

A novel PCA-based microstructure descriptor for heterogeneous material design



Chao Xu, Shuming Gao, Ming Li*

State Key Laboratory of CAD&CG, Zhejiang University, Hangzhou, China

ARTICLE INFO

Article history:

Received 23 August 2016
Received in revised form 18 December 2016
Accepted 21 December 2016
Available online 13 January 2017

Keywords:

Microstructure design
PCA
Microstructure descriptor
Heterogeneous

ABSTRACT

A novel PCA (Principal Component Analysis)-based microstructure descriptor is proposed for heterogeneous material design given a database of material microstructures. The PCA-based descriptor captures the important geometric features of the database, and provides a compact representation of the design space. Based on the PCA-based descriptor, a polynomial expression is ultimately derived that links the microstructures in the design space with their associated physical performances. Compared with previous approaches using first-order or second-order descriptors, the novel approach integrates PCA and Combined Approximation (CA) in mechanics, provides more precise presentation of the microstructure and accelerates the optimization and reconstruction of the result microstructure. The performance of the proposed approach is also tested in various aspects for a linear elasticity analysis problem.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Designing materials of desired physical properties is a constant pursuit of human being but the process is significantly time and effort-consuming. In order to accelerate the process of discovering and designing new materials, the concept of *microstructural mediated design* [1–3] has been proposed, which follows Olson's design framework covering *process*, *structure*, *property*, and *performance* [2]. This new idea emphasizes the microstructures of materials, viewing it as the determining factor which dictates its physical and chemical performance. This novel strategy implements the material design process in the form of finding the optimal microstructure so that the material exhibits desired performance, an idea which revolutionized the material design strategy.

The key issues of *microstructural mediated design* are: how to describe the microstructure, and how to build structure-property and property-performance relations. To build structure-property relation, the traditional homogenization approaches [4–6] provide homogenized property of an RVE (representative volume element) according to its microstructure. To build the property-performance relation, the popular FEM (finite element method) simulates material performance under numerous physical conditions. Further, Xu et al. [7–9] took into account the meta-modeling method and provided an approximate expression of the structure-performance

relations. This computational costs of these approaches are usually high.

How to describe the microstructure is an even more important issue, because the description of the microstructure reveals its key information of the microstructure, determines the design parameters, and is also involved in the formation of all the process-structure-property-performance relations. A direct and precise description is an image obtained from scanning the material sample using SEM/TEM (scanning/transmission electron microscopy), which can capture the microstructure's composition, phases, morphologies and particularly geometries. However, during the design process the image description includes too many design parameters if all the pixel values need determined carefully.

To avoid redundant design parameters, one solution is to extract only statistical information contained within a scanning image. Among the numerous studies on this issue, the first- and second- (or higher) order descriptors [7,10] are the most popular. The first-order descriptors capture the one-point statistics of the microstructure image, such as volume fractions and nearest neighbor distances. However, the first-order descriptors are not suitable for describing the microstructure geometries due to missing spatial information. The second- (or higher) order descriptors solve this issue by considering the distribution statistics, such as 2-point correlation functions describing the distribution of line segments with a fixed length.

Excellent studies have been done using these first- and second- (or higher) order descriptors. Kalidindi et al. [3,10,11] proposed a performance oriented microstructure-sensitive design system with

* Corresponding author.

E-mail address: liming@cad.zju.edu.cn (M. Li).

a simple and effective process-structure-property-performance relationship. This was achieved by projecting microstructure distribution functions like size distribution, orientation distribution, and some correlation functions into a spectral space. Then, the distribution functions were reduced to a compact representation of a few parameters in the spectral space. Further extending the homogenization theory, the structure-property relation was built and the design problem was transformed to an optimization problem that can be easily solved. However, this structure-property relation maintains only some microstructure information and has limited accuracy.

Other statistical descriptors of microstructure composition, dispersion and geometry were also applied by Xu et al. [7–9], and it was proved that these descriptors allow the reconstruction of microstructures as in the cases of high order descriptors [7]. Besides, the machine learning theory [8,9] was introduced to extract the key descriptors with respect to the properties. Then, a design approach was proposed with the structure-property relation built using a combination of FEM and meta-modeling method.

Another study on the microstructure geometry descriptor was done by Wang et al. [12]. The microstructure geometry was modeled and controlled accurately using periodic surfaces, while the structure was related to some properties such as porosity and permeability. This study provide a new perspective to microstructure modeling.

However, the disadvantages of these descriptors are obvious: first, some of them are computationally expansive and their cost increases exponentially with the order; second, they only capture statistical information and lose much, so, key information that determines the material performance might be lost since we have no direct structure-property and property-performance relations to refer to; third, it is computationally expensive to reconstruct the microstructure image from these descriptors and the reconstructed microstructures are not unique, despite the great efforts devoted towards this end, such as genetic algorithms [13,14], gaussian random fields [15,16], branch and bound [17] and simulated annealing [18,19].

Viewed from another perspective, the extraction of these descriptors from microstructure images can be taken as information denoising, i.e. extracting the critical information in the design process and avoiding redundancy. To achieve this goal, we introduced the dimension reduction technique formerly used in image [20] and graphic [21] analysis. The dimension reduction technique has been developed into two categories: linear and nonlinear. Linear dimension reduction methods like PCA (principal component analysis) [22] and LDA (linear discriminant analysis) [23] transform the high dimension data to low dimension data while maintaining the linear relations of the pairwise distances. Nonlinear dimension methods like Isomap (isometric mapping) [24], SNE (stochastic neighbor embedding) [25], t-SNE [26] and LargeVis (large-scale visualization) [27] were conceptualized as manifold learning for preserving the manifold shape after projecting high dimension data to a low dimension.

Among these methods, PCA has the special characteristic of capturing the key components, putting them in a linear space and linking the redundant information to them. Due to this advantage, PCA has already been applied in the material design process. Kalidindi [28] applied PCA in property spaces to approximate the state solution for performance optimization. Recently, Xu et al. [29] employed PCA to reduce the dimension of the material distribution space, in which the material distributions are not always smooth. Also, Li et al. [30] used PCA in both the geometric and physical space to reduce the computational costs, and Niezgodna et al. [31] made use of PCA in the space of second-order descriptors to accelerate the descriptor extraction.

Different from these previous approaches, PCA was applied in this study directly in the microstructure image space. In this space, we can exploit the ability of PCA which maintains the shape features of the database similar to face recognition [32]. Further, with the linear relation to the eigenvector of PCA and the introduced CA (combined approximation) approach [33], we derived a novel polynomial relation between the PCA descriptor and the performance. This polynomial relation enabled fast property and performance prediction and thus fast microstructure optimization.

The major contributions of this paper are as follows. First, we proposed a novel PCA-based microstructure descriptor for a given database, which is easier to extract from microstructure, faster for microstructure reconstruction and preserves shape character better than previous descriptors. Second, based on such descriptor, an analytical approximate structure-performance relation was derived, and can be applied directly in performance-oriented structure optimization. Performance of the proposed descriptor was also tested in various aspects via numerical experiments.

This work is a modified version of our previous conference paper [34] with the following major improvements: First, the parallel analysis method [35] was introduced after the PCA process for selecting proper PCA bases. Second, the Ashby model [36] has been introduced to provide a solid and accurate structure-property relation, which has proved effectiveness in numerous physical experiments. Third, the overall numerical approach was carefully designed and improved for further computational efficiency improvement. Lastly, all the numerical examples were redesigned and tested based on the newly proposed approach.

This paper is arranged as follows: first, the novel PCA-based descriptor definition and its relation with the existing descriptors are presented in Section 2; and then the problem and overall approach is summarized in Section 3 and the structure-property relation is derived in Section 4. Furthermore, the property-performance relation is provided by CA, and based on this relation, a polynomial approximation of the compliance supporting the microstructure design approach is formulated in Section 5; we perform numerical tests on the proposed approach in Section 6, and all the results are concluded in Section 7.

2. Microstructure and descriptors

In this section, we first explain the representation of microstructure and its descriptors, and then introduce the novel PCA-based microstructure descriptor. Further, we deduce the relation of the novel PCA-based descriptor and the classical first- or second- order descriptors.

2.1. Microstructure

A microstructure is represented as a binary digital image in this study, obtained by scanning a material sample using SEM/TEM. The pixel value in the image represents the material phase. Limited by the instrumental errors of SEM/TEM, the microstructure image can be gray images (shown as the left image of Fig. 1) even when photographing bi-phase materials [7]. To eliminate the errors and provide a binary image, we can use image processing techniques like denoising and binarizing to get a strictly binarized image, as shown on the right of Fig. 1. In addition, the microstructure image is also taken as a FE mesh in the rest of the paper when performing FE analysis on the microstructure.

The following notations are used in this paper:

I : a microstructure image under study;

ρ : a density value in I representing the material phase, specifically, '0' for void phase and '1' for solid phase.

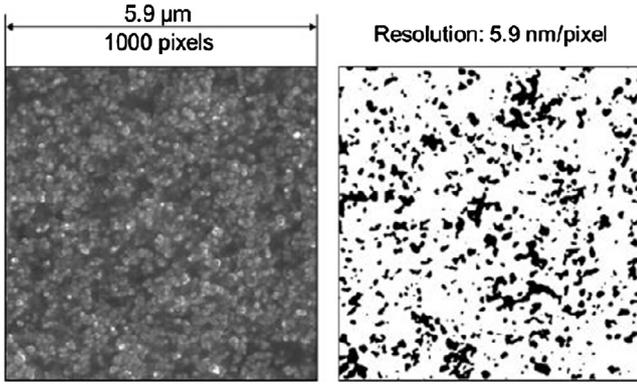


Fig. 1. The SEM image (left) and the bi-phase image (right) after binarization of polymer nanocomposites [7].

2.2. Descriptors

A microstructure descriptor is given as a scaler or a vector which identifies this microstructure from others. Taking the density ρ of each element in I as a dimension, the microstructure image is of high dimension. A microstructure descriptor aims to reduce the dimension to a vector or even a scaler with reasonable information reduction, and is used as the microstructure design parameter.

2.2.1. Definition of PCA-based microstructure descriptor

We introduce the linear dimension reduction technique PCA in image processing, to define the novel microstructure descriptor, which consequently gives a compressed representation of a microstructure.

Given a database of the binarized microstructure images as given in Fig. 1, we can obtain the linear decomposition of any microstructure image I under study via PCA,

$$I \approx \sum_{k=0}^n a_k I_k, \quad (1)$$

where \approx means the decomposition is a nice approximation, $\{I_k\}_{k=1}^n$ are the PCA bases, and $\{a_k\}_{k=1}^n$ are the coefficients of I with respect to the bases. In particular, the first item of the summation, $a_0 I_0$, is preset before the PCA decomposition as $a_0 = 1$ and I_0 the average image of the database. The coefficients $\{a_k\}_{k=1}^n$ give the PCA-based microstructure descriptor, and Eq. (1) builds a bijection between an arbitrary image I and the descriptor $\{a_k\}_{k=1}^n$.

2.2.2. Selection of PCA-based microstructure descriptor

After PCA decomposition, the decision of which and how many bases to retain is critical. Noticing that PCA is a factor analysis technique, we apply the parallel analysis [35], one of the most accurate factor retention methods, to resolve the issue.

The idea of parallel analysis is to find the PCA bases of which the eigenvalues are greater than 95% of the randomly generated ones. The process is formed by four main steps,

1. Perform PCA on the given database;
2. Generate random data similar to the database and perform PCA on them;
3. Repeat the Step 2 for 100 times and calculate the average of the eigenvalues;
4. Select the PCA bases of which the eigenvalues are larger than 95% of the above average.

We note here that 95% is an empirical percentage and needs further experimental verification.

2.2.3. Relation with previous descriptors

Previous descriptors can be classified as the first-, second- and higher order descriptors. The first-order descriptors depict the percentages of overall geometric microstructure features [8]. The second-order descriptors are statistical quantities representing two-point relations, such as the possibility of the two points with fixed distance r placed in the same phase [7]. Other descriptors such as lineal path function and higher order descriptors which focus on more points' relations are similarly defined, and are not further discussed in this paper.

The volume fraction, one of the common first-order descriptors, is defined as the percentage of volume that one phase occupies in the material. Mathematically, it can be calculated as

$$vf(I) = \frac{1}{V} \sum_{x \in I} I(x), \quad (2)$$

here $vf(I)$ denotes the volume fraction of I and V is the volume of area I .

Substituting the decomposition Eq. (1) of I into Eq. (2), we further have,

$$vf(I) = \frac{1}{V} \sum_{x \in I} \sum_{k=0}^n a_k I_k(x) = \sum_{k=0}^n a_k \frac{1}{V} \sum_{x \in I} I_k(x) = \sum_{k=0}^n a_k \cdot vf_k, \quad (3)$$

where vf_k is the volume fraction of I_k and constant for a fixed I_k . In this case, we find the linear relation of the first-order descriptor and the PCA-based descriptor.

The two-point correlation function [7], one of the common second-order descriptors, is defined as

$$P_2(r) = \langle I(x_1)I(x_2) \rangle, \quad (4)$$

where x_1, x_2 are two points with the fixed distance $|x_1 - x_2| = r$ and $\langle \rangle$ indicates the average over all two-point pairs.

By substituting the PCA decomposition Eq. (1) into Eq. (4), we have

$$\begin{aligned} P_2(r) &= \langle I(\mathbf{x})I(\mathbf{x} + \mathbf{r}) \rangle \\ &= \langle \left[\sum_{k_1=0}^n a_{k_1} I_{k_1}(\mathbf{x}) \right] \left[\sum_{k_2=0}^n a_{k_2} I_{k_2}(\mathbf{x} + \mathbf{r}) \right] \rangle \\ &= \langle \sum_{k_1=0}^n \sum_{k_2=0}^n [a_{k_1} I_{k_1}(\mathbf{x})] [a_{k_2} I_{k_2}(\mathbf{x} + \mathbf{r})] \rangle \\ &= \langle \sum_{k_1=0}^n \sum_{k_2=0}^n a_{k_1} a_{k_2} [I_{k_1}(\mathbf{x})] [I_{k_2}(\mathbf{x} + \mathbf{r})] \rangle \\ &= \sum_{k_1=0}^n \sum_{k_2=0}^n a_{k_1} a_{k_2} \langle I_{k_1}(\mathbf{x}) I_{k_2}(\mathbf{x} + \mathbf{r}) \rangle. \end{aligned} \quad (5)$$

Consequently, we have

$$P_2(r) = \sum_{k_1=0}^n \sum_{k_2=0}^n a_{k_1} a_{k_2} P_{2,k_1,k_2}(r), \quad (6)$$

where $P_{2,k_1,k_2}(r) = \langle I_{k_1}(\mathbf{x}) I_{k_2}(\mathbf{x} + \mathbf{r}) \rangle$ is constant when r is assigned. The analytical expression (6) gives the relation between 2-point correlation function and the new PCA-based descriptor. Higher order descriptors can be related to the PCA-based descriptor in a similar manner.

Once the relation with the existing first- and second- descriptors is built above, the previous results of material design involving the existing descriptors can be easily reformulated using the PCA-based descriptor as the design parameters, that is, the novel PCA-based descriptor can substitute the existing ones.

Also, we will explain later that the PCA-based descriptor is easier to calculate, especially compared to the 2-point correlation functions. In the material design process, the microstructure varies and thus the descriptors need to be recalculated in each iterative optimization steps. Introducing the novel PCA-based descriptor can save a great computational cost in the design process.

3. Problem statement and approach overview

3.1. Problem statement

In this paper, we focus on the specific case of linear elasticity. The physics of the microstructure is studied via performing FEM on the microstructure. Accordingly, we have the stiffness matrix K associated to the microstructure, and the equilibrium equation defined as $K\mathbf{u} = \mathbf{b}$.

Now the problem of microstructure design is defined below:

Inputs: the microstructure database of a certain kind of material and its PCA bases $\{I_k\}_{k=1}^n$; the physical boundary conditions (loads \mathbf{b} and fixed boundaries);

Outputs: the optimal microstructure, represented by its descriptor $\{a_k\}_{k=1}^n$ of minimal compliance $c = \mathbf{u}K\mathbf{u}$;

In addition, we also limit the value of ρ in the range $[0, 1]$ for phase presentation. Correspondingly, similar to the decomposition of I , we have $\rho = \sum_{k=0}^n a_k \rho_k$ for each density vector ρ_k in I_k .

To summarize, the problem of optimal microstructure design under study is,

$$\left. \begin{array}{l} \min_{\{a_k\}} : c = \mathbf{u}K\mathbf{u} \\ \text{subject to} : K\mathbf{u} = \mathbf{b} \\ : \rho = \sum_{k=0}^n a_k \rho_k \\ : 0 \leq \rho \leq 1 \end{array} \right\}, \quad (7)$$

We also comment here that the density ρ is now required to lie between 0 and 1. In order to produce a binary image, the value of ρ should better be close to 0 or close to 1. A further strategy to achieve this will be explained later in Section 5.1.

3.2. Approach overview

The proposed approach follows the Olson's design framework that links microstructure, property and performance. In this case, the property is taken as the associated stiffness matrix K to the microstructure, as involved in FEM computation process, and the performance is set as the compliance of the microstructure.

The proposed approach is shown in Fig. 2. In Step 1, the microstructure is represented by its PCA-based descriptors. Then, in Steps 2 and 3, we build a polynomial structure-property-performance relation to approximately link the descriptor with its ultimate physical performance, that is the compliance. Based

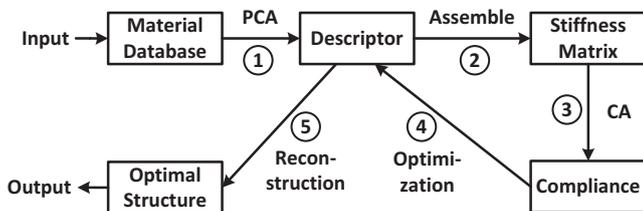


Fig. 2. The design flowchart: (1) using the input material database to build the PCA bases for the descriptor; (2) building the structure-property (descriptor-stiffness matrix) relation; (3) building the property-performance (stiffness matrix-compliance) relation; (4) optimization to find best descriptor; (5) structure reconstruction and output.

on this polynomial relation, an optimization process is performed to find the optimal microstructure descriptors. Finally, we reconstruct the optimal microstructure from the derived optimal descriptors and output it.

Algorithmic details are explained in the following sections.

4. Relation between pca-based descriptor and material performance

Based on the description in Section 3.2, in this section we build the relation between the descriptor, the stiffness matrix and the microstructure's compliance.

4.1. Relation of density and element stiffness matrix

Ashby et al. [36] presented the physical verification of the usage of the penalty weight ρ^p . They tested cells of different densities and plot the stress-strain curves. After numerous tests over different materials and cell microstructures, they concluded that the overall Young's modulus E and Poisson's ratio ν of these cells are as follows,

$$E(\rho) \approx \begin{cases} \alpha \rho^2 E_0, & \text{for open cell,} \\ (\alpha \rho^2 + \beta \rho) E_0, & \text{for closed cell,} \end{cases} \quad (8)$$

$$\nu \approx 0.3, \quad (9)$$

where α, β are determined by the interior structure geometry of the cell, E_0 is the constant Young's modulus of the cell when $\rho = 1$.

In the case of an isotropic cell with a fixed Poisson's ratio ν , the relation of the Young's modulus E and stiffness matrix C is linear, which consequently gives

$$C(\rho) \approx \begin{cases} \alpha \rho^2 C_0, & \text{for open cell,} \\ (\alpha \rho^2 + \beta \rho) C_0, & \text{for closed cell,} \end{cases} \quad (10)$$

where C_0 is the constant element stiffness matrix for $\rho = 1$.

The Ashby model builds an approximation to the element stiffness matrix C with respect to ρ . This approximation on metal foams has been found reasonable and accurate when tested on MetFoam '97 database [36].

4.2. Relation of density distribution and stiffness matrix after assembling

Based on the above analysis, the following lemma states that the final stiffness matrix can always be written as the sum of a constant matrix and a difference matrix in terms of the design density ρ .

Lemma 1. Suppose a microstructure image I is given in a form following Eq. (1), for any element stiffness matrix $C(\rho)$ we have

$$C(\rho) = (\alpha \rho^2 + \beta \rho) C_0. \quad (11)$$

Let K, K_k respectively be the stiffness matrix for I_0, I_k . Then, the assembled stiffness matrix K have a quadratic decomposition,

$$K = K_0 + \beta \sum_{k=1}^n a_k K_k + \alpha \sum_{s=0}^n \sum_{t=0}^n a_s a_t K_{s,t} \quad (12)$$

where k, s, t are the different indexes for the same descriptor, which satisfy $a_k = a_s = a_t$ if and only if $k = s = t$.

Proof. Following Galerkin finite method, the sub-matrix $K(1)$ to the cell with $\rho = 1$ is

$$K(1) = \int_I \nabla \phi_i^T C_0 \nabla \phi_j dV, \quad (13)$$

where ϕ_i, ϕ_j are the test functions. Thus we have the sub-matrix $K(\rho)$ to the corresponding cell with density ρ ,

$$\begin{aligned} K(\rho) &= \int_I \nabla \phi_i^T C(\rho) \nabla \phi_j dV \\ &= \int_I \nabla \phi_i^T (\beta \sum_{k=0}^n a_k \rho_k + \alpha \sum_{s=0}^n \sum_{t=0}^n a_s a_t \rho_s \rho_t) C_0 \nabla \phi_j dV, \\ &= \beta \sum_{k=0}^n a_k \int_I \nabla \phi_i^T \rho_k C_0 \nabla \phi_j dV + \\ &\quad \alpha \sum_{s=0}^n \sum_{t=0}^n a_s a_t \int_I \nabla \phi_i^T \rho_s \rho_t C_0 \nabla \phi_j dV \\ &= K_{0,\rho} + \beta \sum_{k=1}^n a_k K_{k,\rho} + \alpha \sum_{s=0}^n \sum_{t=0}^n a_s a_t K_{s,t,\rho}, \end{aligned} \quad (14)$$

where we denote

$$\left. \begin{aligned} K_{0,\rho} &= \int_I \nabla \phi_i^T C_0 \nabla \phi_j dV \\ K_{k,\rho} &= \int_I \nabla \phi_i^T \rho_k C_0 \nabla \phi_j d\Omega \\ K_{s,t,\rho} &= \int_I \nabla \phi_i^T \rho_s \rho_t C_0 \nabla \phi_j d\Omega \end{aligned} \right\}. \quad (15)$$

We assemble $K_0, K_k, K_{s,t}$ using $K_{0,\rho}, K_{k,\rho}, K_{s,t,\rho}$ respectively. In this case, we have the same polynomial decomposition of every sub-matrix and thus the same decomposition of the matrix. \square

Since we have built the quadratic decomposition of stiffness matrix K with respect to the microstructure descriptor $\{a_k\}_{k=0}^n$, and the state solution \mathbf{u} can be derived by solving the linear equation system $K\mathbf{u} = \mathbf{b}$, then the idea of deriving the analytical expression of \mathbf{u} to $\{a_k\}_{k=0}^n$ comes up via the further usage of CA approach. Details are explained below.

4.3. Approximation of state solution

From the results in Lemma 1, suppose we have

$$K_0 \mathbf{u}_0 = \mathbf{b}, \quad K\mathbf{u} = (K_0 + \Delta K)\mathbf{u} = \mathbf{b}, \quad (16)$$

where

$$\Delta K = \beta \sum_{k=1}^n a_k K_k + \alpha \sum_{s=0}^n \sum_{t=0}^n a_s a_t K_{s,t} \quad (17)$$

is determined by the PCA-based descriptor.

The aim here is to approximate state solution \mathbf{u} using the descriptor $\{a_k\}$ and other fixed quantities $(\alpha, \beta, \mathbf{u}_0, K_0, K_k, K_{s,t}, \mathbf{b})$. The CA (combined approximation) approach, one of the structure reanalysis method [33], is introduced to resolve this issue. CA can produce solution approximation for low-rank or moderately high-rank structure changes by using the modified part of K as the terms in the binomial series expansion.

Mathematically, the CA approach give the approximation of \mathbf{u} as

Lemma 2. [33] The polynomial approximation of \mathbf{u} is

$$\mathbf{u} = \mathbf{u}_0 - \sum_{i=1}^m t_i \mathbf{u}_i = \mathbf{u}_0 - U\mathbf{t}, \quad (18)$$

where \mathbf{u}_i can be iteratively calculated for each i as

$$K_0 \mathbf{u}_i = -\Delta K \mathbf{u}_{i-1}, \quad i = 1, \dots, m, \quad (19)$$

and $U = [\mathbf{u}_1, \dots, \mathbf{u}_m]$. The coefficients \mathbf{t} are determined by solving the reduced system

$$U^T (K_0 + \Delta K) U \mathbf{t} = U^T \Delta K \mathbf{u}_0. \quad (20)$$

We can also set $i = 1$, in which case the equality is not strictly satisfied, however, the computational costs are reduced.

Note here that when calculating \mathbf{u}_i numerically via Eq. (19), K_0 is fixed and can be reused. Thus the trick which can accelerate the numerical computation of \mathbf{u}_i is to avoid solving the linear equation system in each step. Instead, we perform Cholesky decomposition on K_0 as pre-processing to obtain the decomposition

$$K_0 = PQ, \quad (21)$$

where P is lower triangular matrix and Q is upper triangular matrix. In this case, solving Eq. (19) is converted to doing Gaussian elimination twice, which costs very little computationally.

Another numerical acceleration trick here is to avoid symbolic computation at this stage. Note here ΔK has a linear decomposition of $\{a_k\}_{k=1}^n$ with the ‘coefficients’ $\{K_k\}_{k=1}^n$, but we can circumvent computations involving the symbols $\{a_k\}_{k=1}^n$. Taking the computation of \mathbf{u}_1 as an example,

$$\begin{aligned} \mathbf{u}_1 &= \beta \sum_{k=1}^n a_k K_0^{-1} K_k \mathbf{u}_0 + \alpha \sum_{s=0}^n \sum_{t=0}^n a_s a_t K_0^{-1} K_{s,t} \mathbf{u}_0 \\ &= \beta \sum_{k=1}^n a_k \mathbf{y}_k + \alpha \sum_{s=0}^n \sum_{t=0}^n a_s a_t \mathbf{z}_{st}, \end{aligned} \quad (22)$$

here \mathbf{y}_k is calculated from solving $K_0 \mathbf{y}_k = K_k \mathbf{u}_0$, and \mathbf{z}_{st} is calculated from solving $K_0 \mathbf{z}_{st} = K_{s,t} \mathbf{u}_0$. In this sense, no symbols are involved in this computation. Also, the $\{\mathbf{u}_i\}_{i=1}^m$ here are polynomial expressions of $\{a_k\}_{k=1}^n$.

4.4. Approximation of compliance

In linear elasticity, the state solution \mathbf{u} is the key to calculating other physical quantities. Once we have the approximation of \mathbf{u} , we focus on the mechanical compliance c , the important physical quantity which measures the material’s resistance to deformation caused by external forces, and is defined in an energy form $c = \mathbf{u}^T K \mathbf{u}$.

We firstly substitute $K\mathbf{u} = \mathbf{b}$ to Eq. (16) and obtain

$$\mathbf{u}^T K \mathbf{u} = \mathbf{u}^T \mathbf{b}, \quad (23)$$

and further substitute the results of Lemma 2,

$$c = \mathbf{u}^T \mathbf{b} = \mathbf{u}_0^T \mathbf{b} - \mathbf{t}^T U^T \mathbf{b} = c_0 - \sum_{k=1}^m \mathbf{t}^T c_k, \quad (24)$$

taking into account Eq. (18), and denote $\mathbf{u}_k^T \mathbf{b}$ as c_k .

The following theorem summarize the above results:

Theorem 1. The compliance $c = \mathbf{u}^T K \mathbf{u}$ can be expressed in the following polynomial equation:

$$c = \mathbf{u}^T \mathbf{b} = (\mathbf{u}_0^T \mathbf{b} - \mathbf{t}^T U^T \mathbf{b}) = c_0 - \sum_{k=1}^m \mathbf{t}^T c_k. \quad (25)$$

Here we notice that \mathbf{u}_k are polynomial expressions of $\{a_k\}_{k=1}^n$, thus c is also polynomial and denoted as $c(\{a_k\})$.

5. Optimize the microstructure

The aim of our approach is to find the optimal microstructure with the minimal compliance. With the relation of compliance and descriptor derived above, all we need is to add some necessary constraints and modifications to obtain the mathematical form of the optimization problem.

5.1. Mathematical form of the optimization problem

The constraints we add on $\{a_k\}_{k=1}^n$ are $0 \leq a_k \leq 1, k = 1, 2, \dots, n$. Considering the decomposition of I in Eq. (1), after performing PCA the bases $\{I_k\}_{k=0}^n$ are fixed. If the upper and lower bounds of $\{a_k\}_{k=1}^n$ were not limited, it is conceivable that abnormally large/small points in the microstructure image I would occur. Further noticing that the first item I_0 is sufficiently larger than other bases $\{I_k\}_{k=1}^n$, the constraint $0 \leq a_k \leq 1, k = 1, 2, \dots, n$ will ensure a suitable microstructure image I .

Additionally, the microstructure studied here is bi-phase (solid/void), so the resulted microstructure image should be binarized. We thus modify the optimization target as $c + s \int_I (\rho^2(1 - \rho)^2) dV$, where $\int_I (\rho^2(1 - \rho)^2) dV$ is the added penalty function and s is its tuning weight. The minimum of the penalty function is 0 when $\rho = 0$ or 1. When it is approaching 0, ρ is thus approaching 0 or 1.

The weight s of the added penalty function is a trade-off between computational cost and penalty effect. Similar to the problem studied in [37], we set $s = 10$ here accordingly to achieve a cost increase of less than 5% while s increases 10 times. Also, $s = 10$ here is only empirical and needs only small manual modifications in some cases.

To sum up, the final optimal problem is defined as

$$\begin{aligned} \min_{\{a_k\}} & : c(\{a_k\}) + s \int_I (\rho^2(1 - \rho)^2) dV, \\ \text{subject to} & : \left. \begin{aligned} 0 \leq a_k \leq 1, \\ \rho = \sum_{k=0}^n a_k \rho_k, \end{aligned} \right\} \end{aligned} \quad (26)$$

The above optimization can be easily solved using any optimization tool, such as ‘fmincon’ in *Matlab*[®]. In practise and in order to further improve the computational efficiency, we perform Cholesky decomposition for K_0 and re-use it, instead of directly solving the linear system involved in each iterative step in Section 4.3.

6. Performance test

In this section, we aim to test the approach’s performance from various aspects. Following the approach proposed above, the preparations for the tests are: we generate a database containing 1000 microstructure images. Each microstructure image is of resolution 100×100 , and has 5–7 random black holes in it. For simplification, the PCA bases are selected as the first 100 bases with the largest eigenvalues according to the parallel analysis [35] in Section 2.2.2. Consequently, we have a PCA-based descriptor represented in a 100-dimension vector. The physical problem under study is shown in Fig. 3(a), where the left side of the cell is fixed and the load is applied on the bottom right corner. The simulation results from FEA (finite element analysis) performed on a quadri-

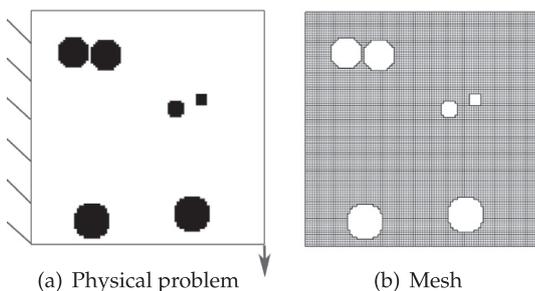


Fig. 3. One random image (a) in the database with the added physical conditions and its mesh (b).

lateral mesh in Fig. 3(b) are taken as the ground truth for accuracy evaluation.

6.1. Approximation accuracy of compliance

Two main types of approximation errors are involved in our design approach. The first comes from the geometric loss when applying PCA decomposition explained in Section 2.2.1, and the second comes from the physical loss when simulating using CA explained in Section 4.3. These accuracy losses add up to the final error of the target compliance. The simulation accuracy determines the performance of the proposed approach and was thus tested here by comparing it with the benchmark results obtained from FEA.

6.1.1. Accuracy under randomly sampling

We first tested the simulation accuracy of our approach for randomly generated sample microstructures. These samples were easily generated here via a random sampling of the PCA-based descriptor space considering their one-to-one mapping to the microstructure, as illustrated in Fig. 4.

A set of 100 sample microstructures was used. For each sample, we calculate the compliance respectively using the expression given in Eq. (24) (denoted as $c_{expr.}$) and FEA (denoted as c_{FEA}). Results of the microstructure samples of the maximum, minimum and average compliances are shown in Fig. 5. Also the values of $c_{expr.}$ and c_{FEA} are shown underneath, together with their relative error calculated via

$$E_c = \frac{|c_{expr.} - c_{FEA}|}{c_{FEA}}. \quad (27)$$

All the 100 tested errors are within an acceptable range 1.2–5.2%.

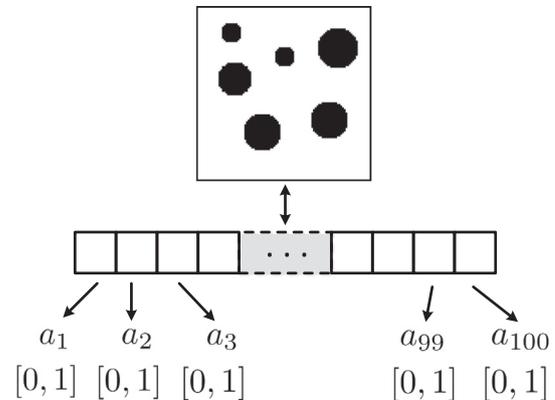


Fig. 4. The one-to-one mapping between the microstructure and PCA-based descriptor: one microstructure and its associated descriptor (a_1, \dots, a_{100}) .

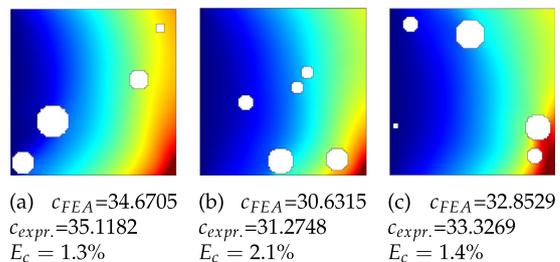


Fig. 5. The three selected microstructure images with compliances as listed: (a) with the maximum compliance; (b) with the minimum compliance; (c) with the average compliance.

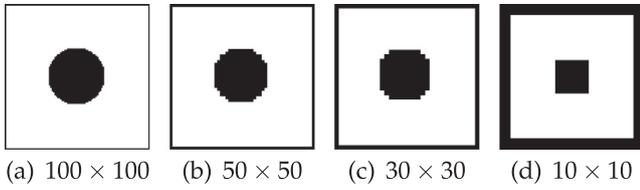


Fig. 6. The same microstructure degrades from higher to lower resolutions (the outer frame not included): (a) 100×100 , (b) 50×50 , (c) 30×30 , (d) 10×10 .

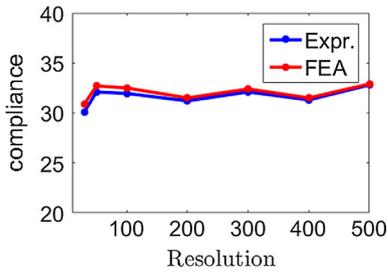


Fig. 7. The compliances comparison with FEA and our approach under different resolutions.

6.1.2. Accuracy under different resolutions

We also test the accuracy of our approach under different resolutions considering that the resolution is limited by the microscopes used in reality. In addition, different resolutions may also show the same phase distribution in different ways. For example in Fig. 6, the round shape in resolution 100×100 can degrade to a square in a low resolution of 10×10 .

In this test, we randomly generated a microstructure of resolution 500×500 and degraded it to lower resolutions: 400×400 , 300×300 , 200×200 , 100×100 , 50×50 , 30×30 , meanwhile preserving the relative size and position of the holes. The compliances computed via Eq. (24) and FEA are compared in Fig. 7. We see that the two compliance curves are becoming closer from left to right all within a small range of differences, indicating that the accuracy is acceptable in all resolutions, and noticeably better in higher resolutions.

6.2. Database shape influences on the optimal results

Next, We assessed the influence of the databases containing microstructures of different phase shapes on the final optimal microstructure, as it is important that after the optimization the derived microstructure maintains expected geometric characters of the database.

In order to quantify the preservation of the geometric features of the databases in the optimal structure, we define a “fitting error” to measure E_{fit} . It is defined as the relative ratio of the number of difference pixels with respect to the expected shape to the pixel number within the smallest square bounding box, or mathematically,

$$E_{fit} = (DP_{in} + DP_{out}) / P_{total}, \quad (28)$$

where DP_{in} is the number of void element within the expected shape and DP_{out} is the number of solid element out of the expected shape, and P_{total} is the total number of pixels in the square bounding box.

Taking the circular shape in Fig. 8 as an example, the fitting error for each shape in the figure relative to the circular shape is calculated, and the corresponding fitting error is shown below each figure.



Fig. 8. A fitting example: fitting the five different shapes into a round shape with the fitting errors underneath respectively.

Five different types of databases were used for this test, each formed by microstructures containing randomly generated circular, square, upwards and downwards pointing triangle or cross-shaped holes respectively. We list six randomly selected samples of each type of database and our corresponding optimal microstructure in Fig. 9. The associated fitting error of each obtained optimal microstructure with respect to the five hole shapes is also given in Table 1. The fitting error in the gray box is that of the optimal microstructure with respect to its associated database. We can see from the table that the optimal microstructure always gives the smallest error, with a maximum of 4.4%. Otherwise the fitting errors are more than 17.1% when fitted with other shapes. The expected shapes in our optimal results are well preserved.

6.3. PCA performance

6.3.1. Optimal PCA base number

A key issue in our approach is to select the proper number of PCA bases to balance the accuracy and the computational cost, which is solved by the parallel analysis described in Section 2.2.1. To test its performance in our approach, we compared the different approximations to a given microstructure under different numbers of PCA bases. The approximate microstructure was obtained using Eq. (1).

In this case, the parallel analysis selected the first 100 bases of largest eigenvalues, as plotted in Fig. 10. We also manually selected other number of bases: 40, 60, 80, 120. The corresponding microstructures are shown in Fig. 11. As can be observed from the results, when the PCA base number increases, the protrusion noises decrease. In particular, when the number of PCA bases reaches 100, only shapes very approximate to disks are left. The improved results can be explained by the fact that more details of the microstructures are captured as the increase of the number of PCA bases, thus resulting in an optimal microstructure resembling the shape character of the database. However, when the number is raised to 120, the reconstructed microstructure is still close to the one obtained via 100 bases, with only minor improvements, and is thus not necessary. These results are consistent with the expectation that the parallel analysis will select an appropriate number of PCA bases.

6.3.2. PCA descriptor extraction speed

In this section, we compare the speed of extracting the PCA-based descriptor with those of extracting the first- and second-order descriptors.

Extracting PCA-based descriptor includes two main steps:

1. Preprocessing: performing PCA on the given microstructure database or solving an eigenvalue problem;
2. Extracting: decomposing the microstructure image into a linear combination of its microstructure of the PCA bases as defined in Eq. (1), which gives the descriptor $\{a_k\}_{k=0}^n$.

On the other hand, extracting the volume fraction (vf) and 2-point correlation ($P_2(1), P_2(2), \dots$) function is performed using Eqs. (2) and (4), called *previous descriptors* for short. A descriptor

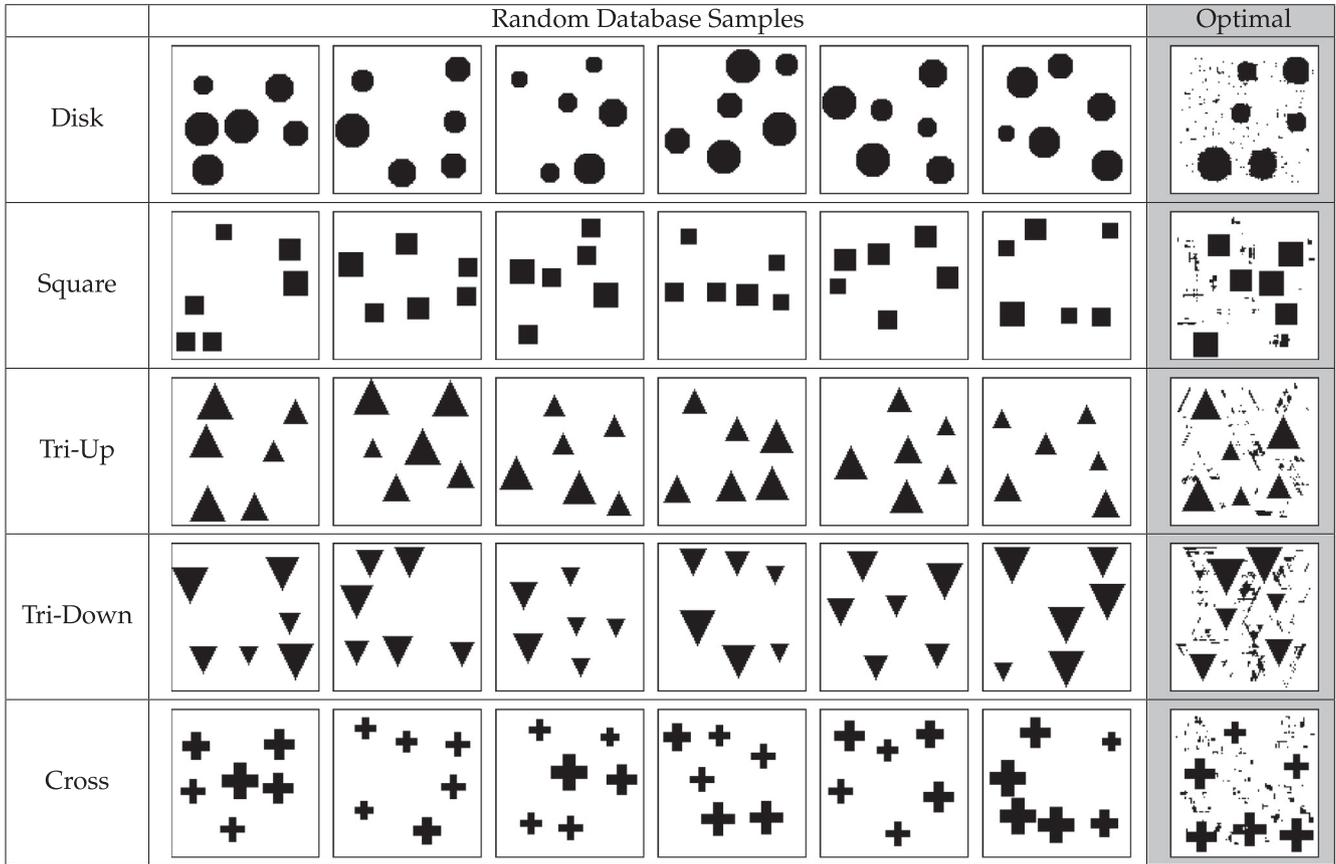


Fig. 9. We present the five different databases respectively with six random samples to show the different geometric characters (disk, square, triangle pointing up/down and cross) of the images in these databases. Also, we present the optimal result images using the corresponding database to show the shape influence on the results.

Table 1

Average error of fitting holes in the optimal result microstructure to a regular shape. Rows show the optimized results using for each database and each column corresponds to one fitting shape.

Error \ Result \ Shape	Disk	Square	Tri-Up	Tri-Down	Cross
Disk	4.4%	17.1%	38.3%	39.1%	23.9%
Square	21.5%	0%	50%	50%	44.4%
Tri-Up	36.1%	48.3%	2.1%	51.1%	23.2%
Tri-Down	36.7%	49.1%	52.9%	2.9%	22.9%
Cross	24.6%	44.4%	23.3%	22.8%	2.6%

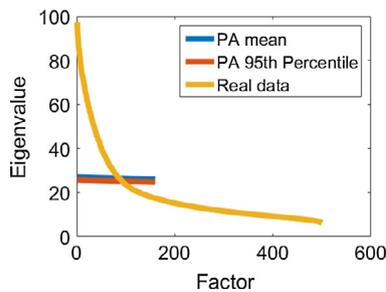


Fig. 10. The descending curve of sorted eigenvalues when using the parallel analysis to select PCA bases. Parallel analysis selects the PCA bases of which the eigenvalues are greater than the 95% of the random generated ones, and in this case 100 bases.

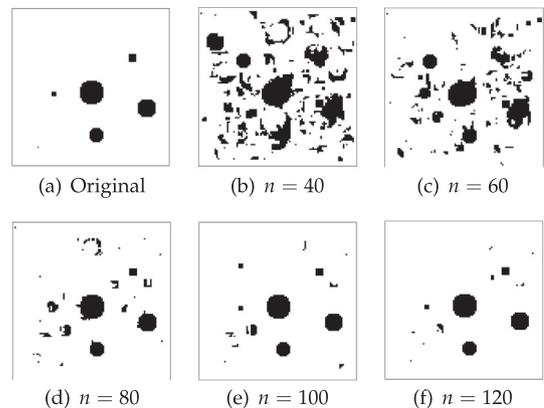


Fig. 11. The target microstructure image and five approximating images under different numbers of PCA bases.

Table 2

Time cost comparison on extracting PCA-based and previous descriptors. 'Prepro.' denotes the preprocessing step of PCA and 'Pre.' denotes the previous descriptors.

	Prepro.	Extract	Total	Speed up	Extract 10 times	Speed up	Extract 100 times	Speed up
PCA	0.52s	0.02s	0.54s	21.9×	0.72s	164.2×	2.52s	469.1×
Prev.	0s	$0.01 + 0.15 + 0.24 + 0.31 + 0.46 + \dots + 3.7 + 5.9s$	11.82s	/	118.21s	/	1182.13s	/

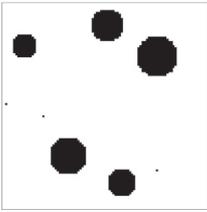
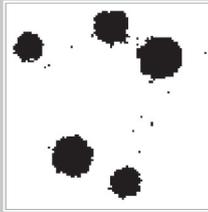
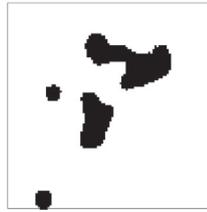
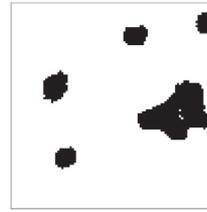
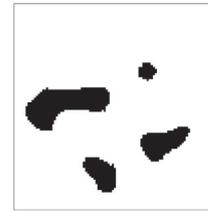
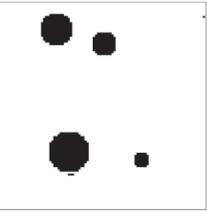
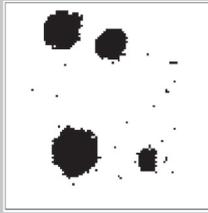
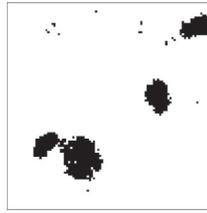
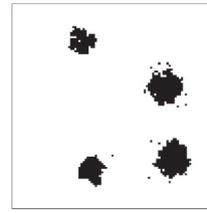
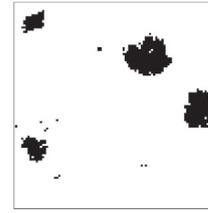
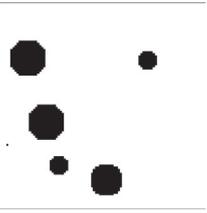
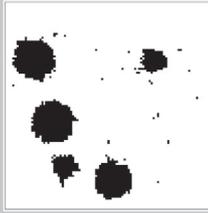
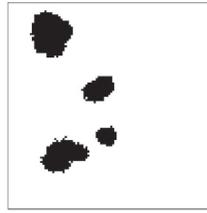
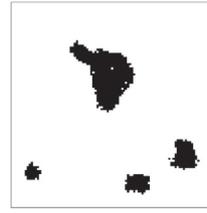
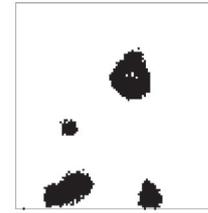
	Target	Ours	MC 1	MC 2	MC 3
Test 1 (Result)					
Test 2 (Result)					
Test 3 (Result)					
Test 1 (Time)	/	0.01s	137.2s	108.3s	91.8s
Test 2 (Time)	/	0.01s	95.1s	121.6s	89.4s
Test 3 (Time)	/	0.01s	96.4s	112.3s	118.5s

Fig. 12. Reconstruction results and time comparisons using our approach and Monte Carlo approach running three times (MC 1, 2, 3) on three different test examples (Test 1, 2, 3).

vector of 100-dimension is used here for both approaches for a randomly selected image. Specifically, the PCA-based descriptor uses its 100 highest eigenvalues, and the previous descriptor uses $v_f, P_2(1), P_2(2), \dots, P_2(99)$ as defined in Eqs. (2) and (4).

The time costs for extracting the descriptors for 1, 10 and 100 microstructure(s), respectively, are summarized in Table 2. In case of a single microstructure, the time of extracting previous descriptors is 21.9 times larger than that of our approach. In case of 10 and 100 images, the speed-up is respectively 164.2 times and 469.1 times, thus demonstrating the efficiency of the proposed descriptor.

6.4. Microstructure reconstruction speed and accuracy

The descriptor can also be mapped to its associated microstructure via a reconstruction process for later use. Numerous reconstruction approaches [13–19] have been proposed from the

first- and second- descriptors, and the proposed descriptor can also give a reconstructed microstructure via Eq. (1). Two different reconstruction approaches were also compared in this section, a recent approach using Monte Carlo approach [38,39] for comparison of aspects of computational time and reconstruction accuracy, as well as our approach.

6.4.1. Experimental settings

Using PCA descriptors for reconstruction consists of the two following steps:

1. Extract the PCA descriptors from the target microstructure image;
2. Construct the microstructure from the descriptor via Eq. (1).

Using previous descriptors for reconstruction consists of the two following steps:

1. Extract the first- and second- order descriptors from the target microstructure image, such as volume fraction (vf), cluster number (N) [7] and sampled 2-point correlation function ($P_2(1), P_2(2), P_2(3), \dots$) [7];
2. The structure starts from a random one, evolves using the probability given by the current and target descriptors' distance and stops when the distance is lower than a preset constant.

For the two approaches, we randomly select three reconstruction target microstructure from the database, and let the numbers of descriptors be 100 in both cases. As the reconstruction results using the first- and second- order descriptors are different every time we run the process, we thus ran the reconstruction three times (denoted respectively MC 1,2 or 3), all shown in Fig. 12.

6.4.2. Reconstruction time and accuracy comparisons

The reconstruction accuracy here is defined using the Chebyshev distance $L_{res,tar}^2$,

$$L_{res,tar}^2 = |(\mathbf{x}_{res} - \mathbf{x}_{tar})^T (\mathbf{x}_{res} - \mathbf{x}_{tar})|^{\frac{1}{2}},$$

where $\mathbf{x}_{res}, \mathbf{x}_{tar}$ are the density distributions of the reconstructed microstructure image and the target microstructure image, respectively.

As can be observed from these results, the reconstructed microstructure of our result is visually more similar to the target than the results obtained using the previous approach. Taking a specific examination of Test 1, and denoting the pixel value distribution respectively as $\mathbf{x}_{tar}, \mathbf{x}_{ours}, \mathbf{x}_{MC1}, \mathbf{x}_{MC2}, \mathbf{x}_{MC3}$, we have

$$L_{ours,tar}^2 = 3,$$

$$L_{MC1,tar}^2 = 25.1, \quad L_{MC2,tar}^2 = 33.2, \quad L_{MC3,tar}^2 = 30.4.$$

i.e. the error of our approach is much smaller.

Furthermore, the distances between the reconstructed results of the same approach were

$$L_{MC1,MC2}^2 = 46.8, \quad L_{MC2,MC3}^2 = 33.8, \quad L_{MC3,MC1}^2 = 42.9,$$

which is due to the random character of the Monte Carlo approach and is hard to produce a unique result.

In addition, the reconstruction time using the first- and second-order descriptors is around 100s which is more time-consuming compared to 0.01s taken by our approach. This can be explained from the fact that the proposed PCA-based descriptor only needs some additions for reconstruction, while the Monte Carlo approach needs numerous iterations.

7. Conclusions and future work

A material design approach is proposed in this paper based on a novel PCA-based descriptor. This descriptor is derived from a database of binary images, which can be obtained from SEM/TEM images after image processing. The novel descriptor has the following advantages over previous studies: First, it is computationally easier to extract than the classical 2-point correlation functions. Second, a microstructure can be directly reconstructed without expensive computational cost, and yields unique results, unlike previous reconstruction approaches. Third, a polynomial structure-performance relation is deduced. Based on this, the iterative simulation process was avoided, and thus the optimization can be solved much faster. Numerical experiments were also performed to testify these advantages.

The proposed approach is to be extended in the future in the following aspects. First, we are going to compare the optimization results with those obtained with classical topology optimization

approaches. Second, we are going to explore other dimension reduction based descriptors to explore their usages in microstructure design. Data mining approaches are also to be introduced to extract the appropriate microstructure descriptor that are essential in determining their physical properties.

Acknowledgment

The work described in this paper is partially supported by the NSF of China (Nos. 61472356, 61210007), and the National Key Research and Development Program (2016YFC1101302) from the MIST of China.

References

- [1] D.L. McDowell, G.B. Olson, Concurrent design of hierarchical materials and structures, in: *Scientific Modeling and Simulations*, Springer, 2009, pp. 207–240.
- [2] G.B. Olson, Computational design of hierarchically structured materials, *Science* 277 (5330) (1997) 1237–1242.
- [3] J.H. Panchal, S.R. Kalidindi, D.L. McDowell, Key computational modeling issues in integrated computational materials engineering, *Comput.-Aid. Des.* 45 (1) (2013) 4–25.
- [4] N. Charalambakis, Homogenization techniques and micromechanics: a survey and perspectives, *Appl. Mech. Rev.* 63 (3) (2010) 030803.
- [5] P. Ming, X. Yue, Numerical methods for multiscale elliptic problems, *J. Comput. Phys.* 214 (1) (2006) 421–445.
- [6] G. Pavliotis, A. Stuart, *Multiscale Methods: Averaging and Homogenization*, vol. 53, Springer, 2008.
- [7] H. Xu, Y. Li, C. Brinson, W. Chen, Descriptor-based methodology for designing heterogeneous microstructural materials system, in: *ASME 2013 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, American Society of Mechanical Engineers, 2013, pp. V03AT03A049–V03AT03A049.
- [8] H. Xu, R. Liu, A. Choudhary, W. Chen, A machine learning-based design representation method for designing heterogeneous microstructures, in: *ASME 2014 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, American Society of Mechanical Engineers, 2014, pp. V02BT03A009–V02BT03A009.
- [9] H. Xu, D.A. Dikin, C. Burkhart, W. Chen, Descriptor-based methodology for statistical characterization and 3d reconstruction of microstructural materials, *Comput. Mater. Sci.* 85 (2014) 206–216.
- [10] B.L. Adams, S.R. Kalidindi, D.T. Fullwood, *Microstructure sensitive design for performance optimization*, Butterworth-Heinemann, 2012.
- [11] D.T. Fullwood, S.R. Niezgod, B.L. Adams, S.R. Kalidindi, Microstructure sensitive design for performance optimization, *Prog. Mater. Sci.* 55 (6) (2010) 477–562.
- [12] C. Qi, Y. Wang, Feature-based crystal construction in computer-aided nano-design, *Comput.-Aid. Des.* 41 (11) (2009) 792–800.
- [13] D. Basanta, P.J. Bentley, M.A. Miodownik, E.A. Holm, Evolving cellular automata to grow microstructures, in: *Genetic Programming*, Springer, 2003, pp. 1–10.
- [14] N.C. Kumar, K. Matouš, P.H. Geubelle, Reconstruction of periodic unit cells of multimodal random particulate composites using genetic algorithms, *Comput. Mater. Sci.* 42 (2) (2008) 352–367.
- [15] A.P. Roberts, Statistical reconstruction of three-dimensional porous media from two-dimensional images, *Phys. Rev. E* 56 (3) (1997) 3203.
- [16] A.P. Roberts, M. Teubner, Transport properties of heterogeneous materials derived from gaussian random fields: bounds and simulation, *Phys. Rev. E* 51 (5) (1995) 4141.
- [17] B.L. Hansen, B.L. Adams, M.E. Lyon, A.J. Henrie, On the reconstruction of polycrystalline microstructures from two-point correlation statistics, *J. Comput.-Aid. Mater. Des.* 10 (3) (2005) 163–173.
- [18] M.S. Talukdar, O. Torsæter, M.A. Ioannidis, Stochastic reconstruction of particulate media from two-dimensional images, *J. Colloid Interface Sci.* 248 (2) (2002) 419–428.
- [19] C.L.Y. Yeong, S. Torquato, Reconstructing random media, *Phys. Rev. E* 57 (1) (1998) 495.
- [20] M. Sonka, V. Hlavac, R. Boyle, *Image processing, analysis, and machine vision*, Cengage Learning, 2014.
- [21] L. Kobbelt, M. Botsch, A survey of point-based techniques in computer graphics, *Comput. Graph.* 28 (6) (2004) 801–814.
- [22] H. Abdi, Principal component analysis, *Wiley Interdisc. Rev.: Comput. Stat.* 2 (4) (2010) 433–459.
- [23] M. Li, B. Yuan, 2D-LDA: a statistical linear discriminant analysis for image matrix, *Pattern Recogn. Lett.* 26 (5) (2005) 527–532.
- [24] K.Q. Weinberger, F. Sha, L.K. Saul, Learning a kernel matrix for nonlinear dimensionality reduction, in: *Proceedings of the Twenty-First International Conference on Machine Learning*, ACM, 2004, p. 106.
- [25] G.E. Hinton, S.T. Roweis, Stochastic neighbor embedding, in: *Advances in Neural Information Processing Systems*, 2002, pp. 833–840.
- [26] L.V.D. Maaten, G.E. Hinton, Visualizing data using t-SNE, *J. Machine Learn. Res.* 9 (Nov) (2008) 2579–2605.

- [27] J. Tang, J. Liu, M. Zhang, Q. Mei, Visualizing large-scale and high-dimensional data, in: Proceedings of the 25th International Conference on World Wide Web, International World Wide Web Conferences Steering Committee, 2016, pp. 287–297.
- [28] S.R. Kalidindi, S.R. Niezgodna, A.A. Salem, Microstructure informatics using higher-order statistics and efficient data-mining protocols, *Jom* 63 (4) (2011) 34–41.
- [29] H. Xu, Y. Li, Y. Chen, J. Barbič, Interactive material design using model reduction, *ACM Trans. Graph. (TOG)* 34 (2) (2015) 1–14.
- [30] S. Li, J. Huang, F.D. Goes, X. Jin, H. Bao, M. Desbrun, Space-time editing of elastic motion through material optimization and reduction, *ACM Trans. Graph.* 33 (4) (2014), Art–No.
- [31] S.R. Niezgodna, Y.C. Yabansu, S.R. Kalidindi, Understanding and visualizing microstructure and microstructure variance as a stochastic process, *Acta Mater.* 59 (16) (2011) 6387–6400.
- [32] J. Yang, D. Zhang, A.F. Frangi, J. Yang, Two-dimensional PCA: a new approach to appearance-based face representation and recognition, *IEEE Trans. Pattern Anal. Machine Intell.* 26 (1) (2004) 131–137.
- [33] M.A. Akgün, J.H. Garcelon, R.T. Haftka, Fast exact linear and non-linear structural reanalysis and the sherman–morrison–woodbury formulas, *Int. J. Numer. Methods Eng.* 50 (7) (2001) 1587–1606.
- [34] C. Xu, S. Gao, M. Li, Heterogeneous material design using a PCA-based microstructure representing method, in: ASME 2015 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, American Society of Mechanical Engineers, 2015, pp. V02BT03A009–V02BT03A009.
- [35] J.C. Hayton, D.G. Allen, V. Scarpello, Factor retention decisions in exploratory factor analysis: a tutorial on parallel analysis, *Organiz. Res. Methods* 7 (2) (2004) 191–205.
- [36] M.F. Ashby, A. Evans, N.A. Fleck, L.J. Gibson, J.W. Hutchinson, H. Wadley, F. Delale, Metal foams: a design guide, *Appl. Mech. Rev.* 23 (6) (2002), 119–119.
- [37] S. Amstutz, A penalty method for topology optimization subject to a pointwise state constraint, *Esaim Control Optim. Calculus Variations* 16 (3) (2010) 523–544.
- [38] C.J. Gommers, Y. Jiao, S. Torquato, Microstructural degeneracy associated with a two-point correlation function and its information content, *Phys. Rev. E: Stat. Phys. Plasmas Fluids Relat. Interdisc. Top.* 85 (5) (2012) 747–761.
- [39] M. Baniassadi, M. Safdari, H. Garmestani, S. Ahzi, P.H. Geubelle, Y. Remond, An optimum approximation of n-point correlation functions of random heterogeneous material systems, *J. Chem. Phys.* 140 (7) (2014) 530–541.