

RESEARCH ARTICLE | NOVEMBER 17 2023

Measuring many-body distribution functions in fluids using test-particle insertion **EP** **FREE**

Special Collection: **JCP Editors' Choice 2023**

Adam Edward Stones  ; Dirk G. A. L. Aarts  




J. Chem. Phys. 159, 194502 (2023)


<https://doi.org/10.1063/5.0172664>



CrossMark




Lock-in Amplifier



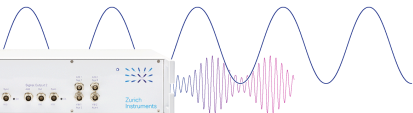
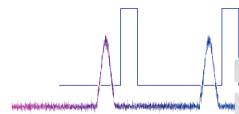
**Zurich
Instruments**

[Find out more](#)



Boxcar Averager

Boost Your Optics and
Photonics Measurements

Measuring many-body distribution functions in fluids using test-particle insertion

Cite as: J. Chem. Phys. 159, 194502 (2023); doi: 10.1063/5.0172664

Submitted: 17 August 2023 • Accepted: 19 October 2023 •

Published Online: 17 November 2023



Adam Edward Stones  and Dirk G. A. L. Aarts ^{a)} 

AFFILIATIONS

Department of Chemistry, Physical and Theoretical Chemistry Laboratory, University of Oxford, Oxford OX1 3QZ, United Kingdom

^{a)} Author to whom correspondence should be addressed: dirk.aarts@chem.ox.ac.uk

ABSTRACT

We derive a hierarchy of equations, which allow a general n -body distribution function to be measured by test-particle insertion of between 1 and n particles. We apply it to measure the pair and three-body distribution functions in a simple fluid using snapshots from Monte Carlo simulations in the grand canonical ensemble. The resulting distribution functions obtained from insertion methods are compared with the conventional distance-histogram method: the insertion approach is shown to overcome the drawbacks of the histogram method, offering enhanced structural resolution and a more straightforward normalization. At high particle densities, the insertion method starts breaking down, which can be delayed by utilizing the underlying hierarchical structure of the insertion method. Our method will be especially useful in characterizing the structure of inhomogeneous fluids and investigating closure approximations in liquid state theory.

Published under an exclusive license by AIP Publishing. <https://doi.org/10.1063/5.0172664>

I. INTRODUCTION

The distribution functions $g^{(n)}$ are central to the statistical mechanical description of classical fluids, characterizing the short-ranged order of their particles' positions.¹ It is widely appreciated that the knowledge of the pair distribution function $g^{(2)}$, where $n = 2$, provides access to the thermodynamics of a fluid by integration, yielding the compressibility and, for pairwise-additive systems, the pressure and energy.^{1–3} While this is correct, even simple monatomic liquids, such as argon, cannot be described adequately with pairwise interactions alone,^{4–8} and fully incorporating many-body interactions requires the knowledge of higher-order distribution functions, where $n \geq 3$.⁹ Moreover, theoretical calculations of $g^{(2)}$ require consideration of many-body correlations even in the additive case,^{10,11} and expressions for higher-order distribution functions in terms of lower ones can be used to close the Yvon–Born–Green (YBG) hierarchy in integral equation theories.^{11–18} Properties depending on thermodynamic derivatives, such as the heat capacity, are expressed as integrals over many-body distribution functions,^{9,11,19} as is the configurational entropy.^{20–22} Finally, a description of the fluid structure at the pairwise level alone is inherently limited, with the effect of many-body correlations increasingly important with increasing density: simulations^{9,23} and colloidal studies^{24–26} have found that the triplet distribution

function $g^{(3)}$ shows clear signatures of crystal-like order even in the fluid phase. The experimental work further shows that the signatures appear in $g^{(3)}$ at lower densities than in $g^{(2)}$. Disruption of this incipient order by impurities²⁶ or locally preferred fivefold structures^{27–31} may play an important role in glass formation.

Previous measurements of many-body distribution functions using particle coordinates from simulations^{23,32–39} and colloidal experiments^{24–26} have used the conventional distance-histogram approach. While this method usually provides adequate resolution for $g^{(2)}$ in homogeneous fluids, which only depends on the interparticle separation r , for many-body distribution functions, its shortcomings are increasingly pronounced. Even in the homogeneous case, $g^{(3)}$ depends on three variables, and doubling the spatial resolution reduces the bin volumes by a factor of 8, significantly increasing the statistical noise. For configurations where the particles approach closely, which are especially important in evaluating closure approximations,^{9,36–42} the distribution functions vary rapidly and approximate interpolations must be performed—in this work, we find both linear and cubic interpolations to be inadequate for $g^{(3)}$. Normalization of the histograms is also increasingly complex: e.g., for $g^{(3)}$, the interparticle separations must satisfy triangle inequalities, which intersect the bins used in a naive distance-histogram scheme.^{34,38}

In this article, we address these problems by deriving a hierarchy of equations for measuring $g^{(n)}$ by test-particle insertion.^{43,44} The hierarchy allows a given $g^{(n)}$ to be obtained by multiple routes, corresponding to insertion of between 1 and n test particles. For insertion of n or $n - 1$ test particles, the measurements are exact: $g^{(n)}$ can be measured with arbitrary resolution and for specific particle configurations. Measurements based on inserting fewer test particles require a histogram in fewer variables than the full distance-histogram method, considerably reducing the statistical noise, while the normalization of the insertion measurements is straightforward in all cases. We demonstrate the application of the hierarchy to measure $g^{(2)}$ and $g^{(3)}$ in a simple homogeneous fluid, comparing the results obtained using different numbers of insertions with each other and to those of the distance-histogram method. In particular, we find that the insertion-based methods remove the need for an approximate interpolation of $g^{(3)}$ and are, therefore, particularly advantageous when examining closure approximations.

This paper is organized as follows. In Sec. II A, we derive the hierarchy and show how it can be applied to the cases of two-body and three-body distribution functions (Sec. II B). Section III describes the computational methods employed, with further details given in the Appendix. We then present the insertion results for low (Sec. IV A) and higher densities (Sec. IV B), comparing them with each other and with results from the conventional distance-histogram method. We conclude this paper in Sec. V.

II. THEORY

A. General case

The n -body density in the grand ensemble is written as¹

$$\rho^{(n)}(\mathbf{s}^n) = \frac{1}{\Xi} \sum_{N=n}^{\infty} \frac{z^N}{(N-n)!} \int d\mathbf{r}^{(N-n)} \exp[-\beta U_N(\mathbf{s}^n, \mathbf{r}^{(N-n)})], \quad (1)$$

where U_N is the potential energy of N interacting particles (including interactions with external fields), the inverse temperature $\beta = 1/k_B T$, and z is the reduced activity $\exp(\beta\mu)/\Lambda^d$, with μ being the chemical potential, Λ being the thermal wavelength, and d being the system dimensionality. The grand partition function Ξ is given by

$$\Xi = \sum_{N=0}^{\infty} \frac{z^N}{N!} \int d\mathbf{r}^N \exp[-\beta U_N(\mathbf{r}^N)].$$

We use \mathbf{r}_i to denote a coordinate that varies under integration and \mathbf{s}_i to denote a fixed position in space and adopt the notational conventions $\mathbf{r}^N \equiv \mathbf{r}_1, \dots, \mathbf{r}_N$ and $\mathbf{r}^{(N-n)} \equiv \mathbf{r}_{n+1}, \dots, \mathbf{r}_N$.¹ Inspecting Eq. (1), we see that $\rho^{(n)}$ is the ratio of the sum of all microstates with any n particles at \mathbf{s}^n to the total sum of microstates, with the microstates weighted by their probability densities in the grand ensemble. The n -body density is, therefore, the marginal probability density of having any n particles at the positions \mathbf{s}^n .

The conditional probability density $\rho^{(n-m)}(\mathbf{s}^{(n-m)}|\mathbf{s}^m)$ of having $(n-m)$ particles at $\mathbf{s}^{(n-m)}$ given that there are m particles at

\mathbf{s}^m is, therefore, given by the ratio of $\rho^{(n)}(\mathbf{s}^n)$ to $\rho^{(m)}(\mathbf{s}^m)$.^{45–47} Using Eq. (1),

$$\frac{\rho^{(n)}(\mathbf{s}^n)}{\rho^{(m)}(\mathbf{s}^m)} = \frac{\sum_{N=n}^{\infty} \frac{z^N}{(N-n)!} \int d\mathbf{r}^{(N-n)} \exp[-\beta U_N(\mathbf{s}^n, \mathbf{r}^{(N-n)})]}{\sum_{N=m}^{\infty} \frac{z^N}{(N-m)!} \int d\mathbf{r}^{(N-m)} \exp[-\beta U_N(\mathbf{s}^m, \mathbf{r}^{(N-m)})]}. \quad (2)$$

Extending the arguments of Widom,^{43,48–50} we split U_N in the numerator of Eq. (2) into a term representing all the interactions between particles at \mathbf{s}^m and $\mathbf{r}^{(N-n)}$ and the additional potential energy $\Psi(\mathbf{s}^{(n-m)}; \mathbf{s}^m, \mathbf{r}^{(N-n)})$ arising from interactions of the particles at $\mathbf{s}^{(n-m)}$ with each other and the other particles of the system. The numerator becomes

$$\begin{aligned} & \sum_{N=n}^{\infty} \frac{z^N}{(N-n)!} \int d\mathbf{r}^{(N-n)} \exp[-\beta U_{N-(n-m)}(\mathbf{s}^m, \mathbf{r}^{(N-n)})] \\ & \quad \times \exp[-\beta \Psi(\mathbf{s}^{(n-m)}; \mathbf{s}^m, \mathbf{r}^{(N-n)})] \\ & = z^{n-m} \sum_{M=m}^{\infty} \frac{z^M}{(M-m)!} \int d\mathbf{r}^{(M-m)} \exp[-\beta U_M(\mathbf{s}^m, \mathbf{r}^{(M-m)})] \\ & \quad \times \exp[-\beta \Psi(\mathbf{s}^{(n-m)}; \mathbf{s}^m, \mathbf{r}^{(M-m)})], \end{aligned} \quad (3)$$

where in the second line, we have made the substitution $M = N - (n - m)$ and renumbered $\mathbf{r}^{(N-n)}$ as $\mathbf{r}^{(M-m)}$. When this is substituted into Eq. (2), the right-hand side has the form of an ensemble average of $\exp(-\beta \Psi)$ over the states with particles at \mathbf{s}^m , multiplied by z^{n-m} . Hence,

$$\begin{aligned} \frac{\rho^{(n)}(\mathbf{s}^n)}{\rho^{(m)}(\mathbf{s}^m)} & = z^{n-m} \left\langle \exp[-\beta \Psi(\mathbf{s}^{(n-m)}; \mathbf{s}^m, \mathbf{r}^{(N-m)})] \right\rangle_{\mathbf{s}^m} \\ & = z^{n-m} P^{(n-m)}(\mathbf{s}^{(n-m)}|\mathbf{s}^m), \end{aligned} \quad (4)$$

where in the second line, we have simplified the notation by using P for the ensemble average.

Introducing the definition of the distribution functions¹

$$g^{(n)}(\mathbf{s}^n) = \frac{\rho^{(n)}(\mathbf{s}^n)}{\prod_{i=1}^n \rho^{(1)}(\mathbf{s}_i)} \quad (5)$$

and rearranging Eq. (4), we obtain

$$g^{(n)}(\mathbf{s}^n) = z^{n-m} g^{(m)}(\mathbf{s}^m) \frac{P^{(n-m)}(\mathbf{s}^{(n-m)}|\mathbf{s}^m)}{\prod_{i=m+1}^n \rho^{(1)}(\mathbf{s}_i)}. \quad (6)$$

Finally, we set $n = 1$ and $m = 0$ in Eq. (4) to obtain $\rho^{(1)}(\mathbf{s}_i) = z P^{(1)}(\mathbf{s}_i)$, which is Widom's result for inhomogeneous systems.⁵¹ Using this result in Eq. (6) yields

$$g^{(n)}(\mathbf{s}^n) = g^{(m)}(\mathbf{s}^m) \frac{P^{(n-m)}(\mathbf{s}^{(n-m)}|\mathbf{s}^m)}{\prod_{i=m+1}^n P^{(1)}(\mathbf{s}_i)}. \quad (7)$$

This hierarchy of equations is the central result of this article and allows an n -body distribution function to be expressed as a product of a lower-order m -body distribution function and a ratio of ensemble averages, which may be evaluated by test-particle insertion.

Equation (7) is entirely general, applying to the following:

- *All n -body distribution functions:* in this work, we will illustrate the result for $n = 2$ and $n = 3$, but in principle, Eq. (7) also allows for measurements of higher-order distribution functions, e.g., $n = 4$. These higher-order functions have attracted recent attention in characterizing the structure of simple fluids.^{52,53}
- *Inhomogeneous systems:* in such cases, distribution functions not only depend on the relative positions of a subset of n particles but also on the position and orientation of the subset in the overall fluid. The inhomogeneous pair distribution function is of interest in the study of fluid–fluid interfaces,⁵⁴ for example.

In both cases, the dependence of the distribution functions on an increasing number of variables makes them increasingly cumbersome to visualize, but we will demonstrate for $g^{(3)}$ that insertion-based methods offer advantages in terms of computation and resolution: these are expected to be of some use when approaching the more complicated cases too. Handling the constraints imposed by the triangle inequality is also anticipated to be challenging in cases where $n \geq 4$, but remains, in principle, possible.

B. Application to two-body and three-body distribution functions

In this section, we explicitly work out Eq. (7) for two-body and three-body distribution functions, which properly illustrates the underlying hierarchy of the solution. Substituting $n = 2$ into Eq. (7) gives

$$g^{(2)}(\mathbf{s}_1, \mathbf{s}_2) = \frac{P^{(2)}(\mathbf{s}_1, \mathbf{s}_2)}{P^{(1)}(\mathbf{s}_1)P^{(1)}(\mathbf{s}_2)}, \quad m = 0, \quad (8a)$$

$$= \frac{P^{(1)}(\mathbf{s}_2|\mathbf{s}_1)}{P^{(1)}(\mathbf{s}_2)}, \quad m = 1, \quad (8b)$$

where we have noted that $g^{(0)} = g^{(1)} = 1$. In a homogeneous system, $P^{(1)}(\mathbf{s}_i) \equiv P$ is spatially uniform, while $g^{(2)}(\mathbf{s}_1, \mathbf{s}_2)$, $P^{(2)}(\mathbf{s}_1, \mathbf{s}_2)$, and $P^{(1)}(\mathbf{s}_2|\mathbf{s}_1)$ only depend on the separation,

$$|\mathbf{s}_2 - \mathbf{s}_1| = r,$$

such that

$$g^{(2)}(r) = \frac{P^{(2)}(r)}{P^2}, \quad m = 0, \quad (9a)$$

$$= \frac{P^{(1)}(r)}{P}, \quad m = 1. \quad (9b)$$

There are, therefore, two possibilities for measuring $g^{(2)}$ by test-particle insertion. According to Eq. (9a), we first evaluate $P^{(2)}(r)$ by test-insertions of a pair of particles with separation r at random positions in the fluid; dividing by the square of P , which can be measured by one-particle insertions, then yields $g^{(2)}$. Alternatively, we can perform test-insertions of one particle at a fixed separation r from an existing particle of the fluid; this yields $P^{(1)}(r)$, which can be used in Eq. (9b). Note that to preserve the simplicity of the notation, the

conditional nature of $P^{(1)}(r)$, $P^{(2)}(r, s, t)$, and $P^{(1)}(r, s, t)$ is taken to be implicit. In contrast with the distance-histogram method, both insertion methods are exact and $g^{(2)}$ can be obtained at arbitrary resolution.⁴⁴

In a homogeneous fluid, $g^{(3)}$ is a function of the triangle formed with a particle at each vertex, which here is specified by the side lengths r , s , and t , as shown in Fig. 1. In this case, Eq. (7) gives

$$g^{(3)}(r, s, t) = \frac{P^{(3)}(r, s, t)}{P^3}, \quad m = 0, \quad (10a)$$

$$= \frac{P^{(2)}(r, s, t)}{P^2}, \quad m = 1, \quad (10b)$$

$$= \frac{P^{(1)}(r, s, t)}{P} g^{(2)}(r), \quad m = 2, \quad (10c)$$

with the corresponding insertion methods summarized in Figs. 1(a)–1(c). The first and second cases are straightforward. In the first case, we evaluate $P^{(3)}(r, s, t)$ by test-insertions of a triangle of three particles, with the triangle having a random position and orientation in the fluid [Fig. 1(a)]. In the second case, we measure $P^{(2)}(r, s, t)$ using two-particle insertions with an existing particle as the third vertex [Fig. 1(b)].

In the third case, we obtain $P^{(1)}(r, s, t)$ by test-insertions of one particle around pairs of existing particles approximately separated by one of the side lengths [Fig. 1(c)]. Note that we must also multiply by $g^{(2)}(r)$, where r is the separation of the pair of particles already in the system. Since $g^{(2)}(r)$ may also be measured by one-particle insertions, this illustrates the principle that Eq. (7) can be used to express $g^{(n)}$ as products of ensemble averages corresponding only to one-particle insertions. Although the first two cases were exact, allowing $g^{(3)}$ to be measured precisely for specific triangles, the third case only requires a histogram in the side length r . This still significantly improves the statistics compared with the full

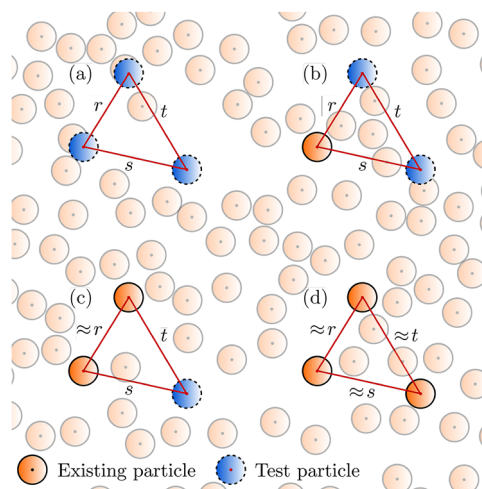


FIG. 1. Applying Eq. (7) permits $g^{(3)}(r, s, t)$ to be measured by insertion of up to three test particles (a)–(c) in addition to the conventional distance-histogram approach (d). Methods (a) and (b) are exact, while (c) and (d) require histograms in one and three variables, respectively.

distance-histogram approach. In all cases, normalization is just a straightforward division by the appropriate power of P .

III. METHODS

A. Simulation

We performed Monte Carlo simulations in the grand ensemble. The interparticle interactions were pairwise according to a Weeks–Chandler–Anderson (WCA) potential,⁵⁵

$$u_{\text{WCA}}(r) = \begin{cases} 4\epsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right] + \epsilon, & r \leq 2^{\frac{1}{6}}\sigma, \\ 0, & r > 2^{\frac{1}{6}}\sigma, \end{cases} \quad (11)$$

where r is the interparticle separation. We used a square simulation box with side length 50σ and chose the interaction parameters $\sigma = 1$, $\epsilon = 1$, and the inverse temperature $\beta = 1/k_{\text{B}}T = 1$. We set the simulation chemical potential

$$\mu' = \mu - 2k_{\text{B}}T \ln(\Lambda/\sigma)$$

such that $\mu' = 1.25$, $\mu' = 4$, or $\mu' = 8.5$. These gave rise to mean numbers of particles $\bar{N} \approx 1017$ (reduced density $\rho\sigma^2 \approx 0.41$), $\bar{N} \approx 1507$ ($\rho\sigma^2 \approx 0.60$), and $\bar{N} \approx 1936$ ($\rho\sigma^2 \approx 0.77$), respectively.

We performed particle displacement, deletion, and insertion moves in the ratio 3:1:1. For the displacements, the trial particle position was randomly chosen from a square box around the original position, with a maximum displacement in each direction of 0.6 ($\rho\sigma^2 \approx 0.41$, acceptance ratio 52%), 0.28 ($\rho\sigma^2 \approx 0.60$, 51%), and 0.19 ($\rho\sigma^2 \approx 0.77$, 42%). We initially placed ten particles at random in the box and then allowed the system to equilibrate for 10^7 moves such that the equilibrium density was attained. We subsequently took snapshots every 10^4 moves and saved them for analysis. The total number of analyzed configurations was 5×10^4 .

B. Distance-histogram method

For both $g^{(2)}$ and $g^{(3)}$ measurements, distance criteria were used to avoid considering pairs and triangles of particles with large separations or side lengths, where the short-ranged structure of the fluid has essentially decayed away. For $g^{(2)}(r)$, pairs with a separation $r \leq 10\sigma$ were found and a histogram of the separations was made. This was normalized by the volumes of the annular bins and the number of particles considered to yield $\rho^{(1)}(r|\mathbf{0})$, the conditional one-body density, given a particle at the origin $\mathbf{0}$. Finally, this was divided by ρ to give $g^{(2)}(r)$.

For $g^{(3)}$, triangles were found with at least two sides $\leq 5\sigma$; using the triangle inequality, the maximum length of the third side is then constrained to be $\leq 10\sigma$. A 3D histogram of the side lengths was made with a bin-size of 0.05σ in each variable. In principle, the symmetry of $g^{(3)}$ means each triangle can be included six times (once for each permutation of $\{r, s, t\}$), though, in practice, those triangles with the third side length $> 5\sigma$ were only included twice due to the distance criteria. Care must be taken when normalizing the histogram to only include the region of each bin, which satisfies the triangle inequalities—we used a two-dimensional version²⁴ of the procedure outlined in Ref. 34. Normalization by this volume and the number

of particles considered gave $\rho^{(2)}(r, s, t|\mathbf{0})$, which was divided by ρ^2 to give $g^{(3)}(r, s, t)$. More technical details are given in the Appendix.

C. Test-particle insertion methods

For $g^{(2)}$, the two-particle and one-particle test-insertion methods are relatively straightforward, as are the three-particle and two-particle test-insertion methods for $g^{(3)}$. In the latter case, the one-particle test-insertion method needs more care, as consideration needs to be given to the binning of the two particles chosen from the system. Details are provided in the Appendix.

IV. RESULTS AND DISCUSSION

We next test these methods using two-dimensional simulations in the grand ensemble, with the particles interacting via a WCA pair potential.⁵⁵ We performed three simulations, with reduced densities $\rho\sigma^2 \approx 0.41$, $\rho\sigma^2 \approx 0.60$, and $\rho\sigma^2 \approx 0.77$. We will first present and discuss the results for a fluid at relatively low density, illustrating the formal equivalence of the insertion methods. We will then consider higher density fluids.

A. Low density

We show the results for $g^{(2)}$ in Fig. 2(a). Agreement between the results of the one-particle insertion and distance-histogram methods is perfect, as previously found for hard disks away from contact.⁴⁴ Here, we demonstrate that agreement also extends to continuous interactions and that these results agree with those of the two-particle insertion method.

Since $g^{(3)}$ is a function of three variables, it is more difficult to visualize than $g^{(2)}$. Although the insertion methods can be used to measure $g^{(3)}$ in full, results are typically only shown for a subset of triangles: either the side lengths are chosen to be functions of a single variable, e.g., equilateral or isosceles triangles,^{23,32–35,37–39} or two particles are held at a fixed separation and $g^{(3)}$ is plotted as a function of the third.^{24–26,36} The insertion approach is efficient in both cases since $g^{(3)}$ can be measured directly for the subset of triangles and the entire $g^{(3)}$ does not need to be calculated unless it is required. Here, we opt for the first case, showing $g^{(3)}$ measured for equilateral and isosceles triangles in Figs. 2(b)–2(d). For each plot, the ordinate is chosen to facilitate a comparison with $g^{(2)}$, which is the prediction of the Kirkwood Superposition Approximation (KSA) closure,²

$$g^{(3)}(r, s, t) = g^{(2)}(r)g^{(2)}(s)g^{(2)}(t), \quad (12)$$

which assumes that the correlation of any two particles is not affected by the presence of the third.⁵⁷

For the equilateral triangles [Fig. 2(b)], agreement of the insertion results with each other and those of the distance-histogram method is essentially perfect, verifying the validity of Eq. (7) and the insertion methods described above. The two- and three-particle insertion methods are formally exact, while the one-particle insertion method has a resolution comparable to the distance-histogram measurement of $g^{(2)}$ since a histogram is only required in the separation r of the existing particles [Fig. 1(c)]. The implementation of the one-particle insertion method is highly efficient: we generally attempt fewer insertions than with the other insertion methods,

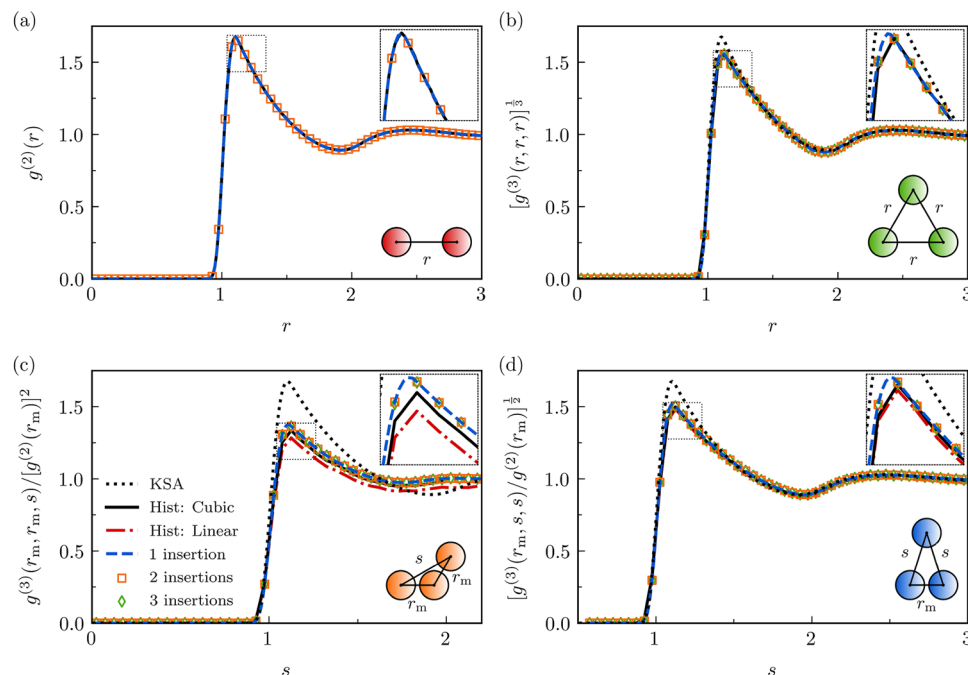


FIG. 2. Comparison of $g^{(2)}$ (a) and $g^{(3)}$ for equilateral and isosceles triangles (b)–(d) measured by the distance-histogram method and each of the insertion methods in the fluid with $\rho\sigma^2 \approx 0.41$. The insets show a zoom of each plot around the first maximum. The ordinates of (b)–(d) facilitate a comparison with $g^{(2)}$, which is the KSA prediction. In (c) and (d), the side length r_m , corresponding to the first maximum of $g^{(2)}$, was not at the center of one of the bins used in the distance-histogram method—instead, the results of linear and cubic interpolation are shown.⁵⁶ See the Appendix for more details.

with attempts only made around existing pairs of particles, which have approximately the correct separation. Although this does not allow for measurements when r is inside the core region, since there are no existing pairs with these separations, this limitation has no practical consequence—at such separations, $g^{(2)} = 0$, and so $g^{(3)} = 0$ according to Eq. (10c).⁵⁸

Figures 2(c) and 2(d) show $g^{(3)}$ measured for isosceles triangles, with either one or two side lengths fixed as $r_m = 1.105$, corresponding to the first maximum in $g^{(2)}$. Note that the range of values for the other side length s is restricted by the triangle inequality. Agreement of the insertion measurements with each other is again excellent, but these results highlight a key deficiency of the distance-histogram approach. To obtain adequate statistics, larger bins are required when measuring $g^{(3)}$ than when measuring $g^{(2)}$; consequently, r_m does not lie at the center of a $g^{(3)}$ bin, and an approximate interpolation must be performed. Although a cubic interpolation⁵⁹ outperforms a simple linear scheme, there is still a marked deviation from the insertion results, especially in Fig. 2(c) (see the Appendix for more details). By contrast, no issues with resolution arise for the two-particle and three-particle insertion methods, which are formally exact, while the one-particle method also allows for a much higher resolution than the pure distance-histogram approach. The poor resolution of the distance-histogram method is expected to be still more pronounced when measuring higher-order distribution functions, where the dependence on more variables further

compromises the resolution and makes interpolation schemes more challenging to implement.

Finally, we note that the greatest deviations from the KSA are found when all three particles are close together; this is especially clear in Fig. 2(c), where the isosceles triangles are analogous to rolling-contact configurations in hard-sphere systems^{16,42,60–62} and where the poor resolution of the distance-histogram method is also most evident. Hence, while comparisons with more sophisticated closures of the YBG hierarchy are possible,^{9,36–41} comparing with the KSA already highlights the advantages of the insertion approach in examining such approximations.

B. Higher densities

Figures 3 and 4 show the same measurements as in Fig. 2, but for the higher-density fluids with $\rho\sigma^2 \approx 0.60$ and $\rho\sigma^2 \approx 0.77$, respectively. As with the lower-density fluid, in both cases, $g^{(3)}$ is best-captured by the one-particle insertion method, which we performed at a resolution 5× that of the other measurements due to its efficient implementation and low noise. Once again, r_m did not lie at the center of one of the bins used for the distance-histogram measurement of $g^{(3)}$, and so we had to use an approximate interpolation in the isosceles cases ($r_m = 1.085$ for $\rho\sigma^2 \approx 0.60$; $r_m = 1.065$ for $\rho\sigma^2 \approx 0.77$). As with the lower-density fluid, the deviation between the distance-histogram and one-particle insertion results is greatest in panel (c), where all three particles are close together at all allowed

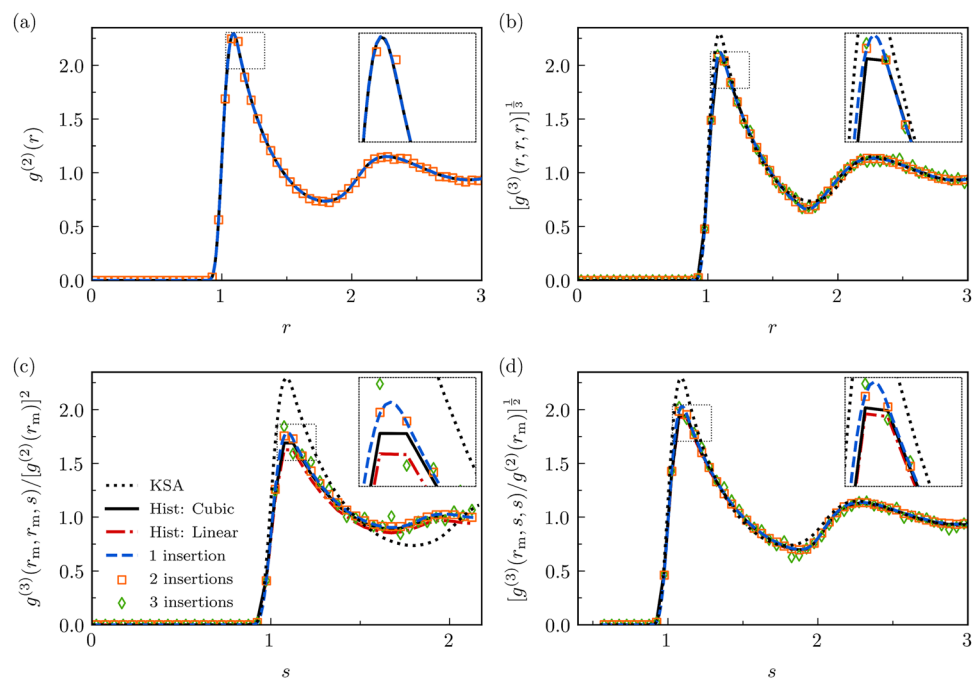


FIG. 3. The same measurements as shown in Fig. 2, but performed for the system with $\rho\sigma^2 \approx 0.60$.

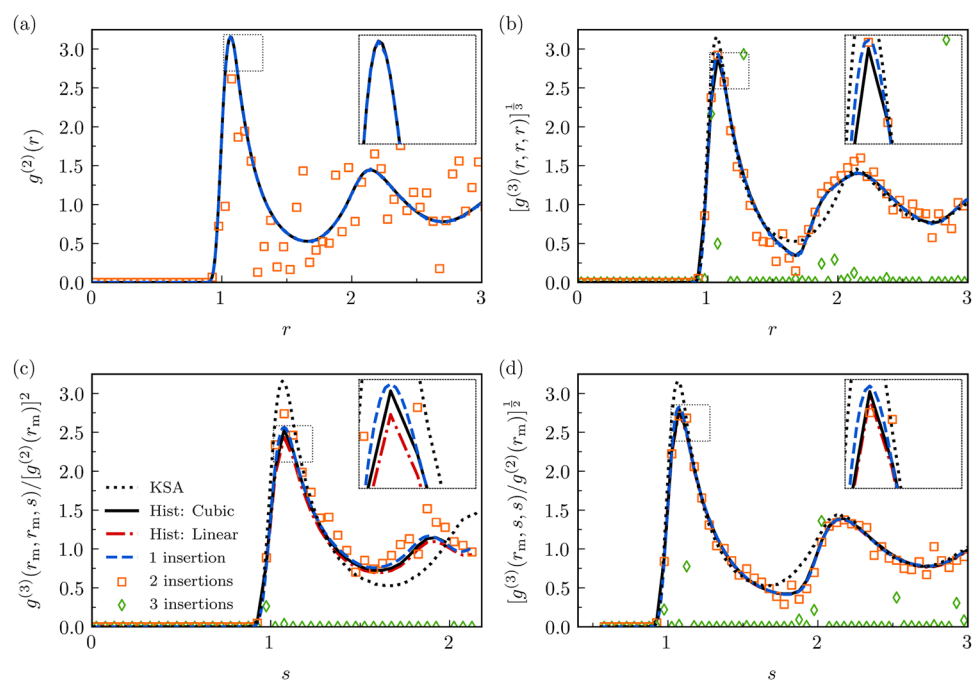


FIG. 4. The same measurements as shown in Fig. 2, but performed for the system with $\rho\sigma^2 \approx 0.77$. The three-particle insertion method fails completely, and some points are outside of the plotted range.

values of s and the deviation remains even after cubic interpolation. Note also that the Kirkwood Superposition Approximation (KSA) does not perform as well as in the lower-density fluid, reflecting the increased importance of three-body correlations at higher densities: this is clearly evident in panels (b) and (d), which now show sizeable deviations at larger separations as well as at the first maximum. Finally, the equilateral $g^{(3)}$ at $\rho\sigma^2 \approx 0.77$ (Fig. 4) shows a developing shoulder to the left of the second peak, indicating incipient crystalline order.²⁶

The increased noise of the multi-particle insertion methods is more apparent at the higher densities, especially that of the three-particle insertion method, which fails completely for $\rho\sigma^2 \approx 0.77$ (Fig. 4). The two-particle insertion method also shows increased noise for both the $g^{(2)}$ and $g^{(3)}$ measurements. The breakdown of test-particle insertion methods at high densities is well-known in the context of chemical potential measurements⁶³ and has the same origin here. In brief, it becomes too difficult to insert an additional particle: small or negative values of Ψ are encountered only rarely, despite making an important contribution to the ensemble average. This difficulty is clearly greater still when inserting multiple particles, leading to a larger error in these insertion measurements and the breakdown of the multi-particle insertion methods at lower densities than in the one-particle case.

At $\rho\sigma^2 \approx 0.77$, a small amount of noise is visible even in the one-particle case—while this may lead to the distance-histogram method being the method of choice for obtaining $g^{(2)}$, the superior resolution of the insertion route is still strongly apparent in the $g^{(3)}$ measurements (see the insets of Fig. 4). Nevertheless, this noise draws attention to the inevitable breakdown of even the one-particle insertion method, in this case at a density lower than that of the fluid–solid transition, beyond which we will have no choice but to use the distance-histogram method. We have previously shown that the one-particle insertion method for $g^{(2)}$ can be used to accurately measure contact values in hard-disk systems up to at least packing fraction $\phi \approx 0.65$,⁴⁴ and we might reasonably expect the one-particle method for $g^{(3)}$ to work up to similar densities. Taking σ of the WCA interaction in Eq. (11) to be an approximation to σ_{HD} , the hard-disk diameter $\phi = \pi\rho\sigma_{\text{HD}}^2/4 \approx 0.65$ corresponds to a WCA system with $\rho\sigma^2 \approx 0.83$. We also note again that simulations^{9,23} and colloidal experiments^{24–26} have found that $g^{(3)}$ shows signatures of crystal-like order deep inside the fluid phase: thus, while for hard sphere-like fluids the breakdown at densities close to the fluid–solid transition is an important qualification of insertion-based methods, we still expect to find interesting behavior at densities amenable to study.

V. CONCLUSION

We have derived a hierarchy of insertion-based methods for measuring many-body distribution functions and demonstrated their application in measuring $g^{(2)}$ and $g^{(3)}$. The methods address drawbacks of the conventional distance-histogram approach, offering improved resolution and a more straightforward normalization. They are expected to be particularly advantageous in inhomogeneous systems and when scrutinizing various closure approximations used in integral equation theories. As well as being effective in measuring $g^{(3)}$, they will facilitate investigations into expressions

for higher-order distribution functions, such as $g^{(4)}$, which can be used to close the YBG hierarchy at higher levels.^{9,13–16}

Which insertion method should be used? The one-particle insertion method is particularly efficient and offers excellent resolution when measuring $g^{(3)}$, with the need for a histogram in only one variable overcoming the poor resolution of the distance-histogram method. On the other hand, the remaining histogram may lead to increased noise in lower-density systems, or when measuring higher-order $g^{(n)}$, where a histogram in more than one variable is again required. Although the insertion method with insertion of n or $n - 1$ particles is formally exact, multi-particle insertion methods exacerbate the well-documented breakdown of the test-particle insertion approach at high densities.⁶³ Approaches based on one-particle insertions, although approximate for $g^{(3)}$ and higher-order distribution functions, are expected to work well except in very high-density fluids and solids.⁴⁴ The optimum number of insertions, therefore, depends on the order n of the distribution function, the state of the fluid, and the application in mind.

ACKNOWLEDGMENTS

The authors would like to thank Roel Dullens for critically reading the manuscript. A.E.S. gratefully acknowledges the financial support from the University of Oxford Clarendon Fund.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Adam Edward Stones: Conceptualization (equal); Methodology (lead); Software (lead); Writing – original draft (lead); Writing – review & editing (equal). **Dirk G. A. L. Aarts:** Conceptualization (equal); Supervision (equal); Writing – review & editing (equal).

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

APPENDIX: CALCULATION DETAILS

In this appendix, we provide relevant details used to calculate the distribution functions.

1. Distance-histogram method

Figure 2(a) shows the measured $g^{(2)}(r)$ at values of r corresponding to the center of each bin. In Fig. 2(b), we use $g^{(3)}$ bins with centers corresponding to equilateral configurations. As noted in the main text, finding $g^{(3)}$ is more challenging for the triangles used in Figs. 2(c) and 2(d) since r_m , the separation at the primary maximum of the measured $g^{(2)}$, does not coincide with the center of the bins used when measuring $g^{(3)}$, which have to be larger to provide adequate statistics. We, therefore, have to interpolate the measured

function to provide values for comparison with the insertion methods and the KSA closure. This is particularly important in the case of Fig. 2(c), where the three particles are always close together and $g^{(3)}$ varies significantly on the length-scale of the bin size—a linear interpolation is clearly inadequate. While the cubic interpolation⁵⁹ provides somewhat improved agreement with the insertion results, it comes with the disadvantage that it does not work when s is too close to zero or the limits given by the triangle inequality since in these cases, the scheme requires values for bins where $g^{(3)}(r, s, t)$ is not defined.

Note that $r_m = 1.105$ at $\rho\sigma^2 \approx 0.41$, $r_m = 1.085$ at $\rho\sigma^2 \approx 0.60$, and $r_m = 1.065$ at $\rho\sigma^2 \approx 0.77$.

2. $g^{(2)}$: Two-particle test-insertions

We measured $g^{(2)}$ at separations corresponding to the centers of the bins used in the distance-histogram measurement, with a maximum r of 5.025σ . To measure $P^{(2)}(r)$, 1000 pairs of particles were inserted per configuration for each separation, at random locations and orientations. The inserted pairs were also separately considered as one-particle insertions to measure P , and Eq. (9a) was used to calculate $g^{(2)}(r)$.

3. $g^{(2)}$: One-particle test-insertions

The number of separations was $5\times$ that used in the two-particle insertion case since this method provides better statistics. To measure $P^{(1)}(r)$, one insertion per particle was attempted for each separation in all configurations. In each case, the angle of the inserted particle relative to the existing particle was chosen at random from a uniform distribution. To measure P , a separate measurement was performed where 10^4 test particles per configuration were inserted at random points in the fluid, and Eq. (9b) was used to calculate $g^{(2)}(r)$.

4. $g^{(3)}$: Three-particle test-insertions

Separations (r or s) were chosen to correspond to the bin centers used in the distance-histogram measurement. To measure $P^{(3)}$, 1000 test-particle triangles were inserted for each separation in the configurations specified in the cartoons of Figs. 2(b)–2(d), at random locations and orientations. A separate measurement of P was performed using 10^4 test particles, and Eq. (10a) was used to calculate $g^{(3)}$.

5. $g^{(3)}$: Two-particle test-insertions

The same separations were used as for the three-particle insertion case. For each separation, two particles were inserted around each particle of the fluid in all configurations, to form a triangle with the appropriate side lengths [see Fig. 1(b)]. The triangles were constructed with a random orientation, and the insertions were used to measure $P^{(2)}$. A separate measurement of P was performed using 10^4 test particles, and Eq. (10b) was used to calculate $g^{(3)}$.

6. $g^{(3)}$: One-particle test-insertions

In the equilateral case, all pairs of particles with $r \leq 5\sigma$ are found, and a test-particle insertion was performed at the two points, which complete an equilateral triangle. A histogram was used to

measure $P^{(1)}(r, r, r)$ based on binning each measurement with a bin width of 0.01σ . Note that the measurement was performed with a resolution $5\times$ greater than in the two-insertion and three-insertion cases. In the isosceles cases, where one or two of the side lengths are fixed at, e.g., $r_m = 1.105$ (for $\rho\sigma^2 \approx 0.41$), pairs of particles were found with a separation r between, e.g., 1.100 and 1.110, and four [Fig. 2(c)] or two [Fig. 2(d)] test-particle insertions were attempted around each pair to measure $P^{(1)}$. Note that in Fig. 2(c), the second side length corresponding to r_m was taken to be consistent with the exact separation of the pair, rather than as, e.g., 1.105. In both cases, the triangle inequality limits the possible values of s to $s \leq 2r_m$ and $s \geq r_m/2$, respectively. In all three cases, a separate measurement of P was performed using 10^4 test particles, and this was used with the value for $g^{(2)}$ corresponding to the separation of the existing pair to calculate $g^{(3)}$ in accordance with Eq. (10c). Here, the one-particle insertion result for $g^{(2)}(r)$ was used for this. While pairs of particles with side lengths close to s can also be considered, we do not include them in this analysis. Note that $r_m = 1.085$ at $\rho\sigma^2 \approx 0.60$, and pairs with separation r between 1.080 and 1.090 were found. For $\rho\sigma^2 \approx 0.77$, $r_m = 1.065$ and pairs with separation r between 1.060 and 1.070 were found.

REFERENCES

- J.-P. Hansen and I. R. McDonald, "Theory of simple liquids," in *With Applications to Soft Matter*, 4th ed. (Academic Press, 2013), ISBN: 9780123870339.
- J. A. Barker and D. Henderson, "What is 'liquid'? Understanding the states of matter," *Rev. Mod. Phys.* **48**, 587–671 (1976).
- M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Oxford University Press, 2017).
- B. M. Axilrod and E. Teller, "Interaction of the van der Waals type between three atoms," *J. Chem. Phys.* **11**, 299–300 (1943).
- J. A. Barker, D. Henderson, and W. R. Smith, "Three-body forces in dense systems," *Phys. Rev. Lett.* **21**, 134–136 (1968).
- J. Barker, D. Henderson, and W. Smith, "Pair and triplet interactions in argon," *Mol. Phys.* **17**, 579–592 (1969).
- J. Barker, R. Fisher, and R. Watts, "Liquid argon: Monte Carlo and molecular dynamics calculations," *Mol. Phys.* **21**, 657–673 (1971).
- R. D. Present, "Non-additive interactions," *Contemp. Phys.* **12**, 595–602 (1971).
- H. J. Raveché and R. D. Mountain, "Triplet correlations," in *Progress in Liquid Physics*, edited by C. A. Croxton (Wiley, 1978).
- J. G. Kirkwood, "Statistical mechanics of fluid mixtures," *J. Chem. Phys.* **3**, 300–313 (1935).
- A. D. J. Haymet, S. A. Rice, and W. G. Madden, "Triplet correlations in the Lennard-Jones fluid," *J. Chem. Phys.* **75**, 4696–4706 (1981).
- D. Henderson, "Structure of the triplet distribution function," *J. Chem. Phys.* **46**, 4306–4310 (1967).
- Y.-T. Lee, F. H. Ree, and T. Ree, "Distribution function of classical fluids of hard spheres. I," *J. Chem. Phys.* **48**, 3506–3516 (1968).
- F. H. Ree, Y.-T. Lee, and T. Ree, "Distribution function of classical fluids of hard spheres. II," *J. Chem. Phys.* **55**, 234–245 (1971).
- Y. Ueharab, T. Ree, and F. H. Ree, "Radial distribution function for hard disks from the BGY2 theory," *J. Chem. Phys.* **70**, 1876–1883 (1979).
- Y. Ueharab, Y.-T. Lee, T. Ree, and F. H. Ree, "Triplet distribution functions for hard spheres and hard disks," *J. Chem. Phys.* **70**, 1884–1893 (1979).
- A. D. J. Haymet, S. A. Rice, and W. G. Madden, "An accurate integral equation for the pair and triplet distribution functions of a simple liquid," *J. Chem. Phys.* **74**, 3033–3041 (1981).
- M. P. Taylor and J. E. G. Lipson, "On the Born–Green–Yvon equation and triplet distributions for hard spheres," *J. Chem. Phys.* **97**, 4301–4308 (1992).

- ¹⁹P. Schofield, "Wavelength-dependent fluctuations in classical fluids: I. The long wavelength limit," *Proc. Phys. Soc.* **88**, 149–170 (1966).
- ²⁰R. E. Nettleton and M. S. Green, "Expression in terms of molecular distribution functions for the entropy density in an infinite system," *J. Chem. Phys.* **29**, 1365–1370 (1958).
- ²¹H. J. Raveché, "Entropy and molecular correlation functions in open systems. I. Derivation," *J. Chem. Phys.* **55**, 2242–2250 (1971).
- ²²A. Baranyai and D. J. Evans, "Direct entropy calculation from computer simulation of liquids," *Phys. Rev. A* **40**, 3817–3822 (1989).
- ²³H. J. Raveché, R. D. Mountain, and W. B. Streett, "Freezing and melting properties of the Lennard-Jones system," *J. Chem. Phys.* **61**, 1970–1984 (1974).
- ²⁴C. Ruß, K. Zahn, and H. H. V. Grünberg, "Triplet correlations in two-dimensional colloidal model liquids," *J. Phys.: Condens. Matter* **15**, S3509–S3522 (2003).
- ²⁵K. Zahn, G. Maret, C. Ruß, and H. H. von Grünberg, "Three-particle correlations in simple liquids," *Phys. Rev. Lett.* **91**, 115502 (2003).
- ²⁶H. M. Ho, B. Lin, and S. A. Rice, "Three-particle correlation functions of quasi-two-dimensional one-component and binary colloid suspensions," *J. Chem. Phys.* **125**, 184715 (2006).
- ²⁷F. C. Frank, "Supercooling of liquids," *Proc. R. Soc. A* **215**, 43–46 (1952).
- ²⁸P. J. Steinhart, D. R. Nelson, and M. Ronchetti, "Bond-orientational order in liquids and glasses," *Phys. Rev. B* **28**, 784–805 (1983).
- ²⁹F. Spaepen, "Five-fold symmetry in liquids," *Nature* **408**, 781–782 (2000).
- ³⁰M. Leocmach and H. Tanaka, "Roles of icosahedral and crystal-like order in the hard spheres glass transition," *Nat. Commun.* **3**, 974 (2012).
- ³¹J. Taffs and C. Patrick Royall, "The role of fivefold symmetry in suppressing crystallization," *Nat. Commun.* **7**, 13225 (2016).
- ³²B. J. Alder, "Triplet correlations in hard spheres," *Phys. Rev. Lett.* **12**, 317–319 (1964).
- ³³A. Rahman, "Triplet correlations in liquids," *Phys. Rev. Lett.* **12**, 575–577 (1964).
- ³⁴J. A. Krumhansl and S. Wang, "Triplet correlation in liquid argon by Monte Carlo method: Low densities," *J. Chem. Phys.* **56**, 2034–2041 (1972).
- ³⁵S. Wang and J. A. Krumhansl, "Superposition assumption. II. High density fluid argon," *J. Chem. Phys.* **56**, 4287–4290 (1972).
- ³⁶H. J. Raveché, R. D. Mountain, and W. B. Streett, "Three atom correlations in the Lennard-Jones fluid," *J. Chem. Phys.* **57**, 4999–5006 (1972).
- ³⁷R. Block and W. Schommers, "Triplet correlations in disordered systems: A study for liquid rubidium," *J. Phys. C: Solid State Phys.* **8**, 1997–2002 (1975).
- ³⁸M. Tanaka and Y. Fukui, "Simulation of the three-particle distribution function in a long-range oscillatory potential liquid," *Prog. Theor. Phys.* **53**, 1547–1565 (1975).
- ³⁹W. J. McNeil, W. G. Madden, A. D. J. Haymet, and S. A. Rice, "Triplet correlation functions in the Lennard-Jones fluid: Tests against molecular dynamics simulations," *J. Chem. Phys.* **78**, 388–398 (1983).
- ⁴⁰H. J. Raveché and R. D. Mountain, "Three atom correlations in liquid neon," *J. Chem. Phys.* **57**, 3987–3992 (1972).
- ⁴¹M. C. Abramo and M. P. Tosi, "Triplet correlations in liquid argon," *Lett. Nuovo Cimento* **5**, 1044–1047 (1972).
- ⁴²P. Attard, "An improved kirkwood superposition approximation for three atoms in rolling contact," *Mol. Phys.* **74**, 547–552 (1991).
- ⁴³B. Widom, "Some topics in the theory of fluids," *J. Chem. Phys.* **39**, 2808 (1963).
- ⁴⁴A. E. Stones, R. P. A. Dullens, and D. G. A. L. Aarts, "Communication: Contact values of pair distribution functions in colloidal hard disks by test-particle insertion," *J. Chem. Phys.* **148**, 241102 (2018).
- ⁴⁵J. K. Percus, "Approximation methods in classical statistical mechanics," *Phys. Rev. Lett.* **8**, 462–463 (1962).
- ⁴⁶J. K. Percus, "The pair distribution function in classical statistical mechanics," in *Equilibrium Theory of Classical Fluids*, edited by H. L. Frisch and J. L. Lebowitz (W A Benjamin, Inc., 1964).
- ⁴⁷M. Puoskari, "Exact self-consistent integral equations for the distribution functions of classical fluids," *Phys. Chem. Liq.* **39**, 201–225 (2001).
- ⁴⁸J. Rowlinson, "Intermolecular potentials that are functions of thermodynamic variables," *Mol. Phys.* **52**, 567–572 (1984).
- ⁴⁹L. L. Lee, "A potential distribution approach to fused heterochain molecules. I. Mixtures of hard dumbbells and spheres," *J. Chem. Phys.* **103**, 4221–4233 (1995).
- ⁵⁰J. F. Robinson, F. Turci, R. Roth, and C. P. Royall, "Morphometric approach to many-body correlations in hard spheres," *Phys. Rev. Lett.* **122**, 068004 (2019).
- ⁵¹B. Widom, "Structure of interfaces from uniformity of the chemical potential," *J. Stat. Phys.* **19**, 563–574 (1978).
- ⁵²Z. Zhang and W. Kob, "Revealing the three-dimensional structure of liquids using four-point correlation functions," *Proc. Natl. Acad. Sci. U. S. A.* **117**, 14032–14037 (2020).
- ⁵³I. Pihlajamäe, C. C. L. Laudicina, C. Luo, and L. M. C. Janssen, "Emergent structural correlations in dense liquids," *PNAS Nexus* **2**, pgad184 (2023).
- ⁵⁴R. Evans, "The nature of the liquid-vapour interface and other topics in the statistical mechanics of non-uniform, classical fluids," *Adv. Phys.* **28**, 143–200 (1979).
- ⁵⁵J. D. Weeks, D. Chandler, and H. C. Andersen, "Role of repulsive forces in determining the equilibrium structure of simple liquids," *J. Chem. Phys.* **54**, 5237–5247 (1971).
- ⁵⁶Note that no interpolation was required in the equilateral case [Fig. 2(b)] since the side lengths do correspond with the bin centers.
- ⁵⁷P. Attard, "Spherically inhomogeneous fluids. I. Percus-Yevick hard spheres: Osmotic coefficients and triplet correlations," *J. Chem. Phys.* **91**, 3072–3082 (1989).
- ⁵⁸T. Steinbrecher, I. S. Joung, and D. A. Case, "For continuous pair potentials such as that used here, $g(2)$ in the core region is actually finite but extremely small, and so is measured as zero in practice," *J. Comput. Chem.* **2**(15), 3253–3263 (2011).
- ⁵⁹P. Walker, U. Krohn, and D. Carty, "ARBTools: A tricubic spline interpolator for three-dimensional scalar or vector fields," *J. Open Res. Software* **7**, 12 (2019).
- ⁶⁰P. Attard and G. Stell, "Three-particle correlations in a hard-sphere fluid," *Chem. Phys. Lett.* **189**, 128–132 (1992).
- ⁶¹E. A. Müller and K. E. Gubbins, "Triplet correlation function for hard sphere systems," *Mol. Phys.* **80**, 91–101 (1993).
- ⁶²Y. Kalyuzhnyi, J. Škvára, and I. Nezbeda, "Analytic results for the three- and four-particle correlation functions of the fluid of hard disks," *J. Chem. Phys.* **150**, 034502 (2019).
- ⁶³D. A. Kofke and P. T. Cummings, "Quantitative comparison and optimization of methods for evaluating the chemical potential by molecular simulation," *Mol. Phys.* **92**, 973–996 (1997).