Effect of the image resolution on the statistical descriptors of heterogeneous media

René Ledesma-Alonso*

CONACYT-Universidad de Quintana Roo, Boulevar Bahía s/n, Chetumal, 77019, Quintana Roo, México

Romeli Barbosa† and Jaime Ortegon

Universidad de Quintana Roo, Boulevar Bahía s/n, Chetumal, 77019, Quintana Roo, México

(Dated: November 11, 2018)

The characterization and reconstruction of heterogeneous materials, such as porous media and electrode materials, involve the application of image processing methods to data acquired by SEM or other microscopy techniques. Among them, binarization and decimation are critical in order to compute the correlation functions that characterize the micro-structure of the abovementioned materials. In this study, we present a theoretical analysis of the effects of the image size reduction, due to the progressive and sequential decimation of the original image. Three different decimation procedures (random, bilinear and bicubic) were implemented and their consequences on the discrete correlation functions (two-points, line-path and pore-size distribution) and the coarseness (derived from the local volume fraction) are reported and analyzed. The chosen statistical descriptors (correlation functions and coarseness) are typically employed to characterize and reconstruct heterogeneous materials.

A normalization for each of the correlation functions has been performed. When the loss of statistical information has not been significant for a decimated image, its normalized correlation function is forecast by the trend of the original image (reference function). In contrast, when the decimated image does not represent the statistical evidence of the original one, the normalized correlation function diverts from the reference function. Moreover, the equally weighted sum of the average of the squared difference, between the discrete correlation functions of the decimated images and the reference functions, leads to a definition of an overall error. During the first stages of the gradual decimation, the error remains relatively small and independent of the decimation procedure. Above a threshold defined by the correlation length of the reference function, the error becomes a function of the number of decimation steps. At this stage, some statistical information is lost and the error becomes dependent of the decimation procedure.

These results may help us to restrict the amount of information that one can afford to lose during a decimation process, in order to reduce the computational and memory cost, when one aims to diminish the time consumed by a characterization or reconstruction technique, yet maintaining the statistical quality of the digitized sample.

I. INTRODUCTION

The advancements on image acquisition technology, which direct consequence is the enhanced resolution of scanners, probes and sensors, allows to obtain a huge amount of information and a detailed description of everyday objects. In addition, the application of this technological progress is not restricted to human scale, since the connection with instruments enables the observation and analysis of objects of scientific interest, in either a much larger scale with a telescope or a much smaller scale with a microscope. For instance, a scanning electron microscope (SEM) can produce high-resolution images of the microscopic structure of materials, such as porous media, composites and random heterogeneous media (RHM). Each SEM image, which corresponds to a two-dimensional grayscale map, provides a projection of the microscopic configuration of the phases that constitute the material under examination. With the employment of the proper techniques [1, 2], this microscopic portrait can be used to characterize the original material and to reconstruct a representative three-dimensional sample that mimics the structure of the heterogeneous media.

3D materials characterization and reconstruction methods based on correlation functions have been broadly accepted since 20 years ago [3–5], and they continue to be used [6–15]. Such reconstructions are of great value in a wide variety of fields, mainly for the recovery of the effective properties of the heterogeneous materials [16–18]. The knowledge of these effective properties is required for applications in physics, materials science, engineering, geology and biology, among other unexpected domains such as food industry [19] and functional clothing [20]. A particular case of interest is the characterization, 3D reconstruction and determination of effective properties of the catalytic layer (CL) of a Proton Exchange Membrane Fuel Cell (PEMFC) [8, 21–23], which will eventually lead to the design of CLs that maximize the separation, exchange and conduction of chemical species.

Recent studies focus on the determination of the size of the representative volumes to characterize the microstructure of the materials [24, 25] and to simulate and recover the effective properties of heterogeneous materials in different fields, for instance in mechanics [26, 27]. In previous studies [28], it has been also found that the size

* rledesmaalonso@gmail.com
† romelix1@gmail.com

of the observation window Ω, within the domain space $\mathcal{V}$ of the entire two-phase RHM, determines the fluctuation of the local volume fraction τ_j, with respect to the global volume fraction ϕ_j of phase j. Special interest has been paid to the determination of the correct size of the image, required to generate an adequate characterization and a representative 3D sample of the material. Nevertheless, these efforts have been only focused on setting the physical size of the sample, neglecting a fundamental issue: the image resolution, which defines the amount of data and the computing time during the characterization, reconstruction and simulation processes.

High-resolution images, their storage and processing stages require high computing power. In this context, it is necessary to optimize the amount of information that an image, representative of a material sample, can provide for its satisfactory characterization and 3D reconstruction. This path is explored in the present study, since we seek to reduce the computational load by decreasing the resolution (number of pixels) of the image, while preserving the statistical information that characterizes the original material sample. The statistical resemblance between the original image and a reduced-size image is verified by means of statistical descriptors (correlation functions [16]).

Herein we propose a method based on the correlation length of low-level normalized correlation functions , which indicates the maximum number of steps of a gradual decimation process, that are allowed in order to obtain an image of minimal size (in pixels) without losing statistical information.

This manuscript is organized as follows. In section II we describe briefly the classic and normalized statistical descriptors and the local volume fraction, which are used to characterize the material samples. The generation of original (full-size) images representing heterogenous materials, the gradual decimation process, the corresponding characterization and the computation of the deviation of the reduced-size images with respect to the original image and the coarseness are presented in section III. The implementation of these ideas and procedures on several types of numerically generated binary images of heterogeneous materials has been done and the results are exposed in section IV. The main outcome of this manuscript is described in section V, an algorithm that performs a statistically sensitive decimation process. Additionally, an example of a real case application of this methodology, the characterization and decimation of an image representing the catalytic layer of a PEMFC, is presented is section VI. Finally, conclusions are given in sec. VII.

II. CHARACTERIZATION

A. Statistical descriptors

Consider a material with domain $\mathcal{V} \subseteq \mathbb{R}^d$, of volume V and composed by $J = 2$ disjoint random phases. Each phase $j \in [0, J-1]$ presents a particular domain $\mathcal{V}_j$, a volume fraction ϕ_j and an interfacial area per unit volumen s. Therefore, it turns out that

$$\bigcup_{j=0}^{J-1} \mathcal{V}_j = \mathcal{V} \ , \qquad \bigcap_{j=0}^{J-1} \mathcal{V}_j = \emptyset \ , \qquad \sum_{j=0}^{J-1} \phi_j = 1 \ . \qquad (1)$$

The indicator function $\mathcal{I}_j(\boldsymbol{x})$ of phase j is defined as

$$\mathcal{I}_j(\boldsymbol{x}) = \begin{cases} 1 & \text{if the point } \boldsymbol{x} \text{ lies on } \mathcal{V}_j \\ 0 & \text{otherwise} \end{cases} \ , \qquad (2)$$

where $\boldsymbol{x}$ is any position in $\mathcal{V}$. In addition, the set of indicator functions must satisfy the relationship

$$\sum_{j=0}^{J-1} \mathcal{I}_j(\boldsymbol{x}) = 1 \ . \qquad (3)$$

In this work, three low-level correlation functions are used as statistical descriptors: two-point correlation function, line-path correlation function and pore-size distribution function. The definitions of these statistical descriptors [16], which will be recalled in the following, are based on the previously described indicator functions $\mathcal{I}_j$. Even though these expressions are given for a three-dimensional media, the equivalent formulas for a 2D media are straightforward.

1. Two-point correlation function

The two-point correlation function $S_j^{(2)}(\boldsymbol{x}_a, \boldsymbol{x}_b)$ is the probability that the two end points $\boldsymbol{x}_a$ and $\boldsymbol{x}_b$ of a line segment of length r fall in the same phase j. Considering a statistically homogeneous and isotropic media, this statistical descriptor is no longer a function of the specific positions $\boldsymbol{x}_a$ and $\boldsymbol{x}_b$, but becomes only a function of the distance r, thus being represented by

$$S_j^{(2)}(r) \equiv \left\langle \mathcal{I}_j(\boldsymbol{x}) \mathcal{I}_j(\boldsymbol{x} + \boldsymbol{r}) \right\rangle \ , \qquad (4)$$

where $\langle \cdot \rangle$ is the expectation value in the domain $\mathcal{V} \subseteq \Re^d$, $\boldsymbol{x}$ is any position in $\mathcal{V}$ and $\boldsymbol{r}$ is a vector of relative position with magnitude r and an arbitrary direction. In addition, the absence of long-range order allows us to forecast the following values

$$S_j^{(2)}(0) = \phi_j \ , \qquad S_j^{(2)}(\infty) \to \phi_j^2 \ . \qquad (5)$$

From this correlation function, we can also define the autocovariance function

$$\chi_j(r) = S_j^{(2)}(r) - \phi_j^2 \ , \qquad (6)$$

which presents the corresponding values

$$\chi_j(0) = \phi_j[1 - \phi_j] \ , \qquad \chi_j(\infty) \to 0 \ . \qquad (7)$$

2. Line-path correlation function

The line-path correlation function $L_j(\boldsymbol{x_a}, \boldsymbol{x_b})$ is the probability that a line segment of length r and with end points $\boldsymbol{x_a}$ and $\boldsymbol{x_b}$ lies wholly in phase j. Considering a statistically homogeneous and isotropic media, this statistical descriptor is no longer a function of the specific positions $\boldsymbol{x_a}$ and $\boldsymbol{x_b}$, but becomes only a function of the distance r, thus reading

$$L_j(r) \equiv \left\langle \left\lfloor \int_0^1 \mathcal{I}_j(\boldsymbol{x} + \alpha \boldsymbol{r}) \, d\alpha \right\rfloor \right\rangle , \qquad (8)$$

where $\langle \cdot \rangle$ is the expectation value in the domain $\mathcal{V} \subseteq \Re^d$, $\boldsymbol{x}$ is any position in $\mathcal{V}$ and $\boldsymbol{r}$ is a vector of relative position with magnitude r and an arbitrary direction. In addition, the absence of long-range order allows us to predict the following values

$$L_j(0) = \phi_j , \qquad L_j(\infty) \to 0 . \qquad (9)$$

3. Pore-size probability density function

The pore-size probability density function $P_j(r)$ is the likelihood of finding the nearest point on the ji interface (with $i \neq j$) at a distance r from a point $\boldsymbol{x}$ in phase j. This statistical descriptor is described by the expression

$$P_j(r) \equiv \left\langle \left\lfloor \frac{1}{V_D} \int_{V_D} \mathcal{I}_j(\boldsymbol{x} + \alpha \boldsymbol{r}) \, dV_D \right\rfloor \right\rangle , \qquad (10)$$

where V_D and dV_D are the volume of a D-dimensional unit sphere and a differential volume, given either by $dV_2 = \alpha d\alpha d\theta$ in a local ($D = 2$)-dimensional domain, or by $dV_3 = \alpha^2 \sin(\theta) \, d\alpha d\theta d\varphi$ in a local ($D = 3$)-dimensional domain, in both cases considering that $\alpha \in [0,1]$, $\theta \in [0, 2\pi]$ and $\varphi \in [0, \pi]$. Particularly in this expression, $\langle \cdot \rangle$ is the expectation value in the domain $\mathcal{V}_j$, $\boldsymbol{x}$ is any position in $\mathcal{V}_j$ and $\boldsymbol{r}$ is a vector of relative position with magnitude r and an arbitrary direction.

In addition, the absence of long-range order allows us to estimate the following values

$$P_j(0) = \frac{s}{\phi_j} , \qquad P_j(\infty) \to 0 . \qquad (11)$$

B. Normalized statistical descriptors

From the previously mentioned statistical descriptors, the normalized statistical descriptors $F_{\beta,j}$, which are functions of the distance r, are defined as

$$F_{1,j}(r) = \frac{\chi_j(r)}{\phi_j(1 - \phi_j)} = \chi_j^*(r) , \qquad (12a)$$

$$F_{2,j}(r) = \frac{L_j(r)}{\phi_j} = L_j^*(r) , \qquad (12b)$$

$$F_{3,j}(r) = \frac{\phi_j P_j(r)}{s} = P_j^*(r) . \qquad (12c)$$

The value of the index β is respectively associated with the type of descriptor: the autocovariance function for $\beta = 1$, the normalized line-path correlation function for $\beta = 2$ and the normalized pore-size distribution function for $\beta = 3$.

C. Local volume fraction

While the volume fraction ϕ_j of phase j is considered to be a global quantity, at a local level the volume fraction of phase j, represented by $\tau_j(\boldsymbol{x}, l)$, depends on the position $\boldsymbol{x}$ and size l of a given observation window Ω_l. Considering an observation window defined by

$$\psi(\boldsymbol{x}, l) = \begin{cases} 1 , & \text{for } \boldsymbol{x} \in \Omega_l \\ 0 , & otherwise , \end{cases} \qquad (13)$$

the local volume fraction reads

$$\tau_j(\boldsymbol{x}, l) = \frac{1}{V_l} \int \mathcal{I}_j(\boldsymbol{z}) \, \psi(\boldsymbol{z} - \boldsymbol{x}, l) \, d\boldsymbol{z} , \qquad (14)$$

where $V_l = V_l(l)$ is the volume of the observation window. Departing from these definitions, the variance of the local volume fraction is recovered with

$$\sigma_j^2(l) = \left\langle \tau_j^2(\boldsymbol{x}, l) \right\rangle - \phi_j^2 , \qquad (15)$$

where $\langle \cdot \rangle$ denotes the average among different positions of the observation window within the global domain. Finally, the coarseness C_j, which is the normalized standard deviation from the global volume fraction of phase j for a given size l of the observation window Ω_l, is defined as

$$C_j(l) = \frac{\sigma_j(l)}{\phi_j} . \qquad (16)$$

This quantity gives a measure of the local fluctuations of the volume fraction [28].

For the particular case of two phases media, we find the following relations

$$\sigma(l) = \sigma_0(l) = \sigma_1(l) , \quad C_0(l) = \left[\frac{\phi_1}{\phi_0} \right] C_1(l) . \quad (17)$$

In addition, an infinitesimally small window and an infinitely large window yield

$$C_1(0) = \frac{\sqrt{\phi_0 \phi_1}}{\phi_1} , \qquad C_1(\infty) \to 0 , \qquad (18)$$

respectively. Finally, considering this properties, a normalized coarseness for a two-phase material is defined as follows

$$C^*(l) = \frac{C_j(l)}{C_j(0)} = \frac{\sigma(l)}{\sqrt{\phi_0 \phi_1}} . \qquad (19)$$

III. METHODOLOGY

Consider two phases planar materials formed by a void phase ($j = 0$) and a solid phase ($j = 1$). A material will be represented by a 2D binary image, a matrix $\mathbf{A}$ with M rows and N columns, representing a discretized binary field. Thus, each element or pixel $A(m, n)$, with $m, n \in \mathbb{Z}^+$ in the ranges $m \in [0, M-1]$ and $n \in [0, N-1]$, presents a value of either 0 or 1.

First, a realization w of the possible states of the material, a configuration with domains $\mathcal{V}_j(w)$, is generated. The result is an original or full-resolution image of the material, represented by the matrix $\mathbf{A}_0 = \mathbf{A}_0(w)$, with $w \in [1, W]$ and $W = 20$ being the total number of random configurations of the same material. From $\mathbf{A}_0$, the normalized statistical descriptors $F_{\beta,j,0}$ are computed and taken as the reference functions of the realization w. Afterwards, the full-resolution image $\mathbf{A}_0$ is subjected to a progressive and sequential decimation process. At the kth decimation stage, with $k \in [1, 8]$, the resulting matrix $\mathbf{A}_k$ is recovered and the corresponding normalized statistical descriptors $F_{\beta,j,k}$ are determined. A comparison of $F_{\beta,j,k}$ with the reference $F_{\beta,j,0}$ leads to estimate the deviation of the specific descriptor β and the phase j, at the k-th decimation stage. As well, the overall error of the decimated image $\mathbf{A}_k$, at the kth stage, with respect to the full-resolution image $\mathbf{A}_0$ is obtained by taking the addition of the deviations of all the descriptors $\beta = 1, 2, 3$ and all the phases $j = 0, 1$.

In this section, details about the generation of full-resolution images, the decimation process and the calculation of the descriptors deviations and the overall error are presented.

A. Full-resolution images

The full-resolution images were constructed following two processes leading to different types of images. An example of each type of the generated images is shown in Figure 1. The first process involves the positioning of disks in two variants: 1a) impenetrable disks, and 1b) partially overlapping disks. The second produces random heterogenous materials (RHM) by coarse graining a matrix of randomly generated values. Different sets of parameters for the coarse graining were employed to obtain RHMs with different values of the surface fractions ϕ_j. All these processes are described in the following.

1. Impenetrable and partially overlapping disks

A number I of disks were placed, which radii have been chosen following a discrete version of the Folded-Normal distribution. The proposed discrete distribution is de-
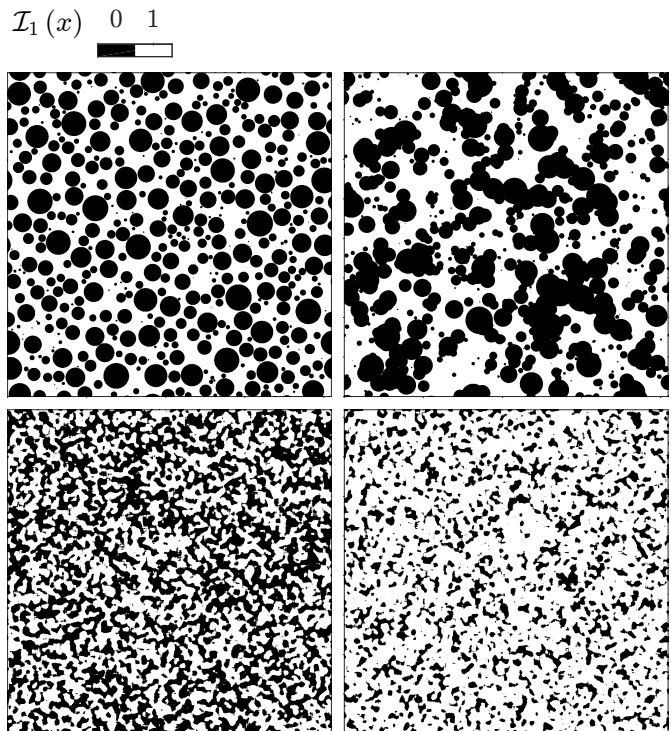
$\mathcal{I}_1(x)$ 0   1

Figure 1 Statistically isotropic heterogenous materials: impenetrable disks (ID - top left), partially overlapping disks (OD - top right) and coarsed grained random heterogenous materials obtained with a Laplacian of Gaussian kernel (LoGK1 - bottom left and LoGK3 - bottom right), for two different values of the level-cut threshold. The corresponding surface fractions, as well as the parameters used for their conformation, are given in Tables I and II.

Table I Average specifications of the two phases planar materials generated with impenetrable (ID) and partially overlapping (OD) disks, using $r_{max} = 250$, $r_{min} = 1$, $\mu = 50$ and $\sigma = 60$ for the radii distribution given by eq. (20).

Type	# of disks I	Surface fraction $\langle \phi_0 \rangle$	$\langle \phi_1 \rangle$	σ_ϕ
ID	491	0.4999	0.5001	7.86×10^{-6}
OD	640	0.5059	0.4941	6.11×10^{-3}

scribed by

$$\mathcal{P}_r\{r_t\} = \frac{1}{I}\left\lfloor I\sqrt{\frac{2}{\pi\sigma^2}}\exp\left(-\frac{r_t^2+\mu^2}{2\sigma^2}\right)\cosh\left(\frac{r_t\mu}{\sigma^2}\right)\right\rceil,$$

(20)

where $\lfloor \cdot \rceil$ indicates the application of the nearest integer function, μ and σ are the location and scale parameters, and $r_t = r_{min} + t$ is the discrete radius of bin t, which range is $t \in [0, r_{max} - r_{min}]$ and with r_{max} and r_{min} being the maximum and minimum radii, respectively. Therefore, the radius r_i of the ith disk is randomly chosen to be equal to some radius r_t.

The space region $\mathbb{R}^2$ in which the disks can be placed

Table II Average specifications of the two phases planar materials generated by coarse graining with a Laplacian of Gaussian kernel (LoGK) of size $a = 75$, specified in eqs. (24).

LoGK	Threshold	Surface fraction		
#	a_0	$\langle\phi_0\rangle$	$\langle\phi_1\rangle$	σ_ϕ
1	127.50	0.4969	0.5031	6.81×10^{-3}
2	127.00	0.3272	0.6728	5.63×10^{-3}
3	126.75	0.2475	0.7525	6.14×10^{-3}

has been chosen to be a square, of area L^2, within the ranges $x \in [0, L]$ and $y \in [0, L]$. For the ith disk, its center coordinates (x_i, y_i) have been selected randomly within the ranges $x_i \in [0, L]$ and $y_i \in [0, L]$, surveying that the distance between its center and that of the previously placed disks, given by

$$d = \sqrt{(x_h - x_i)^2 + (y_h - y_i)^2} \ , \quad \text{for } 1 \le h < i \ , \quad (21)$$

remains with a value

$$
\begin{aligned}
d &\ge (r_h + r_i) \ , &\text{for impenetrable disks ,} \quad (22a) \\
d &> |r_h - r_i| \ , &\text{for partially overlapping disks .} \quad (22b)
\end{aligned}
$$

The matrix $\mathbf{A}_0$ with equal number of rows and columns, $M_0 = N_0 = 4096$ is created with pixel values $A_0(m, n) = 0$ for all $m \in [0, M_0 - 1]$ and $n \in [0, N_0 - 1]$. Now, considering the coordinates of each pixel $A_0(m, n)$, being given by $x_{mn} = n + 1/2$ and $y_{mn} = m + 1/2$, the value of the pixel, which is initially assigned to be $A_0(m, n) = 0$, may change according to the following statement: if $\sqrt{(x_{mn} - x_i)^2 + (y_{mn} - y_i)^2} \le r_i$ then $A_0(m, n) = 1$; otherwise $A_0(m, n)$ keeps its previous value. Additionally, periodic boundary conditions were considered in both x and y directions. The process is repeated for all the disks, from $i = 0$ to $i = I - 1$, and the binary matrix $\mathbf{A}_0$ is obtained.

Following abovementioned process, a total of $W = 20$ random configurations of impenetrable disks (ID) and partially overlapping disks (OD), which specifications are presented in table I, have been gathered. The number of disks I and average surface fractions $\langle\phi_0\rangle$ and $\langle\phi_1\rangle$ for phase $j = 0$ and $j = 1$, respectively, and their standard deviation σ_ϕ are given in the abovementioned table.

2. Random heterogeneous material

A matrix $\mathbf{A}_0^*$ is created with an equal number of rows and columns, $M_0 = N_0 = 4096$. An intensity, given by the discrete random variable $a \in [0, 255]$ with a uniform probability function $\mathcal{P}_a\{a\} = 1/256$, is assigned to each pixel $A_0^*(m, n)$, with $m \in [0, M_0 - 1]$ and $n \in [0, N_0 - 1]$. A blurred matrix $\mathbf{A}_0^\square$ is obtained by performing the convolution of the matrix $\mathbf{A}_0^*$ and a square-shaped kernel $\mathcal{K}$:

$$A_0^\square(m, n) = \sum_h \sum_i \mathcal{K}(h, i, b) A_0^*(m + h, n + i) \quad (23)$$

where b is the size of the kernel, an odd positive integer $\mathbb{Z}^+$, which yields the ranges $h \in [-\lfloor b/2 \rfloor, \lfloor b/2 \rfloor]$ and $i \in [-\lfloor b/2 \rfloor, \lfloor b/2 \rfloor]$. Herein, the discrete Laplacian of Gaussian kernel (LoGK), described by

$$\mathcal{L}(h, i, b) = \left(1 - \frac{h^2 + i^2}{2[b/2]^2}\right) \exp\left(-\frac{h^2 + i^2}{2[b/2]^2}\right) \ , \quad (24a)$$

$$\mathcal{K}(h, i, b) = \frac{\mathcal{L}(h, i, b)}{\displaystyle\sum_h \sum_i \mathcal{L}(h, i, b)} \ , \quad (24b)$$

has been employed. Finally, a binary matrix $\mathbf{A}_0$ is obtained by level-cutting the matrix $\mathbf{A}_0^\square$, using a threshold value a_0 as follows

$$A_0(m, n) = \begin{cases} 0 & \text{if } A_0^\square(m, n) < a_0 \ , \\ 1 & \text{if } A_0^\square(m, n) \ge a_0 \ . \end{cases} \quad (25)$$

Both, the threshold value, which is restricted to the range $a_0 \in [0.0, 255.0]$, and the kernel size b are directly responsible for the structural characteristics of the binary matrix $\mathbf{A}_0$.

Following abovementioned process, two sets of parameters were proposed to generate two different types of LoGK coarsed grained RHMs, both with $b = 75$ but differing in their value of a_0. A total of $W = 20$ random configurations of each LoGK material, which specifications are presented in table II, have been gathered. The kernel size b, threshold value a_0 and average surface fractions $\langle\phi_0\rangle$ and $\langle\phi_1\rangle$ for phase $j = 0$ and $j = 1$, respectively, and their standard deviation σ_ϕ are presented.

B. Decimation

As it has been previously mentioned, a progressive and sequential decimation process has been implemented, starting from the step $k = 1$ and increasing k by one integer, until the final step K is performed. At each kth decimation step, the matrix $\mathbf{A}_{k-1}$, with M_{k-1} rows and N_{k-1} columns, evolves towards a new matrix $\mathbf{A}_k$, where the number of rows and columns is reduced to $M_k = M_{k-1}/2$ and $N_k = N_{k-1}/2$. In other words, each k-decimation step provokes a reduction of the number of rows and columns to $M_k = M_0/(2^k)$ and $N_k = N_0/(2^k)$, respectively. In the present work, defining $K = 8$ and taking $M_0 = N_0 = 4096$ leads to images at the final decimation step of dimensions $M_K = N_K = 16$.

Each decimation process is performed by the application of one of the three methods: random, bilinear or bicubic decimation [29, 30]. An example of the decimation process, in particular with the application of the random method, is shown in Figure 2. The three decimation methods are briefly explained in the following.

1. Random

The element in the mth row and nth column of the matrix $\mathbf{A}_k$ is obtained by performing the kth decimation

a) $k = 1$
$N_1 = M_1 = 2048$

b) $k = 2$
$N_2 = M_2 = 1024$

c) $k = 3$
$N_3 = M_3 = 512$

d) $k = 4$
$N_4 = M_4 = 256$

e) $k = 5$
$N_5 = M_5 = 128$

f) $k = 6$
$N_6 = M_6 = 64$

g) $k = 7$
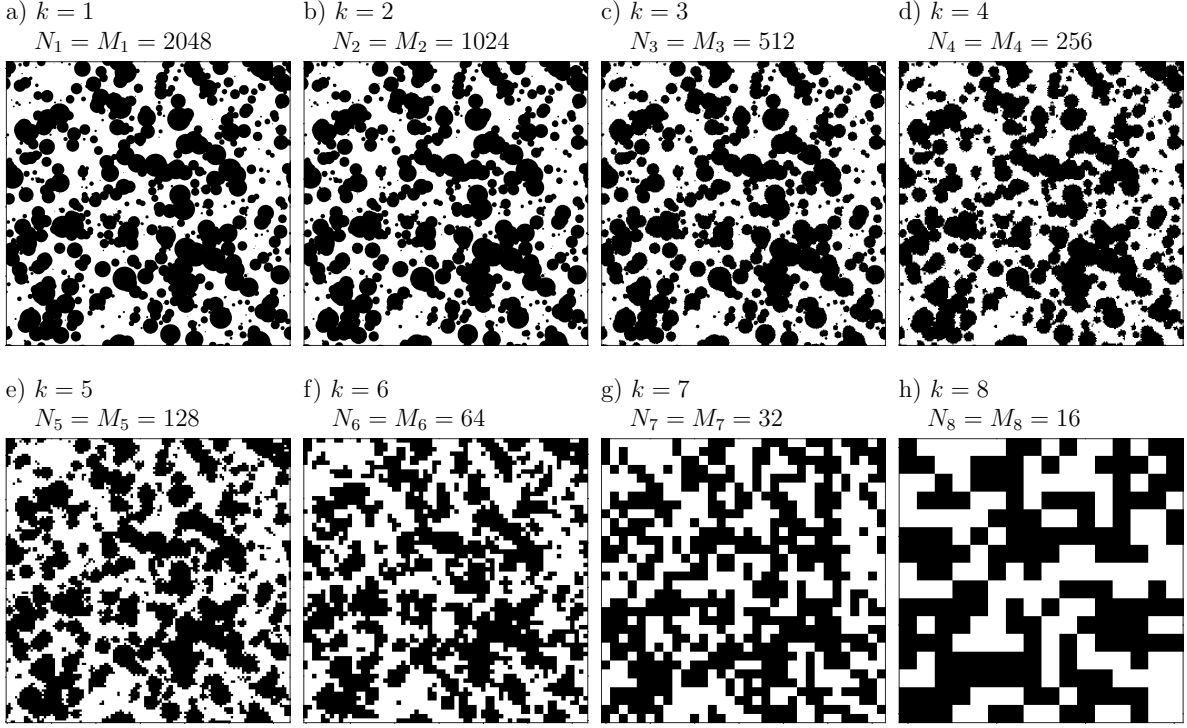$N_7 = M_7 = 32$

h) $k = 8$
$N_8 = M_8 = 16$

Figure 2 Consecutive stages [from a) $k = 1$ to h) $k = 8$] following a random decimation process, departing from the full resolution image ($k = 0$ with $N_0 = M_0 = 4096$) of partially overlapping disks presented in Figure 1.

step expressed by

$$A_k (m, n) = A_{k-1} (2m + \xi_m, 2n + \xi_n) \ , \qquad (26a)$$

with

$$\xi_m = \lfloor \xi/2 \rfloor \ , \qquad \xi_n = \xi - 2 \lfloor \xi/2 \rfloor \ , \qquad (26b)$$

where $\lfloor \cdot \rfloor$ indicates the application of the floor function, ξ is a discrete random variable in the range $\xi \in [0, 3]$, with a discrete uniform distribution $\mathcal{P}_\xi \{\xi\} = 1/4$.

2. Bilinear

The element in the mth row and nth column of the matrix $\mathbf{A}_k$ is the resulting value from a four elements average, followed by a rounding operation, described by the kth decimation step formula

$$A_k (m, n) = \begin{cases} 0 & \text{if } \alpha_{mn} < 1/2 \\ 1 & \text{if } \alpha_{mn} \geq 1/2 \end{cases} \ , \qquad (27a)$$

with

$$\alpha_{mn} = \frac{1}{4} \sum_{i=0}^{1} \sum_{j=0}^{1} A_{k-1} (2m + i, 2m + j) \ . \qquad (27b)$$

3. Bicubic

The element in the mth row and nth column of the matrix $\mathbf{A}_k$ is the outcome of the bicubic interpolation,

performed in the kth decimation step, described by the following expressions

$$A_k (m, n) = \begin{cases} 0 & \text{if } \alpha_{mn} < 1/2 \\ 1 & \text{if } \alpha_{mn} \geq 1/2 \end{cases} \ , \qquad (28a)$$

with

$$\alpha_{mn} = G (q_0, q_1, q_2, q_3) \ , \qquad (28b)$$

where, the function G is the linear combination of its arguments given by

$$G (g_0, g_1, g_2, g_3) = \frac{1}{16} [-g_0 + 9g_1 + 9g_2 - g_3] \ . \qquad (28c)$$

Herein, $q_0 = G(A_{k-1})$ is the value obtained from the application of the function G to the four elements of A_{k-1} in the columns from $2n - 1$ to $2n + 2$ and in the $(2m - 1)$th row. In turn, $q_1 = G(A_{k-1})$, $q_2 = G(A_{k-1})$ and $q_3 = G(A_{k-1})$ are obtained equivalently from the same columns of A_{k-1}, but for the $(2m)$th, $(2m + 1)$th and $(2m + 2)$th rows, respectively. The previously formulas are valid to obtain the values of $\{q_0, q_1, q_2, q_3\}$ when $m \in [1, M_{k-1} - 2]$ and $n \in [1, N_{k-1} - 2]$. During the bicubic decimation procedure, particular attention has to be made for the borders of the image, for which $m = 0$, $m = M_{k-1} - 1$, $n = 0$ or $n = N_{k-1} - 1$, as it is explained in detail elsewhere (see Supplementary Material).
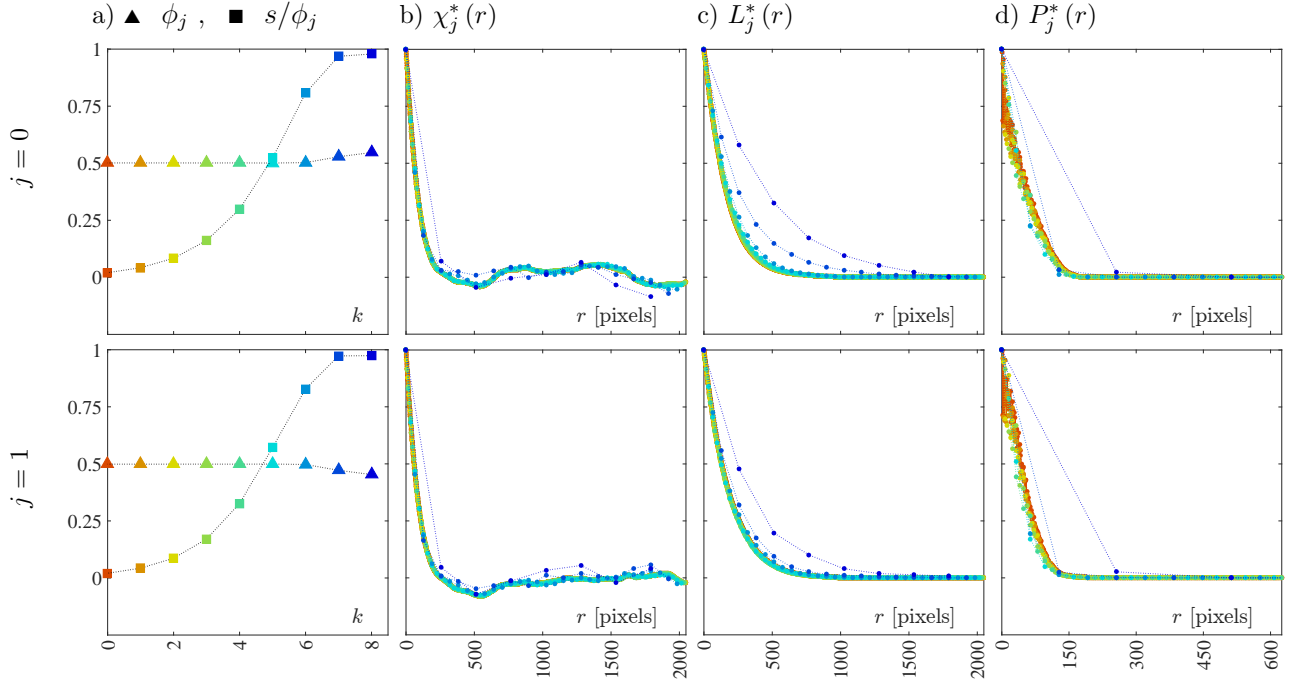
Figure 3 Statistical descriptors along the random decimation process presented in Figure 2. Top and bottom rows represent the descriptors for phases $j = 0$ and $j = 1$ respectively, whereas each column indicates the type of descriptor: a) surface fraction ϕ_j and specific interface area per phase surface fraction s/ϕ_j as functions of the k-th decimation step, b) normalized autocovariance function $F_{1,j} = \chi_j$, c) normalized line-path correlation function $F_{2,j} = L_j$ and d) normalized pore-size distribution function $F_{3,j} = P_j$. The statistical descriptors in columns b)-d) are presented as functions of the distance r. Each color of the points in column a) and the curves in columns b)-d) correspond to a different k-th decimation step, given by the horizontal position of the point in column a).

C. Discrete normalized statistical descriptors

For all the types of materials, disks and RHMs, the computation of the statistical descriptors $\{\chi_j, L_j, P_j\}$ of all the images, full-resolution $\mathcal{A}_0$ and decimated $\mathcal{A}_k$, was performed following a well-know method proposed for isotropic digitized systems [3, 4]. Under the prescribed procedure, the distance is measured in units of pixels, and thus the retrieved distance data is represented by the discrete variable r_l, that contains only integral values, with the range $l \in \left[1, N_k^\dagger\right]$ and considering the limit $N_k^\dagger \approx N_k/2$ to avoid the effects of periodic boundaries or finite size domains. Therefore, the pixel size of each image must be taken into account in order to recover r, an invariant distance variable, from $r_{k,l}$, the distance variable associated with the kth decimation step. For the decimated image $\mathcal{A}_k$, the pixel size is given by $\mathcal{L}_k = N_0/N_k$, whereas the reference scale $\mathcal{L}_0 = 1$ is the pixel size in the full-resolution image $\mathcal{A}_0$. Consequently, one finds that $r = \mathcal{L}_k r_{k,l}$ for the kth decimation step and $r = r_{0,l}$ for the full-resolution image.

Afterwards, the normalization of the statistical descriptor was carried out according to eqs. 12. The notation for each descriptor was adapted to account for the decimation process, i.e. $F_{\beta,j,k}(r)$ denotes the normalized statistical descriptor β for the phase j, obtained at the

kth decimation step. In Figure 3, an example of the evolution of the normalized statistical descriptors is shown, from the full-resolution image to the last decimation step, corresponding to the images depicted in Figure 2.

D. Deviation from the full-resolution image

As it can be discerned from the example presented in Figure 3, the quality of the statistical descriptors reflects the information that is lost along the decimation process. Due to the normalization of the statistical descriptors, the aforementioned quality can be separated into two components: 1) a constant shift on the surface fraction ϕ_j and the interfacial area per unit volume s/ϕ_j of the corresponding phase, and 2) a shape discrepancy with respect to the trend displayed by the statistical descriptors of the full-resolution image. The former is directly measured from the difference between the full-resolution value and the surface fraction at the kth decimation step. The latter is quantified by computing a general deviation, i.e. the square of the difference between the statistical descriptors of $\mathcal{A}_0$ and $\mathcal{A}_k$, both at the same distance $r = \mathcal{L}_k r_{k,l}$, and then taking the average over all the distances. This deviation of the descriptor β of phase j and from the image generated at the kth decimation step, with respect to
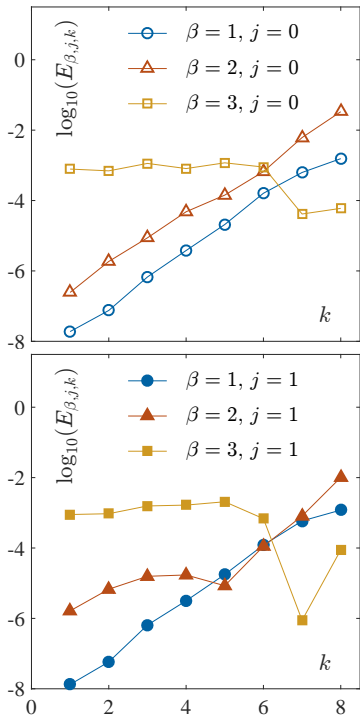
Figure 4 Deviation of statistical descriptors, with respect to the full-resolution image, as a function of the k-th step along the random decimation process, presented in Figure 2. These deviations were obtained by applying Eq. (29). The subindices β and j correspond to the normalized statistical descriptor and phase, respectively, specified in section II B.

the full-resolution image, is defined as follows

$$E_{\beta,j,k} = \frac{1}{N_k^\dagger} \sum_{l=1}^{N_k^\dagger} \left\{ F_{\beta,j,0}\left(\mathcal{L}_k\, r_{k,l}\right) - F_{\beta,j,k}\left(\mathcal{L}_k\, r_{k,l}\right) \right\}^2 .$$
(29)

In Figure 4, computed for both phases of the OD material which decimation is displayed in Figure 2, the corresponding deviation from the full-resolution image is presented.

Moreover, the sum over the three statistical descriptors and the two phases gives rise to an overall (global) error for the kth decimation step

$$\mathcal{E}_k = \sum_{\beta=1}^{3} \sum_{j=0}^{1} E_{\beta,j,k} .$$
(30)

This global error was calculated for each of the realizations, leading us to define an ensemble average and the corresponding standard deviation

$$\langle \mathcal{E}_k \rangle = \frac{1}{W} \sum_{w=1}^{W} \mathcal{E}_k\left(w\right) ,$$
(31a)

$$\sigma_{\mathcal{E}_k} = \sqrt{\frac{1}{W} \sum_{w=1}^{W} \mathcal{E}_k^2\left(w\right) - \langle \mathcal{E}_k \rangle^2}$$
(31b)

which gathers the statistical information, over the $W = 20$ configurations of each material, of the deviation observed with respect to the full-resolution images. In Figure 5, the ensemble average is shown as a function of the decimation step, while the corresponding standard deviation appears as error bars. The interpretation of the data presented in Figures 2–5 is the central part of our analysis, thus more details and an interpretation of these results will be presented and discussed in the following section.

E.   Characteristic length-scale

In addition, the ensemble average of each normalized statistical descriptor $\langle F_{\eta,j,0} \rangle$ is computed for the full-resolution images $\mathcal{A}_0$

$$\langle F_{\beta,j,0} \rangle\left(r\right) = \frac{1}{W} \sum_{w=1}^{W} F_{\beta,j,0}\left(r\right) ,$$
(32)

Once more, as an example, the resulting data for the partially overlapping disks is shown in Figure 6.

Some useful information can be extracted from these ensemble average descriptors. For instance, one can define a correlation length $\ell_{\beta,j}$ as the range over which the descriptor β of the phase j approaches, to a certain extent, the horizontal axis $\langle F_{\beta,j,0} \rangle = 0$ for the first time. Indeed, one can define the approach extent as a margin of width 2% of the maximum value of the function (0.02 since the correlation functions had been normalized) around the horizontal line $\langle F_{\beta,j,0} \rangle = 0$. In general, for statistically homogeneous two phases materials, $[0, \ell_{\beta,j}]$ indicates the range over which the statistical descriptor $F_{\beta,j,0}$ is non-negligible. Consequently, one can take either the average or any order statistic (of the two phases and the three statistical descriptors) as a characteristic length-scale ℓ of the structural features of a given material, and the observed details on these features depend on the number of pixels that are used to depict a region of size ℓ. Further insight on the adequate choice of the characteristic length-scale ℓ will be given in the results section of this manuscript.

F.   Loss of statistical information

The application of a decimation process to an image is always ensued by a loss of information. In the case of a two-phase material configuration, the details below the size of the new pixels are lost, and this effect is reflected on the quality of the information given by the corresponding statistical descriptors.

Consider the variance of the local surface fraction within a square observation window with the size of a pixel at the k-th decimation step, which surface within the full-resolution image corresponds to an area of 2^{2k}. One should recall and adjust Eqs. (13–15), for which $\boldsymbol{x}$
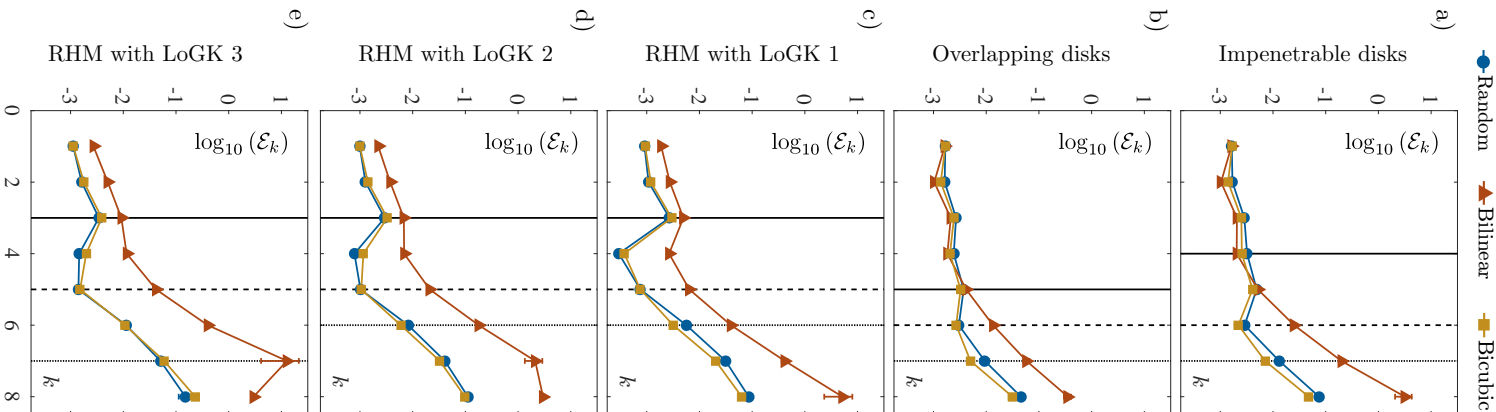
Figure 5 Ensemble average global error $\langle \mathcal{E}_k \rangle$, for each material described in Tables I and II, as a function of the k-th step along the 3 proposed decimation processes. The average and standard deviations (error bars) were obtained by applying Eqs. (31). The statistically sensitive decimation step Z was computed with Eqs. 34, applied to the extrema, [—] $\min(\ell_{\beta,j})$ and [··] $\max(\ell_{\beta,j})$, and the average, [- -] $\ell_{\beta,j}$, of the correlation lengths of the different descriptors β and phases j.

a) Impenetrable disks
b) Overlapping disks
c) RHM with LoGK 1
d) RHM with LoGK 2
e) RHM with LoGK 3

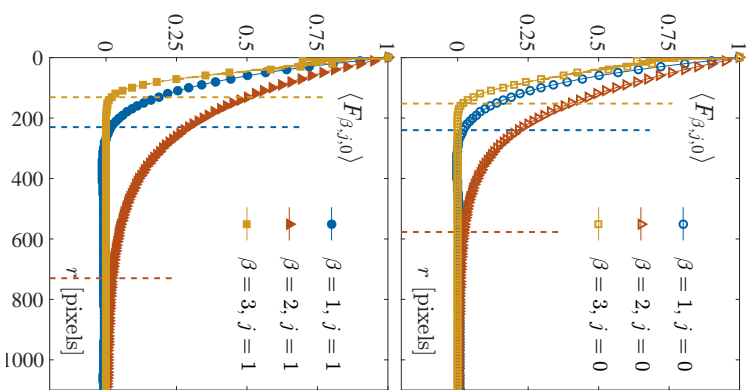Random — Bilinear — Bicubic

$\log_{10}(\mathcal{E}_k)$

Figure 6 Average statistical descriptors of the original images (with $k = 0$, $M_0 = N_0 = 4096$) of the $W = 20$ different configurations of partially overlapping disks, with the same characteristics as the configuration presented in the top row, right column of Figure 2. Each average statistical descriptor was computed with Eq. (32) and its corresponding correlation length is shown as a vertical dashed line. The subindices β and j correspond to the normalized statistical descriptor and phase, respectively, specified in section IIB. For better visualization, only 1/10 of the data points are shown.

now indicates the center of the square observation window, with a side-length $l = 2^k$. If, at each decimation step, the normalized coarseness, given by

$$C_k^* = C^* \left(2^k\right) = \frac{\sigma \left(2^k\right)}{\sqrt{\phi_0 \phi_1}}, \qquad (33)$$

is computed for an observation window of the size that a pixel of the reduced-size image should cover on the full-resolution image, then the amount of information that we are losing at each decimation step is related to the complement of the normalized coarseness $1 - C_k^*$. Therefore, the closer is C_k^* to one, the more information the k-th step decimated image retains from the original full-size image. In contrast, if C_k^* approaches zero, most of the information has been lost at the k-th decimation step. In Figure 7, the evolution of the normalized coarseness is shown as a function of the k-th step along the random decimation process, for all the materials generated for this study. The behavior of C_k^* is discussed in the following section.
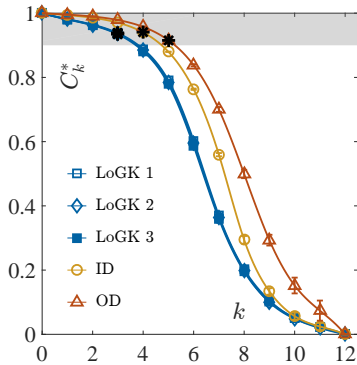
Figure 7 Normalized coarseness (average and standard deviation of $W = 20$ configurations) as a function of the k-th step along the decimation process, for each material described in Tables I and II. The curves were generated with the use of Eq. (33). Note that the curves corresponding to the LoGK configurations overlap. The $*$ symbol at each curve indicates the optimal decimation step Z computed from its characteristic length-scale $\ell = \min(\ell_{\beta,j})$ and Eqs. (34). The gray surface indicates the region for which $C_k^* > 0.9$.

IV. RESULTS

A total of $W = 20$ random configurations for each of the five different types of materials, which characteristics are described in Tables I and II, were generated. Each configuration is represented by a full-resolution image, to which the methodology described in the first paragraph of section III was applied. Along the first half of this section, the described results correspond specifically to the configuration of partially overlapping disks presented in Figure 1, which is subjected to a random decimation process. Nevertheless, the whole analysis is valid for all the materials and all the decimation methods.

Each time a decimation step is performed, and consequently the number of pixels decreases, some information about the configuration of the phases is lost. This effect is clearly observed in Figure 2, where the characteristic structures of the material are distinguishable during most of the decimation process, for $k < 6$. Despite the loss of details and the apparent increase of interfacial roughness, one can still identify the $k = 0$ configuration shown in the full-resolution image (top right configuration in Figure 1) within the decimated image $k = 5$ (Figure 2e). A further step $k \geq 6$ in the decimation process (Figure 2f-k) complicates the recognition of the original features of the material. This qualitative analysis is supported by the quantitative results presented in Figure 3. For instance, as it is shown in Figure 3a, the surface fractions ϕ_0 and ϕ_1 are nearly constant for $k < 6$. For $k \geq 6$, ϕ_0 increases and ϕ_1 decreases, since $\phi_1 = 1 - \phi_0$, both with a monotonous and moderate behavior.

In contrast, the interface between the two phases is significantly modified along the whole decimation process. From being smooth (up to the pixel resolution) for $k = 0$, as the decimation process goes on, the interface becomes

sharper, as the pixel size grows. The rounded and well defined boundaries of the phase $j = 0$ presented in Figure 1, become sharp and rough, more strongly for $k \geq 4$. This is exhibited by the monotonic increase of s/ϕ_0 and s/ϕ_1 observed in Figure 3a, which presents its steepest slope around $k = 5$. For the presented case, since $\phi_0 \sim \phi_1$ and both being close to 0.5 despite the decimation stage, the curves for s/ϕ_0 and s/ϕ_1 are quite similar, but in a general situation, for which $\phi_0 \nsim \phi_1$, their behavior is generally different from each other.

Additionally, when the number of pixels used to describe the material configuration decreases, the normalized statistical descriptors evolve. For instance, the normalized autocovariance $F_{1,j,k} = \left[\chi_j^*\right]_k$ and line-path $F_{2,j,k} = \left[L_j^*\right]_k$ functions throughout the decimation process, depicted in Figure 3, separate from the corresponding trends indicated by the correlation functions of the full-resolution image $F_{\beta,j,0}$ with $\beta = 1, 2$ respectively. This discrepancy from the reference curve $F_{\beta,j,0}$ is negligible for $k \leq 6$, but becomes significant and increasingly exposed as k grows beyond $k = 6$. The effect of the decimation on the normalized pore-size distribution function $F_{3,j,k} = \left[P_j^*\right]_k$ is not discerned easily as compared with the two previously mentioned functions. This is manly due to the fact that $F_{3,j,k}$ presents a noisy signal at small distances $r < 150$ pixels, due to the pixelization of the images and the consequent discretization of the correlation functions. In addition, for considerably decimated images ($k \geq 6$), the gradual loss of resolution in the curves hinders the rift from the reference function $F_{3,j,0}$.

The quantitative deviations $E_{\beta,j,k}$ of the statistical descriptor $F_{\beta,j,k}$ from the reference descriptor $F_{\beta,j,0}$, for the functions $\beta = 1, 2, 3$ and the phases $j = 0, 1$, throughout the random decimation process, are computed by means of Eq.(29) and represented in Figure 4, for the configuration of partially overlapping disks (top-right sample in Figure 1 and Figure 2). For $\beta = 1$ and $\beta = 2$ (corresponding to χ_j^* and L_j^*, respectively), the general trend of the deviations $E_{\beta,j,k}$ is an exponential growth in terms of the decimation step k. In contrast, the deviation for $\beta = 3$ (corresponding to P_j^*) is nearly constant along the first steps of the decimation process, due to its noisy behavior for $k \leq 6$, whereas for $k > 6$, the poor resolution of the statistical descriptor provokes a magnitude drop of the deviation $E_{3,j,k}$.

For each of the materials described in Tables I and II, the full-resolution image was gradually decimated and the deviations of each descriptor β of the phase j were computed, at each kth decimation step. Using Eq. 30, the global error was recovered, and then the evolution of the ensemble average $\langle \mathcal{E}_k \rangle$ and the standard deviation $\sigma_{\mathcal{E}_k}$, in terms of the step k, were obtained with Eqs. 31. The corresponding results are presented in Figure 5. Despite the particular characteristics of each material and the applied decimation method, a general trend is observed: an almost constant value of $\langle \mathcal{E}_k \rangle$ along the first decimation steps, followed by a nearly exponential growth of $\langle \mathcal{E}_k \rangle$ when k increases. The standard deviation $\sigma_{\mathcal{E}_k}$ is barely perceptible in the presented semi-log graph for $k < 6$,

since its order of magnitude is much smaller than that of the average value. It is also interesting to notice that, for small values of k where $\mathcal{E}_k$ is almost constant, the trend followed by the global error is nearly independent of the decimation method. Nevertheless, one realizes that the random and bicubic methods produce lower errors than the bilinear method.

In Figure 6, the ensemble average of the statistical descriptors are displayed, which were obtained from $W = 20$ configurations represented by full-resolution images of partially overlapping disks. The typical behavior of statistically homogeneous and isotropic materials is observed: monotonically decreasing functions with the values of $\langle F_{\beta,j,0} \rangle = 1$ at $r = 0$ pixels and $\langle F_{\beta,j,0} \rangle = 0$ at $r \to 0$ pixels, due to the corresponding normalization stage. From each curve, the correlation length $\ell_{\beta,j}$ is determined, as stated in subsection III E, and displayed in Figure 6.

Now, if one may accurately describe the descent of a statistical descriptors, for instance with a cubic spline, the sampling spatial rate must be at least $3/\ell_{\beta,j}$. This means that the distance between two consecutive pixels in the decimated image should not be larger than $r = \ell_{\beta,j}/3$, a length measured in units of pixels of the full-resolution image. Hence, the gradual decimation process should be stopped at the optimal decimation step Z, given by

$$Z = \left\lfloor \min \left\{ \log_2 \left(\frac{M_0}{M_Z} \right) , \log_2 \left(\frac{N_0}{N_Z} \right) \right\} \right\rfloor , \quad (34\text{a})$$

$$M_Z = \left\lceil \frac{3M_0}{\ell_{\beta,j}} \right\rceil , \qquad N_Z = \left\lceil \frac{3N_0}{\ell_{\beta,j}} \right\rceil , \quad (34\text{b})$$

in order to provide a good statistical representation, via the statistical descriptor $F_{\beta,j,k}$, of the full-resolution image. Briefly, $k = Z$ is the decimation step up to which one may proceed to reduce the size of the full-resolution image, preserving an acceptable amount of statistical information in the descriptor $F_{\beta,j,k}$ with respect to the reference function $F_{\beta,j,0}$.

Additionally, the extrema and the average values of $\ell_{\beta,j}$ were deduced and the corresponding decimation step Z, for each statistic, was computed with Eqs .(34). The different values of Z are included in Figure 5, for the materials analized in this study. In all the cases, it is clearly observed that the step Z obtained with $\min(\ell_{\beta,j})$ corresponds to a decimation step just before the rise of the average global error $\langle \mathcal{E}_k \rangle$ occurs, whereas the ones obtained with $\overline{\ell_{\beta,j}}$ and $\max(\ell_{\beta,j})$ are sequentially located at the already developped increasing trend of $\langle \mathcal{E}_k \rangle$. Thence, in order to limit the deviation from the original statistical information by preserving the behavior of all $F_{\beta,j,0}$, it is reasonable to chose the minimum $\ell = \min(\ell_{\beta,j})$ as the characteristic length-scale. In this way, ℓ corresponds to the maximum range over which all the normalized statistical descriptors $F_{\beta,j,k}$ are non-negligible. Not surprisingly, in most cases, the minimum value of $\ell_{\beta,j}$ corresponds to the correlation length of the normalized pore-size distribution function $F_{3,j,0} = P_j^*$.

In Figure 7, the coarseness C_k^* is shown as a function of the k-th decimation step. A reverse sigmoid function describes accurately the behavior of the curves for all the materials. For the full-size image, the normalized coarseness is $C_0^* = 1$, since the complete statistical information is available for the reference image. As it was expected thanks to the qualitative results, most of the statistical information is preserved for the first decimation steps, which extent depends on the particular material, and the coarseness remains in the range $0.9 < C_k^* < 1$. This is the case of the selected optimal number of decimation steps Z, obtained with $\ell = \min(\ell_{\beta,j})$ and Eqs. 34, which coarseness is $C_k^* > 0.9$. Following the decimation process, the coarseness is abruptly diminished as k increases, slowing down its descent as k approaches its last stage $k = 12$, finally losing the statistical information when $C_k^* = 0$.

V. ALGORITHM

From the presented results, and mainly due to the behavior of the global error $\mathcal{E}_k$ and the coarseness C^* with regards on the characteristic length-scale ℓ, we find that Z is the last decimation step before a significant rise of the global error $\mathcal{E}_k$ and an important loss of statistical information occurs. Therefore, according to the gathered experience, a statistically sensitive decimation process is proposed, consisting in the following sequence of steps:

1. Compute the statistical descriptors $F_{\beta,j,0}$ of the full-resolution image $\mathbf{A}_0$.

2. Obtain the correlation lengths $\ell_{\beta,j}$ for each statistical descriptor β and each phase j.

3. Determine the characteristic length-scale as $\ell = \min(\ell_{\beta,j})$.

4. Compute the optimal number of decimation steps Z using Eqs. (34).

5. Apply Z sequential steps of the selected gradual decimation process (random, bilinear or bicubic) to generate the decimated image $\mathbf{A}_Z$.

This algorithm ensures the reduction in size of the image from $A_0(M_0, N_0)$ to $A_Z(M_Z, N_Z)$, with $M_0 > M_Z$ and $N_0 > N_Z$, while preserving the quality of the information stored in the statistical descriptors of the final image $\mathbf{A}_Z$ with respect to the original image $\mathbf{A}_0$.

VI. APPLICATION ON SEM IMAGES

To validate the presented analysis and the statistically sensitive decimation process, proposed in this study, we tested the methodology on images representing real material samples, which were obtained by SEM.

A sample of the catalytic layer of a Proton Exchange Membrane Full Cell (PEMFC) was scanned and digitized with a conventional SEM, retrieving a 768×768 grayscale

image. This image was binarized, following the procedure given in a previous study [31], yielding a binary image of the same size as the grayscale one. Afterwards, the algorithm that has been exposed in the previous section was applied, and a decimated image with optimal size was obtained after $k = Z = 2$ decimation steps. An additional decimation step $k = 3$ was performed, in order to verify that our methodology yields the desired results. In Figure 8, the original SEM image, the full-size binary image and the reduced-size images after $Z = 2$ and $k = 3$ decimation steps are presented with their corresponding normalized statistical descriptors. The details on the full-size image (reference) are replicated by the $Z = 2$ decimated image, which is supported by the almost indistinguishable difference between each of the statistical descriptors (χ_j^*, L_j^* and P_j^*) computed from the two images, for both phases. In contrast, despite the likeness of the $k = 3$ decimated image with the full-size image, fine details had disappeared, mainly at the interface where smoothness has been lost. This is reflected on the statistical descriptors, essentially on the line-path and the pore-size distribution functions of the $j = 1$ phase, which separate substantially from their corresponding reference descriptors. These qualitative insights are verified by the quantitative information presented in Figure 9, where the deviation of each descriptor is studied along the entire decimation process from the first step $k = 1$ to a maximum and final step $k = 7$. The evolution of the surface fractions ϕ_j and the specific interface area per phase surface fraction s/ϕ_j, for each phase j are shown. This values indicate the shift and dilation of the curves representing the statistical descriptors of the k-th decimation step, with respect to the reference descriptors. One observes a reduction of the amount of scarce phase ϕ_0, while for the abundant phase ϕ_1 increases, according to the relation $\phi_0 + \phi_1 = 1$, as the decimation process evolves. In turn, s/ϕ_0 grows from its initial value close to 0.25 and saturates to 1 for $k \geq 4$, whereas $s\phi_1$ increases from a close to zero value towards a maximum at $k = 5$ and then drops to zero at the end of the decimation process. The variation of the deviation in shape $E_{\beta,j,k}$ of each descriptor β and phase j at the k-th decimation step is also reported. The deviations of χ_j^* and L_j^* show an increase as the decimation process advances, with a faster tendency for the scarce phase $j = 0$ than for the abundant phase $j = 1$. P_0^* shows an initial descent, but above $k = 4$ it becomes negligible due to the reduced amount of data and resemblance with the reference curve. P_1^* presents a significant rise in magnitude as k advances, which is comparable with the trend shown by χ_1^* and L_1^*. As a consequence of the described behaviors of all the statistical descriptors, the overall error $\mathcal{E}_k$ displays a nearly constant value in the range $1 \leq k \leq 3$, and a sudden exponential growth for $k > 3$. From the correlation lengths of each statistical descriptor, which are shown in Figure 8, the optimal number of decimation steps $Z = 2$ was found. This value $Z = 2$ points at a decimation step before the sudden rise of the global error occurs, as it was expected thanks to the analysis presented in the previous sections of this manuscript.

The statistically sensitive decimation process exposed in this study should be applied directly to full-resolution images with numbers of rows and columns which are $M_0 \propto 2^Z$ and $N_0 \propto 2^Z$, respectively. If it is not the case, an interpolation method at each decimation step must be implemented. In any case, a gradual size reduction up to Z decimation steps can be applied, being confident of the preservation of the statistical information, whenever the number of decimation steps Z has been obtained by following the methodology hereinabove described.

VII. CONCLUSIONS

In this manuscript, a gradual decimation process for the size reduction of images, that represent two-phases material configurations, has been developed. We present a methodology to define the maximum number of stages of the gradual decimation process that one can apply to an image, in order to maintain the statistical information, but to reduce its amount. It is convenient to reduce the size of an image to perform calculations in limited computing resources, time and memory capacities. For instance, the reconstruction of 3D representations of random heterogeneous materials requires a memory allocation space of N^3 bits and approximately N^{3x} operations, where x represents an exponent that encloses the combined effect of the number of statistical descriptors employed to characterize the original 2D sample and the computational cost of calculating them all. Therefore, it is essential to reduce N as much as possible. As it has been tested in this study, an unprocessed SEM image of around 800^2 pixels will be transformed into a $800^3 = 5.12 \times 10^8$ pixels 3D-reconstruction image. The reduction of size after $k = 2$ decimation steps, without any substantial loss of statistical information, allows us to manage a 3D-reconstruction image of only $200^3 = 8 \times 10^6$ pixels, and to reduce the number of operations from 10^{8x} to 10^{6x}.

The proposed normalization of the statistical descriptors allows us to decouple the deviation on the statistical descriptors, of gradually decimated images from the full-resolution images, in two parts: 1) a constant shift (and dilation for the two-point correlation function) on ϕ_j and s/ϕ_j, and 2) a shape discrepancy, represented by the corresponding error $E_{\beta,j,k}$ for each of the correlation functions (with $\beta = 1, 2, 3$). Moreover, the presented analysis allows us to define a methodology to calculate an optimal number of decimation steps, based on the correlation length of the statistical descriptors, to reduce the size of an image and to minimize the time consumed for the characterization of the image, but preserving the statistical information. Once 1) Z, the optimal number of decimation steps, has been defined, 2) the gradual decimation process has been performed and 3) the normalized correlation functions of the reduced-size image have been computed, one can reconstruct the statistical descriptors of the full-resolution image just by de-normalizing the reduced-size correlation functions, us-
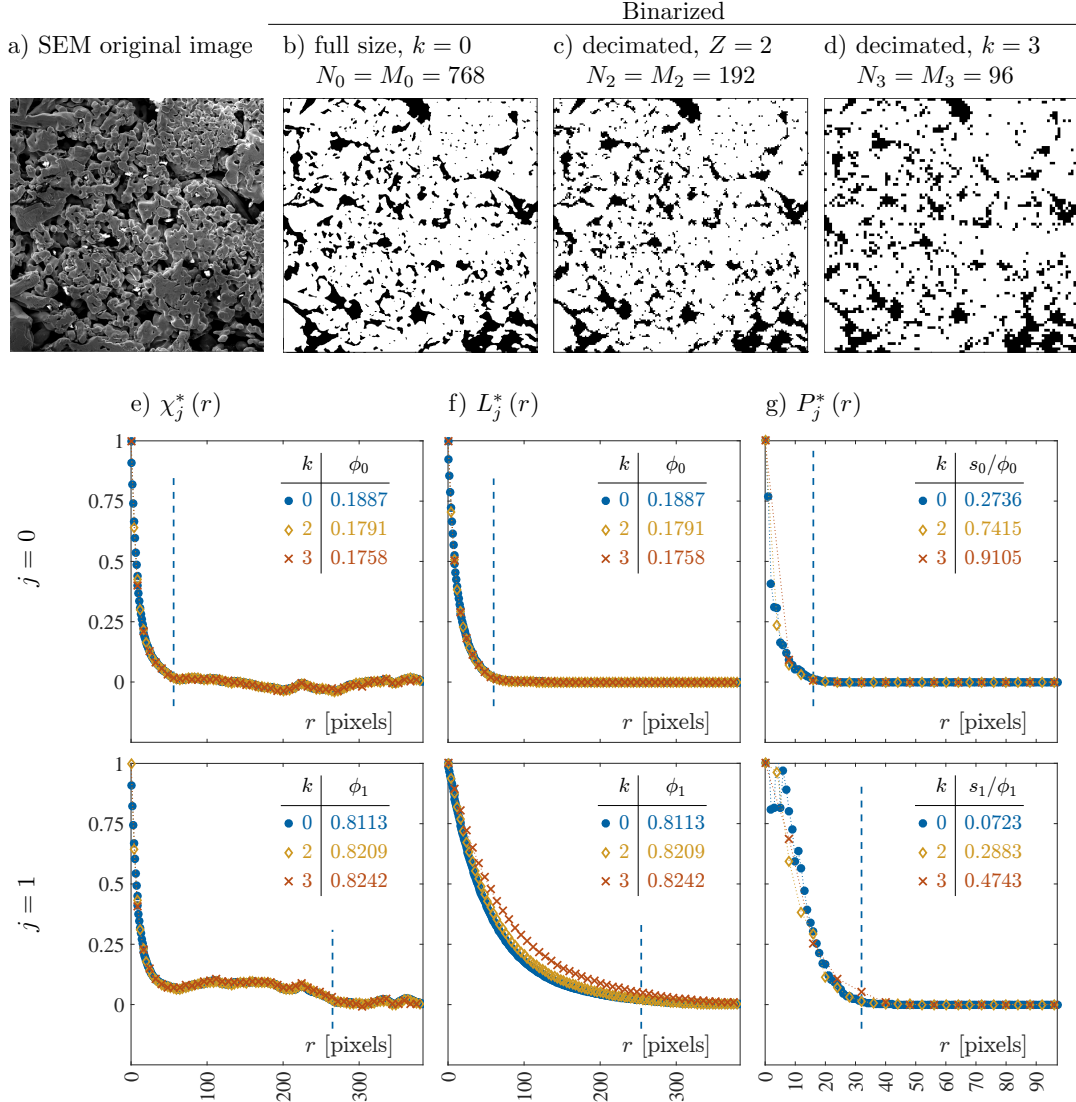
Figure 8 a) Original image of a catalytic layer obtained by SEM. b) Full size binarized image. Decimated image after c) 2 steps and d) 3 steps of a bicubic process. e)-g) Statistical descriptors for the full size and the decimated images. Top and bottom rows represent the descriptors for phases $j = 0$ and $j = 1$ respectively, whereas each column indicates the type of descriptor: e) normalized autocovariance function $F_{1,j,k} = \chi_j$, f) normalized line-path correlation function $F_{2,j,k} = L_j$ and g) normalized pore-size distribution function $F_{3,j,k} = P_j$. In columns e)-g), filled circles (blue) correspond to the descriptors computed for the full size binarized image, whereas empty diamonds (yellow) and crosses (red) correspond to the descriptors computed for a 2-steps and 3-steps decimated binarized images, respectively.

ing the full-resolution values of ϕ_j and s/ϕ_j, which are computed in a very fast and simple way.

In summary, the exhibited analysis and methodology may help to restrict the amount of information that one can afford to lose during a size reduction (decimation) process, when one aims to diminish the computational and memory cost. The main objective is to decrease the time consumed by a characterization or reconstruction technique, yet maintaining the statistical quality of the digitized sample image.

[1] S. Torquato, Annu. Rev. Mater. Res. **40**, 101 (2010).

[2] D. Li, JOM **66**, 444 (2014).

[3] C. Yeong and S. Torquato, Phys. Rev. E **57**, 495 (1998).

[4] C. Yeong and S. Torquato, Phys. Rev. E **58**, 224 (1998).

[5] D. Cule and S. Torquato, J. Appl. Phys. **86**, 3428 (1999).

[6] Y. Jiao, F. Stillinger, and S. Torquato, Phys. Rev. E **76**, 031110 (2007).

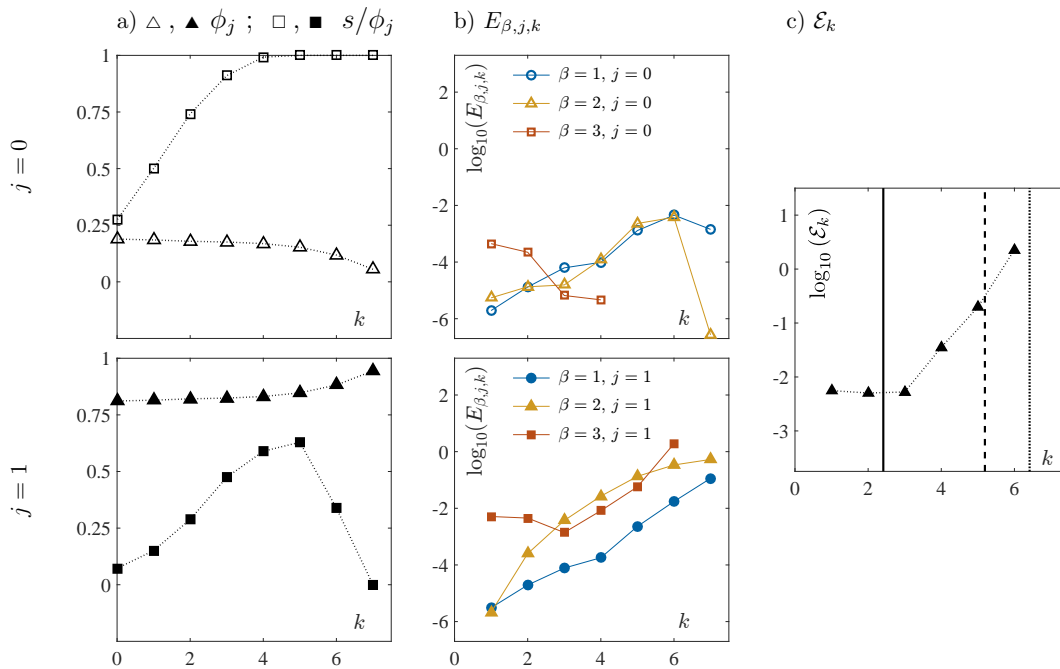[7] H. Garmestani, M. Baniassadi, M. Fathi, and S. Ahzi,

Figure 9 Quantitative information about the deviation from the full-resolution SEM image presented in Figure 8. a) Surface fraction ϕ_j and specific interface area s/ϕ_j, b) deviation of statistical descriptors $E_{\beta,j,k}$ (for each descriptor β and phase j, specified in section II B.) and c) global error as functions of the k-th step of a bicubic decimation process.

Int. J. of Theor. Appl. Mult. Mech. **1**, 134 (2009).

[8] R. Barbosa, J. Andaverde, B. Escobar, and U. Cano, J. Power Sources **196**, 1248 (2011).

[9] M. Baniassadi, B. Mortazavi, H. Amani Hamedani, H. Garmestani, S. Ahzi, M. Fathi-Torbaghan, D. Ruch, and M. Khaleel, Comput. Mater. Sci. **51**, 372 (2012).

[10] M. Huang and Y. Li, Comput. Mater. Sci. **67**, 63 (2013).

[11] P. Tahmasebi and M. Sahimi, Phys. Rev. E **85**, 066709 (2012).

[12] P. Tahmasebi and M. Sahimi, Phys. Rev. Lett. **110**, 078002 (2013).

[13] P. Tahmasebi and M. Sahimi, Phys. Rev. E **91**, 032401 (2015).

[14] Y. Liu, M. Greene, W. Chen, D. Dikin, and W. Liu, Computer-Aided Design **45**, 65 (2013).

[15] L. Pant, S. Mitra, and S. M., Phys. Rev. E **90**, 023306 (2014).

[16] D. Torquato, *Random Heterogeneous Materials microscopic and macroscopic properties* (Springer, 2002).

[17] M. Sahimi, *Heterogeneous Materials I: Linear transport and optical properties* (Springer, 2003).

[18] A. Snarskii, I. Bezsudnov, V. Sevryukov, A. Morozovskiy, and J. Malinsky, *Transport processes in macroscopically disordered media* (Springer, 2016).

[19] T. Kanit, F. NGuyen, S. Forest, D. Jeulin, M. Reed, and S. Singleton, Comput. Methods Appl. Mech. Engrg. **195**, 3960 (2006).

[20] R. Sinclair, *Textiles and Fashion - Materials, Design and Technology* (Elsevier, 2015), woodhead publishing series in textiles: number 126 ed.

[21] A. Suzuki, T. Hattori, R. Miura, H. Tsuboi, N. Hatakeyama, H. Takaba, M. Williams, and A. Miyamoto, Int. J. Electrochem. Sci. **5**, 1948 (2010).

[22] M. Shojaeefard, G. Molaeimanesh, M. Nazemian, and M. Moqaddari, Int. J. Hydrogen Energy **179**, 1 (2016).

[23] M. Che and J. Vedrine, *Characterization of Solid Materials and Heterogeneous Catalysts: From Structure to Surface Reactivity, Volume 1&2* (Wiley-VCH Verlag GmbH & Co. KGaA, 2012).

[24] W. Harris and W. Chiu, J. Power Sources **282**, 552 (2015).

[25] W. Harris and W. Chiu, J. Power Sources **282**, 622 (2015).

[26] C. C.Pelissou, J. Baccou, Y. Monerie, and F. Perales, Int. J. Solids Struct. **46**, 2842 (2009).

[27] S. Mirkhalaf, F. Andrade Pires, and R. Simoes, Finite Elem. Anal. Des. **119**, 30 (2016).

[28] C. E. Zachary and S. Torquato, J. Stat. Mech. **12**, P12015 (2009).

[29] A. Bovik, *Handbook of image and video processing* (Academic Press, 2000).

[30] R. Gonzalez and R. Woods, *Digital image processing* (Pearson-Prentice Hall, 2008), 3rd ed.

[31] J. Ortegon, R. Ledesma-Alonso, R. Barbosa, J. Vazquez Castillo, and A. Castillo Atoche, Submitted to Computational Materials Science (2017).