JOURNAL OF COLLOID AND INTERFACE SCIENCE **186,** 467–476 (1997)
ARTICLE NO. CS964675

Reconstruction of the Structure of Dispersions

MARK D. RINTOUL AND SALVATORE TORQUATO[1]

Princeton Materials Institute and Department of Civil Engineering and Operations Research, Princeton University, Princeton, New Jersey 08540

Received July 12, 1996; accepted October 31, 1996

To what extent can the structure of a disordered heterogeneous material be reconstructed using limited but essentially exact structural information about the original system? We formulate a methodology based on simulated annealing to reconstruct both equilibrium and non-equilibrium dispersions of particles based only on correlation functions which statistically characterize the system. To test this method, we reconstruct dispersions from the radial distribution function (RDF) associated with the original system. Other statistical correlation functions are evaluated to compare how well the reconstructed system matches the original system. We show that for low-density systems or high-density systems with little particle aggregation, our reconstruction from the RDF reproduces the system fairly well. However, for dense systems with extensive clustering, RDF information is somewhat inadequate in being able to reconstruct the original system. We also show that one can produce a system with an RDF that is similar to the reference system, but with appreciably different structure. Finally, system-size effects are analyzed analytically. © 1997 Academic Press

Key Words: **reconstruction; dispersions; correlation functions; simulated annealing.**

I. INTRODUCTION

Disordered materials are ubiquitous in nature and in man-made situations. Examples of such media include fiber-reinforced composites, colloidal dispersions, porous media, granular media, and cell membranes, to mention but a few. A question of fundamental interest that remains largely unstudied is the following inverse problem: To what extent can the disordered material be reconstructed given limited but essentially exact structural information (statistical correlation functions) on the material? The success of the reconstruction can be measured by how well the reconstructed system reproduces not only the correlation functions from which the reconstruction was performed but other correlation functions as well. Clearly, even if the correlation functions of the reference and reconstructed systems are in good agreement, this does not ensure that structures of the two systems will match very well. This interesting question of nonuniqueness can also be probed.

A general procedure to carry out reconstructions for arbitrary structures is too intractable at present. Instead we will focus our attention on an important class of random media, namely dispersions of particles. The method that we propose to reconstruct the particle dispersions from given structural information is a variation of the simulated annealing method that is commonly used to solve a variety of general problems relating to finding a state of minimum ''energy'' among a set of many local minima (1). The energy does not necessarily have to be a physical energy of the system, but can be any relevant objective function (e.g., time in a production process or material in a topological optimization problem). In these systems the simulated annealing procedure is used to sample the different states of the system weighted by the probability of the occurrence of that state.

An important by-product of this method is that a system in an initial state which is far from its limiting state will move through the phase space toward its limiting state, so the simulated annealing method is used not so much to sample, but to find minimum energy states. We can use this notion of moving ''toward'' the desired state to evolve a system of particles toward a system which has some specific correlation function nearly identical to a given reference correlation function which represents the system one is attempting to reconstruct. If we consider the states whose correlation function is more similar to the trial correlation function to have a lower ''energy,'' the system will eventually move toward a state in which the difference between the two functions is minimized.

Given the prescription above for the reconstruction procedure, one must now select the appropriate set of correlation functions. In Section II, we present a general formulation that enables one to incorporate a set of different n-point correlation functions. In practice, it is desired to use a single correlation function that contains the salient microstructural features for a particular application. If one is interested in the shape of the space not occupied by the particles, then a single function such as the pore-size distribution $P(r)$ (2) could be used. Alternatively, if one is interested in the clustering of particles which may be occurring, one could use the pair-connectedness function (3) or the two-point cluster function (4). Practical concerns could also dictate the choice

[1] To whom correspondence should be addressed.

0021-9797/97 $25.00

of correlation function, such as the case in which one is given a specific function which is measured from an experimental system.

The radial distribution function (RDF), $g(r)$, is perhaps the most common statistical correlation function used when characterizing statistically isotropic particle systems in D dimensions. Roughly speaking, it gives the probability associated with finding any particle a radial distance r from the center of another particle. The RDF is of central importance in the study of equilibrium liquids in which the particles interact with pairwise-additive forces since all of the thermodynamics can be expressed in terms of the RDF. Furthermore, the RDF can be ascertained from scattering experiments, which makes it a likely candidate for the reconstruction of a real system.

Although $g(r)$ is a useful quantity, it is inherently a two-particle measure and does not account for any cooperative behavior between groups of more than two particles. This effect is significant for dense systems. Since the RDF of a system is commonly used to characterize its structure, it is natural to ask questions about how much structural information is contained within the RDF, and what sort of information is lost. We have chosen to study this question *quantitatively* by attempting to recreate the structure of various dispersions of particles from their respective radial distribution functions. The reconstructed systems are then compared with the original systems, both visually and by measuring other statistical correlation functions associated with original and reconstructed system.

As a test of this method we consider the reconstruction of two different model systems. The first is a random sequential adsorption system (5, 6) of hard disks, and the second is a sticky-disk system in which the disks can form interparticle contacts (8) and thus cluster. From our studies of these systems, we have observed the following salient points:

- At low densities, the reconstruction from the RDF captures much of the structural information.
- At high densities, the RDF can give a reasonable reconstruction if there is no tendency for the particles to aggregate.
- If there is significant aggregation of the particles, the RDF is usually not adequate to reconstruct the system.
- Unless care is taken, one can produce a system which has an RDF that is similar to the reference system but with an appreciably different structure.

In Section II, we formulate the reconstruction procedure and define relevant statistical correlation functions. This procedure is applied to random sequential adsorption systems in Section III and then to sticky-disk systems in Section IV. We discuss important features of this method in Section V and suggest extensions to the model. Moreover, the Appendix gives an analytical expressions for the relationships between the number of particles in the system and the resolution of the statistical quantities.

II. FORMULATION OF THE RECONSTRUCTION PROCEDURE

A. General Procedure

For reasons of simplicity, we first formulate our reconstruction methodology by considering a single correlation function associated with the original, or ''reference'' system. We then generalize the procedure to incorporate a set of different n-point correlation functions.

Consider a reference two-point correlation function of the form $f_0(r)$ that depends only on the distance r between two points in the system, implying that the system of interest is statistically isotropic. The same correlation function that is associated with the reconstructed system of disks will be denoted as $f_s(r)$, and it is this system which we will attempt to evolve toward $f_0(r)$. Because $f_s(r)$ of our hard-disk system is based on the distribution of the distances between disks in a finite system, it is necessary to discretize $f_s(r)$ through a process of binning. In this process, the relevant values of the correlation functions at different interparticle separation distances r_{ij} are calculated and are placed in equal sized bins of width w_b centered at r_k, such that

$$r_k - w_b/2 \leqslant r_{ij} < r_k + w_b/2. \qquad [1]$$

Once $f_s(r)$ is known at the specific values of the bins, one can define a variable E which plays the role of the energy in the simulated annealing calculation as

$$E = \sum_i \beta_i (f_s(r_i) - f_0(r_i))^2, \qquad [2]$$

where β_i is an arbitrary weight which is implicitly a function of r through its dependence on the ith bin, and the sum is over all bins. Defined in this manner, E has the property of decreasing when the difference between any two bins decreases. The variable β_i can be varied depending on whether one wants to weight small values of the radial distance r greater than larger values of r. This is sometimes useful since the small r values of $f_s(r)$ are more important in determining the system structure.

In order to evolve the system toward $f_0(r)$, we choose a particle i and give it a random displacement $\delta\mathbf{r}$. If this causes the particle to overlap, the move is rejected, and another particle is tried. Otherwise, the energy E' of this state is calculated. We can now calculate the energy difference between two states $\Delta E = E' - E$, and define the acceptance probability of the move $p(\Delta E)$ via the method originally proposed by Metropolis *et al.* (and used in many simulated annealing calculations) as

$$p(\Delta E) = \begin{cases} 1, & \Delta E \leqslant 0 \\ \exp(-\Delta E/kT), & \Delta E > 0. \end{cases} \qquad [3]$$

This method causes $f_s(r)$ to move gradually closer to $f_0(r)$. The value of T is chosen to allow the system to evolve to the desired state as quickly as possible, without getting trapped in any local energy minimum. For most simulated annealing problems, T is varied as a function of time in the simulation in order to reach the desired result as quickly as possible. This variation of T as a function of time is known as the *cooling schedule.*

This procedure can be naturally generalized to a more complicated dependence on the positions of the particles. This is done by considering a reference n-point correlation function $f_0(\mathbf{r}^n)$, which depends upon n different positions $\mathbf{r}^n = \mathbf{r}_1, \ldots, \mathbf{r}_n$. The binning procedure is then a multidimensional one, but is essentially the same. Finally, one can even extend the process to m different n-point correlation functions by defining a reference function $f_0(\mathbf{r}^n)$, where

$$f_0(\mathbf{r}^n) = f_0^{(1)}(\mathbf{r}^n) + f_0^{(2)}(\mathbf{r}^n) + \cdots + f_0^{(m)}(\mathbf{r}^n), \quad [4]$$

(and a similar accompanying $f_s(\mathbf{r}^n)$) giving an energy

$$E = \sum_i \beta_i (f_s(\mathbf{r}^n) - f_0(\mathbf{r}^n))^2, \quad [5]$$

where the sum is a multidimensional one over all bins, and again, β_i depends implicitly on $\mathbf{r}^n$ through the ith bin.

B. Procedure Using the RDF

For purposes of illustration, we specialize in this paper to the case when the reference correlation function is taken to be the well-known radial distribution function (RDF) $g(r)$. The quantity $\rho 2\pi r g(r) dr$ gives the average number of particle centers in an annulus of thickness dr at a radial distance of r from the center of a particle (where ρ is the number density). In this case, we just must calculate all of the interparticle distances and bin them. In this study, the bin width w_b was always chose to be 0.002 times the diameter of a disk. For the case of using the RDF as the reconstruction function, we have found that setting $\beta_i = r_i^{-(D-1)}$, causes the system to converge the fastest in most cases, where D is the dimension of the system and is equal to 2 for hard disks.

We attempted to find an ideal cooling schedule for this problem, but found that the behavior was not significantly improved from the constant T case for the various cooling schedules tested. The particles chosen for test moves can either be chosen randomly or sequentially. In the results presented here, the particles were chosen sequentially.

C. Accuracy Tests

To test the accuracy of the reconstruction procedure, we will not only compare the images of the systems but will compare other correlation functions. Two such quantities that we consider are the *two-point probability function* $S_2(\mathbf{r}_1, \mathbf{r}_2)$ (9), and the *pore-size distribution function,* $P(r)$ (2). The two-point correlation function is defined as the probability of having two points, located at $\mathbf{r}_1$ and $\mathbf{r}_2$ which both lie in the phase of interest (the disk phase in this case). Since we will be dealing with an isotropic media in all cases, we can write this simply as $S_2(r)$, where r is the distance between the two points. For all isotropic media in which there is no long-range order, we know that

$$S_2(0) = \phi \quad \text{and} \quad \lim_{r\to\infty} S_2(r) = \phi^2. \quad [6]$$

It should be emphasized that the two-point probability function is different from $g(r)$ in that it emphasizes the correlations between all points in the disks and not just the centers.

The quantity $P(r)dr$ is defined as the probability that a randomly chosen point in the void phase lies at a distance between r and $r + dr$ of the nearest point on the interface between the two phases. It follows that the extreme limits of the pore-size distribution function as

$$P(0) = s/\phi \quad \text{and} \quad P(\infty) = 0, \quad [7]$$

where s is the specific surface of the system. The pore-size distribution function gives information about the ''width of the pores'' in the system. Because it is a measure of pore size, it also indirectly contains connectedness information. While both $S_2(r)$ and $P(r)$ are related to $g(r)$, they emphasize different features of the system; systems with similar RDF's will not necessarily have similar values of $P(r)$.

III. RECONSTRUCTION OF RANDOM SEQUENTIAL ADSORPTION SYSTEMS FROM EQUILIBRIUM SYSTEMS

The first system studied was a non-equilibrium system created by the random sequential adsorption process (RSA) (5, 6). In this process disks are placed randomly and sequentially on a surface such that each disk is adsorbed if it does not overlap any of the disks already adsorbed. The geometrical blocking effects and the irreversible nature of the process result in structures that are distinctly different from corresponding equilibrium configurations, except for low densities (5). The jamming limit (the final state of this process whereby no particles can be added) occurs at a volume fraction $\phi \approx 0.547$ (7).

The reconstruction of the RSA systems was done at two volume fractions. The first, $\phi = 0.20$, was chosen because it was a system which was fairly similar to the initial equilibrium system, but with enough differences for the change between the two to be noticeable. The second volume fraction used was $\phi = 0.543$, which is very close the jamming limit in two-dimensions. In this case, there is a significant difference between the correlation functions associated with RSA and the equilibrium system. It is also of interest due

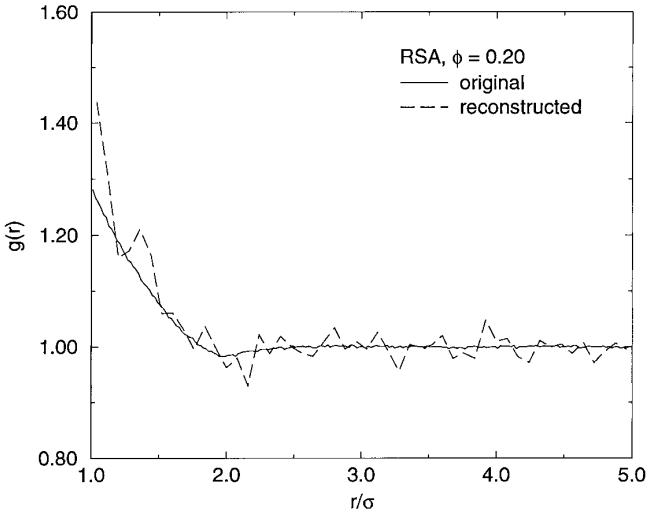

FIG. 1. The radial distribution function (RDF) $g(r)$ vs the dimensionless radial distance r/σ for an RSA system with $\phi = 0.20$ compared to the reconstructed system.

to the difficulty in obtaining the RSA configuration at such a high density. In this case, we might even expect to see the difference visually, by simply looking at a picture of the disks themselves.

In each case, an initial configuration of 5000 disks was obtained by first starting with 5000 random overlapping disks which were randomly (Poisson) distributed, and then moving them to eliminate overlap. They were equilibrated using a hard-disk potential until they reached an equilibrium state and the RDF no longer changed as a function of time. The densities used in this study were below the freezing density for the two-dimensional hard-disk system, so crystallization was not a problem. We have found that at higher densities, starting with a random configuration (as opposed to a crystal) leads to shorter equilibration times, especially when the number of particles chosen is not a number associated with a vacancy-free crystal.

A. RSA at $\phi = 0.20$ from an Initially Random Array

For this first case, the reference RDF is shown along with the reconstructed RDF in Fig. 1. Here, the reconstruction quickly reproduces the reference RDF, although the fluctuations in the RDF due to the low density of disks are significant. One can show that the differences between the two curves are indeed due to the fluctuations caused by a finite number of particles in the bins by applying Eq. (A3) with $N = 5000$, $\phi = 0.2$, and $w_b/\sigma = 0.08$ (as was used in this case for the sake of clarity) to get a predicted fluctuation of about 0.04. This is very much in line with what is seen in Fig. 1. Both $S_2(r)$ and $P(r)$ also are indistinguishable when plotted together. This case was interesting not only as being the most easily reconstructed system, but as a test for a choice of temperature. Because the RDF in this system was very close to the initial equilibrium RDF, the temperature had to be kept very low. If it was not kept low enough, the reconstructed system would show more equilibrium behavior, and this would be reflected in the RDF.

B. RSA at $\phi = 0.543$ from an Initially Random Array

Figure 2 shows that the reference RDF and reconstructed RDF still agree fairly well at $\phi = 0.543$ in the RSA system, although the agreement is not quite as good as the case where $\phi = 0.20$. There is a slight discrepancy near the minima and maxima of the original RDF, with the peaks being slightly larger in the reconstructed RDF. The reason for this is partially due to the finite temperature used in the reconstruction. This temperature, which allows some moves to occur regardless of their effect on the RDF, causes the system to have some small amount of equilibrium type behavior. This behavior is reflected by the more pronounced extrema in the reconstructed system. However, the actual disk systems themselves are difficult to distinguish visually. Figure 3 shows two systems of disks, the first of which is an actual RSA system at $\phi = 0.543$, while the second is the reconstructed system at the same volume fraction.

Like the $\phi = 0.20$ case, the plots of $S_2(r)$ for the original and reconstructed cases are practically identical. This is to be expected due to the fact that the radial distribution functions are so similar, and $S_2(r)$ essentially contains less information than $g(r)$ for the hard-disk system. However, the plots of the two pore-size distributions, $P(r)$, show a bit of a difference, with the reconstructed $P(r)$ having a slightly longer range. This is due to the fact that the reconstructed system tends to ''cluster'' the particles together in order to get the required contact values, while the original RSA sys-

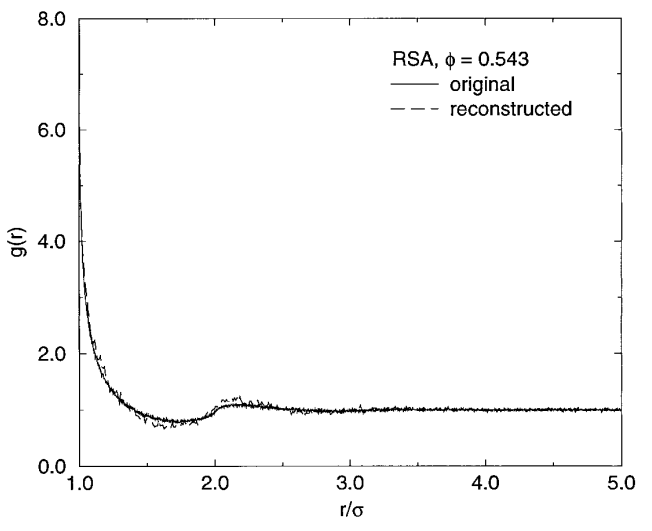

FIG. 2. The radial distribution function (RDF) $g(r)$ vs the dimensionless radial distance r/σ for an RSA system with $\phi = 0.543$ compared to the reconstructed system.

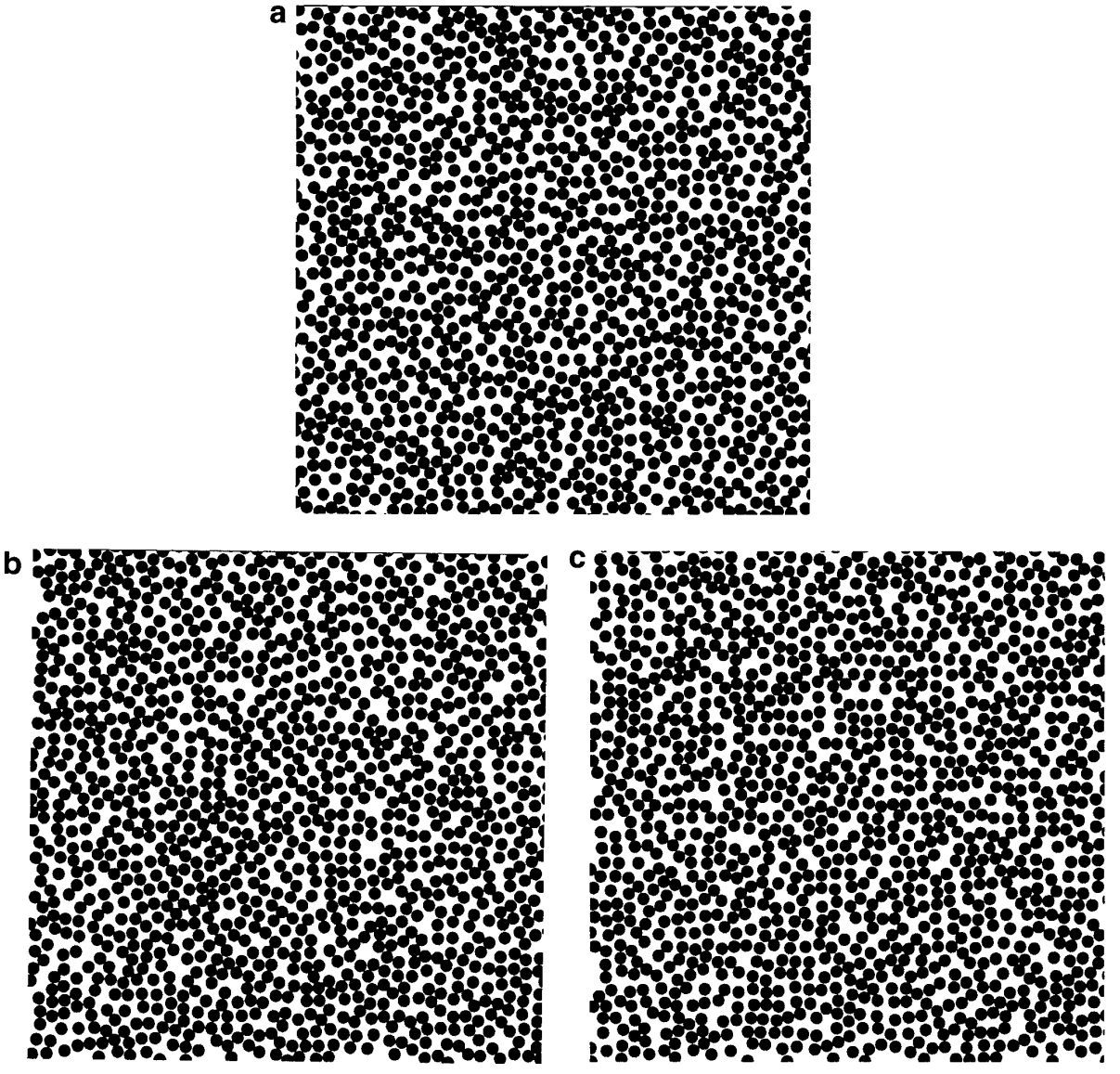

FIG. 3. (a) A portion of a sample RSA system at $\phi = 0.543$. (b) A portion of the reconstructed RSA system at $\phi = 0.543$. (c) Same as b, but with the initial condition being a square lattice. The systems are fairly similar, but b occasionally contains large void areas which are not seen in the true RSA system. Also, the initial lattice structure is clearly evident in c.

tem has very few large pores, due to the nature of the adsorption process.

C. RSA at $\phi = 0.543$ from an Initially Ordered Array

A final interesting example is that of reconstructing a given disordered system from an initial ordered system. In this case we started with a system of 5041 particles (71^2) at a volume fraction of 0.543 initially arranged in a square lattice, and attempted to evolve the system toward the RSA system at $\phi = 0.543$ used previously.

The final RDF is shown along with the original RDF in Fig. 4. The primary difference between the two is the slight oscillation seen at longer distances for the reconstructed system that is not seen in the original system. To see where this arises from, we can look at a picture of the reconstructed system in part c of Fig. 3 and see that there remains some remnant of the initial lattice structure as evidenced by the preponderance of vertical and horizontal organization of the disks. *This example gives strong evidence that the reconstructed system can match the original RDF very well and has a significantly different structure.*

IV. RECONSTRUCTION OF STICKY-DISK SYSTEM FROM EQUILIBRIUM SYSTEMS

The sticky-disk model used here is the one for which Baxter (8) obtained the exact solution to the Percus–Yevick

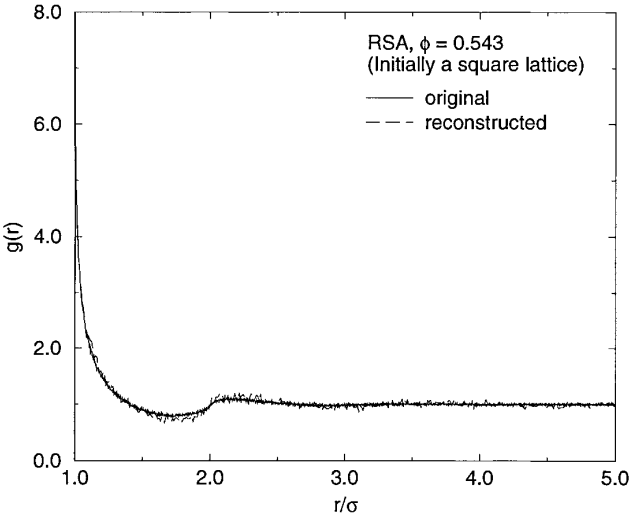

FIG. 4. The RDF $g(r)$ for an RSA system vs the dimensionless radial distance r/σ with $\phi = 0.543$, compared to the reconstructed version, in which the reconstructed system began with the disks originally arranged in a square lattice.

equation in three dimensions and calculated the equation of state. In this model, the interparticle potential is defined by

$$\beta u(r) = \begin{cases} \infty, & 0 < r < \sigma \\ -\ln[d/12\tau(d-\sigma)], & \sigma < r < d \\ 0, & r > d \end{cases} \quad [8]$$

in the limit that $\sigma \to d$. The quantity τ^{-1} represents the adhesiveness of the potential, and the hard disk limit is recovered in the limit that $\tau \to \infty$. Thus, clustering occurs because of the adhesion in the potential. The variable adhesion strength can be tuned to form clusters of larger and larger size (as τ becomes smaller and the adhesion becomes stronger) in order to describe various systems. The algorithm used to create these systems on the computer is the one given by Seaton and Glandt (11). It should be noted that this algorithm is a low-density approximation, and the resulting systems are not true equilibrium configurations corresponding to the potential in (8). However, we are using this model only as a means of producing aggregates of particles and the low-density approximation will produce this effect.

The sticky-disk model, despite its simplicity, turns out to be a very useful representation of many experimental colloidal systems. Indeed, this model closely resembles the experimental dispersions produced by Trau *et al.* (12) in which monolayers of small spheres are deposited on surfaces via electrohydrodynamic effects.

A picture of what a typical sticky-disk system looks like is shown in Fig. 5, where $\phi = 0.4$ and $\tau = 0.05$. Note that partial crystallization is induced by the stickiness. Reconstructions of these systems are made especially tricky by the ''singularities'' that form in the RDF's of the original systems. This is due to the lattice that forms with the touching sticky particles. In a finite system, where binning is used, there are no singularities as such, but there are bins which contain one or two orders of magnitude more points than their neighboring bins. We chose to study this case because it is interesting to see whether or not the computer will automatically discover the partial crystallization from the RDF, or will try to make some other arrangement to obtain the reference RDF. Note that the method we are using is not necessarily the best method one would use to reproduce a sticky-disk system but is useful in getting a feel for how much information is in the RDF of the sticky-disk system.

A. Sticky Disk Model with $\phi = 0.543$, $\tau = 0.10$

The reference RDF and final reconstructed RDF are shown together in Fig. 6. Although it cannot be displayed on the graph, the singularities of the reconstructed system are significantly smaller than those of the original system. (Note that these heights are a function of the bin width, which in this case is 0.002.) The reconstructed RDF looks significantly different from the reference RDF. This is due to the fact that, as the algorithm attempts to fill up the bins corresponding to the singularities, it removes points from the nearest bins, resulting in a significant deficit in those bins. This curve was obtained after many iterations, and appears to be the stable solution to the problem.

The curves comparing $S_2(r)$ for the two cases are shown in Fig. 7. In this case, the agreement is somewhat better than the agreement for the RDF comparison. This is not surprising since $S_2(r)$ contains less information than $g(r)$ and so would

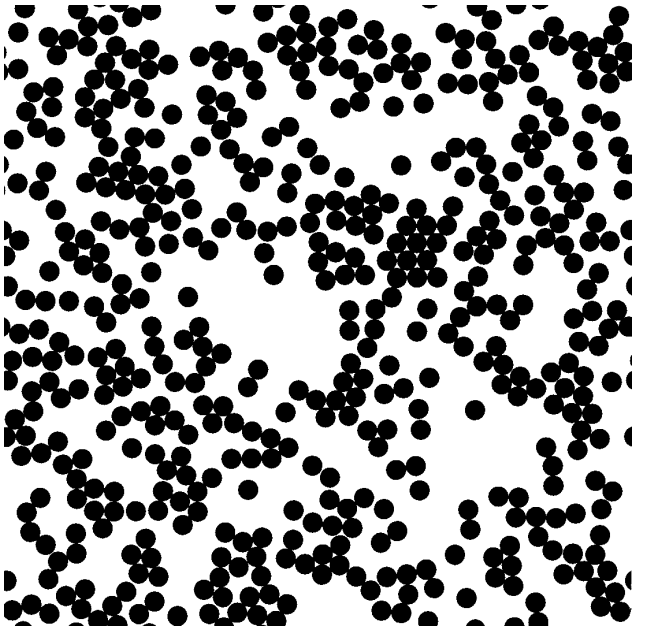

FIG. 5. A sample sticky-disk system with $\phi = 0.4$ and $\tau = 0.05$.

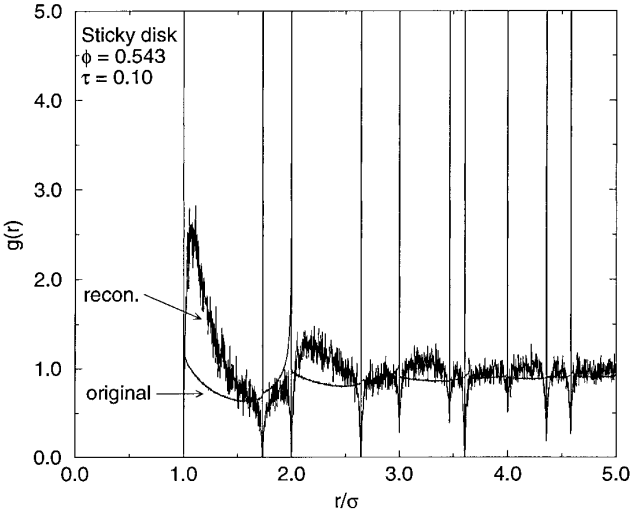

FIG. 6. The RDF $g(r)$ vs the dimensionless radial distance r/σ for the sticky-disk system with $\phi = 0.543$ and $\tau = 0.10$ compared to the reconstructed version.

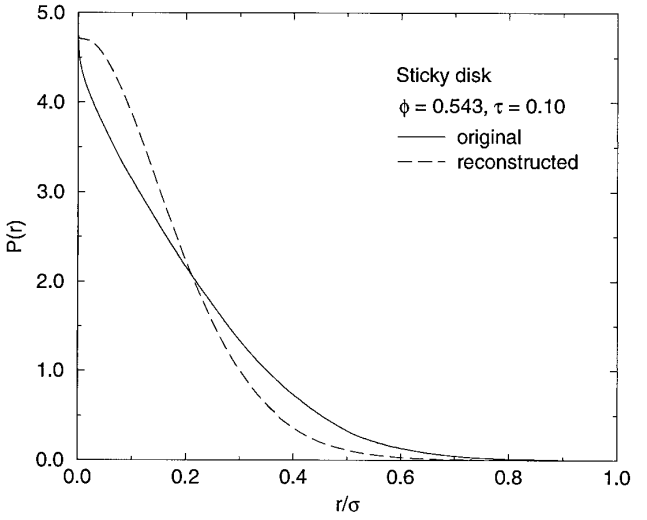

FIG. 8. The pore-size distribution function $P(r)$ vs the dimensionless radial distance r/σ for the sticky-disk system with $\phi = 0.543$ and $\tau = 0.10$ compared to the reconstructed system.

be less able to distinguish between the two cases. The reconstructed image shows slightly stronger correlations than the original case. This is due to the fact that there is better short distance packing in the real sticky-disk system than in the reconstructed system.

One of the more significant differences between the reconstructed image and the original image is seen in Fig. 8 from comparing the pore-size distribution of the two. The pore-size distribution is much longer ranged in the original image than in the reconstructed image. This is due to the fact that many of the disks in the original image were arrange in tightly packed clusters, as opposed to the reconstructed image which had the disks more spread out. Because of this, there were many large areas in the original image which had no disks, while this was not usually the case in the reconstructed image.

B. Sticky Disk Model with $\phi = 0.543$, $\tau = 0.05$

In this case, the disks are stickier, and the peaks associated with the partial crystallization are higher, and constitute a large part of the radial distribution function. This RDF's are compared in Fig. 9. For this situation we find an even worse agreement between the RDF's than we saw for the $\tau = 0.10$ case. This case is notable however, because the two-point correlation function $S_2(r)$ of the reconstructed system also

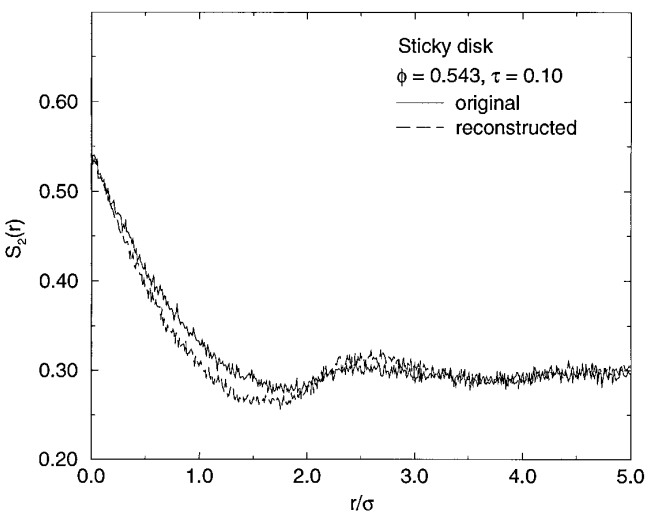

FIG. 7. The two-point correlation function $S_2(r)$ vs the dimensionless radial distance r/σ for the sticky-disk system with $\phi = 0.543$ and $\tau = 0.10$ compared to the reconstructed system.

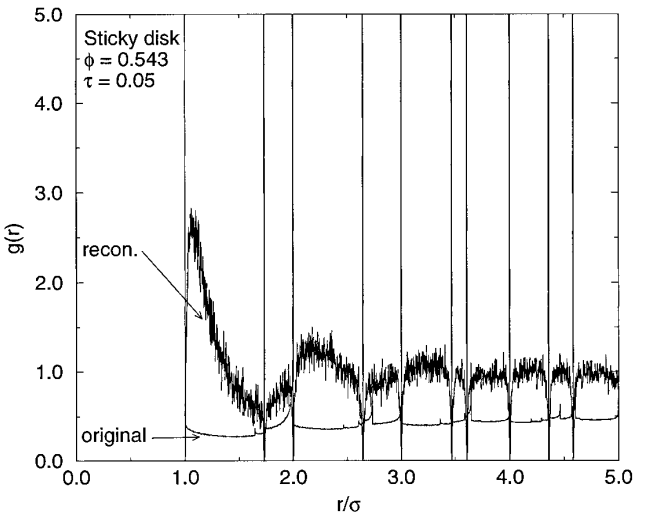

FIG. 9. The RDF $g(r)$ vs the dimensionless radial distance r/σ for the sticky-disk system with $\phi = 0.543$ and $\tau = 0.05$ compared to the reconstructed version.

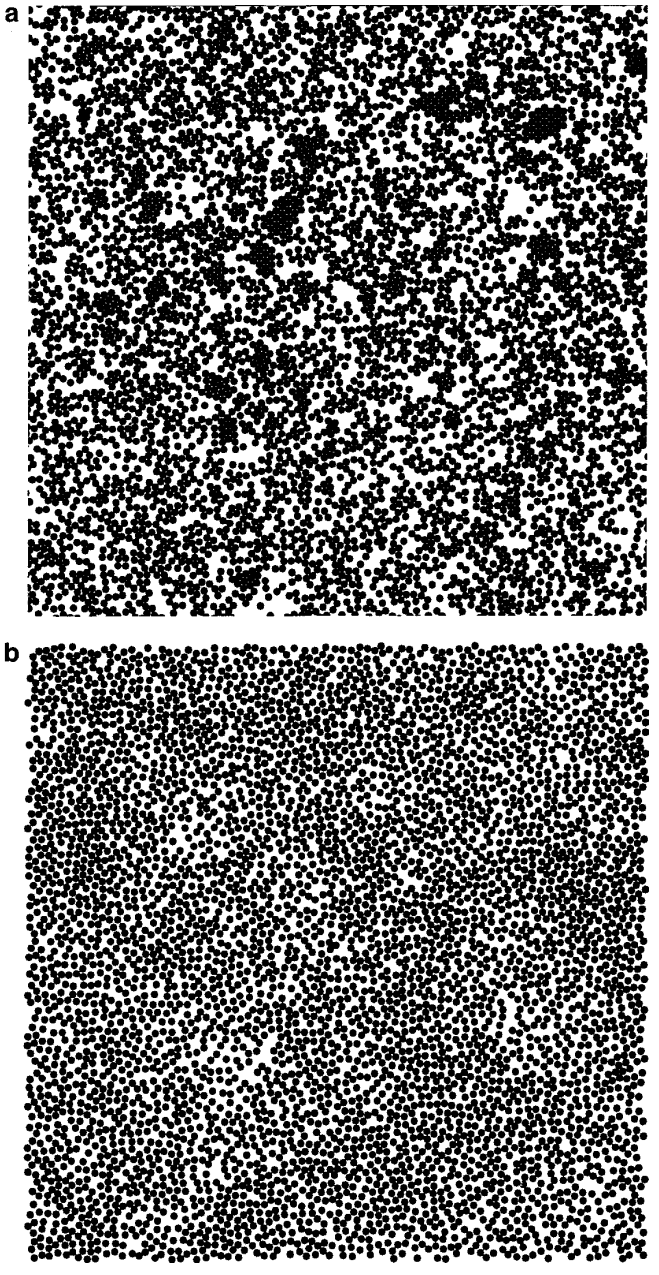

FIG. 10. (a) A portion of the real sticky-disk system at $\phi = 0.543$ and $\tau = 0.05$. (b) A portion of the reconstructed version of the same system as a. The reconstruction cannot capture the crystallization in this case, although there are much larger pores than one would expect in an equilibrium system.

does not match up well with the $S_2(r)$ associated with the original system. The plot of $S_2(r)$ for the original image decreases fairly monotonically to it long range value, while oscillations are seen for the reconstructed image. Again, this is due to the exclusions induced by the mostly disordered reconstructed system. A real sticky-disk system at $\phi = 0.543$ and $\tau = 0.05$ is compared against the final reconstructed system in Fig. 10. Note that the extensive crystallization seen in the true sticky-disk system is not seen in the reconstructed version here.

Finally, the differences in range seen between the pore-size distribution functions of the two systems shown in Fig. 11 are similar to those seen in the $\tau = 0.10$ case but are magnified significantly. There is another difference between the two images that stands out in this case. The partial lattice nature of the original system causes a very specific size and shape of pore to appear very often, corresponding to the space between the disks which are arranged in a triangular lattice. These pores make a strong contribution to the pore-size distribution, and discontinuities appear in the pore-size distribution because of these contributions. Because the reconstructed model doesn't have these pores which occur very frequently, the distribution is much more continuous.

C. Discussion of Sticky-Sphere Results

The reason the curve does not approach the reference RDF any more than it does is related to the fact that eventually, many of the disks become part of a disk pair that makes up a bin at a singularity distance. Because these bins values are still far away from the reference RDF values, it becomes very expensive energy-wise to move them out of that bin, so they effectively get stuck there. Soon, most of the particles are associated with one of these singularity bins, and it becomes difficult to move them without having a temperature in the system which is prohibitively large.

This final result is definitely not at all similar to the original sticky-disk system. The primary difficulty in reaching the sticky-disk system is that it becomes stuck in these intermediate states, which is a standard problem of simulated annealing type of calculation. However, the reason that it is so easy to get stuck in these states is that the radial distribu-

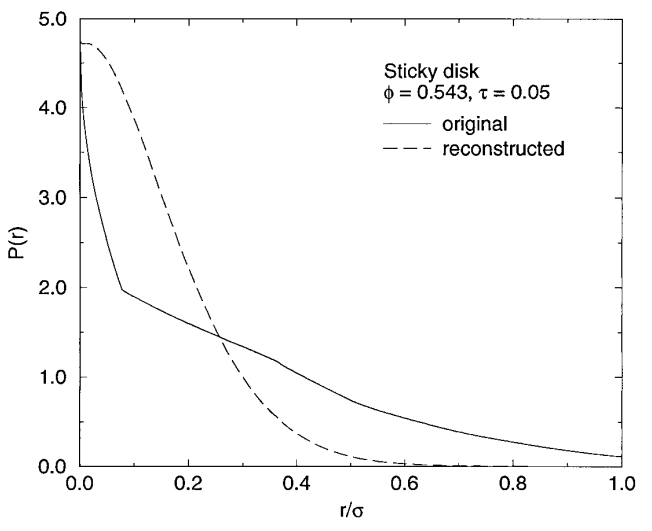

FIG. 11. The pore-size distribution function $P(r)$ vs the dimensionless radial distance r/σ for the sticky-disk system with $\phi = 0.543$ and $\tau = 0.05$ compared to the reconstructed system.

tion function is a two-particle correlation function and does not measure any sort of collective behavior. If one used, for example, a three-particle distribution function, the triangular lattice that underlies many of the smaller clusters would be immediately obvious. When at two-particle distribution function is used, there is very little incentive for the system to collect particles in a lattice shape.

V. DISCUSSION OF METHOD

As stated earlier, there is an implicit assumption of ergodicity in the system which is being reconstructed when the reference function is only a function of r. In terms of this problem, this means that we can use a single large system to reconstruct the system instead of devising a scheme which involves averages over different realizations. For an ergodic system, the spatial average over all of the particles of one large system replaces the average over realizations. In a similar manner, any reference function obtained from a statistically isotropic experimental system will also be ergodic in the limit of an infinitely large experimental system. In practice, one deals with finite experimental systems, but we have found that this simply causes the reference function to be slightly noisy and doesn't significantly change the structure of the reconstructed system.

One modification that could have been made to the sticky-disk case (and other similar cases in which ones has hard disks with attractive forces) was to change the step choosing algorithm to sometimes attempt to move a particle to be adjacent to one or more other particles. The relative fraction of moves of this type and totally random moves could be adjusted constantly by the program as it progresses so as to maximize the convergence toward the final result. Written in this manner, the program could be applied to general cases in which there was no stickiness and quickly realize this by the lack of a delta function at the origin and adjust its parameters accordingly.

We again emphasize the fact that the method defined here could be used for any correlation function $f(\mathbf{r})$. We have specialized here to when $f(\mathbf{r})$ is equal to $g(r)$, but other functions would also be natural for the systems studied here. The collective behavior exhibited by the sticky-disk system could have been better captured by the *three-particle distribution function*, $g_3(\mathbf{r}_1,\mathbf{r}_2,\mathbf{r}_3)$. It would have also been interesting to use the pore-size distribution $P(r)$ as given information. This is a different type of distribution function from the point of view that it does not provide n-point information, but gives the distribution of minimum distances from a randomly chosen point in the void space to the nearest surface. Because of this definition, it also contains some degree of connectedness information. The partially crystalline structure of the sticky-disk system is seen by this function also due to the preponderance of distances associated with the small pores between the particles when they are arranged in a lattice. But like the reconstruction with $g(r)$, the particles would likely try to arrange themselves in a different fashion in order to match this distribution.

If one was especially interested in the clustering in the system, functions which reflect clustering properties would be more useful than $g(r)$, such as the pair-connectedness function (3) and the two-point cluster function (4). The difficulties that we encountered while trying to reconstruct the sticky-disk systems were related to the fact that the RDF did not contain this information, and were not related to the simulated annealing method itself. The only drawback of using cluster oriented correlation functions is that they are normally somewhat more difficult to compute, and therefore increase the time used in the simulation. In order for the algorithm to work effectively, one must be able to easily calculate the change in the correlation function when a disk is moved, and most of the reference correlation functions suggested fall into that category.

Although we specialized to equi-sized disks in this simulation, our procedure can treat disks or spheres with a size distribution. Moreover, if the reference system was not a particle system, it might be more appropriate to use a pixel-based model in which a discrete grid of squares or cubes is used as the system on which the reconstruction is done. A random selection of pixels could be assigned as being occupied, corresponding to the volume fraction of the system being reconstructed. Then, the reconstruction could be performed in a manner similar to the disk system by taking an occupied pixel, and exchanging its position with an unoccupied square. The move would be rejected or accepted based on whether or not the new correlation function for the system was ''closer'' to the reference function. This method is appropriate for digitized media, in which there could be a one-to-one correspondence between the pixel in the original digitized image and the squares in the reconstruction grid. For the grid reconstruction, one would not be able to use n-particle distribution functions as the correlation function, but one could still use n-point correlation functions.

VI. CONCLUSIONS

The reconstruction algorithm which we have presented here gives a quantitative means by which a system can be reconstructed from one or more correlation functions that are associated with the system. This algorithm is a straightforward application of simulated annealing, which has proven to be highly effective in many other applications which require one to ''search'' through many solutions to find the optimal one for a problem. For low density systems, this method can quickly reproduce most of the structural properties in the system, given a reference RDF. At high densities, the method still gives a reasonable reconstruction of the original system, assuming there is no tendency for the particles to aggregate. Although the specific method was

not able to reproduce the high density aggregated systems which were tested, we believe this to be more related to the specific implementation than the general simulated annealing algorithm itself. Finally, we were able to generate two different systems with greatly varying amounts of disorder but practically identical RDF's. This important result shows the limitations of using the RDF alone to reconstruct all structural information in a system.

APPENDIX: FINITE-SIZE RECONSTRUCTION PROBLEM

In all of the specific reconstructions studied in this paper, we have started with a reference RDF which was assumed to be ''perfect.'' Although the reference RDF's were simply obtained through simulations of a finite number of particles themselves, one could average over an arbitrarily large number of ensembles until the fluctuations had been reduced to a prescribed small amount. However, in a real experimental system, one often has a single realization of a limited number of particles. Because of the finite number of inter-particle distances, the widths of the bins used to obtain the RDF must be kept large enough so that fluctuations don't overwhelm the RDF. On the other hand, if the bins are too wide, there is not enough information to accurately reproduce the experimental system.

One can choose the optimal bin width based on the size of the fluctuations that one is willing to tolerate. If we define w_b as the width of the bins, and N as the number of disks in the system, we can calculate the approximate number n of inter-particle distances which would fill the bin centered at r as

$$n \approx 8\left(\frac{r}{\sigma}\right)\left(\frac{w_b}{\sigma}\right)\phi N g(r). \qquad \text{[A1]}$$

Since the fluctuations will go approximately as $\sqrt{n}$, the relative error in n will go approximately as $1/\sqrt{n}$. If we set α equal to the relative error that we desire, then we know that our bins must be on the order of

$$\frac{w_b}{\sigma} \approx \frac{1}{8g(r)(r/\sigma)\alpha^2\phi N} \qquad \text{[A2]}$$

where $g(r) \approx 1$ for all values of r except those near σ. Since we want w_b to be large enough to handle the worst case, we must choose the maximum value of the right hand side of the previous equation. For most cases, this occurs when $r/\sigma \approx 1.5$ and $g(r)$ is somewhat less than 1, giving their product to be approximately unity. Thus, we can write that the best choice for w_b is

$$\frac{w_b}{\sigma} \approx \frac{1}{8\alpha^2\phi N} \qquad \text{[A3]}$$

In a similar manner, if one is starting with ''perfect'' reference RDF, then the above formula gives the relation between the number of disks and the bin width that one uses in reconstruction algorithm.

ACKNOWLEDGMENTS

We thank Roger French for very useful conversations. We gratefully acknowledge the support from the U.S. Department of Energy, Office of Basic Energy Sciences under Grant DE-FG02-92ER14275, and from the MRSEC Program of the National Science Foundation under Award Number DMR-9400362.

REFERENCES

1. Brooks, S. P., and Morgan, B. J. T., *Statistician* **44,** 241 (1995).
2. Torquato, S., and Avellaneda, M., *J. Chem. Phys.* **95,** 6477 (1991).
3. Coniglio, A., De Angelis, U., and Forlani, A., *J. Phys. A* **10,** 1123 (1977).
4. Torquato, S., Beasley, J. D., and Chiew, Y. C., *J. Chem. Phys.* **88,** 6540 (1988).
5. Widom, B., *J. Chem. Phys.* **44,** 3888 (1966).
6. Feder, J., *J. Theor. Biol.* **87,** 237 (1980).
7. Hinrichsen, E., Feder, J., and Jøssang, T., *J. Stat. Phys.* **44,** 793 (1986).
8. Baxter, R. J., *J. Chem. Phys.* **49,** 2770 (1968).
9. Torquato, S., and Stell, G., *J. Chem. Phys.* **82,** 980 (1985).
10. Pomeau, Y., *J. Phys. A* **13,** L193 (1980). Swendsen, R. H., *Phys. Rev. A* **24,** 504 (1981).
11. Seaton, N. A., and Glandt, E. D., *J. Chem. Phys.* **84,** 4595 (1986).
12. Trau, M., Saville, D. A., and Aksay, I. A., *Science* **272,** 706 (1996).