

Aspects of correlation function realizability

Jeness Crawford, Salvatore Torquato,^{a)} and Frank H. Stillinger
*Department of Chemistry and Princeton Materials Institute, Princeton University,
 Princeton, New Jersey 08544*

(Received 6 June 2003; accepted 16 July 2003)

The pair-correlation function $g_2(r)$ describes short-range order in many-particle systems. It must obey two necessary conditions: (i) non-negativity for all distances r , and (ii) non-negativity of its associated structure factor $S(k)$ for all k . For the elementary unit step-function g_2 form, previous work [F. H. Stillinger, S. Torquato, J. M. Eroles, and T. M. Truskett, *J. Phys. Chem. B* **105**, 6592 (2001)] indicates that (i) and (ii) could be formally satisfied, but only up to a terminal density at which the covering fraction of particle exclusion diameters equaled 2^{-d} in d dimensions. To test whether the unit step g_2 is actually achievable in many-particle systems up to the apparent terminal density, a stochastic optimization procedure has been used to shift particles in large test systems toward this target g_2 . Numerical calculations for $d=1$ and 2 confirm that the step function g_2 is indeed realizable up to the terminal density, but with substantial deviation from the configurational preferences of equilibrium hard-rod and hard-disk models. We show that lineal statistical measures are particularly sensitive to this difference. Our results also illustrate the characteristics of “closest approach” to the step function g_2 above the terminal density. © 2003 American Institute of Physics. [DOI: 10.1063/1.1606678]

I. INTRODUCTION

Conventional applications of the principles of statistical mechanics (the “forward” problems), start with particle interaction potentials, and proceed to deduce local structure and macroscopic properties.^{1,2} Other applications (that may be classified as “inverse” problems), begin with low-order correlation functions that characterize local particle order, and attempt to back out full-system configurations^{3–6} and/or interaction potentials.^{7–9} The work reported below constitutes an example of the inverse problem. In particular, we consider the realizability of a simple pair-correlation function form in terms of actual many-particle-system configurations. This inverse problem is approached against a background of prior theory.^{8–10}

A reconstruction algorithm based on stochastic optimization techniques has been developed to generate realizations (configurations) of particle systems or heterogeneous materials that possess a targeted set of lower-order correlation functions.^{3,4} Of course, a finite set of lower-order correlation functions will not uniquely specify a disordered system in the thermodynamic limit, but the extent to which the actual realization can be reconstructed reveals the level of information embodied within the correlation functions. Reconstructions can also shed light on the mathematical properties that realizable correlation functions must possess. However, the same procedure has been used to “construct” configurations for a target correlation function in order to determine if such a correlation function is realizable.^{4,5} In this mode, the procedure is referred to as a *construction* algorithm.

The purpose of this paper is to utilize the construction technique to explore the realizability of a pair-correlation

function $g_2(r)$, namely, the unit step function $U(r-D)$, i.e.,

$$g_2(r) = U(r-D) = \begin{cases} 0, & 0 \leq r < D, \\ 1, & r \geq D, \end{cases} \quad (1)$$

for a statistically homogeneous and isotropic system of identical hard d -dimensional spheres of diameter D at number density ρ . This choice of the pair-correlation function was introduced by Stillinger, Torquato, Eroles, and Truskett⁸ to study so-called “iso- g_2 ” processes. In an “iso- g_2 ” process, the pair potential is changed so that the g_2 is invariant as density is changed. Our interest is in determining whether the step-function pair-correlation function, defined by Eq. (1), is realizable by hard-particle configurations for the apparent allowable covering fraction range $0 \leq \phi \leq \phi_c$, where $\phi_c = 2^{-d}$,⁸ ϕ is the covering fraction, and d is the space dimension. This means a terminal covering fraction of $\phi_c = 0.5$ in one dimension ($d=1$), $\phi_c = 0.25$ in two dimensions ($d=2$), and $\phi_c = 1/8$ in three dimensions ($d=3$). (In three dimensions, Markov and Willis¹¹ used a different approach to show that $\phi_c = 1/8$ is a singular limit of the step function g_2 .) By using the aforementioned construction method to generate configurations corresponding to a unit step function g_2 , this paper provides strong numerical evidence that the step function is indeed achievable in this range and only in this range. The precise determination of the pair potential that achieves this step function will be the subject of a future paper.

Aside from these density limits specific to the step function, there are general nonnegativity conditions that g_2 must obey for it to be realizable by a many-particle system. Since $g_2(r)$ is proportional to the probability of finding particles separated by the distance r , it cannot be negative, i.e.,

$$g_2(r) \geq 0. \quad (2)$$

^{a)}Corresponding author: torquato@princeton.edu

The structure factor $S(k)$ is related to the Fourier transform of the total correlation function $h(r) = g_2(r) - 1$ via the relation²

$$S(k) = 1 + \rho \int e^{-i\mathbf{k}\cdot\mathbf{r}} h(r) d\mathbf{r}, \quad (3)$$

where \mathbf{k} is the wave vector and $k \equiv |\mathbf{k}|$. The second fundamental constraint is the nonnegativity of $S(k)$, i.e.,

$$S(k) \geq 0, \quad (4)$$

which must be obeyed for all real values of k .⁶ The nonnegativity of the structure factor physically arises from the fact that $S(k)$ is proportional to the intensity of scattered radiation from the macroscopic sample. Both of these nonnegativity conditions are not restricted to states at thermal equilibrium, but are more general. It is known that these conditions are necessary,¹⁰ and this work shows that they are sufficient in the case of the step function.

It is useful to note that for an equilibrium hard-sphere system characterized by the pair potential

$$u(r) = \begin{cases} +\infty, & 0 \leq r \leq D, \\ 0, & r > D, \end{cases}$$

the step-function pair-correlation function defined by Eq. (1) is exact in the infinitely dilute limit ($\rho \rightarrow 0$). However, for small positive values of ρ , a peak develops at $r = D$ for the equilibrium g_2 , which becomes more pronounced as the density is increased. The physical origin of this peak is a “shielding” effect; when a second particle is between one and two diameters from a central particle located at the origin, the second one suffers fewer collisions (on average) on the side facing the central one than on the opposite side, resulting in an effective attraction toward the central one. For typical liquid densities, the equilibrium g_2 also reveals short-range order, i.e., it displays finite-amplitude oscillations about unity that decay to zero with increasing r on a length scale roughly comparable to D . Therefore it will be meaningful and interesting to compare our results for the step-function pair-correlation function at positive feasible values of ρ to corresponding equilibrium results. It was shown in Ref. 8 that an effective potential that could generate the former must have a finite repulsion beyond the hard core, i.e., $r > D$. This repulsion that acts beyond $r = D$ serves to completely suppress the peak at $r = D$ and the accompanying short-range order that would otherwise characterize the pure hard-sphere model.

We note in passing that it has been shown that the system generated by the step function is a *hyperuniform* system at the terminal density.¹² Hyperuniformity is concerned with a certain type of behavior of local-density fluctuations. Consider density fluctuations within a regular domain, or “window,” Ω centered at point \mathbf{x}_0 . For a very large class of particle systems, the number variance within the window, $\langle N_\Omega^2 \rangle - \langle N_\Omega \rangle^2$, grows asymptotically as the window volume. A hyperuniform system is one in which the number variance grows only as the surface area of the window in three dimensions, or as its perimeter in two dimensions. In other words, infinite-wavelength density fluctuations vanish.

In Sec. II, we describe the construction procedure used to generate configurations of the step function g_2 , and some caveats for this procedure when applied to a finite system. In Sec. III, we discuss the results found in the one- and two-dimensional cases. Here we also discuss the measurement of other statistical descriptors of the generated configurations. Concluding remarks and discussion of future work is described in Sec. IV.

II. CONSTRUCTION ALGORITHM

We seek to determine whether the step function $g_2(r)$ defined by Eq. (1) is realizable by configurations of spheres at number density ρ with hard cores of diameter D for particle covering fractions in the range $0 \leq \phi \leq \phi_c$, where $\phi_c = 2^{-d}$. The covering fraction of particle exclusion diameters ϕ is defined by⁶

$$\phi = \rho v_1(D/2), \quad (5)$$

where $v_1(r)$ is the d -dimensional volume of a single sphere of radius r given by

$$v_1(r) = \frac{\pi^{d/2}}{\Gamma(1+d/2)} r^d, \quad (6)$$

and $\Gamma(x)$ is the gamma function. In the specific cases of one and two dimensions, our focus in this paper, relation (5) gives

$$\phi = \rho D \quad (d=1), \quad (7)$$

$$\phi = \frac{\rho \pi D^2}{4} \quad (d=2). \quad (8)$$

In one dimension and two dimensions, we will refer to these hard particles as hard *rods* and *disks*, respectively. In all of our simulations, the particles are contained within a d -dimensional cube with sides of length L , so that $\rho = N/L^d$, and periodic boundary conditions are employed.

Following Rintoul and Torquato³ and Yeong and Torquato,⁴ we use the method of simulated annealing^{13,14} to construct configurations of particles from a set of target correlation functions. Our interest here is in a target pair-correlation function $g_2(r)$, which in the case of statistically homogeneous and isotropic systems is referred to as the radial distribution function (RDF). None of our results will be ensemble averaged. Instead we will place the stringent demand that a single configuration for a particular density matches the step-function RDF.

Starting from some initial configuration (e.g., a regular array or a random configuration), the construction technique finds the realization in which the calculated RDF best matches the target RDF defined by the step function (1). This is accomplished by introducing a fictitious energy

$$E = \sum_{r_i} [g_2(r_i) - g_2^0(r_i)]^2, \quad (9)$$

where $g_2^0(r_i)$ is the target step-function RDF, defined by Eq. (1), $g_2(r_i)$ is the RDF at any time step in the simulation, and the sum is over all distances up to a specified limit, which we call the *sampling distance*. The energy is calculated for the initial configuration \mathbf{r}^N of the N particles, then a new con-

figuration is generated by the following rules. A particle is moved by displacing it along each axis by amounts randomly and uniformly distributed in the interval $[-\delta, \delta]$, where δ is the maximum step size. The energy is computed for this new configuration as above, and the move is accepted or rejected with probability $P(\Delta E)$ according to a Metropolis acceptance rule:

$$P(\Delta E) = \begin{cases} 1, & \Delta E \leq 0, \\ \exp(-\Delta E/T), & \Delta E > 0, \end{cases} \quad (10)$$

where T is a fictitious temperature.⁶ The acceptance of uphill moves with a Boltzmann factor probability allows the configuration to escape from local minima. Each particle is moved sequentially in the same manner until the system is equilibrated at a particular value of the temperature. The reasoning behind the method of simulated annealing is that if a system is heated to a high temperature T , then slowly cooled to absolute zero, it will equilibrate in its ground state. In this ground state, the energy can be viewed as the least-squares error between the generated configuration's $g_2^0(r)$ and the target function $g_2(r)$.⁶ It therefore becomes crucial to choose the fastest possible cooling schedule that will allow the system to reach its global minimum without getting stuck in local minima. We found that a maximum step size of $\delta = 0.5D$ worked effectively in this study.

At any particular step of the construction procedure, the RDF is computed from a histogram of the average number of particle centers $n(r)$ contained in a concentric shell of finite thickness (bin width) Δr at radial distance r from a reference particle center.⁶ The radial distance r is taken to be halfway between the inner radius $(r - \Delta r/2)$ and the outer radius $(r + \Delta r/2)$ of each shell. Let $n_k(r)$ be the number of pairs of particles in bin k corresponding to a radial distance r . Then

$$n(r) = \frac{n_k(r)}{N} \quad (11)$$

and the RDF is calculated from the formula

$$g_2(r) = \frac{n(r)}{\rho v_{\text{shell}}(r)}, \quad (12)$$

where v_{shell} is the volume of the d -dimensional shell given by

$$v_{\text{shell}}(r) = v_1(r) \left[\frac{(r + \Delta r/2)^d - (r - \Delta r/2)^d}{r^d} \right],$$

and $v_1(r)$ is given by Eq. (6). For one and two dimensions, we specifically have

$$g_2(r) = \frac{n_k(r)}{2N^2\Delta r} \quad (d=1), \quad (13)$$

$$g_2(r) = \frac{n_k(r)}{2\pi N^2 r \Delta r} \quad (d=2). \quad (14)$$

Because of the finite system size, the resolution of the pair-correlation function becomes a problem. We calculate $g_2(r)$ via relations (13) and (14). By necessity, $n_k(r)$ must be an even integer (the distance between each pair of particles is counted twice, once for each of the particles involved). Consequently, $g_2(r)$ can only achieve certain val-

ues, dependent on the value of the denominator. For $d=1$, the denominator of Eq. (13) has a fixed value, independent of r , and only dependent on the bin width, the system size, and the covering fraction. Substituting in relation (7), and setting $D=1$ (so that r is in terms of diameter),

$$\Delta r \phi N = \text{integer}, \quad (15)$$

for $g_2(r)$ to have an integer value. Since we are trying to match the step function, the program will be unable to do so unless $g_2(r)$ can equal unity.

For $d=2$, ensuring that $g_2(r)$ can equal unity, given the system parameters, is considerably more difficult since the denominator of relation (14) depends on r , and is therefore different for each bin. For $g_2(r)$ to be an integer, the bin width must be adjusted for each bin. If the bin width is kept constant, the program will do its best to match the step function, resulting in a function that jumps above unity, then gradually decreases as r increases, then jumps above unity again, in order to attain a minimum distance from the target function (see the first figure described in Sec. III B). Taking the data generated from a run of constant bin size, we can use the r and $g_2(r)$ values for each bin to adjust Δr and r (r is dependent on Δr since it must be in the middle of the bin) so that $g_2(r)=1$. Using these adjusted values for Δr and r , another simulation is performed, this one generating the desired flat pair-correlation function. Within these constraints for $d=1$ and $d=2$, the bin widths must still be chosen so that the final $g_2(r)$ is an accurate reflection of the system structure. If the widths of the bins are too small, the fluctuations overwhelm the RDF, if the bins are too wide, there is not enough information to accurately reproduce the target function.³

The construction procedure was used to examine the step function up to the terminal density of $\phi=0.5$ for $d=1$ and $\phi=0.25$ for $d=2$. At low densities, starting from a regular array (a square lattice for $d=2$ and evenly spaced rods for $d=1$) was found to get the system stuck in local minima, and to take much longer to equilibrate at a particular temperature T . Starting from an equilibrium hard rod or hard disk configuration alleviated this problem. Consequently, at all densities, we started the annealing program from equilibrium configurations for all results reported below.

Despite previous results to the contrary for a different correlation function,⁵ a "great deluge" algorithm, where only downhill moves are accepted (corresponding to a fictitious temperature $T=0$), was ineffective. In all cases, it led to immediate trapping in a local minimum close to the initial configuration. The most effective cooling schedule was found to be an incrementally decreasing function of temperature. At high initial temperatures, few Monte Carlo (MC) cycles are needed, as the purpose of heating to these temperatures is simply to shake the system out of its initial configuration. At each temperature, the energy of the system rapidly takes on a Gaussian-like distribution (see Fig. 1). The system reaches this equilibrium energy distribution after a finite number of MC cycles. As the temperature is lowered, the mean energy is also lowered, and the distribution narrows. The key to a successful cooling schedule appears to be approaching $T=0$ in temperature steps that are small

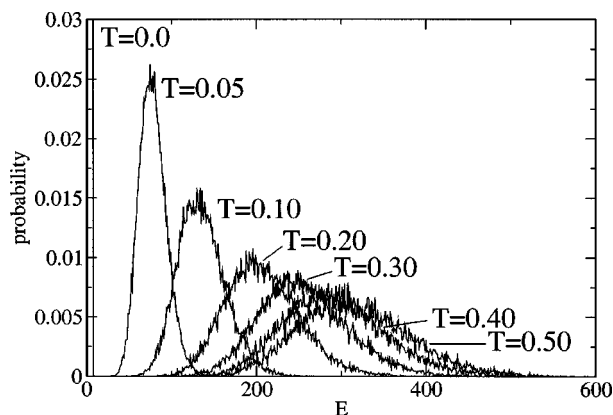


FIG. 1. Energy histograms for one simulation done in one dimension at $\phi = 0.45$. The x axis is the energy in terms of number of misbinned pairs. The y axis is the probability at each temperature of finding the configuration at each energy. The curves for $T=0.0, 0.05, 0.10, 0.20, 0.30, 0.40, 0.50$ are shown. It is clear that as the temperature is lowered, the mean energy is lowered, and the distribution becomes narrower. At $T=0.0$, each configuration has a tendency to be trapped at a single energy, so the curve shown is an average over 16 configurations. The mean of this curve is at $E=4$, with a probability of 0.21.

enough. If the increments between temperatures are too large, the system will tend to get stuck in local minima. At the end of the cooling schedule, the system is run through several thousand MC cycles with $T=0$, to ensure that it is at an absolute energy minimum.

It is noteworthy that for $d=1$, the energy of the system can only take certain incremental values, corresponding to the number of pairs that are the “wrong” distance apart (misbinned pairs), i.e., pair distances that do not give a value of unity for the RDF. As the system parameters Δr , N , and ϕ are changed, the size of these increments changes. Since the uphill moves are accepted with the Boltzmann probability $\exp(-\Delta E/T)$, the effective temperature changes with Δr , N , and ϕ . For example, in a one-dimensional system with $\Delta r = 0.1$ and $\phi = 0.2$ and $N = 200$, one misbinned pair of centers is equivalent to $E = 0.0625$. For a one-dimensional system at $\phi = 0.45$ with the same bin widths and the same number of particles, the energy of one misbinned pair is $E = 0.012\,345\,68$. Because of this, the temperature T used to evaluate the Boltzmann probability is effectively six times hotter for the $\phi = 0.45$ system than the $\phi = 0.2$ system. The temperatures in the cooling schedule need to be selected as the system parameters are changed to preserve the same effective temperatures.

Two-dimensional systems do not encounter this problem. The energy for $d=2$ is dependent on r as well as the parameters listed above. Therefore the energy value for a misbinned pair is different for each bin. Moreover, energy histograms for $\phi = 0.1$ and 0.2 are virtually identical for the same temperature, indicating that the two systems, although at different densities, experience the same effective temperature.

Computational time increases as the square of system size because to calculate $g_2(r)$ for every MC cycle, the program must measure the distance between each particle and the other $(N-1)$ for each of the N particles, leading to

$N(N-1)$ calculations per MC cycle. Only smaller system sizes were computationally feasible. Early trials involved $N = 500$ and 1000 , but all later simulations had smaller sizes ($N = 200$ for $d=1$ and $N = 289$ for $d=2$).

The RDF for an equilibrium hard-sphere fluid system equals unity at large r because the spheres are uncorrelated at large distances. This being the case, we do not need to match the simulated correlation function to the target function at large r while annealing, since we can assume them to have the same value. It is only the short-range order of the equilibrium system that the simulated-annealing program must suppress to make the simulated system’s RDF match the target step function. But how far out should we try to match the systems? Obviously, the farther out the systems are matched, the more accurate the configurational results will be, but computational time increases with the square of the sample distance, making it prudent to limit the sampling distance to only what is absolutely necessary to closely approach the target function.

Early trials in one dimension all used a sampling distance of five diameters ($r=5$). With promising results from these, the sampling distance was increased to $r=10$, leading to great difficulties. For an equilibrium hard rod initial configuration, the mean particle separation is given by $\ell_C + 1$, where ℓ_C is the mean chord length (gap size) given by Eq. (21). Therefore, for densities less than $\phi = 0.4$, only the nearest neighbor is included (on average) in the sampling distance. Given the numerical protocol, the program then works to flatten the pair-correlation function by adjusting only the nearest-neighbor distances. As soon as two particles draw closer together to flatten out the function, they pull away from the particles on either side of them, and these distances are not counted. This leads to a configuration in which particles have paired. It is therefore a simple matter of optimizing each of the distances between the two particles in a pair, with no regard to their relative positions to any other particles, and the program easily matches the target step function. Increasing the sampling distance to 10 or 20 diameters suddenly makes the optimization problem much more difficult. Moving each particle becomes a tug of war between its distances to nearest neighbors and to neighbors farther away, thus creating energy barriers for lowering the energy of the system, i.e., the energy surface becomes more rugged, and the system becomes more easily trapped in local minima. It is costly for the system to move away from its initial configuration, so the simulated annealing program will have difficulty eliminating large differences between the initial RDF and the step function when the sampling distance includes several nearest neighbors. At low sampling distances, the pairing phenomenon leads to meaningless results. It is therefore important to choose a sampling distance at each density that includes several nearest neighbors, as indicated by the mean particle separation.

This problem of sampling distance was not as severe in two dimensions. The reason is that a sampling distance that includes the nearest-neighbor shell of a given particle will include more than two of the particle’s near neighbors, which prevents the simple pairing phenomenon found in one dimension.

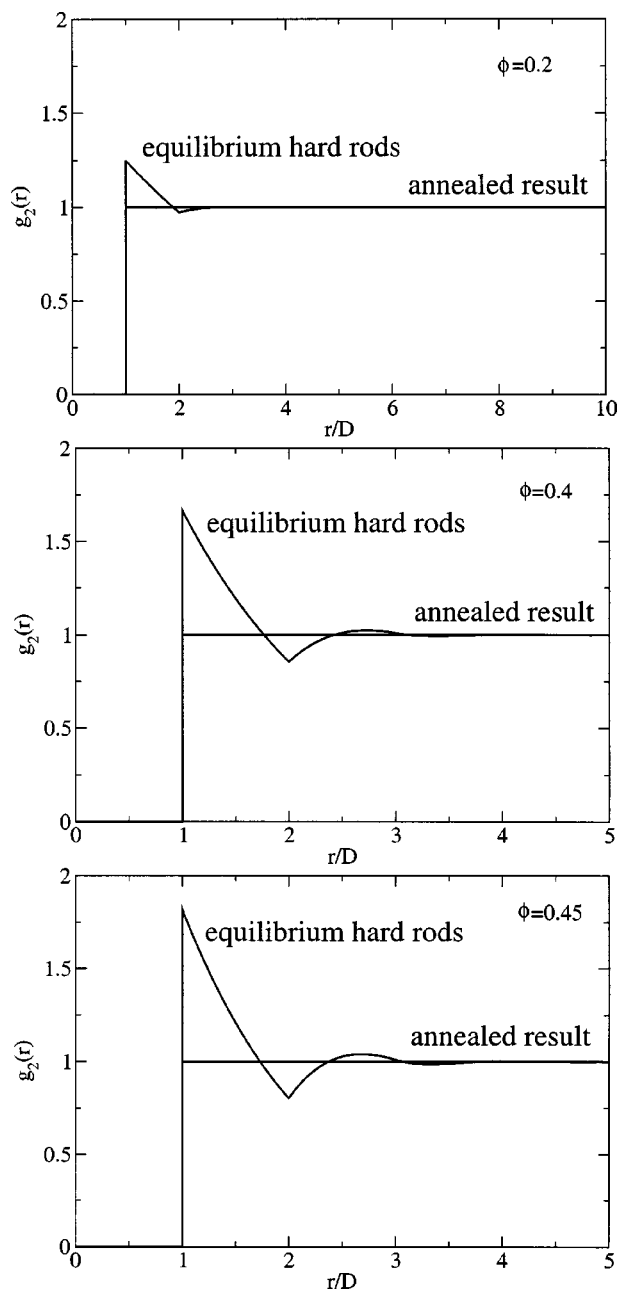


FIG. 2. RDF's in one dimension for densities of $\phi=0.2$ (top), $\phi=0.4$ (middle), and $\phi=0.45$ (bottom). The annealing result for a single configuration is compared to the equilibrium hard rod RDF in each case. The value for each bin is plotted in the center of the bin for this and all following equivalent graphs.

Finally, we remark that the most effective way to lower the energy of systems in one and two dimensions turned out to be repetition of the algorithm. Even when an appropriate cooling schedule had been designed, many initial configurations had to be processed before the global minimum was found. Adjusting the cooling schedule can increase the probability of perfectly matching the unit-step function, but it by no means makes this the certain outcome of every simulation.

III. RESULTS

Our results support the conclusion that the unit step function $g_2(r)$ is realizable up to the terminal density ϕ_c

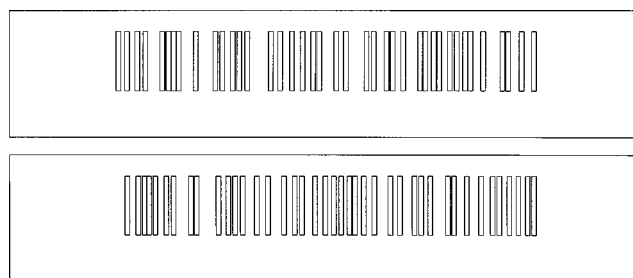


FIG. 3. The equilibrium configuration (top) for $\phi=0.45$ is compared to the annealed step function system for the configuration whose RDF is shown in Fig. 1 (bottom). The equilibrium system shows larger gaps and more clustering than the step-function g_2 system, as expected. The rods have been stretched vertically to enhance ease of pattern visualization.

$=2^{-d}$ in both one and two dimensions. Moreover, satisfaction of the nonnegativity conditions $g_2(r) \geq 0$ and $S(k) \geq 0$ is sufficient in one and two dimensions to ensure that the step function g_2 is realizable by hard-sphere configurations. Nothing suggests that these conclusions would not hold in higher dimensions. It is important to emphasize that all of our annealed results for the step function g_2 reported below are for a single configuration at a particular density, i.e., they are not ensemble averaged.

A. One dimension

For $d=1$, simulations were run at $\phi=0.20, 0.30, 0.40, 0.45, 0.50$, i.e., up to and including the terminal density. A perfect result ($E=0$) was achieved for $\phi=0.2$ up to a sampling distance of 10, and $\phi=0.30, 0.40, 0.45$ to a sampling distance of 5. Figure 2 shows several of these annealed cases and compares them to the corresponding equilibrium g_2 .

As can be seen in Fig. 3, the particles of the equilibrium systems have more of a tendency to cluster than the annealed systems. For simplicity, we take all distances r to be in terms of diameters, so that the contact distance occurs at $r=1$. In order to suppress the short-range order of the system at equilibrium, and eliminate the initial peak at the contact distance, the rods in the step-function g_2 system must have a repulsive interaction potential, forcing them to spread out more. The difference between the equilibrium and the annealed system is most acute at higher densities, as this is where the equilibrium hard rod RDF differs most from the unit step function. At lower densities, the difference is less acute, and indeed in the limit $\rho \rightarrow 0$, the equilibrium hard rod RDF is the step function.⁶

We also achieved a perfect step function at the terminal density $\phi = \frac{1}{2}$ for a sampling distance of 2.5. For a sampling distance of 5 at the terminal density, the annealed result was nearly perfect; there were only a few misbinned pairs near $r=D$, resulting in a small peak at contact. We believe that this imperfection is a finite-size effect, which would vanish in the infinite system limit. It is important to observe that the value of $\frac{1}{2}$ for the terminal density applies in the thermodynamic limit.⁸

Contrary to the expectation that configurations would become more difficult to anneal to the step function g_2 as the terminal density was approached, we found that annealing "equivalent" systems at all densities was comparable. In or-

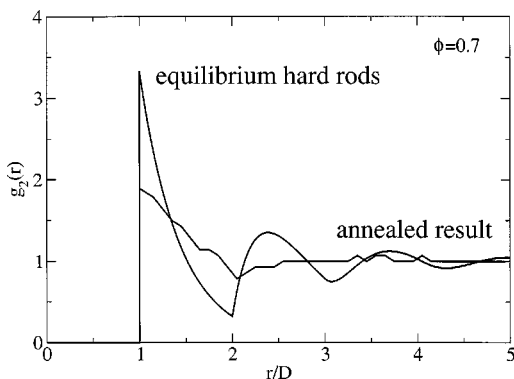


FIG. 4. RDF for $\phi=0.7$ in one dimension. The RDF for one annealed configuration is compared to the analytical equilibrium hard rod RDF. Although it is impossible to match the step function at this density, the simulated annealing program outputs a function significantly different from the equilibrium hard rod RDF at this density.

der for systems to be equivalent at different densities, one must hold constant the number of particles, use the same effective temperatures in the cooling schedule, while adjusting the sampling distance to reflect the difference in mean pair separation.

Above the terminal density, the simulated annealing program struggled to produce a physically unrealizable configuration, with some interesting results. At densities of $\phi=0.55, 0.6$, the RDF generated by the program shows a buildup to a low peak at $r=1$, then a flat curve beyond that, markedly different from the damped oscillations of the equilibrium hard rod system. At $\phi=0.7, 0.8$, the peak at $r=1$ is followed by a dip below 1 (see Fig. 4). At a density of $\phi=0.9$, with very little room to maneuver, the program produces a slightly perturbed regular array.

B. Two dimensions

For $d=2$, simulations initially were run at $\phi=0.10, 0.15$, and 0.20 . Perfect results were achieved at $\phi=0.1$ and 0.15 at a sampling distance of 5, after adjustments were made to the binwidths (see Fig. 5). At $\phi=0.20$ and a sampling distance of 3, we also achieved a perfect result (see Fig. 6). Not surprisingly, Fig. 7 shows that an equilibrium hard-disk configuration exhibits more clustering and larger pores than a configuration for an annealed step-function system. We note, however, $\phi=0.20$ and a sampling distance of 5, there were small deviations from the step function RDF, resulting in an energy of $E=0.0068374$, calculated using relation (9). As in the one-dimensional case, we believe that this imperfection is a finite-size effect, which would vanish

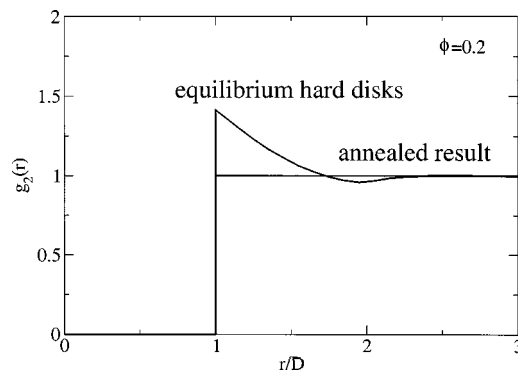


FIG. 6. RDF for $\phi=0.2$ in two dimensions. The annealed step function for one configuration is contrasted with the equilibrium hard disk RDF.

in the infinite system limit, and expect that step function is achievable up to the terminal density $\phi=0.25$.

At densities above the terminal density for step function realizability, $\phi=0.25$, the program produced results similar to those in one dimension. At a density of $\phi=0.30$, there was a peak at $r=1$ that decreased to a flat line with no further oscillations. At $\phi=0.50$, the initial peak is followed by a trough (see Fig. 8). In contrast to one dimension, in two dimensions, there is a disorder-order phase transition at $\phi=0.69$.^{18,19} The highest density achievable in two dimensions for an ordered system $\phi=\pi/\sqrt{12}\cong 0.907$ is for the close-packed triangular lattice.⁶ So the simulations run at $\phi=0.50$ are below the freezing density, and still within the disordered phase. Even with these difficulties, the annealed systems show much less clustering of disks than the equilibrium hard disk system, as reflected by the relative heights of the peaks at the contact distance.

C. Lineal statistical descriptors

1. Theory

There is a variety of other statistical descriptors available to characterize the structure of many-particle systems; see, for example, the book by Torquato.⁶ Here we utilize lineal descriptors, such as the lineal-path function $L(z)$ and related chord-length probability density function $p(z)$ to further characterize the systems. For statistically isotropic media, $L(z)$ is the probability that a line segment of length z lies wholly in a single phase, in this case, the pore space, when randomly thrown into the sample. The pore space is the space exterior to that occupied by the particles. The lineal-path function $L(z)$ for any homogeneous and isotropic system of interacting identical d -dimensional spheres involves

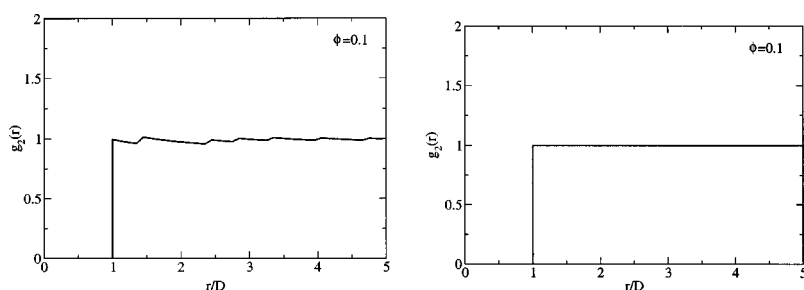


FIG. 5. Output RDF's for the simulated annealing program trying to match the step function in two dimension at $\phi=0.1$. Left panel: Output RDF for one configuration with constant bin width $\Delta r=0.1$, showing the inability of the program to match the step function [see the paragraph immediately below Eq. (15)]. Right panel: The output for one configuration once the bin widths have been adjusted to allow the program to perfectly match the step function.

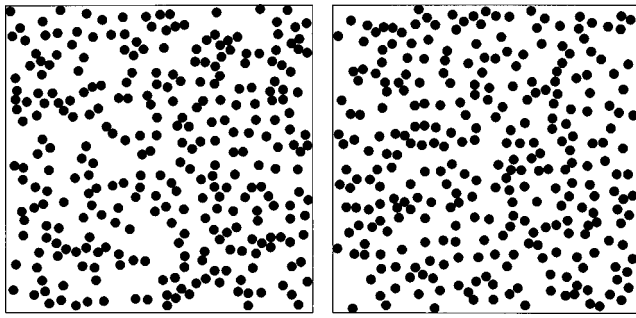


FIG. 7. Configurations of 289 particles for $\phi=0.2$ in two dimensions. The equilibrium hard disk system (left) shows more clustering and larger pores than the annealed step function system (right).

integrals over g_2 and all the higher-order correlation functions $g_3(r), g_4(r), \dots$ and is exactly given by^{6,15}

$$L(z) = 1 + \sum_{k=1}^{\infty} \frac{(-1)^k}{k!} \rho^k \int g_k(\mathbf{r}^k) \times \prod_{j=1}^k m(\mathbf{x} - \mathbf{r}_j; z, R) d\mathbf{r}_j, \quad (16)$$

where the exclusion region indicator function m is given by⁶

$$m(\mathbf{x}; z, R) = \begin{cases} 1, & \mathbf{x} \in \Omega_E(z, R), \\ 0, & \text{otherwise,} \end{cases} \quad (17)$$

and Ω_E is the volume excluded to a spherical inclusion of radius R by a line of length z . For all $z, L(z) \geq 0$. In addition, since $L(z)$ is a probability function, it is a monotonically decreasing function of z .

The lineal-path function is related to the chord-length probability density function $p(z)$.^{6,16} This has also been called the chord-length distribution function. For general two-phase random media, ‘‘chords’’ are the line segments between the intersections of an infinitely long line with the two-phase interface. The quantity $p(z)dz$ is the probability of finding a ‘‘chord’’ of length between z and $z + dz$ in the pore space.¹⁶ Since it is a probability density function, $p(z) \geq 0$ for all z and we have the normalization

$$\int_0^{\infty} p(z) dz = 1. \quad (18)$$

The chord-length density function is related to the lineal-path function by¹⁶

$$p(z) = \frac{\ell_c}{(1 - \phi)} \frac{d^2 L}{dz^2}, \quad (19)$$

where ℓ_c is the mean chord length given generally by

$$\ell_c = \int_0^{\infty} z p(z) dz. \quad (20)$$

In the special case of hard-sphere systems in any ensemble, the mean chord length (or, equivalently, the mean gap size between the particles) is exactly given by⁶

$$\ell_c = \frac{(1 - \phi) \omega_d}{2 \phi \omega_{d-1}} D, \quad (21)$$

where ω_d is the volume of a single sphere in d dimensions of unit radius given by Eq. (6) with $r = 1$.

One can show that if g_2 is the unit step function, then $L(z)$ in one dimension is given by

$$L(z) = \begin{cases} 1 - \phi - \phi \left(\frac{z}{D}\right) + \frac{\phi^2}{2} \left(\frac{z}{D}\right)^2, & 0 \leq z \leq D, \\ f(z), & z \geq D, \end{cases} \quad (22)$$

where $f(z)$ is the functional form of $L(z)$ for $z \geq D$, which is uniquely determined by the infinite set of correlation functions g_2, g_3, g_4, \dots . Therefore combining Eqs. (19), (21), and (22) for the step function g_2 yields in one dimension

$$p(z) = \begin{cases} \frac{\phi}{D}, & 0 \leq z \leq D, \\ \frac{D}{\phi} \frac{d^2 f}{dz^2}, & z \geq D. \end{cases} \quad (23)$$

Due to the nonnegativity of $p(z)$, the second derivative $d^2 f/dz^2$ must also be non-negative.

It is interesting to compare the lineal path function Eq. (22) and chord-length density function $p(z)$ Eq. (23) for the step function g_2 to the corresponding functions for an equilibrium systems of hard spheres at the covering fraction ϕ , which are given by⁶

$$L(z) = (1 - \phi) \exp\left[-\frac{\phi}{(1 - \phi)} \frac{z}{D}\right], \quad z \geq 0, \quad (24)$$

$$p(z) = \frac{\phi}{(1 - \phi) D} \exp\left[-\frac{\phi}{(1 - \phi)} \frac{z}{D}\right], \quad z \geq 0. \quad (25)$$

In contrast to the step-function g_2 counterparts, the equilibrium $L(z)$ and $p(z)$ are purely exponential functions. Of course, expressions (22) and (23) must approach the equilibrium functions (24) and (25) in the limit $\phi \rightarrow 0$.

In addition to the nonnegativity of $L(z)$ and $p(z)$ and the monotonicity of $L(z)$, which implies that $f(z)$ is a non-

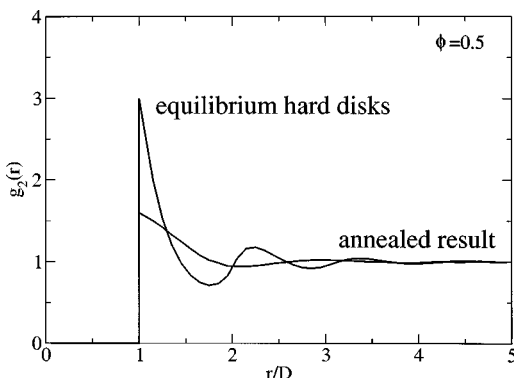


FIG. 8. RDF for $\phi=0.5$ in two dimensions. Like the $\phi=0.7$ case in one dimension, the simulated annealing program, in trying to achieve the physically unrealizable step function creates an RDF significantly flatter than the equilibrium. The annealed result RDF is the average of eight configurations.

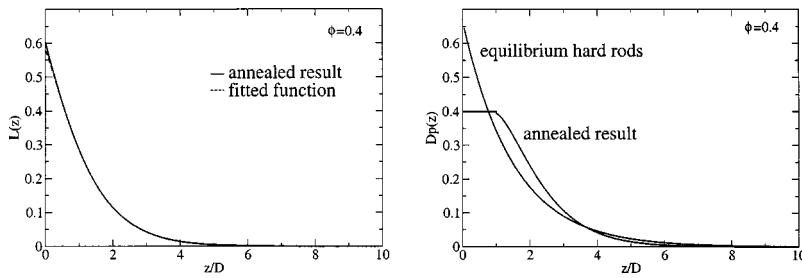


FIG. 9. Left panel: $L(z)$ determined exactly from the one-dimensional annealing results, averaged over 55 configurations, compared to best-fit curve for $z > D$, $f(z) = a \exp(-bz^c)$, $a = 0.57911$, $b = 0.71028$, $c = 1.1878$. Right panel: The annealing result is represented by $p(z) = \phi^{-1}[df^2(z)/dz^2]$ for $z > D$. For $z < D$, experimental results show that $p(z) = \phi$. The annealing result is compared to $p(z)$ for the equilibrium hard rod system.

negative, concave, and monotonic function of z , $f(z)$ must obey five exact conditions for the step function g_2 . The lineal-path function and its first derivative must be continuous at $z = D$ and so

$$f(D) = 1 - 2\phi + \frac{\phi^2}{2}, \quad (26)$$

$$D \left. \frac{df}{dz} \right|_{z=D} = -\phi(1 - \phi). \quad (27)$$

The third condition follows from the continuity of $p(z)$ at $z = D$ [cf. Eq. (23)]:

$$D^2 \left. \frac{d^2f}{dz^2} \right|_{z=D} = \phi^2. \quad (28)$$

The integral conditions (18) and (20) in conjunction with Eqs. (21) and (23) yields the fourth and fifth conditions on $f(z)$ as

$$\phi + \frac{D}{\phi} \int_D^\infty \frac{d^2f}{dz^2} dz = 1, \quad (29)$$

$$\frac{\phi}{2} + \frac{1}{\phi} \int_D^\infty z \frac{d^2f}{dz^2} dz = \frac{(1 - \phi)}{\phi}. \quad (30)$$

2. Evaluation

Here we determine $L(z)$ and $p(z)$ for our constructed configurations in one dimension. One can use a stochastic approach to compute $L(z)$ by tossing line segments of length z into the system for each z , and counting the number that land entirely in the pore space. However, we use a procedure that leads to an exact evaluation of $L(z)$ for a given configuration. The midpoint of any line segment of length z can occupy any position in the pore space that is more than $z/2$ away from any rod. The value of $L(z)$ is therefore exactly the pore space “available” to the system of N rods with diameters $D + z$. The shells of length $z/2$ that must be added to each rod end may overlap, however, since the rods are required to be only farther than D apart. We calculate $L(z)$ as the sum of all the gaps left when the rods are of size $D + z$:

$$L(z) = \frac{1}{D} \sum_{i=1}^N (\delta_i - z) U(\delta_i - z), \quad (31)$$

where δ_i is the gap between the i th and the $(i + 1)$ th particle and $U(x)$ is the unit step function. A technique like this has been used in two dimensions to compute the related nearest-neighbor probability distribution function.¹⁷

Our exact evaluation of $L(z)$ produces smooth curves. In contrast, “binned” determinations of $p(z)$ from experimental data resulted in noisy curves, even when averaged over a number of configurations. Since $f(z)$ cannot be determined exactly because we do not know the higher-order correlation functions for the step-function g_2 configurations that we have constructed, we assume that $f(z)$ can be approximated as

$$f(z) = a \exp(-bz^c), \quad (32)$$

which has the necessary non-negative and monotonic form. Here a , b , and c are positive fit parameters. Of the functional forms tried, this one satisfied the five necessary conditions most satisfactorily. The corresponding chord-length density function $p(z)$ is given by

$$p(z) = \begin{cases} \phi, & 0 \leq z \leq D, \\ \frac{abc}{\phi} \exp(-bz^c) [bcz^{2(c-1)} - (c-1)z^{(c-2)}], & z \geq D. \end{cases} \quad (33)$$

To illustrate the goodness of the fit, we consider a covering fraction $\phi = 0.4$, a value near the terminal density. We see from Fig. 9 that the fit to the simulation data is extremely good; $a = 0.57911$, $b = 0.71028$, $c = 1.1878$, giving a chi squared value of $\chi^2 = 2.8 \times 10^{-5}$. This fit satisfies the five conditions (26)–(30) to excellent approximations:

$$f(D) = 0.2846,$$

$$\left. \frac{df}{dz} \right|_{z=D} = -0.2401,$$

$$\left. \frac{d^2f}{dz^2} \right|_{z=D} = 0.1574,$$

$$\int_0^\infty p(z) dz = 1.000,$$

$$\int_0^\infty zp(z) dz = 1.447.$$

The exact values are given by 0.28, -0.24 , 0.16 , 1 , and 1.5 , respectively, at the density $\phi = 0.4$. The corresponding chord-length density function is also plotted in Fig. 9.

The dramatically different shape of $p(z)$ from the equilibrium function that can be seen in Fig. 9 corresponds to the suppression of the peak at the contact distance by the step function g_2 . With fewer particles clustering, it is expected

that the number of gaps of small size would be fewer, and this is reflected in the shape of $p(z)$. This tendency toward a more spread-out, less clustered configuration for the annealed step-function system can be seen in Fig. 3. Therefore the chord-length density function is a sensitive indicator of the difference between configurations having the step-function g_2 and the ensemble of equilibrium configurations at the same density.

IV. CONCLUSION AND DISCUSSION

Any pair-correlation function $g_2(r)$ and its associated structure factor $S(k)$ for a many-particle system are trivially constrained by non-negativity conditions: $g_2(r) \geq 0$ and $S(k) \geq 0$. Unfortunately, no general theory has been available to indicate whether any arbitrary candidate function g_2 that satisfies these two inequalities is actually attainable as the descriptor of short-range order in a specific many-particle system. In other words, is the pair of necessary inequalities also sufficient, or does there exist an as-yet undiscovered additional set of constraints on candidate g_2 's that must be satisfied to represent an attainable particle system? The work reported herein provides some modest evidence in support of the proposition that only the two non-negativity conditions are required, provided that the particle number density is not too large. If this proposition is true, it would stand in direct contrast to the properties of a related two-point correlation function $S_2(r)$ that arises in characterization of two-phase random media (i.e., binary stochastic spatial processes), where it is known that the two analogous non-negativity conditions are only necessary for realizability.⁶

The specific cases examined involved a unit step function form for $g_2(r)$, in one and two space dimensions. A simulated annealing stochastic numerical procedure served to move particles about, starting from random initial positions, so as to minimize the mean-square error between the pair-correlation function of intermediate particle configurations, and the target unit step function. Results in both one and two dimensions indicate that indeed the target was achievable up to a terminal covering fraction $\phi_c = 0.5$ in one dimension, and $\phi_c = 0.25$ in two dimensions in the infinite system limit.²⁰ These upper density limits agree with the general result for d -dimensional space, implied by a previous study,⁸ that the terminal upper density is $\phi_c = 2^{-d}$. The particle system becomes "hyperuniform" at ϕ_c ; at that density $S(0) = 0$ and long-range density fluctuations are severely inhibited.

Configurational patterns exhibited by the many-particle system subject to a unit step function g_2 are perceptibly different from those dominating the corresponding equilibrium system of hard-core particles. In particular, the distribution of distances between near neighbors, especially upon approach to the terminal density, is significantly different from that of the equilibrium case. Results show that the chord-length density function $p(z)$ is a particularly sensitive indicator of the difference.

In spite of the geometric impossibility to achieve the unit step function for $\phi > \phi_c$, our simulated annealing numerical procedure nevertheless manages to produce a closest approach to that target. Results show a pair-correlation function

that displays smaller amplitude deviations from unity beyond the exclusion diameter, compared to the equilibrium pair-correlation function at the same ϕ . However, even these diminished oscillatory deviations increase in magnitude as ϕ increases beyond ϕ_c , and our numerical procedure encountered increasing difficulty in attaining an optimal solution.

The limited, but useful, results obtained in this study encourage the aggressive pursuit of a variety of extensions. Perhaps the most obvious is to examine realizability of the unit step function $g_2(r)$ in three dimensions. Straightforward extrapolation of the results reported in this paper suggest that indeed realizability is possible up to terminal covering fraction $\phi_c = 0.125$, but not beyond. Furthermore, the chord-length density function $p(z)$ should continue to be a sensitive indicator of differences in patterns of short-range order between the constrained and the equilibrium systems.

Expanding the examination of realizability beyond just the unit step function $g_2(r)$ is of course also a vital direction for future investigation. Two cases might profitably be distinguished, depending upon what k value is involved when the condition $S(k) \geq 0$ is about to be violated, thus defining a terminal density. The cases examined in this paper have this occurring at $k=0$, giving rise to hyperuniform systems¹² at that terminal density. The other case would have $k \neq 0$, and the terminal-density systems would not be hyperuniform. It is conceivable that candidate g_2 's in the former class might always be realizable at and below their terminal densities, while those in the latter class may not be realizable to the same extent.

In addition to the extensions already mentioned, arguments can be made in favor of examining multicomponent (mixture) systems, as well as cases involving structured particles (e.g., those requiring orientational coordinates). All of these open problems, if attacked by numerical means, are likely to be even more demanding than the cases considered in this paper. Consequently, it might be advantageous to consider alternative numerical algorithms. Some that seem to be relevant are the simulated tempering, or parallel tempering, techniques that have the capacity to surmount high barriers in the configurational search space.^{21,22} Another kind of numerical advantage might stem from exploitation of the hypernetted chain (HNC) integral equation,² since it supplies a pair potential that, under equilibrium conditions, should correspond (at least approximately) to the given target $g_2(r)$. By using that HNC pair potential in a Monte Carlo or molecular-dynamics simulation, it should be possible to generate a set of good starting configurations for further refinement by the other procedures mentioned.

ACKNOWLEDGMENTS

This work was supported in part by the Petroleum Research Fund as administered by the American Chemical Society, MRSEC Grant at Princeton University, NSF DMR-0213706, and NSF under Grant No. DMS-0312067.

¹T. L. Hill, *Statistical Mechanics* (McGraw-Hill, New York, 1956).

²J. P. Hansen and I. R. McDonald, *Theory of Simple Liquids* (Academic, New York, 1986).

- ³M. D. Rintoul and S. Torquato, *J. Colloid Interface Sci.* **186**, 467 (1997).
- ⁴C. L. Y. Yeong and S. Torquato, *Phys. Rev. E* **57**, 495 (1998).
- ⁵D. Cule and S. Torquato, *J. Appl. Phys.* **86**, 3428 (1999).
- ⁶S. Torquato, *Random Heterogeneous Materials: Macroscopic Properties* (Springer-Verlag, New York, 2002).
- ⁷W. Kunkin and H. L. Frisch, *Phys. Rev.* **177**, 282 (1969).
- ⁸F. H. Stillinger, S. Torquato, J. M. Eroles, and T. M. Truskett, *J. Phys. Chem. B* **105**, 6592 (2001).
- ⁹H. Sakai, F. H. Stillinger, and S. Torquato, *J. Chem. Phys.* **117**, 297 (2002).
- ¹⁰S. Torquato and F. H. Stillinger, *J. Phys. Chem. B* **106**, 8354 (2002); **106**, 11406(E) (2002).
- ¹¹K. Z. Markov and J. R. Willis, in *Mathematical Models and Methods in Applied Sciences* (World Scientific, Singapore, 1998), Vol. 8.
- ¹²S. Torquato and F. H. Stillinger, *Phys. Rev. E* (in press).
- ¹³S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, *Science* **220**, 671 (1983).
- ¹⁴E. Aarts and J. Korst, *Simulated Annealing and Boltzmann Machines* (John Wiley and Sons, New York, 1989).
- ¹⁵B. L. Lu and S. Torquato, *Phys. Rev. A* **45**, 922 (1992).
- ¹⁶S. Torquato and B. L. Lu, *Phys. Rev. E* **47**, 2950 (1993).
- ¹⁷M. D. Rintoul and S. Torquato, *Phys. Rev. E* **52**, 2635 (1995).
- ¹⁸H. Reiss and A. D. Hammerich, *J. Chem. Phys.* **31**, 369 (1986).
- ¹⁹W. G. Hoover and F. H. Ree, *J. Chem. Phys.* **49**, 3609 (1968).
- ²⁰It should be noted that we also computed the pair-correlation function beyond the sampling distance for each density reported here and found that although it fluctuated irregularly about unity, as expected, there was no indication of a systematic deviation from the target value of unity.
- ²¹A. Mitsutake and Y. Okamoto, *Chem. Phys. Lett.* **332**, 131 (2000).
- ²²K. Makrodimitris, G. K. Papadopoulos, C. Philippopoulos, and D. N. Theodorou, *J. Chem. Phys.* **117**, 5876 (2002).