A superior descriptor of random textures and its predictive capacity

Y. Jiao[a], F. H. Stillinger[b], and S. Torquato[a,b,c,1]

[a]Department of Mechanical and Aerospace Engineering Princeton University, Princeton, NJ 08544; [b]Department of Chemistry, Princeton University, Princeton, NJ 08544; and [c]Princeton Center for Theoretical Physics, Princeton University, Princeton, NJ 08544

Edited by Hans C. Andersen, Stanford University, Stanford, CA, and approved August 6, 2009 (received for review May 28, 2009)

Two-phase random textures abound in a host of contexts, including porous and composite media, ecological structures, biological media, and astrophysical structures. Questions surrounding the spatial structure of such textures continue to pose many theoretical challenges. For example, can two-point correlation functions be identified that can be manageably measured and yet reflect nontrivial higher-order structural information about the textures? We present a solution to this question by probing the information content of the widest class of different types of two-point functions examined to date using inverse "reconstruction" techniques. This enables us to show that a superior descriptor is the two-point cluster function $C_2(r)$, which is sensitive to topological connectedness information. We demonstrate the utility of $C_2(r)$ by accurately reconstructing textures drawn from materials science, cosmology, and granular media, among other examples. Our work suggests a theoretical pathway to predict the bulk physical properties of random textures and that also has important ramifications for atomic and molecular systems.

two-point cluster function | reconstruction

Two-phase random textures are ubiquitous in nature and synthetic situations. Examples include heterogeneous materials (e.g., composites, porous media and colloids) (1, 2), geologic media (e.g., rock formations) (3), ecological structures (e.g., tree patterns in forests) (4), cosmological structures (e.g., galaxy distributions) (5, 6), and biological media (e.g., animal and plant tissue) (7). Over a broad range of length scales, two-phase random textures exhibit a rich variety of structures with varying degrees of disorder and complex bulk properties (8–10), and questions concerning their quantitative characterizations continue to present many fundamental and practical challenges.

It is well known that an infinite set of n-point correlation functions is generally required to completely statistically characterize such textures and their physical properties in the infinite-volume limit. A variety of different types of correlation functions arise in rigorous theories of structure/property relations (1). One such basic quantity is the standard n-point correlation function $S_n(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$, which gives the probability of finding n points at positions $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ all in one of the phases (1, 11). Because the information contained in such an infinite set of functions is generally unattainable in practice, a natural starting point is to characterize the structure and bulk properties of random textures by using lower-order versions. The two-point function S_2, which is experimentally accessible via scattering of radiation (12), provides information about the distribution of pair separations. The three-point function S_3 reveals information about how these pair separations involved in S_2 are linked into triangles. The four-point function S_4 controls the assembly of triangles into tetrahedra. A natural question is how much additional useful information do S_3 and S_4 contain over and above S_2? We will show that triangular and tetrahedral statistics do not significantly increase information content over and above pair statistics for textures possessing no long-range order.

Therefore, an outstanding problem in condensed matter theory is to identify other two-point correlation functions that can be both manageably measured and yet reflect nontrivial higher-order structural information about the textures. The aim of this article is to provide a solution to this problem using inverse techniques; specifically, "reconstruction" methods. The purpose of a reconstruction (construction) technique is to reconstruct (construct) realizations of two-phase random textures that match limited structural information on the textures in the form of lower-order correlation functions (1, 13), which are called the "target" functions. An effective reconstruction procedure enables one to generate accurate renditions of random textures at will (14–20), and subsequent analysis can be performed on the reconstruction to obtain desired macroscopic properties of the texture nondestructively (14, 18). Here, we use this inverse methodology to determine the amount of structural information that is embodied in a set of targeted correlation function by quantifying the extent to which the original structure can be accurately reconstructed by using those target functions. We quantify the accuracy of a reconstruction by measuring unconstrained (untargeted) correlation functions and comparing them with those of the original medium.

We adapt the inverse reconstruction method of Yeong and Torquato (13, 14) to show that a superior two-point signature of random textures is the two-point cluster function $C_2(\mathbf{r})$ (21), which is sensitive to topological connectedness information. We demonstrate that $C_2(\mathbf{r})$ contains not only appreciably more information than S_2, but more information than a variety of other "two-point" quantities, including the surface–surface correlation function F_{ss}, the surface–void correlation function F_{sv}, the pore-size function F, lineal-path function L, and the chord-length density function p (1). (All the aforementioned correlation functions are defined precisely below.) Such a comprehensive study that investigates and compares all of the aforementioned statistical descriptors via inverse techniques has heretofore not been undertaken. Our results have practical implications for materials science, liquid- and solid-state physics, biological systems, cosmology, hydrology, and ecology.

Definitions of the Two-Point Correlation Functions

In the theory of random media, a variety of different types of two-point correlation functions naturally arise (1). Here, we define seven different ones that will be employed in this article. Consider d-dimensional two-phase textures in which phase i has volume fraction ϕ_i ($i = 1, 2$) and is characterized by the indicator function

$$\mathcal{I}^{(i)}(\mathbf{x}) = \begin{cases} 1, & \mathbf{x} \in \mathcal{V}_i, \\ 0, & \text{otherwise}, \end{cases} \qquad [1]$$

where $\mathcal{V}_i$ is the region occupied by phase i (equal to 1 or 2). Note that here "phase" is used in a general sense in that it can refer

Author contributions: S.T. designed research; Y.J., F.H.S., and S.T. performed research; Y.J., F.H.S., and S.T. contributed new reagents/analytic tools; Y.J., F.H.S., and S.T. analyzed data; and Y.J., F.H.S., and S.T. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. E-mail: torquato@electron.princeton.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0905919106/DCSupplemental.

to a solid, liquid or even void. The standard two-point correlation function is defined as

$$S_2^{(i)}(\mathbf{x}_1, \mathbf{x}_2) = \left\langle \mathcal{I}^{(i)}(\mathbf{x}_1)\mathcal{I}^{(i)}(\mathbf{x}_2) \right\rangle, \quad [2]$$

where angular brackets denote an ensemble average. This function is the probability of finding two points $\mathbf{x}_1$ and $\mathbf{x}_2$ both in phase i. Henceforth, we will drop the superscript "i" and only consider the correlation functions for the phase of interest. For statistically homogeneous and isotropic microstructures, which is the focus of the rest of the article, two-point correlation functions will only depend on the distance $r \equiv |\mathbf{x}_1 - \mathbf{x}_2|$ between the points and hence $S_2(\mathbf{x}_1, \mathbf{x}_2) = S_2(r)$.

The surface–void F_{sv} and surface–surface F_{ss} correlation functions are respectively defined as

$$F_{sv}(r) = \langle \mathcal{M}(\mathbf{x}_1)\mathcal{I}(\mathbf{x}_2) \rangle, \quad F_{ss}(r) = \langle \mathcal{M}(\mathbf{x}_1)\mathcal{M}(\mathbf{x}_2) \rangle, \quad [3]$$

where $\mathcal{M}(\mathbf{x}) = |\nabla \mathcal{I}(\mathbf{x})|$ is the two-phase interface indicator function. By associating a finite thickness with the interface, F_{sv} and F_{ss} can be interpreted, respectively, as the the probability of finding $\mathbf{x}_1$ in the "dilated" interface region and $\mathbf{x}_2$ in the void phase and the probability of finding both $\mathbf{x}_1$ and $\mathbf{x}_2$ in the dilated interface region but in the limit that the thickness tends to zero (1).

The lineal-path function $L(r)$ is the probability that an entire line of length r lies in the phase of interest, and thus contains a coarse level of connectedness information, albeit only along a lineal path (1, 22). The chord-length density function $p(r)$ gives the probability associated with finding a "chord" of length r in the phase of interest and is directly proportional to the second derivative of $L(r)$ (23). (Chords are the line segments between the intersections of an infinitely long line with the two-phase interface.) The pore-size function $F(\delta)$ is related to the probability that a sphere of radius r can lie entirely in the phase of interest (1) and therefore is the three-dimensional "spherical" version of the lineal measure L.

The two-point cluster function $C_2(r)$ gives the probability of finding two points separated by a distance r in the same cluster of the phase of interest (21). Note that a cluster of a phase is any topologically connected subset of that phase. The two-point cluster function can be measured experimentally using any appropriate three-dimensional imaging technique (e.g., tomography, confocal microscopy and MRI) (1). The fact that C_2 contains intrinsic three-dimensional topological information is to be contrasted with S_2, which can be obtained from a planar cross-section of the texture. In general, C_2 is expected to embody a much greater level of three-dimensional connectedness information than either L or F, but the degree to which this is true has yet to be quantitatively demonstrated, which is one of the aims of this article.

Inverse Reconstruction Technique

The stochastic optimization reconstruction algorithm for digitized media formulated by Yeong and Torquato (13) is ideally suited to carry out the aforementioned analysis because it can incorporate different types of target statistical descriptors. This algorithm is both robust and simple to implement (15–20). In this method, one starts with an initial realization of a random medium and a set of target correlation functions $\hat{f}_n^1(\mathbf{R}), \hat{f}_n^2(\mathbf{R}), \hat{f}_n^3(\mathbf{R}), \ldots$, which are obtained (i.e., measured) from the medium of interest. Here, $\hat{f}_n^\alpha(\mathbf{R})$ is an n-point correlation function of type α, $\mathbf{R} \equiv \mathbf{r}_1, \mathbf{r}_2, \ldots \mathbf{r}_n$, and $\mathbf{r}_i$ denotes the position vector of the ith point. The method proceeds to find a realization in which calculated correlation functions $f_n^1(\mathbf{R}), f_n^2(\mathbf{R}), f_n^3(\mathbf{R}), \ldots$ best match the corresponding target functions. This is achieved by minimizing an "energy"

$$E = \sum_{\mathbf{R}} \sum_{\alpha} \left[f_n^\alpha(\mathbf{R}) - \hat{f}_n^\alpha(\mathbf{R}) \right]^2, \quad [4]$$

which is defined to be the sum of squared differences between the calculated and target functions, via a simulated annealing method in which a sequence of trial realizations is generated and accepted with the probability $\min\{\exp(-\Delta E/T), 1\}$, where ΔE is the energy difference between the new and old realizations and T is a fictitious temperature. The initially "high" temperature is lowered according to a prescribed annealing schedule until the energy of the system approaches its ground-state value within a very small tolerance level.

Degeneracy of Ground States Using S_2 Alone. It is instructive to consider the energy defined by Eq. **4** when only the standard two-point correlation function S_2 is used. In this special instance, the energy is given by

$$E = \sum_r \left[S_2(r) - \hat{S}_2(r) \right]^2, \quad [5]$$

where $\hat{S}_2$ and S_2 are the two-point correlation functions of the target and reconstructed medium, respectively. We note that most previous reconstruction studies have only tried to match S_2. However, it is now well established that $S_2(\mathbf{r})$ is not sufficient information to generally get an accurate rendition of the original microstructure (13–15, 19, 20). In other words, the ground states when only S_2 is incorporated in the energy (Eq. **5**) are highly degenerate due to the nonuniqueness of the information content of this two-point function, which is clearly illustrated by the subsequent examples in the paper. The reader is also referred to supporting information (SI) *Text*, which provides a rigorous explanation for the degeneracy of the ground states for digitized representations of textures.

Insufficiency of Conventional Three-Point Information. An obvious additional set of correlation functions that could be incorporated in the reconstruction is the higher-order versions of S_2, namely, S_3, S_4, etc. However, not only is the three-point correlation function S_3 more difficult to compute, it is not at all clear that its incorporation will result in appreciably better reconstructions because it only introduces local information about triangles when there is no long-range order, the most common occurrence. We can quantitatively verify the insufficiency of conventional triangular information by reconstructing a one-dimensional equilibrium distribution of equal-sized hard rods (24) using S_3 (see *SI Text* for technical details). Fig. 1 compares this reconstruction with those involving S_2 alone and a combination of S_2 and C_2. It is clear that the reconstruction using S_2 only results erroneously in a highly clustered "rod" phase. Although incorporating S_3 provides an improved reconstruction, it still contains large clusters and isolated rods that can be much smaller than actual rod size. On the other hand, the S_2–C_2 hybrid reconstruction produces the most accurate rendition of the target medium. Although this one-dimensional example is suggestive that C_2 contains nontrivial structural information in excess to what is contained in S_2, one must investigate this problem in higher dimensions, which presents algorithmic challenges, as we will describe in the next section.

Efficient Algorithmic Implementation of the General Problem

Here, we present a general methodology that enables one to efficiently incorporate a wide class of lower-order correlation functions in the Yeong–Torquato reconstruction procedure. The aforementioned probabilistic interpretations of the correlation functions enable us to develop a general sampling method for reconstruction of statistically homogeneous and isotropic digitized textures based on the "lattice-gas" formalism, which was introduced in ref. 20 and has been generalized here. In the generalized formalism, pixels with different values (occupying the lattice sites) correspond to distinct local states and pixels with the same value are considered to be "molecules" of the same

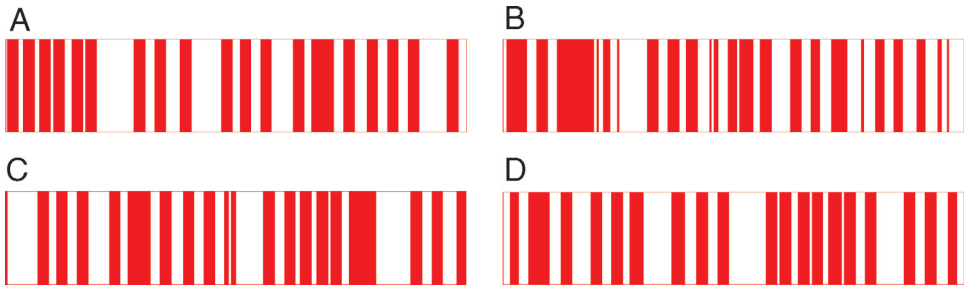

Fig. 1. Equilibrium hard rods. (*A*) Target: an equilibrium hard-rod system in which the equal-sized rods cover 50% of the space. Each rod in the system is 10 pixels in length. (*B*) S_2-alone reconstruction. (*C*) S_3 reconstruction. (*D*) S_2–C_2 hybrid reconstruction. For visualization purposes, the one-dimensional rod systems are artificially extended in the vertical direction.

"gas" species (20). The correlation functions of interest can be obtained by binning the separation distances between the selected pairs of molecules from particular species.

In the case of S_2, all molecules are of the same species. We denote the number of lattice-site separation distances of length r by $N_S(r)$ and the number of molecule-pair separation distances of length r by $N_P(r)$. Thus, the fraction of pair distances with both ends occupied by the phase of interest, i.e., the two-point correlation function, is given by $S_2(r) = N_P(r)/N_S(r)$. To obtain C_2, one needs to partition the molecules into different subsets Γ_i (species) such that any two molecules of the same species are connected by a path composed of the same kind of molecules, i.e., molecules that form a cluster, which is identified using the "burning" algorithm (25). The number of pair distances of length r between the molecules within the same subset Γ_i is denoted by $N_P^i(r)$. The two-point cluster function is then given by $C_2(r) = \sum_i N_P^i(r)/N_S(r)$. The calculation of F_{ss} and F_{sv} requires partitioning the molecules into two subsets: the surface set κ_S containing only the molecules on the surfaces of the clusters and the volume set κ_V containing the rest. In a digitized medium, the interface necessarily has a small but finite thickness determined by the pixel size. Thus, the surface–surface and surface–void correlation functions can be regarded as probabilities that are given by $F_{ss} = N^{ss}(r)/N_S(r)$ and $F_{sv} = N^{sv}(r)/N_S(r)$, respectively; where $N^{ss}(r)$ gives the number of distances between two surface molecules with length r and N^{sv} is the counterpart for pairs with one molecule on the surface and the other inside the cluster.

The lineal path function L can be obtained by computing the lengths of all digitized line segments (chords) composed of pixels of the phase of interest, and for each chord incrementing the counters associated with the distances equal to and less than that chord length (13, 14). The chord-length density function p can then be easily obtained by binning the chord lengths that are used to compute L (1). The pore-size function F can be computed by finding the minimal separation distances of pixels within the phase of interest to those at the two-phase interfaces. The minimal distances are then binned to obtained a probability distribution function, the complementary cumulative distribution function of which is F (3).

We have also devised methods to track clusters and surfaces that enable one to quickly compute the correlation functions of the new realization based on the old ones (see *SI Text* for technical details) and thus make the Yeong–Torquato reconstruction procedure much more efficient than methods that directly resample the correlation functions for each trial realization.

We have used this general procedure here to reconstruct a wide spectrum of random textures, including model microstructures, such as the cherry-pit model, equilibrium hard spheres, Debye random media and symmetric cell materials (1), as well as the digitized representations of sandstones (3), metal-ceramic composites (26), concrete microstructures (27), laser speckle patterns (20) and galaxy distributions. Using the aforementioned largest set of correlation functions utilized to date, our analysis reveals that the best reconstructions always incorporate the two-point cluster function $C_2(r)$. In what follows, we will present specific results for only a subset of the correlation functions that we used, namely, various combinations of S_2, F_{ss}, F and C_2, for the nonpercolating phases of a concrete microstructure (27), a distribution of galaxies, and a three-dimensional hard-sphere packing.

Concrete Microstructure. The wide range of structural features in concrete, from nanometer-sized pores to centimeter-sized aggregates, makes it a wonderful example of a multi scale microstructure (28). Fig. 2*A* shows a binarized digitized image of a concrete sample cross-section. We have thresholded the original image so that the blue phase represents the stones and the lighter gray phase is the cement paste. The "stone" phase is characterized by a dense dispersion of "particles" of various sizes: a nontrivial situation to reconstruct. Using S_2 alone overestimates clustering in the system and indeed incorrectly yields a percolating particle phase. Thus, although S_2 of the reconstruction matches the target one with very small error (see Fig. S1), such information is insufficient to get a good reconstruction. Incorporating both S_2 and surface–surface function F_{ss} leads to a better rendition of the target system but the reconstruction still overestimates the degree of clustering. On the other hand, incorporating C_2 yields an excellent reconstruction in that the stone phase clearly appears as a particle dispersion with a size distribution that closely matches that of the target structure. As noted in the introduction, we can quantitatively test the accuracy of the reconstructions by measuring unconstrained correlation functions and comparing them with the corresponding quantities of the target system. Here, we choose to compute the unconstrained lineal-path function L. Fig. 2*E* reveals that the lineal-path function of the reconstruction that incorporates C_2 matches the target function L well and it clearly is appreciably more accurate than the other reconstructions.

Galaxy Clusters. Correlation functions have been used to understand the formation of galaxies and the large scale structures of the Universe (4, 5). Such characterizations are becoming increasingly important with the advent of high-quality surveys in cosmology. We suggest that the reconstruction procedure can provide an important tool in cosmological and astrophysical applications, especially in helping to determine the lower-order correlation functions that reflect a priori information about nontrivial structural features, such as multi scale clustering and filamentary structures in the Universe. Fig. 3*A* shows a binarized image of a portion of the Abell 1689 galaxy cluster. We have chosen the binarizing threshold such that the "galaxy" phase (bright spots in Fig. 3*A*) exhibits different sized clusters. As one can see, the "primary" (single largest) cluster in the target system has been reproduced by all the reconstructions. This is because the volume fraction of the "galaxy" phase is relatively small, whereas the target S_2 has a relatively long tail (see Fig. S2), which requires the clustering of a majority of galaxy phase. However, the "secondary" (smaller

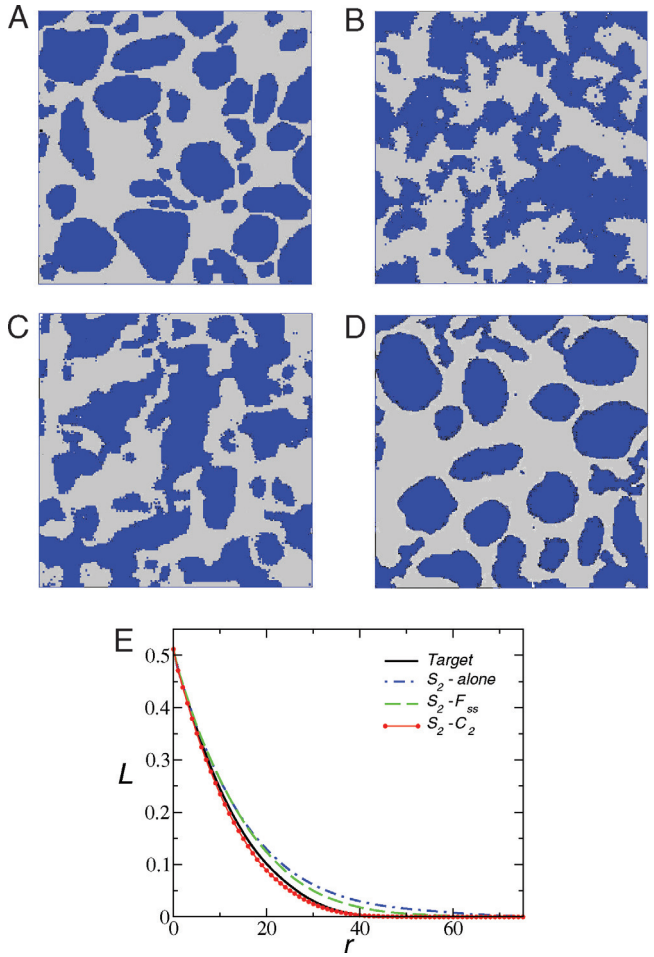

Fig. 2. Concrete microstructure. (*A*) Target system: a binarized image of a cross-section of concrete (27). The linear size of the digitized texture is $N_L = 170$ pixels. (*B*) S_2-alone reconstruction. (*C*) S_2–F_{ss} hybrid reconstruction. (*D*) S_2–C_2 hybrid reconstruction. All the reconstructions are associated with a final energy (error) $E \approx 10^{-8}$. (*E*) The unconstrained lineal-path function L of the reconstructions and the target image. Pixel size supplies the unit for the distance r.

compact) clusters are significantly different for different reconstructions. The reconstruction using S_2 only produces large elongated secondary clusters. Incorporating F_{ss} enables one to obtain a better rendition; however, it still contains a few elongated secondary clusters. The incorporation of C_2 again provides the most accurate reconstruction. This is visually verified by examination of both the sizes and shapes of the primary and secondary clusters, and is quantitatively confirmed by the comparison of the unconstrained L of the reconstructed and target systems, as shown in Fig. 3*E*.

Sphere Packing. As an application of our methodology to three dimensions, we have reconstructed a digitized realization of an equilibrium distribution of equal-sized hard spheres, as shown in Fig. 4*A*. This packing is generated using the standard Metropolis Monte Carlo technique for a canonical ensemble of hard spheres in a cubical box under periodic boundary conditions (1). A visual comparison of the hybrid reconstruction involving the two-point cluster function reveals that it accurately yields a dispersion of well-defined spherical inclusions of the same size, in contrast to the S_2 reconstruction, which again grossly overestimates clustering of the "sphere" phase. In contrast to the previous examples, here we incorporate the pore-size function F (not the surface correlation functions) in one of the reconstructions. Although the reconstruction incorporating F provides improvement over the rendition of the S_2-alone reconstruction, it is still inferior to the S_2–C_2 reconstruction in reproducing both the size and shape of the "sphere" phase. (The target and reconstructed correlation functions are shown in Fig. S3 and the *SI Text*.) The accuracy of the S_2–C_2 hybrid reconstruction can also be seen by comparing the unconstrained lineal-path function L of the target system with those of the reconstructed media (see Fig. 4*E*).

Discussion

In summary, although it was known that the information content of the standard two-point function S_2 of a random texture is far from complete, we have demonstrated here that the next higher-order version S_3 generally does not contain appreciably greater information. The fact that this natural extension to incorporate higher-order S_n, which has been pursued in the last century in statistical mechanics, is not a fruitful path motivated us to inquire whether there exist sensitive two-point statistical descriptors that embody nontrivial structural information. We probed the information content of seven different types of two-point functions using inverse reconstruction methods. For all of the examples studied here, reconstructions that include the two-point cluster function C_2 were always found to be significantly more accurate than those involving any of the combinations of pairs of the other functions.

More precisely, the incorporation of C_2 significantly reduces the number of compatible microstructures as compared to the

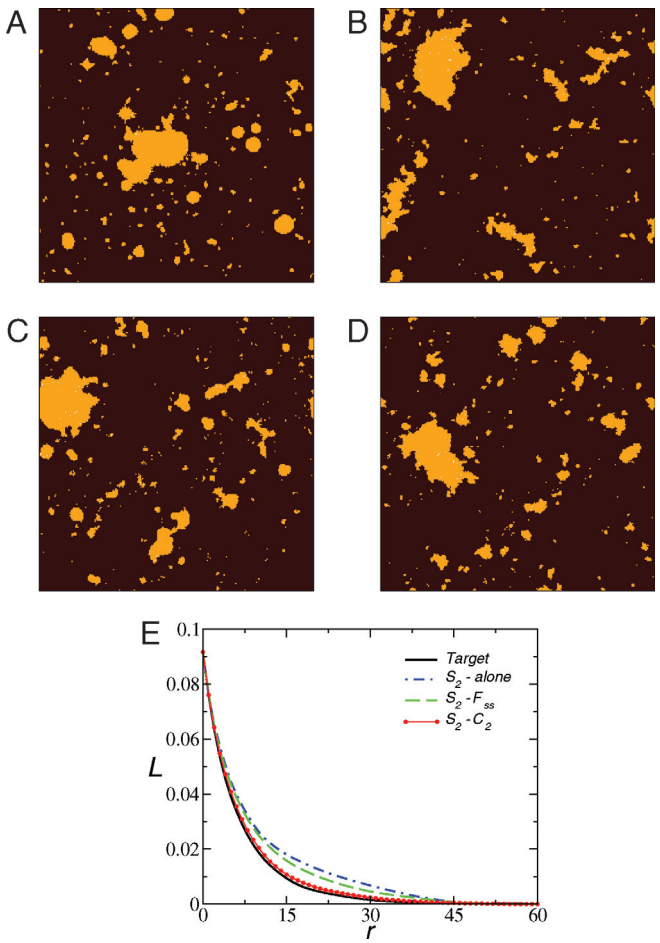

Fig. 3. Galaxy clusters. (*A*) Target system: a portion of the Coma Cluster. The linear size of the digitized texture is $N_L = 238$ pixels. (*B*) S_2-alone reconstruction. (*C*) S_2–F_{ss} hybrid reconstruction. (*D*) S_2–C_2 hybrid reconstruction. All the reconstructions are associated with a final energy (error) $E \approx 10^{-8}$. (*E*) The unconstrained lineal-path function L of the reconstructed and target systems. Pixel size supplies the unit for the distance r.

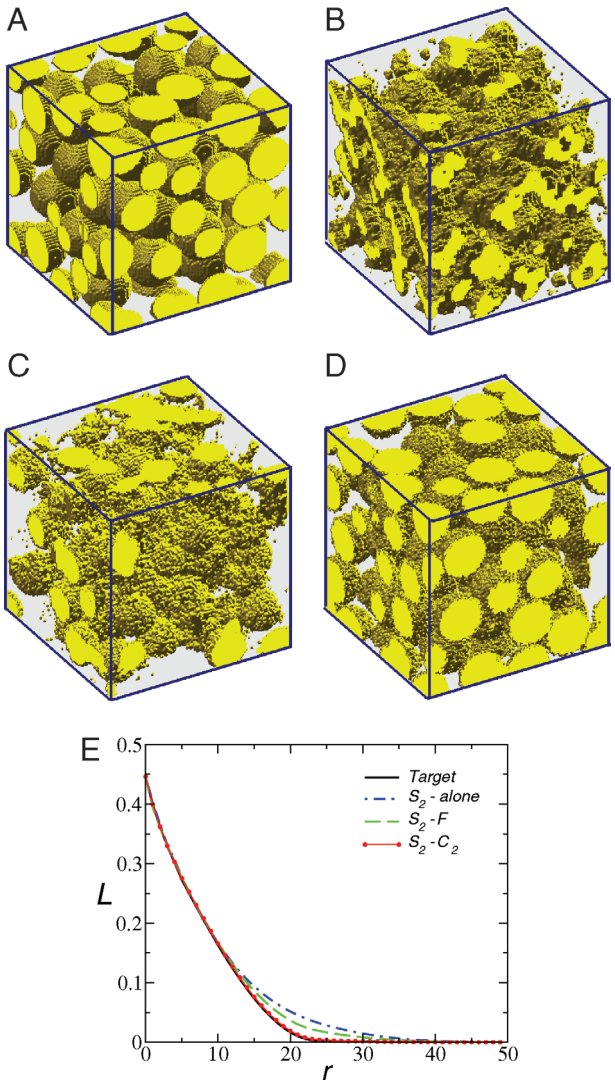

Fig. 4. Sphere packing. (*A*) Target system: a digitized realization of a hard-sphere packing in which the spheres occupy 44.6% of space. The linear size of the digitized texture is $N_L = 100$ pixels. (*B*) S_2-alone reconstruction. (*C*) S_2–*F* reconstruction. (*D*) S_2–C_2 hybrid reconstruction. All the reconstructions are associated with a final energy (error) $E \approx 10^{-11}$. (*E*) The (unconstrained) lineal-path function sampled from target and reconstructed realizations. Pixel size supplies the unit for the distance *r*.

compatible microstructures consistent with the same three-point function S_3, which is schematically indicated in Fig. 5. The two-point cluster function is an especially sensitive structural signature when clustering and phase connectedness are present, the most difficult situations to treat. This can be seen from the reconstruction of the hard-rod system that incorporated C_2 (see Fig. 1). More importantly, we showed the utility of $C_2(\mathbf{r})$ in higher dimensions by accurately reconstructing galaxy distributions, concrete microstructures and dense hard-sphere packings, among other examples.

Why is $C_2(r)$ a superior two-point structural signature? To answer this question, it is useful to first compare it to $S_2(r)$. The latter, unlike $C_2(r)$, does not distinguish between events in which the end points of the line segment of length r fall in the same cluster of a particular phase and those that do not involve the same cluster of that phase (21). More precisely, C_2 is the "connectedness" contribution to the standard two-point correlation function, i.e.,

$$S_2(r) = C_2(r) + D_2(r), \qquad [6]$$

where D_2 measures the probability that the end points of a line segment of length r fall in different clusters of the phase of interest. Therefore, whereas $S_2(r)$ is insensitive to clustering and percolation, $C_2(r)$ becomes a progressively longer-ranged function as clusters grow in size such that its volume integral diverges at the percolation threshold (1, 21). By contrast, the quantities L, p, F, F_{ss} and F_{sv} are insensitive to crossing the percolation threshold. Indeed, for particle systems, one can show that C_2 is a functional of the infinite set of n-particle "connectedness" functions (21).

Thus, although C_2 is a "two-point" quantity, it actually embodies higher-order structural information in a way that makes it a highly sensitive statistical descriptor over and above S_2. This ability to "leapfrog" past the usual approach of incorporating additional information via higher-order versions of S_2 has important ramifications for new structure/property relation of random textures. Specifically, our work suggests that theories should be developed that relate the transport, mechanical, chemical and optical properties of random textures to functionals that incorporate C_2. It is clear that such theories would be highly predictive, because it is well established that the presence of clusters in textures can dramatically alter their macroscopic physical properties (1).

It should not go unnoticed that our work also has important implications for atomic and molecular systems where the analogous standard correlation functions that arise are the two-body correlation function g_2, three-body correlation function g_3, etc. (29–32) The three-body function g_3, for example, has been the focus of great attention, e.g., integral equations and approximations have been devised for g_3 (30–32). The spatial structure of disordered atomic or molecular systems may be regarded to be special cases of random textures. For example, the particles that comprise simple atomic systems are fully specified by their center-of-mass coordinates. These point distributions can be decorated in an infinite number of ways to produce random textures in the sense that we have defined them in this article. For example, one could circumscribe the points by spheres of size dictated by the physics of the problem (e.g., electron hopping distance). For such systems, our work suggests that the pair-connectedness function P_2 [i.e., the connectedness contribution to g_2 (1)] should contain far greater information than g_3 beyond that contained in g_2.

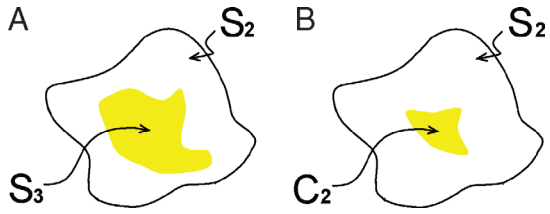

Fig. 5. The set of all microstructures associated with a particular S_2 is schematically shown as the region enclosed by the solid contour in *A* and *B*. The shaded region in *A* shows the set of all microstructures associated with the same S_2 and S_3. The shaded more restrictive region in *B* shows the set of all microstructures associated with the same S_2 and C_2.

ACKNOWLEDGMENTS. This work was supported by the American Chemical Society Petroleum Research Fund and the Office of Basic Energy Sciences, U.S. Department of Energy, under Grant DE-FG02-04-ER46108.

1. Torquato S (2002) *Random Heterogeneous Materials: Microstructure and Macroscopic Properties* (Springer, New York).
2. Sahimi M (2003) *Heterogeneous Materials* (Springer, New York).
3. Coker DA, Torquato S, Dunsmuir J (1996) Morphological and physical properties of Fountainebleau sandstone from tomographic analysis. *J Geophys Res* 101:17497–17510.
4. Pommerening A, Stoyan D (2008) Edge-correction needs in estimating indices of spatial forest structure. *Can J For Res* 38:1110–1122.
5. Peebles PJE (1993) *Principles of Physical Cosmology* (Princeton Univ Press, Princeton).
6. Gabrielli A, Sylos Labini F, Joyce M, Pietronero P (2005) *Statistical Physics for Cosmic Structures* (Springer, New York).

7. Kherlopian AR, et al. (2008) A review of imaging techniques for systems biology. *BMC Syst Biol* 2:1–18.
8. Torquato S (1997) Exact expression for the effective elastic tensor of disordered composites. *Phys Rev Lett* 79:681–684.
9. Zohdi TI (2006) On the optical thickness of disordered particulate media. *Mech Mater* 38:969–981.
10. Mejdoubi A, Brosseau C (2007) Numerical calculations of the intrinsic electrostatic resonances of artificial dielectric heterostructures. *J Appl Phys* 101:084109.
11. Torquato S, Stell G (1983) Microstructure of two-phase random media. II. The Mayer-Montroll and Kirkwood-Salsburg hierarchies. *J Chem Phys* 78:3262–3272.
12. Debye P, Bueche AM (1949) Scattering by an inhomogeneous solid. *J Appl Phys* 20:518–525.
13. Yeong CLY, Torquato S (1998) Reconstructing random media. *Phys Rev E* 57:495–506.
14. Yeong CLY, Torquato S (1998) Reconstructing random media: II. Three-dimensional media from two-dimensional cuts. *Phys Rev E* 58:224–233.
15. Sheehan N, Torquato S (2001) Generating microstructures with specified correlation functions. *J Appl Phys* 89:53–60.
16. Wu K, et al. (2006) 3D stochastic modelling of heterogeneous porous media - applications to reservoir rocks. *Trans Porous Media* 65:443–467.
17. Basanta D, Miodownik MA, Holm EA, Bentley PJ (2005) Investigating the evolvability of biologically inspired CA. *Metal Mater Trans A* 36:1643–1652.
18. Kumar H, Briant CL, Curtin WA (2006) Using microstructure reconstruction to model mechanical behavior in complex microstructures. *Mech Mater* 38:818–832.
19. Jiao Y, Stillinger FH, Torquato S (2007) Modeling heterogeneous materials via two-point correlation functions: Basic principles. *Phys Rev E* 76:031110.
20. Jiao Y, Stillinger FH, Torquato S (2008) Modeling heterogeneous materials via two-point correlation functions. II. Algorithmic details and applications. *Phys Rev E* 77:031135.
21. Torquato S, Beasley JD, Chiew YC (1988) Two-point cluster function for continuum percolation. *J Chem Phys* 88:6540–6546.
22. Lu B, Torquato S (1992) Lineal path function for random heterogeneous materials. *Phys Rev A* 45:922–929.
23. Torquato S, Lu B (1993) Chord-length distribution function for two-phase random media. *Phys Rev E* 47:2950–2953.
24. Tonks L (1936) The complete equation of state of one, two and three dimensional gases of hard elastic spheres. *Phys Rev* 50:955963.
25. Stauffer D, Aharony A (1994) *Introduction to Percolation Theory* (Taylor and Francis, London).
26. Torquato S, Yeong CLY, Rintoul MD, Milius DL, Aksay IA (1999) Characterizing the structure and mechanical properties of interpenetrating multiphase cermets. *J Am Ceram Soc* 82:1263–1268.
27. Askeland DR, Phule PP (2005) *The Science and Engineering of Materials* (Cengage, Florence, KY).
28. Garboczi EJ, Bentz DP (1998) Multi-scale analytical/numerical theory of the diffusivity of concrete. *J Adv Cement-Based Mater* 8:77–88.
29. Widom B (1968) Random Sequential Addition of Hard Spheres to a Volume. *J Chem Phys* 44:3888–3894.
30. Henderson D (1967) Structure of the Triplet Distribution Function. *J Chem Phys* 46:4306–4310.
31. Rice SA, Gray P (1966) *The Statistical Mechanics of Simple Liquids* (Wiley, New York).
32. Hansen JP, McDonald IR (2006) *Theory of Simple of Liquids* (Academic, New York).

Supporting Information

Jiao et al. 10.1073/pnas.0905919106

SI Text

Degeneracy of Ground States Associated with the Two-Point Correlation Function. In the main article, we have pointed out that a large number of studies (most of which are numerical) have established that S_2 is not sufficient information to generally get an accurate rendition of the original microstructure. Here, we provide a rigorous analysis that explains the reasons why the ground states of reconstructions of digitized random textures using S_2 alone are degenerate (i.e., nonunique).

A d-dimensional digitized two-phase random medium (texture) can be represented by a d-dimensional array $I_{i_1\cdots i_d}$, which is essentially the discrete one-point indicator function of the medium (see Eq. **1** in the main article). For example, in two dimensions for a texture with $N \times N$ pixels, we have

$$\mathbf{I} = \begin{bmatrix} I_{11} & I_{12} & \dots & I_{1N} \\ I_{21} & I_{22} & \dots & I_{2N} \\ \vdots & \ddots & & \vdots \\ I_{N1} & I_{N2} & \dots & I_{NN} \end{bmatrix}, \qquad \text{[S1]}$$

where the entries I_{ij} $(i,j = 1,\dots,N)$ can only take the value of 0 or 1, which correspond to one of the two phases, respectively.

Mathematically, the reconstruction of the digitized texture from a prescribed set of correlation functions amounts to recovering the discrete one-point indicator function of the phase of interest from a set of algebraic equations involving $I_{i_1\cdots i_d}$ and the specified correlation functions. In particular, for the reconstruction using S_2 alone, the discrete indicator function $I_{i_1\cdots i_d}$ satisfies the following equations for every value of $r \le N/2$ such that r^2 is an integer:

$$\sum_{(j_1,\dots,j_d)\in\Omega} \left[\sum_{i_1=1}^{N} \cdots \sum_{i_d=1}^{N} I_{i_1\cdots i_d} I_{(i_1+j_1)\cdots(i_d+j_d)} \right] - \omega N^d S_2(r) = 0, \qquad \text{[S2]}$$

where

$$\Omega = \left\{ (j_1,\dots,j_d) | j_1^2 + \cdots + j_d^2 = r^2, r \le N/2 \right\}, \qquad \text{[S3]}$$

and ω is the number of elements in Ω, N is the linear size of the digitized texture. Since the number of equations $(N/2)$ is significantly smaller than the number of unknowns (N^d), i.e., all the entries $I_{i_1\cdots i_d}$, it is clear that Eq. S2 possesses a large number of solutions, all of which are the degenerate ground states for the "energy" defined by Eq. **5** in the main article.

Reconstructing One-Dimensional Random Textures Using S_3. In the main article, we have shown that S_3, the three-point version of S_2, generally does not contain appreciably greater information through the reconstruction of one-dimensional equilibrium hard-rod system from S_3. Here, we provide some technical details concerning this reconstruction.

It is generally highly nontrivial to design and implement efficient random texture reconstruction procedures that incorporate higher-order correlation functions such as S_3, which necessarily involves tracking the changed triangular configurations of pixels due to the trial move of a randomly selected pixel. However, in one dimension, the system size (i.e., total number of pixels) is not large, which makes the "brute force" method—resampling the entire system to obtained the updated S_3—applicable in practice. We note that in higher dimensions, direct resampling always turns out to be extremely time consuming.

In particular, the number of possible triangle configurations formed by triplets of pixels is significantly reduced in one dimension. The three edges with length r,s,t of a "crashed" one-dimensional triangle is linearly dependent, i.e., when r and s are specified, t can only take the value $|r+s|$ or $|r-s|$, depending on the position of the third pixel. Thus, to obtain S_3, one only needs to move specified triangle configurations (completely determined by r and s) through the whole system and compute the fraction of times when all the vertex pixels are occupied by the phase of interest.

The aforementioned procedure is repeated for each trial configuration. The three-point function S_3 obtained is then used to compute the "energy," which determines whether or not the trial configuration is accepted.

Cluster and Surface Algorithms. In the main article, we employed the Yeong–Torquato reconstruction procedure, which requires generating and sampling a large number of realizations of the textures because the system must evolve from an initial realization guess to the realization that matches the targeted correlation functions. Here, we provide the technical details of the algorithms to handle clusters and surfaces in the texture in order to efficiently recompute the desired correlation functions of the new realization based on the old ones. This makes the incorporation of those functions computationally feasible in a manageable time. Direct resampling is computationally too expensive to implement in practice.

A distance matrix **D** that stores the separation distances of all "molecule" pairs is established when the system is initialized and the molecules are partitioned into different "species" depending on their positions, as discussed in the main article. The quantities N_P, N_P^i, N^{ss} and N^{sv} (all defined in the main article) can be obtained by binning the separation distances of selected pairs of molecules from particular species. Recall that the molecules are pixels associated with different values indicating the species, i.e., the clusters or the surface/volume set they belong to.

In the reconstruction procedure, a trial realization is generated by moving a randomly selected pixel of the phase of interest to an unoccupied site (1). This results in changes of the separation distances between the moved the pixel and all of the other pixels, and causes two kinds of possible species events. The first kind is a "cluster" event, which involves breaking and combining clusters. For example, if the selected pixel happens to be a "bridge" connecting several sub-clusters, removing the bridge will make the original single cluster break into smaller pieces, i.e., new species (clusters) are generated. Similarly, the reverse of the above process can occur, i.e., a randomly selected pixel moved to a position where it connects several small clusters to form a larger cluster, which leads to combination of clusters and annihilation of species. The other kind of species event is the transition of pixels between surface and volume sets. If a pixel originally on the surface is removed, certain volume pixels (i.e., those inside the phase of interest) will now constitute the new surface and vice versa. The selected pixel itself could also undergo such a transition, depending on its original and new positions, e.g., a volume pixel originally inside the phase of interest could be moved to the interface to become a surface pixel.

The contributions of the number of pair distances to N_P, N_P^i, N^{ss} and N^{sv} from the pixels in old realization involved in the species events are computed and subtracted accordingly. The new contributions can be obtained by binning the separation distances of pixel pairs in the new realization involved in the species event,

which are then added to the corresponding quantities N_P, N_P^i, N^{ss} and N^{sv}. This method only requires operations on a small number of pixels, including retrieving and binning their separation distances and updating the species sets (i.e., the clusters and surface/volume set). The use of the distance matrix **D** speeds up the operations involving distances. However, for very large systems (e.g., those including millions of pixels), storing **D** requires very a large amount of computer memory. An alternative is to re-compute the separation distances of the pixel pairs involved in the species events for every trial realization, instead of explicitly storing all the distances in **D**. This may slightly slow down the reconstruction process but make it easy to handle very large systems. Correlation functions of the new realization can then be obtained from the updated N_P, N_P^i, N^{ss} and N^{sv} (i.e., dividing those quantities by N_S), as discussed in the main article. Importantly, the complexity of the algorithm is linear in the total number of pixels (molecules) within the system.

Correlation Functions of the Example Textures. In the main article, we provided the images of the target and reconstructed textures, i.e., the concrete microstructure, the galaxy clusters and hard sphere packing. Moreover, we showed that the correlation functions of the reconstructions match the target ones very well by giving the final "errors," which are extremely small positive numbers. Here, we provide plots of the various target and reconstructed two-point correlation functions, which show that they match each other almost identically. The significance of the shape of the correlation functions is also briefly discussed.

Concrete Microstructure. Fig. S1 shows S_2, F_{ss} and C_2 of the target and reconstructed concrete microstructures. The two-point function S_2 of the stone phase does not properly account for the size distribution of the stones and thus leads to clustering of the stone phase in the S_2-alone reconstruction, as discussed in the main article. Incorporation of other nontrivial two-point information results in better renditions, but inclusion of C_2 provides the best reconstruction.

Galaxy Clusters. Fig. S2 shows S_2, F_{ss} and C_2 of the target and reconstructed galaxy clusters. Though the volume fraction of the "galaxy" phase is small, S_2 has a relatively long tail, which indicates the existence of large compact clusters in the system. This is also consistent with the rapid decay of C_2.

Sphere Packing. Fig. S3 shows S_2, F and C_2 of the target and reconstructed hard-sphere packings. The oscillation of S_2 is a manifestation of the short-range order in the system due to impenetrable nature of the spheres. C_2 has a clear cutoff beyond the diameter of the sphere, which indicates that there are no physically connected clusters of spheres. The pore size function F of the disconnected sphere phase is also short-ranged.

1. Jiao Y, Stillinger FH, Torquato S (2007) Modeling heterrogeneous materials via two-point correlation functions: basic principles. *Phys Rev E* 76:031110.

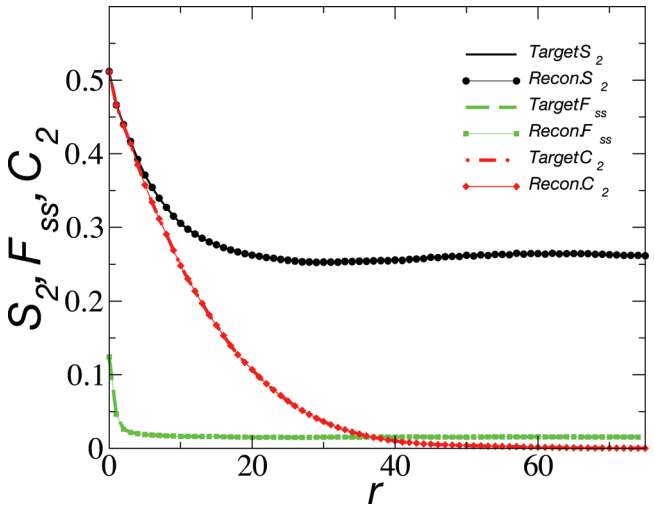

Fig. S1. S_2, F_{ss} and C_2 of the target and reconstructed concrete microstructures. It is clear that the target and reconstructed correlation functions match each other almost identically. As indicated in the main paper, the final error is $E \approx 10^{-8}$ for all cases. Pixel size supplies the unit for the distance r.

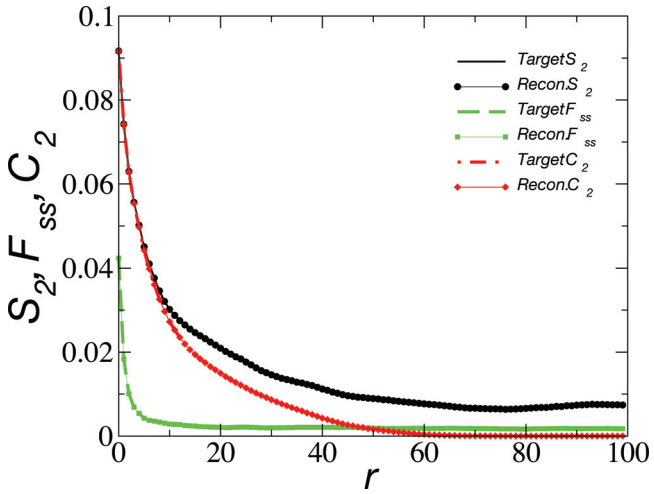

Fig. S2. S_2, F_{ss} and C_2 of the target and reconstructed galaxy clusters. It is clear that the target and reconstructed correlation functions match each other almost identically. As indicated in the main paper, the final error is $E \approx 10^{-8}$ for all cases. Pixel size supplies the unit for the distance r.

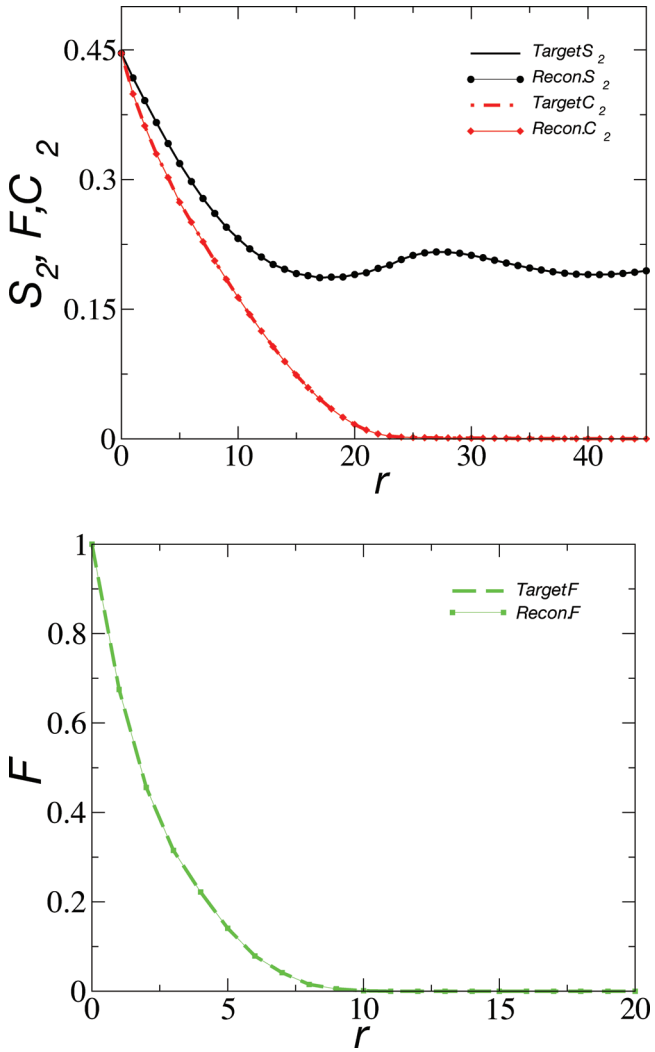

Fig. S3. S_2, C_2 (*Upper*) and F (*Lower*) of the target and reconstructed hard-sphere packing. The diameter of the digitized sphere is 24 pixels. It is clear that the target and reconstructed correlation functions match each other almost identically. As indicated in the main paper, the final error is $E \approx 10^{-11}$ for all cases. Pixel size supplies the unit for the distance r.