

IDETC2014-34570

A MACHINE LEARNING-BASED DESIGN REPRESENTATION METHOD FOR DESIGNING HETEROGENEOUS MICROSTRUCTURES

Hongyi Xu

Department of Mechanical Engineering,
Northwestern University
Evanston, IL, US

Alok Choudhary

Department of Electrical Engineering and
Computer Science, Northwestern University
Evanston, IL, US

Ruoqian Liu

Department of Electrical Engineering and
Computer Science, Northwestern University
Evanston, IL, US

Wei Chen*

Department of Mechanical Engineering,
Northwestern University
Evanston, IL, US

ABSTRACT

In designing microstructural materials systems, one of the key research questions is how to represent the microstructural design space quantitatively using a descriptor set that is sufficient yet small enough to be tractable. Existing approaches describe complex microstructures either using a small set of descriptors that lack sufficient level of details, or using generic high order microstructure functions of infinite dimensionality without explicit physical meanings. We propose a new machine learning-based method for identifying the key microstructure descriptors from vast candidates as potential microstructural design variables. With a large number of candidate microstructure descriptors collected from literature covering a wide range of microstructural material systems, a 4-step machine learning-based method is developed to eliminate redundant microstructure descriptors via image analyses, to identify key microstructure descriptors based on structure-property data, and to determine the microstructure design variables. The training criteria of the supervised learning process include both microstructure correlation functions and material properties. The proposed methodology effectively reduces the infinite dimension of the microstructure design space to a small set of descriptors without a significant information loss. The benefits are demonstrated by an example of polymer nanocomposites optimization. We compare designs using key microstructure descriptors versus using empirically-chosen microstructure descriptors to validate the proposed method.

Keywords: material design, machine learning, microstructure descriptors, informatics.

1. INTRODUCTION

Trial-and-error procedures are the traditional way of material design, which has been mostly guided by experiences and heuristic rules in materials classification, selection, and property predictions. Applying heuristic rules to existing materials databases for searching combinations of processing procedure or material constituents [1, 2] is time-consuming and resource intensive, however, microstructure information is often not considered in this process. Constituent-based design approach has relied on heuristic search to choose proper material compositions from materials databases [3, 4], but this approach no longer suffices in designing complex microstructural materials systems. To fully explore the potential of computational material design and accelerate the development of advanced materials, “microstructural-mediated design of materials” [5, 6] has gained more attention. With this new paradigm, materials are viewed as a complex structural system that has design degrees of freedom in choices of composition, phases, and microstructure morphologies, which can be optimized for achieving superior material properties. In particular, the morphology of microstructure (i.e., the spatial arrangements of local microstructural features) has a strong impact on the overall properties of a materials system. Taking polymer nanocomposites as an example, microstructure percolation determines the electrical conductivity, and the quantity of fillers’ surface area determines the damping properties [7]. Furthermore, heterogeneity in microstructure is the root cause of material randomness at multiple length scales.

There are two major categories of methods: correlation functions and physical descriptors, for quantifying the morphology and heterogeneity of microstructures (also known as “statistical characterization”). The microstructure

* Correspondence to: Dr. Wei Chen, Wilson-Cook Professor in Engineering Design and Professor of Mechanical Engineering
Email: weichen@northwestern.edu

information of heterogeneous materials can be accurately captured via N -point correlation functions [8-11]. As a balance between computational cost and accuracy, the 2-point correlation function (autocorrelation) [12] is widely adopted in practice. However, correlation functions lack clear physical meanings. It is inconvenient to design an optimal correlation functions as they are infinite dimensional [13, 14]. Furthermore, correlation function-based microstructure reconstructions are either computationally expensive (when using the pixel moving optimization algorithm [15]), or lacking of stochasticity (when using the phase recovery algorithm [16]). With the physical descriptor-based approach, microstructures are represented by physically meaningful structural parameters (descriptors), such as volume fraction, particle number, and particle size. In our recent research [13], we classified microstructure descriptors into three categories: composition, dispersion, and geometry. The major strengths of physical descriptors are the clear physical meanings they offer and meaningful mappings to processing parameters [17]. We have developed a descriptor-based methodology for characterization and reconstruction of polymer nanocomposites [14, 18]. However, the descriptors were chosen based on experiences. A systematic approach of identifying key microstructure descriptors as material design variables is needed.

Material informatics [19, 20] is a growing area that leverages information technology and data science to represent, parse, store, manage, and analyze the material data. The goal is to share and mine the data for uncovering the essence of materials, and accelerate the new material discovery and design [21]. Data mining and machine learning techniques have been applied to exploit material databases and discover trends and mathematical relations for material design. To manage the information complexity of using large-dimensional representations of microstructures, recent work has attempted unsupervised microstructure dimensionality reduction via manifold learning [9] and kernel principal components [22]. However, dimension reduction of microstructure parameters considering the microstructure only does not reflect its impact on material properties of interest so that the reduced parameter set does not address the direct need of material design. Supervised learning [23], a concept in machine learning where labeled training data are used to infer a relationship, has been employed in establishing the Process-Composition-Property relation for metals [1, 2] and predicting polymer composites' properties based on the composition-property database [24-26]. However, limited efforts have been made on modeling the microstructure-property relation using statistical learning and further reducing the high dimensionality of microstructure representations obtained from analyzing microscopic images.

High dimensionality is handled in machine learning by feature selection and extraction, to reduce the number of variables in a system by either selecting a subset of relevant features, or transforming the original high-dimensional feature space into a space of fewer dimensions. Both selection and extraction can be either supervised or unsupervised. The transformation incurred by extraction methods usually refers to

a linear or nonlinear combination of the original variables, in order to construct new features for improved description of data. In this regard, extraction methods are not suitable for our needs. We rather want to retain the clear physical meanings of features (descriptors) so as to use them as design variables. Feature selection, on the other hand, chooses a subset of more informative features from the original set and well fits our scenario. Only looking at the microstructure descriptors forms an unsupervised learning process. If the corresponding responses (behavior) of microstructures, in our case the morphology and properties are also available, supervised learning provides more insights in the selection process.

Existing supervised feature selection methods typically involve developing heuristics or measures to evaluate the worth of features. Examples of heuristics developed in literature include information gain [27], Gini index [28], Chi-square and other distance measures. The limitation is that they can only handle discrete variables as the supervisory signal, as when the desired output is within a set of a small number of known labels. However, if the supervisory signal is presented as continuous, many feature selection methods fail to work properly. In our case, both microstructure correlation functions and properties are provided as continuous values, and therefore pose challenges for the feature selection procedure. What's more, distance measure based heuristics do not take into account the feature interactions and dependencies, for example, the surface area of filler phase and that of matrix phase in our descriptor group have a high dependency, which cannot be appropriately addressed by common distance measures. The family of Relief algorithms, beginning with the basic form of Relief [29] and being later adapted into RelifF [30] and RReliefF, are efficient and effective heuristic measures that correctly estimate the quality of features considering their capability of differentiating opposite-class training examples. The first two in the family are developed for discrete problems. RReliefF [31], the algorithm employed in this research, accounts particularly for continuous problems. For simplicity we refer it as Relief.

In this paper, we propose a 4-step machine learning methodology for identifying the key microstructure descriptors as potential material design variables. In Step 1, image analysis is applied to gather an initial set of potential microstructure descriptors (Section 2), to understand the dependencies among the descriptors and the topological constraints of the microstructure morphology (Section 3.1). In Step 2, an image analysis-based supervised learning further reduces the descriptor set by analyzing each descriptor's influence on the microstructure morphologies represented by the correlation functions (Section 3.2). In Step 3, material property-based supervised learning is employed using data obtained from physics-based simulations or from literature for further dimension reduction to identify the key set of descriptors (design variables) that have the largest impact on properties of interest (Section 3.3). In Step 4, microstructure design variables are selected from key descriptors (Section 3.4) by maximizing the impact score and minimizing the dependency. We

demonstrate the strength of the proposed method with the design of polymer nanocomposites (Section 4).

2. TECHNICAL BACKGROUND OF STATISTICAL MICROSTRUCTURE REPRESENTATIONS

In this section, we provide the technical background of microstructure characterization with the example of bi-phase nanoparticle-reinforced polymer composite. Statistical microstructure characterization enables a quantitative understanding of the microstructure–property relationship through physics-based simulations and the sensitivity of the heterogeneity of microstructure morphology with respect to their impacts on the prediction of bulk properties. The heterogeneity is represented by statistical characteristics, which can be further utilized as variables for material design to shape the microstructure morphology for achieving target material properties. The microstructure reconstruction techniques are needed in the prediction of 3D microstructure morphology based on 2D information, when three dimensional (3D) imaging techniques like the X-ray or micro-tomography are not affordable or unavailable [18, 32]. This section introduces the technical background of two types of microstructure characterization techniques: correlation function-based method (Section 2.1) and descriptor-based method (Section 2.2). We also summarize a list of commonly used descriptors covering composition, dispersion status, and geometry information of the inclusions.

2.1 Correlation function-based microstructure characterization

A wide range of microscopic imaging techniques such as scanning electron microscopy (SEM) [7, 33] and transmission electron microscopy (TEM) [34] are applicable to obtain the digital microstructure images for statistical characterization. In the step of image preprocessing, the bi-phase microstructure images are denoised and binarized with the volume fraction of each phase maintained. In the binary image, pixels in the matrix phase are marked by “0” and pixels in the filler phase are marked by “1”. Figure 1 illustrates the transformation of the grey scale SEM image (left) to a binary image (right), where black pixels represent nanoparticle filler and white pixels represent polymer matrix. The binary pixelated images are used in both correlation function-based and descriptor-based characterization.

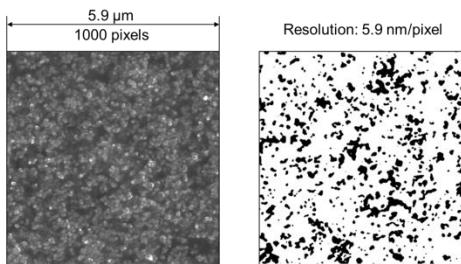


Figure 1: SEM and binary images of polymer nanocomposites

In this work, we collect four types of correlation functions [35, 36]: 2-point correlation function, (2-point) surface

correlation function, lineal path function and radial distribution function (Figure 2). These four types of correlation functions are widely used for an accurate representation of microstructure with affordable computational costs.

The meaning of the 2-point correlation function is the probability of finding two points with a given distance r in the same phase of the random media. Therefore, two-point correlation is a function of distance r . r can be any value from 0 to infinity. Similarly, we can define other types of correlation functions (Figure 2). The surface correlation is defined as the probability that finding two points with a distance r located on the boundary of the filler phase. Surface correlations capture the morphology of interphase. The lineal path correlation is the probability that an entire line of length r lies in the filler phase. Lineal path correlations describe the connectedness of fillers. Radial distribution function is the probability of finding fillers on a circle of radius r around filler cluster centers. These four correlation functions are complementary to each other. Different correlation function emphasizes on different aspects of microstructure features. We assume that high-dimensional microstructure images are fully represented by these four correlation functions [37].

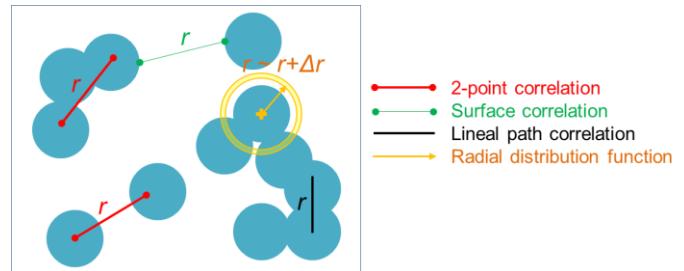


Figure 2: Definitions of different types of correlation functions

2.2 Descriptor-based microstructure characterization

A descriptor-based approach is proposed in our prior work to represent microstructure morphologies using three levels of microstructure features [14]: composition, dispersion, and geometry (Figure 3). Composition descriptors distinguish different phases and describe their volume/weight percentage in the material, such as volume fraction of filler in polymer composites. Dispersion status descriptors depict the inclusions’ spatial relation and their neighbor status, such as the nearest neighbor distance, number of filler clusters [11, 33, 38, 39], etc. Geometry descriptors are on the lowest length scale, which describe the inclusions’ shapes. Geometry descriptors include the inclusions’ size distribution, surface area, surface-to-volume fraction, roundness, eccentricity, elongation, rectangularity, tortuosity, aspect ratio, etc. [8-10, 33, 39-43]. The descriptor-based methodology is featured by three strengths: the well-defined physical meaning of microstructure characteristics, the low computational cost in characterization/reconstruction, and the low dimensionality of parameterized microstructure characteristics that enables parameter-based optimal microstructure design. On the largest length scale, composition is the lowest order of microstructure information as it only captures the homogenized response of an entire material

system. One scale down, dispersion is capable to capture material properties induced by local features. For example, the nearest distance between inclusions and voids in alloy determines the fracture behaviors, and the orientation of the fibers determines the strength of the fiber composites. On the lowest length scale, filler clusters' geometry features also play an important role. For example, the shape of voids in porous materials determines the yield point or the critical load in microvoid-induced microbuckling. Descriptors are statistical, quantified by mean and higher order moments such as variance, skewness, kurtosis, etc. With a sufficient descriptor set, high orders of microstructure information can be captured [14].

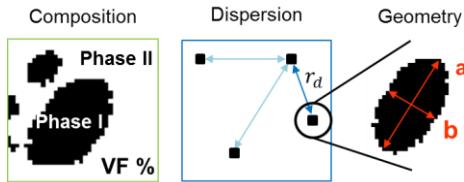


Figure 3: Illustration of three levels of microstructure descriptors: composition, dispersion and geometry

Table 1: Collected microstructure descriptors. Statistical information includes 1 to 4th orders of moments: mean, variance, skewness and kurtosis

Descriptor	Definition	Type
<i>Composition</i>		
VF	Volume Fraction	Deterministic
<i>Dispersion</i>		
r_{nsd}	Cluster's nearest surface distance	Statistical
r_{ncd}	Cluster's nearest center distance	Statistical
θ	Principle axis orientation angle [44]	Statistical
I_{filter}	Surface area of filler phase	Deterministic
I_{matrix}	Surface area of matrix phase	Deterministic
N	Cluster number	Deterministic
V_{VF}	Local VF of Voronoi cells [40]	Statistical
<i>Geometry</i>		
r_p	Pore sizes (inscribed circle's radius) [45]	Statistical
A	Statistical	Statistical
r_c	Equivalent radius, $r_c = \sqrt{A/\pi}$	Statistical
δ_{cmp}	Compactness [46]	Statistical
δ_{rnd}	Roundness [47]	Statistical
δ_{ecc}	Eccentricity [47]	Statistical
δ_{asp}	Aspect ratio [34, 42]	Statistical
δ_{rec}	Rectangularity [47]	Statistical
δ_{tor}	Tortuosity [47]	Statistical

In this paper, we collect a large set of descriptors from literature as candidates of microstructure design variables. This section covers descriptors used in polymer nanocomposites, alloy, fiber composites and ceramic composites, etc. In previous works, different descriptors are chosen for different materials based on expertise. Often times, the descriptors used in a single work only capture the microstructure features that are highly related to the interested properties, while all the other microstructure features are neglected. Therefore, to avoid bias in the key descriptor learning, it is necessary to include a wide range of descriptors from different types of materials. The full candidate descriptor set is referred to as the “full descriptor set”

in this paper. The collection of descriptor titles and their definitions are provided in Table 1. There are 17 descriptors in the list, in which each statistical descriptor is represented by 4 parameters (1 to 4th order moments). In total, the 17 microstructure descriptors are represented using 56 descriptor parameters.

3. MACHINE LEARNING-BASED IDENTIFICATION OF KEY DESCRIPTORS

In the presence of a large number of microstructure descriptors, the key research questions is how to represent the microstructural design space quantitatively using a descriptor set that is sufficient yet small enough to be tractable. A 4-step machine learning-based method is proposed to exploit the microstructure-property database (Figure 4). The four steps include: (1) Elimination of redundant descriptors using descriptor-descriptor correlation analysis; (2) Microstructure correlation function-based supervised learning for further dimension reduction; (3) Property-based supervised learning to identify key descriptors; (4) Determination of microstructure design variables based on the optimization criteria of maximizing the impact score and minimizing the within-group correlations of the selected descriptor set. Steps 1 and 2 are image analysis-based procedures, which don't require expensive Finite Element Analysis (FEA) simulations. These two steps will provide a fast reduction of the size of a candidate descriptor set. Both steps 1 and 2 involve supervised learning. Step 3 needs structure-property data from either high-fidelity simulations or from literature. Step 4 is an optimization-based descriptor subset selection process.

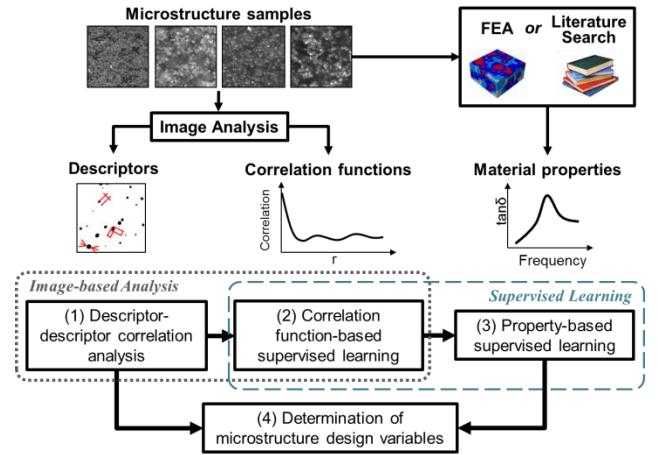


Figure 4: Framework of machine learning-based microstructure descriptor identification

To build a rich set of data, multiple microstructure images are collected for the type of materials of interest. For each material sample, one Representative Volume Element (RVE) size image or multiple Statistical Volume Element (SVE) size images should be collected [48]. For each image, a full set of microstructure representations (correlation functions and

descriptors) are evaluated using the characterization techniques introduced in Section 2.

3.1 Descriptor-descriptor correlation analysis for identifying redundant descriptors

In Step 1 of the proposed framework, redundant descriptors are identified by the pair-wise descriptor-descriptor correlation analysis. Some descriptors may be strongly correlated due to the pre-existing relations. For example, geometry descriptor cluster area A and major radius r of the fillers in Microstructure I (Figure 5) follow a strict mathematical relation of $A = \pi r^2$, so these two descriptors A and r are interchangeable to each other, therefore one of them becomes redundant. However, in Microstructure II (Figure 5) there is no clear mathematical relation between A and r, so both descriptors are to be kept for designing Microstructure II.

Since the mathematical relation may not necessarily be linear (such as the example given in Figure 6), rank correlation is preferred as the measure of descriptor-descriptor correlation to the widely-used correlation coefficient, which only measures the linear dependence between variables. In statistics, rank correlation (Kendall's τ) measures the degree of similarity between two rankings, and is used to assess the significance of the two variables' relation. The formula computing Kendall's τ is shown below.

$$\tau = \frac{a-b}{\frac{1}{2}n(n-1)} \quad (1)$$

where a is the number of concordant pairs, and b is the number of discordant pairs. Any pair of observations (x_i, y_i) and (x_j, y_j) are said to be concordant if the ranks for both elements agree: that is, if both $x_i > x_j$ and $y_i > y_j$ or if both $x_i < x_j$ and $y_i < y_j$. The pair of observations (x_i, y_i) and (x_j, y_j) are said to be discordant if $x_i > x_j$ and $y_i < y_j$ or if $x_i < x_j$ and $y_i > y_j$. They are neither concordant or discordant if $x_i = x_j$ and $y_i = y_j$.

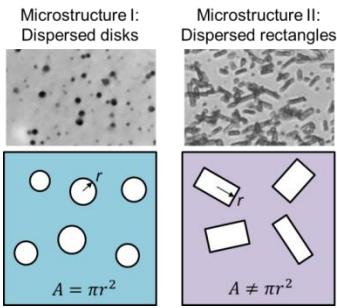


Figure 5: Illustration of redundant microstructure descriptors. In Microstructure I, area A and major radius r can replace each other; In Microstructure II, both are needed for a full microstructure representation

The advantage of using rank correlation is demonstrated in Figure 6. x_1 and y_1 are defined with a linear relation. Both the correlation coefficient and the rank correlation between x_1 and y_1 are 1, which indicates that there exists a perfectly interchangeable relation between the two variables. One of

them can be eliminated as redundant design variables without any information loss. In the case of x_2 and y_2 , they are defined by a non-linear relation. Still only one of them is needed to be used as the design variable. The value of rank correlation is 1, which indicates redundancy in design variables; On the other hand, the value of correlation coefficient is smaller than 1, because correlation coefficient fails to capture non-linear mathematical relations.

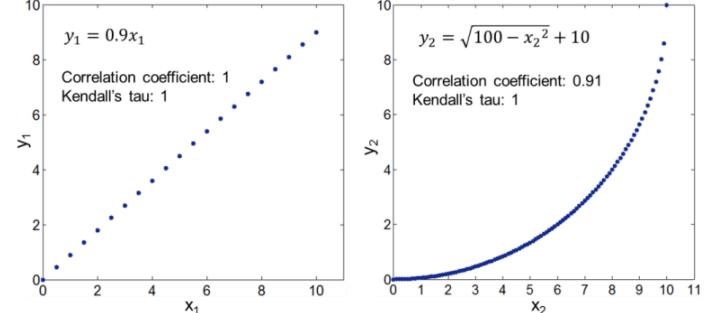


Figure 6: Comparison of correlation coefficient and rank correlation. In both cases there exists a perfect mathematical relation (monotone increasing) between x and y. Rank correlation Kendall's tau identifies both cases, while correlation coefficient only identifies the linear case.

3.2 Correlation function-based supervised learning

Step 1 of the proposed framework eliminates a few redundant microstructure descriptors based on the descriptor-descriptor correlations, but it doesn't provide any information on the significance of each descriptor to the properties of interest. Supervised learning is needed to search the key descriptors. However, it is not realistic to directly conduct property-based supervised learning on the large set of microstructure descriptors. The high dimensionality of descriptor set requires a great amount of microstructure samples (e.g. 10 times of the number of dimension) in structure-property simulations for supervised learning. This process may not be affordable due to the high computational costs of simulations. For example, a high fidelity damping property simulation that explicitly models the microstructures of an $80 \times 80 \times 80$ voxel size 3D microstructure takes over 80 hours [49, 50]. Therefore, a simulation-free, image analysis-based supervised learning step (Step 2) is proposed to further reduce the number of candidate descriptors before property-based supervised learning in Step 3.

In Step 2, each descriptor's influence on microstructure morphology is evaluated based on their influences on the four correlation functions introduced in Section 2.1. For each descriptor, four impact scores (on four correlation functions) are evaluated using supervised learning algorithm. Their average is taken as the descriptor's final score. Relief [31] is employed as the supervised learning algorithm, which takes descriptors as input features and the sum of correlation function values as the supervisory signal. We take the sum of first 50 points of correlation functions, which represents the homogenized high-strength correlation within a distance of 50 pixels, 295 nm.

Relief uses a statistical method and avoids heuristic search. Only statistically relevant features are selected. The key idea of the Relief algorithm is to estimate the quality of attributes according to how well their values distinguish between instances that are near to each other within a local context. Figure 7 gives the pseudo code of the basic Relief that handles the discrete class case. Given a randomly selected case R (line 3), two nearest neighbors are searched. They can either be from the same class, called *hit* H, or from the different classes, called *miss* M (line 4). A quality estimation vector W is updated for all attributes A (line 5 and 6). The process is repeated for m times, where m is a user-defined parameter.

For categorical attributes, the outcome of the function $\text{diff}(\text{Attribute}, \text{Instance1}, \text{Instance2})$ is a binary value, 0 being the values of Attribute agree between Instance1 and Instance2 and 1 otherwise. For continuous attributes, the function $\text{diff}(\text{Attribute}, \text{Instance1}, \text{Instance2})$ is defined as:

$$\text{diff}(A, I_1, I_2) = \frac{|\text{value}(A, I_1) - \text{value}(A, I_2)|}{\max(A) - \min(A)} \quad (4)$$

The above function calculates the difference between the values of Attribute for two instances, where Instance1 is a random instance, and Instance2 can be either hit H or miss M.

Algorithm Relief

Input: for each training instance a vector of attribute values and the class value
Output: the vector W of estimations of the qualities of attributes

```

1. set all weights W[A] := 0.0;
2. for i := 1 to m do begin
3.     randomly select an instance R;
4.     find nearest hit H and nearest miss M;
5.     for A := 1 to #all_attributes do
6.         W[A] := W[A] - diff(A,R,H)/m + diff(A,R,M)/m;
7. end;
```

Figure 7: Pseudo code of the basic Relief algorithm for discrete classes

To handle regressive cases, instead of the above difference functions, a kind of probability is introduced to address how much the predicted values of two instances are different. This probability can be modelled with the relative distance between the predicted (class) values of two instances. The output of this algorithm, after going through all instances, is the quality estimation vector (impact factors) W that represents the estimations of the qualities of each feature.

Finally, according to the obtained quality estimation vector W, features are ranked, and how many are to be selected from the ranked list is a decision subject to the user. Relief requires linear time in the number of given features and the number of instances regardless of the target concept to be learned.

3.3 Property-based supervised learning

The end goal of the machine learning framework is to identify key microstructure descriptors as design variables to optimize for achieving target material properties. In the third step of the framework, supervised learning is employed to study descriptors' influences on properties of interest. The reduced descriptor set from the first two steps is used as inputs, and material properties are taken as the supervisory signal. The properties of microstructure samples are either obtained from advanced FEA or collected from literature. The Relief algorithm is employed again to calculate the score of each descriptor on each property of interest. The learning result is normalized such that the scores of all microstructure descriptors are in the range of [0, 1] and add up to 1. If multiple properties are considered in material design, the supervised learning is applied on each property for all descriptors, and then the scores are added together to determine the final ranking of the microstructure descriptors.

3.4 Determination of microstructure design variables

A small set of microstructure descriptors are chosen from the key descriptors as microstructure design variables. It is not realistic to include all key descriptors as design variables because the strong descriptor-descriptor correlations may lead to unrealistic (infeasible) designs. The microstructure design variables should have high contribution to material properties (high ranking from machine learning) and high dependency (low descriptor-descriptor correlation). A combinatorial search is conducted to determine the most proper subset of descriptors by formulating the problem as a two-objective heuristic search:

Given the number of design variables n, find descriptors d_1, d_2, \dots, d_n , s.t.:

Min: $\sum C_{ij}$, where $i = 1, 2, \dots, n, j = 1, 2, \dots, n, i \neq j$;

Max: $\sum_{k=1}^n S_k$

C_{ij} is the correlation between any two descriptors. S_k is the k -th descriptor's contribution to the properties (impact score).

4. DESIGN OF POLYMER COMPOSITES USING REDUCED DESCRIPTOR SET

The addition of reinforcing particles to polymer nanocomposites' matrix can lead to significant improvements in homogenized mechanical properties even at a very low filler concentration [51]. High impact of the quantity and morphologies of nanoparticle fillers on damping property makes it an interesting design problem for microstructure optimization. This section demonstrates how to use the proposed method to determine the key microstructure descriptors as design variables for carbon black nanoparticle filled polymer elastomers. All material samples collected have the same type of fillers, but are produced under different processing conditions, which directly impact the morphology of nanoparticle clusters in the polymer matrix. 56 microstructure images are collected on materials produced under 11 different processing conditions. The pixel size of the SEM images is

1000×1000. The physical size is 5.9×5.9 μm, which can be considered as RVE. Sample images are shown in Figure 1 and Figure 4, representatively. The microstructure information includes 4 types of correlation functions and 17 types of microstructure descriptors discussed in previous section (also see Appendix). The interested material properties are the damping property, which is defined as:

$$\tan \delta(\omega) = \frac{G''(\omega)}{G'(\omega)} , \quad (5)$$

where G'' is the shear loss modulus (GPa), G' is the shear storage modulus (GPa). $\tan \delta$, G' , and G'' are all functions of ω , the frequency of excitation (Hz).

To achieve a long wear life, low rolling resistance, and high wet traction of tire materials, it is desired that the nanocomposites have a $\tan \delta$ curve with low value in the low frequency domain (smaller than 1×10^{-1} Hz), high value in the normal (from 1×10^{-1} to 1×10^3 Hz) and high frequency domains (larger than 1×10^3 Hz) [14]. Shown in Figure 8, we choose three property characteristics as the design criteria: value of the first point on $\tan \delta$ (L, representing damping property in low frequency domain), value of the peak on $\tan \delta$ (P, representing normal frequency domain), and value of the last point on $\tan \delta$ (H, representing high frequency domain). L, P and H represent three different desired properties of tire material respectively: low L leads to low rolling resistance; high P leads to high wet traction; high H leads to low wear. The goal is to achieve good performances on all three properties simultaneously. Thus, the design problem is formulated as a multiobjective optimization problem:

Find a set of microstructure descriptors, s.t.: Min L; Max P (Min -P); Max H (Min -H);

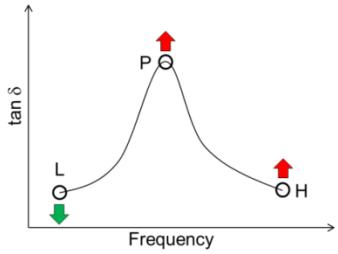


Figure 8: Three design criteria defined by three points on $\tan \delta$

4.1 Results of correlation-based feature selection

In Step 1 of the proposed framework (descriptor-descriptor correlation analysis), the rank correlation (Kendall's tau) is calculated for each pair of descriptor parameters and written into a 56×56 symmetric matrix. The correlation matrix is reordered using Cuthill-McKee algorithm [52], which permutes the symmetric correlation matrix to obtain a band matrix form with a small bandwidth. The permuted correlation matrix is shown in Figure 9. The correlation values are represented by colors. Darker color means a higher correlation, and lighter color means a lower correlation. Numbers on the X and Y axis represent different microstructure descriptor parameters (refer

to Appendix). Figure 9 indicates several highly intra-correlated descriptor groups. Group 1 includes the composition descriptor (VF) and five dispersion descriptors (r_{nsd} and r_{ncd}). Group 2 incorporates five dispersion descriptors related to the quantity of surface area (I_{matrix} , N and V_{VF}). Group 1 and 2 are inter-correlated. Group 3 incorporates three geometry descriptors (δ_{asp} and δ_{ecc}). Group 3 is independent from the other two groups. The intra-correlations exist between descriptor groups for two reasons. (1) Microstructure features are correlated inherently. For example, given the sample volume of fillers, increasing the number of filler cluster N will lead to larger surface area of the filler phase I_{matrix} . Therefore, a high correlation (0.8360) can be observed between N and I_{matrix} . (2) Some descriptors describe the same microstructure feature in different ways. For example, three highly correlated descriptors, cluster number N , local volume fraction of each Voronoi cell V_{VF} , and Surface area I_{matrix} , all describe the quantity of fillers' surface area from different perspectives.

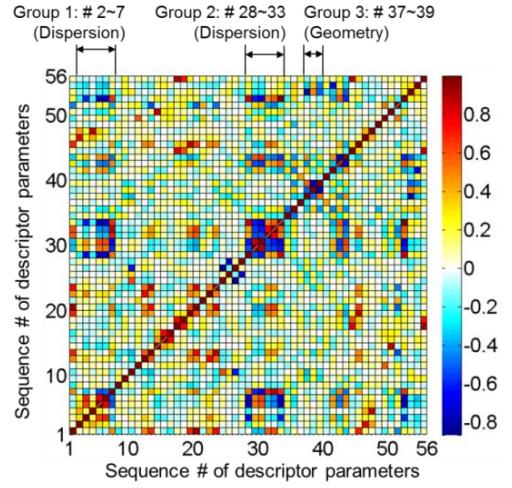


Figure 9: The permuted rank correlation matrix. Larger correlations are marked by darker colors. White color means the correlation is 0 (no correlation). The permuted matrix shows several descriptor groups of high intra-correlations. The sequential numbers on X, Y axis represent different descriptors. Refer to Appendix for the meanings of sequential numbers.

To determine the redundant microstructure descriptors, a threshold is set on the correlation matrix to distinguish “strongly-correlated” descriptor pairs and “weakly-correlated” descriptor pairs. Two different threshold values (0.9 and 0.8) are tried to study the threshold’s influence on dimension reduction. In the binarized correlation matrix (Figure 10), light yellow represents correlation values passing the threshold, and dark green represents those failing the threshold. The highly correlated descriptors (Table 2) are considered as interchangeable, as the result, one of each pair can be eliminated to reduce the dimension. With a higher threshold (larger than 0.9 or smaller than -0.9), the dimension is reduced from 56 to 54. Redundancy exists in one pair of dispersion descriptors (I_{matrix} and I_{filler} are the measurements of surface area) and one pair of geometry descriptors (A_{-1} and r_{c-1}

describe the size of each filler cluster). When using a lower threshold (larger than 0.8 or smaller than -0.8), the size of the candidate descriptor set can be reduced by 15 (redundancy exists in 6 pairs of dispersion descriptors and 9 pairs of geometry descriptors).

In addition to dimension reduction, correlation analysis among microstructure descriptors leads to other important conclusions. Geometry descriptors tend to be independent with composition and dispersion descriptors. High correlations exist between descriptors of the same category. This observation validates our method of classifying all microstructure descriptors into three categories [14] according to the different levels of details (global versus local).

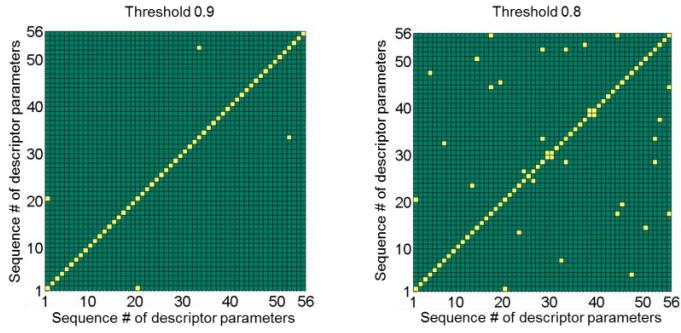


Figure 10: The binarized correlation matrix. Light yellow means a correlation larger than or equal to the threshold; dark green means a correlation smaller than the threshold. Two thresholds are tested: 0.9 and 0.8

Table 2: The highly correlated descriptors identified with different thresholds

Threshold = 0.9	
$I_{matrix} \leftrightarrow I_{filler}$	Dispersion
$A_1 \leftrightarrow r_c-1$	Geometry
Threshold = 0.8	
$VF \leftrightarrow V_{VF}-1$	Composition, Dispersion
$I_{matrix} \leftrightarrow I_{filler} \leftrightarrow N$	Dispersion
$\theta_2 \leftrightarrow \theta_4$	Dispersion
$r_{nsd_3} \leftrightarrow r_{nsd_4}$	Dispersion
$V_{VF_3} \leftrightarrow V_{VF_4}$	Dispersion
$A_1 \leftrightarrow r_c-1$	Geometry
$A_4 \leftrightarrow A_3 \leftrightarrow r_c-4$	Geometry
$\delta_{rec_3} \leftrightarrow \delta_{rec_4}$	Geometry
$\delta_{rnd_1} \leftrightarrow r_p-1$	Geometry
$A_2 \leftrightarrow r_c-2$	Geometry
$\delta_{asp_1} \leftrightarrow \delta_{ecc_1}$	Geometry
$\delta_{rec_3} \leftrightarrow \delta_{rec_4}$	Geometry
$\delta_{ecc_3} \leftrightarrow \delta_{ecc_4}$	Geometry

4.2 Supervised learning of key descriptors

The results of Step 2 correlation function-based learning and Step 3 property-based learning are listed in Table 3. In Step 2, we calculate the scores of all 56 microstructure descriptors based on their influences on correlation functions. The scores add up to 1. The size of candidate descriptor set is reduced to 20, when a threshold of 0.5 is set on the sum of scores of a reduced descriptor set. The top 20 descriptors are identified as significant descriptors. Next in Step 3, property-based learning,

is conducted to further evaluate the significant descriptors' impacts on material properties. The final ranking and scores of the 20 significant descriptors are listed in the right half of Table 3. In addition, this study leads to another two important observations:

(1) For the type of materials studied in this paper, the composition and dispersion descriptors have strong influences on both correlation function and properties. Composition and dispersion descriptors are on higher (global) levels than the geometry (local) descriptors. The material property of interest (damping property) is the averaged response of the bulk material, so it is reasonable that the property has higher correlations with the high level dispersion descriptors than the low level geometry descriptors.

(2) For this particular type of materials, we observed a high similarity between the correlation function-based descriptor ranking and the property-based descriptor ranking. The two rankings have the same top 5 descriptors, share 9 descriptors out of top 10, and share 17 descriptors out of top 20.

Table 3: Results of Step 2 correlation function-based supervised learning, and Step 3 property-based supervised learning. The meanings of the symbols are listed in Table 1. The number after each symbol indicates the statistical moment of the descriptor (1st: mean; 2nd: variance; 3rd: skewness; 4th: kurtosis). Different descriptors from the two lists are highlighted with dark background color.

Rank	Step 2: Correlation function-based			Step 3: Property-based		
	Descriptor	Category	Score	Descriptor	Category	Score
1	I_{filler}	Dispersion	0.0362	I_{matrix}	Dispersion	0.0623
2	I_{matrix}	Dispersion	0.0358	I_{filler}	Dispersion	0.0623
3	VF	Composition	0.0348	N	Dispersion	0.0618
4	V_{VF_1}	Dispersion	0.0340	VF	Composition	0.0587
5	N	Dispersion	0.0335	V_{VF_1}	Dispersion	0.0584
6	V_{VF_2}	Dispersion	0.0286	A_1	Geometry	0.0491
7	δ_{rnd_1}	Geometry	0.0263	V_{VF_2}	Dispersion	0.0491
8	A_1	Geometry	0.0246	δ_{rnd_1}	Geometry	0.0485
9	r_p-1	Geometry	0.0243	r_p-1	Geometry	0.0484
10	V_{VF_3}	Dispersion	0.0243	θ_2	Dispersion	0.0483
11	r_c-1	Geometry	0.0242	V_{VF_3}	Dispersion	0.0479
12	r_{nsd_1}	Dispersion	0.0232	r_c-1	Geometry	0.0476
13	r_p-2	Geometry	0.0223	r_c-2	Geometry	0.0474
14	r_{nsd_2}	Dispersion	0.0222	r_{nsd_1}	Dispersion	0.0461
15	r_{ncd_1}	Dispersion	0.0222	r_{nsd_2}	Dispersion	0.0457
16	θ_2	Dispersion	0.0219	θ_4	Dispersion	0.0456
17	δ_{cmp_1}	Geometry	0.0215	r_{ncd_1}	Dispersion	0.0438
18	r_c-2	Geometry	0.0213	V_{VF_4}	Dispersion	0.0435
19	V_{VF_4}	Dispersion	0.0212	r_p-2	Geometry	0.0433
20	θ_4	Dispersion	0.0206	δ_{cmp_1}	Geometry	0.0422

The benefits in computational efficiency are concluded from the perspective of problem dimensionality (number of design variables). The size of candidate descriptor set is reduced from 56 to 41 after Step 1 (when threshold is 0.8). According to our recommendation that the sample number should be at least 10 times of the number of design variables, it requires 410 microstructure samples (410 high fidelity simulations) when property-based machine learning is directly applied on the 41 descriptors. On contrary, the correlation function-based learning (Step 2) eliminate low-impact descriptors. The dimension is further reduced to 20. It means

that 210 less samples (210 high fidelity simulations) are required for Step 3 property-based supervised learning.

4.3 Design validation: optimization of polymer nanocomposites' microstructure

A comparative study of microstructure design is conducted to validate the effectiveness of using the machine learning-identified descriptors as microstructure design variables. In Step 4, three microstructure design variables are determined ($N=3$) by maximizing the descriptor set's impact score and minimizing the within-group correlation. The design variable set includes one composition descriptor (volume fraction, VF), one key dispersion descriptor (number of particle clusters, N), and one key geometry descriptor (mean of roundness, δ_{rnd_1}). These three descriptors have a strong impact on the damping properties and are weakly correlated with each other. For the purpose of comparative study, we choose another design variable set empirically (referred as "empirical descriptor set"), with three descriptors (VF, r_{nsd_1} , and δ_{rec_1}). These three descriptors have low correlations (to guarantee design feasibility), but they are not necessarily strongly correlated with the properties. The value ranges of the descriptors are listed in Table 4.

Table 4: Two design variable sets. Set 1 is chosen based on the results of machine learning; Set 2 is chosen empirically

Design variable set 1 (statistical learning)	Lower bound	Upper bound
VF	0.1	0.3
N	100	300
δ_{rnd_1} , mean	1	4
Design variable set 2 (empirical)	Lower bound	Upper bound
VF	0.1	0.3
r_{nsd_1} , mean	10	40
δ_{rec_1} , mean	$\pi/4$	1

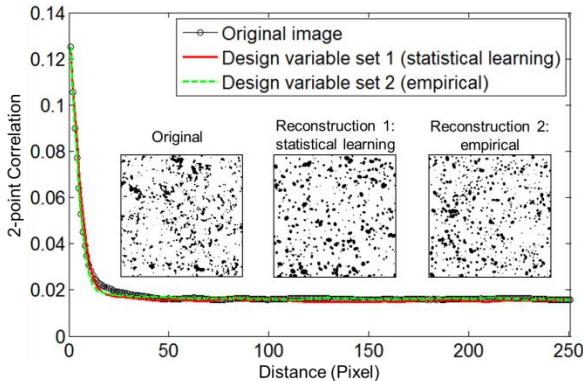


Figure 11: Microstructure reconstructions using statistically learned descriptor set and empirical descriptor set

We validate that the statistically learned microstructure descriptors can capture microstructure information accurately. The descriptor-based microstructure reconstructions are compared with the original image (Figure 11). Both reconstructions using statistically-learned descriptor set and empirical descriptor sets respectively match well with the

original image in visual comparison and correlation functions. The statistically-learned descriptor set is more accurate, as it has a smaller Sum of Squared Error in 2-point correlation function compared with the reconstruction from empirical descriptor set (1.83×10^4 vs. 6.26×10^4).

Microstructure optimization is conducted using a DOE/metamodeling-based optimization strategy. For both Descriptor Sets 1 and 2, Design of Experiment (DOE) is applied to explore the design space formed by microstructure descriptors. Each DOE point represents one microstructure design, for which we reconstruct one or multiple statistically equivalent microstructures [14]. The properties of reconstructions are simulated using FEA. Metamodels, also known as surrogate models [53], are created to replace the computationally expensive FEA models in optimization. The optimal designs are verified by running simulations on the reconstructed microstructures.

We make two sets of comparisons to demonstrate that the performances of microstructure designs are improved by using key descriptors as microstructure design variables. (1) Single-objective optimizations. Microstructure descriptors are used as design variables in single-objective optimizations to minimize L, maximize P and maximize H respectively. We compare the single-objective optimal designs using the key descriptor set and the optimal designs using the empirical descriptor set in Figure 12. Significant improvements in properties are achieved by using statistically-learned descriptors as design variables. Figure 13 shows two examples of optimal microstructure designs, which are obtained using different design variable sets. The optimal microstructure design by the key descriptor set is more dispersed (have larger surface area) compared with the design by the empirical descriptor set. Larger surface area leads to a higher value of property H.

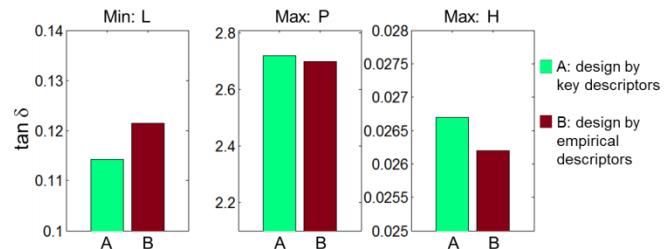


Figure 12: Comparison of single-objective optimal designs using key descriptors and empirical descriptors

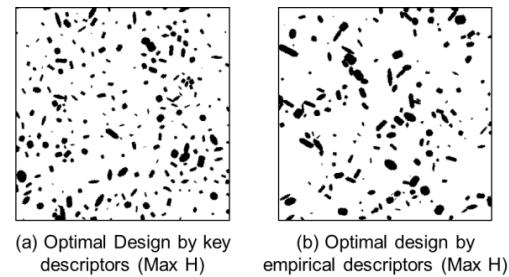


Figure 13: Examples of optimal microstructures (Max H) obtained by key descriptor set and empirical descriptor set

(2) Multi-objective optimization. Microstructure descriptors are used as design variables in multi-objective optimizations which minimize L, maximize P and maximize H simultaneously. Equal weights are assigned to the three normalized objectives. Better designs can be obtained by using key descriptors in multi-objective optimization as well (Figure 14). It is observed that the optimal design using the key descriptor set has lower L, roughly the same P and higher H compared with the optimal design using empirical descriptor set as design variables.

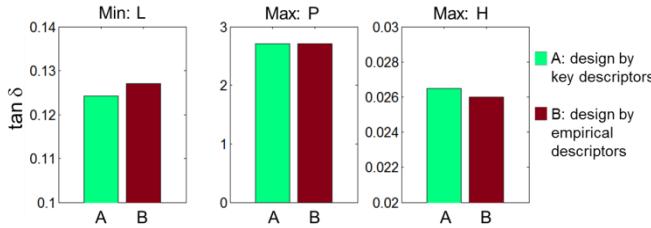


Figure 14: Comparison of multi-objective (Min: L, Max: P, Max: H) optimal designs using key descriptors and empirical descriptors. Equal weights are assigned to three normalized objectives.

5. CONCLUSION

This paper presents a machine learning-based method for identifying key microstructure descriptors as microstructure design variables. It facilitates low dimensional descriptor-based microstructure representations. Starting from a complete set of microstructure descriptors collected from literature, we reduce the redundant descriptors via descriptor-descriptor correlation analysis and correlation function-based supervised learning. These two steps are computationally efficient as only image analyses are involved. Furthermore, a property-based supervised learning is conducted to identify the key descriptors. Microstructure design variables are chosen from the key descriptors based on their contributions to material properties as well as the need for minimizing descriptor dependency. This 4-step method enables parametric optimization of heterogeneous microstructures using a small set of physically meaningful descriptors to achieve target properties. We verify this method using a case study of designing polymer nanocomposites' microstructures. This proposed method leads to better designs compared with designs using descriptors chosen randomly or empirically.

Our research contributes to the computational design of microstructural materials system in the following three aspects. First, this method provides a rigorous way for material scientists to choose microstructure descriptors in analyzing and designing new materials. It is demonstrated that optimization using key descriptors obtained by machine learning significantly improves the performance of microstructure designs. The proposed method identifies descriptors that are important to both microstructure morphology and properties. Second, this method effectively cuts down the computation costs by introducing image analysis-based prescreening. The prescreening steps reduce the dimension of the candidate descriptor set before applying the property-based supervised

learning. Conducting property-based supervised learning on the reduced descriptor set requires less number of samples (simulations) than on the full descriptor set. Third, this method enables parametric optimization of the microstructure with a small set of design variables. State-of-art computational design methods (e.g. DOE, metamodeling) are applied to explore the microstructural design space to achieve optimal material properties.

In future works, the process-microstructure relation will be included into the material design process to cover the whole spectrum of material design and to ensure that material engineers can fabricate the optimal microstructure. The same machine learning approach will be applied to the process-microstructure database to establish mathematical relations between processing parameters and the resultant microstructures.

APPENDIX: DESCRIPTOR PARAMETERS' SEQUENCE NUMBERS IN FIGURE 9

This table lists the number and name of the descriptors in Figure 9. The three intra-correlated descriptor groups are highlighted using grey background color.

The number after each symbol represents the order of the moment (1st: mean; 2nd: variance; 3rd: skewness; 4th: kurtosis).

Number	Descriptor Name	Number	Descriptor Name	Number	Descriptor Name
1	A_1	20	r_c_1	39	δ_{ecc_3}
2	r_ncd_3	21	r_p_4	40	δ_{ecc_2}
3	r_ncd_1	22	r_p_2	41	δ_{cmp_4}
4	r_nsd_4	23	r_p_1	42	δ_{cmp_2}
5	r_nsd_2	24	θ_4	43	δ_{cmp_1}
6	r_nsd_1	25	θ_3	44	A_4
7	VF	26	θ_2	45	A_2
8	δ_{tor_4}	27	θ_1	46	r_ncd_4
9	δ_{tor_2}	28	N	47	r_nsd_3
10	δ_{tor_1}	29	VVF_4	48	δ_{tor_3}
11	δ_{rnd_4}	30	VVF_3	49	δ_{rnd_3}
12	δ_{rnd_2}	31	VVF_2	50	δ_{rec_4}
13	δ_{rnd_1}	32	VVF_1	51	r_p_3
14	δ_{rec_3}	33	I_matrix	52	I_filler
15	δ_{rec_2}	34	δ_{asp_4}	53	δ_{ecc_1}
16	δ_{rec_1}	35	δ_{asp_3}	54	δ_{cmp_3}
17	r_c_4	36	δ_{asp_2}	55	A_3
18	r_c_3	37	δ_{asp_1}	56	r_ncd_2
19	r_c_2	38	δ_{ecc_4}		

ACKNOWLEDGMENTS

This work is supported by NSF grant (CMMI-0928320). The authors also sincerely thank Prof. Linda Schadler from Rensselaer Polytechnic Institute for providing the SEM images.

REFERENCES

- [1] Y. Li, "Predicting materials properties and behavior using classification and regression trees," *Materials Science and Engineering a-Structural Materials Properties Microstructure and Processing*, vol. 433, pp. 261-268, Oct 15 2006.

- [2] Doreswamy, K. S. Hemanth, C. M. Vastrand, and S. Nagaraju, "Data Mining Technique for Knowledge Discovery from Engineering Materials Data Sets," *Advances in Computer Science and Information Technology, Pt I*, vol. 131, pp. 512-522, 2011.
- [3] S. Broderick, C. Suh, J. Nowers, B. Vogel, S. Mallapragada, B. Narasimhan, *et al.*, "Informatics for combinatorial materials science," *Jom*, vol. 60, pp. 56-59, Mar 2008.
- [4] M. Ashby, *Materials Selection in Mechanical Design*. Burlington, MA: Butterworth-Heinemann, 2005.
- [5] D. L. McDowell and G. B. Olson, "Concurrent Design of Hierarchical Materials and Structures," *Scientific Modeling and Simulation SMNS*, pp. 1-34, 2008.
- [6] J. H. Panchal, S. R. Kalidindi, and D. L. McDowell, "Key computational modeling issues in Integrated Computational Materials Engineering," *Computer Aided Design*, 2012.
- [7] L. Karasek and M. Sumita, "Characterization of dispersion state of filler and polymer-filler interactions in rubber carbon black composites," *Journal of Materials Science*, vol. 31, pp. 281-289, Jan 15 1996.
- [8] S. Torquato, *Random Heterogeneous Materials: Microstructure and Macroscopic Properties*. New York: Springer-Verlag, 2002.
- [9] V. Sundaraghavan and N. Zabaras, "Classification and Reconstruction of Three-dimensional Microstructures using Support Vector Machines," *Computational Materials Science*, vol. 32, pp. 223-239, Feb 2005.
- [10] D. Basanta, M. A. Miodownik, E. A. Holm, and P. J. Bentley, "Using genetic algorithms to evolve three-dimensional microstructures from two-dimensional micrographs," *Metallurgical and Materials Transactions a-Physical Metallurgy and Materials Science*, vol. 36A, pp. 1643-1652, Jul 2005.
- [11] A. Borbely, F. F. Csikor, S. Zabler, P. Cloetens, and H. Biermann, "Three-dimensional characterization of the microstructure of a metal-matrix composite by holotomography," *Materials Science and Engineering a-Structural Materials Properties Microstructure and Processing*, vol. 367, pp. 40-50, Feb 25 2004.
- [12] Y. Liu, Greene, M.S., Chen, W., Dikin, D., Liu, W.K., "Computational Microstructure Characterization and Reconstruction to Enable Stochastic Multiscale Design," *Computer Aided Design*, 2012.
- [13] H. Xu, Y. Li, L. C. Brinson, and W. Chen, "Descriptor-based Methodology for Designing Heterogeneous Microstructural Materials System," presented at the ASME 2013 International Design Engineering Technical Conference & Computers and Information in Engineering Conference, IDETC/CIE 2013, 2013.
- [14] H. Xu, Y. Li, L. C. Brinson, and W. Chen, "A Descriptor-based Design Methodology for Developing Heterogeneous Microstructural Materials System," *Journal of Mechanical Design*, 2013, in press.
- [15] Y. Liu, M. S. Greene, W. Chen, D. A. Dikin, and W. K. Liu, "Computational Microstructure Characterization and Reconstruction for Stochastic Multiscale Material Design," *Computer-Aided Design*, vol. 45, pp. 65-76, Jan 2013.
- [16] D. T. Fullwood, S. R. Niezgoda, and S. R. Kalidindi, "Microstructure reconstructions from 2-point statistics using phase-recovery algorithms," *Acta Materialia*, vol. 56, pp. 942-948, Mar 2008.
- [17] V. Vaithyanathan, C. Wolverton, and L. Q. Chen, "Multiscale modeling of precipitate microstructure evolution," *Physical Review Letters*, vol. 88, Mar 25 2002.
- [18] H. Xu, D. Dikin, C. Burkhart, and W. Chen, "Descriptor-based Methodology for Statistical Characterization and 3D Reconstruction for Polymer Nanocomposites," *Computational Material Science*, 2014, in press.
- [19] J. R. Rodgers, "Materials informatics: Knowledge acquisition for materials design," *Abstracts of Papers of the American Chemical Society*, vol. 226, pp. U302-U303, Sep 2003.
- [20] K. F. Ferris, L. M. Peurrung, and J. Marder, "Materials informatics: Fast track to new materials," *Advanced Materials & Processes*, vol. 165, pp. 50-51, Jan 2007.
- [21] Q. Y. Wei, X. D. Peng, X. G. Liu, and W. D. Xie, "Materials informatics and study on its further development," *Chinese Science Bulletin*, vol. 51, pp. 498-504, Feb 2006.
- [22] X. Ma and N. Zabaras, "Kernel principal component analysis for stochastic input model generation," *Journal of Computational Physics*, vol. 230, pp. 7311-7331, Aug 10 2011.
- [23] M. Mohri, A. Rostamizadeh, and A. Talwalkar, *Foundations of Machine Learning*: The MIT Press, 2012.
- [24] Doreswamy, "A Survey for Data Mining Frame Work for Polymer Matrix Composite Engineering Materials Design Applications," *International Journal of Computational Intelligence Systems*, vol. 1, pp. 313-328, Dec 2008.
- [25] C. Ortiz, O. Eriksson, and M. Klintenberg, "Data mining and accelerated electronic structure theory as a tool in the search for new functional materials," *Computational Materials Science*, vol. 44, pp. 1042-1049, Feb 2009.
- [26] Y. Saad, D. Gao, T. Ngo, S. Bobbitt, J. R. Chelikowsky, and W. Andreoni, "Data mining for materials: Computational experiments with AB compounds," *Physical Review B*, vol. 85, Mar 6 2012.
- [27] E. B. Hunt, J. Martin, and P. J. Stone, *Experiments in Induction*: New York: Academic Press, 1996.
- [28] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*: Belmont, California: Wadsworth Inc, 1984.

- [29] K. Kira and L. A. Rendell, "The Feature-Selection Problem - Traditional Methods and a New Algorithm," *Aaaai-92 Proceedings : Tenth National Conference on Artificial Intelligence*, pp. 129-134, 1992.
- [30] I. Kononenko, "Estimating attributes: Analysis and extensions of Relief," in *In L. De Raedt, & F. Bergadano (Eds.), Machine Learning: ECML-94 (pp. 171–182). Springer Verlag.*, 1994.
- [31] M. Robnik Sikonja and I. Kononenko, "An adaptation of relief for attribute estimation in regression," in *In D. H. Fisher (Ed.), Machine Learning: Proceedings of the Fourteenth International Conference (ICML'97) (pp. 296–304). Morgan Kaufmann.*, 1997.
- [32] C. L. Y. Yeong and S. Torquato, "Reconstructing random media. II. Three-dimensional media from two-dimensionmal cuts," *Physical Review E*, vol. 58, pp. 224-233, Jul 1998.
- [33] A. D. Rollett, S. B. Lee, R. Campman, and G. S. Rohrer, "Three-dimensional characterization of microstructure by electron back-scatter diffraction," *Annual Review of Materials Research*, vol. 37, pp. 627-658, 2007.
- [34] A. Jean, D. Jeulin, S. Forest, S. Cantournet, and F. N'Guyen, "A multiscale microstructure model of carbon black distribution in rubber," *Journal of Microscopy*, vol. 241, pp. 243-260, Mar 2011.
- [35] S. Torquato, "Optimal Design of Heterogeneous Materials," *Annual Review of Materials Research*, Vol 40, vol. 40, pp. 101-129, 2010.
- [36] S. Yang, A. Tewari, and A. M. Gokhale, "Modeling of non-uniform spatial arrangement of fibers in a ceramic matrix composite," *Acta Materialia*, vol. 45, pp. 3059-3069, Jul 1997.
- [37] D. S. Li, M. A. Tschopp, M. Khaleel, and X. Sun, "Comparison of reconstructed spatial microstructure images using different statistical descriptors," *Computational Materials Science*, vol. 51, pp. 437-444, Jan 2012.
- [38] A. Tewari and A. M. Gokhale, "Nearest-neighbor distances between particles of finite size in three-dimensional uniform random microstructures," *Materials Science and Engineering a-Structural Materials Properties Microstructure and Processing*, vol. 385, pp. 332-341, Nov 15 2004.
- [39] M. Steinzig and F. Harlow, "Probability distribution function evolution for binary alloy solidification," *Solidification 1999*, pp. 197-206, 1999.
- [40] M. Thomas, N. Boyard, L. Perez, Y. Jarny, and D. Delaunay, "Representative volume element of anisotropic unidirectional carbon-epoxy composite with high-fibre volume fraction," *Composites Science and Technology*, vol. 68, pp. 3184-3192, Dec 2008.
- [41] S. Holotescu and F. D. Stoian, "Prediction of particle size distribution effects on thermal conductivity of particulate composites," *Materialwissenschaft Und Werkstofftechnik*, vol. 42, pp. 379-385, May 2011.
- [42] C. Klaysom, S. H. Moon, B. P. Ladewig, G. Q. M. Lu, and L. Z. Wang, "The effects of aspect ratio of inorganic fillers on the structure and property of composite ion-exchange membranes," *Journal of Colloid and Interface Science*, vol. 363, pp. 431-439, Nov 15 2011.
- [43] J. Gruber, A. D. Rollett, and G. S. Rohrer, "Misorientation texture development during grain growth. Part II: Theory," *Acta Materialia*, vol. 58, pp. 14-19, Jan 2010.
- [44] V. V. Ganesh and N. Chawla, "Effect of particle orientation anisotropy on the tensile behavior of metal matrix composites: experiments and micro structure-based simulation," *Materials Science and Engineering a-Structural Materials Properties Microstructure and Processing*, vol. 391, pp. 342-353, Jan 25 2005.
- [45] B. Kenney, M. Valdmanis, C. Baker, J. G. Pharoah, and K. Karan, "Computation of TPB length, surface area and pore size from numerical reconstruction of composite solid oxide fuel cell electrodes," *Journal of Power Sources*, vol. 189, pp. 1051-1059, Apr 15 2009.
- [46] I. A. Morozov, B. Lauke, and G. Heinrich, "A Novel Method of Quantitative Characterization of Filled Rubber Structures by AFM," *Kgk-Kautschuk Gummi Kunststoffe*, vol. 64, pp. 24-27, Jan-Feb 2011.
- [47] C. P. Prakash, V. D. Mytri, and P. S. Hiremath, "Classification of Cast Iron Based on Graphite Grain Morphology using Neural Network Approach," *Second International Conference on Digital Image Processing*, vol. 7546, 2010.
- [48] M. Ostoja-Starzewski, "Material spatial randomness: From statistical to representative volume element," *Probabilistic Engineering Mechanics*, vol. 21, pp. 112-132, Apr 2006.
- [49] H. Deng, Y. Liu, D. Gai, D. A. Dikin, K. Putz, W. Chen, et al., "Utilizing Real and Statistically Reconstructed Microstructures for the Viscoelastic Modeling of Polymer Nanocomposites," *Submitted to Composite Science and Technology*, 2012.
- [50] H. Y. Xu, M. S. Greene, H. Deng, D. Dikin, C. Brinson, W. K. Liu, et al., "Stochastic Reassembly Strategy for Managing Information Complexity in Heterogeneous Materials Analysis and Design," *Journal of Mechanical Design*, vol. 135, Oct 2013.
- [51] W. Zheng and S. C. Wong, "Electrical conductivity and dielectric properties of PMMA/expanded graphite composites," *Composites Science and Technology*, vol. 63, pp. 225-235, 2003.
- [52] E. Cuthill and J. McKee, "Reducing the bandwidth of sparse symmetric matrices," in *Proceedings of the 1969 24th national conference. ACM*, 1969.
- [53] R. Jin, W. Chen, and T. W. Simpson, "Comparative studies of metamodeling techniques under multiple modelling criteria," *Structural and Multidisciplinary Optimization*, vol. 23, pp. 1-13, Dec 2001.