructure

function (CbMF). For example, in a two-phase composite material microstructure in a uniform cuboidal tessellation of the spatial domain, the size of local distinct conformations based on the two local states and the six adjoining nearest neighboring cells (two in each direction) is $H = 2^7 (= 128)$ [24]. If these local conformations are used to define the local state, then the size of the local state space increases from $H = 2$ in the simple description of the microstructure function to $H = 128 = 2^7$ in the seventh-order description of the microstructure function. The order corresponds to seven bits of information used to define distinct conformation: the local state and states of six adjoining neighbors. The size of the local state space increases rapidly with either the addition of more states (for example, if 3 local states are allowed in each cell, the value of H increases from 3 to $3^7 = 2,187$) and/or the addition of more neighbors (with only the first six neighbors considered, if second nearest neighbors are considered H increases to $2^{15} = 32,768$). This combinatorial explosion in size is the main difficulty with this form of microstructure function. The second difficulty comes from the large discrete jumps in the size of the local state description (i.e. from 2 to 2^7 to 2^{15} in the examples described above) that leads to the significantly increased storage requirements.

In this work, we propose a new microstructure description to mitigate the above challenges. We pose the following requirements for the proposed description: (i) it must enable a hierarchical, flexible description of higher order features; (ii) while mitigating combinatorial explosion of the local state space size to ensure computationally viable approach.

The basic idea is to explicitly include information on the local microstructure gradients in the local state descriptions (instead of considering more neighbors when defining distinct local conformations). In essence, when for a given cell, the local state along with gradients are known, all local conformations from previous approach can be uniquely identified. In other words, information encoded in new local state space are equivalent to the ones from previous approach [24]. We call this function gradient-based microstructure function (GbMF). This is akin to replacing a function in a region by all its derivative a one point. The evaluation is done via the Taylor series expansion.

We are interested in accounting for higher-order microstructure features in the discrete representation m_s^h of the morphology $m(\mathbf{x}, \mathbf{h})$ while keeping H relatively low. For example, we want to account for the spatial gradient of the microstructure function denoted as $\nabla m(\mathbf{x}, \mathbf{h})$ (or $m_{,i}(\mathbf{x}, \mathbf{h})$ using the standard convention for spatial derivatives). Therefore, we introduce a discretized representation of $m_{,i}(\mathbf{x}, \mathbf{h})$ and we denote it as ${}_i\tilde{m}_s^h$ using the tessellation schemes described earlier [1]. We note that using the same discretization scheme, we can also introduce analogous discretized representation of higher-order gradients (or combinations of certain gradients). For example, the Hessian $\nabla(\nabla m(\mathbf{x}, \mathbf{h}))$ would be discretized as ${}_{ij}\tilde{m}_s^h$, and scalar field describing lapacian $\nabla^2 m(\mathbf{x}, \mathbf{h})$ would be discretized as ${}_{ii}\tilde{m}_s^h$.

Flexibility to express various higher-order gradient is a very important advantage of the proposed approach. Another important feature is the guarantee that the extended discrete representation is bounded. Since the values of m_s^h are bounded (see Eq. 1, also compare with CbMF), the use of a uniform tessellation in the spatial domain will ensure that the discretized representations of the gradients are also bounded.

Finally, we conceptually combine discrete values of microstructure m_s , and its derivatives ${}_i m_s$ and ${}_{ij} m_s$ to define the enriched local state of the microstructure function. We denote the enriched discretized representation of microstructure as $\tilde{m}_s^{\tilde{h}}$. The local state index $\tilde{h} = \{1, 2, \dots, \tilde{H}\}$ now enumerates the distinct combinations of discrete values of the microstructure function and its selected derivatives. We next consider two ways to construct $\tilde{H}$: by Multiplicative Augmentation and by Additive Augmentation.

In Multiplicative Augmentation, the microstructure and its gradient (i.e. $\tilde{h} = (h, {}_i h)$, where $h \in H$ and ${}_i h \in {}_i H$) represents the local states as two independent variables. The enriched local state description is the product of two respective spaces (i.e. $\tilde{H} = H \otimes {}_i H$). Note that the constraints of Eq. 1 continue to apply to this higher-order microstructure function as:

$$\sum_{\tilde{h}=1}^{\tilde{H}} \tilde{m}_s^{\tilde{h}} = 1, \quad 0 \leq \tilde{m}_s^{\tilde{h}} \leq 1 \quad (2)$$

In Additive Augmentation, the enriched local state description is simply the sum of two respective spaces, i.e. $\tilde{H} = H + {}_i H$. Note that the constraint of Eq. 1 needs to be applied separately to each variable used to

describe the local state:

$$\sum_{h=1}^H m_s^h = 1, \quad \sum_{i=1}^H m_s^{i^h} = 1, \quad 0 \leq (m_s^h, m_s^{i^h}) \leq 1 \quad (3)$$

Table 1: Equivalence between states in conformation (CbMF) and gradient based microstructure function (GbMF) for 1D two-state composite material when only second order conformations are considered (left) and when third order conformations are considered (right).

CbMF	GbMF		CbMF		GbMF	
	${}_1\bar{m}_s$	${}_{11}\bar{m}_s$	${}_1\bar{m}_s$	${}_{11}\bar{m}_s$	${}_1\bar{m}_s$	${}_{11}\bar{m}_s$
0 0	0	0	0 0 0	0	0	0
0 1	1	1	0 0 1	1	1	1
1 0	-1	-2	0 1 0	-1	-2	-2
1 1	0	1	0 1 1	0	1	1
			1 0 0	0	0	0
			1 0 1	1	1	1
			1 1 0	-1	-2	-2
			1 1 1	0	1	1

The main concepts introduced above are illustrated through a simple 1-D example of two-state composite material with 3 adjoining voxels. Since there are only two distinct local states, the microstructure function needs to be defined only for one local state (0 or 1). For convenience, we therefore drop the superscript h (as it has only one value in this simple problem), and simply represent the discretized microstructure function as m_s . In Table 1, we compare previous microstructure function with the one proposed in this work. In particular, Table 1 lists all possible conformations used in the previous approach. The order of the description is defined based on the total number of cells considered, i.e. second order description for 2 cells (left), third order for 3 cells (right). When a second-order description of the microstructure is considered, the conformation is defined based on the spatial cell of interest and the cell adjoining it on the right. Consequently there are four distinct conformations that are shown in the left table. For the third-order description (now including both adjoining cells) four additional conformations are included, as shown in right table. The total number of conformations corresponds to $2^2 = 4$ and $2^3 = 8$, respectively, and the size of the state suffers the combinatorial explosion. Table 1 list corresponding values of gradients with the proposed approach. The discretized representation of the first and second microstructure gradients are denoted as ${}_1\bar{m}_s$ and ${}_{11}\bar{m}_s$ with values listed in Table 1 (right). Gradient of microstructure ${}_1\bar{m}_s$ can take three distinct values: $\{0, 1, -1\}$, while laplacian ${}_{11}\bar{m}_s$ can take five distinct values $\{0, 1, -1, 2, -2\}$. For convenience, the voxel lengths are assumed to be unity in gradient computations. Size of corresponding state space in GbMF is 3 and 5, respectively and no longer suffers due to the combinatorial explosion. Note that the four second order distinct conformations from previous approach are equivalently described by m_s and ${}_1\bar{m}_s$. This is because ${}_1\bar{m}_s$ along with m_s uniquely identifies each of the four distinct second-order conformations. In a similar manner, one can see that all eight distinct third-order conformations can be uniquely mapped to the values of m_s , ${}_1\bar{m}_s$ and ${}_{11}\bar{m}_s$.

This example serves as simple illustration of the proposed microstructure function, however it does not emphasize the significant gains that this function delivers. We compare the size of local states between previous representation (CbMF) and the one proposed in this work (GbMF, both for Multiplication and Additive Augmentation) in Table 2. We make the comparison for local states with more discretized phases (2, 3, 10 phases) or more neighbors considered (as a function of spatial extent: 1D, 2D and 3D). It is interesting to notice that when small number of phases/bins is considered, the new approach does not show significant gains, especially for Multiplicative Augmentation. However, when 10 bins or phases need to be included, the gain of the new approach is enormous. This is especially clear for 3D when 7 cells are considered, where total number of states reduces from 10^7 to 63,270. It is also important to note that when discretization of microstructure is relatively fine (the absolute value of the gradient between neighboring cells is smaller than 2), the size of local state is reduced even further. In these cases, the computational gain is even higher, as 10 bin/phase structure can be defined with state space size comparable to the 3-phase case (for GbMF-MM total size reduces from 10^7 to 945, which is 4 orders of magnitude!). Small gradients

between neighbors are common for microstructure with fine resolution, where interface between phases is resolved with a few grid points. Better resolution of the sample (inherent to the continuous local states) reduces the size of the local state in the discrete microstructure function.

Table 2: Comparison between total number of states in configuration (CbMF) and gradient based microstructure function (GbMF).

	CbMF: H^S	
1D	2nd order (S=2)	3rd order (S=3)
2 phase	$2^2 = 4$	$2^3 = 8$
3 phase	$3^2 = 9$	$3^3 = 27$
10 phase	$10^2 = 100$	$10^3 = 1,000$
2D	3rd order (S=3)	5th order (S=5)
2 phase	$2^3 = 8$	$2^5 = 32$
3 phase	$3^3 = 27$	$3^5 = 243$
10 phase	$10^3 = 1,000$	$10^5 = 100,000$
3D	4rd order (S=4)	7th order (S=7)
2 phase	$2^4 = 16$	$2^7 = 128$
3 phase	$3^4 = 81$	$3^7 = 2,187$
10 phase	$10^4 = 10,000$	$10^7 = 10,000,000$

Note: H= 2, 3, 10 for 2 phase, 3 phase and 10 phase case, respectively.

	GbMF			
	1st order gradient		1st and 2nd order gradients	
	AM ($H + {}_iH$)	MM($H \otimes {}_iH$)	AM ($H + {}_iH + {}_{ii}H$)	MM($H \otimes {}_iH \otimes {}_{ii}H$)
1D				
2 phase	$2 + 3 = 5$	$2 \cdot 3 = 6$	$2 + (3 + 5) = 10$	$2 \cdot 3 \cdot 5 = 30$
3 phase	$3 + 5 = 8$	$3 \cdot 5 = 15$	$3 + (5 + 7) = 15$	$3 \cdot 5 \cdot 7 = 105$
10 phase	$10 + 19 = 29$	$10 \cdot 19 = 190$	$10 + (19 + 37) = 66$	$10 \cdot 19 \cdot 37 = 7,030$
2D				
2 phase	8	12	18	120
3 phase	13	30	27	420
10 phase	48	380	122	28,120
3D				
2 phase	11	18	26	270
3 phase	18	45	39	945
10 phase	67	570	178	63,270

Note: number of 1st and 2nd gradients is dimension dependent:
 1D: 1st (${}_1H$) and 2nd gradient spaces (${}_{11}H$)
 2D: 1st (${}_1H, {}_2H$) and 2nd gradient - Hessian (${}_{11}H, {}_{22}H, {}_{12}H$)
 3D: 1st (${}_1H, {}_2H, {}_3H$) and 2nd gradient - Hessian (${}_{11}H, {}_{22}H, {}_{33}H, {}_{12}H, {}_{13}H, {}_{23}H$)
 Size of microstructure state (H), 1st (${}_iH$) and 2nd gradient spaces (${}_{ij}H$) are
 2 phase: 2 $\{0, 1\}$, 3 $\{-1, 0, 1\}$, 5 $\{-2, -1, 0, 1, 2\}$
 3 phase: 3 $\{0, 1, 2\}$, 5 $\{-2, -1, 0, 1, 2\}$, 7 $\{-3, -2, -1, 0, 1, 2, 3\}$
 10 phase: 10 $\{0, 1, \dots, 9\}$, 19 $\{-9, -8, \dots, 0, 1, \dots, 9\}$, 37 $\{-18, -17, \dots, 0, 1, \dots, 17, 18\}$
 General formula for total number of states
 AM: $\sum_{o=0}^m H_o$, MM: $\prod_{o=0}^m H_o$
 where o is the gradient order with zero-order being the microstructure state

3. Analytic Systems

Broadly speaking, Analytic Systems quantify the statistics of the spatial distribution of the local states in the internal structure. These typically serve as a basis to establish process-structure-property relationship [3, 19, 27].

There are number of different measures of the spatial correlations in the microstructure, e.g. lineal-path functions [28, 29, 4] and chord-length distributions [30, 31], n -point statistics [2, 3, 19, 5, 6, 32, 33]. However, only the n -point spatial correlations provide an extensible and flexible set of measures that are naturally organized by increasing amounts of microstructure information. For example, 1-point statistics reflects the probability density associated with finding a specific local state of interest at any randomly selected single point in the microstructure. The 2-point statistics capture the probability of finding the local states of interest at the tail and head of randomly tossed vectors of specified length and orientation. Higher-order correlations (3-point and higher) are defined in an analogous manner and additionally enrich the morphology

quantification. For this reason, we utilize the concepts of the n -point spatial correlations (simply referred to as n -point statistics) for the rigorous quantification of the material microstructure.

3.1. Spatial Correlations

In the framework introduced earlier for the discretized representation of microstructure, the 2-point correlations of interest can be expressed as

$$f_r^{\bar{h}\bar{h}'} = \frac{1}{S} \sum_{s=1}^S \tilde{m}_s^{\bar{h}} \tilde{m}_{s+r}^{\bar{h}'} \quad (4)$$

where r enumerates the discretization of the space of vectors using the same uniform binning scheme that was used for binning the spatial domain [1]. There is equivalence between the 2-point correlations of the above augmented microstructure functions introduced in this work and the higher-order correlations (3-point and higher) obtained using the original microstructure function, m_s^h .

The central challenge in dealing with a rigorous quantitative description of the spatial statistics of the microstructure is that they constitute an extremely large set. However, there are redundancies and interdependencies between the higher-order spatial correlations. For example, Niezgoda et al. [33] recently showed that only $(H - 1)$ of the 2-point correlations are actually independent by exploiting known properties of the discrete Fourier transform (DFT) representation of the correlations. Furthermore, the higher-order microstructure description introduced in the previous section leads to additional interdependencies among the corresponding 2-point statistics. The physical origin and nature of these additional interdependencies, along with suitable mathematical protocols for identifying them, have been discussed in detail in recent work from our group [24].

3.2. Objective Reduced-Order Representations: Principal Component Analysis

It was recently demonstrated that techniques such as principal component analysis (PCA) [34, 35, 36, 37], can be used to obtain objective low dimensional representations of the 2-point statistics [2, 15]. Here, we employ the same PCA dimensionality reduction techniques on the augmented spatial correlations introduced in this work.

Let f_r denote the truncated set of independent 2-point statistics of interest. Let R denote the dimensionality of f_r , i.e. $r = 1, 2, \dots, R$. Let us assume that we are investigating an ensemble of microstructures, whose elements are enumerated $i = 1, 2, \dots, I$. It is generally expected that $I \leq R$. In such situations, PCA identifies a maximum of $(I - 1)$ orthogonal directions in the R -dimensional space that are arranged by decreasing levels of variance. In other words, the first direction exhibits the highest variance in the given ensemble of microstructures, and the second direction exhibits the second highest variance while being orthogonal to the first direction, and so on. Let ϕ_{ri} denotes the components of the orthogonal set of directions identified by PCA. Note that these are dependent on the specific microstructures in the selected ensemble. If $I > R$, the maximum of orthogonal directions identified by PCA will be R .

Mathematically, the PCA representation of any member of the selected ensemble (of microstructures), labeled by superscript (k) , can be expressed as:

$$f_r^{(k)} = \sum_{i=1}^{\min((I-1), R)} \alpha_i^{(k)} \phi_{ri} + \bar{f}_r \quad (5)$$

where $\bar{f}_r$ is simply the average 2-point statistics for the entire ensemble, and $\alpha_i^{(k)}$ (referred as PCA weights) provide an objective representation of the specific microstructure identified by label (k) in the orthogonal reference frame defined by ϕ_{ri} .

An important output from the PCA is the significance of each principal component, b_i , obtained in the eigenvalue decomposition performed as a part of the PCA. The values of b_i provide important measures of the inherent variance among the members of the ensemble of microstructures [2]. More importantly, by retaining only the components associated with the highest (or the most significant) eigenvalues, it is often

possible to obtain an objective reduced- order representation of the microstructure with only a handful of parameters [2, 15]. Mathematically, this reduced-order representation can be expressed as

$$f_r^{(k)} \approx \sum_{i=1}^{R^*} \alpha_i^{(k)} \phi_{ri} + \bar{f}_r \quad (6)$$

where $R^* \ll \min((I - 1), R)$. Selection of R^* will depend on the specific application. Note also that the concepts described above can be easily extended to include higher-order statistics of the microstructure (e.g., 3-point spatial correlations).

4. Data-mining Systems

Data-mining System is a set of tools that extracts the embedded high value information in the given dataset (here an ensemble of material structures). The major objective of this work is to develop a framework to classify various types of microstructure into distinctive groups and subsequently provide a microstructure taxonomy. Classification and taxonomy are closely related terms [38, 39]: classification is a way to organize species into classes, while taxonomy is a name for systems that classify different species. Most generally, taxonomy is independent of the type of objects. Taxonomy typically has been associated with biology, where it has allowed a rigorous expression of the relationship between different species, by encoding their characteristics. Recently, taxonomy has found utility in materials science, e.g. to build the materials tree of processing [40]. Classification and taxonomy have been shown to be natural tools to cluster and classify individuals into clearly distinguishable classes. We build upon this idea to develop a classification strategy to organize various classes of microstructures and ultimately build the taxonomy. We use 2-point correlations of microstructure function as a characteristics - the basis of the classification. We build the classification based on statistical descriptors because visually different microstructures can exhibit identical morphology descriptors.

Microstructure classification is an important step towards a quantitative link between process and structure. This is an important quest in materials science as different process protocols may lead to statistically similar structure and thus properties. Identifying such ranges of process and system variables opens new possibilities to find more economical manufacturing routes to desired properties. For example, our previous work on processing titanium alloys showed that different processing treatments lead to distinguishable structures [15]. In that work [15], interclass and intraclass variations between structures have been naturally revealed in the PCA-based reduced dimensional embedding of the microstructure spatial statistics. In the current work, we use this observation and we extend it to automatically classify microstructures, with the ultimate goal of creating a microstructure taxonomy.

In this work, we use clustering as a classification method. Cluster analysis is a class of automated data-mining tools that group data points, or objects, into clusters. Numerous heuristics exist for performing cluster analysis (e.g. Support Vector Machines [41, 42], Gaussian Cluster Models [43, 44], and k-means Clustering [45, 46]). We employ k-means clustering here.

4.1. Automated Data-mining: K-means Clustering

We employ k-means clustering which partitions each observation into clusters based on the nearest Euclidean mean of each cluster. [47] We perform clustering on the reduced PCA embedding produced by the workflow for each Higher-Order Descriptor. For each of the microstructure descriptors, we assess the degree of clustering of microstructure points in the PCA space based on 2-point correlations. A k-means clustering algorithm is trained in Matlab to partition the PCA embedding in each of the considered classes. Once the k-means algorithm is trained, the same PCA embedding is passed through the trained k-means clustering algorithm to reclassify each point. The training data provides the actual classification while the reclassification provides the predicted classification. In the best case, the classification for the actual and predicted classes will be the same. This will not always be true because of either the topology of the PCA embedding or the inner workings of the k-means clustering algorithm.

4.2. Sensitivity and Specificity Metrics in K-means Clustering

Cluster analysis allow clustered data to be mined and the degree of clustering to be qualified numerically. Sensitivity and specificity are common metrics in performance classification tests that compare the classes of the training data and the predicted classifications. [48, 49] Sensitivity measures the proportion of correct classifications (true positive rate), while specificity measures ability to exclude a condition correctly (true negative rate). An ideal classification technique will have sensitivity and specificity measurements of unity. Both metrics are used in our analysis.

5. Classification of Microstructure Taxonomies: Application to Spinodal Decomposition and Grain Coarsening

We studied microstructure evolution during spinodal decomposition. We start with a well mixed two component system that undergoes spinodal decomposition. Depending on the specifics of the blend and processing conditions (for example, annealing time [50, 51], solvent type [52, 16]) different morphologies are typically formed. There is a rich and complex collection of interacting phenomena that direct this morphology evolution - the mixture initially separates into phases very rapidly (phase-separation), followed by slow, and sporadic coarsening events. Remarkably, the complex dynamics is described by the governing equation called Cahn-Hilliard equation [17] that encompasses the kinetics of both phase separation and coarsening:

$$\frac{\partial \phi}{\partial t} = \nabla \cdot (M \nabla (\frac{\partial f}{\partial \phi} - C n^2 \nabla^2 \phi)) \quad (7)$$

where $\phi(x, t)$ is the atomic fraction, which evolves in time and space, M is the mobility and f is the free energy. By solving this equation, we predict the spatio-temporal morphology evolution. The Cahn-Hilliard equation is a fourth order nonlinear partial differential equation, where the inherent multiple temporal and spatial scales make this equation difficult to solve. Solving this equation to obtain long term behavior of system of interest is a very difficult computational problem [18]. Hence, screening large phase space of process and system variables becomes intractable. At the same time, long term behavior is critical for analysis of extended stability of heterogenous systems [53]. To mitigate this modeling challenge, we have developed a computationally efficient, massively parallel framework based on the finite element method [18]. To address multiple temporal scales involved, we utilize an adaptive time stepping scheme (Explicit Single Diagonal Implicit Runge-Kutta method of order 3 (ESDIRK3)). In particular, adaptive time stepping strategy allows to efficiently and accurately capture fast interface creation between two phases by lowering the time-step and subsequently utilizing larger time- steps for the very slow coarsening events. With this solution, for the first time we were able to obtain all five hypothesized steady state solutions of morphology in binary blends. This is however still a computationally intensive process - even with such highly sophisticated computational framework. Therefore, we are interested in leveraging data mining tools to correlate the intermediate morphology with the final morphology, or rather class of morphologies. In this way, screening can be optimized by introducing a two step process. In the first step, we perform detailed screening of the phase space of process and system variables over a short period of time to get the intermediate morphology. In the second step, we use classification tool to identify the class to which the final morphology belongs. This screening strategy serves as a motivation to develop the the data mining tools that we discuss in this paper.

5.1. Spinodal Decomposition and Grain Coarsening Simulation

We use the full scale physics-based framework to generate large data sets of morphologies. We uniformly sample 25 different initial concentrations ranging from $\phi_0 \in (0.5, 0.76)$. We model morphology evolution starting from random initial field around ϕ_0 with a maximum fluctuation of 0.05. For each realization of morphology evolution we store three characteristic morphologies: initial concentration field, concentration with boundaries created and the final steady-state. For each initial concentration, we generate a series of snapshots of the morphology evolution consisting of 30 images. We use a unit computational domain $L = 1$, with $M = 1$. In 3D, we assume Cahn number $C n^2 = 0.015$ and discretize the domain into 20 elements along

each direction. In 2D we assume $Cn^2 = 0.001$ with 60 elements in each direction. We utilize a Flory-Huggins free energy as a description of free energy of mixing $f = \phi \ln(\phi) + (1 - \phi) \ln(1 - \phi) + 3\phi(1 - \phi)$. Figure 1 (a) depicts morphology evolution typical for spinodal decomposition, from early phase separation to the final steady state solution.

Figure 1: **Morphologies representative for spinodal decomposition** (a) Typical morphology evolution: from early phase separation to the final steady state solution. (b) Isosurfaces of the each of the possible final topologies following the completion of the simulation..

There are 11 classes that demarcate each of the 11 topologies that can be realized (Figure 1 (b)). Figure 1 illustrates the genus of final topologies that are generated from the simulation. This figure shows the isosurfaces of the possible topologies.

5.2. Clustering in the Reduced Principal Component Decomposition Embedding

From the simulation described above, approximately 1050 final simulation microstructures were generated. These structures were then passed through a work flow that includes the aforementioned microstructure informatics tools in the following order:

1 Segmentation System Module:

Higher-Order Microstructure Descriptor of Continuous Materials States

2 Statistics System Module:

2-pt statistics of a Higher-Order Microstructure Descriptor from the Segmentation System

3 Data-mining System:

Reduced Order representation of 2-pt Statistics via PCA

4 Automation of Classification Taxonomy:

K-means clustering of the PCA embedding

i Quantification of Clustering

ii Qualification of Microstructure Descriptors for Spinodal Decomposition and Grain Coarsening Simulation

Each final microstructure generated from the simulation is passed through the modules 1-3 in the workflow above. The clustering is performed on the reduced PCA embedding produced by the workflow for each Higher-Order Descriptor (Figures 2 and 3). The final product of the microstructure informatics workflow is illustrated in Figure 2. In our analysis, we study various microstructure functions and effect of number of bins used in given description. Figure 2 shows the PCA embedding of the final structures, where each datapoint corresponds to one microstructure with color denoting the class of the final structure. We analyzed three different microstructure descriptors: the atomic fraction (left column), the first gradient of the atomic fraction (middle column), and the second gradient of the atomic fraction (right column). For each of the

Figure 2: **PCA: effect of total number of bins** Principal Component Analysis of the two-point correlation functions for the atomic fraction (first column), first gradient of atomic fraction (second column), and the second gradient of atomic fraction (third column) with 2 bins (top row) and 10 bins (bottom row). The color of each point corresponds to the different classes of structures.

Figure 3: **PCA: comparison between additive and multiplicative augmentation** Principal Component Analysis of the two-point correlation functions for additive and multiplicative local state spaces that combine all permutations of atomic fraction, first gradient of atomic fraction, and second gradient of atomic fraction. The color of each point corresponds to the different classes of structures.

descriptor we changed the discretization of their local state space. Figure 2 shows the PCA embedding for tessellations of 2 and 10 bins (top and bottom row).

Our analysis reveals that each microstructure descriptor and the binning scheme give significantly different PCA embedding. Some of the data plotted in Figure 2 shows hard clustering, with data points divided into

distinct clusters. The atomic fraction and its first gradient exhibit a distinct hard clustering behavior, both for 2 and 10 bins. In both cases, the spread of the points is more reduced for 10 bins when compared to 2 bins. Finally, the second gradient descriptor displays some degree of soft clustering wherein data points corresponding to different topologies overlap each other.

As discussed earlier, the higher order microstructure descriptors can be augmented either additively as a dual space (top row of Figure 3) or multiplicatively as a product space (bottom row of Figure 3). We study all permutation of: the atomic fraction, its first gradient, and its second gradient and results are plotted in Figure 3. For all studied permutations, 10 bins in the local state space were used. Similar to Figure 2, distinct cases of hard clustering emerges from PCA. Both augmentations of the atomic fraction and its second gradient (second column) along with the multiplicative combination of all of the descriptors (bottom right) demonstrate soft clustering behavior. Finally, the data shows that for this particular material system and informatics workflow devised, the first-order gradient is the most salient microstructure descriptor.

Figure 4: **Clustering: effect of total number of bins** Simultaneous visualization of the sensitivity and specificity of the k-means hard clustering of the PCA representations of unmixed atomic fraction, first gradient of atomic fraction, and second gradient of atomic fraction for binning schema of two, five, and ten.

5.3. K-means clustering in the PCA domain

In this paper, we not only explore a workflow to classify microstructures, but we also aim to identify the microstructure representations that are the most useful in this task. Figure 4 illustrates the multi-class sensitivity and specificity of the k-means clustering algorithm for all microstructure descriptions. Going counter clockwise from right to left, we show metrics for the atomic fraction, first gradient of atomic fraction, and the second gradient of atomic fraction, respectively. Results correspond to the clustering shown in the Figure 2. In each case, we study local spaces tessellated into two, five, and ten bins that are indicated by different markers shown in the legend. In each slice, the radial axis express the sensitivity and specificity of the clustering algorithm for given class. With such visualization, a microstructure descriptor with the most accurate classification is one with markers located on the left most edge of the slice. Figure 4 indicates the first-order gradient of the atomic fraction, middle slice, to be the most accurate with high recall rate in the k-means clustering algorithm. Interestingly, for the representation based only on the first order gradients, the clustering is insensitive to the number of bins. Consequently, the microstructure representation based on the first order gradients with two bins only offers the most optimal option in terms of computational complexity. Moreover, it also suggests that when the total number of bins is fixed (e.g. limited by the resolution of the experimental set up), microstructure function should be based on the first gradient rather than on the atomic fraction, as then the classification is insensitive to the total number of bins.

We perform similar analysis for the two augmentation schemes to study the soundness of aggregating various orders of information about the microstructure. In Figure 5, we illustrate the sensitivity and specificity

Figure 5: **Clustering: comparison between additive and multiplicative augmentation** Simultaneous visualization of the sensitivity and specificity of the k-means hard clustering of the PCA representations of the additive and multiplicative mixed atomic fraction, first gradient of atomic fraction, and second gradient of atomic fraction for binning schema of two, five, and ten. Going counter clockwise from right to left, we show both metrics for (i) the atomic fraction and first gradient of atomic fraction, (ii) atomic fraction and the second gradient of atomic fraction, (iii) first and second gradient and (iv) atomic fraction, first and second gradients augmented, respectively.

of k-means clustering algorithm for two augmentation schemes: additive and multiplicative augmentation. Going counter clock wide from right to left, we show both metrics for (i) the atomic fraction and first gradient of atomic fraction, (ii) atomic fraction and the second gradient of atomic fraction, (iii) first and second gradient and (iv) atomic fraction, first and second gradients augmented, respectively. The sensitivity and specificity plots provide quantification of the quality of the clustering of the Figure 3. Results indicate the additive augmentation based on atomic fraction and the second gradient of atomic fraction as the representation ensure good clustering. Regardless of total number of bins used, the sensitivity and specificity is the highest for this case. In general, additive augmentation demonstrates much better clustering metrics than multiplicative. Moreover, as the order of microstructure representation increases with increasing number of bins and higher order local states included, and presumably more information being encoded in the discrete representation of the microstructure, the worse metrics of clustering we observe.

Summary

In this paper, we developed new modules and enhanced preexisting modules in basic systems of microstructure informatics (segmentation system, analytic system, data-mining system). In particular, we built an automated tool to classify microstructure taxonomies shown in the context of coarsening behavior. We proposed generic higher-order microstructure functions that allowed us to hierarchically include local microstructure gradients in the local state description (segmentation system). Principle Component Analysis was used to generate reduced-order representation of the n -point statistics (Data-mining System). The k-means algorithm was harnessed to identify clusters in the reduced order representation. The flexibility and advantages of these novel data science concepts was shown in our application of coarsening of binary polymer blends. We performed a series of data experiments and identified two microstructure representations that deliver excellent basis for the clustering:

- when only one order of microstructure local state is allowed, the first order gradient ensured the best

clustering

- when enriched local state of the microstructure is possible (additive vs multiplicative augmentation), the additive augmentation based on the atomic fraction and second gradient ensured the best clustering.

As a general conclusion, in order to build the effective classification protocol, one needs to be careful of the level of information used to represent the microstructure.

Acknowledgment

TF and SK acknowledge support from AFOSR award FA9550-12-1-0458 for this work. BG and OW acknowledge support from awards NSF 1437636 and NSF 1149365 for this work.

References

References

- [1] B. L. Adams, X. C. Gao, S. R. Kalidindi, Finite approximations to the second-order properties closure in single phase polycrystals, *Acta Materialia* 53 (13) (2005) 3563–3577.
- [2] S. R. Niezgoda, Y. C. Yabansu, S. R. Kalidindi, Understanding and visualizing microstructure and microstructure variance as a stochastic process, *Acta Materialia* 59 (16) (2011) 6387–6400.
- [3] D. T. Fullwood, S. R. Niezgoda, B. L. Adams, S. R. Kalidindi, Microstructure sensitive design for performance optimization, *Progress in Materials Science* 55 (6) (2010) 477–562.
- [4] S. R. Niezgoda, S. R. Kalidindi, Applications of the phase-coded generalized hough transform to feature detection, analysis, and segmentation of digital microstructures, *Computers, Materials, & Continua* 14 (14) (2009) 79–97.
- [5] S. R. Niezgoda, D. M. Turner, D. T. Fullwood, S. R. Kalidindi, Optimized structure based representative volume element sets reflecting the ensemble-averaged 2-point statistics, *Acta Materialia* 58 (13) (2010) 4432–4445.
- [6] D. T. Fullwood, S. R. Niezgoda, S. R. Kalidindi, Microstructure reconstructions from 2-point statistics using phase-recovery algorithms, *Acta Materialia* 56 (5) (2008) 942–948.
- [7] D. Fullwood, S. Kalidindi, S. Niezgoda, T. Fast, N. Hampson, Gradient-based microstructure reconstructions from distributions using fast fourier transforms, *Materials Science and Engineering: A* 494 (1) (2008) 68–72.
- [8] B. Ganapathysubramanian, N. Zabaras, Modeling diffusion in random heterogeneous media: Data-driven models, stochastic collocation and the variational multiscale method, *Journal of Computational Physics* 226 (1) (2007) 326 – 353.
- [9] B. Ganapathysubramanian, N. Zabaras, A non-linear dimension reduction methodology for generating data-driven stochastic input models, *Journal of Computational Physics* 227 (13) (2008) 6612–6637.
- [10] A. Gupta, A. Cecen, S. Goyal, A. K. Singh, S. R. Kalidindi, Structure–property linkages using a data science approach: Application to a non-metallic inclusion/steel composite system, *Acta Materialia* 91 (2015) 239–254.
- [11] S. R. Kalidindi, Data science and cyberinfrastructure: critical enablers for accelerated development of hierarchical materials, *International Materials Reviews* 60 (3) (2015) 150–168.

- [12] A. Çeçen, T. Fast, E. Kumbur, S. Kalidindi, A data-driven approach to establishing microstructure–property relationships in porous transport layers of polymer electrolyte fuel cells, *Journal of Power Sources* 245 (2014) 144–153.
- [13] S. R. Niezgoda, A. K. Kanjarla, S. R. Kalidindi, Novel microstructure quantification framework for databasing, visualization, and analysis of microstructure data, *Integrating Materials and Manufacturing Innovation* 2 (1) (2013) 1–27.
- [14] S. Kalidindi, J. Gomberg, Z. Trautt, C. Becker, Application of data science tools to quantify and distinguish between structures and models in molecular dynamics datasets, *Nanotechnology* in press, invited article for a special issue.
- [15] S. R. Kalidindi, S. R. Niezgoda, A. A. Salem, Microstructure informatics using higher-order statistics and efficient data-mining protocols, *Jom* 63 (4) (2011) 34–41.
- [16] O. Wodo, B. Ganapathysubramanian, Modeling morphology evolution during solvent-based fabrication of organic solar cells, *Computational Materials Science* 55 (2012) 113–126.
- [17] J. W. Cahn, J. E. Hilliard, Free energy of a nonuniform system. i. interfacial free energy, *The Journal of Chemical Physics* 28 (2) (1958) 258–267.
- [18] O. Wodo, B. Ganapathysubramanian, Computationally efficient solution to the cahn–hilliard equation: Adaptive implicit time schemes, mesh sensitivity analysis and the 3d isoperimetric problem, *Journal of Computational Physics* 230 (15) (2011) 6037–6060.
- [19] S. Torquato, *Random heterogeneous materials: microstructure and macroscopic properties*, Vol. 16, Springer, 2002.
- [20] G. Landi, S. R. Kalidindi, Thermo-elastic localization relationships for multi-phase composites, *Computers, Materials, & Continua* 16 (3) (2010) 273–293.
- [21] S. R. Kalidindi, S. R. Niezgoda, G. Landi, S. Vachhani, T. Fast, A novel framework for building materials knowledge systems, *Computers, Materials, & Continua* 17 (2) (2010) 103–125.
- [22] G. Landi, S. R. Niezgoda, S. R. Kalidindi, Multi-scale modeling of elastic response of three-dimensional voxel-based microstructure datasets using novel dft-based knowledge systems, *Acta Materialia* 58 (7) (2010) 2716–2725.
- [23] T. Fast, S. R. Niezgoda, S. R. Kalidindi, A new framework for computationally efficient structure–structure evolution linkages to facilitate high-fidelity scale bridging in multi-scale materials models, *Acta Materialia* 59 (2) (2011) 699–707.
- [24] T. Fast, S. R. Kalidindi, Formulation and calibration of higher-order elastic localization relationships using the mks approach, *Acta Materialia* 59 (11) (2011) 4595–4605.
- [25] O. Wodo, S. Tirthapura, S. Chaudhary, B. Ganapathysubramanian, A graph-based formulation for computational characterization of bulk heterojunction morphology, *Organic Electronics* 13 (6) (2012) 1105 – 1113.
- [26] O. Wodo, S. Tirthapura, S. Chaudhary, B. Ganapathysubramanian, Computational characterization of bulk heterojunction nanomorphology, *Journal of Applied Physics* 112 (6).
- [27] G. W. Milton, *The theory of composites*, Vol. 6, Cambridge University Press, 2002.
- [28] B. Lu, S. Torquato, Lineal-path function for random heterogeneous materials, *Physical Review A* 45 (2) (1992) 922.

- [29] H. Singh, A. Gokhale, S. Lieberman, S. Tamirisakandala, Image based computations of lineal path probability distributions for microstructure representation, *Materials Science and Engineering: A* 474 (1) (2008) 104–111.
- [30] W. Gille, D. Enke, F. Janowski, Pore size distribution and chord length distribution of porous vycor glass (pvg), *Journal of Porous Materials* 9 (3) (2002) 221–230.
- [31] S. Torquato, B. Lu, Chord-length distribution function for two-phase random media, *Physical Review E* 47 (4) (1993) 2950.
- [32] W. F. Brown Jr, Solid mixture permittivities, *The Journal of Chemical Physics* 23 (8) (1955) 1514–1517.
- [33] S. Niezgoda, D. Fullwood, S. Kalidindi, Delineation of the space of 2-point correlations in a composite material system, *Acta Materialia* 56 (18) (2008) 5285–5292.
- [34] K. Person, On lines and planes of closest fit to system of points in space. *philosophical magazine*, 2, 559–572 (1901).
- [35] H. Hotelling, Analysis of a complex of statistical variables into principal components., *Journal of educational psychology* 24 (6) (1933) 417.
- [36] R. A. Johnson, D. W. Wichern, et al., *Applied multivariate statistical analysis*, Vol. 4, Prentice hall Englewood Cliffs, NJ, 1992.
- [37] I. Jolliffe, *Principal component analysis*, Wiley Online Library, 2002.
- [38] P. H. Sneath, R. R. Sokal, et al., *Numerical taxonomy. The principles and practice of numerical classification.*, 1973.
- [39] R. Van Rees, Clarity in the usage of the terms ontology, taxonomy and classification, *CIB REPORT* 284 (432) (2003) 1–8.
- [40] M. F. Ashby, H. Shercliff, D. Cebon, *Materials: engineering, science, processing and design*, Butterworth-Heinemann, 2013.
- [41] V. Vapnik, *The nature of statistical learning theory*, Springer Science & Business Media, 2013.
- [42] C. Cortes, V. Vapnik, Support-vector networks, *Machine learning* 20 (3) (1995) 273–297.
- [43] J. A. Hartigan, *Clustering algorithms*, John Wiley & Sons, Inc., 1975.
- [44] C. Fraley, A. E. Raftery, How many clusters? which clustering method? answers via model-based cluster analysis, *The computer journal* 41 (8) (1998) 578–588.
- [45] J. A. Hartigan, M. A. Wong, Algorithm as 136: A k-means clustering algorithm, *Applied statistics* (1979) 100–108.
- [46] J. MacQueen, et al., Some methods for classification and analysis of multivariate observations, in: *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, Vol. 1, Oakland, CA, USA., 1967, pp. 281–297.
- [47] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, A. Y. Wu, An efficient k-means clustering algorithm: Analysis and implementation, *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 24 (7) (2002) 881–892.
- [48] N. Japkowicz, M. Shah, *Evaluating learning algorithms: a classification perspective*, Cambridge University Press, 2011.

- [49] T. Fawcett, An introduction to {ROC} analysis, Pattern Recognition Letters 27 (8) (2006) 861 – 874, {ROC} Analysis in Pattern Recognition.
- [50] G. Li, V. Shrotriya, J. Huang, Y. Yao, T. Moriarty, K. Emery, Y. Yang, High-efficiency solution process-able polymer photovoltaic cells by self-organization of polymer blends, Nature materials 4 (11) (2005) 864–868.
- [51] J. Jo, S.-S. Kim, S.-I. Na, B.-K. Yu, D.-Y. Kim, Time-dependent morphology evolution by annealing processes on polymer: Fullerene blend solar cells, Advanced Functional Materials 19 (6) (2009) 866–874.
- [52] K. Kawano, J. Sakai, M. Yahiro, C. Adachi, Effect of solvent on fabrication of active layers in organic solar cells based on poly (3-hexylthiophene) and fullerene derivatives, Solar Energy Materials and Solar Cells 93 (4) (2009) 514–518.
- [53] M. Jorgensen, K. Norrman, F. C. Krebs, Stability/degradation of polymer solar cells, Solar Energy Materials and Solar Cells 92 (7) (2008) 686 – 714.

Microstructures

Spatial correlations

$$f_r^{\tilde{h}\tilde{h}'} = \frac{1}{S} \sum_{s=1}^S \tilde{m}_s^{\tilde{h}} \tilde{m}_{s+r}^{\tilde{h}'}$$

Classification