



## Missing information and data fidelity in digital microstructure acquisition

**Bao, Chongbing; Ji, Chuanyi; Poulsen, Henning Friis; Li, Mo**

*Published in:*  
Acta Materialia

*Link to article, DOI:*  
[10.1016/j.actamat.2019.05.012](https://doi.org/10.1016/j.actamat.2019.05.012)

*Publication date:*  
2019

*Document Version*  
Peer reviewed version

[Link back to DTU Orbit](#)

*Citation (APA):*  
Bao, C., Ji, C., Poulsen, H. F., & Li, M. (2019). Missing information and data fidelity in digital microstructure acquisition. *Acta Materialia*, 173, 262-269. <https://doi.org/10.1016/j.actamat.2019.05.012>

---

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

## **Missing information and data fidelity in digital microstructure acquisition**

Chongbing Bao<sup>a,b</sup>, Chuanyi Ji<sup>c</sup>, Henning Friis Paulsen<sup>d</sup>, Mo Li<sup>b,e,\*</sup>

<sup>a</sup>School of Mathematics and Statistics, Central South University, Changsha, Hunan, 410083,  
China

<sup>b</sup>State Key Laboratory for Powder Metallurgy, Central South University, Changsha, Hunan,  
410083, China

<sup>c</sup>School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta,  
Georgia 30332, United States

<sup>d</sup>Department of Physics, Technical University of Denmark, Lyngby 2800 kgs, Denmark

<sup>e</sup>School of Materials Science and Engineering, Georgia Institute of Technology, Atlanta, Georgia  
30332, United States

**ABSTRACT:** Microstructure is a pillar supporting the structure-property relations in materials science and engineering. The digitized format referred to as the digital microstructure plays an increasingly vital role in materials genome, new architecture design and advanced manufacturing of functional materials. However, one basic issue facing digital acquisition of microstructures is left untouched, i.e., how much information is missing in the very first step of data gathering and processing. The missing information leads to a series of issues in data fidelity from reconstruction of digital microstructures to prediction of material properties and new material design. Here using a Laguerre-Voronoi model and Xu-Li microstructure characterization method, we define and establish this problem by showing quantitatively the missing information and this impact. Depending on the experimental resolution and the nature of the microstructures, the lost microstructural data become significant. The missing information leads to distortion to the actual microstructure and eventually limits the prediction of material properties.

**KEYWORDS:** **missing information**, microstructure, 3D digital microstructure, structure-property relation

## **Introduction**

Microstructure, along with atomic and electronic structures, plays a fundamental role in the structure-property relations of a vast number of materials, ranging from composites to polycrystals [1]. Microstructure refers to the geometric and topological arrangement made by various aggregates made of crystallites, phases, and inclusions separated by the interfaces [1-3]. For polycrystalline materials (PMs), microstructures manifest as grain boundary networks made of a set of geometric objects, vertex points, triple junction lines, grain boundary areas, and grain volumes or cells. Collectively these microstructure entities contribute to material properties, often with dramatic outcomes. For example, reorienting grain boundaries to perpendicular to the direction of the electrical current flow along the nanoscale wires can reduce the damage from electromigration in Cu and Al interconnects, which results in a significant increase in lifetime of microelectronics [4,5]. The evolution of the grain shapes in the Ni-based superalloy in jet engine turbine, from equiaxed to columnar and finally to single crystal, is another example of microstructure engineering that made wonders in modern aviation [6,7]. However, these revolutionary changes in material performance through manipulating microstructure-property relations are often time consuming and costly. To achieve a desired property, hundreds to thousands of samples need to be made and tested, from which vast amount of microstructure information is acquired and examined to correlate to the property improvement in each cycle of trial-and-error [1-3].

Modern techniques made this process much easier, especially the digitization of microstructures. The patterns in the microstructures are acquired and stored as pixels or voxels [8-10]. Visualization of a digital microstructure (DM) is now a routine, and importantly, one can relate the microstructures to material properties much easier and faster. These modern approaches include electron backscattering diffraction (EBSD) combined with focused ion beam (FIB)

sectioning which can be conducted with commercial equipment. The same can be done with designated research facilities such as synchrotron 3D x-ray diffraction (3DXRD) and scattering [7-11]. Combined with computer software and storage, DM acquisition has become a powerful approach in accelerating materials research and application.

While the current technology enables us to acquire digital microstructures faster, some of the basic and more delicate issues in the data gathering and processing stage has not received due attention. One such question is how much microstructure information we actually acquire, or conversely, how much information we have missed. The question touches upon the fundamental principle regarding data fidelity, especially when an object or a source, in this case the microstructure, is unknown *a priori*. If our objective is just visualize or quickly screen microstructures, as we have done in the past, the information in the currently existing DMs can provide qualitative estimation that is often sufficient. However, as explained below, as the roles of the DMs become increasingly important in material property prediction, design and quantitative analysis for structure-property relations, one must face the issue of data fidelity seriously [12].

There are many sources that contribute to the missing information in digital microstructure acquisition. The first and foremost is *instrumental resolution*. For example, the finest features in the microstructure resolvable under the traditional optical microscopy is about the multiple of the wavelength of lights [2]. In serial sectioning, the scale is also limited by the layer thickness in sectioning. For surface based methods, the penetration depth of electron beam from the surface and the lateral electron beam spot interval used in diffraction sampling on the sample surface are other resolution limiting length scales. These length scales naturally set the limit for microstructure features that can be acquired with confidence. Although 100 nm scale or less can be reached in both EBSD-FIB and synchrotron processes, larger scales on the order of micron or

bigger are often used in practice, for efficiency and low cost [7-11]. The same is seen in synchrotron scattering [13,14].

The second is related to *operation*, which can occur even when the abovementioned instrumental resolution is not an issue. For example, when two neighboring grains happen to have a misorientation that is below certain angle, i.e., about 1 degree for EBSD and 0.2 degree for 3DXRD respectively. The small scale renders the experimental operation difficult for resolving certain microstructure features in time as well as space.

The third is related to the basic *nature of the microstructures*, such as grain size distribution, grain shape and misorientation. Polycrystals have varying grain sizes which typically obey Gaussian and lognormal distributions [1]. For a typical sample with the mean grain size of 1 micron in a lognormal grain size distribution, there are about 50% smaller grains below the mean grain diameter (see below for more details). The small grains are among those potentially lost in digital microstructure acquisition. In general, therefore, depending on the instrument and operation resolutions discussed above, the number of these small grains lost in digital microstructure acquisition can increase dramatically if the effective resolutions are below the mean grain size. We shall demonstrate this analytically as a classic case here.

Undoubtedly, the missing microstructure information in the very first stage of digital microstructure acquisition has impacts on the subsequent data processing as well as applications. Like any other information processing in Big Data, *data fidelity* is a central issue in DM. Proper assessment of missing information provides the needed confidence via a complement null value to the available data in downstream applications. For example, *microstructure reconstruction* is based on the available DM raw data gathered in the first stage. The goodness of the reconstructed digital microstructure depends on the fidelity of the input data. This leads to the question of *error*

*estimation* for the reconstructed digital microstructures. Furthermore, when using these DMs in material design and prediction of material property, the uncertainties in the predicted properties and performance from the reconstructed DMs depend on the information of the input data, which includes the null values of missing information. The uncertainty is mandatory for DMs and their applications but up to date, there are no answers to these questions.

## **Theory and Polycrystal Model**

Here we address these questions by using a model polycrystal sample, the Laguerre-Voronoi (LV) diagram [15]. Our objective is to bring the important issue of missing information to the attention of a broader community including materials science, big data and digital manufacturing through a simple and quantitative example that is still challenging to have in experiments. To obtain the digital microstructures from this sample with a range of resolutions, we design a *simulation of digital microstructure acquisition*, that is, we perform a simulated digital microstructure acquisition and obtain DM on this theoretical polycrystal sample. The digitized LV polycrystal sample obtained under the given resolution in the simulation allows us to examine quantitatively the changes in the microstructure features of the DMs. We find that depending on the imposed “experimental” resolution and the nature of the microstructures, the lost microstructural features such as the number of grains can become significant. More importantly, we capture the missing information through quantitative characterization of the geometric and topological properties of the microstructures that are still impossible to obtain at present time through experiment-based approaches. These missing features can have a series of effects on the reconstructed DMs and microstructure related properties predicted from these partial DM data sets.

The choice of the Laguerre-Voronoi diagram is for two considerations. One is its close resemblance to real microstructure features [13]. LV models have been used extensively for

polycrystalline microstructures. Their rich and complex topological features and relations allow us to perform quantitative microstructure characterization and analysis, as well as develop numerical algorithms relevant to real DMs. The other is that LV model provides a *perfect microstructure that has no missing components*. Therefore, the “ground truth” is provided by such a synthetic microstructural model. By comparing the “ground truth” with the digitized microstructures through the simulation, we can define and determine quantitatively the missing microstructural entities in the digitization of the polycrystals as we vary the imposed resolution. The justification for using the theoretical polycrystal sample also comes from the fact that the experimental DM data set that is available to date cannot be qualified as the “ground truth”. The reason is not because the data set may (or may not) contain missing microstructure information but because we simply do not know whether and what information is missing. Such an unknown in the existing experimental data sets is another reason that motivates this work.

The Laguerre-Voronoi diagram is generated by distributing  $N$  number of disks on a plane that follow a certain distribution for the disk diameters. Here we use a lognormal distribution with the mean at 0.01 and the variance  $0.035^2$ [14,15] for  $N=10,000$  disks (Fig. 1). The number of the disks is large enough to obtain sufficient statistics for data analysis. In this work, all length scales are normalized by the sample size, that is, the side length of the sample is kept at unity and the length scales of the microstructure entities are between 0 and 1.

The polycrystal model is established by constructing the “grain boundary” lines surrounding each disk to form grain cells as space filling polyhedra (Fig. 1). Each polygon in the LV polycrystal sample is a grain cell colored differently from its neighbors, the lines are grain boundaries, and the intersections by the grain boundary lines are triple junctions. The 2D LV diagram resembles the microstructure acquired in 2D slices in EBSD and 3DXRD. Note that in the LV model, the grain boundaries are obtained analytically. The other microstructural



information is available also analytically with their locations known. Therefore, we know the exact information of such a perfect microstructure, i.e. the spatial location and their geometric and topological properties of the boundary lines, vertex points, and grain cells. Therefore we can use the LV model as a reference to define missing microstructure features in DM acquisition. As shown below, we also use the LV model as a baseline sample to create various digital microstructures and perform simulation of digital microstructure acquisition with different resolutions.

### **Simulation of Digitization of Microstructure Acquisition**

Once the LV model is established, the next step is to obtain the “digital” microstructures from this sample with varying imposed resolution. The digitation process is simulated by the following steps. To explain the algorithms, we use an illustration from EBSD-FIB.

(a) Form a mesh or grid of the size  $a$  on top of the LV sample containing the ground truth microstructure (Fig. 2). The grids serve as the digitized electronic presentation of the microstructure, or voxels, and the size of the grids is related to the resolution.

The simulation procedure mimics experiment where the digital acquisition in FIB-EBSD or 2DXRD imposes grids naturally on the samples that are formed by the sectioned layers in depth and the discrete diffraction spots in lateral directions [7-11] (see Fig. 2). The microstructure information for each voxel is from different diffraction patterns of the grains with different orientations or lattice parameters. The same grain gives the same pattern in each diffraction spot that defines a voxel with a specific “grain index”, a numerical value used to identify one grain from the others. In our simulated DM data gathering, each square grid with size  $a$  is a “voxel” and the grid size defines a natural “resolution” or the length scale. Since we know all the microstructural information of the sample *a priori*, we can transfer the information of the underlying grain to each mesh grid voxel that is on top of the grain (see Fig. 2). In other words,

we do not need to perform EBSD or 2DXRD in order to obtain the attributes of the grains to assign to each voxel a grain index. Instead, we can identify each grain exactly from the information already known. This replaces what has to be done in experiments via data collection from EBSD or 2DXRD in the samples. The simulation procedure mimics in principle the digital acquisition of microstructure voxels from the LV model that already has all microstructural information known and available. Of course, if needed we can also obtain diffraction patterns from the grains in the LV model when crystal structure and misorientation are assigned to them. For now we forsake this step since the information needed to distinguish each grain is known in our LV polycrystal model.

(b) In our simulation, the grain indexing for each voxel is determined by the following conditions:

(i) If a grid is located entirely within a grain from the LV sample, we assign the grain index to this voxel; (ii) if a grid abuts two or more grains, we assign the grain index to the voxel by the grain that has the largest occupancy of the area of the grid, which resembles the procedure used in experiment acquisition; and (iii) if a grid has equal area occupancy of the grains underneath, we select one grain randomly and assign the grain index to this voxel [7-11].

The above procedure allows us to label all  $n \times n$  voxels with proper and consistent grain indices by the end. The domain formed by the contiguous voxels with the same grain index which is colored with the same color in digital presentation is a grain in the *digital microstructure* of the polycrystal (see Fig. 3). In the end, the 2D analog or image of the microstructure is transformed into a set of digitized voxels labeled with different grain indices. The collection of these voxels forms the so-called digital microstructure.

(c) To investigate the resolution effects on DMs, we follow the convention by defining the resolution  $R = n \times n$  as the number of grids per unit area, where  $n$  is the number of grids along each sample direction with length  $a$  (Fig. 2). Clearly, the larger  $n$  is, or the smaller  $a$  is, the larger

the resolution; and vice versa. Under different resolutions, that is, different grid size  $a$  or number of grid  $n$ , we can perform the “digitization” process with the techniques described above to obtain different sets of DMs. In this work, we take a series of  $n$  from 3000 to 100, over an order of magnitude change in linear scale.

## Results

Fig. 3(a) shows parts of the LV microstructure in Fig. 1 under the resolution of  $n=3000$ , or  $R=9\times 10^4$  voxels per unit area. For such a high resolution, one can hardly see the grid lines for the voxels. In other words, one can no longer tell the difference between the DM voxles and these of the computer screen for online view. These vast numbers of voxels with different grain indices form a set of digital microstructures: The same grain is grouped as a polyhedral domain of the voxels with the same color or grain index.

The quality of the DM under such a high resolution is reflected by the smooth boundaries surrounding each grain formed by different colored voxels of the neighboring grains. The boundaries between the voxels of different colors nearly overlap the grain boundary lines determined analytically in the ground truth LV model in Fig. 3(a). All voxels that have the same grain index or color fall completely inside the polygon grain cell from the LV model. Therefore, each voxel has the right grain index of the grains underneath. No microstructural feature of grain cell area, boundary lines and vertex points are missing for the DM data set composed of all  $3000\times 3000$  voxels.

As a comparison, we show in Fig. 3(b) and 3(c) the digitized microstructure with coarser resolutions at  $n=600$  and  $n=200$ . The voxels or square grids are much bigger than these in the DM with  $n=3000$  and become visible easily as compared with the finer ones in Fig. 3a. The domains of the same colored voxels form the new “grain cells” in the corresponding digital microstructures,

while the exact, “ground truth” microstructure is marked by the polygons surrounded by the grain boundary lines.

Direct comparison of the DMs in Fig. 3(b) and 3(c) with that in Fig. 3(a) reveals that the digitized microstructures under different resolutions have marked differences. The first is *coarsening of microstructure features*. The *grain boundaries* formed by the voxels with different colors appear to be more rugged. Particularly in Fig. 3(c), one can see the extruding square voxels across the boundary lines that are the real grain boundaries in the LV polycrystal sample. In addition, many of the *triple junction points* formed by three or more intercepting grains or domains of same-colored voxels (see Fig. 3(a) and (3(b))) become those of the intercepting corners of the square voxels of different colors. As we show below, the coarsening of these microstructure features under a decreasing resolution leads to significant changes in geometric and topological quantities of the DMs.

The second marked difference is *missing small grains*. When the voxel size is approaching that of the grains, the likelihood of these small grains missing from the DMs becomes larger. In Fig. 3(c) we mark some of these grains that disappear from the “digital microstructure” completely. The missing grains can be easily estimated using analytical approaches as shown below. However, other missing microstructure features associated with the missing grains are much harder to access, that requires us to perform simulation.

The third marked difference is shown by *missing vertices and boundaries*. Along with the missing grains, the associated boundary lines and vertices disappear accordingly. This can be seen clearly in Fig. 3(c) in contrast to the original grain boundary lines and vertices from the LV polycrystal sample. In addition, vertices and boundaries can vanish in DM even without missing

the grains, due simply to their lower dimensions or size. We will witness these consequences later in computing the geometric and topological properties of DMs.

The fourth is the *distortion of grain shapes*. In DM, grains are represented by domains of the discrete voxels, grain boundary lines by two abutting voxels from different grains, and the vertex points by three or more intercepting voxels with different colors. The discretization of the continuous geometric features makes them sensitive to the decrease in the number and size of the voxels. One of the most obvious outcome is grain shape distortion. First, when the number of voxels inside of a grain is getting fewer, the shape of the grains becomes less smooth with the zigzag grain boundaries made of extruding voxels (Fig. 3(c)). Second, the reduction in the number of the voxels and resulting missing grains at lower resolution exaggerates certain geometric features, including shape distortion in these grains. Shape distortion is further exacerbated if the original shape of the grains is already anisotropic. These changes become more severe and dramatic at coarser and coarser resolutions. As we show below, these changes in geometric and topological properties in the DMs can be quantified accurately in the LV polycrystal sample.

Having acquired the DMs in the format of voxels from our simulated digitization process and observed missing microstructures and their features, we proceed to quantify their geometric and topological properties as functions of the imposed resolution  $R$ . By comparing these properties with the microstructure properties to those from the exact “ground truth” sample we can see how much the missing information is, and importantly, which properties are more prone to being lost.

For the 2D slices, the topological properties include the number of grains ( $N_C$ ), the number of the triple junction points ( $N_V$ ), and the relations among them, i.e., the Euler relations. In addition, we have geometric properties such as the boundary length ( $L$ ) and the area of the cell ( $S$ ). These properties depict quantitatively each unique digital microstructure under a given resolution. The numerical techniques used in calculating these properties are developed by Xu and Li [16,17].

Each voxel is treated as a unit in the stochastic geometric objects, i.e. grains, boundary lines, and vertex points, labeled by the corresponding microstructure entities. As a result, each voxel is known for belonging to a specific vertex, triple junction, boundary, or grain cell. These labels, besides the grain index normally obtained in experiment or in our simulation, enable us to obtain all geometric and topological properties of these microstructural objects. For the perfect LV microstructure, these properties are obtained from geometric and topological relations through analytical methods [16,17]. By comparing these properties with the results from the DM in the samples with different resolutions, we can obtain quantitative answers of the changes

Fig. 4a shows the number of grain cells  $N_C$  left in the samples as the result of the varying resolution plotted against the inverse of the resolution,  $R^{-1}$ , where the high resolution result appears on the left and low resolution on the right of the figure. In the DMs,  $N_C$  is counted by the number of the domains formed by the voxels with the same grain index. In the initial perfect microstructure model, the total number of grains  $N_0=10000$ . When the resolution is larger than  $n=500$ ,  $N_C$  remains nearly unaffected. However, below  $n=400$ ,  $N_C$  shows a remarkable decrease versus the resolution. Indeed, the smaller the resolution, or the larger  $1/R$ , and the smaller the number of grains that one actually is able to acquire from the sample. The decrease can be significant when the resolution goes below  $n=400$ , or  $a=2.5 \times 10^{-3}$ . When the voxel size is down from  $n=3000$ , which corresponds to the highest resolution used in this work, to the coarsest  $n=100$  or  $a=10^{-2}$ , which is the mean grain diameter used in our case, nearly 40% grains are lost. The critical threshold resolution  $a$  corresponds to  $n=400$  is about 2.5. From these results we can give an estimation of the critical resolutions for real materials: For a polycrystal with the mean grain diameter of 1 micron, one will start to see missing grains if the resolution is about 0.25 microns or bigger; and about 40% of the grains with less than 1 micron in diameters are lost if the resolution is 1 micron.

As mentioned previously, for the sample with lognormal grain size distribution we can obtain analytical result for the missing grains. Therefore, we can compare the simulated missing number of grains  $N_{miss}$  with that from the exact solution using the grain size distribution. Consider a sample with initial  $N_0$  number of grains and the grain size distribution  $P_D(d)$ , where  $d$  is the mean for grain diameter  $D$ . If we set the externally imposed resolution  $a$  equal to  $d$ , the grains potentially missing or not being seen are those with the diameter  $d \leq a + \sigma$ , where  $\sigma$  is the variance in  $P_D(d)$ . This is because in actual DM acquisition, either in an experiment or a numerical simulation, the probability of locating exactly a grain with the diameter  $d = a$  is bounded by the upper limit  $a + \sigma$ . Therefore, we can estimate the number of potential missing grains from the cumulative function  $F_D(d) = \int_0^d P_D(D) \Delta D$ , where  $d = a + \sigma$  and  $P_D(D)$  is the grain size distribution. That is, the number of the missing grains is  $N_{miss} = N_0 F_D(d)$ , while the remaining number of grains in the sample is  $N_C = N_0(1 - F_D(d))$ .

Fig. 4b shows the lognormal grain size distribution  $P_D(d)$  used in this work to generate the LV diagram and the actual grain size distribution obtained after we digitize the LV polycrystal samples. The missing number of grains, or the cumulative function  $F_D(d)$ , along with the remaining number of grains  $N_C$ , are shown in the inset of Fig. 4b. As a comparison, in Fig. 4a and 4b we also plot the number of grains in the samples from the digitization process of the LV microstructures under different values of  $a$ . The numerically evaluated  $N_C$  from the DMs and exact solution agree well. As we can handle only a limited number of grains and size variance are in the simulations, the numerical results can only cover part of the analytical results. However, the general trend is captured within this range of grain sizes (Fig. 4b).

Geometric properties of the polycrystal sample are sensitive to missing microstructural information. However, they cannot be obtained analytically as for the missing number of grains. Therefore, simulation is needed. In Fig. 5a and 5b we show the distributions of the grain areas and the boundary lengths or perimeters for each grain in the DMs under different resolutions. The mean grain boundary area and the perimeter are obtained by averaging the corresponding areas and perimeters of all grains over the entire sample. Both the mean area and the mean length increase with decreasing resolution. This is expected since as the resolution becomes coarser, the number of grains becomes smaller and the grain size per grain gets bigger (Fig. 4). As a result, the grain area partitioned to each grain and its perimeter become larger because the total area of the sample is fixed. At  $n=100$  or when the resolution is reduced to about an order of magnitude, the area is enlarged by 148% and the perimeter length by 128%.

The change in the geometric properties of the DMs not only corroborates well with the finding in missing grains but also supports our claim of the importance of missing microstructural information. As known, both the grain area and perimeter length are extensive properties in defining grain boundary energy and transport properties, such as diffusion and thermal conductivity. Further investigation is need to find out how such large difference introduced in DMs affect the outcome in polycrystals' properties.

In addition to the geometric properties, topological properties are also sensitive to changes of microstructures, including the number of edges ( $N_E$ ) and the number of vertices of triple junction lines ( $N_V$ ). These topological quantities, along with the number of grains ( $N_C$ ), follow certain relations. For example, in two dimensions,  $N_E$ ,  $N_V$ , and  $N_C$  follow Euler relation  $N_V - N_E + N_C = 2$  and  $3N_V = 2N_E$ . We list these quantities and the Euler relations at different resolution  $n$  in Table 1. The general trend shows that all the topological quantities decrease with coarser resolution. For



example, as  $n$  changes by an order of a magnitude, from 3000 to 100, the number of grains  $N_C$  decreases by about 345%, edges  $N_E$  by 50 % and vertices  $N_V$  by nearly 60%.

In addition, we found that the decreasing rates for these microstructure entities are different: the vertices ( $N_V$ ) decrease the fastest, followed by edges ( $N_E$ ) and grain themselves ( $N_C$ ). Such a subtle difference reflects the sensitivity of different microstructure components to the variation of resolution: The vertex, being the one-dimensional point in terms of the geometric feature, is the first to be affected by a varying resolution. In other words, the lower dimensional object such as the vertex is the easiest to be missed during DM acquisition in simulation as well as in experiment. The second is the two-dimensional lines or edges. And, the last is the three-dimensional grain cells. With the sample principle, one should expect to see this hierarchical missing rate of microstructure features in real experimental DMs, which of course needs to be confirmed in the future.

The Euler relations among these quantities are, however, obeyed. That is, the variation of the resolution does not affect the topological relations for not only the total numbers of grains ( $N_C$ ), edges ( $N_E$ ) and vertices ( $N_V$ ), but also the mean number of edges  $\langle E \rangle$  and vertices  $\langle V \rangle$  per grain. This “agreement”, however, comes from the strict requirement in defining grains in DMs. The total area (or volume in 2D) must be partitioned by all grains. Confirmation of the Euler relations does not indicate the “missing” microstructure features, nor the severity of the lost information. Instead, the confirmation of the Euler relations simply reflects the fact that the grain cells are convex polyhedra and full space filling by these polyhedral grains, even with a certain lost features of the microstructures.

## **Discussion and conclusions**

Like traditional metallography that has played important roles in materials science for the past century, digitization of microstructure is part of the modern advances that promises to

revolutionize synthesis, processing, and application of a vast number of materials. Compared to metallography, digital microstructure involves data with unproportioned magnitude of size and scale. Dealing with missing information is the first and most important step in DM data gathering and processing as missing information is expected to impact significantly the downstream applications in microstructure reconstruction, quantification and characterization, as well as the vast number of applications pivoted on (micro) structure-property relations. Data fidelity, therefore, becomes a key issue in ensuring the reliability and final applicability of the digital microstructures. While the current technology enables us to acquire DMs, little attention has been paid to the missing microstructure and its consequences. To raise the awareness, we have performed a first simulation of digital microstructure acquisition to address this issue. The simulation is necessary since no DM data available today are free of the concern and potential influence of the missing information.

We have established a baseline model of a perfect microstructure, the Laguerre-Voronoi polycrystal in two dimensions. The ground truth from this model polycrystal allows us to make rigorous comparison and quantitative measurements for the microstructure quantities that are lost in the digital acquisition. We have shown that both the geometric and topological quantities disappear under decreasing resolution and become significant with nearly half of these features missing. Among the missing features, it is the smallest microstructural entity, i.e. the vertex point, that can be easily missed, followed by line and the volume features. We have also discovered that the Euler relations are obeyed among the topological quantities at all resolutions. This suggests that the topological relations are insensitive to the imposed resolution. As a result, the Euler relations are not the good measure for missing information.

We hope that these new findings and insights may bring awareness to the important issue of missing information in digital microstructure acquisition. As the digital microstructure

acquisition and data processing rapidly growing, the issue of data fidelity is expected to inspire more experimental efforts in treating the missing information rigorously and systematically.

### **Acknowledgement**

M.L. acknowledges the support for this work by the National Thousand Talents Program of China. C.J. and C.B.B. would like to thank Partial support from the 2011 Project from the Collaborative Center for Advanced Structures and Manufacturing of Nonferrous Metals of the Central South University.

**\*Corresponding Author:** Email: [mo.li@gatech.edu](mailto:mo.li@gatech.edu)

## References

- [1] Callister W D, Rethwisch Jr D G. *Materials Science and Engineering: An Introduction*, 8<sup>th</sup> edition(John Wiley & Sons, 2010).
- [2] Kurzydowski K J, Ralph B. *The quantitative description of the microstructure of materials*( CRC press, 1995).
- [3] Cahn R W, Haasen P. *Physical Metallurgy*. ( Elsevier Science B.V., 1996).
- [4] Vaidya S, Sinha A K. Effect of texture and grain structure on electromigration in Al-0.5% Cu thin films. *Thin Solid Films*. 75(3): 253-259 (1981).
- [5] Pierce D G, Brusius P G. Electromigration: A review. *Microelectronics Reliability*. 37(7): 1053-1072 (1997).
- [6] Sczerzenie F E, Maurer G E, Gell M, et al. *Superalloys 1984: Proc. 5th Int. Symp. on Superalloys*, TMS-AIME, Warrendale, PA. 573-82 (1984).
- [7] Bowman R. *Superalloys: A primer and history*. (9th International Symposium on superalloys, 2000).
- [8] Spowart J E. Automated serial sectioning for 3-D analysis of microstructures. *Scripta Materialia*. 55(1): 5-10 (2006).
- [9] Uchic M D, Groeber M A, Rollett A D. Automated serial sectioning methods for rapid collection of 3-D microstructure data. *Jom*. 63(3): 25-29 (2011).
- [10] Kotula P G, Rohrer G S, Marsh M P. Focused ion beam and scanning electron microscopy for 3D materials characterization. *MRS Bulletin*. 39(4): 361-365 (2014).
- [11] Rajan K. Materials informatics. *Materials Today*. 8(10): 38-45 (2005).
- [12] Duda R O, Hart P E, Stork D G. *Pattern classification*. (John Wiley & Sons, 2012).
- [13] X-ray (Henning)
- [14] X-ray (Henning)

- [15] Imai H, Iri M, Murota K. Voronoi diagram in the Laguerre geometry and its applications. *SIAM Journal on Computing*. 14(1): 93-105 (1985).
- [16] Xu T, Li M. Topological and statistical properties of a constrained Voronoi tessellation. *Philosophical Magazine*. 89(4): 349-374 (2009).
- [17] Li M, Xu T. Topological and atomic scale characterization of grain boundary networks in polycrystalline and nanocrystalline materials. *Progress in materials science*. 56(6): 864-899 (2011).

**Table 1** The topological properties and their relations for the 2D Laguerre-Voronoi (LV) diagram at different resolution  $n$ .  $N_C$  is the total number of grains,  $N_E$  the total number of edges, and  $N_V$  vertices.  $\langle V \rangle$  is the average number of vertices per grain and  $\langle E \rangle$  the average number of edges per grain.  $N_V - N_E + N_C = 2$ ,  $2N_V = 3N_E$ , and  $\langle V \rangle = \langle E \rangle$  are the Euler relations among these quantities. For comparison, the exact results from the perfect LV microstructure with infinite resolution, are listed. The change of each quantity compared to the “ground truth” is represented as the percentage (%) shown in the parentheses.

$n$	$N_E$	$N_V$	$N_C$	$N_V - N_E + N_C$	$2N_E - 3N_V$	$2N_E / 3N_V$	$\langle V \rangle$	$\langle E \rangle$	$\langle V \rangle / \langle E \rangle$
LV	30394 (100%)	20395 (100%)	10001 (100%)	2	-397	0.9935	6 (100%)	6 (100%)	1
3000	29685 (100%)	19910 (100%)	9777 (100%)	2	-360	0.9935	5.9850 (100%)	5.9850 (100%)	1
2000	29639 (99.85%)	19865 (97.77%)	9776 (99.99%)	2	-317	0.9946	5.9821 (99.95%)	5.9821 (99.95%)	1
1500	29618 (99.77%)	19848 (99.69%)	9772 (99.95%)	2	-308	0.9948	5.9795 (99.91%)	5.9795 (99.91%)	1
1000	29575 (99.63%)	19809 (99.49%)	9768 (99.91%)	2	-277	0.9953	5.9723 (99.78%)	5.9723 (99.78%)	1
900	29570 (99.61%)	19808 (99.48%)	9764 (99.87%)	2	-284	0.9952	5.9743 (99.82%)	5.9743 (99.82%)	1
800	29537 (99.50%)	19777 (99.33%)	9762 (99.85%)	2	-257	0.9956	5.9688 (99.73%)	5.9688 (99.73%)	1
700	29487 (99.33%)	19731 (99.10%)	9758 (99.81%)	2	-219	0.9963	5.9615 (99.61%)	5.9615 (99.61%)	1
600	29432 (99.15%)	19680 (98.84%)	9754 (99.76%)	2	-176	0.9970	5.9523 (99.45%)	5.9523 (99.45%)	1
500	29293 (98.68%)	19546 (98.17%)	9749 (99.71%)	2	-52	0.9991	5.9270 (99.03%)	5.9270 (99.03%)	1
400	29021 (97.76%)	19282 (96.86%)	9738 (99.60%)	2	187	1.0032	5.8790 (98.23%)	5.8790 (98.23%)	1
300	28315 (95.38%)	18625 (93.55%)	9692 (99.13%)	2	755	1.0135	5.7623 (96.28%)	5.7623 (96.28%)	1
200	25681 (86.51%)	16328 (82.00%)	9355 (95.68%)	2	2378	1.0485	5.4140 (90.46%)	5.4140 (90.46%)	1
150	21981 (74.05%)	13384 (67.22%)	8599 (87.95%)	2	3810	1.0948	5.0399 (84.21%)	5.0399 (84.21%)	1
100	14773 (49.77%)	8366 (42.02%)	6409 (65.56%)	2	4448	1.1772	4.5387 (75.87%)	4.5387 (75.87%)	1

## Figure captions

**Figure 1** The Laguerre-Voronoi diagram generated by distributing 10 000 disks on a plane that follows a lognormal distribution with the mean diameter at 0.010 and the variance  $0.035^2$ . The sample is in a  $1 \times 1$  unit cell; all lengths including the grain cell diameter are normalized by the sample size. The lines partitioning the plane into polygons or grain cells are grain boundaries and their interception points are triple junctions in two dimensions. Different colors represent different grains, or the grain index in digital microstructure. The LV diagram is the perfect microstructure where the “ground truth” microstructural information is known exactly.

**Figure 2** Illustration of digitization process in the simulation of microstructure acquisition of a 3D polycrystal sample. A layer with thickness of  $a$  is formed between two planes. Each plane is meshed into  $n \times n$  lateral grids of the size  $a \times a$ . In 3D, the grids form a voxel of the size  $a \times a \times a$  (the cube formed by white lines) and in 2D the voxel is  $a \times a$ . The grid size  $a$  is determined by the resolution such as the electron beam diffraction or scattering sampling interval on the surface (the red arrow) and the sectioning layer thickness between the two planes. Each voxel contains the information of the grain underneath.

**Figure 3** Part of the microstructure shown in Fig. 1 and its digitized microstructure in 2D as shown in Fig. 2. Different resolutions are used, (a)  $n=3000$ , (b)  $n=600$ , and (c)  $n=200$ . The perfect ground truth microstructure is shown by the polygons surrounded by the grain boundary lines and triple junction points drawn from the analytical solution. The digitized microstructures are shown by the voxels or square grids. The grain cells are the domains of the contiguous voxels with the same color or grain indices. Some missing grains in the digitized microstructures are marked by red circles in (c).

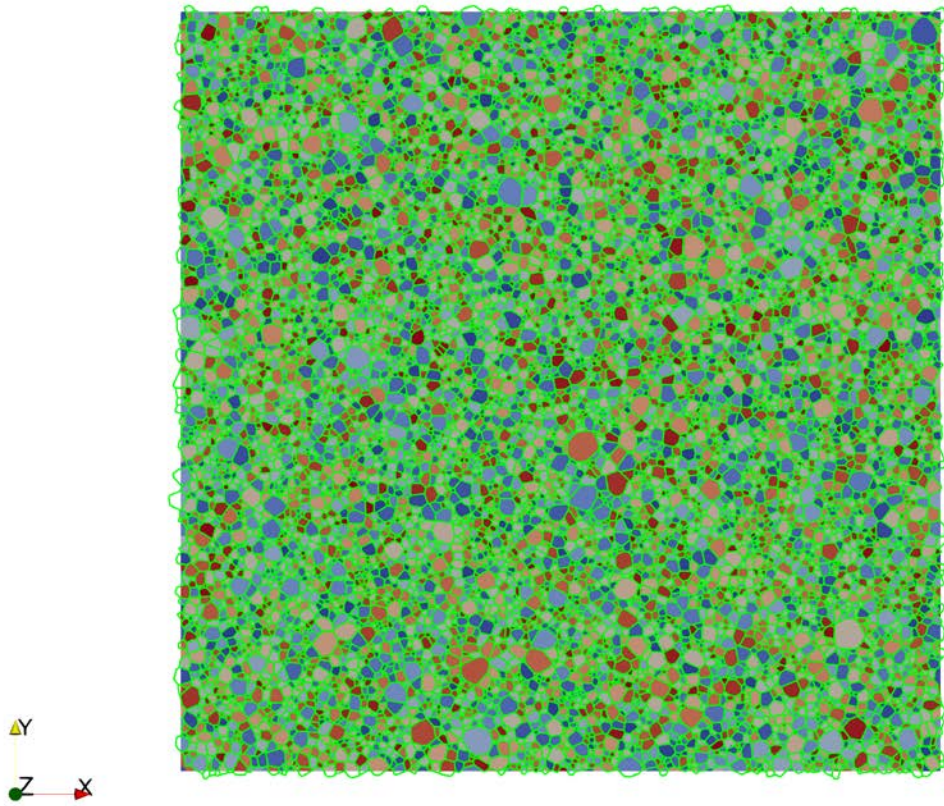
**Figure 4(a)** The log-log plot of the number of grain cells  $N_C$  in the DMs as the result of the varying resolution,  $R^{-1}$ . The number of grains obtained from the analytical solution is also plotted as a comparison. The data points are from the simulated DM and analytical solution and the lines are the guide for eyes.

**Figure 4(b)** The lognormal grain size distribution  $P_D(d)$  used to generate LV microstructure model (green line). The mean is at 0.01 and variance  $0.035^2$ . The actual grain size distribution from the digitized LV model obtained with  $n=3000$  (blue line). The inset is the cumulative function  $F_D(d)$  (thin blue line) and the number of missing grains  $N_C$  (thin red line) obtained from  $N_C = N_0 - F_D(d)$ . The actual number of grains obtained from the digital microstructures of the LV microstructure under different resolution is show as blue squares.

**Figure 5(a)** The distributions of the grain areas for each grain in the perfect LV polycrystal and digitized microstructures under different imposed resolutions. The insets mean are area and the variance plotted in the inset and the zoom-in view for the grain area distribution. The data points are from the simulated DMs and the lines are the best fit. The insets are for the zoom-ins for  $P(s)$  and the mean and variance of the grain area.

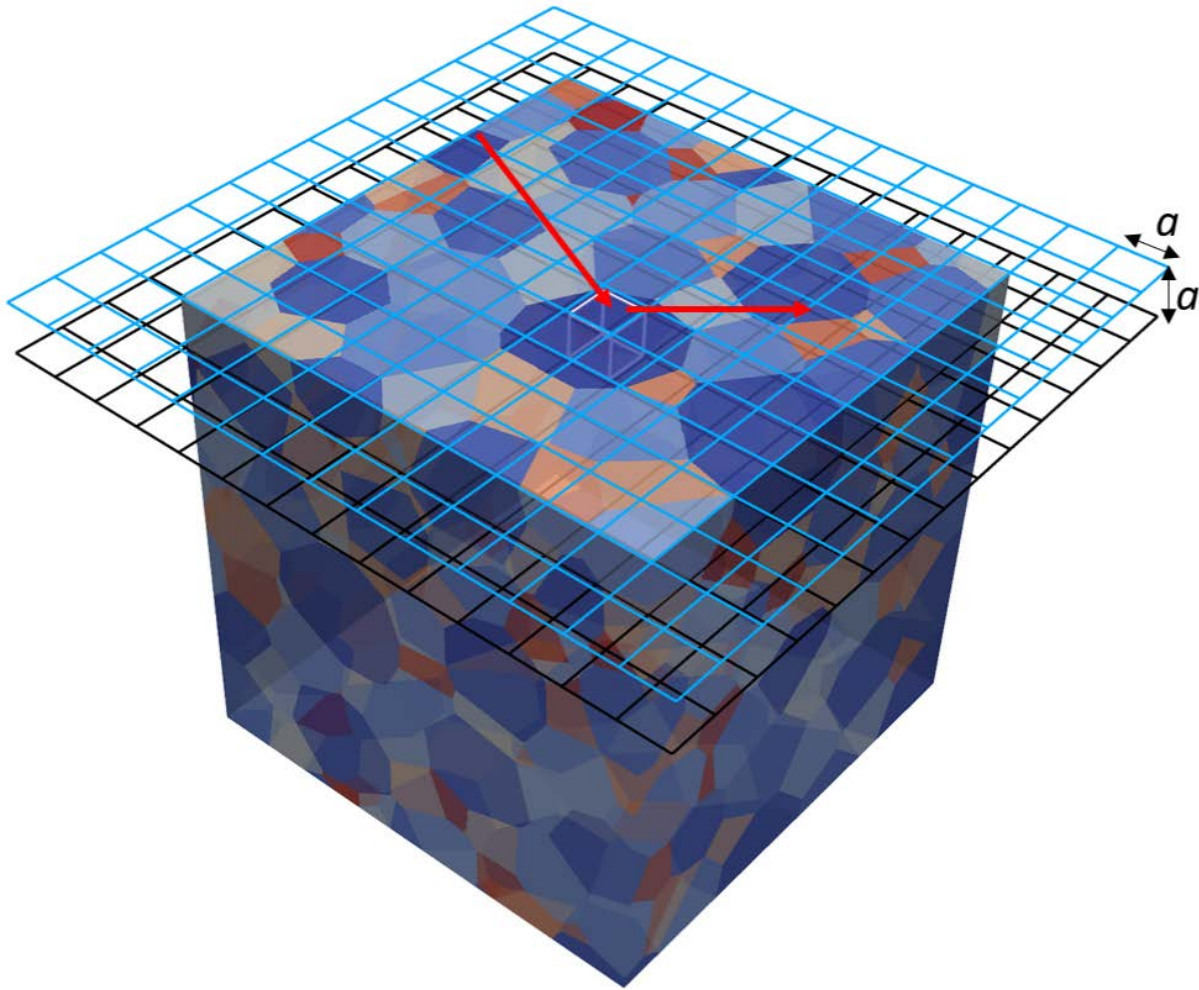
**Figure 5(b)** The distributions of the grain boundary length for each grain in the perfect LV polycrystal and digitized microstructures under different imposed resolutions. The insets are mean boundary length and the variance plotted in the inset and the zoom-in view for the grain area distribution. The data points are from the simulated DMs and the lines are the best fit. The inset is the mean and variance of the grain parameter length.





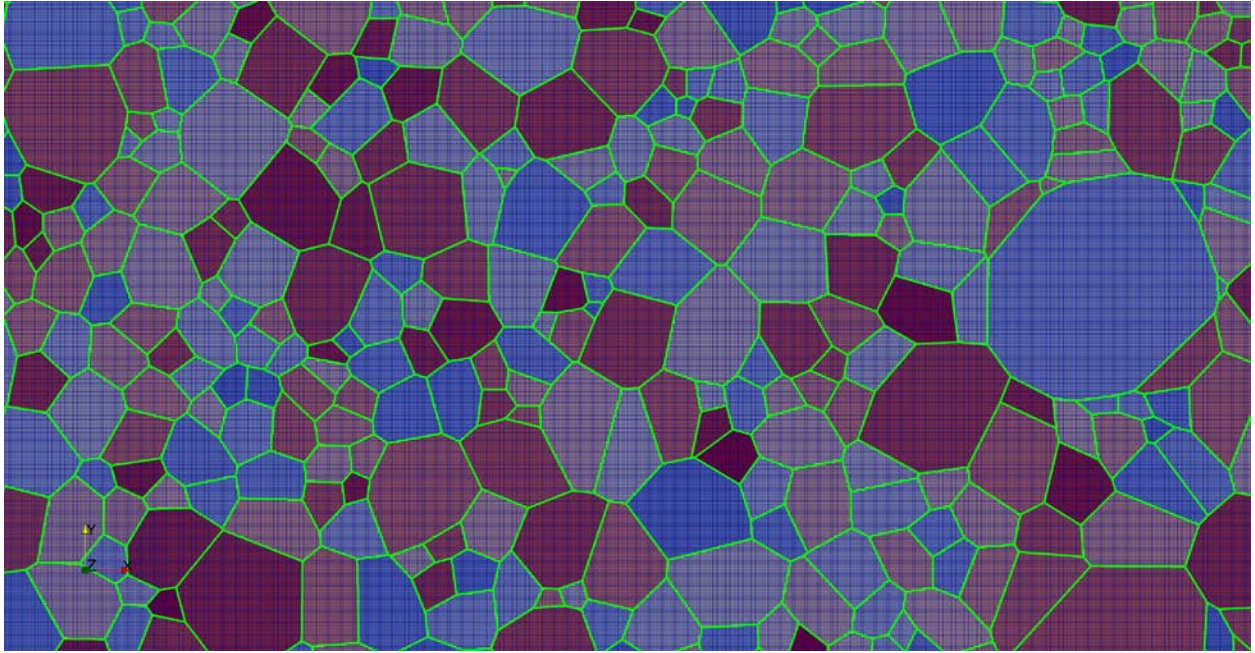
**Figure 1**

**Figure 1** The Laguerre-Voronoi diagram generated by distributing 10 000 disks on a plane that follows a lognormal distribution with the mean diameter at 0.010 and the variance  $0.035^2$ . The sample is in a  $1 \times 1$  unit cell; all lengths including the grain cell diameter are normalized by the sample size. The lines partitioning the plane into polygons or grain cells are grain boundaries and their interception points are triple junctions in two dimensions. Different colors represent different grains, or the grain index in digital microstructure. The LV diagram is the perfect microstructure where the “ground truth” microstructural information is known exactly.

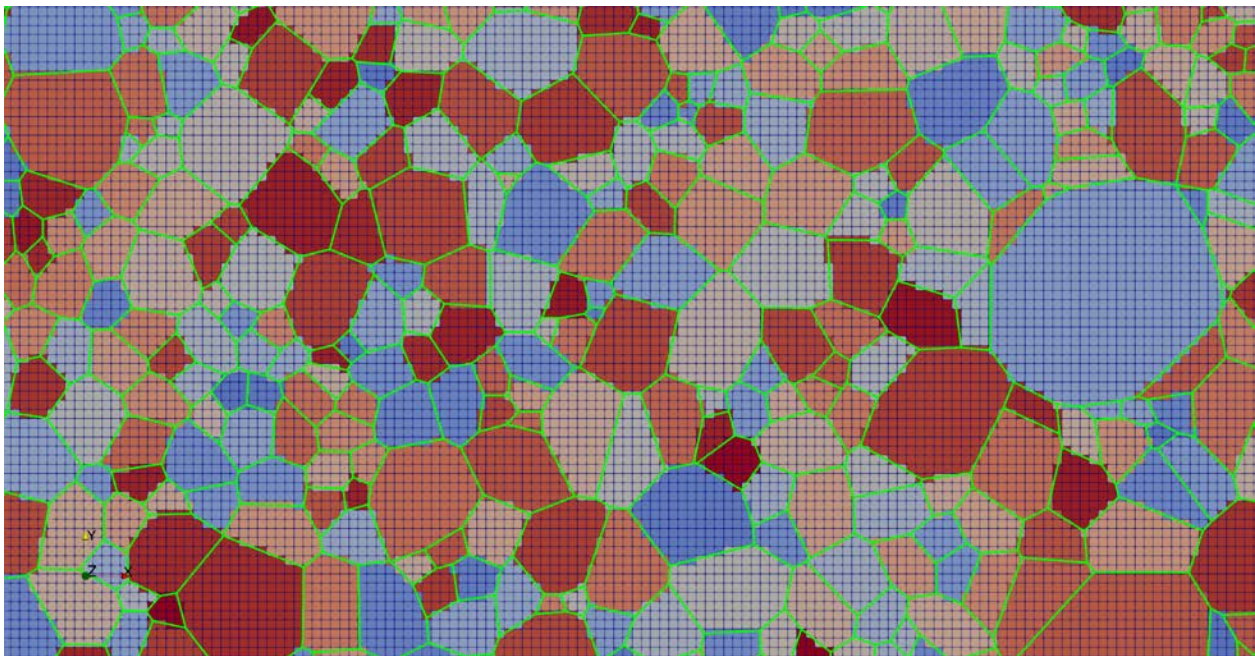


**Figure 2**

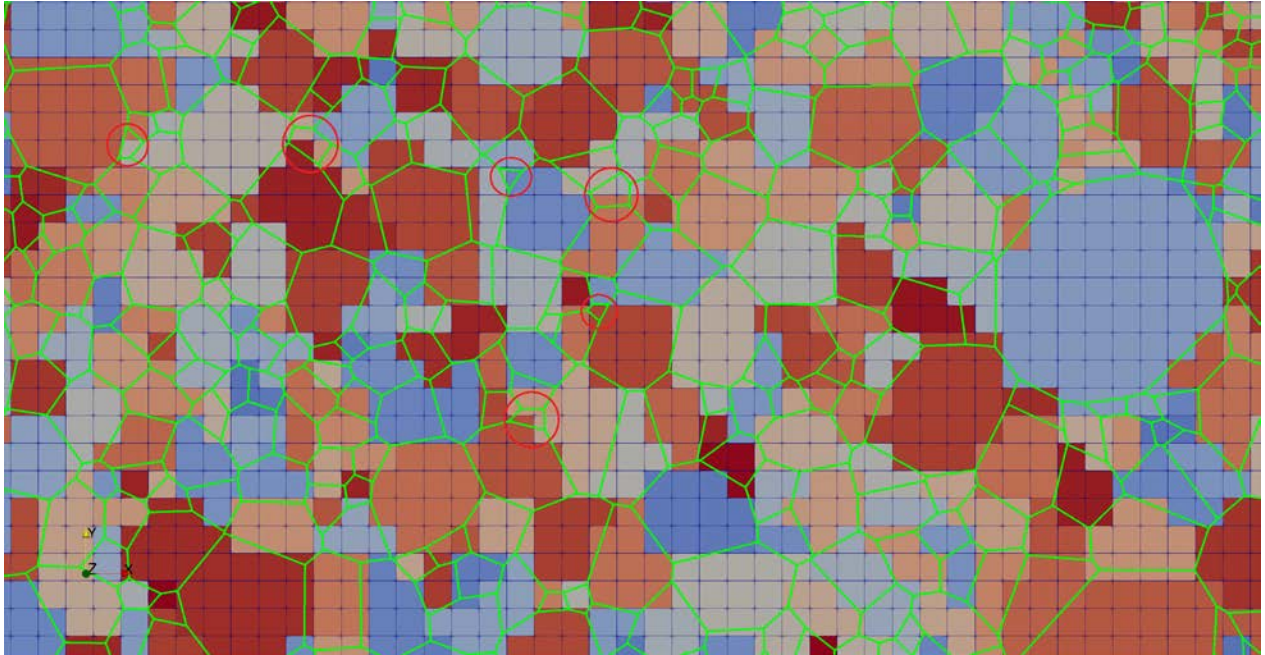
**Figure 2** Illustration of digitization process in the simulation of microstructure acquisition of a 3D polycrystal sample. A layer with thickness of  $a$  is formed between two planes. Each plane is meshed into  $n \times n$  lateral grids of the size  $a \times a$ . In 3D, the grids form a voxel of the size  $a \times a \times a$  (the cube formed by white lines) and in 2D the voxel is  $a \times a$ . The grid size  $a$  is determined by the resolution such as the electron beam diffraction or scattering sampling interval on the surface (the red arrow) and the sectioning layer thickness between the two planes. Each voxel contains the information of the grain underneath.



**Figure 3(a)**

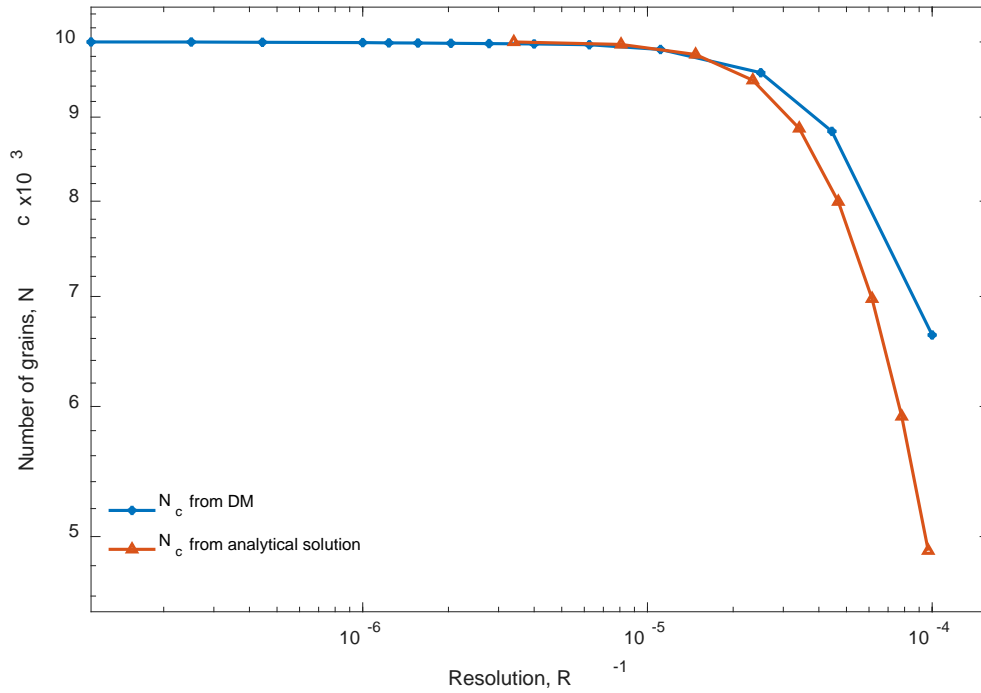


**Figure 3(b)**



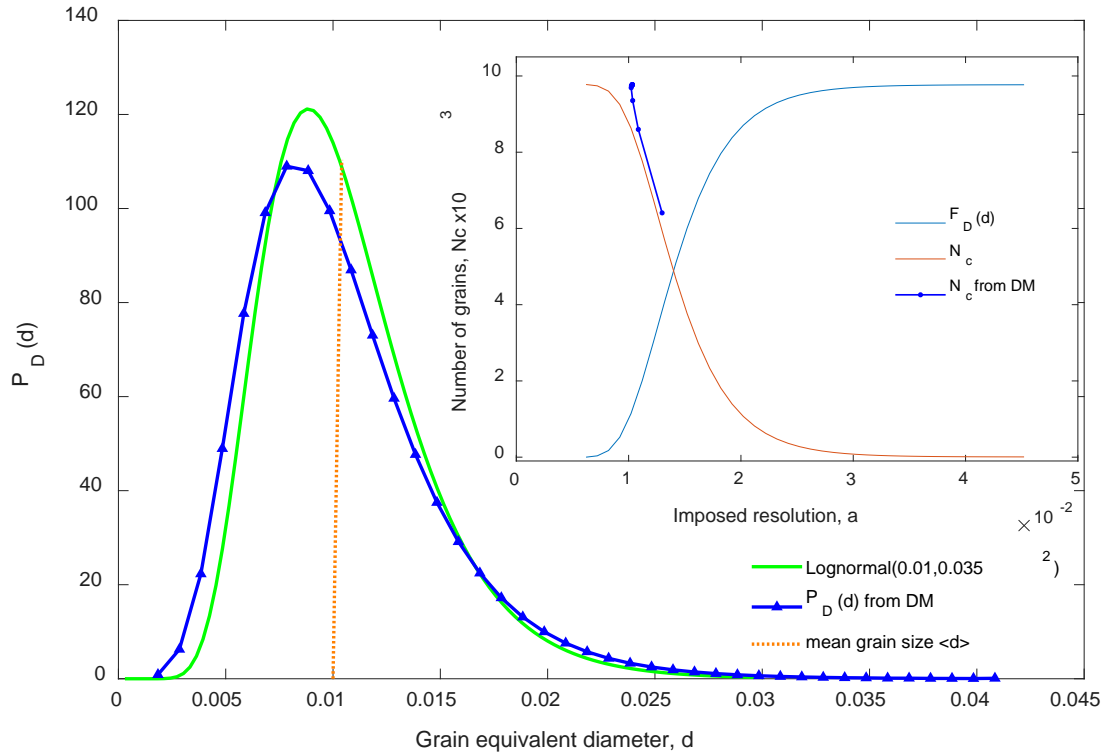
**Figure 3(c)**

**Figure 3** Part of the microstructure shown in Fig. 1 and its digitized microstructure in 2D as shown in Fig. 2. Different resolutions are used, (a)  $n=3000$ , (b)  $n=600$ , and (c)  $n=200$ . The perfect ground truth microstructure is shown by the polygons surrounded by the grain boundary lines and triple junction points drawn from the analytical solution. The digitized microstructures are shown by the voxels or square grids. The grain cells are the domains of the contiguous voxels with the same color or grain indices. Some missing grains in the digitized microstructures are marked by red circles in (c).



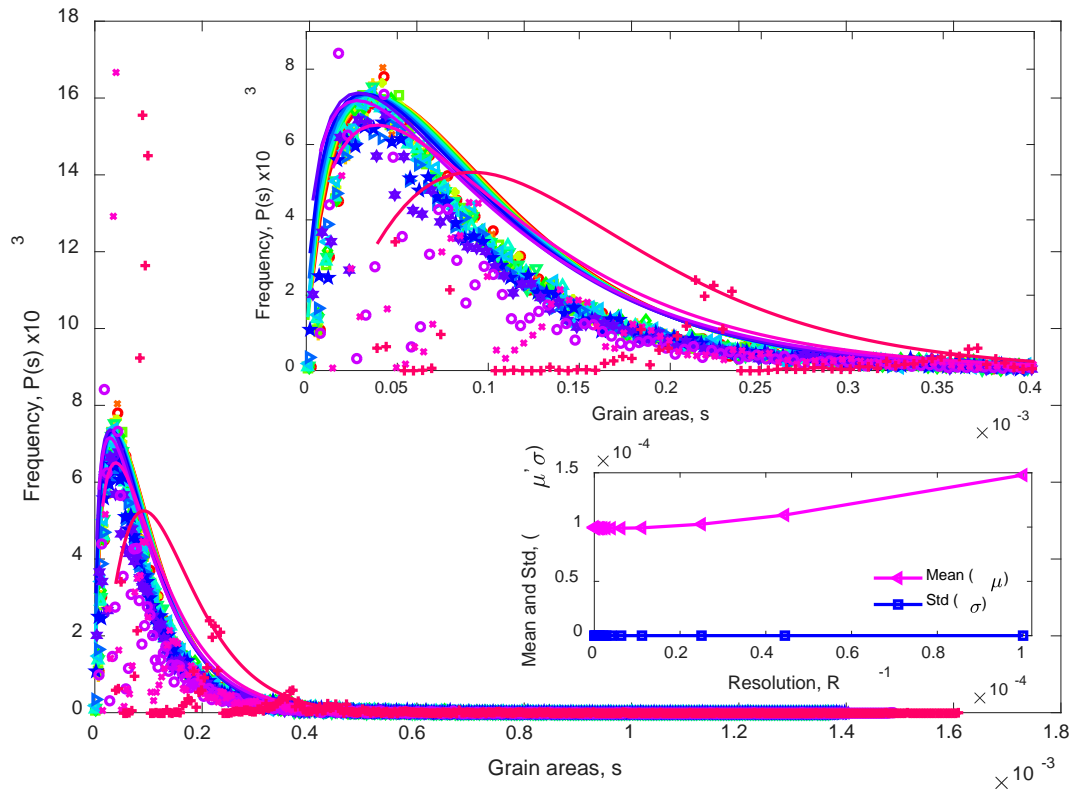
**Figure 4(a)**

**Figure 4(a)** The log-log plot of the number of grain cells  $N_C$  in the DMs as the result of the varying resolution,  $R^{-1}$ . The number of grains obtained from the analytical solution is also plotted as a comparison. The data points are from the simulated DM and analytical solution and the lines are the guide for eyes.



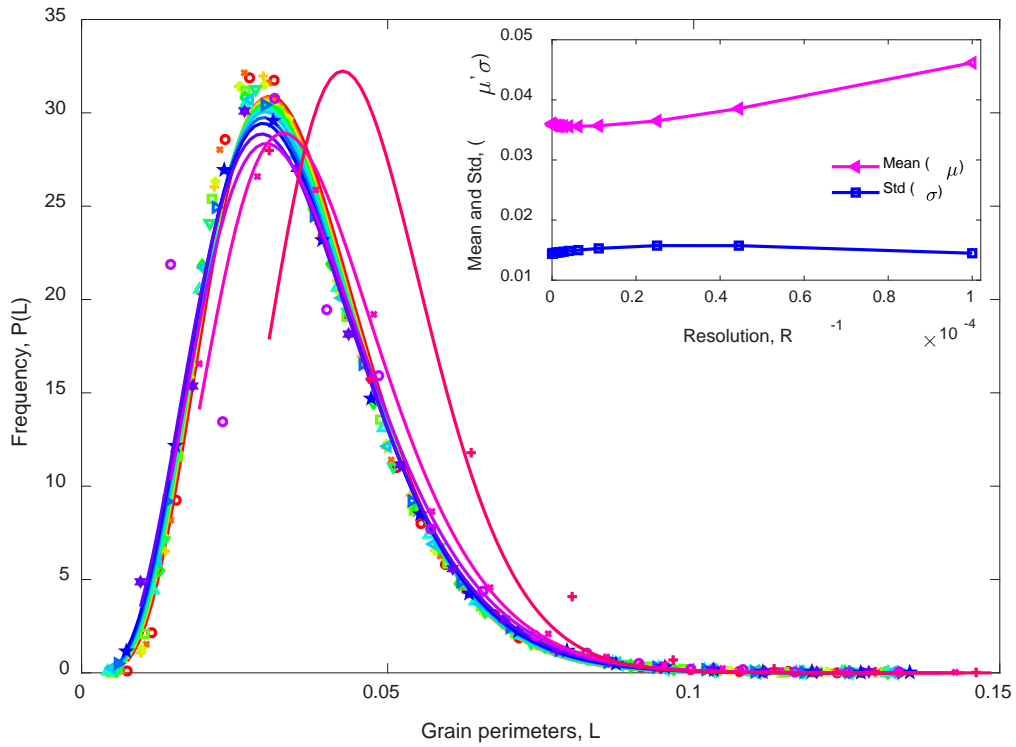
**Figure 4(b)**

**Figure 4(b)** The lognormal grain size distribution  $P_D(d)$  used to generate LV microstructure model (green line). The mean is at 0.01 and variance  $0.035^2$ . The actual grain size distribution from the digitized LV model obtained with  $n=3000$  (blue line). The inset is the cumulative function  $F_D(d)$  (thin blue line) and the number of missing grains  $N_C$  (thin red line) obtained from  $N_C = N_0 - F_D(d)$ . The actual number of grains obtained from the digital microstructures of the LV microstructure under different resolution is show as blue squares.



**Figure 5(a)**

**Figure 5(a)** The distributions of the grain areas for each grain in the perfect LV polycrystal and digitized microstructures under different imposed resolutions. The insets mean are area and the variance plotted in the inset and the zoom-in view for the grain area distribution. The data points are from the simulated DMs and the lines are the best fit. The insets are for the zoom-ins for  $P(s)$  and the mean and variance of the grain area.



**Figure 5(b)**

**Figure 5(b)** The distributions of the grain boundary length for each grain in the perfect LV polycrystal and digitized microstructures under different imposed resolutions. The insets are mean boundary length and the variance plotted in the inset and the zoom-in view for the grain area distribution. The data points are from the simulated DMs and the lines are the best fit. The inset is the mean and variance of the grain parameter length.