

RESEARCH

Open Access



Versatile algorithms for the computation of 2-point spatial correlations in quantifying material structure

Ahmet Cecen¹, Tony Fast² and Surya R. Kalidindi^{1,2*}

* Correspondence:

surya.kalidindi@me.gatech.edu

¹School of Computational Science & Engineering Georgia Institute of Technology, Atlanta, GA 30332, USA

²George W. Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA

Abstract

This paper presents a generalized framework along with the associated computational strategies for a rigorous quantification of the material structure in a range of different applications using the framework of 2-point spatial correlations. In particular, we focus on applications requiring different assumptions about the periodicity and/or involving irregular domain shapes and potentially extremely large datasets. Important details of the computational algorithms needed to address these challenges are developed and illustrated with example case studies. Algorithms developed and presented in this work are available at <http://dx.doi.org/10.5281/zenodo.31329>.

Background

Almost all materials enabling advanced technologies exhibit a richness of hierarchical internal structures at multiple length scales (spanning from atomic to macroscale). Certain salient features of this structure control the performance characteristics of interest for a selected application. Although there is often some intuition about what these salient features might be, validated and automated protocols do not yet exist for reliably identifying these features. Further, efficient computational protocols do not yet exist for tracking their evolution during the various unit processing/synthesis steps employed in the industrial manufacturing of new products/devices. In fact, the modulation of the material structure in order to improve the performance of engineering components is often the main motivation behind all activities in the field of materials science and engineering. Despite its important role, a unified computational framework for the quantification of the material hierarchical structure does not exist currently.

Conventional practices for the quantification of the material microstructure have largely relied on accumulated legacy knowledge by domain experts and intuition. Some examples of such microstructure measures include volume fraction, average particle/grain size, average particle/fiber spacing (mean free path), tortuosity and coordination number [1–12]. However, it is easily seen that this simple set of microstructure measures is unlikely to be the best possible set or even an adequate set, because it is easy to imagine multiple instantiations of microstructures that would exhibit the same values of these simple microstructure measures while displaying vastly different values

of macroscale properties of interest. This is particularly true when establishing structure–property linkages for defect-sensitive, potentially anisotropic, macroscale properties.

In recent papers, Niezgodka et al. [13, 14] presented a rigorous theoretical framework for the stochastic quantification of the material structure at any selected length/structure scale, building on the established concepts of spatial correlation functions [15–26]. Although a number of different measures of the spatial correlations in the microstructure are possible (e.g., lineal path functions [27–30] and radial distribution [30–33] functions), only the n -point spatial correlations (or n -point statistics) [15, 16, 18, 19, 30, 34–39] provide the most complete set of measures that are naturally organized by increasing amounts of structure information. For example, the most basic of the n -point statistics are the 1-point statistics, and they reflect the probability of finding a specific local state of interest at any randomly selected single point (or voxel) in the material structure. In other words, they essentially capture the information on volume fractions of the various distinct local states present in the material system. The next higher level of structure information is contained in the 2-point statistics, which capture the probability of finding specified local states h and h' at the tail and head, respectively, of a prescribed vector r randomly placed into the material structure. It should be noted that there is a tremendous leap in the amount of structure information contained in the 2-point statistics compared to the 1-point statistics. It should also be noted that if the 2-point statistics described above are expressed only as a function of the distance between the two points (i.e., r is treated as a scalar instead of a vector), one recovers the radial distribution functions or the pair correlation functions that have been used extensively in prior literature [21, 30, 40]. Higher-order correlations (3-point and higher) are defined in a completely analogous manner.

It is emphasized that the n -point spatial correlations provide statistical information on the microstructure. For example, 2-point statistics provide the expected (i.e., the average) value of a selected correlation between two points separated by a specified vector. However, they also contain information on the variance in the 1-point statistics [39]. In some special cases, they can provide readily interpretable information such as the average shape of the particle (i.e., a mesoscale constituent), especially when the particles have a dominant shape and orientation. On the other hand, when the distribution of the particle shape and orientation is completely random, the corresponding correlations are indeed isotropic (i.e., do not reveal the particle shape directly). The connections between the n -point statistics and the more traditional measures of microstructure have been detailed in prior literature [30, 39].

The n -point statistics described above are most efficiently computed on digital datasets using fast Fourier transform (FFT) techniques [13, 35, 36, 41]. An implicit benefit of treating the material structure function as a stochastic process is that it allows a rigorous quantification of the associated variance [13, 14]. A second important benefit of the spatial correlations described here is that they lend themselves to objective, low-dimensional, high-value representations (using techniques such as principal component analysis (PCA)) [13, 14, 37, 42–44].

The strongest support for the choice of n -point spatial correlations as the most appropriate measures of material structure comes from the pioneering work of Kroner [45], who has taught us that the effective properties of composite material systems can be conveniently expressed as a series sum with the structure details entering this series

explicitly in the form of n -point spatial correlations. These composite theories have been generalized to a broad range of materials phenomena and have been summarized in several books [30, 39, 46]. There are also several reports in literature, where they have been successfully applied to estimate effective properties (both linear and nonlinear) of a broad range of materials with complex structures [47–52]. Physically, the n -point spatial correlations are very effective in rigorously quantifying the local neighbourhoods in the complex internal structure of most advanced materials. Since the local neighbourhoods control the local response, it is only logical that the n -point spatial correlations are the ideal measures of the material structure in formulating process–structure–property (PSP) linkages of interest in designing high performance engineering components. In recent work, the spatial correlations have been used successfully to establish reliable low-cost surrogate models for capturing the materials core knowledge in the form of process–structure–property linkages [19, 37, 43, 44, 53, 54].

In this work, we focus exclusively on the computations of 2-point spatial correlations, but the concepts presented can be expanded trivially for the computation of higher-order statistics. Much of the prior work on the computations of the 2-point spatial correlations has focused on fairly simple microstructures described on rectangular parallelepiped domains that were uniformly tessellated into cuboids (also referred as pixels or voxels). In these earlier applications, the microstructure domains were mainly assumed to be periodic to take advantage of the computational efficiency of discrete Fourier transforms (DFTs). Furthermore, most computations were demonstrated on relatively small domain sizes. In this paper, we present new enhancements that facilitate the computation of the 2-point spatial correlations in a much broader range of applications. In particular we focus on three challenges: (i) avoiding the need to invoke periodicity while still using DFTs, (ii) application to irregular domains, and (iii) application to extremely large datasets.

Methods: Discretized microstructure function and spatial correlations

A microstructure function expresses spatially resolved material structure information gathered from any source, either experiments or simulations. Conceptually, one can think of the microstructure function as $h(x)$, where h denotes the local state occupying the spatial position x . In this notation, the local state refers to any combination of attributes used to define the material locally (e.g., a combination of elemental composition, phase identifier, crystal lattice orientation, and dislocation density may be used to define the local state in multiphase polycrystalline materials at the mesoscale). Brief reflection will expose the unwieldy nature of such a description, especially when one tries to include a diverse set of local state attributes over multiple hierarchical length scales. In an effort to overcome this challenge, the concept of a stochastic microstructure function was introduced [15]. In this novel concept, the microstructure function is defined as $m(h, x)$, where m denotes the probability density associated with finding the local state h at the spatial position x . Consequently, $m(h, x)dhdx$ captures the corresponding probability measure.

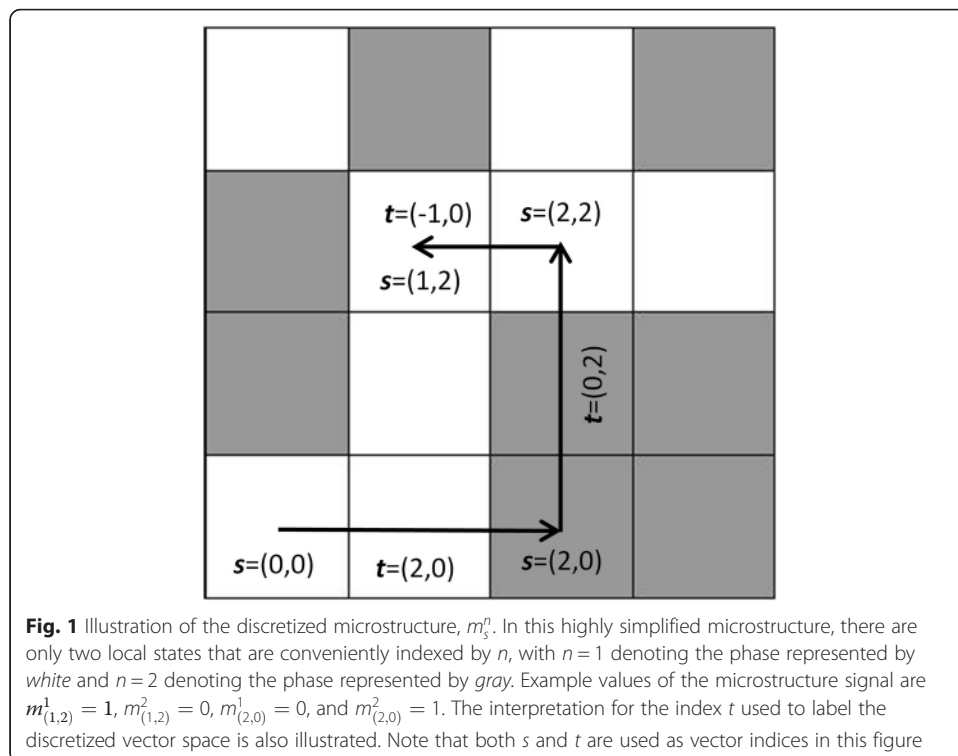
Our interest in this paper, however, rests solely on digital description of the microstructure. Although it is theoretically possible to extract a digital representation of the microstructure function using a multitude of choices in the selection of the basis functions for both the spatial and local state variables [55, 56], we focus our attention here

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

$$m(h, x)dhdx \approx \sum_n \sum_s m_s^n \chi_s(x) \chi^n(h) \tag{1}$$

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

Because of the absence of a natural origin from where one might start indexing the spatial bins, only the relative placement of local states in the material structure contains meaningful information. In other words, only the spatial correlations in the material structure contain high value information. As mentioned earlier, an extensible framework for rigorous quantification of spatial correlations in the material structure is available in the



form of n -point spatial correlations (or n -point statistics) [16, 19, 30, 34–37]. The first-order information on the spatial statistics is actually contained in the 2-point spatial correlations (recall that 1-point statistics capture only the volume fractions) defined as [39]

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

As noted earlier, the 2-point spatial correlation function, $f(h, h' | r)$, reflects the probability density associated with finding local states h and h' at the tail and head, respectively, of a randomly placed vector (includes both a magnitude and a direction) r in the material internal structure. Because the vector r carries both the magnitude and direction in this definition, the spatial correlation function defined in Eq. (2) is directionally resolved. As one can imagine, it is possible to average the statistics over the direction and use $f(h, h' || r)$ instead, where $|r|$ denotes the magnitude of the vector. Indeed, $f(h, h' || r)$ are generally referred to as the pair correlation functions or the radial distributions and contain significantly less spatial information compared to $f(h, h' | r)$. In Eq. (2), $\Omega(r)$ denotes the volumetric domain of the material internal structure analyzed, with $|\Omega(r)|$ denoting the measure of the corresponding volume. It is important to note the dependence of the volumetric domain on the vector itself. This is because material structures studied often have finite domains (except when periodicity is invoked) and the domain available for evaluating the 2-point spatial correlation defined in Eq. (2) depends on the vector r . This is because only those points where it is possible to evaluate both $m(h, x)$ and $m(h', x + r)$ can be included in the evaluation of Eq. (2). As one might imagine, there are certain regions near the boundaries of a given microstructure image where this condition is not met (i.e., either x or $x + r$ fall outside the given image) and therefore the region available for use in Eq. (2) should be expected to show a strong dependence on r (to be discussed in more detail later).

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

$$f(h, h' | r) dh dh' \approx \sum_p \sum_n \sum_t f_t^{np} \chi_t(r) \chi^n(h) \chi^p(h') \tag{3}$$

It is important to recognize that the index t in Eq. (3) effectively bins the vector space associated with r as illustrated in Fig. 1. Starting with the above notions, one can establish the desired relationship between the digital representations of microstructure and the (directionally resolved) 2-point spatial correlations as [35, 36]

$$f_t^{np} = \frac{1}{S_t} \sum_{s=1}^{S_t} m_s^n m_{s+t}^p \tag{4}$$

where S_t captures the r -dependence of $\Omega(r)$ (see Eq. (2)). It is important to recognize that the denominator S_t in Eq. (4) is essentially the total number of trials conducted (where each trial denotes checking what local states exist in spatial bins marked s and $s + t$) and the numerator $\sum_{s=1}^{S_t} m_s^n m_{s+t}^p$ in Eq. (4) denotes an expected measure of total success in these trials (i.e., actually finding the selected local states n and p at the two

bins, respectively). Recognizing this feature of Eq. (4) allows one to make any needed corrections for different situations (will be expanded in later sections).

The computation of f_t^{np} for a specified combination of n and p , essentially requires $O(S^2)$ (i.e., of the order of S^2) computations ($O(S)$ for each value of t and there are $O(S)$ different values of t). Such calculations are generally very expensive and are not easily scalable for datasets with high values of S . In recent years, it has been demonstrated that the angularly resolved n -point statistics computations can be accomplished at $O(\text{Slog } S)$ by employing discrete Fourier transforms (DFTs) [35, 36] (which allow the use of fast Fourier transform (FFT) algorithms) and invoking the convolution theorem. One of the main benefits of these computational schemes is their excellent scalability to large datasets.

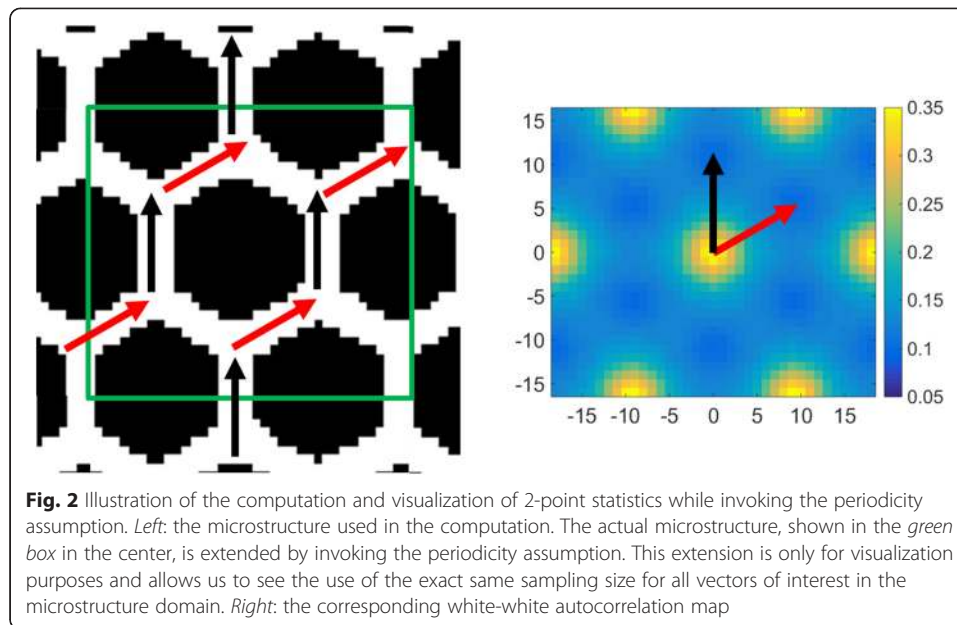
In prior work, the protocols described above have been successfully applied to multi-phase composite systems [13, 14, 19, 43, 44, 57, 58], atomistic datasets [59, 60], and polycrystalline microstructures [42, 61]. However, in all of these applications, the microstructure domains had a simple overall shape (rectangles in 2-D and rectangular parallelepipeds in 3-D), periodicity was generally imposed to take advantage of FFT algorithms, and the studies used relatively small domains. In this work, we present major enhancements to the current protocols that are designed to address these challenges.

As noted earlier, FFT algorithms are central to scalable computation of 2-point statistics. However, they implicitly assume that the microstructure being studied is periodic in all directions (i.e., it can be extended by simply repeating the entire domain as many times as needed). With the assumption of periodicity, S_t in Eq. (4) can be taken to be the same as S (the total number of spatial bins in the microstructure). This is because every spatial bin in the microstructure can be used to place the tail (or equivalently the head) of the vector in evaluating the 2-point statistics. Furthermore, one can simply use the properties of DFTs to compute f_t^{np} . This is because Eq. (4), with the assumption of periodicity, translates to the following in the DFT space via the convolution theorem:

$$F_k^{np} = \frac{1}{S} M_k^{n*} \odot M_k^p, \quad F_k^{np} = \mathfrak{F}(f_t^{np}), \quad M_k^n = \mathfrak{F}(m_s^n) \quad (5)$$

where \odot is the element-wise product operator (also known as Hadamard or Schur product). Throughout this paper, superscript $*$ will denote the complex conjugate and $\mathfrak{F}()$ denotes the DFT transformation of the data to the frequency space enumerated by k (in the context of this paper, this is the spatial frequency space). As a result of Eq. (5), the computation of the 2-point statistics is reduced to computing the DFT of m_s^n , performing requisite products in the frequency space (where they are fully uncoupled), and performing an inverse DFT. For plotting the 2-point statistics, the most intuitive visualizations of 2-pt. statistics would result if $t = 0$ lies in the center of plot. This shift is accomplished trivially by making use of the periodicity implied in the DFT-based computations.

Figure 2 illustrates the above concepts through a simple “honeycomb” microstructure, where each pixel or voxel is colored either white or black. Since there are two local states, we can potentially compute a total of four different 2-point spatial correlations functions: f_t^{11} , f_t^{12} , f_t^{21} , and f_t^{22} , where $n = 1$ refers to the white-colored phase and $n = 2$ refers to the black-colored phase in Fig. 2. Exploiting the known properties of



DFTs, Niezgodą et al. [35] have demonstrated that the number of independent 2-point spatial correlations defined in Eq. (5) is only $H - 1$, where H is the total number of distinct local states present in the material system of interest. Consequently, for two-phase microstructures studied here, we generally need to compute only one of the autocorrelations. Figure 2 shows a plot of white-white autocorrelation.

The autocorrelations presented in Fig. 2 capture a number of salient features of the microstructure. The hexagonal symmetry, the feature shape, and the feature spacing are readily apparent. Furthermore, the periodicity implied in the use of DFTs resulted in the autocorrelations also exhibiting the same periodicity. Note also that the autocorrelation for the zero vector (at the center of the plot) provides the phase volume fraction.

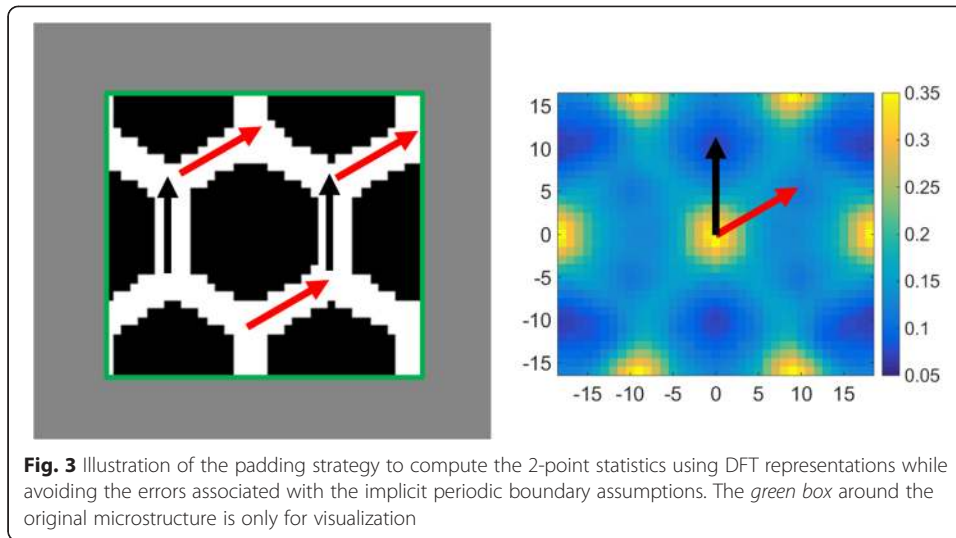
An important consequence of invoking periodicity assumptions is that the number of trials for all vectors is exactly the same and is equal to the number of pixels or voxels in the microstructure studied. In other words, all vectors of interest have been sampled fairly.

Results and Discussion

Application to non-periodic microstructures

As a specific example, we will revisit the same structure illustrated earlier, but without invoking the assumption of periodicity. In other words, our interest is to compute the autocorrelations as defined in Eq. (4), while accounting for the fact that $S_t \neq S$. However, as stated earlier, a direct implementation of Eq. (4) would incur $O(S^2)$ computations. A much better computational strategy would result if one borrows a well-established concept from image analysis [62, 63] and “pads” the microstructure such that only long vectors (larger than the vectors of interest in computing the 2-point statistics) can wrap around from one edge of the original image to the opposite edge when the periodic assumption is implicitly invoked to take advantage of the computational expediency of the DFTs.

The padding strategy described above is illustrated in Fig. 3. Let $S = (S_1, S_2)$ denote the number of spatial bins in the original two-dimensional microstructure



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

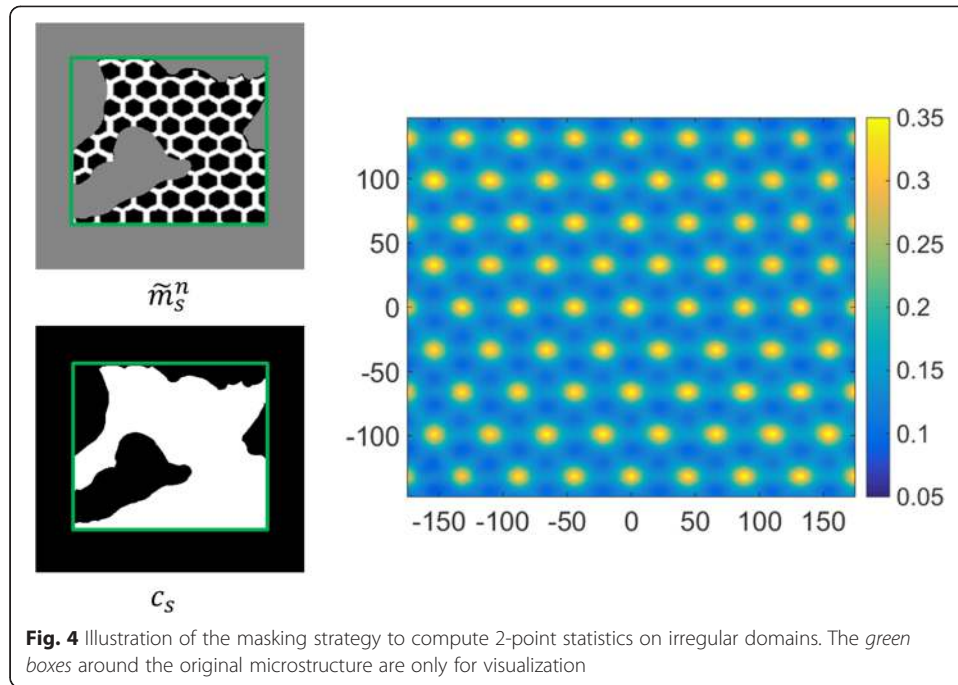
With the padded microstructures, we are now in a position to take advantage of DFTs. Following Eq. (5), we can first compute $\tilde{M}_k^n = \mathfrak{F}(\tilde{m}_s^n)$, and then $\mathfrak{F}^{-1}(\tilde{M}_k^{n*} \odot \tilde{M}_k^p)$, which produces an accurate count of the number of successes in finding local states n and p separated by all vectors $t \leq T$. In fact, the computation described above produces results even for vectors $t > T$, but these results are corrupted by vectors wrapping around the padded region because of the periodicity assumption implicit in the DFTs. However, since our interest here is exclusively in $t \leq T$, we will only take these results from the DFT computation described above. In order to compute the 2-point statistics of interest, we simply need to divide these numbers (equivalent to the numerator in Eq. (4)) with a suitable denominator denoting the total number of trials involved, which is expressed simply as $(S_1 - |t_1|)(S_2 - |t_2|)$. It is pointed out that this strategy provides the exact answer we seek, and not an approximation to it. In fact all of the novel strategies presented in this paper provide the exact answers for the problems posed, but have the advantage that they provide these answers at significantly reduced computational cost compared to direct computations. Furthermore, the padding in Fig. 3 is shown such that it equally envelopes all sides of the original microstructure. This is just for easy visualization and

interpretation. In reality, any placement of the original microstructure inside the overall padded region (i.e., any unequal distribution of the padding as long as the extended microstructure has the same overall size) will produce identical results for the computed 2-point statistics (this is, once again, a consequence of using DFTs).

Figure 3 depicts a plot of the f_t^{11} (white-white) autocorrelations that are not tainted by the periodicity assumptions implied in the use of DFTs. A comparison of the autocorrelations in Figs. 2 and 3 reveals important consequences of the assumption of periodicity. For example, the hexagonal symmetry is no longer evident in the autocorrelations (see the values corresponding to the black and red vectors shown in these figures). This is mainly because the different vectors are no longer sampled the exact same number of times. Although this may not be as important when one deals with a very large image, it clearly has an effect for the relatively small image shown in Fig. 3. In this simple example, one can easily reconcile the different values of the autocorrelations for the red and black vectors depicted in Fig. 3, by noting that we can indeed place many more red vectors with both endpoints in a white pixel, when compared to the similar placement of the black vectors. It is therefore important to recognize that the assumption of periodicity can indeed influence significantly the computed 2-point statistics, especially when one has a limited number of features in the image. Note that the strategy described above can be applied selectively on any of the bounding planes of the image. In other words, one can decide to invoke periodicity assumption on certain bounding planes and employ the padding strategy described above selectively on the other bounding planes.

Masked microstructure domains

As an extension of the idea described above, we now demonstrate a general concept of “masks” that can be used advantageously in many situations related to computing the 2-point statistics. In fact, the padding strategy described above can be considered as a special case of using masks. As an example, consider the microstructure in Fig. 4 which is essentially an extended version of the same microstructure shown in Figs. 2 and 3. However, certain regions of the microstructure have been masked to hide certain irregularly shaped regions where the information is either not available or is of inferior quality (in other words, we do not wish to include that information in the computations of the 2-point statistics). As shown in Fig. 4, these masked regions can be on the boundary of the microstructure (e.g., the microstructure is measured in an irregular domain). But they can also be inside the microstructure (e.g., some regions of the micrograph may not be discernable or reliable). As demonstrated earlier, a mask can also be applied to produce a padded region to impose non-periodic boundaries (see Fig. 3). In this situation, it is convenient to define two microstructure functions (see Fig. 4): (i) an extended microstructure function denoted as \tilde{m}_s^n , where we have introduced an additional fictitious local state (i.e., the third phase colored gray in Fig. 4) in the masked region as well as the boundary padded regions and (ii) a mask function denoted as c_s such that it takes a value of zero for spatial bins (shown as black) in the masked regions and one (shown as white) everywhere else. It is pointed out that the extended \tilde{m}_s^n already contains the information in the c_s . However, we choose to carry this information in the redundant manner described above for ease of discussion and computation.



Following the methodology described in the previous example, we compute $\tilde{M}_k^n = \mathfrak{F}(\tilde{m}_s^n)$ and then $\mathfrak{F}^{-1}(\tilde{M}_k^{n*} \odot \tilde{M}_k^p)$ to accurately count the number of successes in finding local states n and p separated by all vectors of interest (as mentioned earlier, it is important to include padding if we wish to avoid the default assumption of periodicity implicit in the use of DFTs). In order to compute the 2-point statistics of interest, we simply need to divide these numbers (equivalent to the numerator in Eq. (4)) with a suitable denominator denoting the total number of trials involved. For the masked microstructures described here, the denominator can be computed easily as $\mathfrak{F}^{-1}(C_k^* \odot C_k)$, where $C_k = \mathfrak{F}(c_s)$. It should be noted once again that the padding scheme described in the previous case study is essentially a special case of the masking protocol described here.

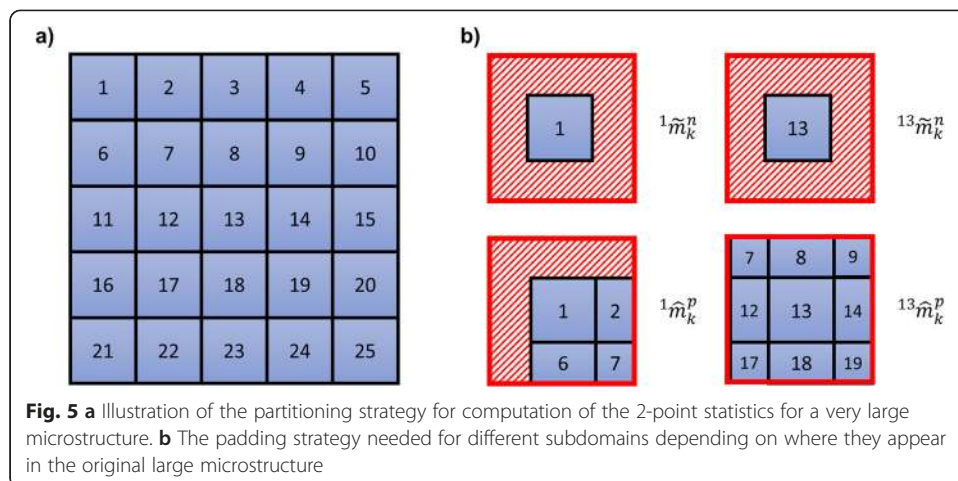
Figure 4 depicts a plot of the f_t^{11} (white-white) autocorrelations where the computations were limited to the unmasked regions (the white region of the mask) using the computationally efficient DFT-based protocols developed and presented in this paper. Furthermore, there was no assumption of periodicity in this computation. However, it is seen that these autocorrelations are indeed very similar to the ones shown in Fig. 2 (performed assuming periodicity and limited to a much smaller range of vectors). This provides unambiguous confirmation that the protocols presented here are doing an excellent job of computing the 2-point statistics for irregular domains without invoking periodicity, while taking full advantage of the computational efficiency of the FFT algorithms.

Large microstructure domains

We have already emphasized the benefits of using FFT algorithms to dramatically reduce the computational time incurred in the calculations of spatial correlations. In this section, we now shift our attention to cases where the datasets are extremely large and present a substantial challenge with their storage requirements. For example, a

microstructure of about $600 \times 600 \times 600$ pixels is likely to prove unwieldy for an average desktop computer, especially since the application of the FFT algorithms would require double precision storage of complex numbers. Consequently, the computation of the non-periodic spatial correlations for a $2000 \times 2000 \times 2000$ voxel dataset can easily demand close to 180 GBs of memory, forcing the use of a supercomputer for such calculations. We address the challenge described above using a strategy that carefully partitions the large domain into smaller subdomains, performs the requisite computations on them, and then assembles correctly the statistics for the original large domain from the computations on the subdomains. Our approach can be compared to various partitioning strategies for efficient computation of convolutions via FFTs that are well known in digital signal processing applications, such as overlap-save, overlap-add, and hybrid schemes [64, 65]. The overall process is illustrated schematically for a 2-D dataset in Fig. 5.

In this specific illustration, the overall domain is broken into 25 subdomains (see Fig. 5a). Let $i m_s^n$ denote the digitized microstructure in the subdomain enumerated by i . The microstructure in each subdomain is then extended by padding in two ways to produce $i \tilde{m}_s^n$ and $i \hat{m}_s^n$ as shown in Fig. 5b for a corner subdomain (labelled 1) and an interior subdomain (labelled 13). The main idea is that $i \tilde{m}_s^n$ is a simply padded version of $i m_s^n$ with the padding size controlled by the largest vector size of interest in the calculation of the 2-point statistics (as we did before for avoiding the assumption of periodicity of the microstructure), while $i \hat{m}_s^n$ is an extended version of $i m_s^n$ that actually captures the real neighborhood information from the original large dataset. As illustrated in Fig. 5b, the treatment for generating $i \hat{m}_s^n$ would have to be somewhat different for interior subdomains versus those that are at the boundary of the original large domain. Furthermore, it is important to ensure that the extensions for both $i \tilde{m}_s^n$ and $i \hat{m}_s^n$ are of the exact same size. Let $i \tilde{M}_k^n$ and $i \hat{M}_k^p$ denote the DFT representations of $i \tilde{m}_s^n$ and $i \hat{m}_s^n$, respectively. Following the ideas presented earlier, it should be clear that $\mathfrak{F}^{-1}(i \tilde{M}_k^{n*} \odot i \hat{M}_k^p)$ will produce an accurate count of the number of successes from the i th subdomain in finding vectors with local states n and p at the tail and the head of the vector, respectively. As before, these counts are only accurate for vectors smaller than the padding size used in $i \tilde{m}_s^n$, which is really our stated interest anyway. Once the



numbers of successes are computed for all of the non-overlapping subdomains, the desired 2-point statistics for the original large domain can be easily recovered using

$$f_t^{np} = \frac{\sum_i \mathfrak{S}^{-1} \left({}^i \tilde{M}_k^{n*} \odot {}^i M_k^p \right)}{\sum_i {}^i S_t} \tag{6}$$

Note that the total number of trials (denominator in Eq. (6)) is actually the same as what we used before in the case of non-periodic domains and is given by $({}^i S_1 - |t_1|)({}^i S_2 - |t_2|)$, where $({}^i S_1, {}^i S_2)$ denote the grid size in the i th subdomain being studied, and (t_1, t_2) denote the components of the vector for which the 2-point statistics are being computed. It is important to also note that the concepts of masking and modification for periodicity/non-periodicity can be combined with this scheme by making suitable adjustments to the algorithm as described in earlier example case studies.

As a demonstration, the scheme is applied to a 3-D (three-dimensional) micro-CT dataset obtained from a sample of reinforced polymer composite. A visualization of the entire dataset and an exemplar subdomain is shown in Fig. 6a, b. For this dataset, we have applied masks on the irregularly shaped overall domain and computed the non-periodic 2-point autocorrelations of the fiber phase. The computed autocorrelations are visualized as 3-D iso-contour surfaces in Fig. 6c. It can be observed that the fibers are predominantly aligned along the xy -plane with a small angular margin confined within a flat, ellipsoid region. There is also visible anisotropy in the in-plane distribution of the fiber orientations. Note that these topological features regarding the placement of

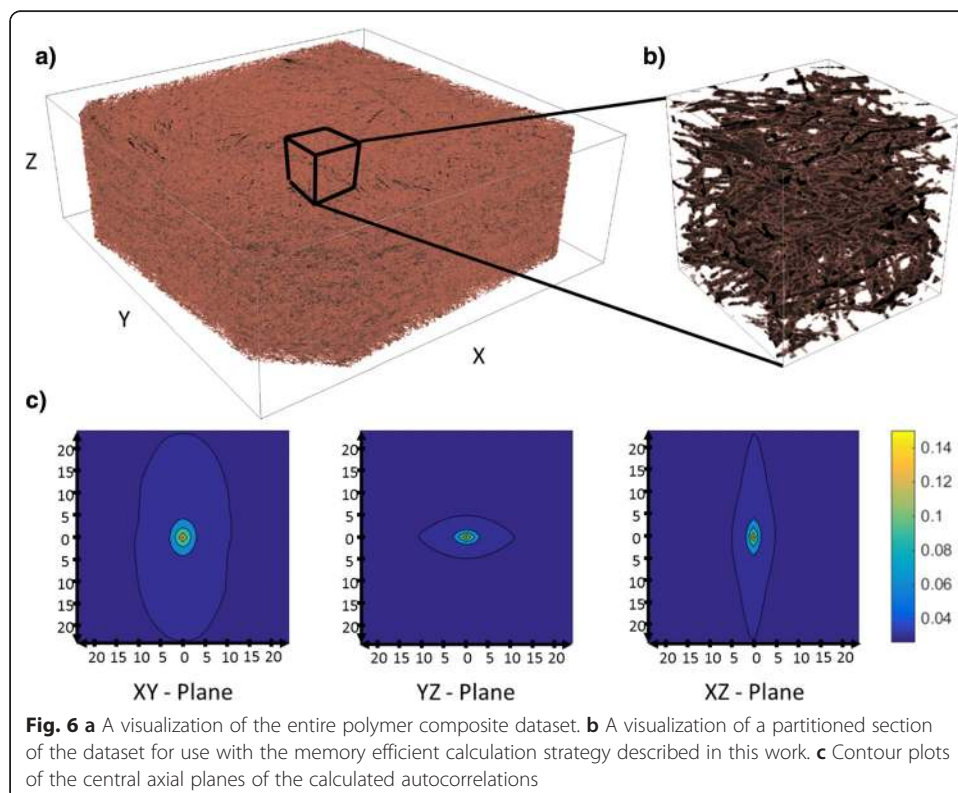


Table 1 Comparison of computation times and memory required for the naive computation and various choices of partition (patch) size for the memory efficient procedure

	Full computation	Patched computation optimized for the following:		
		Desktop	Laptop	Minimum RAM
Memory requirement (GBs)	42	8	4	1.4
Computation time (min)	3	8	14	48

the fibers in the structure cannot easily be inferred from a direct 3-D visualization of the original structure.

It is important to note that suitable trade-offs can be made between the execution speed and the memory usage for the computation on the large microstructure described above. This is accomplished using the partitioning strategy illustrated in Fig. 5. Obviously, using more partitions reduces the memory requirements at the expense of increased overall computation time. Table 1 presents the time and memory cost comparisons for the 2-point autocorrelation calculation for the example dataset, for different partitioning window sizes (i.e., different memory requirements). For this case study, the partitioning window sizes were selected to correspond to commonly available memory choices. For example, at the current time, an average consumer laptop has 4 GBs of DDR3 memory, while an average researcher desktop has 8 GBs of DDR3 memory. All tests were done entirely on a single personal machine utilizing all threads available with an i7-5820K CPU and 48 GBs of DDR4 RAM.

Conclusions

We have presented a rigorous framework for quantification of the material microstructure using directionally resolved 2-point spatial correlations. The use and importance of FFTs for computationally efficient calculation of these spatial correlations have been discussed. Schemes to accommodate non-periodic boundaries, irregular grids, and very large datasets are detailed and demonstrated on simplistic datasets for maximum clarity. Finally, all schemes are simultaneously demonstrated on an experimentally obtained 3D microstructure dataset of very large size displaying an irregular grid with non-periodic boundaries. Algorithms developed and presented in this work are made available at <http://dx.doi.org/10.5281/zenodo.31329>.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

AC and TF developed and implemented one or more of the algorithms described in this work. SK initiated and led this project. All authors read and approved the final submitted manuscript.

Acknowledgements

The authors gratefully acknowledge support from AFOSR award FA9550-12-1-0458.

Received: 7 August 2015 Accepted: 10 December 2015

Published online: 15 January 2016

References

1. Bei H, George EP (2005) Microstructures and mechanical properties of a directionally solidified NiAl–Mo eutectic alloy. *Acta Mater* 53(1):69–77. doi:10.1016/j.actamat.2004.09.003
2. Promentilla MAB, Sugiyama T, Hitomi T, Takeda N (2009) Quantification of tortuosity in hardened cement pastes using synchrotron-based X-ray computed microtomography. *Cement Concrete Res* 39(6):548–557. doi:10.1016/j.cemconres.2009.03.005

3. Holzer L, Münch B, Iwanschitz B, Cantoni M, Hocker T, Graule T (2011) Quantitative relationships between composition, particle size, triple phase boundary length and surface area in nickel-cermet anodes for Solid Oxide Fuel Cells. *J Power Sources* 196(17):7076–7089. doi:10.1016/j.jpowsour.2010.08.006
4. Su H, Gao W, Feng Z, Lu Z (2012) Processing, microstructure and tensile properties of nano-sized Al₂O₃ particle reinforced aluminum matrix composites. *Mater Des* 36:590–596. doi:10.1016/j.matdes.2011.11.064
5. Shearing PR, Howard LE, Jørgensen PS, Brandon NP, Harris SJ (2010) Characterization of the 3-dimensional microstructure of a graphite negative electrode from a Li-ion battery. *Electrochem Commun* 12(3):374–377. doi:10.1016/j.elecom.2009.12.038
6. Van de Lagemaat J, Benkstein KD, Frank AJ (2001) Relation between particle coordination number and porosity in nanoparticle films: implications to dye-sensitized solar cells. *J Phys Chem B* 105(50):12433–12436
7. Cecen A, Wargo E, Hanna A, Turner D, Kalidindi S, Kumbur E (2012) 3-D microstructure analysis of fuel cell materials: spatial distributions of tortuosity, void size and diffusivity. *J Electrochem Soc* 159(3):B299–B307
8. Dimiduk DM, Hazzledine PM, Parthasarathy TA, Seshagiri S, Mendiratta MG (1998) The role of grain size and selected microstructural parameters in strengthening fully lamellar TiAl alloys. *Metallurgical and Materials Transactions A* 29(1):37–47. doi:10.1007/s11661-998-0157-3
9. Baker T, Gave K, Charles J (1976) Inclusion deformation and toughness anisotropy in hot-rolled steels. *Metals Technol* 3(1):183–193
10. Garrison WM Jr, Wojcieszynski AL (2007) A discussion of the effect of inclusion volume fraction on the toughness of steel. *Mater Sci Eng A* 464(1–2):321–329. doi:10.1016/j.msea.2007.02.015
11. Lankford J (1977) (E) Effect of oxide inclusions on fatigue failure. *Int Metals Rev* 22(1):221–228
12. Murakami Y (2002) Effects of Nonmetallic Inclusions on Fatigue Strength. *Metal Fatigue*, 75–127. Elsevier BV, Amsterdam. doi:10.1016/b978-008044064-4/50006-2
13. Niezgoda SR, Yabansu YC, Kalidindi SR (2011) Understanding and Visualizing Microstructure and Microstructure Variance as a Stochastic Process. *Acta Mater* 59:6387–6400
14. Niezgoda SR, Kanjarla AK, Kalidindi SR (2013) Novel microstructure quantification framework for databasing, visualization, and analysis of microstructure data. *Integrating Materials and Manufacturing Innovation* 2:3
15. Adams BL, Gao X, Kalidindi SR (2005) Finite approximations to the second-order properties closure in single phase polycrystals. *Acta Mater* 53(13):3563–3577. doi:10.1016/j.actamat.2005.03.052
16. Brown WF (1955) Solid Mixture Permittivities. *J Chem Phys* 23(8):1514–1517
17. McDowell DL, Choi HJ, Panchal J, Austin R, Allen J, Mistree F (2007) Plasticity-Related Microstructure-Property Relations for Materials Design. *Key Engineering Materials* 340–341:21–30
18. Kalidindi SR (2013) Microstructure Informatics. *Informatics for Materials Science and Engineering*. Elsevier BV. doi:10.1016/b978-0-12-394399-6.00018-7
19. Kalidindi SR (2015) Data science and cyberinfrastructure: critical enablers for accelerated development of hierarchical materials. *Int Mater Rev* 60(3):150–168
20. Torquato S (1991) Random heterogeneous media: microstructure and improved bounds on effective properties. *Applied Mechanics Reviews* 44(2):37–76
21. Gokhale AM, Tewari A, Garmestani H (2005) Constraints on microstructural two-point correlation functions. *Scr Mater* 53:989–993
22. Saheli G, Garmestani H, Gokhale A (2004) Effective Elastic Properties of an Al-SiC Composite Using Two-point Statistical Mechanics Approach. In: NUMIFORM, Ohio State University, Columbus, OH, June 10, 2004. AIP Conference Proceedings., pp 355–359
23. Baniassadi M, Mortazavi B, Hamedani HA, Garmestani H, Ahzi S, Fathi-Torbaghan M, Ruch D, Khaleel M (2012) Three-dimensional reconstruction and homogenization of heterogeneous materials using statistical correlation functions and FEM. *Computational Materials Science* 51(1):372–379. doi:10.1016/j.commatsci.2011.08.001
24. Baniassadi M, Garmestani H, Li DS, Ahzi S, Khaleel M, Sun X (2011) Three-phase solid oxide fuel cell anode microstructure realization using two-point correlation functions. *Acta Mater* 59(1):30–43. doi:10.1016/j.actamat.2010.08.012
25. Li DS, Saheli G, Khaleel M, Garmestani H (2006) Quantitative prediction of effective conductivity in anisotropic heterogeneous media using two-point correlation functions. *Computational Materials Science* 38(1):45–50. doi:10.1016/j.commatsci.2006.01.004
26. Lee PS, Piehler HR, Rollett AD, Adams BL (2002) Texture clustering and long-range disorientation representation methods: application to 6022 aluminum sheet. *Metallurgical and Materials Transactions A*, 33(12), 3709–3718.
27. Yeong C, Torquato S (1998) Reconstructing random media. *Phys Rev E* 57(1):495
28. Schröder J, Balzani D, Brands D (2011) Approximation of random microstructures by periodic statistically similar representative volume elements based on lineal-path functions. *Archive of Applied Mechanics* 81(7):975–997
29. Singh H, Gokhale A, Lieberman S, Tamirisakandala S (2008) Image based computations of lineal path probability distributions for microstructure representation. *Mater Sci Eng A* 474(1):104–111
30. Torquato S (2002) *Random Heterogeneous Materials*. Interdisciplinary Applied Mathematics, 16th edn. Springer-Verlag, New York. doi:10.1007/978-1-4757-6355-3
31. Li G, Liang Y, Zhu Z, Liu C (2003) Microstructural analysis of the radial distribution function for liquid and amorphous Al. *J Phys Condens Matter* 15(14):2259
32. Park SH, Lee DN (1988) A study on the microstructure and phase transformation of electroless nickel deposits. *J Mater Sci* 23(5):1643–1654
33. Rollett AD, Lebensohn RA, Groeber M, Choi Y, Li J, Rohrer GS (2010) Stress hot spots in viscoplastic deformation of polycrystals. *Modelling and Simulation in Materials Science and Engineering* 18(7):Artn 074005. doi:10.1088/0965-0393/18/7/074005
34. Fullwood DT, Niezgoda SR, Adams BL, Kalidindi SR (2010) Microstructure sensitive design for performance optimization. *Progress in Materials Science* 55(6):477–562. doi:10.1016/j.pmatsci.2009.08.002
35. Niezgoda SR, Fullwood DT, Kalidindi SR (2008) Delineation of the space of 2-point correlations in a composite material system. *Acta Mater* 56(18):5285–5292. doi:10.1016/j.actamat.2008.07.005

36. Fullwood DT, Niezgoda SR, Kalidindi SR (2008) Microstructure reconstructions from 2-point statistics using phase-recovery algorithms. *Acta Mater* 56(5):942–948. doi:10.1016/j.actamat.2007.10.044
37. Kalidindi SR, Niezgoda SR, Salem AA (2011) Microstructure informatics using higher-order statistics and efficient data-mining protocols. *JOM* 63(4):34–41
38. Niezgoda SR, Turner DM, Fullwood DT, Kalidindi SR (2010) Optimized structure based representative volume element sets reflecting the ensemble-averaged 2-point statistics. *Acta Mater* 58(13):4432–4445. doi:10.1016/j.actamat.2010.04.041
39. Adams BL, Kalidindi S, Fullwood DT (2013) Microstructure-sensitive design for performance optimization. Butterworth-Heinemann, Waltham
40. Debye P, Anderson HR, Brumberger H (1957) Scattering by an Inhomogeneous Solid. II. The Correlation Function and Its Application. *J Appl Phys* 28(6):679–683. doi:10.1063/1.1722830
41. Fullwood DT, Kalidindi SR, Niezgoda SR, Fast A, Hampson N (2008) Gradient-based microstructure reconstructions from distributions using fast Fourier transforms. *Materials Science and Engineering a-Structural Materials Properties Microstructure and Processing* 494(1–2):68–72. doi:10.1016/j.msea.2007.10.087
42. Qidwai SM, Turner DM, Niezgoda SR, Lewis AC, Geltmacher AB, Rowenhorst DJ, Kalidindi SR (2012) Estimating response of polycrystalline materials using sets of weighted statistical volume elements (WSVEs). *Acta Mater* 60:5284–5299
43. Gupta A, Cecen A, Goyal S, Singh AK, Kalidindi SR (2015) Structure–property Linkages for Non-Metallic Inclusions/Steel Composite System using a Data Science Approach. *Acta Mater* 91:239–254
44. CeCen A, Fast T, Kumbur EC, Kalidindi SR (2014) A Data-driven Approach to Establishing Microstructure-Property Relationships in Porous Transport Layers of Polymer Electrolyte Fuel Cells. *J Power Sources* 245:144–153
45. Kröner E (1986) Statistical Modelling. In: Gittus J, Zarka J (eds) *Modelling Small Deformations of Polycrystals*. Springer, Netherlands, pp 229–291. doi:10.1007/978-94-009-4181-6_8
46. Milton GW (2002) *The theory of composites*. Cambridge monographs on applied and computational mathematics. Cambridge university press, Cambridge
47. Fullwood DT, Adams BL, Kalidindi SR (2008) A strong contrast homogenization formulation for multi-phase anisotropic materials. *J Mech Phys Solids* 56(6):2287–2297. doi:10.1016/j.jmps.2008.01.003
48. Saheli G, Garmestani H, Adams BL (2004) Microstructure design of a two phase composite using two-point correlation functions. *J Computer-Aided Materials Design* 11(2–3):103–115
49. Garmestani H, Lin S, Adams BL, Ahzi S (2001) Statistical Continuum Theory for Large Plastic Deformation of Polycrystalline Materials. *J Mech Phys Solids* 49(3):589–607
50. Mason TA, Adams BL (1999) Use of microstructural statistics in predicting polycrystalline material properties. *Metallurgical and Materials Transactions a-Physical Metallurgy and Materials Science* 30(4):969–979. doi:10.1007/s11661-999-0150-5
51. Adams BL, Olson T (1998) The mesostructure–properties linkage in polycrystals. *Progress Mater Sci* 43(1):1–87
52. Beran MJ, Mason TA, Adams BL, Olsen T (1996) Bounding elastic constants of an orthotropic polycrystal using measurements of the microstructure. *J Mech Physics Solids* 44(9):1543–1563
53. Binci M, Fullwood D, Kalidindi SR (2008) A new spectral framework for establishing localization relationships for elastic behavior of composites and their calibration to finite-element models. *Acta Mater* 56(10):2272–2282. doi:10.1016/j.actamat.2008.01.017
54. Kalidindi SR, Binci M, Fullwood D, Adams BL (2006) Elastic properties closures using second-order homogenization theories: Case studies in composites of two isotropic constituents. *Acta Mater* 54(11):3117–3126. doi:10.1016/j.actamat.2006.03.005
55. Yabansu YC, Patel DK, Kalidindi SR (2014) Calibrated localization relationships for elastic response of polycrystalline aggregates. *Acta Mater* 81:151–160
56. Yabansu YC, Kalidindi SR (2015) Representation and calibration of elastic localization kernels for a broad class of cubic polycrystals. *Acta Mater* 94:26–35
57. Wargo EA, Hanna AC, Cecen A, Kalidindi SR, Kumbur EC (2012) Selection of Representative Volume Elements for Pore-Scale Analysis of Transport in Fuel Cell Materials. *J Power Sources* 197:168–179
58. Steinmetz P, Yabansu YC, Hötzer J, Jainta M, Nestler B, Kalidindi SR (2016) Analytics for microstructure datasets produced by phase-field simulations. *Acta Mater* 103:192–203. doi:10.1016/j.actamat.2015.09.047
59. Kalidindi SR, Gomberg JA, Trautt ZT, Becker CA (2015). Application of data science tools to quantify and distinguish between structures and models in molecular dynamics datasets. *Nanotechnology*, 26(34), 344006.
60. Dong X, McDowell DL, Kalidindi SR, Jacob KI (2014) Dependence of mechanical properties on crystal orientation of semi-crystalline polyethylene structures. *Polymer* 55(16):4248–4257. doi:10.1016/j.polymer.2014.03.045
61. Salem AA, Shaffer JB, Satko DP, Semiatin SL, Kalidindi SR (2014) Workflow for integrating mesoscale heterogeneities in materials structure with process simulation of titanium alloys. *Integrating Materials and Manufacturing Innovation* 3(1):1–22. doi:10.1186/s40192-014-0024-6
62. Gonzalez RC, Woods RE (2008) *Digital Image Processing*. Pearson Prentice Hall, Upper Saddle River, NJ.
63. Pratt WK (2006) *Linear Processing Techniques*. In: *Digital Image Processing*. John Wiley & Sons, Inc., pp 217–244. doi:10.1002/9780470097434.ch9
64. Oppenheim AV, Schaffer RW, Buck JR (1989) *Discrete-time signal processing*, vol 2. Prentice-hall, Englewood Cliffs
65. Svoboda D (2011) Efficient computation of convolution of huge images. *Image Analysis and Processing–ICIAP 2011*. Springer, Berlin, pp 453–462 doi:10.1007/978-3-642-24085-0_47