Finite-size effects in molecular dynamics simulations: Static structure factor and compressibility. I. Theoretical method

J. J. Salacuse

*Science and Mathematics Department, GMI Engineering and Management Institute, 1700 West Third Avenue,
Flint, Michigan 48502-2276*

A. R. Denton* and P. A. Egelstaff

Department of Physics, University of Guelph, Guelph, Ontario, Canada N1G 2W1

(Received 26 January 1995)

A method is presented for determining the static structure factor $S(Q)$ of a uniform bulk fluid, for *all* wave-vector magnitudes Q, from computer simulation data for a finite (N-particle) system. The method is based on the usual Fourier transform relation between $S(Q)$ and the radial distribution function $g(r)$ and on the theory of explicit finite-size effects, in particular a Taylor-series expansion, in powers of $1/N$, for $g_N(r)$, the N-particle counterpart of $g(r)$. A time-dependent generalization of the method to the intermediate scattering function $I(Q,t)$ is also presented. Implicit finite-size effects are also considered, but are found to be negligible for systems to which we have applied the method.

PACS number(s): 61.20.Gy, 61.20.Ja, 02.70.Ns, 05.70.Ce

I. INTRODUCTION

Finite-size effects arise inevitably in any computer simulation and must be carefully accounted for before simulation data may be directly compared with experiment or theory. In addition to the usual random statistical errors associated with averaging over limited numbers of particles, systematic errors may also occur due to the finite size of the model system. In this paper we propose a method of correcting for size effects in the calculation of the static structure factor of a dense fluid by computer simulation. In the following paper [1] (paper II) we proceed to test and demonstrate the utility of the method by applying it to molecular dynamics (MD) simulation data for a model krypton fluid.

Two general types of size effect have been identified. One is the well-known *explicit* (or ensemble) size effect [2,3] that results from fixing the number of particles and is commonly encountered in canonical and microcanonical ensemble simulations. The other, and more subtle, is the *implicit* (or anomalous) size effect [3–5], whose origin may be traced to the (usually periodic) boundary conditions. Both types of size effect can directly influence the microscopic structure of the system, as characterized by the pair- and higher-order correlation functions, and consequently any derived thermodynamic or structural properties.

Of especially fundamental interest for a uniform, isotropic system is the radial distribution function $g(r)$, defined for a bulk fluid of density ρ such that the quantity $(4\pi r^2 dr)\rho g(r)$ is equal to the number of particles in a

spherical shell of thickness dr at radial distance r from a central particle [6,7]. Of more practical concern, since they are experimentally measurable (e.g., by neutron diffraction), are (i) the static structure factor $S(Q)$, which is directly related to $g(r)$ by Fourier transformation, according to

$$S(Q) = 1 + 4\pi\rho \int_0^\infty dr\, r^2 \left[\frac{\sin(Qr)}{Qr} \right] [g(r)-1] , \quad (1a)$$

and (ii) the isothermal compressibility κ_T, which from thermodynamic fluctuation theory is related to the $Q=0$ limit of $S(Q)$ according to

$$\frac{\kappa_T}{\kappa_T^0} = S(0) = 1 + 4\pi\rho \int_0^\infty dr\, r^2 [g(r)-1] , \quad (1b)$$

where κ_T^0 denotes the ideal-gas limit of κ_T.

The extraction of $S(Q)$ from a computer simulation is a longstanding problem. In principle, given $g(r)$ from a simulation, $S(Q)$ can be obtained from Eq. (1). In practice, however, for a cubic simulation cell of length L, $g(r)$ can be determined only over the limited range $r \lesssim L/2$. Furthermore, the finite-size effects mentioned above produce errors in $g(r)$ that are especially significant at long range in compressible systems. For Q sufficiently large, the integrand in Eq. (1a) damps to zero rapidly enough with increasing r as to render the limited range and errors in $g(r)$ negligible. As Q tends to zero, however, any inaccuracies in $g(r)$ become increasingly relevant and, for Q sufficiently small, may produce significant error in $S(Q)$. Thus the low-Q region can be particularly troublesome.

A variety of methods has been proposed and employed to determine the low-Q behavior of $S(Q)$. For example, Verlet [8] has analytically extended a simulated $g(r)$ into the range $r \gtrsim L/2$ by use of the approximate Percus-

*Present address: Institut für Theoretische Physik, Technische Universität, Wiedner Hauptstrasse 8-10, A-1040 Wien, Austria.

Yevick closure relation between $g(r)$ and the direct correlation function $c(r)$. Foiles, Ashcroft, and Reatto [9] and Reatto and Tau [10] have calculated $g(r)$ for arbitrarily large r by means of the very accurate modified hypernetted chain integral equation formalism. More recently, Rovere, Heermann, and Binder [11] have employed Monte Carlo simulation to compute the variance of the number of particles contained within subsystems of the simulation cell. By extrapolation to progressively larger subsystems, the compressibility, and hence [from Eq. (1b)] $S(0)$, of an infinite system can be deduced. Recently Stevenson *et al.* [12] have considered finite-size corrections in MD simulations of polymer blends.

The purpose of this paper is to propose an approach to calculating the static structure factor $S(Q)$ of a bulk fluid for arbitrarily small Q, *including* the $Q=0$ limit, from MD simulations of relatively small systems. The MD method is chosen because we wish to extend this work to time-dependent phenomena. Our approach is based on the defining connection between $S(Q)$ and $g(r)$ in Eq. (1), together with an analytic relation between the infinite- (or bulk-) system function $g(r)$ and its finite-system (N-particle) counterpart $g_N(r)$, which in the asymptotic (large-r) limit takes the well-known form [13]

$$g_N(r) \approx 1 - \frac{S(0)}{N} \text{ for large } r . \tag{2}$$

Equation (2) would appear to permit an estimation of $S(0)$ directly from the asymptotic behavior of $g_N(r)$. Unfortunately, the long-range statistical uncertainties in $g_N(r)$ tend to be comparable in magnitude to $[g_N(r)-1]$ itself, rendering this simple approach impractical. Consequently, we take another route and begin by defining two distribution functions

$$S(Q,R) = 1 + 4\pi\rho \int_0^R dr \, r^2 \left[\frac{\sin(Qr)}{Qr} \right] [g(r)-1] \tag{3a}$$

and

$$S_N(Q,R) = 1 + 4\pi\rho \int_0^R dr \, r^2 \left[\frac{\sin(Qr)}{Qr} \right] [g_N(r)-1] , \tag{3b}$$

where $S(Q,R) \to S(Q)$ as $R \to \infty$. In essence, we then proceed by first computing the finite-system function $S_N(Q,R)$ via simulation, next calculating the infinite-system function $S(Q,R)$ by analytically correcting for explicit finite-size effects, and finally obtaining the desired $S(Q)$ in the limit of large R. We first write down general formulas and then truncate them in a way we show in paper II to work for the states discussed there. A useful property of a MD calculation is that the appropriate approximations may be tested at intermediate stages en route to the final results.

As noted, we base our work upon the asymptotic behavior of $g_N(r)$ for large r, an approach that leads itself to an "r-space presentation" characterized by Eqs. (3a) and (3b). An alternate presentation of our method is obtained by defining $S_N(\mathbf{Q})$ in terms of $\langle |\sum_j \exp(i\mathbf{Q}\cdot\mathbf{r}_j)|^2 \rangle$ and defining $S_N(Q)$ as the average of $S_N(\mathbf{Q})$ over a representative set of $\mathbf{Q}$ vectors with magnitude $|\mathbf{Q}|=Q$.

However, for $Q < 2\pi/L$, L being the edge length of the simulation cube, there is no *unique* way of selecting a representative set of $\mathbf{Q}$ vectors, a fact that modestly complicates this type of presentation. In addition, calculating $S_N(Q,R)$ by Eq. (3b) is equivalent to averaging (integrating) over all $\mathbf{Q}$, with $|\mathbf{Q}|=Q$, and thus is more general than methods based upon averaging over a representative set of $\mathbf{Q}$ vectors. Hence, in order to clarify the description of our method, presented in this paper, and to simplify the calculation technique presented in paper II, we have elected to use an r-space presentation.

The remainder of this paper is organized as follows. In Sec. II we review the theory of explicit size effects and formulate more explicitly our method of computing $S(Q)$. We also develop the time-dependent generalization of the method for the purpose of future applications to the intermediate scattering function $I(Q,t)$. In Sec. III we discuss the approximations on which our method depends and practical restrictions on the size of system to which it may be applied. We also examine the relative magnitude of implicit size effects, presenting MD simulation data to support the neglect of these effects. Finally, in Sec. IV we summarize our results.

II. THEORETICAL METHOD

A. Static structure factor and compressibility

As mentioned in the Introduction, our method of computing the static structure factor $S(Q)$ involves first computing the distribution function $S_N(Q,R)$, defined by Eq. (3b). For practical purposes, we find it convenient to rewrite this equation in the form

$$S_N(Q,R) = \langle N(Q,R) \rangle - \bar{N}(Q,R) , \tag{4a}$$

where

$$\langle N(Q,R) \rangle \equiv 1 + 4\pi\rho \int_0^R dr \, r^2 \left[\frac{\sin(Qr)}{Qr} \right] g_N(r) \tag{4b}$$

and

$$\bar{N}(Q,R) \equiv 4\pi\rho \int_0^R dr \, r^2 \left[\frac{\sin(Qr)}{Qr} \right] = \frac{4}{3}\pi\rho R^3 u(QR) , \tag{4c}$$

with

$$u(x) \equiv \frac{3}{x^3}(\sin x - x \cos x) = \frac{3}{x} j_1(x) , \tag{4d}$$

$j_1(x)$ being the first spherical Bessel function. The choice of notation in Eq. (4a) is suggested by the $Q=0$ limit, in which that relation reduces to

$$S_N(0,R) = \langle N(R) \rangle - \bar{N}(R) , \tag{5a}$$

where

$$\langle N(R) \rangle \equiv 1 + 4\pi\rho \int_0^R dr \, r^2 g_N(r) \tag{5b}$$

may be physically interpreted as the average number of particles contained within a sphere of radius R *centered*

on a particle and

$$\bar{N}(R) \equiv 4\pi\rho \int_0^R dr \, r^2 = \tfrac{4}{3}\pi\rho R^3 \tag{5c}$$

is the average number of particles within a sphere of radius R, irrespective of the location of the sphere's center. Thus, given MD simulation data for the coordinates of the particles, the function $S_N(0,R)$ may be calculated by the simple procedure of counting particles within spheres of radius R, each sphere centered on a particle. Similarly, $S_N(Q,R)$ may be calculated by the same counting procedure, but with each count *weighted* by the factor $[\sin(Qr)/Qr]$, r being the radial distance from the center of the sphere of the particle being counted. Explicit formulas for implementing this procedure are given in paper II.

An equation relating the infinite-system function $S(Q,R)$ [Eq. (3a)] to its N-particle counterpart $S_N(Q,R)$ [Eq. (3b)] may be derived from a corresponding relation between $g(r)$ and $g_N(r)$. Such a relation may be derived, following Lebowitz and Percus [2] by means of a Taylor-series expansion in powers of $1/N$. Here we extend the derivation to the coefficient of the $O(1/N^2)$ term. In the grand canonical ensemble, the pair distribution function (or two-particle density) may be expressed as

$$\rho^{(2)}(\mathbf{r}_1,\mathbf{r}_2) = \sum_{N=0}^{\infty} P(N)\rho^{(2)}(\mathbf{r}_1,\mathbf{r}_2;N) , \tag{6}$$

where $P(N)$ is the probability that in equilibrium the system contains N particles and $\rho^{(2)}(\mathbf{r}_1,\mathbf{r}_2;N)$ is the pair distribution function for a system of N particles. Expanding $\rho^{(2)}(\mathbf{r}_1,\mathbf{r}_2;N)$ about the average number of particles

$$\bar{N} = \sum_{N=0}^{\infty} P(N)N , \tag{7}$$

we have

$$\rho^{(2)}(\mathbf{r}_1,\mathbf{r}_2) = \sum_{N=0}^{\infty} P(N) \left[\rho^{(2)}(\mathbf{r}_1,\mathbf{r}_2;\bar{N}) + (N-\bar{N})\frac{\partial}{\partial \bar{N}}\rho^{(2)}(\mathbf{r}_1,\mathbf{r}_2;\bar{N}) \right.$$
$$\left. + \frac{1}{2}(N-\bar{N})^2\frac{\partial^2}{\partial \bar{N}^2}\rho^{(2)}(\mathbf{r}_1,\mathbf{r}_2;\bar{N}) + \frac{1}{6}(N-\bar{N})^3\frac{\partial^3}{\partial \bar{N}^3}\rho^{(2)}(\mathbf{r}_1,\mathbf{r}_2;\bar{N}) + \cdots \right] . \tag{8}$$

Using the normalization condition

$$\sum_{N=0}^{\infty} P(N) = 1 \tag{9}$$

and the identity

$$\sum_{N=0}^{\infty} P(N)(N-\bar{N}) = 0 , \tag{10}$$

Eq. (8) may be written in the form

$$\rho^{(2)}(\mathbf{r}_1,\mathbf{r}_2) = \rho^{(2)}(\mathbf{r}_1,\mathbf{r}_2;\bar{N})$$
$$+ \frac{1}{2}\overline{(N-\bar{N})^2}\frac{\partial^2}{\partial \bar{N}^2}\rho^{(2)}(\mathbf{r}_1,\mathbf{r}_2;\bar{N})$$
$$+ \frac{1}{6}\overline{(N-\bar{N})^3}\frac{\partial^3}{\partial \bar{N}^3}\rho^{(2)}(\mathbf{r}_1,\mathbf{r}_2;\bar{N}) + \cdots . \tag{11}$$

Now, in the uniform limit $[\rho(\mathbf{r}) \to \rho \equiv \bar{N}/V]$, where $\rho^{(2)}(\mathbf{r}_1,\mathbf{r}_2) \to \rho^2 g(r)$ and $\rho^{(2)}(\mathbf{r}_1,\mathbf{r}_2;\bar{N}) \to \rho^2 g_{\bar{N}}(r)$, Eq. (11) reduces to

$$g(r) = g_{\bar{N}}(r) + \frac{1}{2\bar{N}}\frac{\overline{(N-\bar{N})^2}}{\bar{N}}\frac{\partial^2}{\partial \rho^2}[\rho^2 g_{\bar{N}}(r)]$$
$$+ \frac{\rho}{6\bar{N}^2}\frac{\overline{(N-\bar{N})^3}}{\bar{N}}\frac{\partial^3}{\partial \rho^3}[\rho^2 g_{\bar{N}}(r)] + \cdots . \tag{12}$$

The coefficients in Eq. (12) may be related to the compressibility via fluctuation theory, as follows. From the definition of the grand partition function

$$\Xi = \sum_{N=0}^{\infty} \frac{e^{\beta\mu N}}{N!}Z_N , \tag{13}$$

where $\beta \equiv 1/k_B T$, μ is the chemical potential, and Z_N is the canonical partition function for an N-particle system, the N-particle probability may be expressed as

$$P(N) = \Xi^{-1}\frac{e^{\beta\mu N}}{N!}Z_N . \tag{14}$$

The average number of particles, and the mean-squared and mean-cubed deviations in the number of particles are given by

$$\bar{N} = \Xi^{-1}\sum_{N=0}^{\infty}\frac{N}{N!}e^{\beta\mu N}Z_N = \frac{\partial}{\partial(\beta\mu)}\ln\Xi , \tag{15}$$

$$\overline{(N-\bar{N})^2} = \frac{\partial\bar{N}}{\partial(\beta\mu)} = \bar{N}S(0) , \tag{16}$$

$$\overline{(N-\bar{N})^3} = \frac{\partial^2\bar{N}}{\partial(\beta\mu)^2} = \bar{N}\left[S(0)^2 + \frac{\partial S(0)}{\partial(\beta\mu)}\right] . \tag{17}$$

Finally, substituting Eqs. (16) and (17) into Eq. (12) and for simplicity relabeling $\bar{N}$ with N, we obtain the desired relation

$$g(r) = g_N(r) + \frac{c_1}{N} + \frac{c_2}{N^2} + O(1/N^3) , \tag{18}$$

where the coefficients c_1 and c_2 are given by

$$c_1 = \frac{S(0)}{2} \cdot \frac{\partial^2}{\partial \rho^2} [\rho^2 g(r)] \tag{19a}$$

and

$$c_2 = \frac{\rho}{6} \left[S(0)^2 + \frac{\partial S(0)}{\partial(\beta\mu)} \right] \frac{\partial^3}{\partial \rho^3} [\rho^2 g(r)]$$

$$- \frac{S(0)}{4} \frac{\partial^2}{\partial \rho^2} \left\{ \rho^2 S(0) \frac{\partial^2}{\partial \rho^2} [\rho^2 g(r)] \right\} . \tag{19b}$$

Note that in the asymptotic limit $(r \to \infty)$, where $g(r) \to 1$, Eq. (18) correctly reduces to the asymptotic Eq. (2) to $O(1/N)$ and c_2 simplifies to

$$c_2 \approx \frac{S(0)}{2} \frac{\partial^2}{\partial \rho^2} [\rho^2 S(0)] . \tag{19c}$$

Now substituting Eq. (18) into Eq. (3) and using the identity

$$\frac{\partial^2}{\partial \rho^2} [\rho^2 g(r)] = 2 + \frac{\partial^2}{\partial \rho^2} \{\rho^2 [g(r) - 1]\} , \tag{20}$$

we obtain the general result

$$S(Q,R) = S_N(Q,R) + \rho \frac{S(0)}{N} \left\{ \frac{4}{3}\pi R^3 u(QR) \right.$$

$$\left. + \frac{1}{2} \frac{\partial^2}{\partial \rho^2} [\rho S(Q,R)] \right\} + O(1/N^2) . \tag{21}$$

Clearly, because of the density derivative term, a precise solution of Eq. (21) by MD simulation is possible only through simulations at several densities. As demonstrated in paper II, however, the latter term is relatively small for the fluid states of interest, the more so as R increases. Thus, neglecting terms of $O(1/N^2)$ and higher, an approximation for $S(Q,R)$, which improves in accuracy with increasing R, is

$$S(Q,R) \simeq S_N(Q,R) + \frac{S(0)}{N} \frac{4}{3}\pi\rho R^3 u(QR) . \tag{22}$$

Now for a sufficiently large system, R may be taken large enough that to a good approximation $S(Q,R) \cong S(Q)$ (see paper II). When this approximation is valid at $Q=0$, Eq. (22) may be solved explicitly for $S(0)$, yielding the large-R approximation

$$S(0) \cong \frac{S_N(0,R)}{1 - \frac{1}{N} \frac{4}{3}\pi\rho R^3} . \tag{23}$$

Equation (23) provides a practical means of predicting the infinite-system quantity $S(0)$ from MD data for the N-particle function $S_N(0,R)$, as we demonstrate in paper II. The resulting $S(0)$ may subsequently be substituted into Eq. (22) to determine $S(Q,R)$, and hence $S(Q)$ for large R, from MD data for $S_N(Q,R)$.

We also briefly mention an alternative means by which $S(0)$ may, in principle, be obtained from MD data computed for two different system sizes. Writing Eq. (22) for two particle numbers, say, N_1 and N_2, the N-independent quantity $S(Q,R)$ may be eliminated to yield the "difference formula"

$$S(0) = \frac{S_{N_1}(Q,R) - S_{N_2}(Q,R)}{\frac{4}{3}\pi\rho R^3 \left[\frac{1}{N_2} - \frac{1}{N_1} \right] u(QR)} , \tag{24}$$

which should give a reasonable Q-independent approximation for $S(0)$ at sufficiently large R. In practice of course, N_1 and N_2 should be sufficiently different to minimize errors associated with the subtraction of two comparable numbers. Furthermore, Eq. (24) should be valid, for a given R, only for Q smaller than the first zero of $u(QR)$ because of our neglect of the density-derivative term in Eq. (21).

It should be noted that, in principle, the above method of correcting for finite-size effects applies only to simulation data collected in the canonical ensemble. This is clear from our derivation of the $O(1/N)$ correction for $g_N(r)$ [Eqs. (8)–(12)], in which we consider only particle number fluctuations. Corrections for other ensembles would require considering fluctuations of other extensive variables, e.g., total energy for the microcanonical ensemble, or both total energy and linear momentum for the molecular dynamics ensemble [13–16]. In practice, however, as we demonstrate in the Appendix, fluctuations in these other variables are of $O(1/N^2)$ and thus usually can be ignored.

B. Intermediate scattering function

The method described above for determining the static structure factor $S(Q)$ may easily be generalized to the time-dependent intermediate scattering function [17] $I(Q,t)$. The latter is defined for a uniform, bulk system by

$$I(Q,t) \equiv 4\pi \int_0^\infty dr\, r^2 \left[\frac{\sin(Qr)}{Qr} \right] [G(r,t) - \rho] , \tag{25}$$

where

$$G(\mathbf{r},t) \equiv \frac{\langle \rho(\mathbf{r},t)\rho(0,0) \rangle}{\rho} \tag{26}$$

is the van Hove space-time distribution function [7,13,18], which has the physical interpretation that $G(\mathbf{r},t)d\mathbf{r}$ is proportional to the probability of finding a particle within a volume element $d\mathbf{r}$ at position $\mathbf{r}$ and time t, given that there was an arbitrary reference particle at $\mathbf{r}=0$ and $t=0$. Note that $I(Q,0) = S(Q)$, which follows from Eqs. (1a) and (25), since $G(\mathbf{r},0) = \delta(\mathbf{r}) + \rho g(\mathbf{r})$.

A related distribution function $I(Q,t,R)$ may be defined by limiting the integral in Eq. (25) to a sphere of radius R, such that

$$I(Q,t,R) \equiv 4\pi \int_0^R dr\, r^2 \left[\frac{\sin(Qr)}{Qr} \right] [G(r,t) - \rho] . \tag{27a}$$

The corresponding N-particle distribution function is then defined as

$$I_N(Q,t,R) \equiv 4\pi \int_0^R dr\, r^2 \left[\frac{\sin(Qr)}{Qr} \right] [G_N(r,t)-\rho] \;,$$

(27b)

where $G_N(r,t)$ is the finite-system (N-particle) counterpart of $G(r,t)$. Note that $I_N(Q,0,R)=S_N(Q,R)$, which follows from Eqs. (3b) and (27). By analogy with Eq. (4), we may write

$$I_N(Q,t,R) = \langle N(Q,t,R) \rangle - \overline{N}(Q,R) \;,$$

(28a)

where

$$\langle N(Q,t,R) \rangle \equiv 4\pi \int_0^R dr\, r^2 \left[\frac{\sin(Qr)}{Qr} \right] G_N(r,t) \quad (28b)$$

is the time-dependent generalization of $\langle N(Q,R) \rangle$, with $\langle N(Q,0,R) \rangle = \langle N(Q,R) \rangle$.

As in the static case, a formula relating the infinite-system function $I(Q,t)$ and the N-particle function $I_N(Q,t)$ may be obtained by first establishing a connection between $G(r,t)$ and $G_N(r,t)$. To this end, the previous derivation [Eqs. (6)–(18)] leading to Eq. (18) may be generalized simply by replacing the pair distribution function $\rho^{(2)}(\mathbf{r}_1,\mathbf{r}_2)$ by its time-dependent analog $\rho^{(2)}(\mathbf{r}_1,\mathbf{r}_2;t)$. Thus Eq. (8) becomes

$$\rho^{(2)}(\mathbf{r}_1,\mathbf{r}_2;t) = \sum_{N=0}^{\infty} P(N) \left[\rho^{(2)}(\mathbf{r}_1,\mathbf{r}_2;t\,;\overline{N}) + (N-\overline{N})\frac{\partial}{\partial\overline{N}}\rho^{(2)}(\mathbf{r}_1,\mathbf{r}_2;t\,;\overline{N}) + \frac{1}{2}(N-\overline{N})^2\frac{\partial^2}{\partial\overline{N}^2}\rho^{(2)}(\mathbf{r}_1,\mathbf{r}_2;t\,;\overline{N}) + \cdots \right] \;,$$

(29)

which leads, by the same sequence of steps, to the result

$$G(r,t) = G_N(r,t) + \frac{1}{N}\frac{S(0)}{2}\rho\frac{\partial^2}{\partial\rho^2}[\rho G(r,t)] + O(1/N^2) \;.$$

(30)

Note that in the asymptotic limits of large r, where $G(r,t) \to \rho$, Eq. (30) reduces to

$$G_N(r,t) \approx \rho \left[1 - \frac{S(0)}{N} \right] \quad \text{for large } r \;,$$

(31)

which prescribes the same size correction as in the static case [cf. Eq. (2)].

Now substituting Eq. (30) into Eqs. (27) and using the identity

$$\frac{\partial^2}{\partial\rho^2}[\rho G(r,t)] = 2 + \frac{\partial^2}{\partial\rho^2}\{\rho[G(r,t)-\rho]\} \;,$$

(32)

we obtain the desired relation

$$I(Q,t,R) = I_N(Q,t,R) + \rho\frac{S(0)}{N}\left\{ \frac{4}{3}\pi R^3 u(QR) \right.$$

$$\left. + \frac{1}{2}\frac{\partial^2}{\partial\rho^2}[\rho I(Q,t,R)] \right\}$$

$$+ O(1/N^2) \;,$$

(33)

which may be compared with Eq. (21) (to which it reduces at $t=0$). As in the static case, a precise solution of Eq. (33) is complicated by the presence of the density derivative term. Although we have not numerically determined the magnitude of this term, we expect it to be even smaller than the corresponding term in Eq. (21) since $I(Q,t,R)$ is usually a decreasing function of time, such that $I(Q,t,R) < I(Q,0,R) = S(Q,R)$. It does, how-

ever, contain the time dependence of the correction. Thus, for sufficiently large R and assuming weak time-dependent corrections, we can neglect this term and approximate $I(Q,t,R)$ to $O(1/N)$ by

$$I(Q,t,R) \cong I_N(Q,t,R) + \frac{S(0)}{N}\frac{4}{3}\pi\rho R^3 u(QR) \;.$$

(34)

Given simulation data for $I_N(Q,t,R)$, Eq. (34) dictates the size correction required to obtain $I(Q,t,R)$ and hence the intermediate scattering function $I(Q,t)$ at large R. It is expected to be satisfactory for the states discussed in paper II. However, care should be taken over the time-dependent corrections when calculating time Fourier transforms.

III. DISCUSSION

A. Approximations and limitations of the method

The main results of this paper are expressed by Eqs. (22), (23), and (34), which prescribe analytic corrections for explicit size effects to simulation data for the static structure factor $S(Q)$ and for the intermediate scattering function $I(Q,t)$. In deriving these results, we have made three main approximations. The first is the truncation of the Taylor-series expansion for $g(r)$, in which terms of $O(1/N^2)$ and higher were dropped. For systems with N of order 100 or more we expect this to be an excellent approximation as long as the coefficients of the higher-order terms remain of order unity or smaller. A more careful test would require explicit evaluation of at least the second-order coefficient c_2. Although even c_2 is a rather complicated function [Eq. (19b)], its asymptotic form [Eq. (19c)] does suggest that the neglect of higher-order terms is valid as long as the magnitude of $S(0)$ and the

curvature (with respect to ρ) of $\rho^2 S(0)$ are not unusually large, which should be true of states away from the critical point. In practice, the approximation can be tested by applying the method at two distinct values of N and ensuring that the results are independent of N.

Our second approximation is the neglect of the density derivative terms in Eqs. (21) and (33). Evidently, this should be reasonable at large R as long as the curvature (with respect to ρ) of $\rho S(Q,R)$ or $\rho I(Q,t,R)$ remains sufficiently small, which again precludes states near the critical point. In paper II [1] we describe a numerical test that supports this approximation. Our third approximation is the neglect of implicit size effects, whose relative importance we next discuss.

B. Implicit size effects

As noted in the Introduction, implicit size effects, linked directly to periodic boundary conditions, can influence pair correlations, especially for very small systems [3] (e.g., $N \leq 8$). As a result, the pair distribution function is in principle *anisotropic* and must therefore depend not only on radial distance r but also on angular variables, i.e., $g_N(\mathbf{r})=g_N(r,\theta,\phi)$, where θ and ϕ are the usual polar and azimuthal angles in spherical coordinates.

To test for isotropy in our MD simulations (described in paper II), we have analyzed our data to explicitly compute the angular variation of $g_N(\mathbf{r})$. As Fig. 1 illustrates, the partially averaged functions

$$g_N(r,\theta)=\frac{1}{2\pi}\int_0^{2\pi}d\phi\, g_N(r,\theta,\phi) \qquad (35a)$$

and

$$g_N(r,\phi)=\frac{1}{2}\int_0^{\pi}d\theta\sin(\theta)g_N(r,\theta,\phi) \qquad (35b)$$

show no significant angular variation compared with random statistical errors of about 1%. We expect that when Eq. (35a) is averaged over θ, or Eq. (35b) over ϕ, to produce $g_N(r)$, the statistical errors will fall to $\sim 0.1\%$ of $g_N(r)$.

The question remains, however, whether the periodic boundary conditions might still systematically alter the *magnitude* of $g_N(r)$. This issue can be addressed within the framework of a theory of implicit size effects proposed by Pratt and Haan [5]. Their theory is based on the observation that an infinite ensemble of periodically replicated simulation cells is equivalent to an infinite system of "supermolecules," each of which comprises a single physical atom from the primary cell together with all of its periodic images. Starting from this insight and working in the grand canonical ensemble to avoid explicit size effects, these authors develop a formally exact cluster expansion for $\ln g_N(\mathbf{r})$, expressed in terms of spring bond graphs. By summing exactly a certain class of "unbridged" graphs, which depend on only the infinite-system function $g(r)$, but neglecting entirely another class of "bridged" graphs, they derive an approximate relation between $g_N(\mathbf{r})$ and $g(r)$, which may be written in the concise form

$$g_N\mathbf{r}_{12}\simeq g(r_{12})\prod_i g(|\mathbf{r}_1-\mathbf{r}_{2i}|)\,, \qquad (36)$$

where $\mathbf{r}_{12}\equiv\mathbf{r}_1-\mathbf{r}_2$, $r_{12}\equiv|r_{12}|$, and the product index runs over all periodic images of particle 2. Since the neglected graphs are more highly connected than those included, the approximation is expected to be accurate for sufficiently large system sizes, low densities, and short-ranged interactions.

We have tested the accuracy of Pratt and Haan's theory by using Eq. (36) to numerically study the predicted N dependence of the angular variation of $g_N(\mathbf{r})$ and comparing it with MD data for relatively small systems ($N \simeq 100$) and high densities (near the triple point), where implicit size effects are expected to be strongest and perhaps comparable to explicit size effects. In general, we find the theoretical predictions to compare quantitatively well with corresponding MD data, inspiring
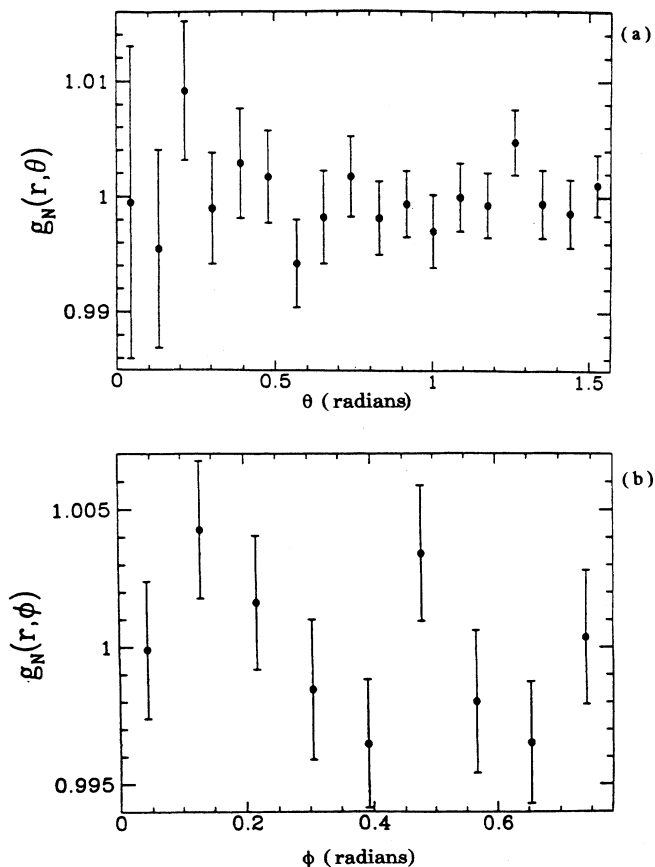
FIG. 1. MD simulation data showing the variation of the partially averaged pair distribution function [Eq. (35)] with respect to (a) polar angle θ and (b) azimuthal angle ϕ. The simulation was performed for a model Aziz potential krypton fluid of $N=706$ atoms at a thermodynamic state defined by reduced density $\rho^*=0.4$ and reduced temperature $T^*=1.51$ (see paper II for details). The data shown are for a radial distance $r=5$ in units of σ_m, where σ_m is the distance of the Aziz pair potential minimum.

confidence in the accuracy of the theory. Details of this study will be presented in a future presentation [19].

For the present purposes, however, it suffices to consider the influence of implicit size effects on $S(Q)$ as determined by the approach described in Sec. II. To this end, we first emphasize that our approach is in practice restricted to system sizes (N) and thermodynamic states for which $g_N(\mathbf{r})$ has essentially reached its asymptotic limit by $r \simeq L/2$. This ensures that $S(0,R)$ has "saturated" (i.e., reached its asymptotic limit with increasing R) by $R \simeq L/2$, thus permitting an accurate estimate of $S(0)$ [by Eq. (23)] and hence of $S(Q)$ [by Eq. (22)]. As shown in paper II, this condition is satisfied by our simulated systems. In general, however, the nearer the state to the critical point, the longer the range of pair correlations and hence the larger the system that must be simulated.

Now, if we consider any two particles in the primary simulation cell separated by a distance $r_{12} \simeq L/2$, then the *shortest* distance between one of the particles and any periodic image of the other can be no less than $L/2$, i.e., $\min\{|\mathbf{r}_1 - \mathbf{r}_{2i}|\} \geq L/2$. Therefore, in the product over images in Eq. (36) the function $g(|\mathbf{r}_1 - \mathbf{r}_{2i}|)$ is evaluated always at a distance that lies within the asymptotic tail of the function. If $g(r)$ decays sufficiently rapidly with increasing r that $g(|\mathbf{r}_1 - \mathbf{r}_{2i}|) \cong 1$, then Eq. (36) reduces to the statement that $g_N(\mathbf{r}_{12}) \simeq g(r_{12})$, which would justify the neglect of implicit size effects in the application of our method. For continuous potentials, however, the asymptotic tail of $g(r)$ is known to be proportional to the pair potential [20]. Thus, for a Lennard-Jones or similarly short-ranged pair potential we would expect $g_N(\mathbf{r}_{12}) \cong g(r_{12})$ to a very good approximation in our simulated systems.

IV. SUMMARY

Although in infinite systems the radial distribution function $g(r) \rightarrow 1$ as $r \rightarrow \infty$, in finite systems (e.g., those modeled by computer simulation) finite-size effects may produce significant relative errors in the asymptotic region of the N-particle function $g_N(r)$. These in turn lead to errors in the Fourier transform [i.e., the N-particle static structure factor $S_N(Q)$], which grow as $Q \rightarrow 0$.

In this paper we have presented a general method of correcting for explicit size effects to compute the bulk static structure factor $S(Q)$ for *all* Q (including $Q=0$)

from simulation data for a fixed number of particles. The practical form of the method is summarized by Eqs. (22), (23), and (34). In practice, the method relies on three main approximations: (i) truncation to first order of a Taylor-series expansion (in powers of $1/N$) of $g_N(r)$ or, equivalently, $S_N(Q)$; (ii) an approximation for the coefficient of the first-order ($1/N$) term in this series; and (iii) neglect of implicit size effects due to periodic boundary conditions. The first two approximations are expected to be valid for thermodynamic states away from the critical point and for system sizes typical of commonly simulated atomic systems (usually several hundred or more particles). As for implicit size effects, we have considered their role and concluded that for the systems considered in the following paper they may be safely ignored.

For future applications we have also presented the time-dependent generalization of the method to the intermediate scattering function $I(Q,t)$. In the following paper, however, we confine ourselves to the static case and proceed to test and demonstrate the method by applying it to a calculation of $S(Q)$ from extensive MD simulation data for a model krypton fluid.

ACKNOWLEDGMENTS

APPENDIX: OTHER ENSEMBLE CORRECTIONS

Here we demonstrate that the finite-size (or ensemble) correction associated with fixing the total energy E (as in the microcanonical ensemble), or additionally the total linear momentum (as in the molecular dynamics ensemble), affects the radial distribution function only to $O(1/N^2)$ and that, for sufficiently large N, this correction therefore may be ignored relative to the $O(1/N)$ correction arising from fixing the particle number N.

Consider first the connection between the pair distribution function in the canonical ensemble (fixed N) $\rho_N^{(2)}(r_1,r_2)$ and the corresponding function in the microcanonical ensemble (fixed N and E) $\rho_N^{(2)}(r_1,r_2;E)$. By analogy with Eq. (6), we can expand the former in a Taylor series about the mean energy $\bar{E}$:

$$\rho_N^{(2)}(r_1,r_2)= \int_0^\infty dE\, n(E)P_N(E)\left[\rho_N^{(2)}(r_1,r_2;\bar{E})+(E-\bar{E})\frac{\partial}{\partial\bar{E}}\rho_N^{(2)}(r_1,r_2;\bar{E})+\frac{1}{2}(E-\bar{E})^2\frac{\partial^2}{\partial\bar{E}^2}\rho_N^{(2)}(r_1,r_2;\bar{E})+ \cdots \right] ,$$

$$\text{(A1)}$$

where $n(E)$ is the density of states and $dE\, n(E)P_N(E)$ is the probability of the system having N particles in an energy range between E and $E+dE$. Now, since

$$\bar{E} = \int_0^\infty dE\, n(E)P_N(E)E , \qquad \text{(A2)}$$

we may write

$$\rho_N^{(2)}(r_1,r_2)=\rho_N^{(2)}(r_1,r_2;\bar{E})$$
$$+\frac{1}{2}\overline{(E-\bar{E})^2}\frac{\partial^2}{\partial\bar{E}^2}\rho_N^{(2)}(r_1,r_2;\bar{E})+ \cdots .$$
$$\text{(A3)}$$

But since N is here fixed, then

$$\rho_N^{(2)}(r_1,r_2)=\rho_N^{(2)}(r_1,r_2;\bar{E})$$
$$+\frac{1}{N}\left[\frac{c_v kT^2}{2}\right]\frac{\partial^2}{\partial\epsilon^2}\rho_N^{(2)}(r_1,r_2;\bar{E})+\cdots,$$
(A4)

where $\epsilon\equiv\bar{E}/N$ is the mean energy per particle and we have used the fluctuation-dissipation relation

$$\overline{(E=\bar{E})^2}=Nc_v kT^2,$$
(A5)

c_v being the specific heat.

In the limit of a uniform system, the canonical-ensemble radial distribution function is then related to its microcanonical counterpart according to

$$g_N(r)=g_N(r;\bar{E})+\frac{1}{N}\left[\frac{c_v kT^2}{2}\right]\frac{\partial^2}{\partial\epsilon^2}g_N(r)+\cdots.$$
(A6)

In the crucial asymptotic regime, however, where

$$g_N(r)\simeq 1-\frac{S(0)}{N},$$
(A7)

we have

$$g_N(r)=g_N(r;E)-\frac{1}{N^2}\left[\frac{c_v kT^2}{2}\right]\frac{\partial^2}{\partial\epsilon^2}S(0)+\cdots.$$
(A8)

Thus the leading-order finite-size correction to the radial distribution function associated with fixing the total energy is only of $O(1/N^2)$. Furthermore, by a completely analogous derivation, it is easily shown that fixing the total linear momentum also results in a leading-order correction of $O(1/N^2)$. These corrections are, however, $O(1/N)$ smaller than the $O(1/N)$ correction arising from fixing the particle number N (Sec. II A). We conclude, therefore, that the correction required to convert simulation data from the molecular dynamics ensemble to the canonical ensemble is, in practice, negligible (for sufficiently large N) compared with the correction involved in converting further from the canonical to the grand-canonical ensemble.

---

[1] J. J. Salacuse, A. R. Denton, P. A. Egelstaff, M. Tau, and L. Reatto, following paper, Phys. Rev. E **53**, 2390 (1996).

[2] J. L. Lebowitz and J. K. Percus, Phys. Rev. **122**, 1675 (1961).

[3] J. L. Lebowitz and J. K. Percus, Phys. Rev. **124**, 1673 (1961).

[4] M. J. Mandell, J. Stat. Phys. **15**, 299 (1976).

[5] L. R. Pratt and S. W. Haan, J. Chem. Phys. **74**, 1864 (1981).

[6] J. K. Percus, Phys. Rev. Lett. **8**, 462 (1962).

[7] P. A. Egelstaff, *An Introduction to the Liquid State*, 2nd ed. (Oxford University Press, New York, 1992).

[8] L. Verlet, Phys. Rev. **159**, 98 (1967); **165**, 201 (1968).

[9] S. M. Foiles, N. W. Ashcroft, and L. Reatto, J. Chem. Phys. **81**, 6140 (1984).

[10] L. Reatto and M. Tau, J. Chem. Phys. **86**, 6174 (1987).

[11] M. Rovere, D. W. Heermann, and K. Binder, Europhys. Lett. **6**, 585 (1988); J. Phys. Condens. Matter **2**, 7009 (1990).

[12] C. S. Stevenson, J. D. McCoy, S. J. Plimpton, and J. G. Curro, J. Chem. Phys. (to be published).

[13] T. L. Hill, *Statistical Mechanics* (McGraw-Hill, New York, 1956).

[14] W. W. Wood, in *Fundamental Problems in Statistical Mechanics*, edited by E. G. D. Cohen (North-Holland, New York, 1975).

[15] F. Lado, J. Chem. Phys. **75**, 5461 (1981).

[16] D. C. Wallace and G. K. Straub, Phys. Rev. A **27**, 2201 (1983).

[17] J. J. Salacuse, W. Schommers, and P. A. Egelstaff, Phys. Rev. A **34**, 1516 (1986).

[18] L. van Hove, Phys. Rev. **95**, 249 (1954).

[19] A. R. Denton and P. A. Egelstaff (unpublished).

[20] J. E. Enderby, T. Gaskell, and N. H. March, Proc. Phys. Soc. London **85**, 217 (1965).